2021 ICCV OCTOBER 11-17 VIRTUAL

Point-set Distances for Learning Representations of 3D Point Clouds

Trung Nguyen[1] Quang-Hieu Pham[3] Tam Le[4]
Tung Pham[1] Nhat Ho[5] Binh-Son Hua[1,2]

[1]VinAI Research, Vietnam
[2]VinUniversity, Vietnam
[3]Woven Planet North America, Level 5
[4]RIKEN AIP, Japan
[5]University of Texas, Austin

VinAI RESEARCH

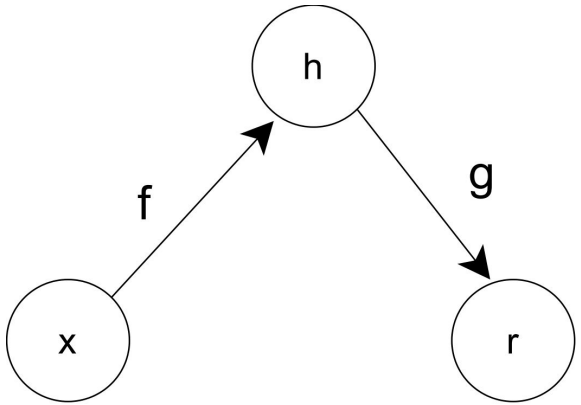
VINUNIVERSITY

Motivation

- Autoencoder: successfully applied to **dimensionality reduction** and **information retrieval** tasks.

- Learning process of an autoencoder:

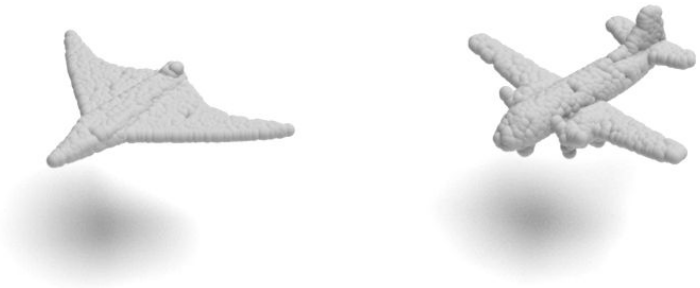
$$\min_{f,g} L(x, g(f(x)))$$



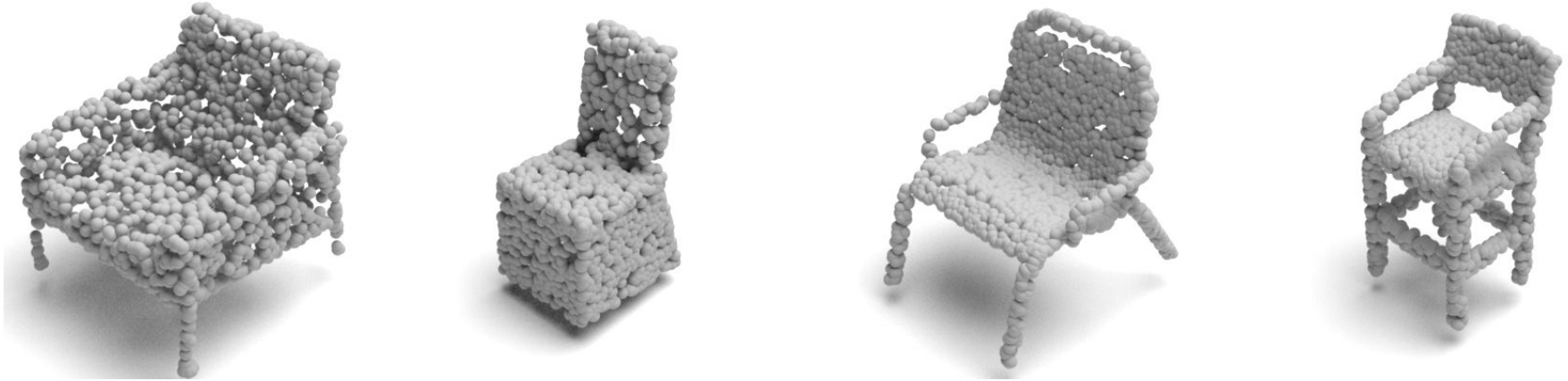
The general structure of an autoencoder

Motivation

- Point cloud data: a set of d-dimensional vector of coordinates, color, normals, etc.

