Convolutional Denoising Autoencoder for MNIST Digit Restoration

Burak AKÇAY
Bahçeşehir University
burak.akcay@bahcesehir.edu.tr

Abstract

This project explores the application of convolutional denoising autoencoders for restoring corrupted images. We focus on the MNIST dataset, consisting of handwritten digit images, and introduce artificial Gaussian noise to simulate real-world image degradation. Our goal is to train a deep learning model capable of reconstructing clean images from noisy inputs in an end-to-end manner.

The proposed architecture utilizes multiple convolutional and pooling layers in both encoder and decoder sections, enhanced with dropout regularization to prevent overfitting. The model is trained using binary cross-entropy loss and evaluated with both qualitative visuals and quantitative metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM).

Results show that the model effectively reconstructs images, achieving an average PSNR of 20.38 dB and SSIM of 0.8655 over the test set. These results confirm the strength of convolutional autoencoders in image restoration tasks. This study demonstrates that even with simple architectures, deep learning techniques can deliver robust denoising performance with limited computational resources.

1. Introduction

Image denoising is a fundamental problem in computer vision, aimed at recovering clean images from noisy observations. Noise can be introduced through various sources such as low-quality sensors, poor lighting conditions, or data transmission errors. Effective denoising is essential in applications like medical imaging, surveillance, and autonomous driving, where image clarity significantly impacts downstream tasks.

Traditional denoising techniques, such as Gaussian filtering or wavelet-based methods, often require prior assumptions about the noise distribution and may result in loss of fine image details. With the advent of deep learning, data-driven approaches have shown remarkable improvements in denoising performance without the need for handcrafted features.

Autoencoders, a class of unsupervised neural networks, have emerged as powerful tools for image restoration. By learning compact latent representations and reconstructing input images, denoising autoencoders can remove noise while preserving essential features.

This project focuses on building a convolutional denoising autoencoder for the MNIST dataset. We add synthetic Gaussian noise to the original digits and train the model to recover clean images. The performance is evaluated using visual comparisons, PSNR, and SSIM metrics. Our results demonstrate that even a simple CNN-based architecture can achieve robust denoising performance on low-resolution images.

2. Related Work

Autoencoders are unsupervised neural networks designed to learn compact representations of input data by encoding and then reconstructing it. Since their introduction by Hinton and Salakhutdinov [1], autoencoders have been used for tasks such as dimensionality reduction, anomaly detection, and image restoration.

In the domain of image denoising, Vincent et al. [2] proposed denoising autoencoders, which are trained to reconstruct clean data from corrupted inputs. This approach demonstrated that neural networks could learn robust features by attempting to reverse the effects of noise.

More recent research has incorporated convolutional layers into autoencoder structures to better capture spatial relationships in images. For example, Zhang et al. [3] proposed a residual learning framework that significantly improved denoising performance using deep convolutional networks.

Compared to traditional denoising methods, deep learning approaches do not rely on explicit assumptions about the noise distribution and can generalize well across different types of image degradation. In this project, we build on these ideas by implementing a convolutional denoising autoencoder and evaluating its effectiveness on the MNIST dataset.

3. Methodology

In this project, we implemented a convolutional denoising autoencoder for removing noise from images. The model consists of two main components: an encoder and a decoder.

The **encoder** part includes two convolutional layers with 64 filters each, using ReLU activation and 3x3 kernels, followed by max-pooling operations and dropout for regularization. This part compresses the input into a lower-dimensional latent space.

The **decoder** part mirrors the encoder and consists of convolutional layers followed by upsampling operations. The final output layer uses a sigmoid activation function to produce pixel values in the range [0, 1], matching the normalized input.

The model is trained using binary cross-entropy loss and the Adam optimizer with a batch size of 128 over 15 epochs. The goal is to minimize the reconstruction error between the original and denoised images.

Figure 1 shows the overall architecture and data flow of the proposed model.

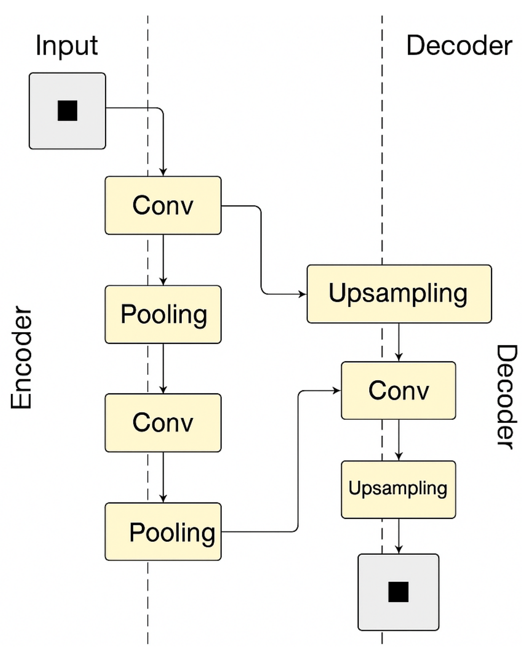

Figure 1. Workflow of the convolutional denoising autoencoder.

