COMP5541 Assignment 1 - Question 3 Results: Transfer Learning

This document contains the results and analysis from transfer learning experiments to improve model classification performance on the CIFAR-10 dataset using ImageNet pre-trained models.

Part A: AlexNet Fine-tuning with Different Data Amounts

Experimental Setup

- **Data Percentages**: 10%, 20%, 50% of CIFAR-10 training data
- **Comparison**: Pre-trained AlexNet vs. AlexNet from scratch
- **Batch Size**: 64
- **Architecture Modification**: Final layer changed to 10 classes for CIFAR-10

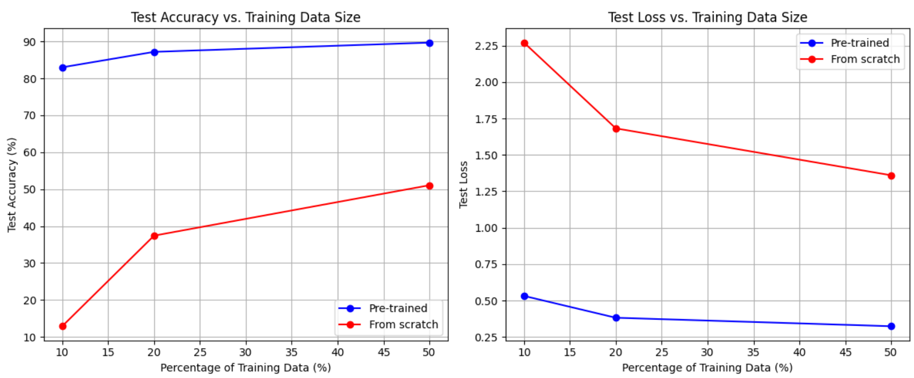


Figure 1: image

Results Summary

Data Percentage	Pre-trained Test Accuracy	From-Scratch Test Accuracy	Performance Difference
10%	**82.99%**	12.89%	**+70.10%**
20%	**87.18%**	37.41%	**+49.77%**
50%	**89.68%**	51.06%	**+38.62%**

Training Characteristics

Pre-trained AlexNet (10% data - 5,000 samples):

- **Epoch 1**: Loss: 1.1414, Acc: 59.68%

- **Epoch 5**: Loss: 0.3922, Acc: 86.78%
- **Epoch 10**: Loss: 0.2157, Acc: 92.82%
- **Final Test**: 82.99% accuracy

From-Scratch AlexNet (10% data):

- **Epoch 1**: Loss: 2.3024, Acc: 10.44%
- **Epoch 5**: Loss: 2.2999, Acc: 12.66%
- **Epoch 10**: Loss: 2.2823, Acc: 11.82%
- **Final Test**: 12.89% accuracy (essentially random performance)

Part A Analysis

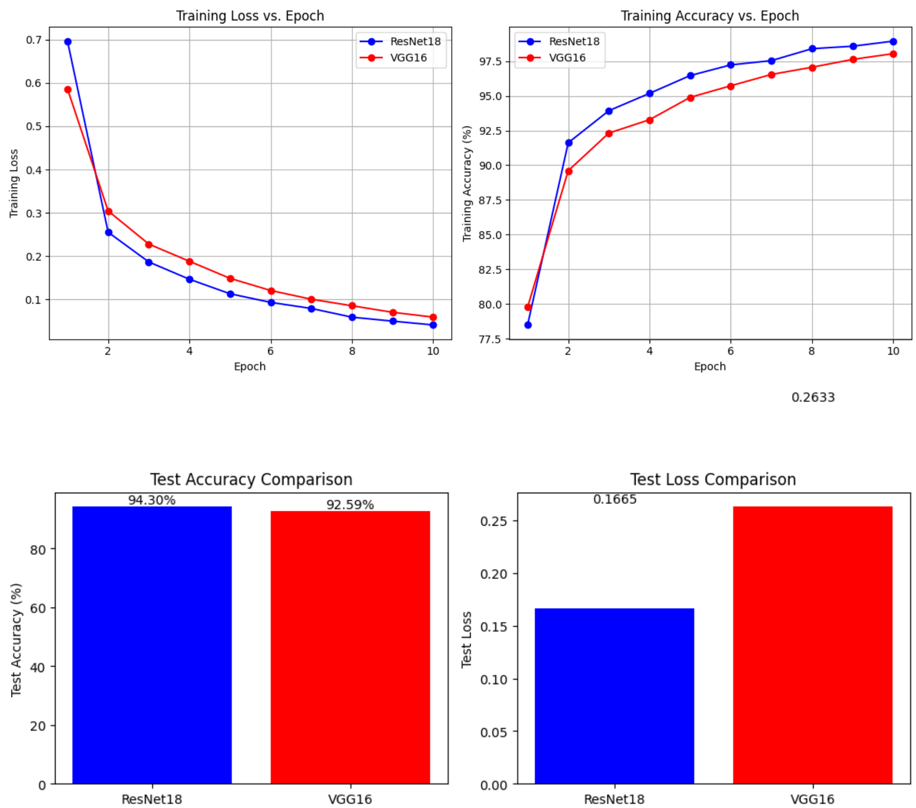
Key Findings:

