Autoencoder Analysis of Zebrafish Embryo Morphogenesis

Pablo Escobar

November 26, 2025

Motivation and Objectives

- Build a compact latent representation of zebrafish embryo images.
- Evaluate reconstruction quality and generalization.
- Analyze latent structure using PCA and UMAP.
- Compare control vs mutant phenotypes.

Dataset

- Training embryos: Control1, Mutant1, Mutant2.
- Test embryos: Control2, Mutant3, Mutant4.
- TIFF stacks ($T \times 200 \times 200$).
- Metadata: embryo_id, frame_idx.

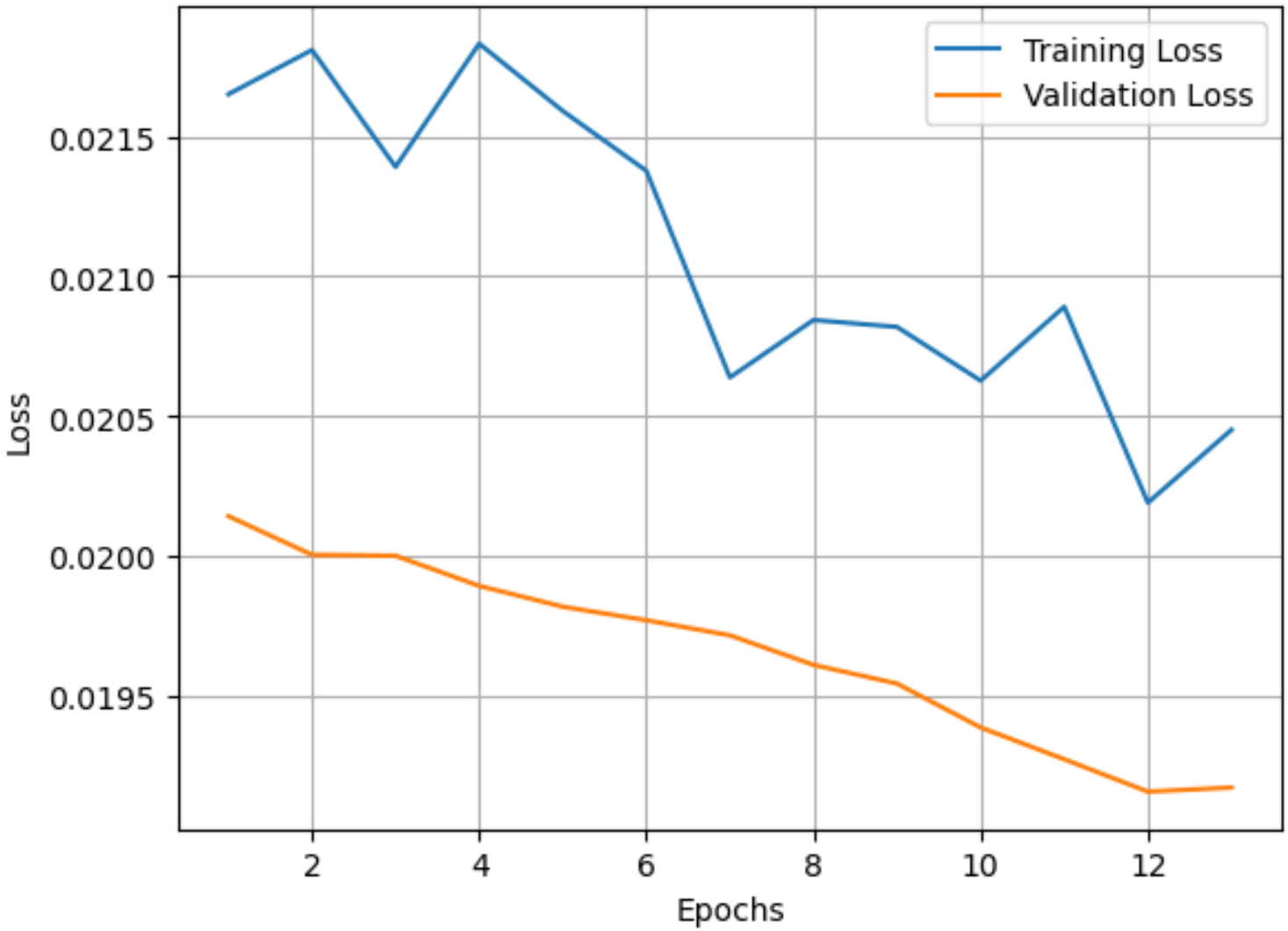
Model Architecture

- Convolutional Autoencoder.
- Encoder: $1 \rightarrow 16 \rightarrow 32 \rightarrow 64$ channels.
- Latent dimension: 36.
- Decoder mirrors encoder.
- Loss: MSE.

