Balanced Chamfer Distance as a Comprehensive Metric for Point Cloud Completion

Anonymous Author(s)
Affiliation
Address
email

Abstract

Chamfer Distance (CD) and Earth Mover's Distance (EMD) are two broadly adopted metrics for measuring the similarity between two point sets. However, CD is usually insensitive to mismatched local density, and EMD is usually dominated by global distribution while overlooks the fidelity of detailed structures. Besides, their unbounded value range induces a heavy influence from the outliers. These defects prevent them from providing a consistent and robust evaluation. To tackle these problems, we propose a new similarity measure named **Balanced Chamfer Distance (BCD)**. It is derived from CD and benefits from several desirable properties: **1)** It can detect disparity of density distributions and is thus a more intensive measure of similarity compared to CD; **2)** it is stricter with detailed structures and significantly more computationally efficient than EMD; **3)** the bounded value range encourages a more stable and robust evaluation over the whole test set. We evaluate the proposed BCD on point cloud completion task, in which BCD is used to measure the similarity between the completion results and the high-fidelity ground truth. Experimental results show that BCD not only evaluates the accuracy of overall structures, but also pays more attention to local geometric features by considering point density distributions. In particular, BCD can provide a more reliable evaluation, as CD and EMD often contradict each other. To enhance the balanced distribution of predicted complete point clouds, we incorporate a balanced design and propose a novel point discriminator module in our network, which achieves noticeable improvements under BCD together with competitive results for both CD and EMD. We hope our work could pave the way for a more robust and practical point cloud similarity evaluation.

1 Introduction

Point cloud as one of the fundamental 3D representations is attracting increasing attentions due to its efficiency, flexibility, and direct connection to real-world objects through 3D scanning devices. It has been employed in a wide range of application scenarios and studied for various tasks [1, 4, 7, 11, 29, 31, 32]. A proper similarity measure between two point clouds is always a crucial aspect for both guiding the learning and providing a fair and rational evaluation. However, this is a challenging design considering the unordered and irregular data form and varying point numbers.

Chamfer Distance (CD) and Earth Mover's Distance (EMD) are two of the most universally acknowledged metrics in various tasks. CD is a nearest-neighbour-based method and benefits from its efficient computation and flexible applicability for point sets with different point numbers. EMD relies on solving an optimization problem to find the least expensive one-to-one transportation flow between two point sets. Although sometimes considered to be more faithful to visual quality than CD [1, 13], it is significantly more computational expensive and only applicable to point sets with equal number

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.