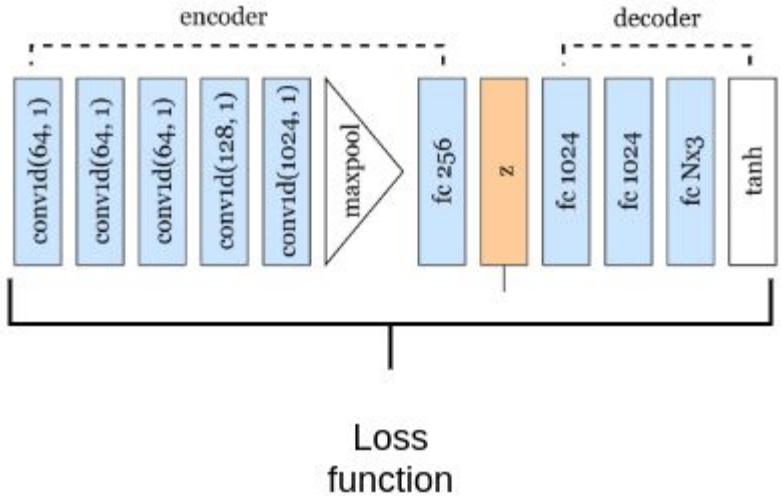


- Invariant under permutation



Problem Statement

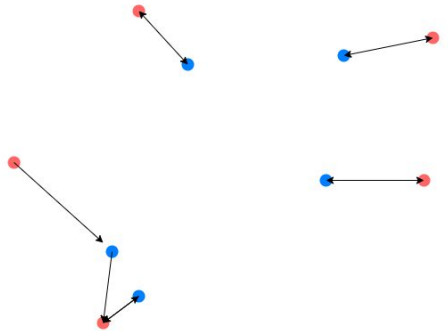
When learning **autoencoders** for **point clouds**, how do different types of **loss functions** affect the **learning process** and the **quality** of **latent codes**?



Background

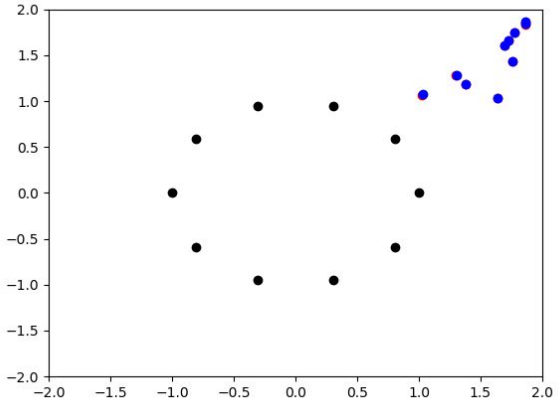
Chamfer discrepancy (CD) [1]

$$d_{\mathrm{CD}}(P,Q) = \frac{1}{|P|}\sum_{x\in P}\min_{y\in Q}||x-y||_2^2 + \frac{1}{|Q|}\sum_{y\in Q}\min_{x\in P}||x-y||_2^2.$$



Chamfer correspondence

Correspondence: **closest point**

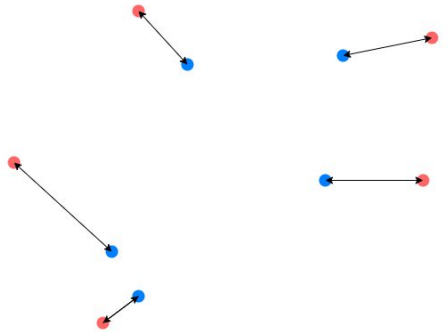


[1] Fan et al - A point set generation network for 3D object reconstruction from a single image - CVPR' 17

Background

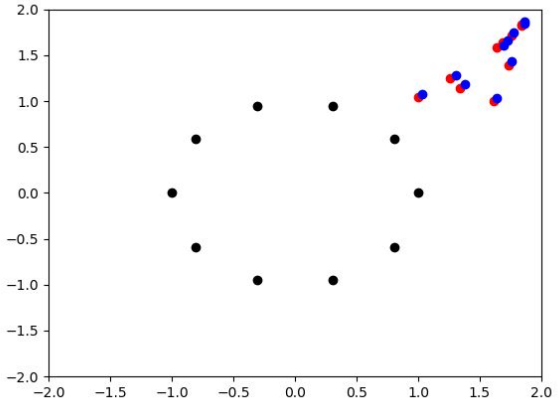
Earth Mover's distance (EMD) [1]

$$d_{\text{EMD}}(P, Q) = \min_{T:P\to Q} \sum_{x\in P} ||x - T(x)||_2.$$



EMD correspondence

Correspondence: **optimal assignment**



[1] Fan et al - A point set generation network for 3D object reconstruction from a single image - CVPR' 17