Training Setup

- 80/20 train/val split.
- Augmentations: flips, rotation, jitter.
- Early stopping with patience 1.
- Best model checkpoint saved.

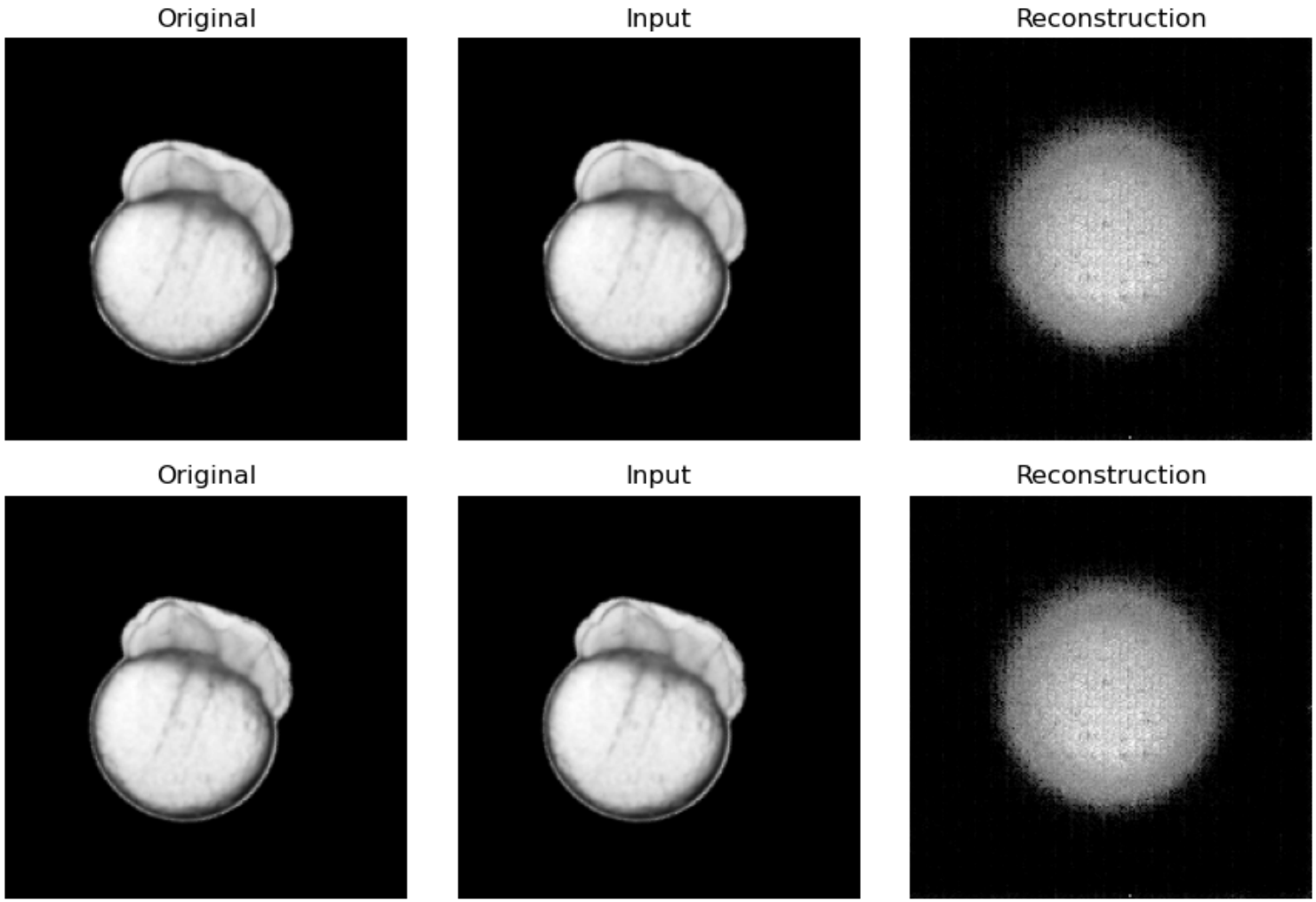
Training and Validation Loss



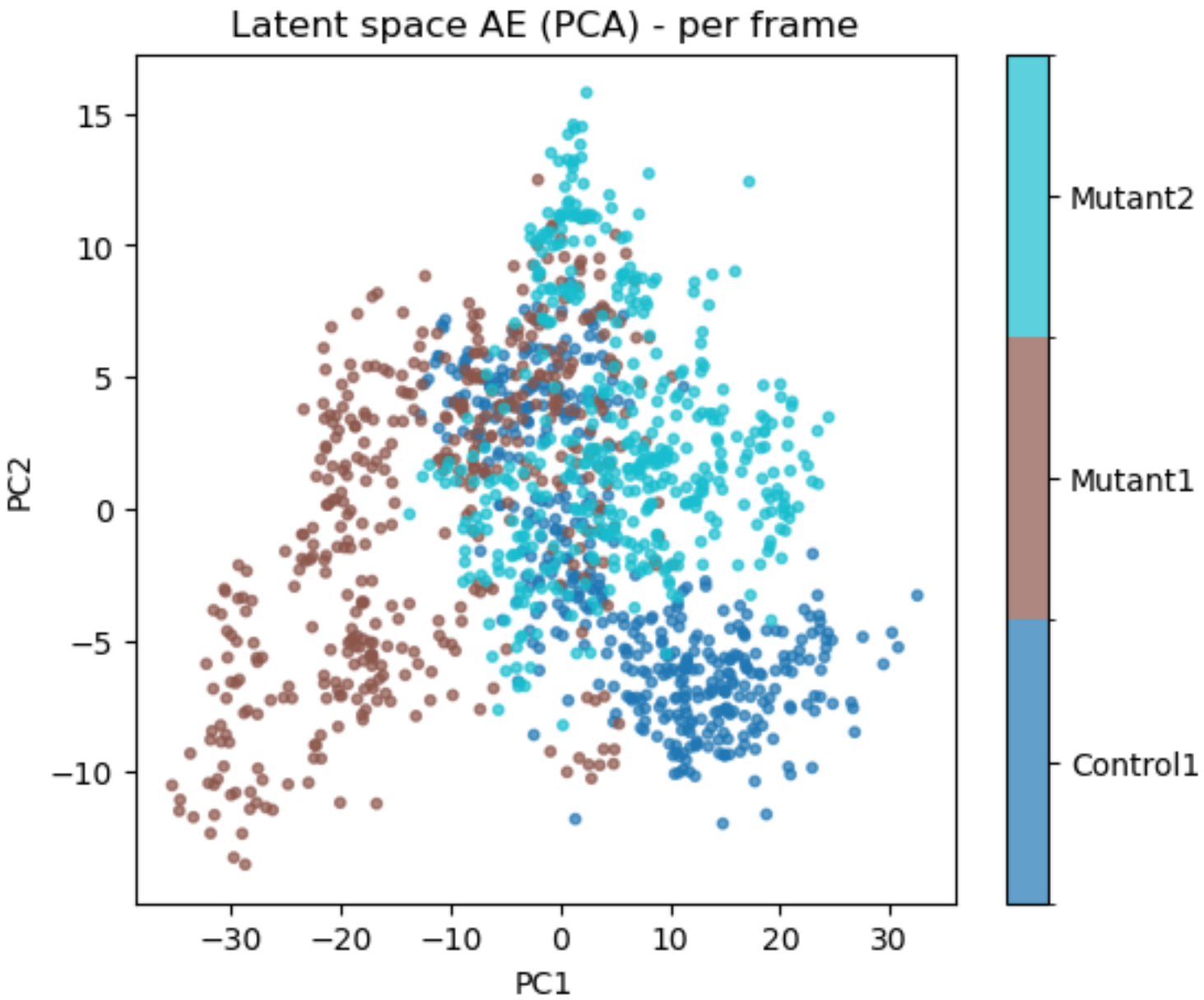
