Denoising Autoencoders for MNIST and Fashion-MNIST

Universidad Carlos III de Madrid

Duarte Moura

Roberto Martínez - Guisasola Guerrero

October 21, 2024

This project explores the effectiveness of denoising autoencoders applied to the MNIST and Fashion-MNIST datasets. By implementing varying architectures and encoded space dimensions, we assess reconstruction quality using Peak Signal-to-Noise Ratio (PSNR) and visualize denoising capabilities.

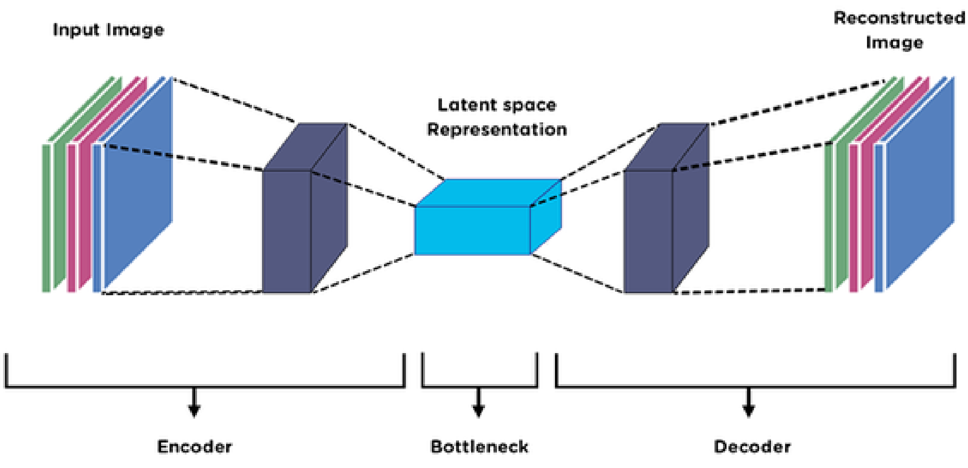


Figure 1: Autoencoder Architecture

1 Introduction

Autoencoders are neural networks designed to learn efficient codings of data by compressing inputs into a latent space and reconstructing the output from this representation. Autoencoders have been traditionally used for dimensionality reduction or feature learning. In this project, we investigate denoising autoencoders applied to the MNIST and Fashion-MNIST datasets, aiming to improve image reconstruction quality despite the presence of noise.

2 Methodology

2.1 Autoencoder Structures

Two distinct autoencoder architectures were implemented to evaluate their performance on denoising tasks.

2.1.1 Autoencoder3

A 3-layer autoencoder with the following structure:

- **Encoder**: 784 neurons that are the images pixel's flattened. After those neurons there are **2 hidden layers** with 512 and 256 neurons each one with a ReLU activation function. Finally, there is the **latent space** that has variable dimensions (15, 30, 50, 100).
- **Decoder**: Mirrors encoder layers with Sigmoid activation at output of the decoder.

2.1.2 Autoencoder5

A more complex 5-layer autoencoder with the following structure:

- **Encoder**: 784 neurons that are the images pixel's flattened. After those neurons there are **4 hidden layers** with 656, 512, 256 and 128 neurons each one with a ReLU activation function. Finally, there is the **latent space** that has variable dimensions (15, 30, 50, 100).
- **Decoder**: Mirrors encoder layers with Sigmoid activation at output.

2.2 Training Setup

- **Datasets**: MNIST and Fashion-MNIST. Each dataset has 60.000 images for training and 10.000 images for testing.
- **Hyperparameters**:
  - Epochs: 20
  - Learning Rate: 0.001
  - Batch Size: 128 and 64
  - L1 Regularization: 1×10^{-5}, 1×10^{-4}, 1×10^{-3}
  - Noise Variance: 0.01, 0.02, 0.05, 0.1
- **Optimizer**: Adam
- **Loss Function**: Mean Squared Error (MSE) with L1 regularization

Gaussian noise was added to input images to train denoising autoencoders, enabling the models to reconstruct clean images from noisy inputs.

