Homework 4

Submitted by Vishantan Kumar
April 29, 2024

Part 1: Image Reconstruction

1.1 Model Architecture

The model architecture was as demonstrated in Figure 1

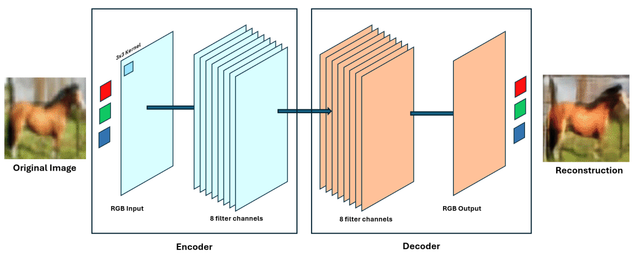

Figure 1. Autoencoder Architecture

The architecture is a compact convolutional auto-encoder tailored for the CIFAR10 dataset. The encoder consists of two convolutional layers with kernel size 3, stride 1, and no padding, mapping 3-channel input to an 8-channel feature space and then to an 8-channel latent representation, each followed by a ReLU activation. The decoder mirrors this structure using transposed convolutions to upsample from the latent space back to the image space, aiming to reconstruct the original input. The model employs a mean squared error loss for training, optimizing for fidelity in image reconstruction.

1.2 Method

The model was trained on 50,000 training datapoints and batch size of 64 for 10 epochs and a learning rate of 1e-3.

1.3 Results

The progressive loss function and Reconstructed images are given below

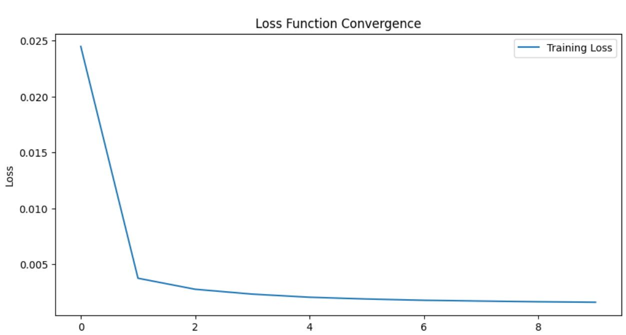

Figure 2. Training loss

We can see that the loss saturates at around 4 epochs, not having much improvement post that. In figure 3 we have demonstrated the reconstructed images from the CIFAR dataset.

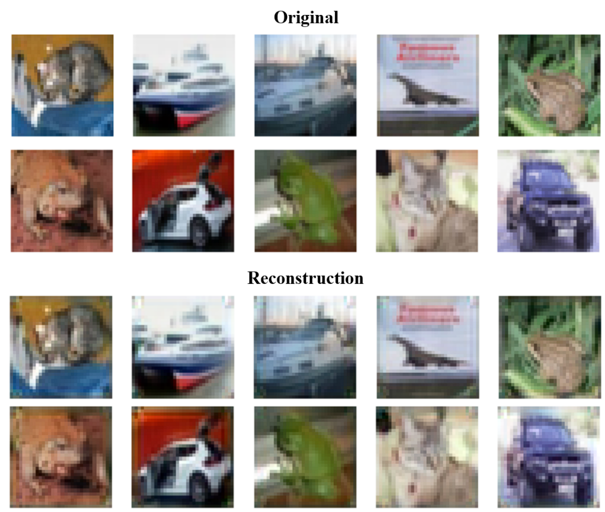

Figure 3. Reconstructed CIFAR Images

The reconstructions are fairly accurate, but have some artifacts in some images. Its possible that those can be removed by increasing the number of training epochs.

Part 2: Image Denoising

2.1 Model Architecture

The model architecture was kept similar to Part 1 with 4 convolutional layers, 2 in the encoder and 2 in the decoder.

2.2 Method

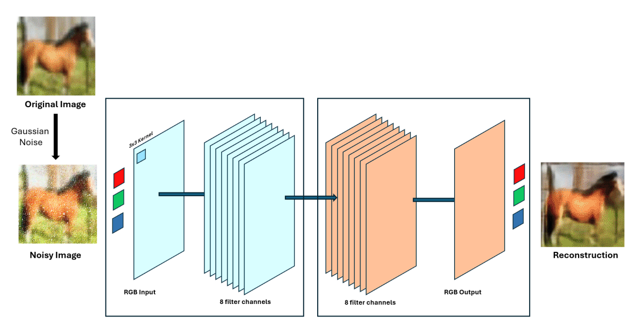

Figure 4. Autoencoder Architecture

The images were first normalised to lie between -1 and 1 and Gaussian noise with $\mu = 0$ and $\sigma^1 = 0.1$ was added to each image before feeding into the model for training. MSE loss was chosen as the optimization criterion with ADAM optimizer. While evaluating the model, we also calculated the Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) to check the model performance.

2.3 Results

The progressive loss function and Reconstructed images are given below

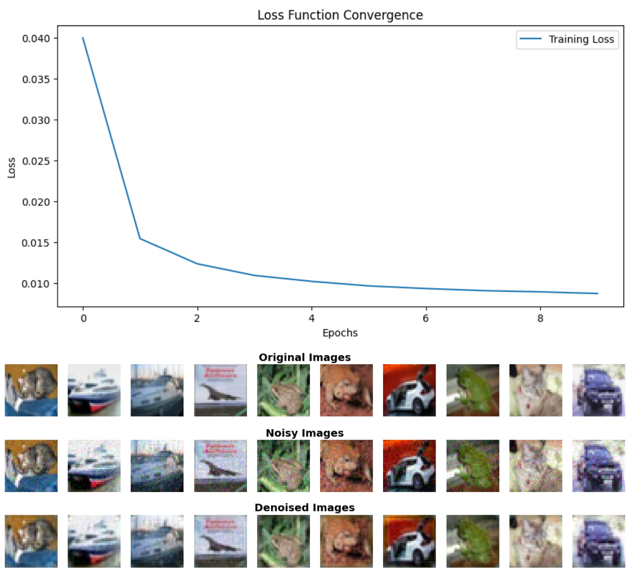

Figure 5. Training loss and Denoised Images

The loss function again saturated around 4-5 epochs. We saw a considerably good performance on the test set with a **PSNR of 26.78** and **SSIM of 0.92**. Since the SSIM was very close to 1, the reconstructed images were of comparable quality to the original images. Additionally, the training time was fairly less clocking around 200s on a 6GB RTX3060 Super GPU. Overall the performance was fairly good for both reconstruction and denoising.