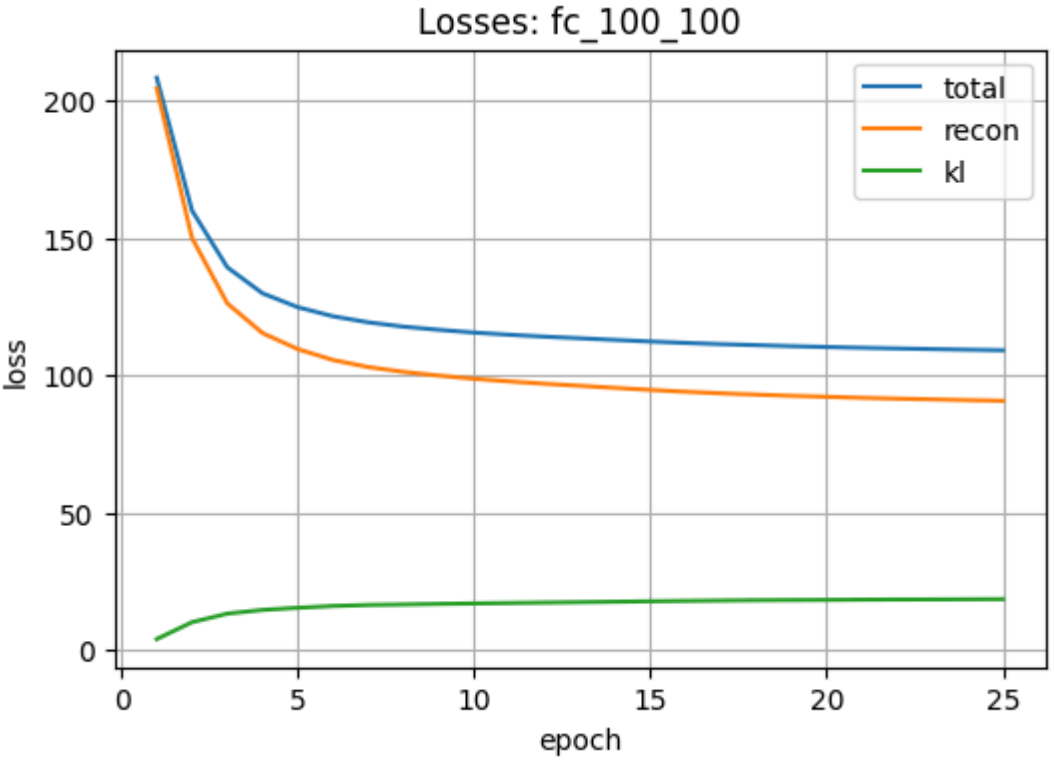
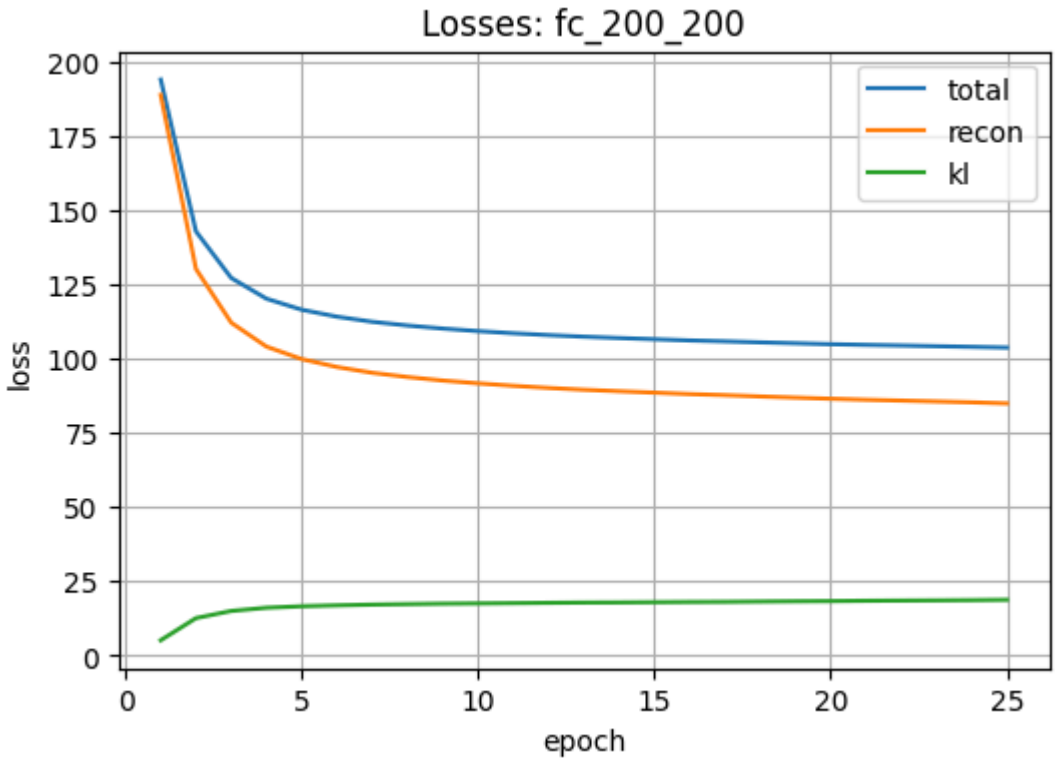