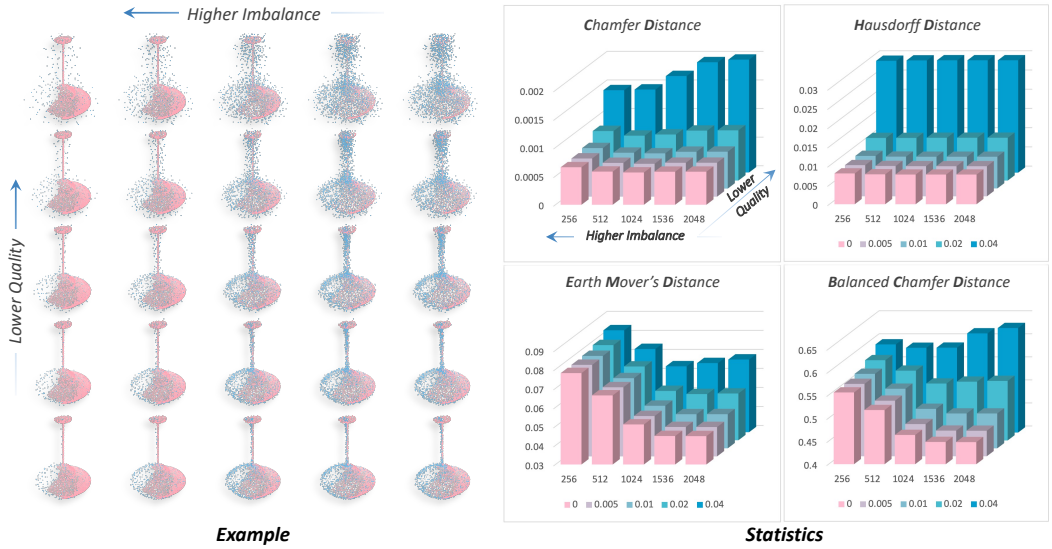

Figure 1: To generate the example matrix on the left, a complete yet noisy shape (blue) with $n \leq 2048$ points is combined with a partial yet clean shape (pink) with 2048 points and then down-sampled to 2048 points via FPS. Four point set distances are calculated between the generated shape and the ground truth. The results are averaged over the whole dataset to get the 3D histogram on the right: CD and HD are not sensitive to imbalance while highly influenced by noise when the intensity achieves a certain level; EMD and BCD share a similar varying pattern as n changes, while BCD is more sensitive to the noise which also represents detailed structures.

of points. The desirable properties of a good similarity metric can be different across application scenarios. **1)** For a discriminative model, it should be robust to sampling strategies and noisy points on detecting the similarity between the continuous surfaces represented by the discrete points. **2)** For a generation model, it should be more strict with the quality of local point distributions so as to depict underlying surfaces of higher quality.

Take a closer look at the two metrics above. The formulation of CD suffers from its intrinsic deficiencies: being insensitive to imbalanced distribution while often violently influenced by outliers [20]. In particular, we specialize one concept closely related to visual quality, namely a matching density distribution between two point sets. Without loss of generality, we assume the ground truth distribution is uniform and use *balance* to denote a perfect matching in this paper. We visualize the issue in Fig. 1 through the task of point cloud completion: 1) under a low noise level, CD hardly changes with the imbalance ratio; 2) it increases violently once the noise intensity raises to some extent; 3) under a high noise level, higher imbalance even results in a lower CD, which leads to a tricky operation to reduce CD for the task: increasing the point density in the seen area with assured accuracy while reducing it in the unseen area to lower down the risk from abnormal points. But severe imbalance also significantly affects the global appearance. On the contrary, EMD can steadily detect the change of balance in Fig. 1. However, it requires the constraint of one-to-one mapping, which is usually over harsh. Consequently, the optima of the transportation problem is dominated by the global distribution while ignoring a fine-structured local details [5], as to be described in Sec. 3. Therefore, both CD and EMD are not perfectly suitable for evaluating the quality of the generated shapes.

In this work, we mainly focus on CD's injustice as an evaluation metric and propose a new similarity measure named **B**alanced **C**hamfer **D**istance (BCD) to tackle the aforementioned challenges. Specifically, BCD is derived from the original CD, while it benefits from a higher sensitivity to distribution quality through a fraction term of query frequency and a higher tolerance to outliers through an approximation of Taylor Expansion. It shares a similar tendency with EMD under a varying point distribution while being more computationally efficient and more sensitive to details. Moreover, due to their different focuses, CD and EMD can often encounter divergence when evaluating different methods, which makes them less reliable as a consistent metric, as to be demonstrated in Sec. 5. We empirically observe that BCD usually provides a more robust evaluation, especially when the results of CD and EMD contradict each other.

Furthermore, using CD as a loss function usually encourages over-populated distributions in common areas shared by the mass of shapes and suffers from Chamfer's blindness [1]. To address the

problem of imbalanced distribution at training time, we also provide a detailed analysis together with an integration of balanced design into the two-stage completion framework as a solution. Extensive investigations are provided for the comparison among different metrics and methods for the task of point cloud completion. Experimental results validate that the proposed metric, Balanced Chamfer Distance, successfully overcomes the aforementioned issues of CD. BCD is more faithful to visual quality than both CD and EMD and provides a more reliable evaluation when CD and EMD contradicts. Our proposed approach also gains noticeable improvement under the new metric together with competitive results for both CD and EMD. Visualization results further validate its superiority.

2 Related Works

Point Cloud Completion. Point cloud completion aims to recover a complete shape based on a partial observation. Earlier works represent shapes with voxels [3, 6, 19], while PCN [31] first proposes to use raw point data and leverages an encoder-decoder structure to generate a global-feature-based coarse shape followed by the folding-based up-sampling [30]. Following works enhance the feature representation by techniques like attention mechanisms [24, 16, 15], hierarchical aggregation [8, 33], and grid structure for cubic feature sampling [28], etc.; the decoding process can also be examined to leverage, for example, a tree-structure [21], iterative refinement [23, 27], multiple patches generation [13], or separated prediction for the seen and unseen [33], etc. These works use either CD or EMD for evaluation, yet the two may not be consistently satisfied due to their different concerns, as to be shown in Sec. 5, and thus a more comprehensive metric is essential for a fair and reliable comparison.

