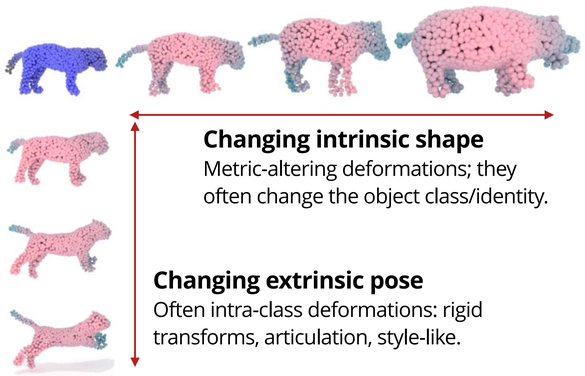
Geometric Disentanglement for Generative Latent Shape Models

Tristan Aumentado-Armstrong, Stavros Tsogkas, Allan Jepson, Sven Dickinson
University of Toronto · Vector Institute for AI · Samsung AI Center, Toronto

Factorizing pose & shape

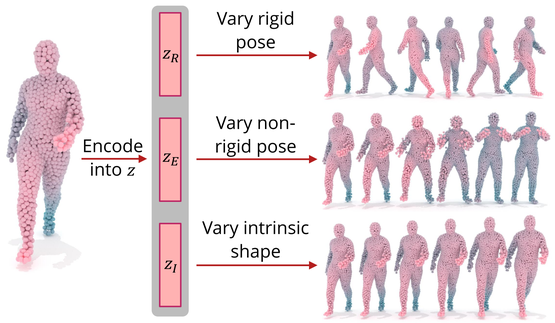
- Enables vision, graphics, & robotics tasks (e.g., pose-invariant recognition, manipulation, constrained inference, retrieval, pose transfer).
- Spectral methods also factorize pose and shape by separating **ext**rinsic from **int**rinsic geometry.



Changing intrinsic shape
Metric-altering deformations; they often change the object class/identity.

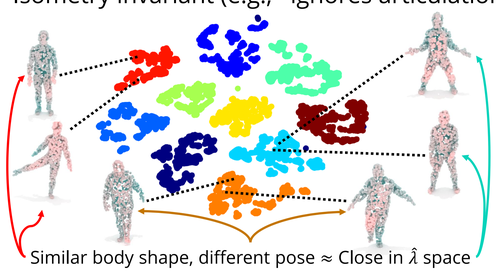
Changing extrinsic pose
Often intra-class deformations: rigid transforms, articulation, style-like.

Goal: disentangle *extrinsic pose* and *intrinsic shape* in the latent representation of a 3D point cloud, *without annotations*.

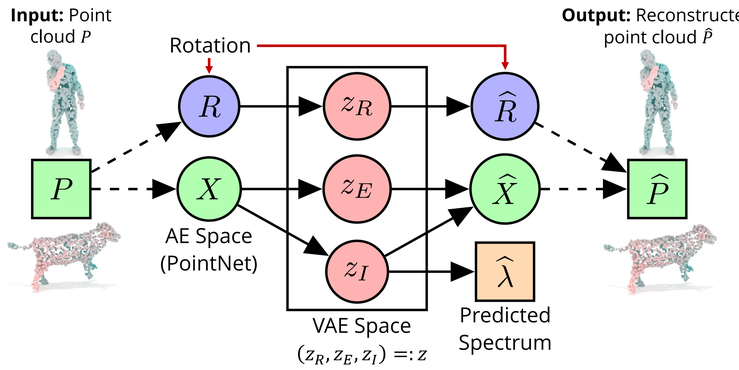


Background: Intrinsic Geometry

- Laplace-Beltrami Operator (LBO): Δ_g
- $LBOSpectrum(shape) \approx \hat{\lambda}$:
  - Captures intrinsic geometry (vector signature).
  - Isometry invariant (e.g., ~ignores articulation).



Geometrically Disentangled VAE: Model Architecture



- Model: two-level VAE (as in [1]).
- Pretrain AE space independently.
- Disentanglement with constricted $\dim(z)$ forces pose into z_E.

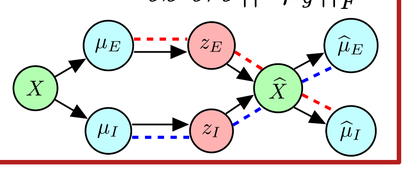
VAE Loss Function

$$\mathfrak{L} = \underbrace{\mathcal{L}_{\text{HF}} + \mathcal{L}_{\text{COV}} + \mathcal{L}_J}_{\text{Disentanglement penalties}} + \underbrace{\mathcal{L}_S}_{\text{Spectral loss}} + \underbrace{\mathcal{L}_V}_{\text{Reconstruction}}$$

Prior Matching

Penalty on Jacobian of each latent group with respect to another:

$$\mathcal{L}_J = \gamma_J \max_{g,\tilde{g}\,:\,g\neq\tilde{g}} \left\| \frac{\partial \hat{\mu}_g}{\partial \mu_{\tilde{g}}} \right\|_F^2$$

