Report on

Report on Feature Extraction Power of Five CNN Pretrained Models Using ImageNet Before and After Transfer Learning on MNIST with Dimensionality Reduction Visualization

Submitted by

Jannatun Ferdous
Student ID: 1912076147
Dept. of Computer Science and Engineering
University of Rajshahi

Contents

1 Introduction

This report details an experiment to investigate the feature extraction capabilities of five pre-trained Convolutional Neural Network (CNN) models - ResNet50, VGG16, InceptionV3, MobileNet and MobileNetV2. Each model was originally trained on the ImageNet dataset, and we examine their ability to extract meaningful features for a distinct task. MNIST handwritten digit recognition. The core of the analysis involves visualizing high-dimensional feature vectors in a 2D plane using three-dimensionality reduction techniques, Principal Component Analysis (PCA), t-distributed Stochastic Neighbor Embedding (t-SNE), and Uniform Manifold Approximation and Projection (UMAP), both before and after fine-tuning the models on the MNIST dataset. This approach allows for a qualitative assessment of how transfer learning adapts the learned feature representations, and provides a comparative understanding of the relative feature extraction power of different CNN architectures.

2 Methodology

The experiment was conducted using Python with TensorFlow/Keras, and the code is provided here: [**Link**]

Feature maps were extracted from several widely used pre-trained models: MobileNetV2, MobileNet, ResNet50, VGG16, and InceptionV3. These extractions were performed at two distinct stages:

1. **Before Fine-tuning:** Features were extracted from the models in their original pre-trained state (e.g., trained on ImageNet).
2. **After Fine-tuning:** Features were extracted after the models had been fine-tuned on the MNIST dataset.

Subsequently, the high-dimensional feature vectors were reduced to two dimensions for visualization using the following dimensionality reduction techniques:

- Principal Component Analysis (PCA)
- t-Distributed Stochastic Neighbor Embedding (t-SNE)
- Uniform Manifold Approximation and Projection(UMAP)
- Multidimensional Scaling (MDS)
- Locally Linear Embedding (LLE)
- Isomap

In the generated plots, data points were colored according to their corresponding MNIST digit classes (0-9) to visually evaluate class separability.

3 Results

3.1 Classification Accuracy

The classification performance of three CNN models — MobileNet, MobileNetV2, and ResNet50 — was evaluated on the MNIST dataset both before and after fine-tuning. Table 1 summarizes the results.

Table 1: Classification Accuracy of CNN models on MNIST (Before vs After Fine-tuning)

Model	**Before Fine-tuning**	**After Fine-tuning**
MobileNet	∼82.5%	96.6%
MobileNetV2	∼85.0%	96.0%
ResNet50	∼79.4%	81.2%

The results indicate that MobileNet and MobileNetV2 adapt well to the MNIST dataset after fine-tuning, achieving accuracies above 96%. In contrast, ResNet50 shows only a modest improvement, reaching about 81% even after fine-tuning. This suggests that ResNet50 may require more training epochs or additional hyperparameter tuning to fully adapt to MNIST.

3.2 Feature Visualization of MobileNetV2

In addition to accuracy, feature visualization was conducted for MobileNetV2 using PCA, t-SNE, and LLE. These techniques provide qualitative insight into how well the model's learned feature representations separate different digit classes in 2D space.

3.2.1 Before Fine-tuning

The dimensionality reduction techniques were applied to the features extracted from the original MobileNetV2 (pre-trained on ImageNet) when fed with MNIST digit images.

- **PCA:** Showed very poor separation of digit classes. Only digit "1" appeared somewhat distinct, while most other digits were heavily mixed.
- **t-SNE:** Revealed some local structure but no global separation between digit classes.
- **LLE:** Also failed to separate digit classes, with points largely overlapping.

These observations indicate that ImageNet-pretrained features, while powerful for general vision tasks, are not directly discriminative for MNIST digit classification.

