University of Rajshahi

Department of Computer Science and Engineering

CSE4261 Neural Networks and Deep Learning

Assignment-3

Submitted By:
Name: Md.Sumon Miah
Roll: 1910676103

Submitted To:
SANGEETA BISWAS
ASSISTANT PROFESSOR

June 21, 2025

Abstract

This report analyzes the feature extraction for MNIST digit recognition through transfer learning. A VGG16 model's feature representations are visualized before and after fine-tuning on the MNIST dataset. We used three dimensionality reduction techniques—PCA, t-SNE, and UMAP—to project the high-dimensional features onto a 2D plane.

1 Methodology

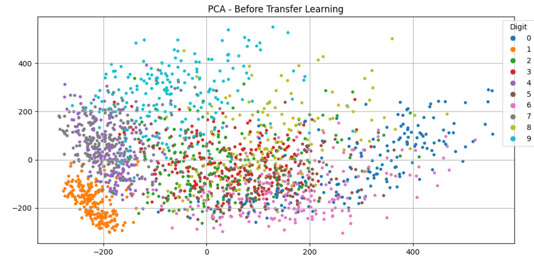
A VGG16 model pre-trained on ImageNet was used as a base. Features for 2000 MNIST digits were extracted from this base model ("Before Transfer Learning"). A new classification head was then added, and the model was fine-tuned on the MNIST data. Features were subsequently extracted from the final hidden layer of this trained model ("After Transfer Learning"). The feature vectors from both stages were reduced to two dimensions using Principal Component Analysis (PCA), t-Distributed Stochastic Neighbor Embedding (t-SNE), and Uniform Manifold Approximation and Projection (UMAP) for comparison.

2 Results and Discussion

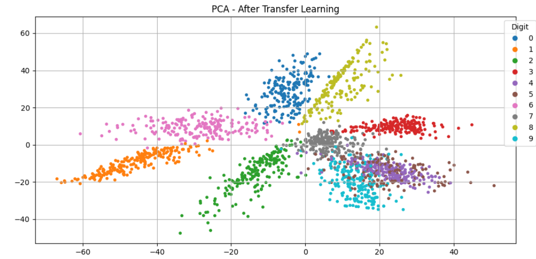
The visualizations below compare the feature spaces before and after the transfer learning process of the three reduction techniques.

Figure 1: 2D feature visualizations Before and After Transfer Learning.

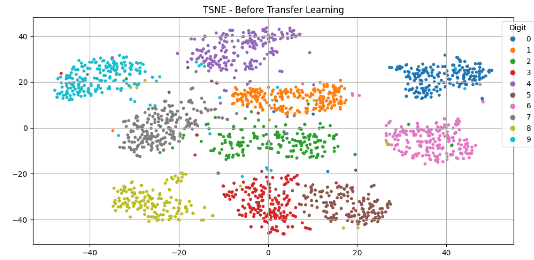
Before Transfer Learning **After Transfer Learning**



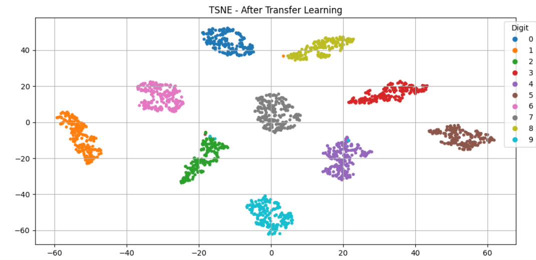
(a) PCA Before



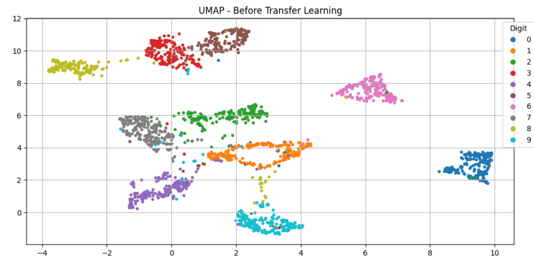
(b) PCA After



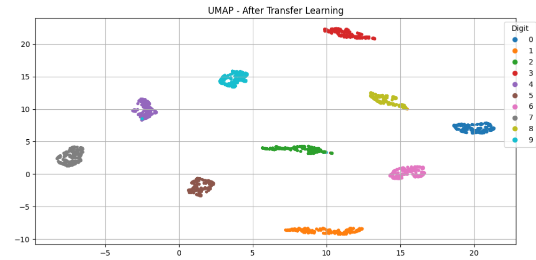
(c) t-SNE Before



(d) t-SNE After



(e) UMAP Before



(f) UMAP After

2.1 Before Transfer Learning

Initially, the features from the pre-trained VGG16 model show poor suitability for digit classification. The PCA plot (Fig. 1a) shows a single, undifferentiated cloud of data points, indicating that the principal components do not align with the digit classes. The t-SNE (Fig. 1c) and UMAP (Fig. 1e) plots show some local clustering, but the classes are heavily mixed and poorly defined.

2.2 After Transfer Learning

A dramatic improvement is seen after fine-tuning the model.

- **PCA (Fig. 1b):** Shows improved separation, with classes beginning to form distinct streams, yet significant overlap remains. This highlights the limitation of a linear technique in capturing the complex, non-linear feature space learned by the model.
- **t-SNE (Fig. 1d):** Reveals a remarkable transformation, with the data now organized into dense, well-separated clusters corresponding to each digit.
- **UMAP (Fig. 1f):** Provides clearly defined clusters.

Both t-SNE and UMAP visualize how the network has learned to map digits to distinct regions of the feature space.

3 Conclusion

Transfer learning effectively repurposed a general-purpose image recognition model for the specific task of digit classification. The fine-tuning process reshaped the feature space to make it highly discriminative, a change clearly visualized by the enhanced clustering shown by all three reduction techniques, and most notably by t-SNE and UMAP. This confirms the value of transfer learning for efficient model development on specialized tasks.