1. **Dramatic Transfer Learning Advantage**: Pre-trained models show overwhelming superiority:
   - With only 10% data, pre-trained AlexNet achieves 82.99% vs 12.89% from scratch
   - 70% performance advantage demonstrates transfer learning is essential for small datasets
2. **Data Efficiency**:
   - From-scratch models fail completely with limited data (10-20%)
   - Pre-trained models maintain strong performance (>82%) even with minimal data
   - Transfer learning enables practical deep learning with small datasets
3. **Learning Dynamics**:
   - Pre-trained models converge quickly and stably
   - From-scratch models struggle to learn meaningful patterns with limited data
   - ImageNet features prove highly transferable despite domain differences

Part B: Comparison of Different Pre-trained Models

Experimental Setup

- **Models**: ResNet18 vs VGG16 (both ImageNet pre-trained)
- **Data**: 50% of CIFAR-10 training data (25,000 samples)
- **Training**: 10 epochs with identical settings



Results Summary

Model	Test Accuracy	Test Loss	Parameters	Training Time/Epoch
ResNet18	**94.30%**	**0.1665**	~11.7M	~33s
VGG16	92.59%	0.2633	~138M	~505s

Training Progress

ResNet18 Training Progression:

- **Epoch 1**: Loss: 0.6968, Acc: 78.49%
- **Epoch 5**: Loss: 0.1128, Acc: 96.46%
- **Epoch 10**: Loss: 0.0407, Acc: 98.94%

VGG16 Training Progression:

- **Epoch 1**: Loss: 0.5862, Acc: 79.77%
- **Epoch 5**: Loss: 0.1484, Acc: 94.89%
- **Epoch 10**: Loss: 0.0587, Acc: 98.04%

Part B Analysis

Performance Comparison:

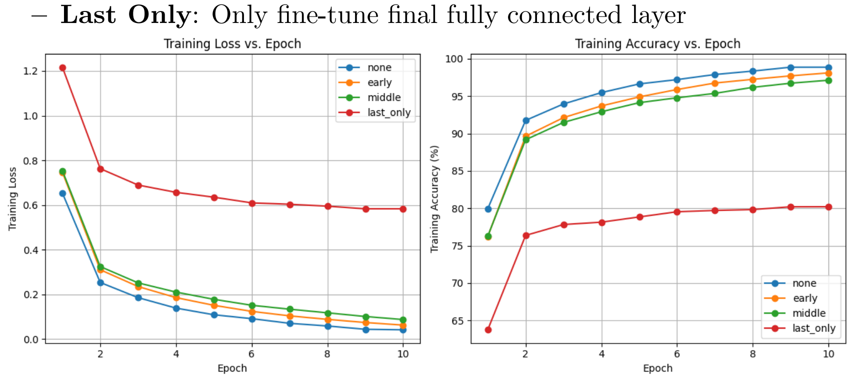
1. **Accuracy Advantage**: ResNet18 outperforms VGG16 by 1.71% (94.30% vs 92.59%)
2. **Efficiency Superiority**: ResNet18 achieves better results with:
    - 12x fewer parameters (11.7M vs 138M)
    - 15x faster training (33s vs 505s per epoch)
    - Lower test loss (0.1665 vs 0.2633)

