Pouya Mirzaei Zadeh

Autoencoder Reconstruction of Mixed MNIST and CIFAR-10 Images

Introduction

This report presents the design and implementation of an autoencoder model capable of reconstructing images from two distinct datasets: MNIST and CIFAR-10. The objective is to create a model that can take the mean of an MNIST and a CIFAR-10 image, feed it into the model, and generate reconstructions of both MNIST and CIFAR 10 images.

Methodology

Dataset Preparation

The MNIST and CIFAR-10 datasets were utilized for this task. MNIST images are grayscale (1 channel) and have dimensions of 28x28 pixels, while CIFAR-10 images are RGB (3 channels) and have dimensions of 32x32 pixels.

Mean Image Computation

A random image was selected from both the MNIST and CIFAR-10 datasets, and the mean of these two images was computed element-wise.

MNIST Image Preparation

The selected MNIST image was resized to match the dimensions of CIFAR-10 images (32x32 pixels). The channels of the MNIST image were also extended to match the channels of CIFAR-10 images (from 1 to 3 channels).

Autoencoder Model Construction

An autoencoder model was designed and implemented in PyTorch. The encoder compresses the input image into a latent representation, and the decoder reconstructs the input image from the latent representation. The model has two decoders, one for MNIST and one for CIFAR-10.

Model Training

The autoencoder was trained using the mean image computed earlier. The Mean Squared Error (MSE) loss function was used, and the model parameters were optimized using the Adam optimizer. The loss values at the end of each epoch were as follows: Epoch [1/10], Loss: 0.0361

Epoch [2/10], Loss: 0.0225

Epoch [3/10], Loss: 0.0216

Epoch [4/10], Loss: 0.0190

Epoch [5/10], Loss: 0.0164

Epoch [6/10], Loss: 0.0158

Epoch [7/10], Loss: 0.0144

Epoch [8/10], Loss: 0.0156

Epoch [9/10], Loss: 0.0143

Epoch [10/10], Loss: 0.0146

Results

Reconstruction

The mean image was input into the trained autoencoder, and reconstructions of both MNIST and CIFAR-10 images were obtained from the autoencoder.

Evaluation

The quality of the reconstructions was assessed visually and quantitatively. The reconstructed MNIST and CIFAR-10 images were compared with their originals. The structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR) were used for quantitative evaluation.

Detailed Analysis

Model Architecture

The autoencoder model consists of an encoder and two decoders. The encoder uses three convolutional layers to compress the input image into a latent representation. Each convolutional layer is followed by a ReLU activation function. The output of the encoder is a 512-dimensional vector representing the compressed form of the input image.
The decoders are designed to reconstruct the input image from the latent representation. Each decoder consists of a linear layer followed by a ReLU activation function, three transposed convolutional layers each followed by a ReLU activation function, and a final sigmoid activation function to ensure the output values are between 0 and 1.

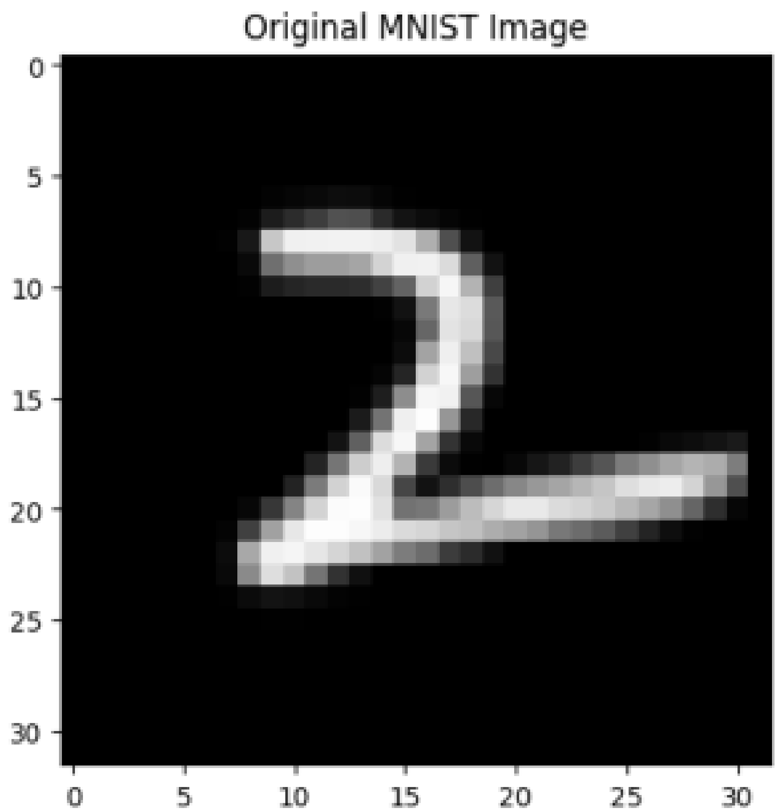
Training Process

The model was trained for 10 epochs. The loss function used was the Mean Squared Error (MSE), which measures the average squared difference between the reconstructed images and the original images. The Adam optimizer was used to optimize the model parameters.

