Graph Geometry-Preserving Autoencoders

Jungbin Lim*[1] Jihwan Kim*[1] Yonghyeon Lee[2] Cheongjae Jang[3] Frank Chongwoo Park[1]

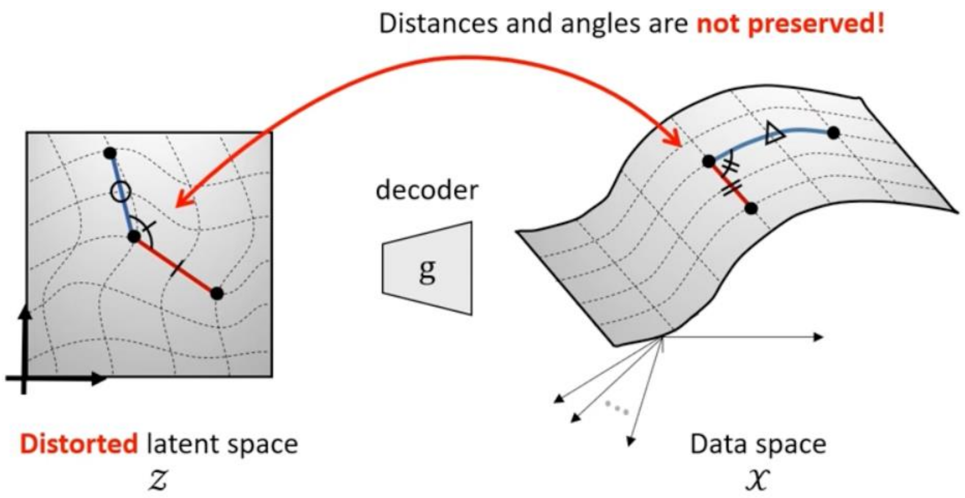
$$\mathcal{L}_{\mathrm{GGAE}}(\theta, \phi) = \underbrace{\frac{1}{N}\sum_{i=1}^{N} \|x_i - (g_\phi \circ f_\theta)(x_i)\|^2}_{\text{standard autoencoder reconstruction}} + \alpha \overbrace{\frac{1}{N}\sum_{i=1}^{N} \mathrm{Tr}\left(\tilde{H}_i(L, f_\theta(X))^2 - 2\,\tilde{H}_i(L, f_\theta(X))\right)}^{\text{distortion term}}$$

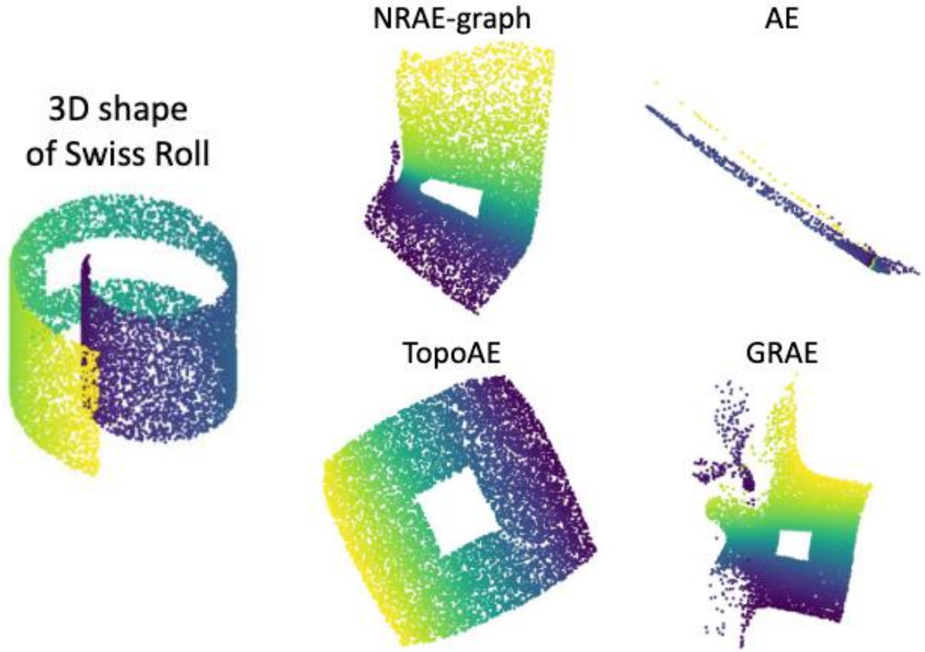
Contents

- Riemann Manifold and Isometry
  - Laplace-Beltrami Operator
  - Estimation by Graph Laplacians
- Distortion Loss
  - Minibatch evaluation of distortion
  - Intuition on Jacobian/Laplacian based loss
- Experiments
- Further Readings
- Possible Applications

Why Geometry-Preserving?

- Naïvely trained autoencoders(AEs) do not preserve locality (or local geometry) in latent space.





What is Geometry-Preserving?

- Preserving of local-geometry can be defined as ...

“Isometry” in Riemannian Geometry

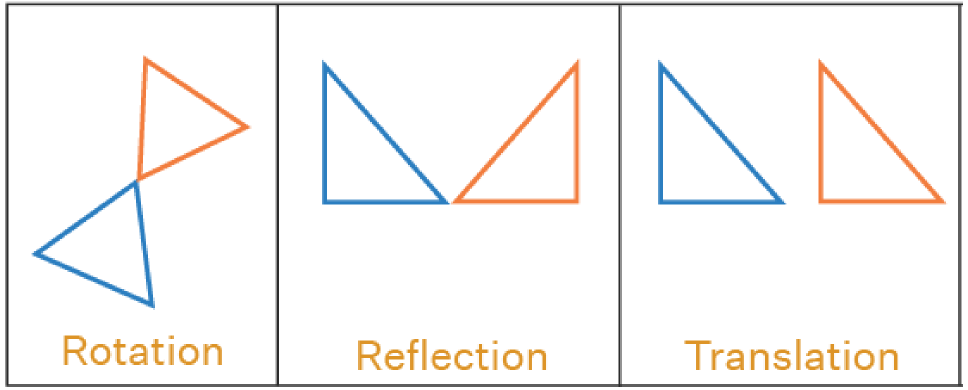
- Prerequisites …
  - Isometry
  - Riemann Geometry, Invariance
  - Graph based discretization
  - Laplacian