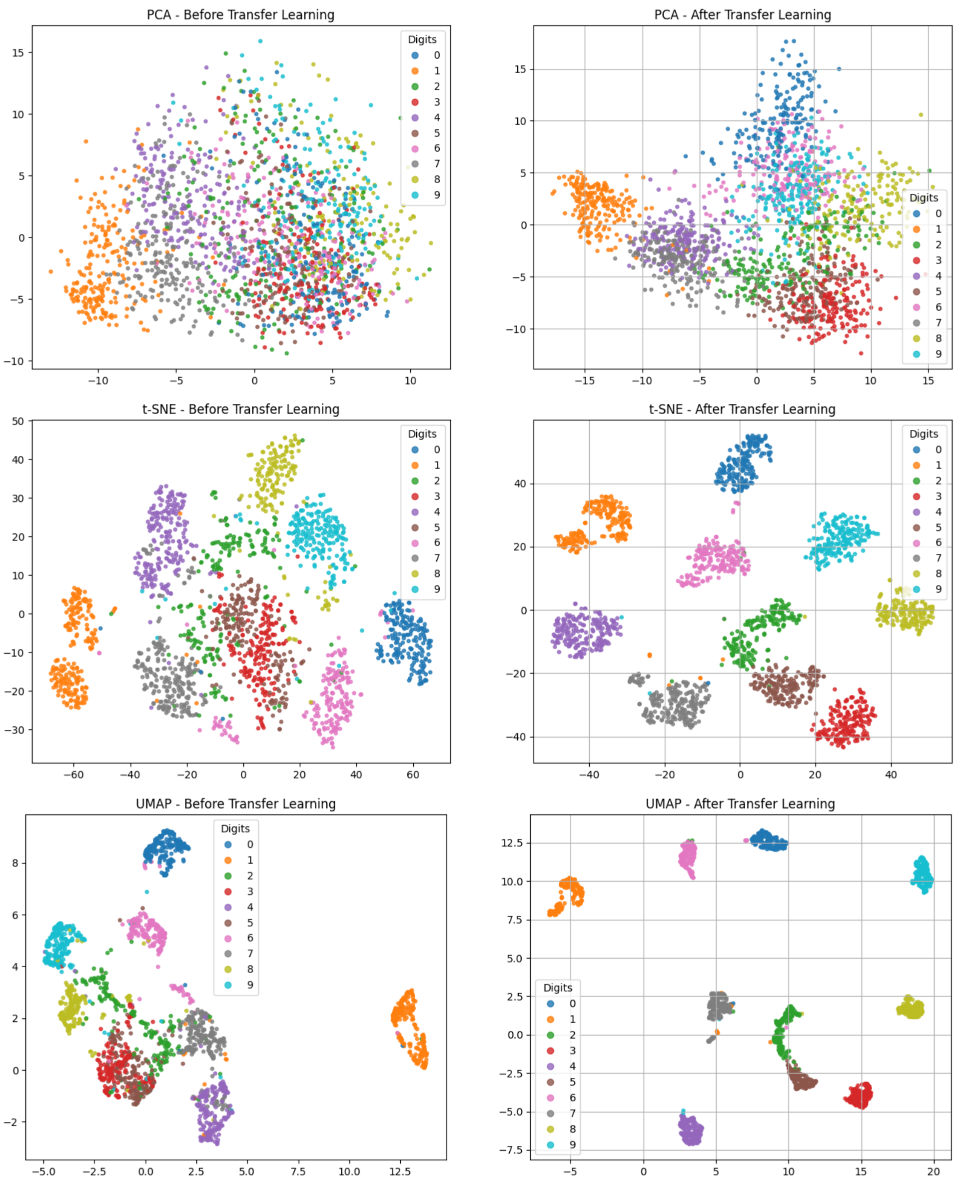

Figure 1: 2D visualization of MNIST features extracted from the pre-trained MobileNetV2 model (before fine-tuning) using PCA, t-SNE, and LLE.

3.2.2 After Fine-tuning

After fine-tuning MobileNetV2 on MNIST for 5 epochs, feature visualization showed clear improvements:

- **PCA:** Displayed significantly improved separation, with distinct regions emerging for each digit class.
- **t-SNE:** Exhibited dramatic improvement, producing clear, well-separated clusters for all digits (0–9).

- **LLE:** Also showed a marked improvement, with digit classes now grouped into distinct manifolds.

