uc3m | Universidad Carlos III de Madrid

Master Degree in ML

2025–2026

*Deep Learning Project II*

# “Deep Autoencoders for Image Reconstruction and Denoising”

Àlex Sánchez Zurita

Santiago Prieto Núñez

Jorge Barcenilla González

02 - 11 - 2025

# 1 Introduction

This project focuses on the implementation and analysis of **autoencoders** for image reconstruction and denoising. Initially, a standard autoencoder is developed to learn efficient low-dimensional representations of the MNIST and Fashion-MNIST datasets by compressing and reconstructing input images in an unsupervised manner. Subsequently, a denoising autoencoder (DAE) is trained to recover clean images from inputs corrupted by Gaussian noise, allowing the evaluation of its robustness and reconstruction fidelity.

An **autoencoder** is a neural network that learns to encode input data $x \in \mathbb{R}^d$ into a compact latent space $z = e(x) \in \mathbb{R}^k$ with $k < d$, and then reconstructs the input as $\hat{x} = d(z) = d(e(x))$. The architecture comprises two main components:

- **Encoder ($e$):** compresses the input into a lower-dimensional latent representation (*bottleneck*) that captures the most relevant features.
- **Decoder ($d$):** reconstructs the original input from the latent representation.

The model is trained by minimizing a **reconstruction loss** $L(x, \hat{x})$, which measures how accurately the output $\hat{x}$ reproduces the input $x$. Although the Mean Squared Error (MSE) is a common choice, this project employs the **Binary Cross-Entropy (BCE)** loss, better suited for pixel values normalized to the $[0, 1]$ range of MNIST and FMNIST datasets. The BCE loss is defined as:

$$\mathcal{L}_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}\left[x_i \log(\hat{x}_i) + (1-x_i)\log(1-\hat{x}_i)\right] \tag{1.1}$$

To quantify reconstruction performance, the **Peak Signal-to-Noise Ratio (PSNR)** is employed, a standard metric that measures the similarity between the original and reconstructed images based on the Mean Squared Error (MSE):

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX^2}{MSE}\right) \tag{1.2}$$

Here, $MAX$ represents the maximum pixel value. Since the datasets are normalized to $[0, 1]$, $MAX = 1$, while for 8-bit grayscale images $MAX = 255$. A higher PSNR value indicates a better reconstruction, reflecting lower distortion relative to the original image.

In summary, this report presents a practical study on the design, training, and evaluation of deep autoencoders, analyzing the effects of architecture depth, latent dimension size, and regularization, as well as the performance of denoising autoencoders under varying noise conditions.

# 2 Experiments and results

## 2.1. Results and Discussion

In this section, we show the results aimed to evaluate how architectural complexity, latent space size, regularization, and noise robustness affect the performance of dense autoencoders on MNIST and Fashion-MNIST datasets.

### Experiment 1: Architecture and Latent Dimension

Results show that **increasing the latent dimension improves reconstruction quality**, with PSNR values rising from approximately 20 dB at 15 dimensions to nearly 27 dB at 100 dimensions for MNIST as we can see in Figure 2.1. Also **with a computational cost of 1s for epoch** between 15 and 100 latent layers.

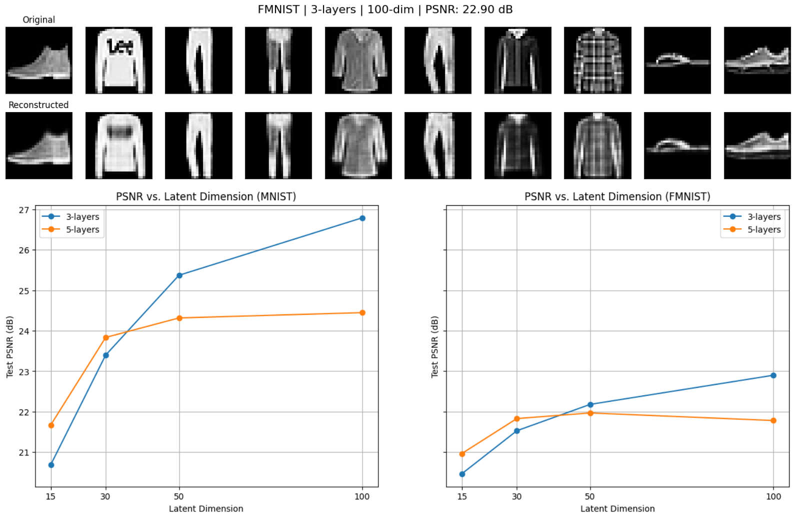

Fig. 2.1. Comparison of results on MNIST and FMNIST datasets. (Top) Output reconstruction for FMNIST. (Bottom) PSNR vs latent dimension across both datasets.

