Unsupervised Feature Learning with Autoencoders

Introduction

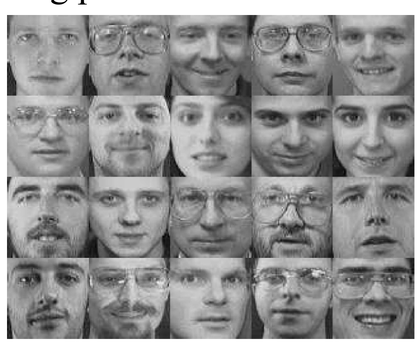
This report details a deep learning project that focuses on reconstructing images using a Convolutional Autoencoder (CAE), followed by unsupervised clustering with K-Means. The Olivetti Faces dataset, consisting of grayscale images of faces, was chosen for this experiment. The primary objective is to train a CAE to compress these images into a lower-dimensional space and reconstruct them with high fidelity, and then perform clustering on the encoded representations.

Dataset Overview

The Olivetti Faces dataset contains a set of face images taken between April 1992 and April 1994. This dataset is a classic in the field of machine learning and computer vision for testing and benchmarking algorithms.

Data Preprocessing:

- Reshaping and Normalization: Images are reshaped to fit the input layer of the CAE and normalized to facilitate learning.
- Train-Test Split: The dataset is split into training and testing sets, with 80% of the data used for training and the remaining 20% for testing.
- Data Loader: Torch's DataLoader is used to efficiently load data in batches during the training and testing phases.



Model Architecture

The Convolutional Autoencoder architecture is designed to capture spatial hierarchies in images through convolutional layers.

Encoder:

- Layer Structure: Consists of three convolutional layers with ReLU activation functions. The layers progressively decrease the image's spatial dimensions while increasing its depth.

- Dimensionality Reduction: The model compresses the input image into a lower-dimensional latent space, capturing essential features.

Decoder:

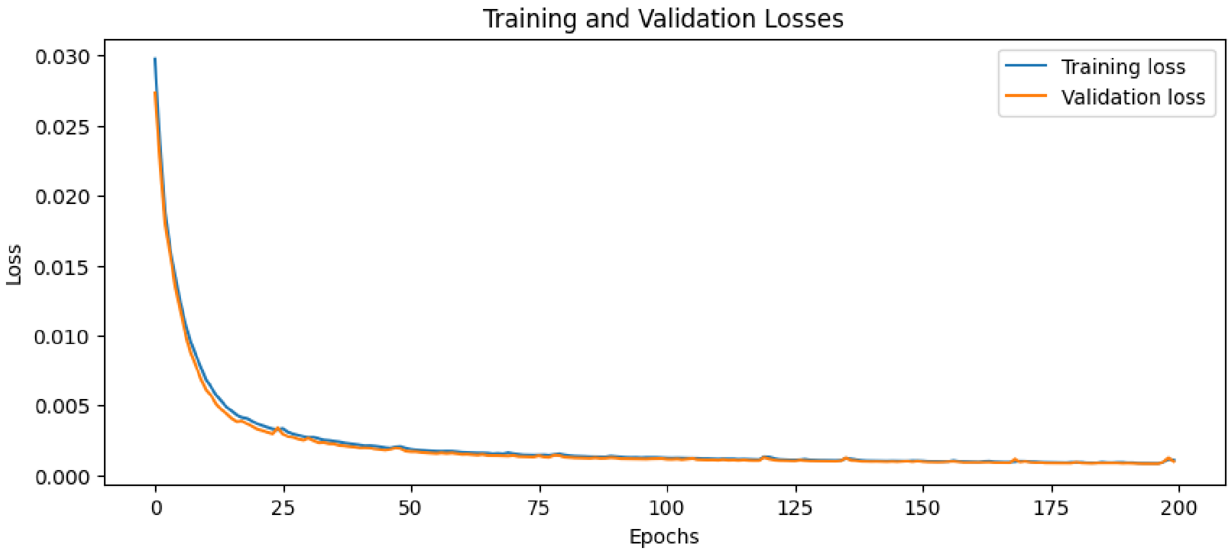
- Layer Structure: Mirrors the encoder with ConvTranspose layers to increase the spatial dimensions while decreasing depth.
- Output Activation: A sigmoid activation function is used in the final layer to scale the output between 0 and 1, matching the input image's normalized pixel values.

Hyperparameters

- Learning Rate: Set to 0.001, suitable for the Adam optimizer used in this context.
- Loss Function: Mean Squared Error (MSE) Loss, as it is effective for regression-like tasks such as image reconstruction.
- Batch Size: 32, balancing the trade-off between computational efficiency and model performance.
- Number of Epochs: 200, determined empirically to allow sufficient learning without overfitting.