Isometry

- In Euclidean Geometry …

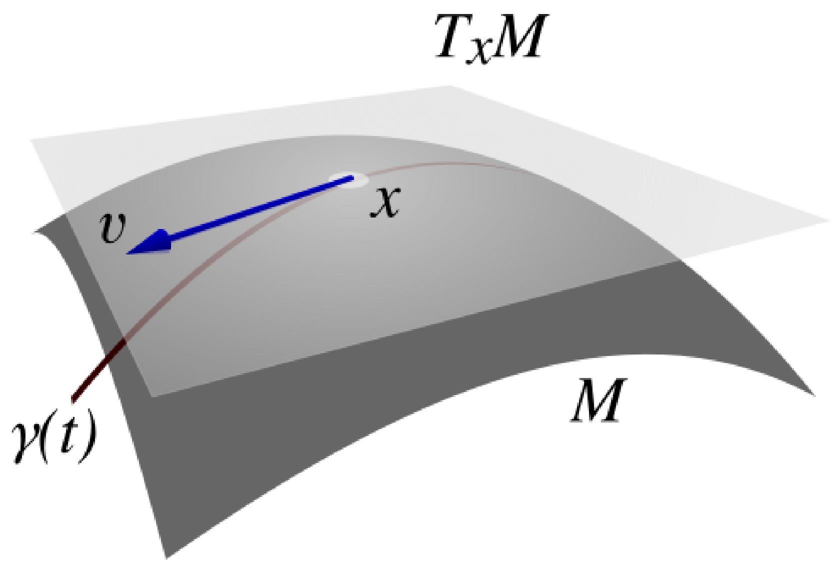
$$\|f(x) - f(y)\| = \|x - y\|$$

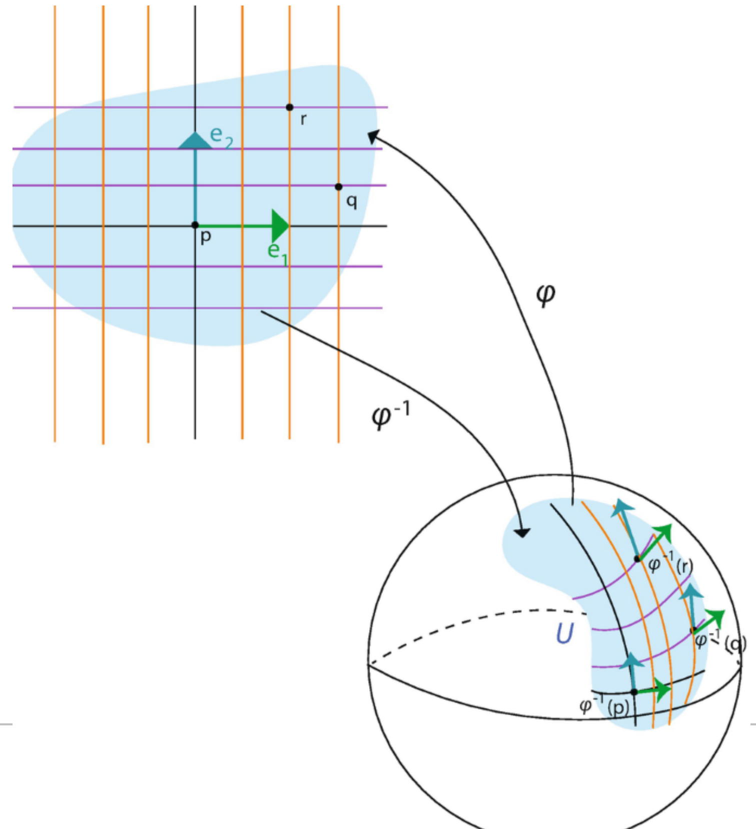
- Rotation, Reflexion, (Translation)



Non-Euclidean Data

- Manifold $\mathcal{M}$, point $x \in \mathcal{M}$
- "Local coordinate" approximating neighboring points of x

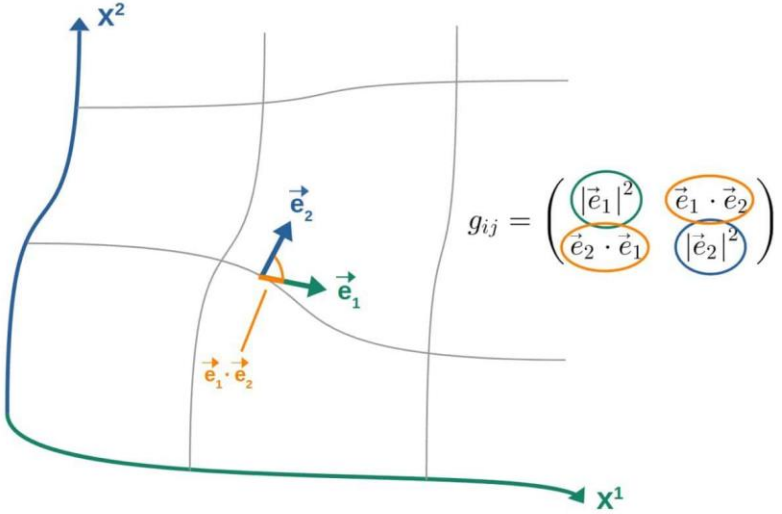




Riemann Metric

- **Riemann metric** defines an **inner-product** within tangential space

$$\{\langle -, - \rangle_x : T_x\mathcal{M} \times T_x\mathcal{M} \to \mathbb{R} | x \in \mathcal{M}\}$$



- Riemann metric can be represented as a **matrix** regarding a selected local coordinate (ie. basis)

$$\phi : \mathbb{R}^m \to \mathcal{M} \subset \mathbb{R}^D,\ x = \phi(q), \qquad \langle \frac{\partial \phi}{\partial q_i}, \frac{\partial \phi}{\partial q_j} \rangle_{\phi(q)} = g_{ij}(q)$$