However, deeper networks (5 layers) did not outperform the simpler 3-layer models. In fact, the 3-layer autoencoder achieved higher PSNR values and lower reconstruction losses, especially for FMNIST, indicating that **additional depth added unnecessary complexity and slightly reduced generalization**. Thus, the **best-performing model was the 3-layer architecture with a latent dimension of 100**.

### Experiment 2: $L_1$ Regularization on the Latent Space

Adding an $L_1$ (Lasso) penalty on the encoder's output effectively encouraged sparsity in the latent representation, as we can see on Figure 2.2. Low regularization values ($\lambda_{L1} \leq 0.01$) had minimal impact on reconstruction quality (PSNR $\approx$ 25 dB), while stronger penalties ($\lambda_{L1} \geq 0.1$) reduced PSNR to around 22 dB due to excessive constraint on the latent activations. This confirms that **mild regularization provides a good balance between interpretability and reconstruction fidelity**.

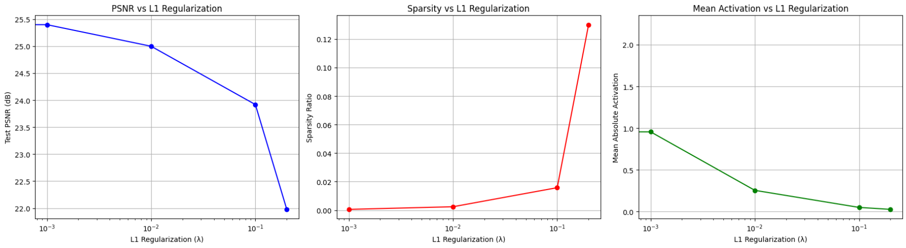

Fig. 2.2. Evolution of PSNR (Right), Sparsity (Center) and Mean values of the latent layer (Left)

## Experiment 3: Denoising Autoencoder (DAE)

We train a denoising autoencoder (DAE) with 3 layers and 100 latent to reconstruct clean images from inputs corrupted with Gaussian noise of variances $[0.01, 0.05, 0.1, 0.2]$.

For **MNIST**, the model achieved slight PSNR gains as noise increased, reaching about $+1$ dB at $\sigma^2 = 0.2$, showing limited denoising ability. In contrast, **Fashion-MNIST** performance degraded with noise (up to $-1.2$ dB), indicating over-smoothing and loss of texture detail as we can see on 2.3.

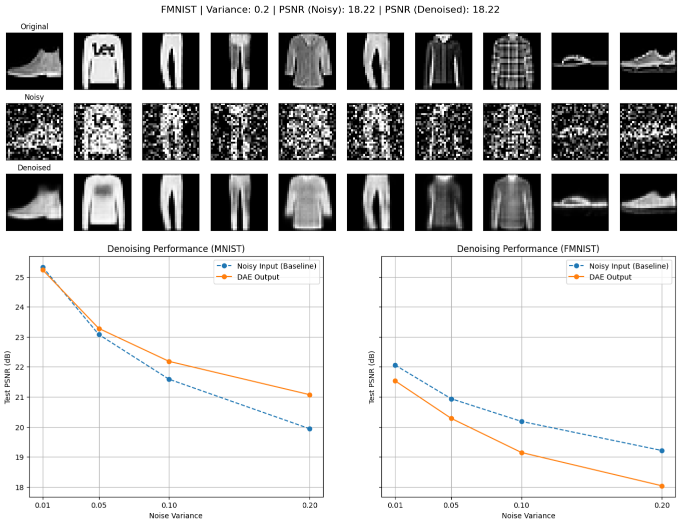

Fig. 2.3. (Top) FMNIST reconstruction under high noise. (Bottom) PSNR comparison between noisy and denoised images.

Overall, the DAE reduced noise slightly in MNIST but failed to generalize to the more complex FMNIST dataset.