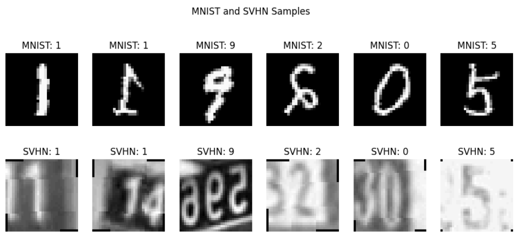
Morgan Yeazel November 21st, 2025

Objective:

The goal of this lab is to implement and analyze a joint variational autoencoder (VAE) capable of learning a shared latent representation between two different image domains, MNIST (handwritten digits) and SVHN (street view house numbers).

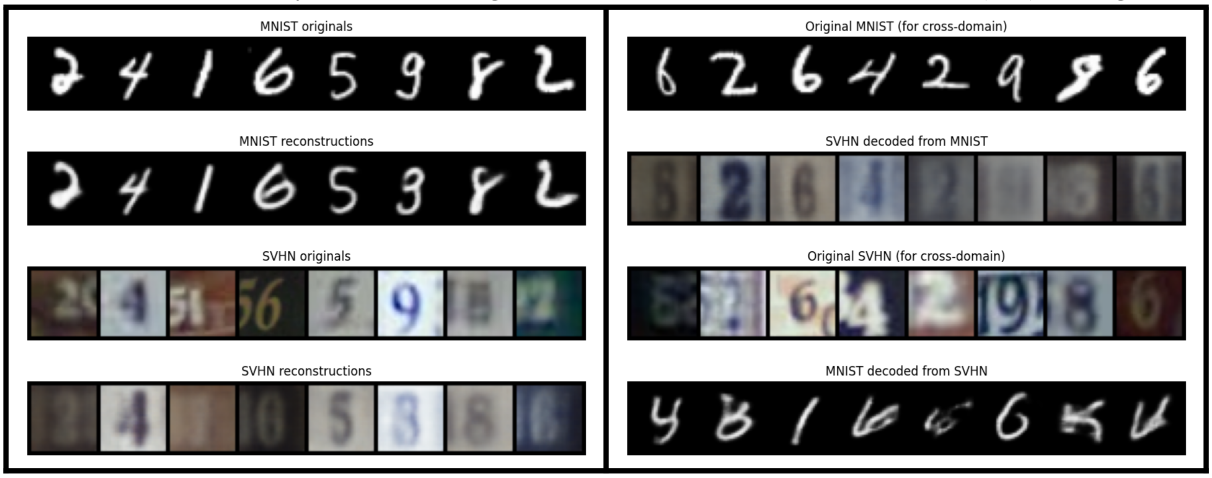
Dataset Preparation:

The MNIST and SVHN datasets were downloaded. MNIST images and converted to three channels to match SVHN's RGB format. Pixel values for both datasets were normalized to [0, 1]. Each MNIST digit was paired with an SVHN image of the same label, creating a multi-view dataset where each sample consisted of a matched MNIST–SVHN pair and its shared digit label.



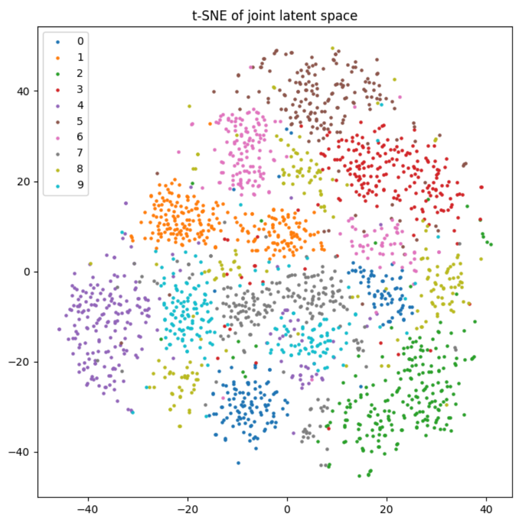
Model Training:

A Joint Variational Autoencoder (Joint VAE) was implemented with separate encoders and decoders for each domain, connected through a shared latent space. Encoders produced Gaussian distributions for each input, which were combined using a Product of Experts (PoE) approach to create a unified latent vector. Decoders reconstructed images in their respective domains. The model was trained to maximize the Evidence Lower Bound (ELBO), balancing reconstruction accuracy with latent regularization via the Kullback–Leibler (KL) divergence.



Display:

After training, the model's performance was evaluated by visualizing reconstructions for both MNIST and SVHN, generating new samples from random latent vectors, and performing cross-domain translations (e.g., encoding MNIST images and decoding as SVHN). The learned latent space was further analyzed using t-SNE, which showed how samples from both datasets clustered according to digit labels and how well the shared representation aligned the two domains.



Key Takeaways:

Effective joint training:
- The Joint VAE successfully learned to reconstruct both MNIST and SVHN images.
- The ELBO loss guided the model to balance reconstruction accuracy with latent regularization.

Latent space alignment:
- t-SNE visualization showed clear clustering of digits across both datasets.
- Shared latent space captures meaningful, domain-independent features.

Cross-domain generation:
- Encoding an image from one domain and decoding it in the other produced plausible cross-domain outputs.
- Demonstrates the latent space supports information transfer between MNIST and SVHN.

Generative and reconstruction quality:
- Reconstructions preserved key digit features.
- Random sampling from the latent space generated realistic images in both domains.

Conclusion:
The Joint VAE successfully reconstructed MNIST and SVHN images while learning a shared latent space that clusters digits across both datasets. Cross-domain generation demonstrated the latent space captures meaningful, domain-independent features. Overall, the model effectively supports reconstruction, generation, and alignment of multi-view data.

Google Colab:
https://colab.research.google.com/drive/1LKDVuiLZu80Gu0P3C9HWwA5HE3IDiwnE?usp=sharing

Github:
https://github.com/mryeazel-729/MLHealth/blob/main/Lab3.ipynb