Quantitative Evaluation

- Train MSE: 0.02046
- Val MSE: 0.01916
- Test MSE: 0.02771

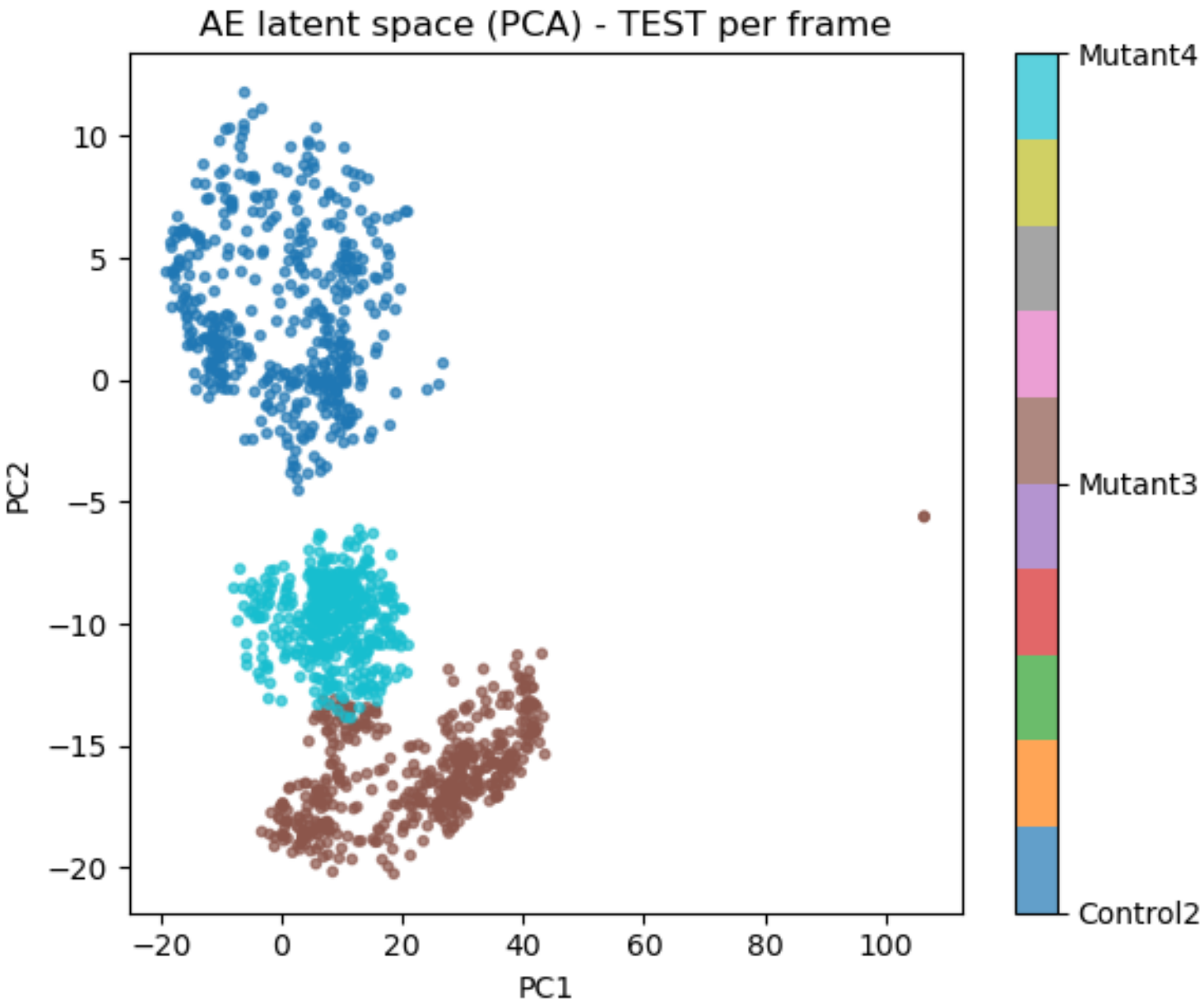
Reconstruction Examples



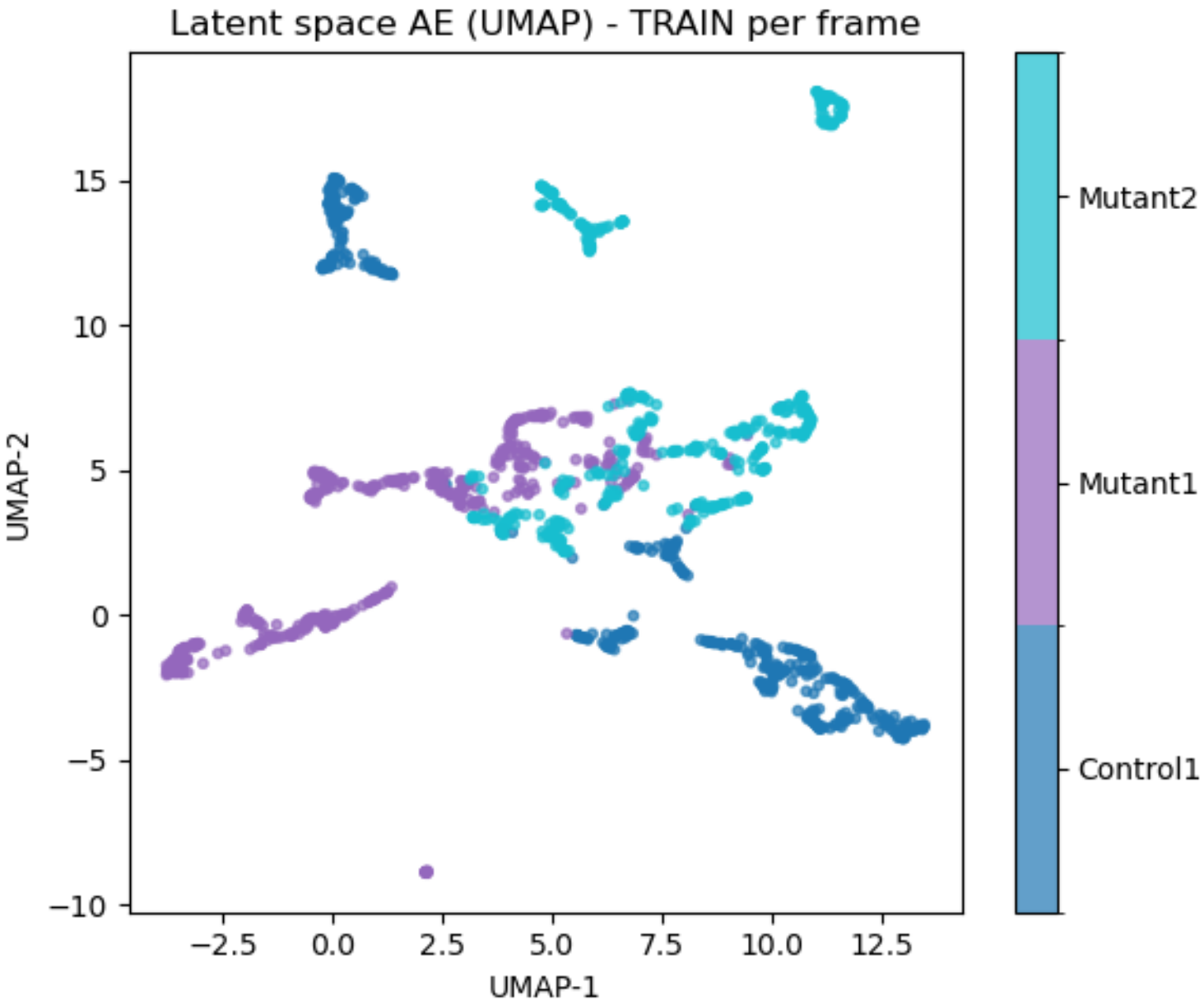
Latent Space: PCA (Train)