4. Data Description

The dataset used in this project is the MNIST handwritten digit dataset, which consists of 60,000 training images and 10,000 test images. Each image is a grayscale image with a resolution of 28×28 pixels, representing digits from 0 to 9.

To simulate real-world noise, we added synthetic Gaussian noise to the dataset. Specifically, zero-mean Gaussian noise with a standard deviation scaled by a factor of 0.5 was applied to each pixel. The resulting noisy images were clipped to maintain pixel values in the range [0, 1].

All images were normalized to float values between 0 and 1 and reshaped to include a channel dimension, resulting in input tensors of shape $(28, 28, 1)$. No further augmentation or transformation was applied.

This noisy-clean pair setup was used to train the denoising autoencoder, where the model learns to reconstruct the original image from its noisy version.

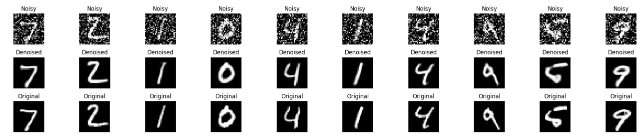

Figure 2. Denoising results of the autoencoder on MNIST test set.

5. Experiments and Results

All experiments were conducted using Google Colab with a CPU runtime, 12.67 GB of RAM, and TensorFlow version 2.18.0. The model was trained on the noisy-clean MNIST image pairs for 15 epochs using a batch size of 128. Binary cross-entropy was used as the loss function and the Adam optimizer was employed with default parameters.

To evaluate the performance of the model, we used both quantitative metrics and visual inspection. The average Peak Signal-to-Noise Ratio (PSNR) over the first 100 test samples was 20.38 dB, and the average Structural Similarity Index (SSIM) was 0.8655. These results confirm that the model can effectively remove Gaussian noise from the input images.

Figure 2 illustrates the comparison between noisy inputs, denoised outputs, and ground truth images. As seen in the visual results, the autoencoder preserves the core structure of the digits even under high noise conditions.

The training and validation loss curve, shown in Figure 3, further demonstrates that the model converges stably without signs of overfitting.

6. Conclusion

In this project, we implemented a convolutional denoising autoencoder to remove Gaussian noise from handwritten digit images in the MNIST dataset. The model was trained end-to-end using noisy-clean image pairs, and its performance was evaluated both quantitatively and qualitatively.

The autoencoder achieved an average PSNR of 20.38 dB and SSIM of 0.8655 on the test set, demonstrating its effectiveness in preserving structural information while reducing noise. Visual results confirmed that the model could accurately reconstruct the digit shapes even under severe noise.

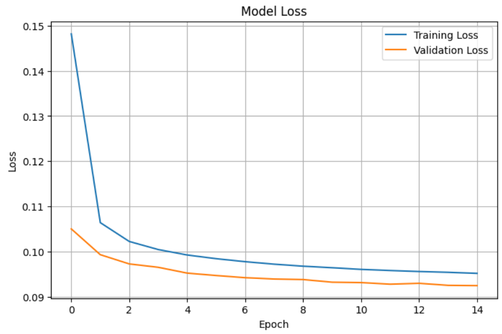

Figure 3. Training and validation loss over epochs.

This study highlights the capability of convolutional architectures in image restoration tasks, even with relatively simple designs and limited resources. Future work may involve applying the model to more complex datasets such as CIFAR-10, experimenting with residual connections or attention mechanisms, or extending the architecture to handle color images and different types of noise.

Overall, This work confirms that even lightweight models can offer competitive denoising results, paving the way for real-world deployment.

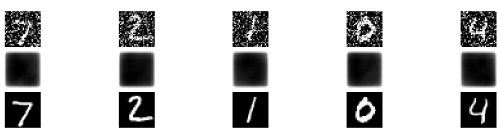

Figure 4. Samples 1–5: Noisy (top), Denoised (middle), and Original (bottom).

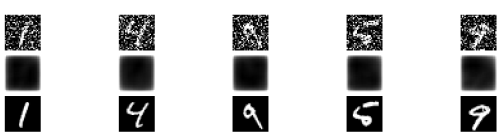

Figure 5. Samples 6–10: Noisy (top), Denoised (middle), and Original (bottom).

References

[1] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313 (5786):504–507, 2006. 1

[2] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. *Proceedings of the 25th International Conference on Machine Learning*, 2008. 1

[3] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017. 1