Point-set Distances for Learning Representations of 3D Point Clouds
– Supplementary Material –

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

Abstract

In this supplemental material, we further test the robustness of the proposed metrics (Section 1), and then report numerical results on more metrics including adaptive-sliced Wasserstein (ASW)*, max sliced Wasserstein* (MSW) *and generalized sliced Wasserstein* (GSW) *(Section 2). We also report additional results on the point cloud registration (Section 3) and the point cloud generation task (Section 4). Additionally, we also provide an evaluation of the number of slices used in the sliced Wasserstein (SSW) on the reconstruction, classification, and registration task.*

1. Robustness

We conduct an experiment to compare robustness between Chamfer, EMD, and SWD. Particularly, we train autoencoder using Chamfer, EMD, and SWD respectively on ShapeNet, with point coordinates in $[-1, 1]$. At test time, we use ModelNet40, and the point clouds are perturbed by Gaussian noise $N(0, \sigma^2)$ with $\sigma \in [0.01, 0.05]$. We use the autoencoder to extract features from noisy point clouds and then input to learn a classifier. Figure 1 shows the performance of the classifier with increasing standard deviation values, where the experiments are carried out three times and then taken average. The solid lines demonstrate the case where we train the autoencoder with clean point clouds, while the dashed lines demonstrate the case where we train with noisy point clouds, i.e., we perturb ShapeNet in the same way as we do with ModelNet40. In both cases, Figure 1 shows that features learned with SWD are the most robust to noise, outperforming both CD and EMD by about 3% of accuracy. We also found that CD is less robust than EMD.

1.1. Performance with respect to batch size

We provide an experiment with batch size in Table 1, which shows negligible change in the Chamfer discrepancy between the input and reconstructed point clouds in ModelNet40 across batch sizes.

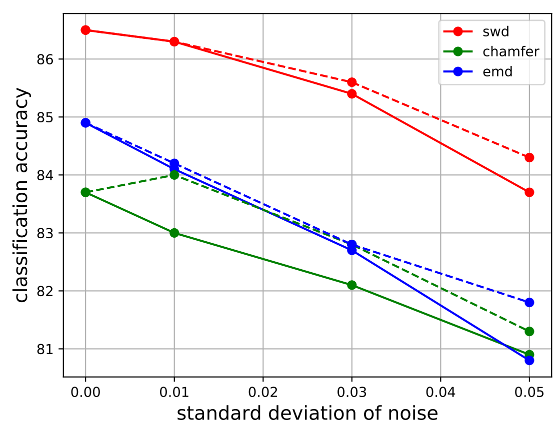

Figure 1: Classification accuracy on ModelNet40 with noisy data.

		Batch size		
		32	128	256
Model	SWD-AE	**0.006**	**0.007**	**0.008**
	CD-AE	0.012	0.014	0.014
	EMD-AE	0.012	0.014	0.013

Table 1: Average of the discrepancy between the input point clouds and their reconstructed versions on ModelNet40 with different batch sizes.

2. Generalized sliced Wasserstein distance

First, we recall briefly the definition of generalized sliced Wasserstein distance [1]. Generalized sliced-Wasserstein distance (GSW) extends the sliced-Wasserstein distance by replacing the inner product $\theta^\top x$ with a defining function $g(\theta, x)$ (cf. Assumptions H1-H4 in [1] for the definition of defining function). Denote $\pi_{g_\theta} \sharp \mu$ the pushforward measure of μ through the mapping g_θ where $g_\theta(x) := g(\theta, x)$ for all

x. Then, for $p \geq 1$, the *GSW* is given by

$$GSW_p(\mu,\nu) := \Big(\int_{\Omega_\theta} W_p^p(\pi_{g_\theta}\sharp\mu, \pi_{g_\theta}\sharp\nu) \Big)^{1/p} \quad (1)$$

where Ω_θ is the compact set of feasible parameters.
In our experiments, $\Omega_\theta := \mathbb{S}^2$ and $g(x,\theta) := ||x-\theta||_2$. To estimate GSW, we use Monte Carlo scheme as follows:

$$GSW_p(\mu,\nu) \approx \Big(\frac{1}{N} \sum_{i=1}^{N} W_p^p(\pi_{g_{\theta_i}}\sharp\mu, \pi_{g_{\theta_i}}\sharp\nu) \Big)^{\frac{1}{p}}. \quad (2)$$

where we set $N := 100$. In Table 2, we provide numerical results for GSW on reconstruction and classification tasks. As we can see, GSW is slightly better than SW and MSW in reconstruction task, while SW is slightly better than other variants in classification task.