Training Process and Observations

- Loss Trend: The loss values decrease steadily over epochs, indicating that the model is learning effectively.
- Overfitting Checks: No significant overfitting observed, as evidenced by a consistent decrease in training loss.
- Model Evaluation: Periodic visual inspection of reconstructed images against the originals during the training phase for qualitative assessment.



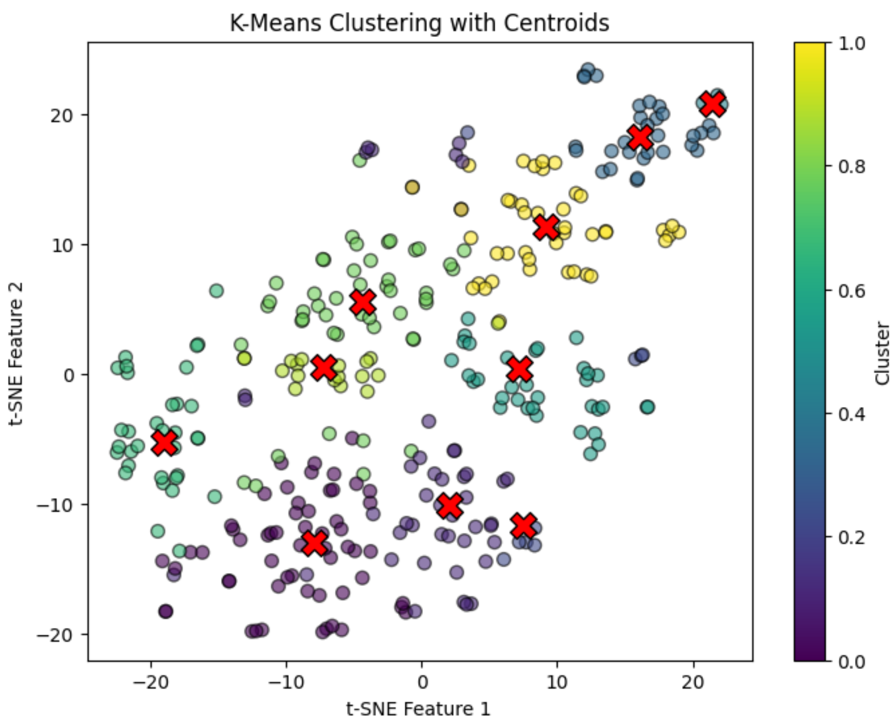
Observations from Learning Curves

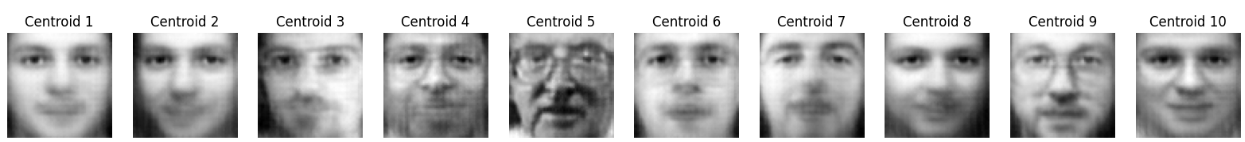
- Training Loss Curve: A plot of the training loss over epochs shows a clear downward trend, validating the effectiveness of the learning process.
- Image Reconstruction: Visual comparison between original and reconstructed images demonstrates the CAE's capability in capturing essential facial features and structures. The reconstructed images, although slightly blurred, maintain a high resemblance to the original images, showcasing the model's efficacy.

Dimensionality Reduction and Clustering with K-Means

After training the CAE, the encoded features of the images were extracted and used as input for K-Means clustering. This allowed for the grouping of similar images in the reduced feature space, providing insights into the natural clustering of the data.

- Feature Extraction: The encoder part of the CAE was used to compress images into a lower-dimensional feature space.
- K-Means Clustering: The encoded features were then clustered using K-Means. The number of clusters was selected based on empirical evaluation and silhouette analysis.
- Visualization with t-SNE: To visualize the clustering, t-SNE was applied to the encoded features, reducing them to two dimensions for easy plotting.





Observations from Clustering

- Cluster Visualization: The t-SNE plot showed distinct clusters, indicating effective grouping by K-Means.
- Cluster Interpretation: Examination of images within each cluster revealed similarities among faces, validating the clustering approach.

Conclusion

This project demonstrated the effective use of a Convolutional Autoencoder for image compression and reconstruction, followed by successful clustering of encoded features using K-Means. The CAE learned to capture significant facial features, and the subsequent clustering revealed meaningful groupings in the data. These results highlight the synergy between deep learning-based feature extraction and traditional clustering techniques, showcasing their combined potential in understanding and processing complex visual datasets.