Isometry in Riemannian manifolds

- Geometry preservation is the preservation of inner products

$$\|f(x) - f(y)\|^2 = \|x - y\|^2$$

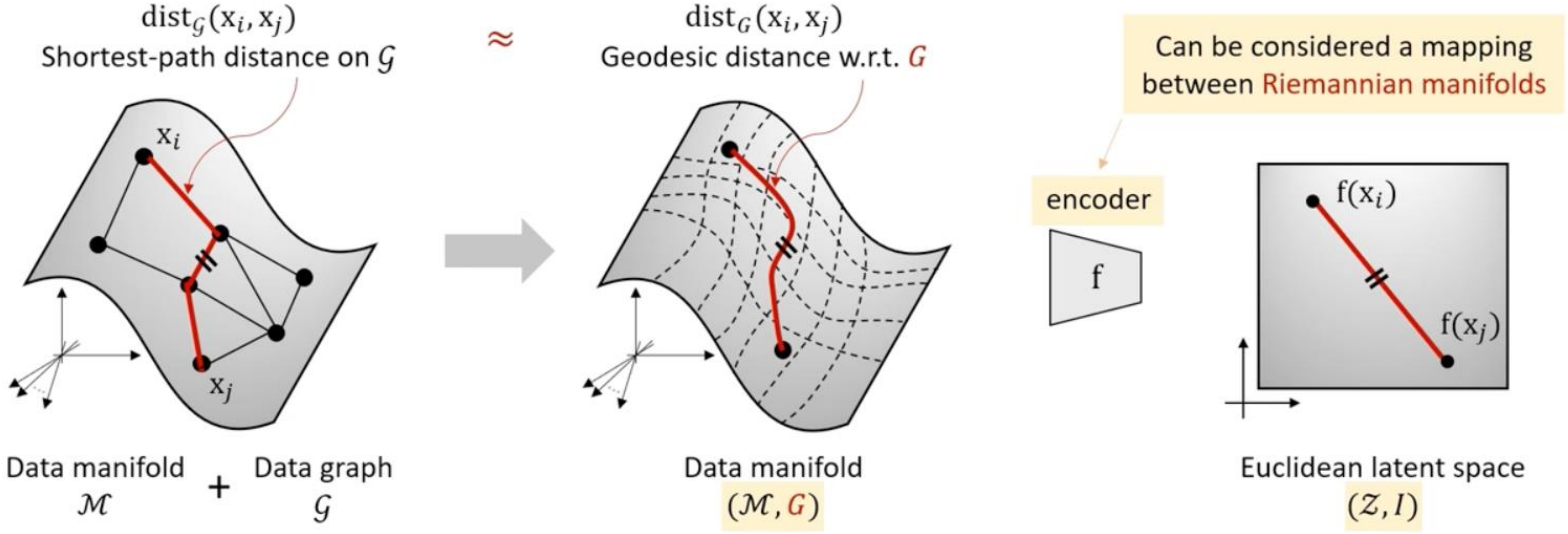
$$\langle f(x) - f(y), f(x) - f(y) \rangle = \langle x - y, x - y \rangle.$$

- Isometry $f: \mathcal{M} \to \mathcal{N}$

$$G(x) = J_f(x)^T H(f(x)) J_f(x), \quad \forall x \in \mathbb{R}^m.$$

$$\langle d\mathrm{f_x}(\mathrm{v}), d\mathrm{f_x}(\mathrm{w}) \rangle_{\mathrm{f(x)}} = v^T J_f(x)^T H(f(x)) J_f(x) w = v^T G(x) w = \langle \mathrm{v}, \mathrm{w} \rangle_{\mathrm{x}},$$

From discretized datapoints …



https://www.youtube.com/watch?v=CNBzomrvtIo

Laplace-(Beltrami) Operator

- ”Diffusion” defining operator…

$$\frac{\partial u(x,t)}{\partial t} = D\,\Delta u(x,t)$$

(Euclidean space)

$$\frac{\partial f}{\partial t} = -\Delta_M f$$

(Riemannian Manifold)

$$\frac{\partial f}{\partial t} = -Lf$$

(Graphs)

Laplace-(Beltrami) Operator

- "Deviance from local average" …

$$\Delta = \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2}$$

(Euclidean space)

$$\Delta_M f = \frac{1}{\sqrt{|g|}} \frac{\partial}{\partial x^i}\left(\sqrt{|g|}\, g^{ij} \frac{\partial f}{\partial x^j}\right)$$

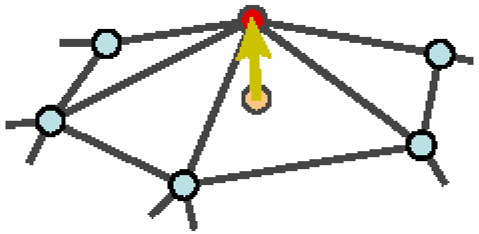
(Riemannian Manifold)

$$L = D - A$$

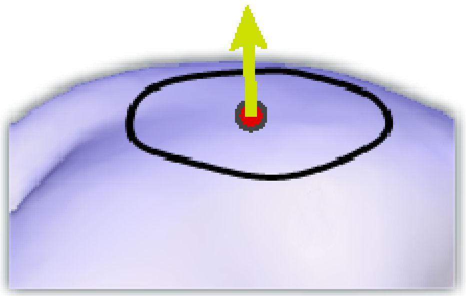
(Graphs)