3 Results

3.1 Performance Evaluation

Model performances were evaluated using Peak Signal-to-Noise Ratio (PSNR). Table 1 summarizes the best and worst PSNR results for each dataset.

Table 1: Best and Worst PSNR Results for MNIST and Fashion-MNIST

Dataset	Performance	PSNR (dB)	Architecture	Encoded Dimension	L1 Penalty
MNIST	Best	24.01	Autoencoder3	100	1×10^{-4}
MNIST	Worst	14.01	Autoencoder5	15	1×10^{-3}
Fashion-MNIST	Best	21.16	Autoencoder3	100	1×10^{-4}
Fashion-MNIST	Worst	15.49	Autoencoder5	30	1×10^{-5}

3.2 Training Dynamics

The training dynamics were analyzed by observing the PSNR progression over epochs for different encoded dimensions. In this figure we are actually looking at the best value over all different values of L1 Regularization.

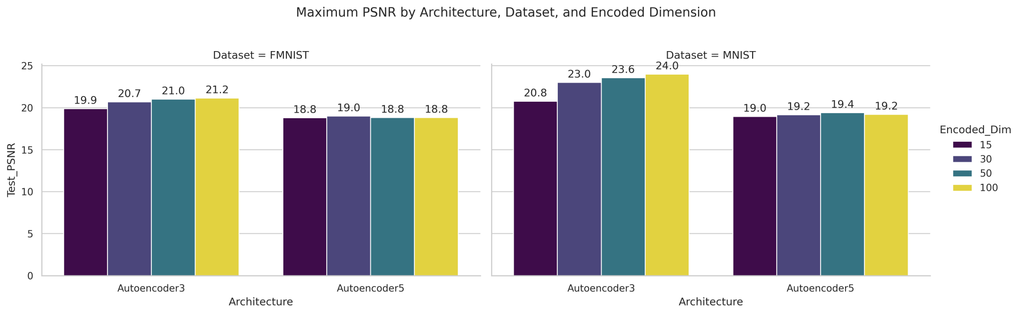


Figure 2: Best PSNR over Epochs

We also analyzed the training versus test loss for the worst-performing architectures to check for potential overfitting. However, there are no significant signs of overfitting. Both the training and test loss decrease in a similar fashion, with no substantial divergence between the two curves, which typically indicates overfitting.

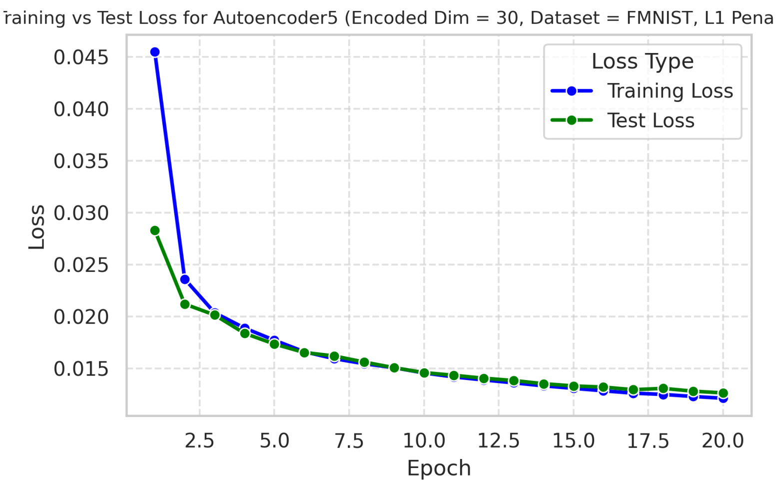
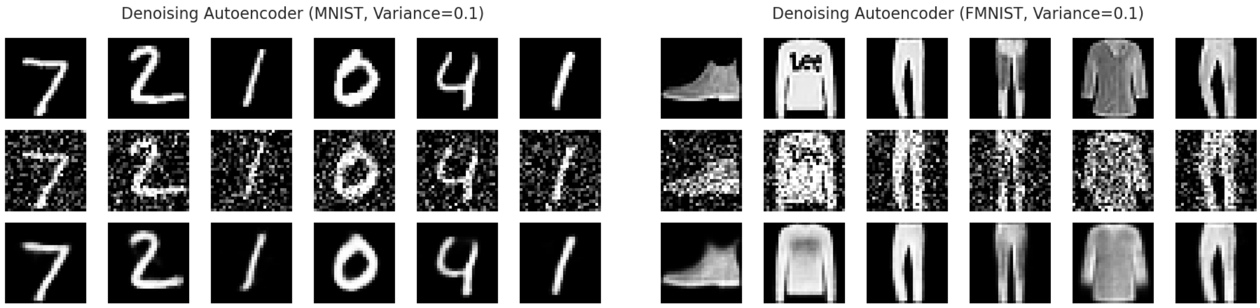