Method	CD	SWD	EMD	Accuracy (%)
CD-AE	0.014	6.738	0.314	83.9
EMD-AE	0.014	2.295	0.114	84.4
SSW-AE (ours)	0.007	0.831	0.091	**86.8**
ASW-AE (ours)	0.007	0.854	0.092	**86.8**
MSW-AE (ours)	0.007	0.865	0.093	86.5
GSW-AE (ours)	**0.006**	**0.816**	**0.090**	85.8

Table 2: Quantitative measurements of the discrepancy between the input point clouds and their reconstructed versions on ModelNet40. The last column is the classification accuracy on ModelNet40.

2.1. Effect of the number of slices

We measure the effect of varying the number of slices when computing sliced Wasserstein distance using Monte Carlo scheme. We denote SSWn − AE the auto-encoders trained using the sliced Wasserstein distance estimated by Monte Carlo estimation with n projections. We provide quantitative results for reconstruction and classification tasks in Table 3. Table 3 shows that increasing the number of slices in Monte Carlo estimation does not affect performance much in reconstruction and classification tasks.

Method	CD	SWD	EMD	Accuracy(%)
CD-AE	0.014	6.738	0.314	83.9
EMD-AE	0.014	2.295	0.114	84.4
SSW1-AE	**0.007**	0.901	0.094	86.5
SSW2-AE	**0.007**	0.865	0.093	86.5
SSW5-AE	**0.007**	0.829	**0.091**	86.7
SSW10-AE	**0.007**	**0.812**	**0.091**	**86.8**
SSW50-AE	**0.007**	0.849	0.092	**86.8**
SSW100-AE	**0.007**	0.831	**0.091**	**86.8**

Table 3: Quantitative measurements of the discrepancy between the input point clouds and their reconstructed versions on ModelNet40. The last column is the classification accuracy on ModelNet40.

3. Point cloud registration

In Table 4, we provide quantitative results as discussed in section Point cloud registration on page 7 of the main paper. Table 4 shows that SSW archives the best recall on average.

As in reconstruction, we measure the effect of varying the number of slices when computing sliced Wasserstein distance for the registration task. The result is shown in Table 5. In the registration task, increasing the number of slices helps improve the performance by more than 2% on average (Table 5).

	ASW-AE	MSW-AE	GSW-AE	SSW-AE
home1	**63.2**	**63.2**	62.3	60.4
home2	49.1	49.1	**52.8**	47.8
hotel1	67.6	66.5	68.1	**69.8**
hotel2	46.2	**48.7**	46.2	**48.7**
hotel3	57.7	53.8	57.7	**65.4**
kitchen	**64.1**	63.3	63.3	62.6
lab	44.4	44.4	46.7	**48.9**
study	57.7	55.6	**58.5**	55.6
Average	56.3	55.6	57.0	**57.4**

Table 4: 3D registration results (recall) on the 3DMatch benchmark.

4. Point cloud generation

In Figure 2 and Table 6, we provide qualitative and quantitative results as mentioned in section Point cloud generation on page 6 of the main paper. As we can see, MSW archives best performance among SW variants in generation tasks.

References

[1] Soheil Kolouri, Kimia Nadjahi, Umut Simsekli, Roland Badeau, and Gustavo Rohde. Generalized sliced wasserstein distances. In *NeurIPS*, 2019. 1

	SSW1-AE	SSW5-AE	SSW10-AE	SSW50-AE	SSW100-AE	EMD-AE	CD-AE
home1	62.3	63.2	61.3	63.2	60.4	60.4	59.4
home2	49.7	48.4	49.1	50.9	47.8	46.5	47.2
hotel1	65.9	68.7	65.9	68.1	69.8	62.1	62.6
hotel2	50.0	43.6	43.6	47.4	48.7	44.9	43.6
hotel3	50.0	57.7	53.8	65.4	65.4	34.6	46.2
kitchen	64.4	62.6	63.7	62.1	62.6	57.0	58.4
lab	42.2	40.0	46.7	48.9	48.9	46.7	42.2
study	56.4	56.0	56.0	55.6	55.6	50.0	50.4
Average	55.1	55.0	55.0	**57.7**	<u>57.4</u>	50.3	51.3

Table 5: Varying number of slices for the 3D registration task. The best scores are highlighted in bold. The second best scores are underlined.