Graph Laplacian and Laplace-Beltrami

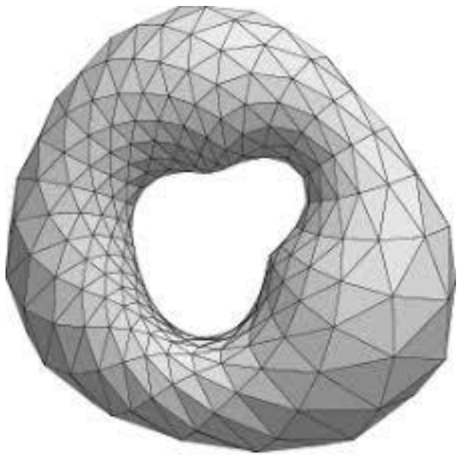
- Laplacian-Beltrami operator can be locally approximated
  - by a well-defined Graph Laplacian
  - constructed on some $\varepsilon - net$ within the manifold



$$\delta_i = \frac{1}{d_i} \sum_{j \in N(i)} (\mathbf{v}_i - \mathbf{v}_j)$$



$$\frac{1}{|\gamma|} \int_{\mathbf{v} \in \gamma} (\mathbf{v}_i - \mathbf{v}) dl(\mathbf{v})$$



Triangulation / Polygonal Mesh

Point Cloud

Sorkine, Olga. "Laplacian mesh processing." *Eurographics (State of the Art Reports)* 4.4 (2005): 1.

Graph Laplacian and Laplace-Beltrami

- Define Graph Laplacian as …

$$
\begin{aligned}
&K_{ij} = k_h(\mathrm{u}_i, \mathrm{u}_j) = k\left(\frac{\mathrm{dist}_{\mathcal{A}}(\mathrm{u}_i, \mathrm{u}_j)^2}{h}\right), \quad i, j = 1, \ldots, N. \\
&K = (K_{ij}), \;\; d_i = \textstyle\sum_j K_{ij}, \;\; D = \mathrm{diag}(d_i). \\
&\tilde{K} = D^{-1} K D^{-1}, \;\; \tilde{d}_i = \textstyle\sum_j \tilde{K}_{ij}, \;\; \tilde{D} = \mathrm{diag}(\tilde{d}_i). \\
&L = \frac{\tilde{D}^{-1}\tilde{K} - I}{c\,h}.
\end{aligned}
$$

- Then, it locally approximates Laplace-Beltrami as …

$$\Delta_{\mathcal{M}} q(\mathrm{x}_i) = \textstyle\sum_j L_{ij} q(\mathrm{x}_j)$$

Riemannian Distortion Metric

- Constraints of Isometry $f: \mathcal{M} \rightarrow \mathcal{N}$ …

$$G = J^T H J$$

$$J^T H J \, G^{-1} = I$$

$$\forall \lambda, \quad \lambda = 1$$

- Reformulate into minimizing a (local) distortion metric …

$$\text{argmin}\left(\sum_i (\lambda_i - 1)^2\right) = \text{argmin}\left(\sum_i \left(\lambda_i^2 - 2\lambda_i\right)\right) \quad (\lambda_i\text{는 } J^T H J G^{-1}\text{의 고유값})$$

Riemannian Distortion Metric

- Global distortion metric as …

$$\int_{\mathcal{M}} \mathrm{Tr}((J_f^T H J_f G^{-1})^2 - 2 J_f^T H J_f G^{-1})\, d\mu,$$

- Using cyclic invariance of Trace …

$$\int_{\mathcal{M}} \mathrm{Tr}((\overbrace{H}^{\boldsymbol{I}\ (desired)} \underbrace{J_f G^{-1} J_f^{\top}}_{(denoted\ as)\ \widetilde{\boldsymbol{H}}})^2 - 2 H J_f G^{-1} J_f^{\top})\, d\mu.$$

Riemannian Distortion Metric

- $\widetilde{H} \equiv J_f G^{-1} {J_f}^T$ is reformulated via Laplace-Beltrami op.

$$(J_f G^{-1} J_f^\top)_{lk} = \frac{1}{2} \Delta_{\mathcal{M}} \left(\mathrm{f}^l - \mathrm{f}^l(\mathrm{x})\right) \left(\mathrm{f}^k - \mathrm{f}^k(\mathrm{x})\right) |_{\mathrm{x}}$$

- Approximate Laplace-Beltrami by Graph Laplacian

$$J_f G^{-1} J_f^\top = \frac{1}{2} \mathrm{f}(X)(\mathrm{diag}(L_i) - e_i e_i^\top L - L^\top e_i e_i^\top) \mathrm{f}(X)^\top$$

- Thus $\widetilde{H} \equiv J_f G^{-1} {J_f}^T$ is (locally) approximated as function of

$$\tilde{H}_i(L, \mathrm{f}(X)) := \tilde{H}(e_i, L, \mathrm{f}(X)) \in \mathbb{R}^{n \times n}$$

Riemannian Distortion Loss

- Global distortion metric is formulated as

$$\mathcal{F}(\tilde{\mathrm{f}}_\theta, L) := \frac{1}{N}\sum_{i=1}^{N} \mathrm{Tr}\left[\tilde{H}_i(L, \tilde{\mathrm{f}}_\theta(X))^2 - 2\tilde{H}_i(L, \tilde{\mathrm{f}}_\theta(X))\right]$$

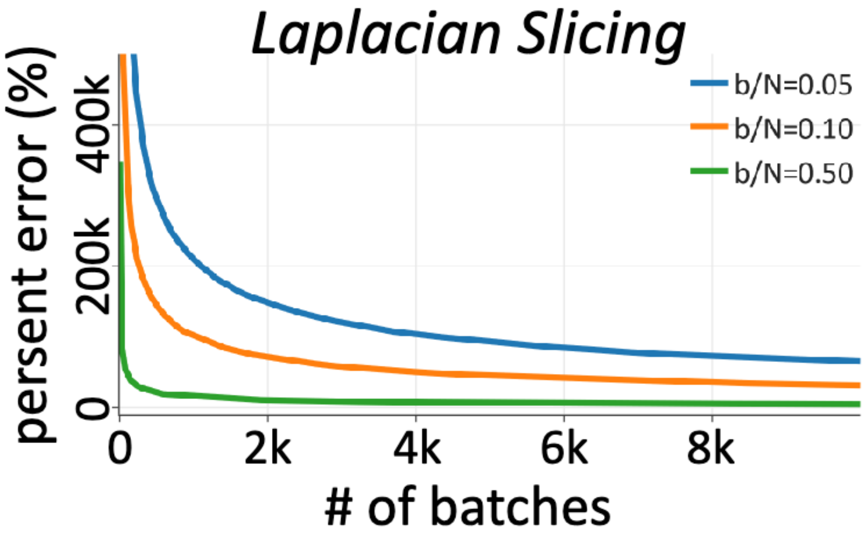
$$\tilde{H} \;=\; \tfrac{1}{2}\, f(X)\left(\mathrm{diag}(L_i)\right) \;-\; e_i\, e_i^T\, L \;-\; L^T\, e_i\, e_i^T\,\left(f(X)\right)^T$$