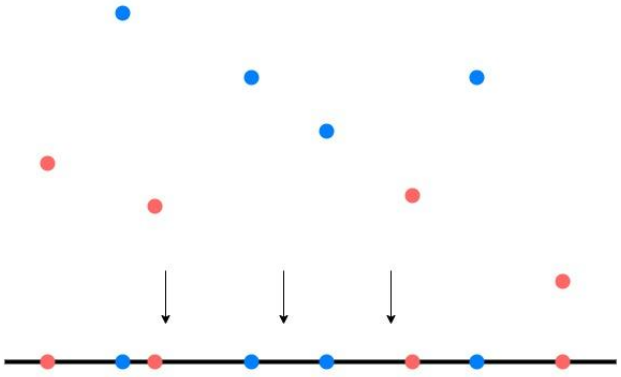
Background

Sliced Wasserstein distance (SWD) [1]

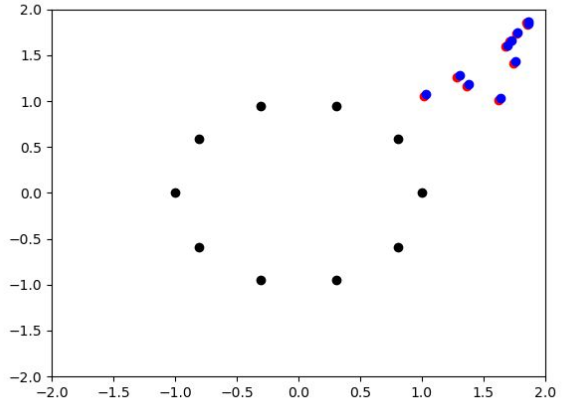
For μ and ν are distributions over $\mathbb{R}^n$, the sliced Wasserstein distance between μ and ν is:

$$SW_p(\mu,\nu) = \left(\int_{\mathbb{S}^{n-1}} W_p^p(\pi_\theta \sharp \mu, \pi_\theta \sharp \nu) d\theta \right)^{\frac{1}{p}}$$

where θ denotes a unit vector on $\mathbb{R}^n$, and $\pi_\theta \sharp \mu$ denotes the projected measure of μ on θ .



Project points to a line



[1] Bonneel et al - Sliced and Radon Wasserstein barycenters of measures - JMIV' 15

Background

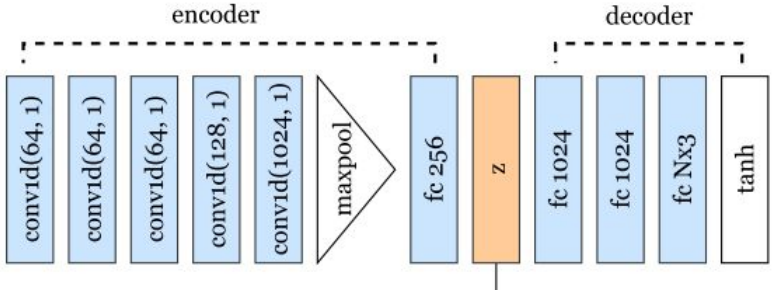
PointNet autoencoder [1]

Theorem 1 *Suppose $f : \mathcal{X} \to \mathbb{R}$ is a continuous set function w.r.t Hausdorff distance $d_H(\cdot,\cdot)$. $\forall \epsilon > 0$, $\exists$ a continuous function h and a symmetric function $g(x_1,\ldots,x_n) = \gamma \circ MAX$, such that for any $S \in \mathcal{X}$,*

$$\left| f(S) - \gamma \left(\underset{x_i \in S}{MAX} \{h(x_i)\} \right) \right| < \epsilon$$

where $x_1,\ldots,x_n$ is the full list of elements in S ordered arbitrarily, γ is a continuous function, and MAX is a vector max operator that takes n vectors as input and returns a new vector of the element-wise maximum.

Theorem 1 [1]



PointNet autoencoder

[1] Qi et al - PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation - CVPR' 17

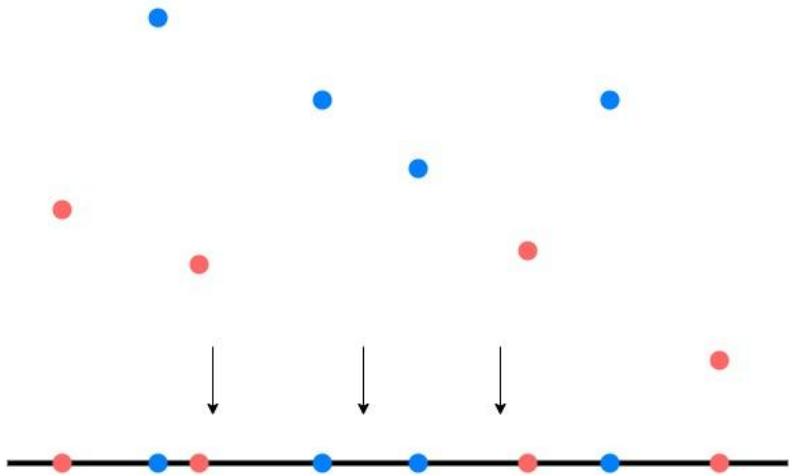
Our findings

- Theoretical relationships between of Chamfer, EMD and SWD
- Experimental results
- Improve approximating SWD