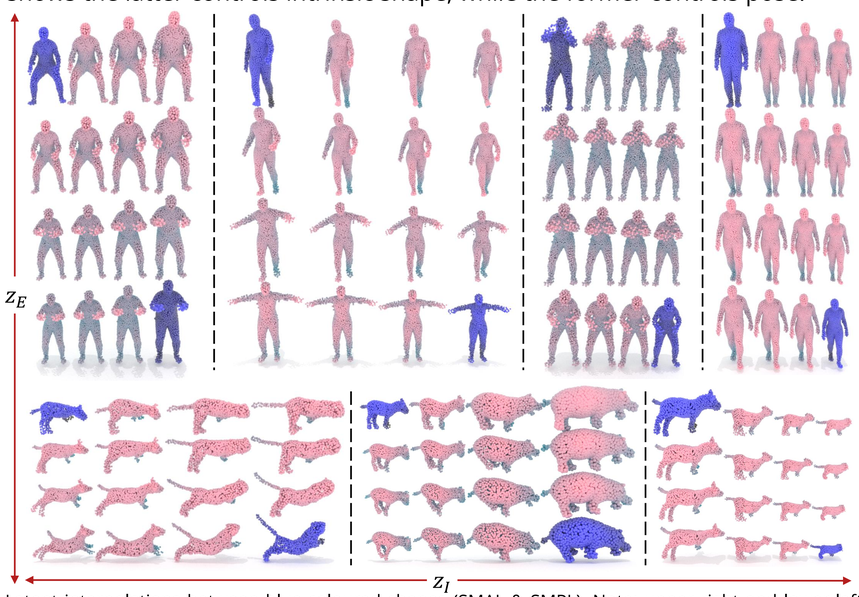


Hierarchically Factorized (HF) VAE loss [2]:

$$\mathcal{L}_{\text{HF}} = \beta_1 \text{TC}_{\text{intra}} + \beta_2 P_{\text{DKL}} + \beta_3\, \mathcal{I}[x; z] + \beta_4 \text{TC}_{\text{inter}}$$

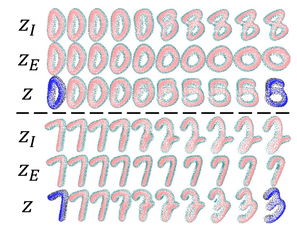
Results: Latent Manipulations

Datasets: point clouds derived from MNIST, DYNA, SMAL, and SMPL.
Interpolations: visualizing movement in z_E and z_I independently largely shows the latter controls intrinsic shape, while the former controls pose.



Latent interpolations between blue-coloured shapes (SMAL & SMPL). Note: upper-right and lower-left shapes are latent *pose/deformation transfers*. Vertical: movement in z_E ; horizontal: movement in z_I.

Latent digit subspace traversal can roughly separate shape and style. Each row traverses the marked set.



VAE Operations

Generation/Sampling

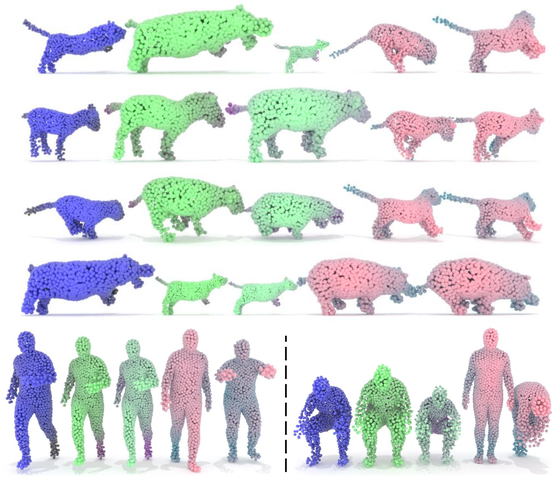
Autoencoding

Results: Pose-Aware Retrieval

- SMAL and SMPL have separate shape β and pose θ parameters, so we can compute separate retrieval errors for each (E_β and E_θ).
- Ideally, z_I should have *high* E_θ and *low* E_β; z_E should have *high* E_β and *low* E_θ.
- Comparisons: full AE X and VAE z latent vectors.

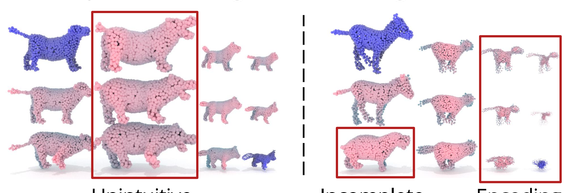
Errors		X	z	z_E	z_I
SMAL	E_β	0.641	0.743	0.975	0.645
	E_θ	0.938	0.983	0.983	0.993
SMPL	E_β	0.856	0.922	0.997	0.928
	E_θ	0.577	0.726	0.709	0.947

Query · Retrieval with z_E · Retrieval with z_I



Failure Modes

Observed trade-off between reconstruction fidelity, prior matching, and disentanglement.



Unintuitive intermediates · Incomplete disentanglement · Encoding failure

References

[1]: Achlioptas et al, *Learning Representations and Generative Models for 3D Point Clouds*, ICLR, 2018.
[2]: Esmaeili et al, *Structured Disentangled Representations,* AISTATS, 2019.