Learning embryonic development as latent morphology trajectories

Pablo Escobar

YachayTech

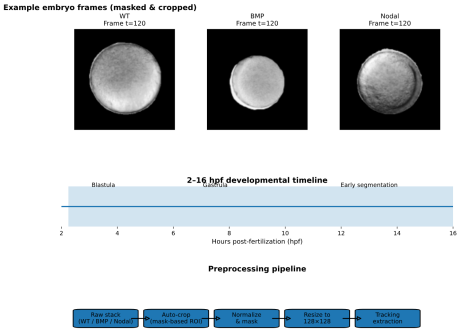
11/12/2025

Motivation

- We analyze zebrafish development from 2–16 hpf: blastula $\rightarrow$ gastrulation $\rightarrow$ early segmentation.
- Can we learn a compact morphology representation that:
  - organizes development as a smooth trajectory, and
  - separates WT vs BMP vs Nodal from morphology alone?
- This is still a challenge, because embryo phenotypes are high-dimensional (images over time), hard to compare directly.

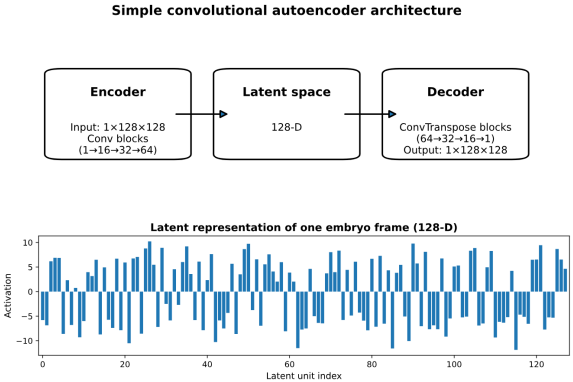
In this part, we build representation learning + latent-space analysis (no prediction).

Figure 1 — Dataset and preprocessing



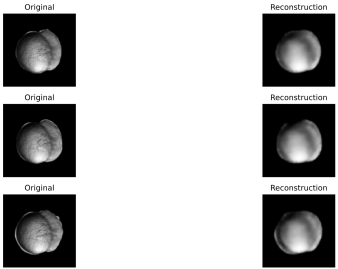
- Input: time-lapse embryo tracks (WT, BMP, Nodal) across 2–16 hpf.
- Preprocessing makes learning consistent: masking, embryo-centered crop, normalization, fixed resolution.
- Model focuses on embryo morphology (not background).

Figure 2 — Representation model (autoencoder)



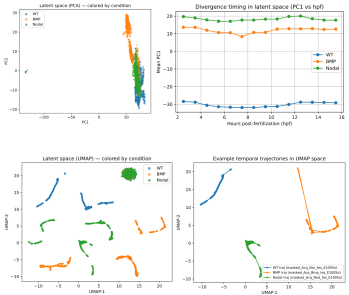
- Convolutional encoder compresses each 128×128 frame into a 128-D latent vector.
- Decoder reconstructs the image (reconstruction is a check for representation quality).
- Latent code is distributed: meaning comes from geometry (PCA/UMAP), not single units.

Figure 3 — Reconstruction quality & stage sensitivity



- Originals and reconstructions show global morphology is preserved.
- Error vs hpf: reconstruction difficulty changes with developmental stage (need to add image)
- The learned representation is sensitive to morphogenetic complexity (e.g., gastrulation).

Figure 4 — Latent space organization (PCA/UMAP)



- PCA/UMAP reveal emergent separation of WT, BMP, and Nodal from morphology.
- Trajectories: development forms smooth paths in latent space.
- Divergence timing: supports biological interpretation:
  - Nodal deviates earlier (gastrulation)
  - BMP deviates later (posterior shaping)

What to remember

- Autoencoder learns a compact morphology representation across 2–16 hpf.
- Latent space organizes development as continuous trajectories.
- Genotypes separate in latent geometry (PCA/UMAP), and divergence timing aligns with known biology.

Latent morphology trajectories provide a practical way to quantify developmental differences under genetic perturbations.