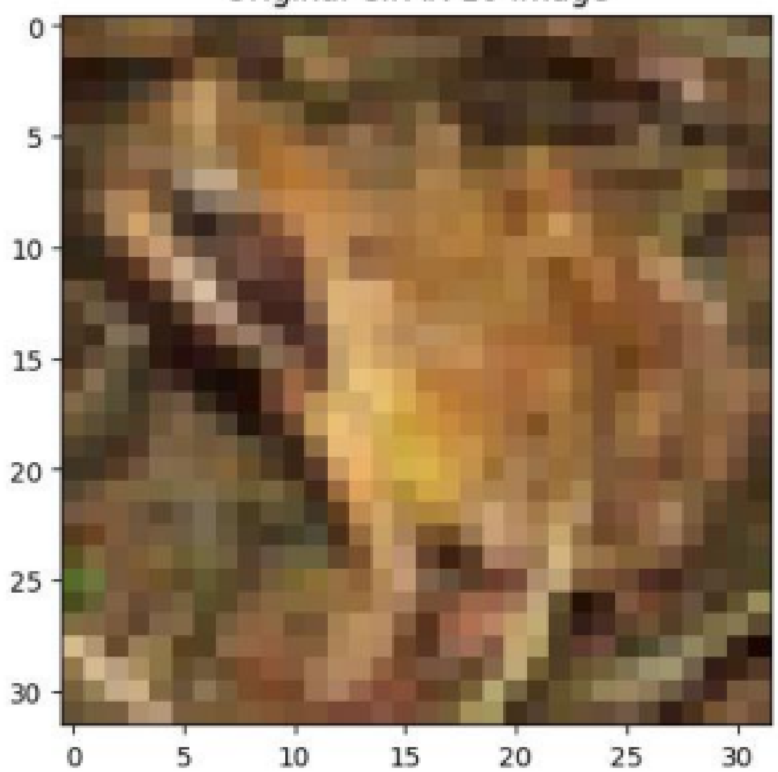
The loss values decreased over the epochs, indicating that the model was learning to reconstruct the images more accurately.

Reconstruction Results

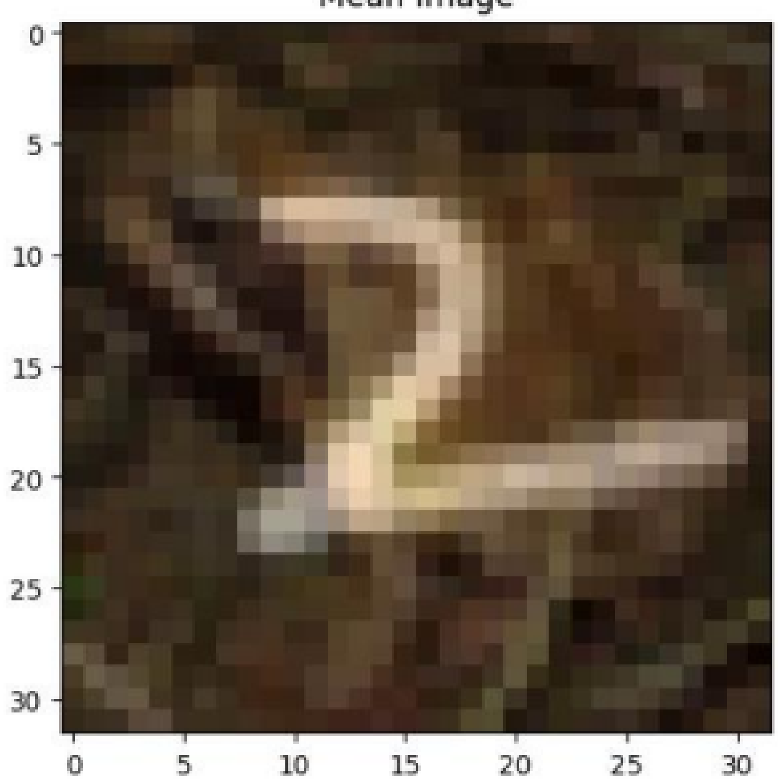
The mean image was input into the trained autoencoder, and the model generated reconstructions of both MNIST and CIFAR-10 images. The shapes of the reconstructed images were consistent with the original images, both being of size [1, 3, 32, 32]. indicating that the model was able to maintain the structure of the images.