SWD on point clouds

- A point cloud can be viewed as a discrete uniform distribution.

- The projection step in SWD will only lead to small loss of information of the original point clouds.



Project points to a line

Relationships between Chamfer, EMD and SWD

- **Theorem**: Given 2 point clouds P, Q. Assume |P|=|Q| and the support of P and Q is bounded in a convex hull of diameter K, then we find that

$$d_{\mathrm{CD}}(P, Q) \leq 2K d_{\mathrm{EMD}}(P, Q)$$

- **Implication**: Minimizing EMD leads to a smaller Chamfer but not vice versa.

- SWD is equivalent to EMD [1].

[1] Bayraktar et al - Strong equivalence between metrics of Wasserstein type

Reconstruction results

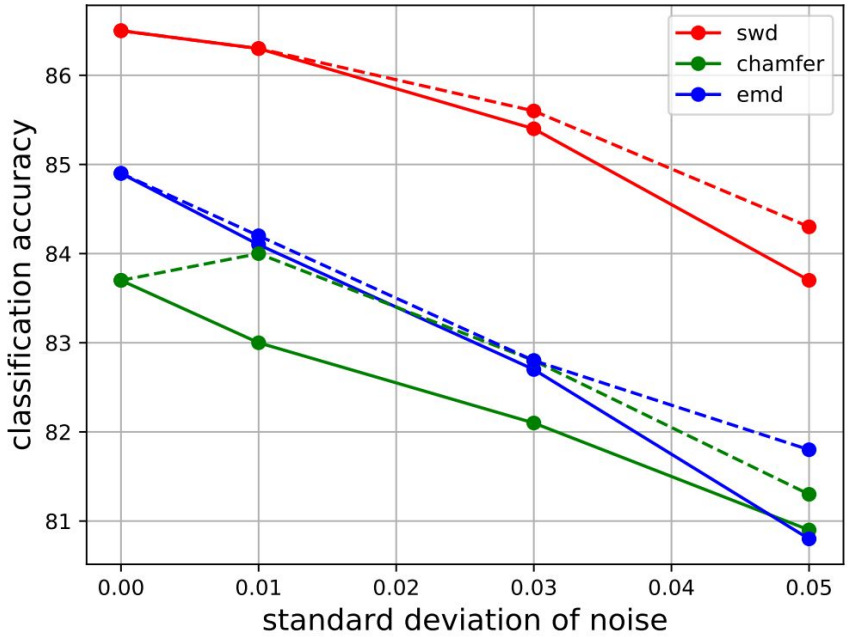
We train our autoencoders on the ShapeNetCore v2 dataset, and test the reconstruction tasks on the ModelNet40 dataset.

Method	CD	SWD	EMD
CD-AE	0.014	6.738	0.314
EMD-AE	0.014	2.295	0.114
SWD-AE	**0.007**	**0.831**	**0.091**

Average dissimilarity between inputs and reconstructions over the ModelNet40.

Classification with noisy data

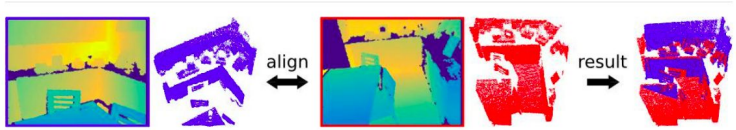
- Solid lines: train the autoencoder with clean point clouds
- Dashed lines: train the autoencoder with noisy point clouds, i.e., perturb ShapeNet in the same way as with ModelNet40.
- SWD is more robust to noise.



Classification accuracy on ModelNet40 with noisy data.

Registration results

Estimate a rigid transformation between two 3D point clouds. We use the autoencoders for local feature extraction, and use RANSAC to estimate the transformation.