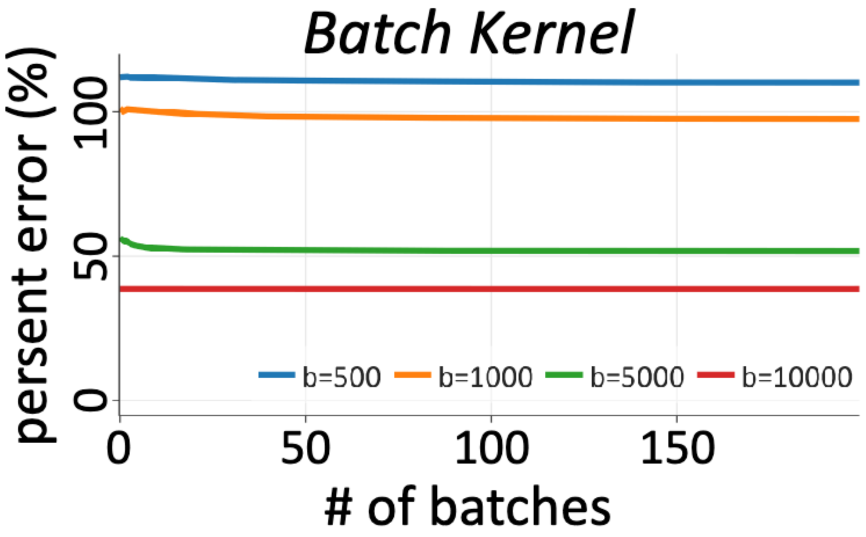
- Enabling gradient-based regularization of distortion with …
  - $\boldsymbol{f}(\boldsymbol{X})$: latent representations of batch data
  - $\boldsymbol{L}$: Graph Laplacian calculated from distance between batch data
- Compatibility to mini-batch-wise SGD?

Minibatch Evaluation

- Minibatch-wise evaluation approximates global distortion
  - Laplacian Slicing: submatrix of the whole Laplacian matrix
  - **Batch Kernel:** Laplacian calculated from submatrix of whole Kernel

$$L = \frac{\tilde{D}^{-1}\tilde{K} - I}{h/4}$$





Jacobian Regularization?

$$\int_{\mathcal{M}} \mathrm{Tr}((J_f^\top H J_f G^{-1})^2 - 2J_f^\top H J_f G^{-1})\, d\mu$$

on ... $\boldsymbol{r} = \boldsymbol{g} \cdot \boldsymbol{f}$ …

- Denoising Autoencoders

$$\min_r \int_{\mathbb{R}^D} E_{q(\tilde{x}|x)} \left[\|r(\tilde{x}) - x\|^2 \right] \rho(x)\, dx,$$

- Contractive regularization

$$\min_r \int_{\mathbb{R}^D} \left(\|r(x) - x\|^2 + \sigma^2 \mathrm{Tr} \left(\left(\frac{\partial r}{\partial x} \right)^\top \left(\frac{\partial r}{\partial x} \right) \right) \right) \rho(x)\, dx$$

Jacobian Regularization?

Both r converges to…

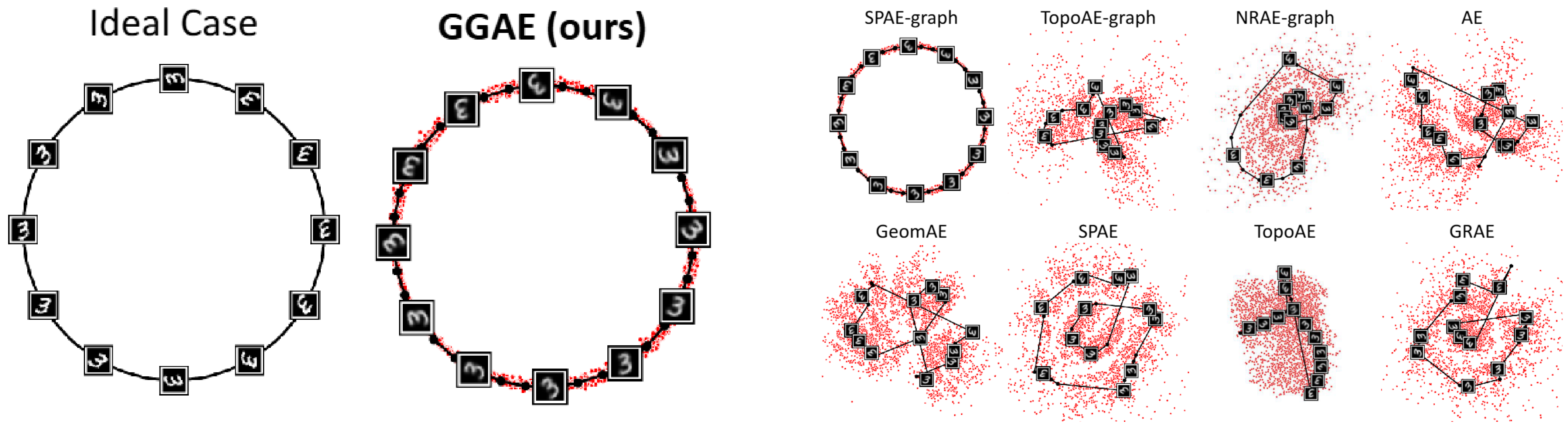
$$r(x) = x + \sigma^2 \nabla_x \log p(x),$$