Point Cloud Distance. The term "distance" refers to a non-negative function that measures the dissimilarity between two point sets. Considering the unordered structure of point clouds, the shape-level distance usually comes from statistics of pair-wise point-level distances based on certain assignment strategy. Chamfer Distance (CD) is one of the most widely used metrics based on the nearest neighbour (Eqn. 1). There are variants of it used for training [30, 4] and some other similarly formed distances like Hausdorff [9, 2, 25]. Another generally adopted metric is Earth Mover's Distance (EMD), which relies on solving an optimization problem to find the best mapping function from one set to the other. It is sometimes considered to be more rational than CD [1, 13], but is much more computationally expensive. Recently, Urbach *et al* [22] propose DPDist, which compares point clouds by measuring the distance between the surfaces that they were sampled on. However, it is estimated by a network rather than a mathematical formulation, making it inconvenient and potentially unstable to be adopted into various tasks. Another closely related work is by Nguyen *et al* [14], who propose the sliced Wasserstein distance that has equivalent properties as EMD and similar computational complexity to CD, with the Monte Carlo method involved for approximation. In comparison, we start from the opposite point of view by deriving a new formulation based on CD and result in a clean and explicit expression. Similarly, our BCD also shares the properties of EMD in many cases (Fig. 1), while it detects the detail preserving issue better than EMD.

3 Balanced Chamfer Distance for Point Sets

3.1 Preliminaries

Chamfer Distance between two point sets S_1 and S_2 is defined as:

$$d_{CD}(S_1, S_2) = \frac{1}{|S_1|} \sum_{x \in S_1} \min_{y \in S_2} ||x - y||_2 + \frac{1}{|S_2|} \sum_{y \in S_2} \min_{x \in S_1} ||y - x||_2. \tag{1}$$

Each point $x \in S_1$ finds its nearest neighbour in S_2 and vice versa; all the point-level pair-wise distances are averaged to produce the shape-level distance. The simple and flexible formulation generalizes well across many tasks. Earth Mover's Distance is defined as:

$$d_{EMD}(S_1, S_2) = \min_{\phi: S_1 \to S_2} \sum_{x \in S_1} ||x - \phi(x)||_2. \tag{2}$$

It relies on solving an optimization problem that finds a one-to-one bijection mapping $\phi : S_1 \to S_2$, thus only applicable when $|S_1| = |S_2|$. The pair-wise distances are then calculated between x and $\phi(x)$. As computing the optimal mapping is computationally expensive and even hardly affordable, several approximation schemes [13, 12] have been developed to relieve the computation burden.

3.2 Balanced Chamfer Distance

Formulation and Interpretation. As discussed in Sec. 1, CD is not a comprehensive metric for evaluating visual quality for the generation tasks, e.g., point cloud completion. We explain this from its formulation: **1)** the square operation makes it intensively influenced by outliers, and the evaluation results have a huge varying range across the dataset; **2)**, the nearest point query operation makes it less sensitive to the issue of imbalanced distribution, and hence resulting in a less discriminative evaluation of visual quality. Therefore, we aim to propose a new metric based on the original formulation that not only preserves the capacity of similarity measure but also highly alleviates the problems above, namely Balanced Chamfer Distance (BCD).

Firstly, in order to prevent the dominance of the worst cases, we aim to use a robust distance function with boundaries for measuring each point pair. The Taylor Expansion of $e^z = \sum_{n=0}^{\infty} z^n/n!$ is leveraged, and we take the first order approximation, $e^z \approx 1 + z$, where $z = -||x - y||_2$. Thanks to the point cloud normalization and the nearest neighbour querying operation, the condition that $z \approx 0$ can be easily matched and the approximation is reasonable. Thus we have:

$$d_{CD}(S_1, S_2) \approx \frac{1}{|S_1|} \sum_{x \in S_1} \min_{y \in S_2} (1 - e^{-||x-y||_2}) + \frac{1}{|S_2|} \sum_{y \in S_2} \min_{x \in S_1} (1 - e^{-||y-x||_2}). \tag{3}$$

Considering the property of the exponential function $e^z (z < 0)$, each point-level distance is mapped to a value between $[0, 1]$. As a result, the formulation also sets a natural boundary of $[0, 1]$ for the overall shape distance. Although the approximation would be less accurate as the point-level distance gets away from zero, it exactly helps to mitigate the over-sensitivity of the outliers by suppressing the square growth. We add another scale factor α as in Eqn. 4 to adjust the sensitivity.