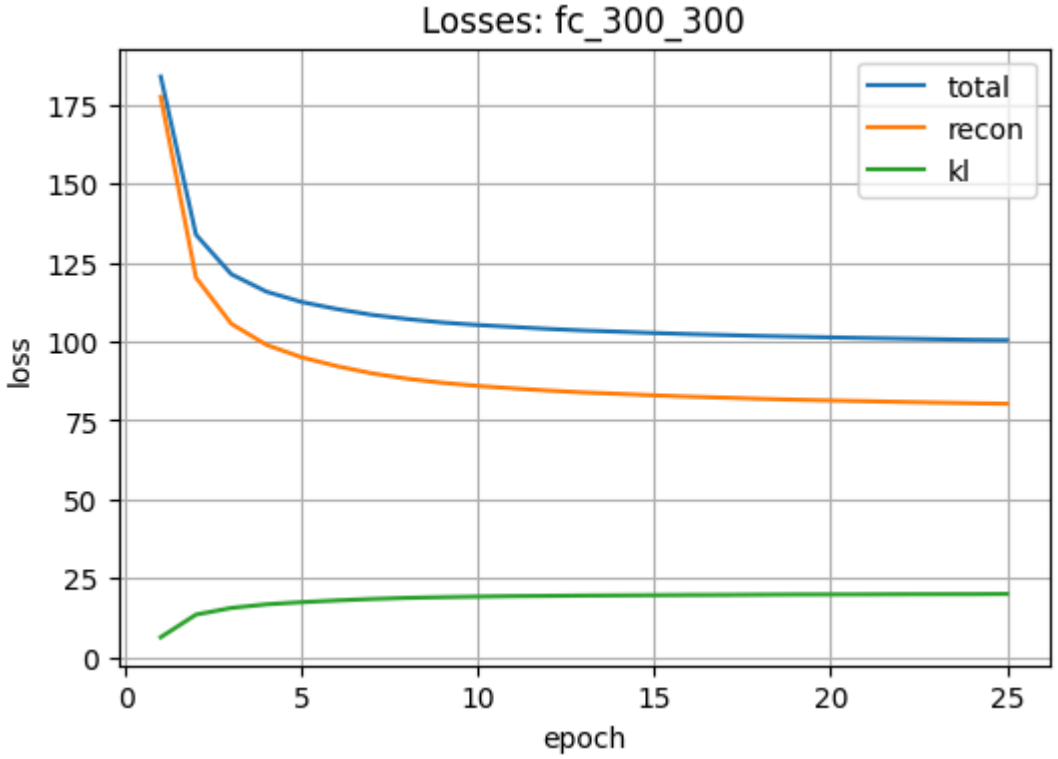
# 1. Fully Connected VAEs (Hidden sizes: 100–100, 200–200, 300–300)

