Latent Geometry with MNIST Torus VAE

One-Page Summary

Key Questions Addressed

Theme	Question
Latent Geometry	Can a 2D toroidal latent space improve interpretability and class separation in variational autoencoders?
Shock Detection	Can toroidal VAEs identify and rank structured shocks using latent geometry and reconstruction loss?
Cross-Domain Potential	Can this approach extend to domains like healthcare or finance, where cyclicity and structured anomalies are key?

Conceptual Ideas Proposed

- A **2D toroidal latent space** is implemented via sine–cosine embeddings of two angular latent variables, enabling structured clustering and interpretable geometry in VAEs.
- **KL annealing with a cosine schedule** balances exploration and regularization, preventing posterior collapse and ensuring effective use of the toroidal latent space.
- **Latent shock modeling** is performed using EMNIST letters as structured out-of-distribution inputs, with latent displacement and reconstruction loss acting as interpretable anomaly scores.
- The method is designed as a **minimal modification** to standard VAE pipelines, retaining Gaussian priors and reparameterization while introducing angular structure.

Key Results

- **Digit classes form distinct clusters** in 2D toroidal latent space, with proximity reflecting visual similarity (e.g., "3", "8", and "5" cluster together).
- **Visually dissimilar letters (K, W, X, Z)** embed in low-density regions with high reconstruction losses, acting as clear out-of-distribution shocks.
- **Structurally similar letters (B, D, P, I)** fall near digit clusters with moderate reconstruction loss, enabling nuanced anomaly ranking.
- **Reconstruction loss histogram** shows clear separation: digits peak at 150–300, similar letters have a broader spread, and dissimilar letters peak higher, validating latent-space-driven anomaly scoring.

Illustrative Figures

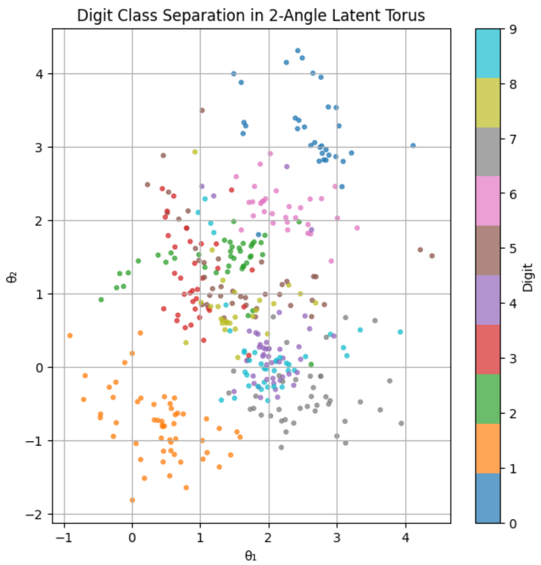

Figure 1: **Digit class separation in 2-angle latent torus.** MNIST digits form structured clusters, with proximity reflecting visual similarity.

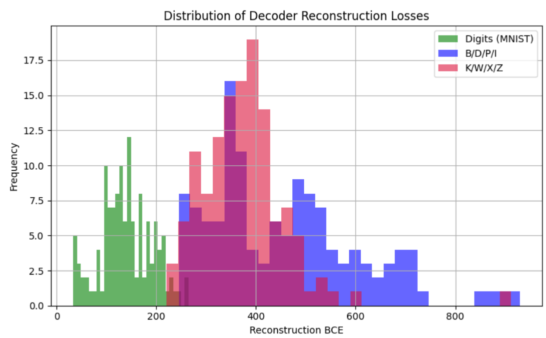

Figure 2: **Distribution of reconstruction losses across input categories.** Digit samples concentrate at lower loss values; structured shocks (B/D/P/I) and dissimilar shocks (K/W/X/Z) show progressively higher loss distributions.