or re-expressed as …

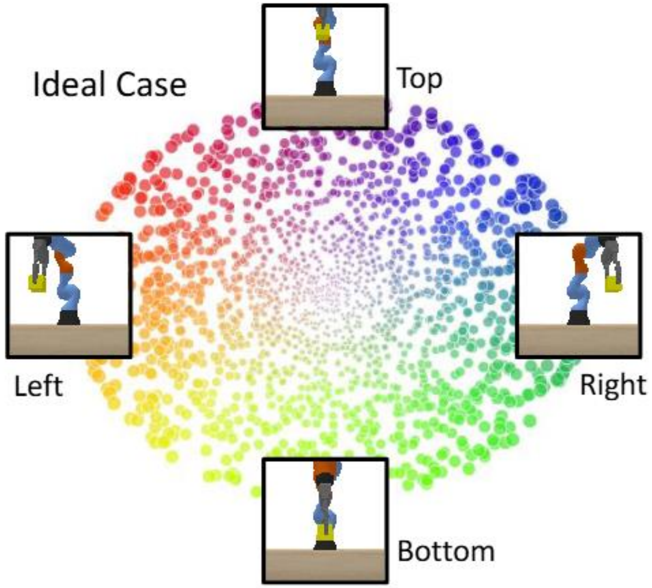
$$\frac{\partial \log \rho(x)}{\partial x} = \frac{1}{\rho} \frac{\partial \rho}{\partial x}(x) = \frac{r(x) - x}{\sigma^2} + \mathcal{O}(\sigma^2).$$

Rotating MNIST

- Rotate MNIST digit-3 by 10 degrees * 36
- GT (ground truth) graph as neighboring images (2-NN)
- 2D latent space



Dynamics learning



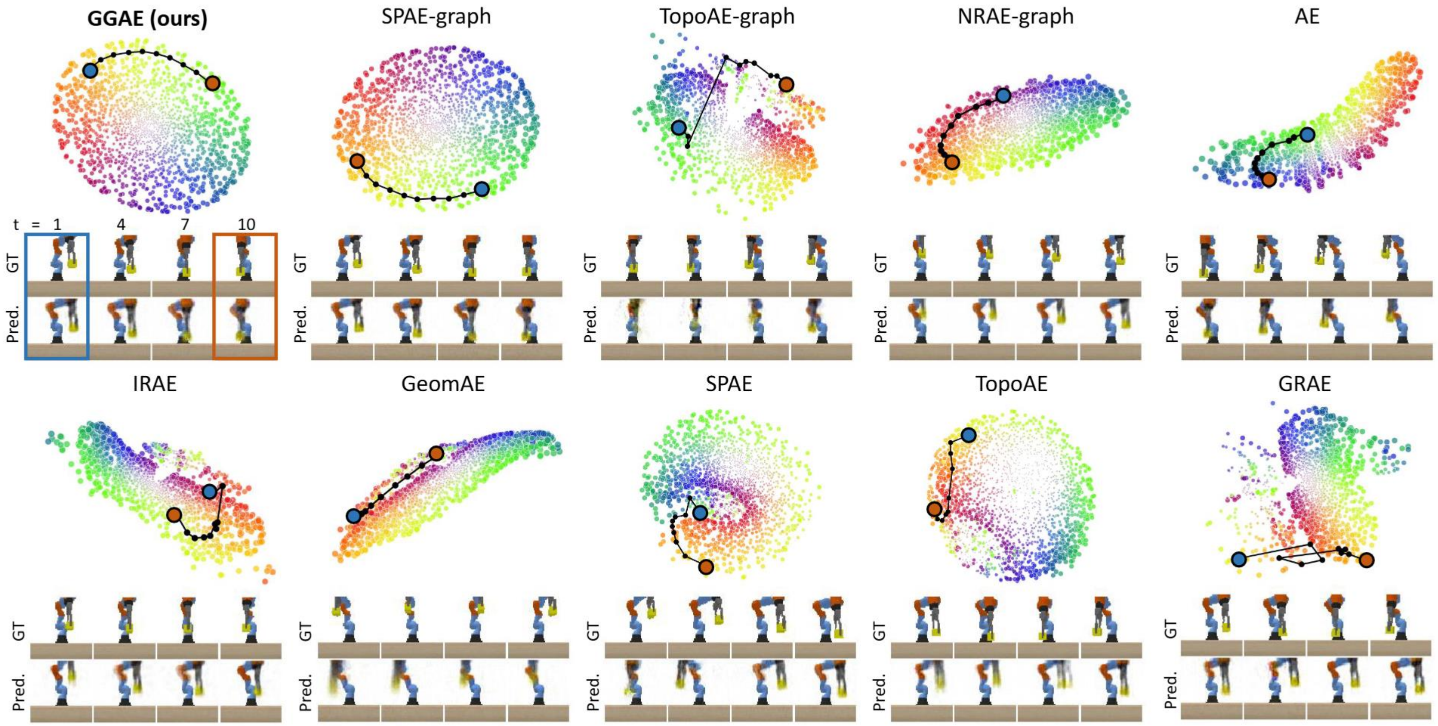
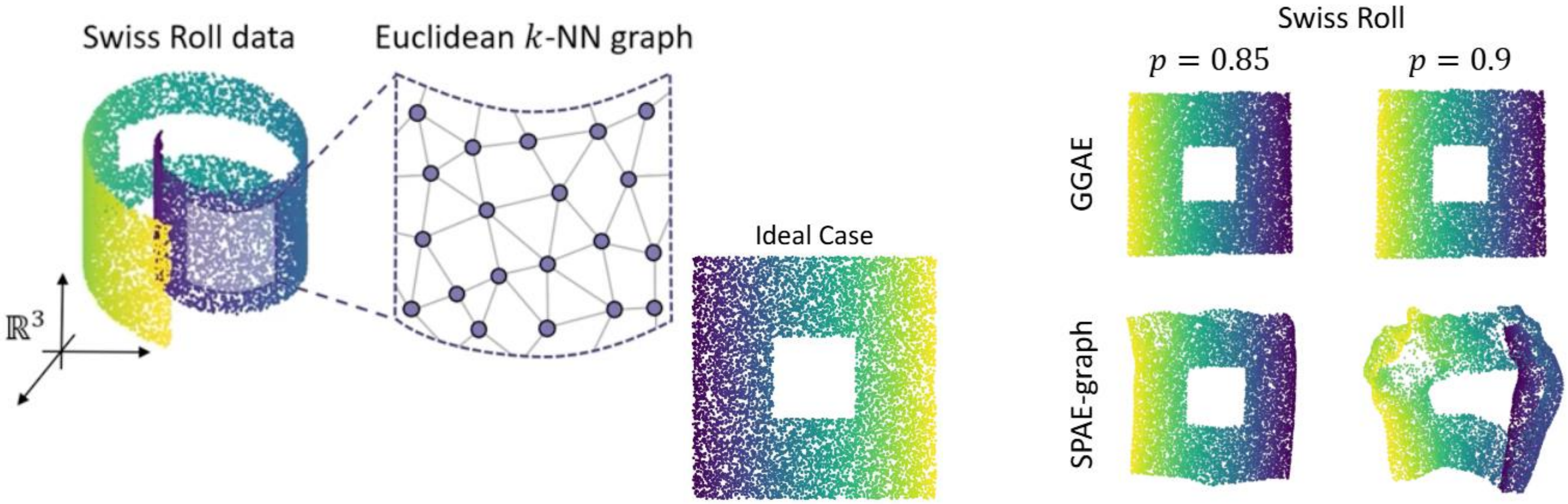