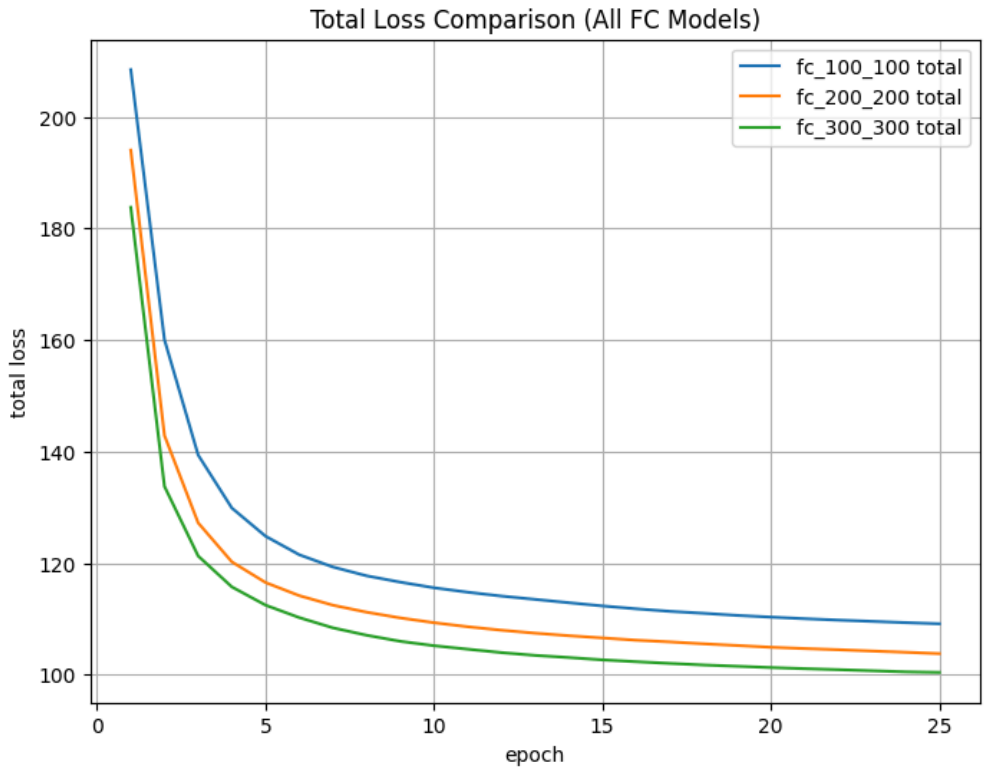
## Training Setup

- Dataset: MNIST (28×28 grayscale)
- Latent dimension: 20
- Three FC-VAEs trained for 25 epochs each with hidden layers:
    - Model A: (100, 100)
    - Model B: (200, 200)
    - Model C: (300, 300)
- Loss Curves

