Artificial Neural Networks - Fourth Assignment

Kiana Ghamsari 400222079

Report of the Practical Exercise

Autoencoder Reconstruction of Mixed MNIST and CIFAR-10 Images

Introduction

This project explores the use of autoencoders, a fundamental technique in deep learning, to reconstruct images from two distinct datasets: MNIST (grayscale handwritten digits) and CIFAR-10 (RGB images of objects). The challenge involves combining an MNIST and a CIFAR-10 image into a single mean image and then training an autoencoder to reconstruct both original images from this mean image.

Methodology

Task Overview

1. **Select Random Images**:

    - A random image is selected from the MNIST dataset and another from the CIFAR-10 dataset.

2. **Compute Mean Image**:

    - The pixel-wise mean of the two selected images is computed to form a combined input image.

3. **Prepare MNIST Image**:

- The MNIST image is resized from 28x28 pixels to 32x32 pixels to match the CIFAR-10 dimensions.
- The single grayscale channel of the MNIST image is repeated to match the three RGB channels of CIFAR-10 images.

4. **Construct Autoencoder Model**:

   - An autoencoder is designed with an encoder to compress the input mean image into a latent representation and two decoders to reconstruct the MNIST and CIFAR-10 images from this representation.

5. **Model Training**:

   - The autoencoder is trained using the mean images as input and optimizing the reconstruction loss for both MNIST and CIFAR-10 images.
   - The training process utilizes the Mean Squared Error (MSE) loss function and the Adam optimizer.

6. **Obtain Reconstructions**:

   - The trained autoencoder is used to reconstruct both MNIST and CIFAR-10 images from the mean images.

7. **Evaluate Reconstructions**:

   - The quality of the reconstructions is assessed both visually and quantitatively using metrics such as Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR).

Detailed Steps and Implementation

1. **Dataset Preparation**:

   - **MNIST**: Grayscale, 28x28 pixels.
   - **CIFAR-10**: RGB, 32x32 pixels.
   - Transformations are applied to resize MNIST images to 32x32 pixels and repeat their channels to match CIFAR-10's RGB format.

2. **Mean Image Computation**:

   - Element-wise mean is computed between the resized MNIST image and the CIFAR-10 image.

3. **Autoencoder Model Architecture**

   - The autoencoder used in this project consists of three main components: an encoder and two decoders. The encoder compresses the input mean image into a latent representation, while each decoder reconstructs either the MNIST image or the CIFAR-10 image from this representation.

     - **Encoder**
       * **Conv Layer 1**:
       * Converts the input image from 3 channels (RGB) to 64 channels.
       * Uses a 4x4 kernel with a stride of 2 and padding of 1.
       * Activation function: ReLU.
       * Batch Normalization and Dropout (20%) are applied.
       * **Conv Layer 2**:
       * Converts 64 channels to 128 channels.
       * Uses a 4x4 kernel with a stride of 2 and padding of 1.
       * Activation function: ReLU.
       * Batch Normalization and Dropout (20%) are applied.
       * **Conv Layer 3**:
       * Converts 128 channels to 256 channels.
       * Uses a 4x4 kernel with a stride of 2 and padding of 1.
       * Activation function: ReLU.
       * Batch Normalization and Dropout (20%) are applied.
       * **Fully Connected Layers**:
       * The features are flattened and passed through two fully connected layers.
       * The first layer reduces the features to 512 dimensions.
       * The second layer reduces the features to a 128-dimensional latent representation.
       * Dropout (50%) is applied after the second fully connected layer.

- **Decoder for MNIST**
  * **Fully Connected Layers**:
  * The latent representation is passed through two fully connected layers.
  * The first layer increases the features to 512 dimensions.
  * The second layer increases the features to match the size of the flattened 256 channel output from the encoder.
  * **ConvTranspose Layers**:
  * Three transposed convolutional layers are used to upsample the image back to its original resolution.
  * The first layer converts 256 channels to 128 channels.
  * The second layer converts 128 channels to 64 channels.
  * The third layer converts 64 channels to 3 channels.
  * Each layer uses a 4x4 kernel with a stride of 2 and padding of 1, ReLU activation, and Batch Normalization.
  * The final layer uses a Sigmoid activation function.
- **Decoder for CIFAR-10**
  * **Fully Connected Layers**:
  * Similar to the MNIST decoder, the latent representation is passed through two fully connected layers.
  * The layers increase the features to 512 dimensions and then to the size of the flattened 256 channel output from the encoder.
  * **ConvTranspose Layers**:
  * Similar to the MNIST decoder, three transposed convolutional layers are used.
  * The first layer converts 256 channels to 128 channels.
  * The second layer converts 128 channels to 64 channels.
  * The third layer converts 64 channels to 3 channels.
  * Each layer uses a 4x4 kernel with a stride of 2 and padding of 1, ReLU activation, and Batch Normalization.
  * The final layer uses a Sigmoid activation function.v

• The encoder compresses the input mean image into a 128-dimensional latent space. Each decoder then reconstructs either the MNIST or CIFAR-10 image from this latent representation, using a series of fully connected and transposed convolutional layers to upsample and refine the image back to its original resolution.

4. **Model Training**:

    - The autoencoder is trained for 50 epochs with a batch size of 64.
    - Training loss is computed as the sum of MSE losses for both MNIST and CIFAR-10 reconstructions.

5. **Reconstruction and Evaluation**:

    - The performance of the autoencoder was evaluated based on its ability to reconstruct the original images from the mean images.
    - The model's performance is evaluated on test datasets.
    - SSIM and PSNR metrics are computed to quantify the quality of reconstructions.