Figure 6. Ten future latent states (black dots) predicted by the trained dynamics model and corresponding reconstructed images at $t = 1, 4, 7, 10$. The dynamics model, when trained in a distorted latent space, exhibits irregular step sizes or jumps in the latent space, leading to its failure to accurately predict future images.

Robustness to Graph Construction error

- Probability p to **drop** an edge in true k-NN graph



https://www.youtube.com/watch?v=CNBzomrvtIo

Weak Supervision by GGAE

- k-NN graph from Euclidean distance of original image
  - Absent GT graph
- Multiply distance at differing classification label (MNIST-3, 8)

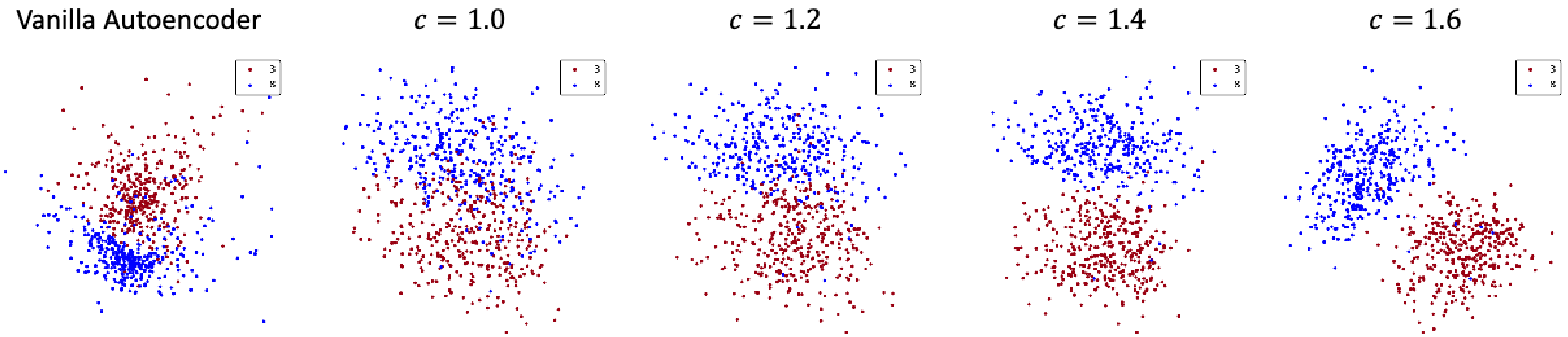


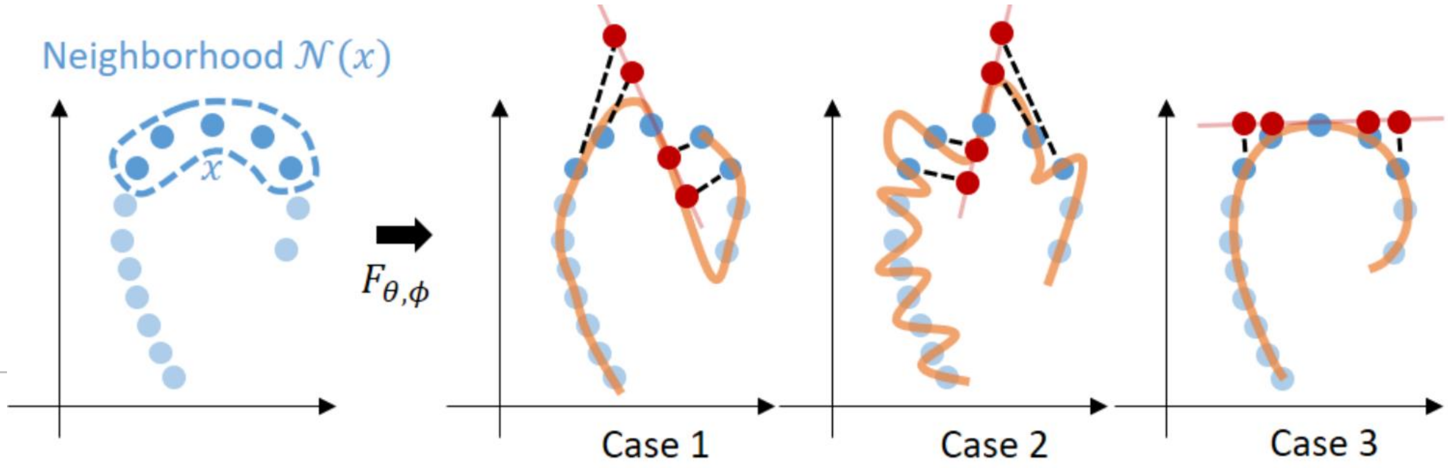
Figure 12. Left: Two-dimensional latent representation learned by vanilla autoencoder. *Others*: Latent representation learned by GGAE, with increasing value of c in graph construction. Larger c results in clearer separation between '3' and '8', as expected.