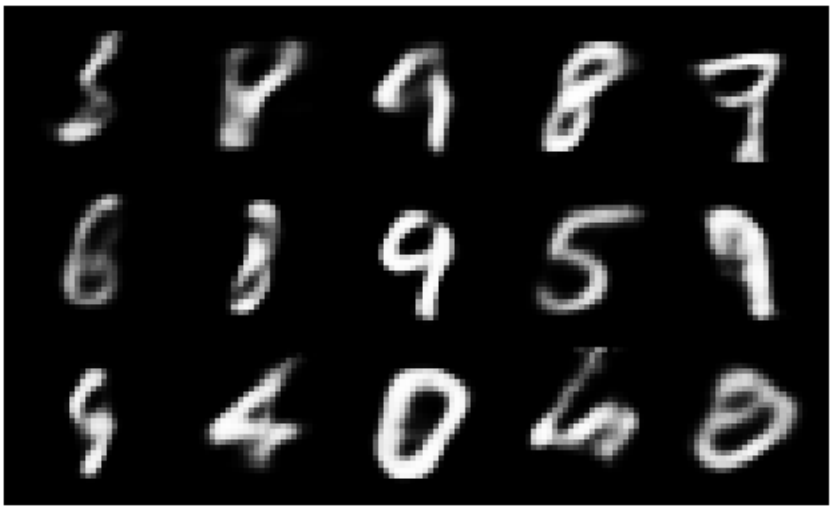



Losses: fc_200_200
total
recon
kl
loss
epoch
0
25
50
75
100
125
150
175
200
0
5
10
15
20
25

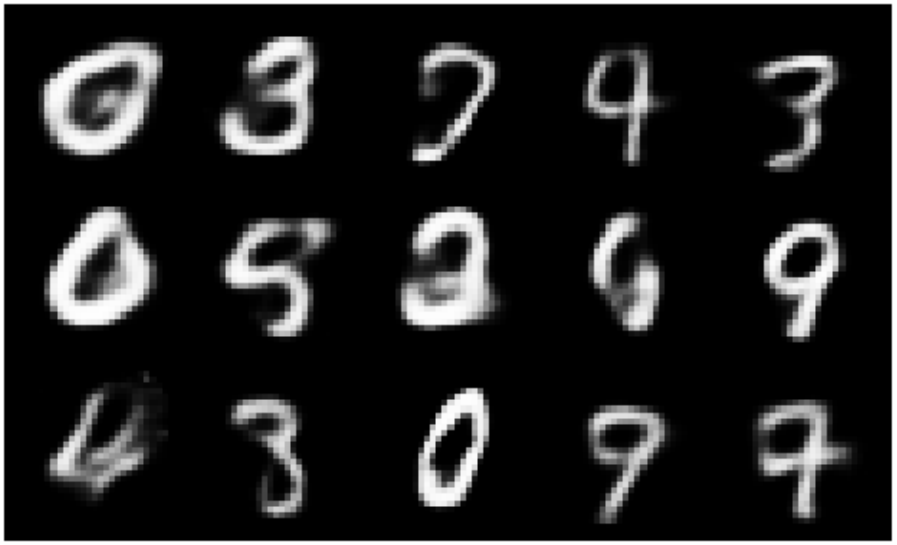

Losses: fc_300_300
total
recon
kl
loss
epoch
0
25
50
75
100
125
150
175
0
5
10
15
20
25