Secondly, the imbalance ambiguity of CD is partially due to its "blindness" that each point only considers its nearest neighbour in the other set while ignoring the surroundings. We denote the nearest neighbour assignment process by "query" for simplicity and present a simple example here: assume S_1 to be a uniform point cloud and S_2 an in-homogeneous one; consider two points $y_1, y_2 \in S_2$ with y_1 located at a sparse area and y_2 in a relatively dense area. The calculation of $d_{CD}(S_1, S_2)$ would likely get y_1 frequently queried by points in S_1 due to the local sparsity, and the case for y_2 can be exactly the opposite. Thus we denote that y_1 and y_2 are not equally critical in representing the shape. Furthermore, assume a subset $S_1^y \subseteq S_1$ so that each point in S_1^y queries y and that $n_y = |S_1^y|$, points in this set are unaware of each other under the formulation of CD, and the contribution of y to each point in S_1^y is not affected as n_y gets larger, which is unreasonable.

Alternatively, we denote that the overall contribution of each point to the evaluation system shall be normalized, and we introduce $1/n_y$ to deal with the case where y is being shared by multiple xs:

$$d_{BCD}(S_1, S_2) = \frac{1}{|S_1|} \sum_{x \in S_1} \left(1 - \frac{1}{n_{\hat{y}}} e^{-\alpha||x-\hat{y}||_2}\right) + \frac{1}{|S_2|} \sum_{y \in S_2} \left(1 - \frac{1}{n_{\hat{x}}} e^{-\alpha||y-\hat{x}||_2}\right), \tag{4}$$

where $\hat{y} = \min_{y \in S_2} ||x - y||_2, \hat{x} = \min_{x \in S_1} ||y - x||_2$, and α denotes a temperature scalar. Finally, consider the first term without loss of generality, each y contributes $|-\frac{1}{n_y} \sum_{x \in S_1^y} e^{-||x-y||_2}| \in [0, 1]$ to the overall distance metric (before averaging). A variants for BCD to deal with point sets with different number of points will be discussed in the supplementary material.

Comparison among the Distance Metrics. We conduct a brief comparison of the properties among the three metrics in Table 1, sketching their assignment strategies and computation schemes in Fig. 2. As discussed above, both CD and BCD assign point pairs by querying the nearest neighbours, while BCD further considers the point-specific query frequency n_y, and incorporates the property of density distribution into the measurement. In comparison, EMD naturally forces an equal query frequency of 1 for each point via the mapping function, and the metric is highly sensitive to the global point distribution. However, the harsh constraint not only imposes a significant increase in computational cost but also tends to sacrifice its attention to visual quality for an optimal mapping. As shown in Fig. 2, the assigned pair of points could be located far from each other, and the distance is less physically meaningful. Experimental results in Sec. 5 would also show that an over compacted and smoothed shape can be favored by EMD despite its loss of detailed structure.

In brief, BCD takes a step from CD and attempts to provide a rationale bridge towards EMD for a better sense of point distribution rather than being blinded by its nearest neighbour. Compared with

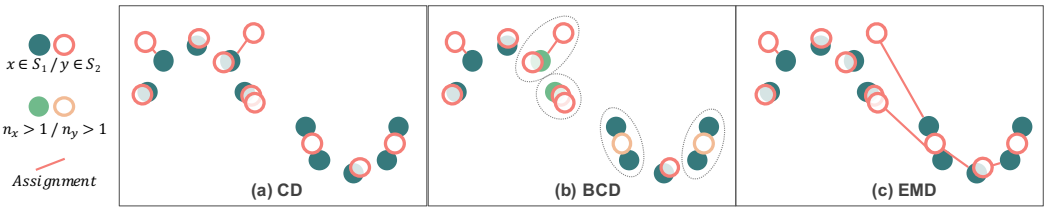

Figure 2: Comparison of assignment strategies and distance calculation. CD and BCD take the nearest neighbour locally, and BCD further considers the point-specific query frequency; EMD forces a one-to-one mapping, and the assigned pair of points may locate far from each other with weaker physical meaning.

