Ayesha Fathima

Afathima20760@ucumberlands.edu

**Building an Autoencoder using Fashion-MNIST**

Professor. David Ostrowski

**Introduction**

The objective of this assignment is to design an autoencoder model that can reconstruct images from the Fashion-MNIST dataset. An autoencoder consists of two main parts: an encoder, which compresses the input image into a latent space representation, and a decoder, which reconstructs the image from this compressed representation. This assignment will focus on building, training, and evaluating an autoencoder using convolutional layers.

**Dataset: Fashion-MNIST**

The Fashion-MNIST dataset is a popular dataset consisting of grayscale images, each sized 28x28 pixels, of different fashion items such as shirts, trousers, shoes, and bags. There are 10 classes in the dataset, but for this assignment, we will not use the labels since the task is unsupervised. The training set contains 60,000 images, and the test set contains 10,000 images.

**Model Architecture**

The autoencoder consists of two parts:

**1. Encoder**

The encoder compresses the input image into a latent space by reducing its spatial dimensions using convolutional layers and max pooling. The architecture of the encoder is as follows:

* A convolutional layer with 32 filters, a kernel size of 3x3, ReLU activation, and 'same' padding.
* Max-pooling layer to reduce the spatial dimensions by a factor of 2.
* A second convolutional layer with 64 filters, a kernel size of 3x3, ReLU activation, and 'same' padding.
* Another max-pooling layer to further reduce the dimensions.
* A third convolutional layer with 128 filters, followed by flattening and a dense layer to map the output to a latent vector of size 64.

**2. Decoder**

The decoder reconstructs the compressed latent representation back to the original image size. The architecture of the decoder is:

* A dense layer to project the latent vector to a 7x7x128 feature map.
* Reshaping the output to match the feature map shape.
* Transposed convolutional layers with up sampling to increase the spatial dimensions back to 28x28.
* A final transposed convolutional layer to output the reconstructed image.

**Results**

The model was trained for 10 epochs, and the following images show the comparison between the original and reconstructed images. For the purpose of this assignment, we selected 5 images from the test set to evaluate the performance of the autoencoder.

**Observations:**

1. **Image 1: Shoes**
   * The general shape of the shoes is well preserved in the reconstructed image.
   * Some finer details, such as the outline, appear slightly blurred.
2. **Image 2: Long-sleeved Shirt**
   * The reconstructed image retains the shape and texture of the shirt.
   * However, the 'LE' text on the shirt is less clear compared to the original.
3. **Image 3: Trousers**
   * The trousers' shape and size are well reconstructed.
   * Minor details, such as folds, are slightly blurred.
4. **Image 4: Trousers**
   * Like Image 3, the overall shape is intact, but small textures are missing.
5. **Image 5: T-shirt**
   * The overall structure is preserved, but the texture and depth of the shirt in the original are more prominent than in the reconstruction.

A collage of images of a person's body

Description automatically generated

The reconstructed images generally capture the essential features of the original images. However, there is a noticeable reduction in the quality of finer details. This is typical behavior for simple autoencoders, which prioritize compressing the most important visual features while sacrificing some of the subtler ones.

A computer screen shot of a program

Description automatically generated

**Conclusion**

In this assignment, we successfully built an autoencoder using convolutional layers to reconstruct images from the Fashion-MNIST dataset. The encoder compresses the input into a 64-dimensional latent space, while the decoder attempts to reconstruct the original image from this compressed representation. The autoencoder was able to recreate the basic shapes and patterns of the fashion items, but with some loss of detail. This assignment demonstrates how autoencoders can be used for image compression and reconstruction tasks.

**REFERENCES**

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Krizhevsky, A., Hinton, G., & Nair, V. (2009). *Learning multiple layers of features from tiny images*. CIFAR-10 Dataset.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. <https://doi.org/10.1038/nature14539>

Agarap, A. F. (2018). Deep learning using rectified linear units (ReLU). *arXiv preprint arXiv:1803.08375*. <https://arxiv.org/abs/1803.08375>