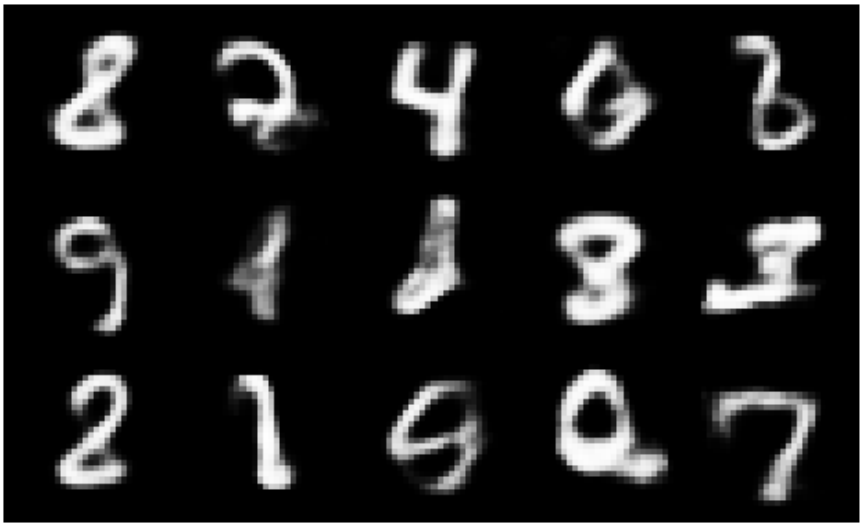

- Randomly generated samples
  - For fc 100,100



- For fc 200,200



- For fc 300,300



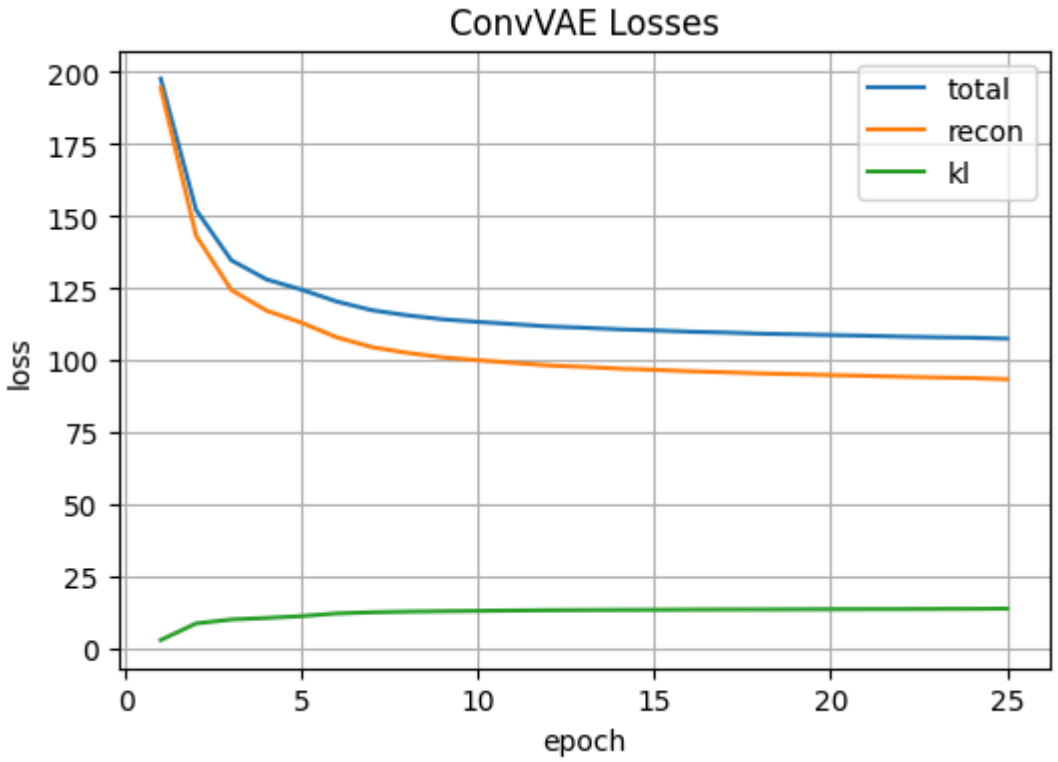
All three fully connected VAEs successfully reconstruct recognizable digits. Models with larger hidden layers show smoother and more detailed generations, indicating higher representational capacity. The 300–300 model maintains the lowest reconstruction loss and produces the cleanest samples. KL divergence stabilizes after ~10 epochs, showing the latent space converges to a Gaussian prior. Overall, increasing network size improves reconstruction fidelity.