Table 1: Comparison of properties among different metrics.

Metrics	assignment	efficient	bounded	balance-sensitive	detail-sensitive
CD	nearest neighbour	✓	×	×	✓
EMD	optimization	×	×	✓	×
BCD	nearest neighbour	✓	✓	✓	✓

EMD, it is not only more efficient but also stricter with local structures. A balanced distribution and a good preservation of detailed structures are both important factors for the visual quality of the completion result. More examples are to be shown in Sec. 5 for better illustration.

4 Incorporating Balanced Design into Point Cloud Completion

4.1 Two-Stage Point Completion Framework

Review of a Typical Two-Stage Pipeline. The two-stage coarse-to-fine completion pipeline was first proposed by PCN [31] and improved by a series of its following works [13, 23, 16]. The first stage takes a partial point cloud as input and extracts a global feature f^g by an MLP, *e.g.*, PointNet [18], followed by a decoder to generate a coarse point cloud P_{coarse}. Though it maintains a relatively reasonable global shape, P_{coarse} usually fails to capture and depict the details. Therefore, the second stage follows up, which aims **1)** to precisely reconstruct the input point cloud without loss of details, denoted by P_{rec}, **2)** to improve the quality of P_{coarse} especially in the unseen part via coordinates adjusting, up-sampling, and probably detail transferring, denoted by P^{+}_{coarse}, and finally **3)** to obtain the final output with desirable visual quality and high fidelity to the original input. Point-wise local features f^l with abundant geometry information encoded are usually involved in this stage.

Discussion. Although CD does not involve any hard assignment between prediction and ground truth point sets, it is observed that the output coordinates from each node of the last Fully Connected (FC) layer have a relatively convergent local distribution. By assuming a Gaussian-like distribution, we visualize their **mean** and **variance** for each category separately across the test set in Fig. 3, where the mean coordinates form the category-specific **mean shapes**, and the color of each node denotes the node variance [1]. This observation indicates that for the coarse shape, there exists an obvious imbalance of density across different regions according to how commonly they are shared across the dataset. This pattern occurs not only for the *statistical results* but also for *each shape instance* [1, 5].

Another thing to emphasize is the trade-off between accuracy and distribution balance. The first few points located in an unseen area significantly reduces the CD loss, while the marginal gain soon decays with more points predicted there. Considering the side effect that inaccurate points would bring in extra corruption to the overall CD loss, lying more points in the seen region with high confidence usually boosts the CD performance. This trick benefits the CD metric, while it violates the overall distribution and hurts the visual quality at the same time.

4.2 Incorporating Balanced Design

Point Discriminator. Given the phenomenon depicted above, the points generated in a shape are not equally "valuable": those in a sparse area would play a more critical role in representing the shape.

[1] The category information is not provided during training, while only used for statistics.

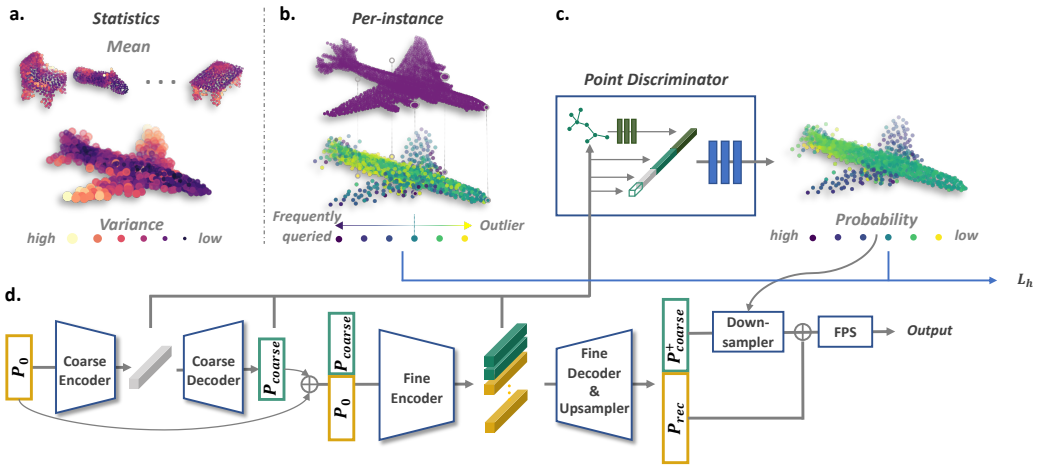

Figure 3: **a.** visualization of mean shape over the dataset; **b.** visualization of $g(x)$ in Eqn. 5 at instance level; **c.** module structure of the point discriminator; **d.** two-stage framework with guided sampling.