Method	JSD (↓)	MMD (↓)		COV (%, ↑)		1-NNA (%, ↓)	
		CD	EMD	CD	EMD	CD	EMD
SSW-AE	3.24	0.79	11.22	28.51	37.96	91.43	91.80
ASW-AE	3.58	0.73	10.65	31.76	36.48	93.57	93.94
MSW-AE	3.96	**0.59**	**9.64**	**35.89**	**40.47**	**89.73**	**89.29**
GSW-AE	**3.06**	0.76	10.98	30.13	37.52	91.21	91.65

Table 6: Quantitative results of point cloud generation task on the chair category of ShapeNet. ↑: the higher the better, ↓: the lower the better. JSD, MMD-CD, and MMD-EMD scores are all multiplied by 10^2.

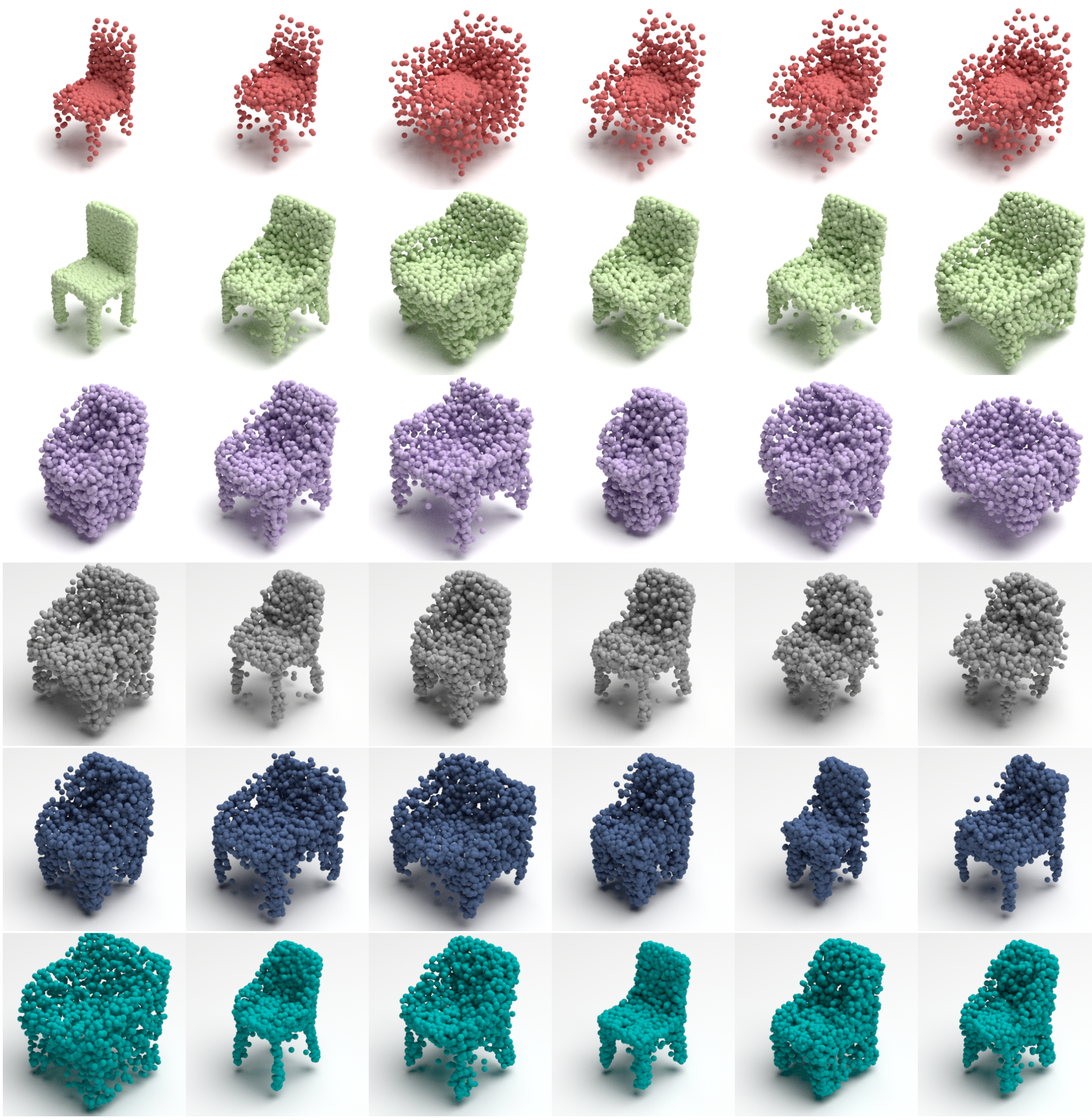

Figure 2: Point cloud generation results of the trained autoencoders on the chair category of ShapeNet. From top to bottom: CHAMFER-AE (red), EMD-AE (green) , SSW-AE (magenta), ASW-AE (gray) and MSW-AE (navy) and GSW-AE (aqua).