# 2 & 3. Convolutional VAE

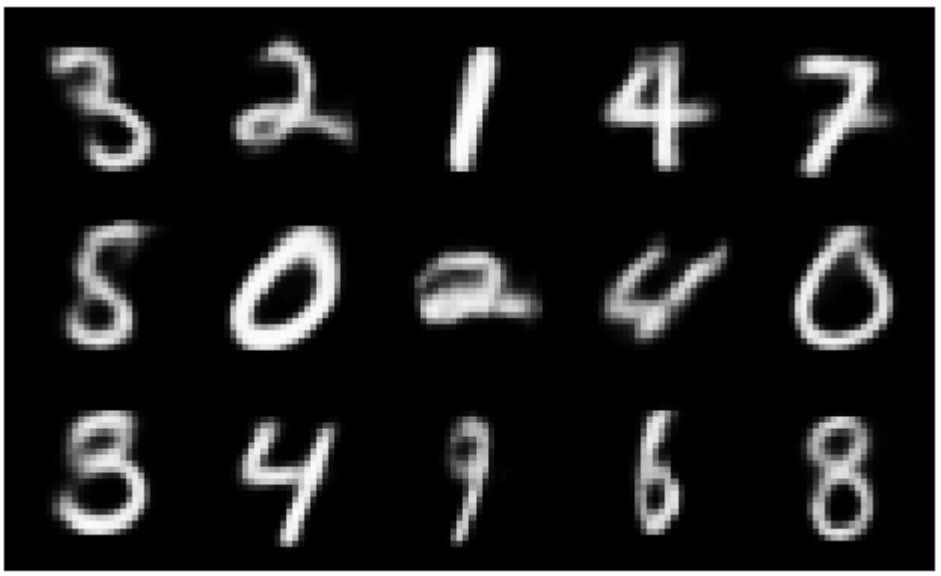
## Architecture

- Conv1: 32 filters, 4×4, stride 2
- Conv2: 64 filters, 4×4, stride 2
- FC encoder: 256 → 128 → μ/logσ²
- FC decoder: 128 → 256 → reshape → two deconvs (32, 1 filters)