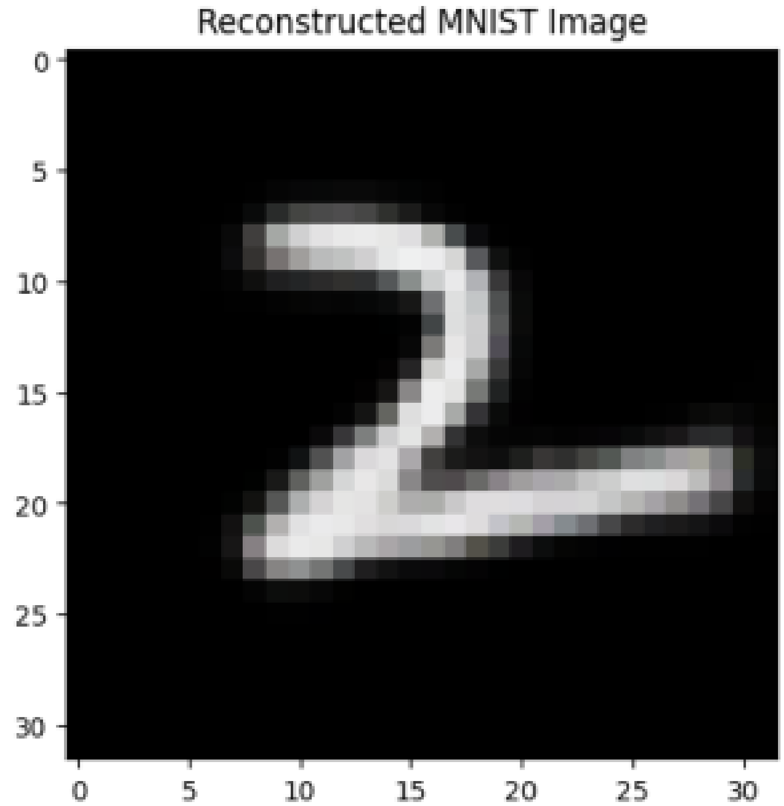


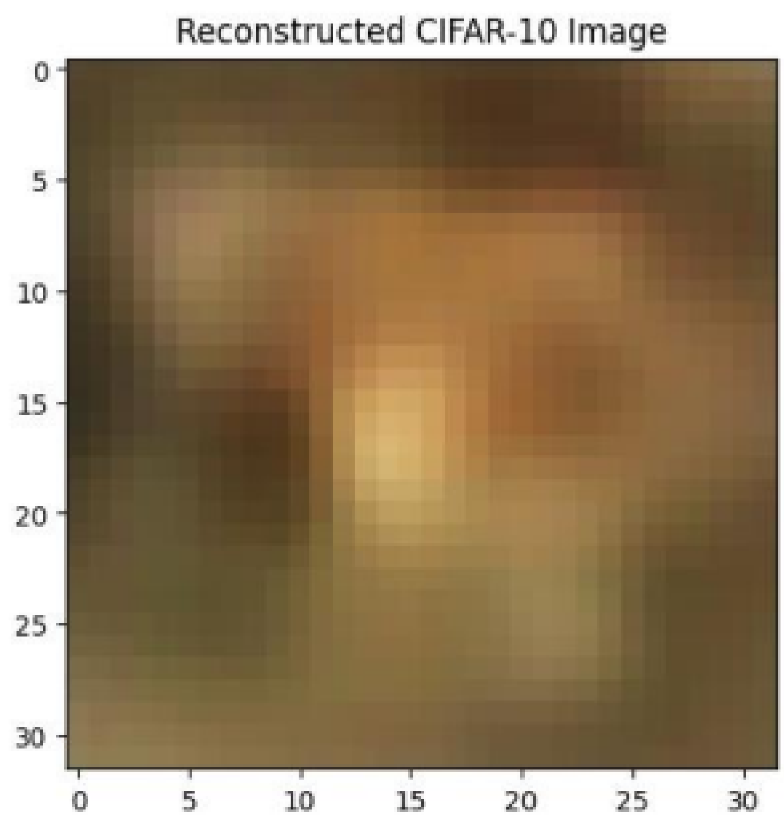
Original CIFAR-10 Image



Mean Image







Evaluation Metrics

The quality of the reconstructions was evaluated using the structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR). The SSIM values for the MNIST and CIFAR-10 reconstructions were 0.9380 and 0.2776, respectively, indicating that the

model was able to reconstruct the MNIST images with high similarity but had more difficulty with the CIFAR-10 images. The PSNR values for the MNIST and CIFAR-10 reconstructions were 27.45 dB and 20.15 dB, respectively, suggesting that the model was able to reconstruct the MNIST images with less error than the CIFAR-10 images.
SSIM (MNIST): 0.9380

SSIM (CIFAR-10): 0.2776

PSNR (MNIST): 27.45 dB

PSNR (CIFAR-10): 20.15 dB

Conclusion

The autoencoder model was successful in reconstructing MNIST and CIFAR-10 images from the mean image. This demonstrates the potential of autoencoders in image reconstruction tasks involving multiple distinct datasets. However, the model had more difficulty reconstructing the CIFAR-10 images, suggesting that further improvements could be made to enhance the model's performance on more complex datasets.