MINI PROJECT: AUTO ENCODERS FOR REPRESENTATION LEARNING

TASK: IMAGE DENOISING

INTRODUCTION

Autoencoders are a type of neural network used for unsupervised learning, where the goal is to compress input data into a lower-dimensional representation and then reconstruct it back to its original form. This structure allows them to learn important features of the data without explicit labels. Image denoising is a process of removing unwanted noise from images while preserving important visual details. Autoencoders are well-suited for this task because, during training, they learn to map noisy input images to their clean versions by capturing essential patterns and structures. As a result, the trained network can effectively reconstruct noise-free images, making autoencoders a powerful tool for image restoration applications.

METHODOLOGY

1. **DATASET:**

   The Fashion-MNIST dataset was used, consisting of 60,000 training and 10,000 test grayscale images of size 28×28 pixels, covering 10 classes of clothing items. Each image was normalized to a pixel intensity range of [0, 1]. To simulate noisy conditions, Gaussian noise with a standard deviation factor of 0.25 was added to each image.

2. **MODEL ARCHITECTURE:**

   A Deep Convolutional Autoencoder was designed to learn noise-free representations. The architecture consists of multiple convolutional layers to extract hierarchical spatial features, followed by pooling for dimensionality reduction and upsampling for reconstruction.

   Encoder

   - Conv2D (32 filters, 3×3, ReLU, padding='same') × 3
   - MaxPooling2D (2×2, padding='same')
   - Conv2D (64 filters, 3×3, ReLU, padding='same') × 3
   - MaxPooling2D (2×2, padding='same')
   - Conv2D (128 filters, 3×3, ReLU, padding='same')

   Decoder

   - UpSampling2D (2×2)
   - Conv2D (64 filters, 3×3, ReLU, padding='same') × 3
   - UpSampling2D (2×2)
   - Conv2D (32 filters, 3×3, ReLU, padding='same') × 2
   - Output Layer: Conv2D (1 filter, 3×3, Sigmoid, padding='same')

The model was compiled using the Adam optimizer and a custom hybrid loss function combining Mean Squared Error (MSE) and Structural Similarity Index (SSIM):

$$L = \alpha \cdot MSE + \beta \cdot (1 - SSIM)$$

with $\alpha = 0.8$ and $\beta = 0.2$, ensuring that both pixel-level accuracy and perceptual similarity are optimized.

3. **TRAINING:**

The model was trained for 15 epochs with a batch size of 64 using Early Stopping and Learning Rate Reduction callbacks to prevent overfitting and improve convergence. The validation data consisted of noisy and clean test pairs from the Fashion-MNIST dataset.

RESULT AND EVALUATION

To evaluate reconstruction quality, two standard metrics were used:

- Peak Signal-to-Noise Ratio (PSNR) — measures overall reconstruction fidelity.
- Structural Similarity Index (SSIM) — measures perceptual image similarity.

Metric used	Value
PSNR	22.70
SSIM	0.8371