Figure 3: Training vs Test Loss for Worst Performing Architecture and Dataset

3.3 Denoising Capabilities

The denoising performance was visualized by comparing original, noisy, and reconstructed images. This was achieved by creating a denoising autoencoder using the best architecture identified in the previous section, which was consistent for both datasets. Figure 4 shows examples from both MNIST and Fashion-MNIST datasets.



(a) MNIST (b) Fashion-MNIST

Figure 4: Denoising Results: Original vs. Noisy vs. Reconstructed Images

Below are the plotted values of the PSNR variation with the variance parameter of the Gaussian noise introduced in the images.

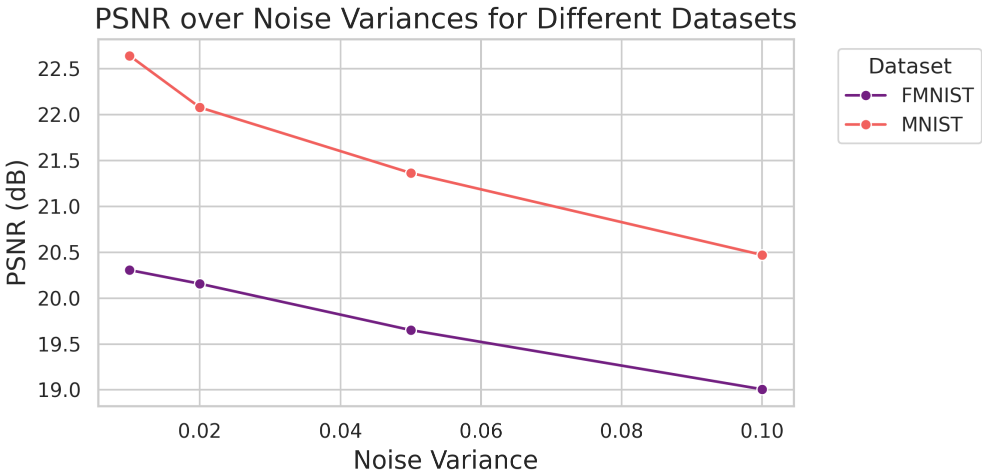


Figure 5: PSNR vs. Noise Variance

4 Conclusions

The experiments revealed that the simpler **Autoencoder3** consistently outperformed the deeper **Autoencoder5** on both MNIST and Fashion-MNIST datasets. This indicates that additional network depth did not benefit, and may have hindered, performance on these simple datasets—possibly due to training challenges like vanishing gradients. A key observation was that larger encoded dimensions (e.g., 100) led to higher PSNR values, meaning better reconstruction quality. This aligns with the expectation that a larger latent space can capture more meaningful features from the input data. Additionally, an L1 regularization penalty of 1×10^{-4} provided a good balance between underfitting and overfitting, contributing to improved generalization. As expected, higher noise levels resulted in lower PSNR values, emphasizing the difficulty of reconstructing heavily corrupted images. Future work could investigate incorporating skip connections or early stopping to enhance training in deeper models. We also tested the AE architectures for 2 different batches (64 and 128) but the results were very similar, so the results in this report are corresponding to the 128 batch notebook.