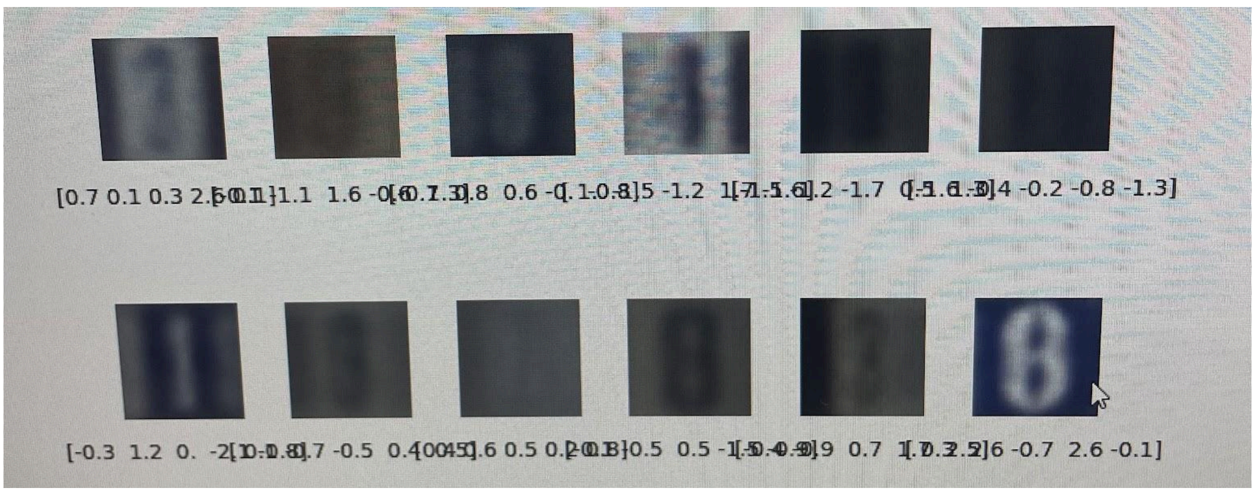
Generative Artificial Intelligence
Assignment 3
Sarah Groark
October 9, 2024

Original Images:

Reconstructed Images:



Discussion:

The variational autoencoder on the SVHN dataset poses some challenges. The model was trained and fit with several iterations of different parameter sets. For example –

Training #1:

- Beta = 1.0
- Epochs = 100
- Batch size = 64
- Learning rate = 0.005

Training #2:

- Beta = 0.5
- Epochs = 50

- Batch size = 128
- Learning rate = 0.001

Training #3:
- Beta = 2.0
- Epochs = 100
- Batch size = 64
- Learning rate = 0.01

The model's architecture featured a set up with an encoder and decoder. The encoder had four convolutional layers with 32, 64, 128, and 256 filters, respectively. Each layer also featured a batch normalization layer, activation layer, and drop out layer (with a rate of 0.3). The decoder, similarly, had three deconvolution layers with 256, 128, 64, and 32 filters. The decoder layers also had the batch normalization, dropout and activation layers.

Several features of a VAE are important to discuss when comparing an original and reconstructed dataset. First, the clarity is the most obvious, as the images provided in the training set will inevitably and most likely appear clearer than the reconstructed ones. Above, it is obvious that the reconstructed images lack clarity in comparison to the original images. However, there are foundational details in the reconstructed images that implies the model was able to learn some features of the original images. In most of the reconstructed images, you can see faint outlines of numbers. Additionally, the bottom right image is the most clear, featuring an easier eight to detect in the image.

Secondly, the reconstructed images also appear slightly distorted from the original image set. This implies that the model could use some fine tuning to better enhance its ability to extract features and details like edges, colors, shapes, and sizes of the images. The combination of parameter values such as beta, the learning rate, batch size, and the amount of epochs could be optimized to output the best possible results.

Next, the diversity of generalization plays an important role in assessing the success of a generative model. In what is seen above, the model attempts to reconstruct images that vary slightly. For example, there are blue numbers (some clearer than others) as well as lighter, more white-toned numbers that are outputted by the decoder. To allow the model to increase its diversification abilities, providing a larger latent space dimensionality encourages the model to explore a wider range of possibilities.

Finally, a model's ability to generalize is also a key factor in its effectiveness. In this specific case, if you take the clearest output as seen above (the blue 8), you can see how the model derived this new image by analyzing the original dataset images just above it. In the string of original images provided above, there are a few instances of the number 8. The model then took those features and manipulated them slightly to output the far bottom right 8 in the reconstructed images. For example, the size of the number increased, the outlines were made bolder, a blue background was given, and the number color changed to white.

Overall, this model, although not perfect by any means, was able to, to an extent, learn the high level features and details of the SVHN dataset. In the future, more experimentation and testing with larger parameters and different combinations of them could result in a higher accuracy of outputted, generated images that are more coherent and similar to the original set.