Further Readings

- **A Riemannian Geometric Framework for Manifold Learning of non-Euclidean Data (2020)**
  - Manifold learning can be generalized into a minimization problem of a well-defined Riemannian distortion metric
    - Metrics composed of eigenvalues λ_i or trace values of matrices
  - Establishment of theoretical backgrounds of GGAE
    - Approximation $JG^{-1}J^T$ via Laplace-Beltrami and Graph Laplacian
    - Minimization of $\Sigma(\lambda - 1)^2$ or $\Sigma\lambda$, for eigen-values of $JG^{-1}J^T \; or \; J^T GJ$

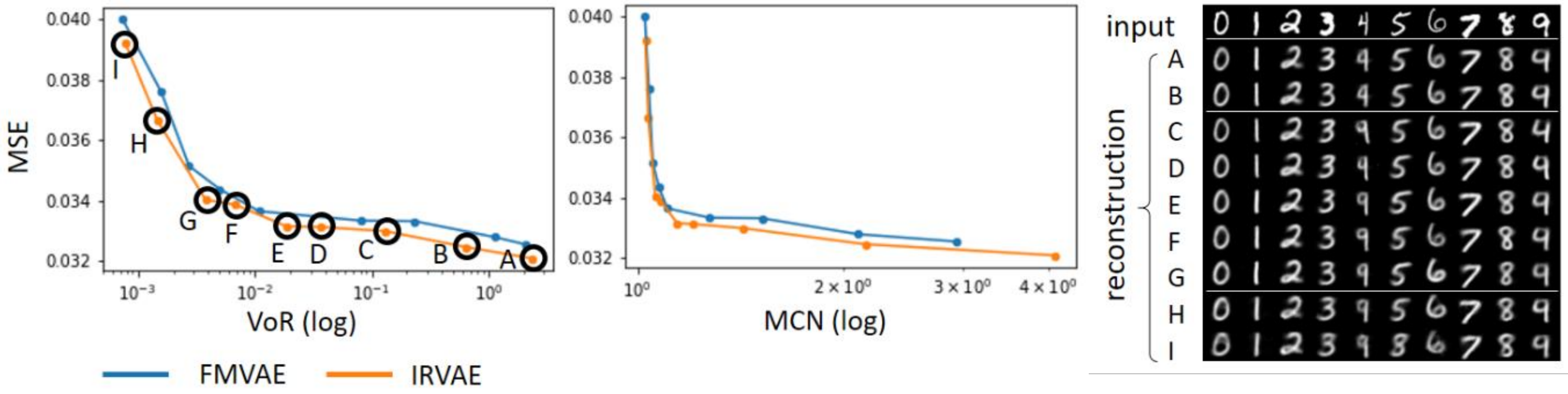
Further Readings

- Neighborhood Reconstructing Autoencoders (NRAE, NeurIPS, 2021)
  - https://neurips.cc/media/neurips-2021/Slides/27723.pdf
  - Equalize tangential space approximated via quadratic approximation to the (true) tangential space
    - Neighboring points are sampled data from the (true) tangential space
  - Regularizations on decoder



Further Readings

- Regularized Autoencoders for Isometric Representation Learning (IRVAE, ICLR, 2022)
  - Regularization function to force “scaled” isometry
  - Adjunct flattening procedure, evaluation by VoR, MCN
  - Trade-off between reconstruction and distortion



Further Readings

- Geometrically Regularized Autoencoders for non-Euclidean Data (GRCAE, ICLR, 2023)
  - DAE, RCAE loss shifts the reconstruction output towards a higher probability density
    - DAE = noise invariance
    - RCAE = Jacobian (or Dirichlet energy) regularization
    - Shift induced by DAE, RCAE loss aligns to the score function
  - Utilization of Riemann metrics provide better AE functionality
  - (Known Riemann metric required)

Further Readings

- **Bronstein, Michael M., et al. "Geometric deep learning: going beyond euclidean data."** ***IEEE Signal Processing Magazine*** **34.4 (2017): 18-42.**
  - Introduction to differential geometry
  - Graph Laplacians, Spectral invariance
- Burago, Dmitri, Sergei Ivanov, and Yaroslav Kurylev. "A graph discretization of the Laplace–Beltrami operator." *Journal of Spectral Theory* 4.4 (2015): 675-714.
  - Approximation of Laplacian-Beltrami and required assumptions
- https://www.youtube.com/watch?v=CNBzomrvtIo

Applications

- Dimension Reduction
- A form of contrastive loss with weak supervision
    - Construct distance graph based on known labels
    - Aligning representations to the geometry of the distance graph may act as a weak supervision

$$\mathcal{L}_{\mathrm{GGAE}}(\theta, \phi) = \underbrace{\frac{1}{N}\sum_{i=1}^{N} \|x_i - (g_\phi \circ f_\theta)(x_i)\|^2}_{\text{standard autoencoder reconstruction}} + \alpha \overbrace{\frac{1}{N}\sum_{i=1}^{N} \mathrm{Tr}\left(\tilde{H}_i(L, f_\theta(X))^2 - 2\,\tilde{H}_i(L, f_\theta(X))\right)}^{\text{distortion term}}$$

Realistic Considerations?

- Convergence characteristics?
  - Dataset size? – coverage of data manifold
  - Batch size? – approximation accuracy of mini-batch Laplacian
  - Minibatch selection strategy
- Graph construction
  - Bandwidth parameter, k-NN
- CPU/GPU requirements
- Trade-off against reconstruction task
- Comparison to established/available methods
  - Rank-based contrastive learning, CLIP