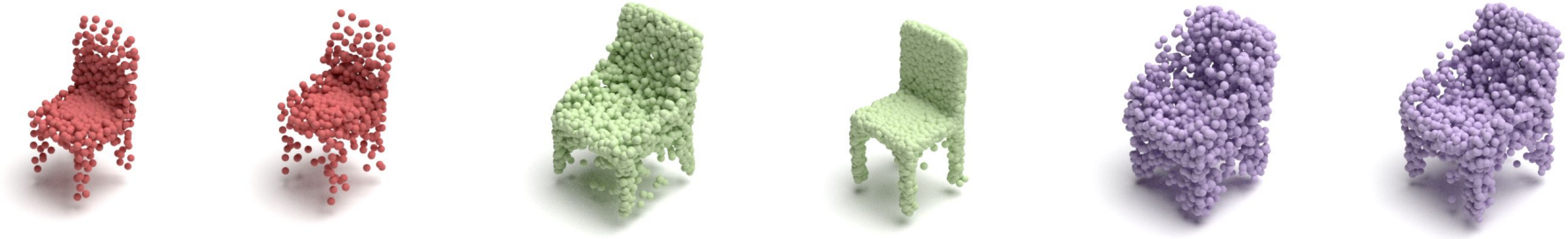


	CD-AE	EMD-AE	SWD-AE
home1	59.4	**60.4**	**60.4**
home2	47.2	46.5	**47.8**
hotel1	62.6	62.1	**69.8**
hotel2	43.6	44.9	**48.7**
hotel3	46.2	34.6	**65.4**
kitchen	58.4	57.0	**62.6**
lab	42.2	46.7	**48.9**
study	50.4	50.0	**55.6**
Average	51.3	50.3	**57.4**

3D registration results (recall) on 3DMatch benchmark

Generation results

	JSD	MMD-CD	MMD-EMD	COV-CD	COV-EMD	NNA-CD	NNA-EMD
CD-AE	38.97	0.65	23.44	31.91	5.47	86.63	100.00
EMD-AE	3.73	**0.61**	**10.44**	**35.75**	35.75	**86.34**	**87.96**
SWD-AE	**3.24**	0.79	11.22	28.51	**37.96**	91.43	91.80



Experiments with other backbone

Training Point-capsule autoencoder [1] is much more computationally expensive than training PointNet autoencoder, so we cannot use EMD to train Point-capsule autoencoder.

	CD	SWD	EMD	Accuracy
PCN-SWD	0.006	**0.761**	**0.084**	**88.78**
PCN-CD	**0.003**	3.035	0.156	88.45

Quantitative measurements of the discrepancy between inputs and reconstructions on ModelNet40. The last column is the classification accuracy on ModelNet40.

[1] Zhao et al - 3D Point Capsule Networks - CVPR' 19

Runtime statistics

SWD is as computationally favorable as Chamfer, while is cheaper than EMD.

Distance	Runtime (ms)
EMD	385
CD	120
SWD	138

Training time per iteration in milliseconds of PointNet autoencoders with different distance functions

Improving approximation of SWD

Provide a statistical guarantee on the approximated value, based on **first** and **second empirical moments**

	CD	SWD	EMD	Acc
SWD	**0.007**	**0.831**	**0.091**	**86.8**
ASW	**0.007**	0.854	0.092	**86.8**

Quantitative measurements of the discrepancy between inputs and reconstructions on ModelNet40. The last column is the classification accuracy on ModelNet40.

Algorithm 1: Adaptive sliced Wasserstein.

Input: Two point sets, positive integers N_0, s; $\epsilon > 0$; maximum number of projections M

Ouput: $\overline{sw}_N$

Sample N_0 projections ;

Compute $\overline{sw} := \overline{sw}_{N_0}, \overline{sw^2} := \overline{sw^2}_{N_0}, N := N_0$;

while $\overline{sw^2} - (\overline{sw})^2 > \frac{(N-1)\epsilon^2}{4}$ & $N \leq M$ **do**

  Sample s projections ;

  Compute $\overline{sw}_s, \overline{sw^2}_s$;

  Assign $\overline{sw} := \frac{N \times \overline{sw} + s \times \overline{sw}_s}{N+s}$;

  Assign $\overline{sw^2} := \frac{N \times \overline{sw^2} + s \times \overline{sw^2}_s}{N+s}$;

  Assign $N := N + s$;

end

Conclusions

- SWD possesses both the statistical benefits of EMD and the computational benefits of Chamfer divergence.
- Latent codes learned by SWD seem to lead to better performance in many downstream tasks than those learned by Chamfer and EMD.

Future work

- Investigate why latent codes of SWD seems "better" than those of EMD. (May related to how we approximate those distances with deep learning.)

Thank you for listening!

Contact: trungnguyen100397@gmail.com
Code: https://github.com/VinAIResearch/PointSWD