## Loss Curve



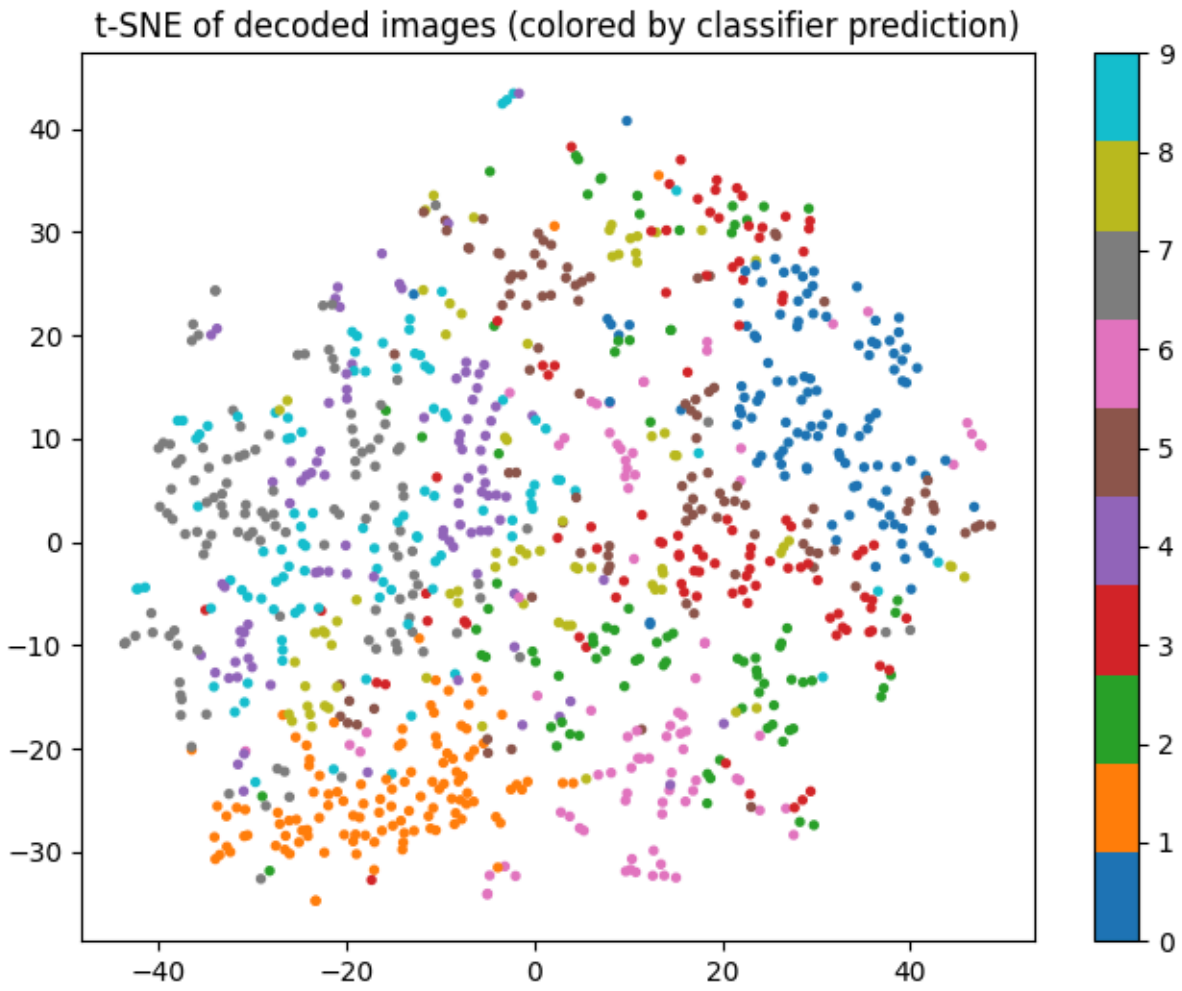

**Generated Samples (15 prior samples)**



The convolutional VAE produces much sharper digits than fully connected VAEs because convolution preserves spatial structure. Loss decreases faster and stabilizes earlier. Reconstructions exhibit clear stroke continuity and realistic MNIST patterns. The learned latent space is more structured, which improves sample quality from the standard Gaussian prior.

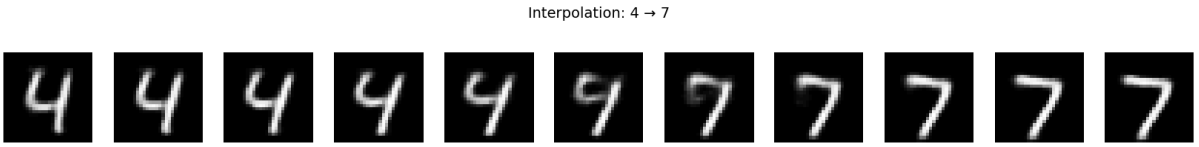
Sampling from z~N(0,I) yields clean digit-like outputs, confirming that the encoder successfully shapes the latent distribution to match the prior during training. Most digits are well-formed, but occasional ambiguous shapes reflect regions of latent space with lower data density. Overall, the model generalizes well to unseen latent points.

**4. t-SNE on Decoded Samples**




The t-SNE embedding forms distinct clusters corresponding to predicted digit classes, showing that the VAE's latent space organizes digits into semantically meaningful regions. Some overlap exists between visually similar digits (e.g., 3 and 8), reflecting shared structural features. The smooth cluster borders indicate that decoding varies gradually with latent changes, confirming continuity of the learned manifold.

**5. Latent Space Interpolation**




The interpolation shows a smooth transformation from digit 4 to digit 7, demonstrating the continuity of the learned latent manifold. Intermediate images gradually change shape without sudden artifacts, suggesting that similar digits lie near each other in latent space. This smooth progression confirms that the VAE has learned a meaningful, structured latent representation rather than merely memorizing samples.

**Conclusion:**

**The VAE models effectively learn compressed latent representations of MNIST digits. Fully connected VAEs produce reasonable reconstructions, but the convolutional VAE performs significantly better, generating smoother and more realistic samples. Latent space exploration through prior sampling, t-SNE, and interpolation confirms that the model learns a coherent, continuous manifold. Overall, the convolutional VAE captures underlying digit structure more effectively than its FC counterparts.**