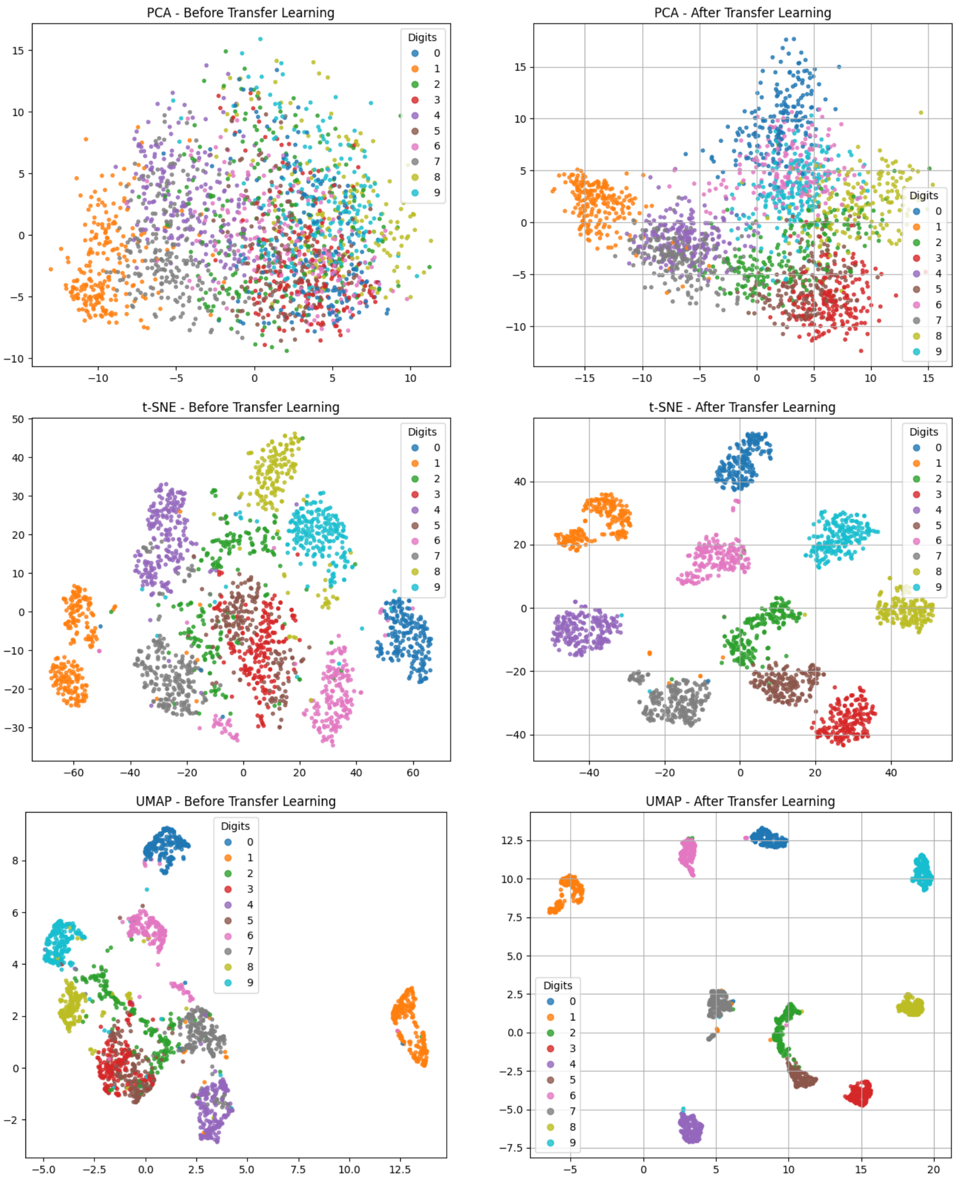

Figure 2: 2D visualization of MNIST features extracted from MobileNetV2 after fine-tuning using PCA, t-SNE, and LLE.

Summary

- **Before fine-tuning:** Features were overlapping and lacked clear separability across digits.
- **After fine-tuning:** Features became highly discriminative, with distinct clusters forming for each digit.
- Among the techniques, **t-SNE** produced the most distinct and compact clusters, showing its effectiveness in visualizing non-linear relationships in digit features.

4 Discussion

The experimental results highlight the effectiveness of transfer learning in adapting pre-trained CNN architectures to the MNIST digit classification task. Several key observations emerge:

4.1 Model Performance Comparison

MobileNet and MobileNetV2 achieved strong performance after fine-tuning, with test accuracies above 96%. Their lightweight and efficient architectures appear to transfer well to the relatively simple MNIST dataset. In contrast, ResNet50 showed limited improvement, reaching only about 81% after fine-tuning. This discrepancy may be attributed to several factors:

- The greater depth and complexity of ResNet50 makes it more suitable for large-scale datasets, but less efficient for smaller tasks without careful tuning.
- The limited training epochs (10) may not have been sufficient for ResNet50 to fully adapt to the MNIST task.
- Lightweight models such as MobileNet are optimized for smaller datasets and demonstrated faster convergence.

4.2 Feature Extraction Power of MobileNetV2

The feature visualization experiments with MobileNetV2 further illustrate the transformation enabled by fine-tuning:

- **Before Fine-tuning:** The pre-trained ImageNet features were not well aligned with the MNIST task. PCA, t-SNE, and LLE visualizations all showed overlapping clusters, indicating weak class separability.
- **After Fine-tuning:** The model's representations became much more discriminative. PCA showed clearer grouping, t-SNE revealed highly compact clusters for each digit, and LLE also indicated improved separation.
- These results demonstrate that transfer learning successfully adapts general-purpose visual features into task-specific discriminative representations.

4.3 Impact of Transfer Learning

Overall, the experiments confirm that transfer learning substantially improves the feature extraction power of CNNs for domain-specific tasks. Pre-trained models capture broad visual hierarchies, but without fine-tuning, their features may lack the discriminative detail needed for specialized datasets like MNIST. Fine-tuning enables the model to emphasize relevant features and suppress irrelevant ones, leading to both improved classification accuracy and clearer feature clustering.

In this study, MobileNet-based models demonstrated superior adaptability compared to ResNet50, highlighting the importance of selecting architectures that balance model complexity with dataset characteristics. These findings suggest that for small, well-structured datasets, lightweight architectures such as MobileNet and MobileNetV2 may be more effective than deeper networks like ResNet50.

5 Conclusion

This experiment successfully demonstrated the feature extraction power of a pre-trained CNN (MobileNetV2) and the transformative effect of transfer learning. Visualizations of feature vectors using dimensionality reduction techniques (PCA, t-SNE, and LLE) showed a marked improvement in class separability after fine-tuning the model on the MNIST dataset. Overall, the findings underscore that fine-tuning significantly enhances the feature extraction capability of pre-trained CNNs, enabling them to produce highly discriminative representations. Lightweight architectures such as MobileNet and MobileNetV2 proved particularly effective for small, structured datasets like MNIST, outperforming deeper networks such as ResNet50 in both adaptability and efficiency.