And it would be helpful to be aware of the characteristic of each point at inference time so that we can perform a more effective sampling. Concretely, the concept of query frequency in Sec. 3 can be a reliable indicator. For each point $x \in P_{coarse}$, n_x points from the ground truth P_{gt} take it as the nearest neighbour, and a larger n_x denotes a higher importance of x for representing the shape. We apply a $\log_2$ mapping function to constrain the value range.

On the other hand, if n_x equals zero, there are two possible situations: the point either locates in an over-populated region that the competition among candidates is fierce, or it may locate far away from P_{gt} and become a harmful outlier. These cases will be estimated by the distance from x to P_{gt} with a scaling factor t. Thus we formulate a two-sided conjoint objective $g(x)$ as:

$$g(x) = \begin{cases} \log_2(\frac{|P_{coarse}|}{|P_{gt}|} n_x + 1) & n_x > 0, \\ -\min\limits_{y \in P_{gt}} ||x - y||_2 \cdot t & n_x = 0. \end{cases} \tag{5}$$

We design a network h to be trained along with the main network that learns to predict $g(x)$. Specifically, it takes the local features of a point and its neighbours for edge convolution; the results are concatenated with the local feature f^l_x, global feature f^g, and point coordinates c_x, denoted by $z = h(f^l_{x'}, f^g, f^l_x, c_x)$, where $x' \in \mathcal{N}_x$ and $\mathcal{N}_x$ denotes the neighbours for point x. It then takes a 3-layer MLP structure and outputs a single scalar z, which is supervised by a regression loss:

$$L_h = \frac{1}{|P_{coarse}|} \sum_{x \in P_{coarse}} ||z - g(x)||_2. \tag{6}$$

The discriminator would be used for deciding point sampling privilege at inference time. A full training loss will be included in the supplementary materials.

Up-sampling and Guided Down-sampling. In view of the imbalanced population issue discussed in Sec 4.1, an intuitive idea to alleviate it would be to allow a larger number of points generated via up-sampling before reduced to the desired number via Furthest Point Sampling (FPS). However, FPS tends to select more points from outer regions, which increases the risk of including more outliers, and thus this operation usually results in a noisier output. As a result, we aim to make good use of the point discriminator above for the down-sampling stage at inference time together with FPS. Specifically, for each point $x \in P_{coarse}$, we define a point-wise existing probability from $h(f, x)$ by:

$$p(x) = \sigma(\beta \cdot z + \gamma) = \frac{1}{1 + e^{(-\beta \cdot z - \gamma)}}. \tag{7}$$

Let s denote the scale for up-sampling, then x is corresponding to s points in P^+_{coarse} and they all share the same $p(x)$, an independent probability of the point being sampled. P^+_{coarse} is first down-sampled in this manner and then combined with P_{rec} before the final FPS. This process acts as a point filter that leverages the prior learned during training to effectively reduce the unreasonably located points at inference time (Fig. 6), relieving the side effect brought by FPS operation.

Table 2: Point cloud completion results in terms of CD $\times 10^4$, EMD $\times 10^2$, and BCD, lower is better.