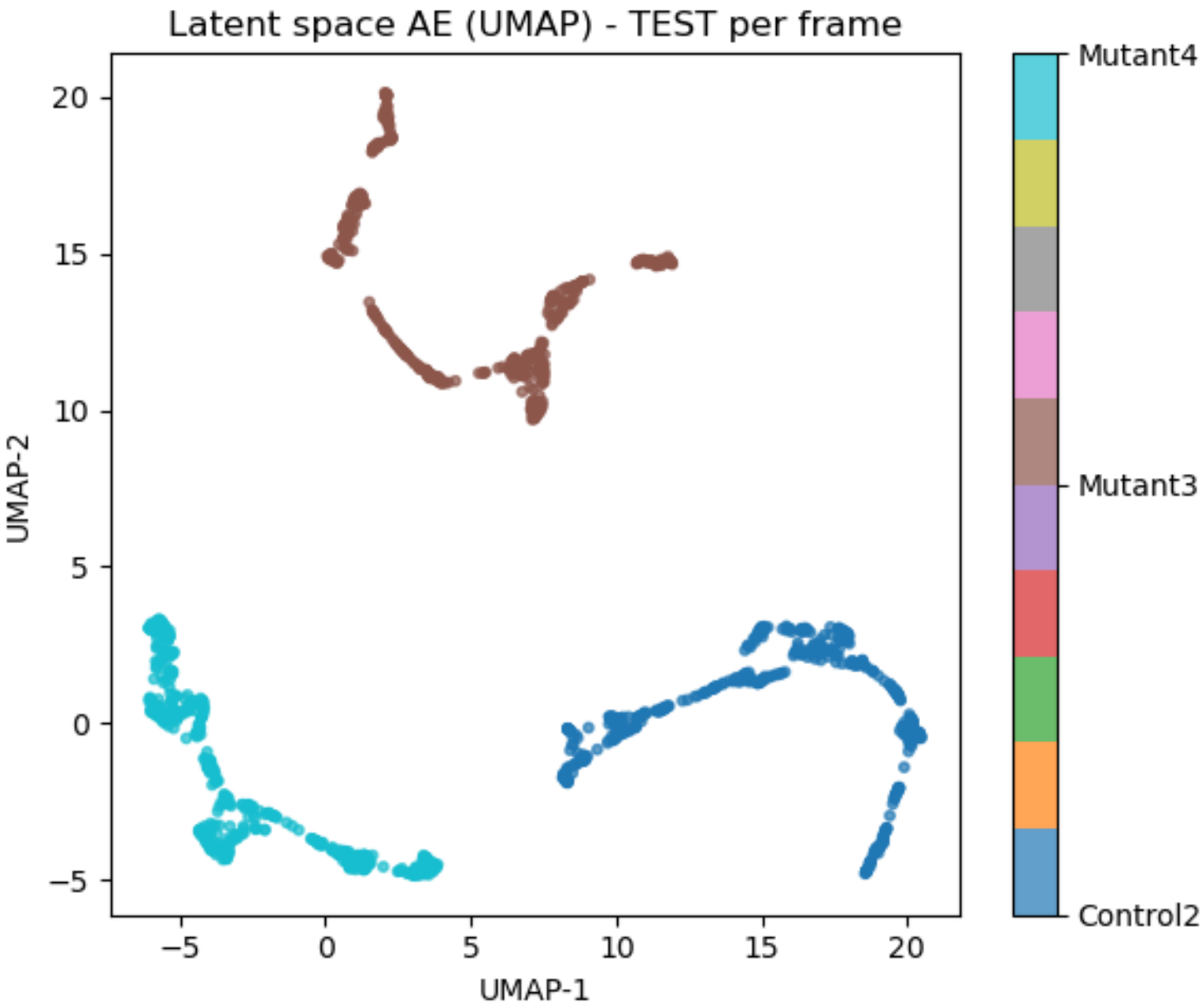
Latent Space: PCA (Test)



Latent Space: UMAP (Train)



Latent Space: UMAP (Test)



Biological Interpretation

- AE learns coarse morphology and developmental progression.
- Latent trajectories follow embryo time evolution.
- Clear separability between mutant and control embryos.

Limitations

- Reconstructions are blurry.
- Test embryos harder to reconstruct (domain shift).
- AE loses fine morphological detail.

Future Work

- Supervised AE - VAE.
- Perceptual and SSIM losses.
- Try and add the complete data for the original work.
- Start with the writing