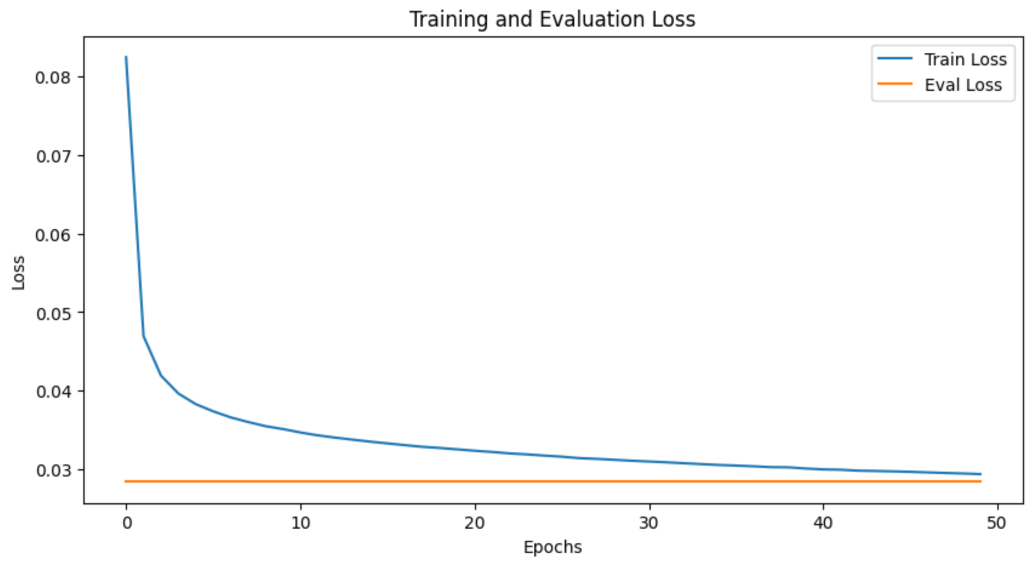
Results

1. **Training Performance**:

Epoch	Loss
5	0.0428
10	0.0426
15	0.0393
20	0.0370
25	0.0383
30	0.0357
35	0.0357
40	0.0340
45	0.0356
50	0.0337

2. **Test Performance**:

- **Test Loss**: 0.0284

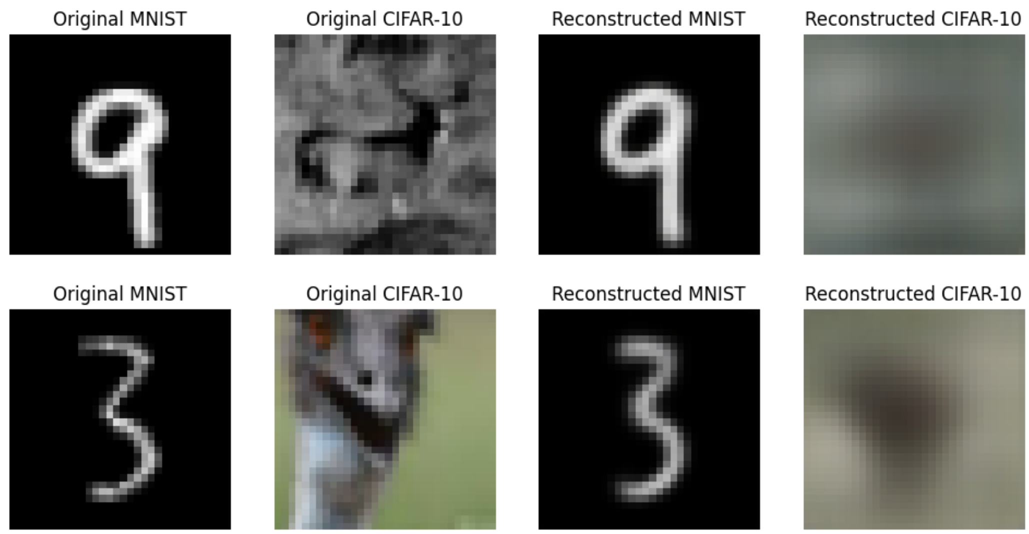


3. **Quantitative Evaluation**:

Metric	MNIST	CIFAR-10
Mean SSIM	0.8998	0.3735
Mean PSNR	22.54 dB	17.01 dB

4. **Visual Results**:

   - Original and reconstructed images are visually inspected to assess the quality of reconstructions.



Discussion

1. **Training Performance**:

    - The training loss decreased consistently over the epochs, indicating that the model was learning to reconstruct the images effectively.
    - The convergence of the training loss suggests that the autoencoder reached a stable state by the end of the training period.

2. **Reconstruction Quality**:

    - **MNIST Reconstructions**: High SSIM and PSNR values indicate that the reconstructed MNIST images preserved the structural and visual quality of the original images. The model effectively separated and reconstructed the MNIST features from the mean input.
    - **CIFAR-10 Reconstructions**: The lower SSIM and PSNR values for CIFAR-10 images suggest that while the overall shapes and colors were preserved, finer details were somewhat blurred. This highlights the complexity of reconstructing more detailed images from a combined input.

3. **Visual Inspection**:

- The visual inspection confirmed that MNIST digits were clear and distinct in the reconstructions, while CIFAR-10 images showed some blurring but remained recognizable.

Conclusion

The project demonstrated the feasibility of using an autoencoder to reconstruct images from a combined mean input of MNIST and CIFAR-10 images. The results highlight the potential of autoencoders in handling heterogeneous data and producing meaningful reconstructions. Despite the challenge posed by the blending of features from two different datasets, the autoencoder effectively separated and reconstructed the original images, particularly excelling with simpler MNIST digits.