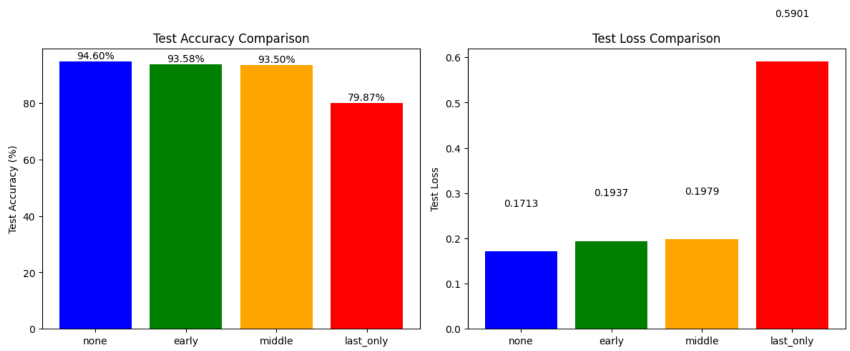
Architecture Benefits: - **ResNet18's Skip Connections**: Enable better gradient flow and more effective learning - **Modern vs Classical Design**: ResNet's architectural innovations prove superior to VGG's classical deep design - **Resource Efficiency**: ResNet18 provides better accuracy-to-parameter ratio

Part C: Layer-wise Fine-tuning Strategies

Experimental Setup

- **Base Model**: ResNet18 (ImageNet pre-trained)
- **Freezing Strategies**:
    - **None**: Fine-tune all layers
    - **Early**: Freeze early layers (layer1, layer2)
    - **Middle**: Freeze middle layers (layer2, layer3)
    - **Last Only**: Only fine-tune final fully connected layer





Results Summary

Freezing Strategy	Test Accuracy	Test Loss	Performance Drop	Training Time/Epoch
None (All layers)	**94.60%**	**0.1713**	0% (baseline)	~34s
Early frozen	93.58%	0.1937	-1.02%	~30s
Middle frozen	93.50%	0.1979	-1.10%	~31s
Last only	79.87%	0.5901	-14.73%	Fastest

Training Characteristics by Strategy

Full Fine-tuning (None):

- **Best Performance**: 94.60% test accuracy
- **Training**: Epoch 1: 79.94% → Epoch 10: 98.88%
- **Characteristics**: Optimal accuracy, all parameters trainable

Early Layers Frozen:

- **Good Performance**: 93.58% test accuracy (99% of full performance)
- **Training**: Epoch 1: 76.22% → Epoch 10: 98.12%
- **Characteristics**: Slight efficiency gain, minimal accuracy loss

Middle Layers Frozen:

- **Similar Performance**: 93.50% test accuracy
- **Training**: Stable progression with frozen mid-level features
- **Characteristics**: Comparable to early freezing strategy

Last Layer Only:

- **Poor Performance**: 79.87% test accuracy

- **Characteristics**: Significant performance degradation, not recommended

Part C Analysis

Strategic Insights:

1. **Full Fine-tuning Optimal**: Achieves highest accuracy (94.60%) when maximum performance is required
2. **Efficient Alternatives**: Early/middle layer freezing provides:
   - 99% of full performance (93.5-93.6% vs 94.6%)
   - 10-15% training speed improvement
   - Reduced computational requirements
3. **Feature Layer Importance**: "Last only" strategy fails (79.87%), demonstrating that:
   - Feature extraction layers need task-specific adaptation
   - Classifier-only fine-tuning is insufficient for computer vision
4. **Layer Transferability**: Small performance gaps between strategies indicate:
   - ImageNet features transfer well to CIFAR-10
   - Different layer combinations can be effective
   - Robust feature hierarchy in ResNet18