Methods	Metrics	airplane	cabinet	car	chair	lamp	sofa	tabel	watercraft	bed	bench	bookshelf	bus	guitar	motorbike	pistol	skateboard	Avg.
PCN	CD	4.50	8.83	6.41	13.01	21.33	9.90	12.86	9.46	20.00	10.26	14.63	4.94	1.73	6.17	5.84	5.76	9.78
	EMD	4.70	7.99	5.75	6.90	11.99	5.32	6.60	5.40	9.84	4.85	7.87	5.24	10.56	4.93	4.86	5.59	6.80
	BCD	0.478	0.519	0.490	0.617	0.710	0.552	0.559	0.580	0.662	0.562	0.608	0.429	0.446	0.548	0.491	0.445	0.553
PCN++	CD	4.06	9.08	6.64	13.11	19.25	9.78	14.36	9.66	22.33	9.73	15.51	5.13	1.86	6.25	5.81	4.99	10.29
	EMD	3.44	3.75	**3.15**	4.65	8.00	3.56	4.69	4.22	6.13	3.85	4.39	**2.62**	2.78	3.60	3.71	3.07	4.27
	BCD	0.428	0.464	0.451	0.574	0.661	0.504	0.517	0.540	0.617	0.524	0.563	0.389	0.369	0.527	0.447	0.393	0.508
TopNet	CD	4.12	9.84	7.44	13.26	18.64	10.77	12.95	8.98	19.99	9.21	16.06	5.47	2.36	7.06	7.04	4.68	10.30
	EMD	4.89	6.30	4.07	7.01	10.75	6.47	7.50	4.68	8.09	6.27	6.80	3.50	4.21	4.26	6.02	3.49	6.18
	BCD	0.536	0.558	0.548	0.650	0.711	0.598	0.599	0.600	0.678	0.588	0.622	0.492	0.487	0.572	0.542	0.496	0.598
MSN	CD	2.73	8.92	6.50	10.75	13.37	9.26	10.17	7.70	17.27	6.64	12.10	5.21	1.37	4.59	4.62	3.38	7.99
	EMD	2.75	4.02	3.47	4.44	6.28	3.74	4.46	3.82	5.27	3.34	4.28	2.92	**2.07**	3.30	3.62	2.21	3.94
	BCD	0.404	0.509	0.516	0.537	0.539	0.532	0.498	0.515	0.574	0.471	0.541	0.458	0.388	0.491	0.463	0.422	0.499
VRC	CD	2.20	**7.92**	5.60	**7.49**	**8.15**	**7.45**	**7.52**	**5.20**	**11.90**	**4.88**	**7.39**	**4.53**	**1.15**	3.90	**3.44**	3.22	**6.09**
	EMD	3.03	7.57	6.14	5.49	6.15	5.80	4.65	4.97	6.58	3.45	5.28	6.59	3.08	4.45	4.56	3.20	5.27
	BCD	0.374	0.509	0.499	0.488	0.475	0.515	0.438	0.478	0.527	0.401	0.470	0.462	0.349	0.452	0.443	0.363	0.462
VRC-EMD	CD	2.72	9.03	6.58	9.93	11.53	9.38	9.80	6.71	17.22	6.88	10.34	5.32	1.39	4.47	4.62	4.69	7.87
	EMD	2.50	**3.65**	3.23	**4.15**	5.31	**3.61**	3.93	**3.58**	**5.17**	3.19	**3.97**	2.69	2.08	**3.06**	3.48	**2.29**	**3.62**
	BCD	0.369	0.483	0.473	0.502	0.499	0.509	0.450	0.478	0.547	0.423	0.487	0.424	0.349	0.445	0.430	0.370	0.461
Our	CD	**2.17**	8.27	**5.48**	7.88	8.20	8.35	9.01	5.45	14.85	5.45	9.07	4.47	1.48	**3.78**	4.06	**2.41**	6.55
	EMD	**2.26**	4.90	3.88	**4.15**	**5.17**	4.39	**3.85**	3.64	5.57	**3.01**	4.09	3.72	2.33	3.13	3.46	2.39	3.88
	BCD	**0.332**	**0.461**	**0.437**	**0.454**	**0.453**	**0.482**	**0.417**	**0.427**	**0.510**	**0.395**	**0.458**	**0.389**	**0.340**	**0.400**	**0.368**	**0.342**	**0.425**

5 Experiments

Dataset. We use the recently proposed MVP Dataset [16] for our study and experiments. It is a multi-view partial point cloud dataset covering 16 categories with 62,400 and 41,600 pairs for training and testing, respectively. It renders the partial 3D shapes from 26 uniformly distributed camera poses for each 3D CAD model selected from ShapeNet [26], and the ground truth point cloud is sampled via Poisson Disk Sampling (PDS).

Comparison Methods and Metrics. We include the following methods for comparison in the main experiments: PCN [31], TopNet [21], MSN [13], and VRCNet [16] (with the PSK module discarded to improve efficiency while slightly scarifying performance). PCN++ is a simple extension of PCN [31] that generates the double number of points for training and down-sampled to the required point number at inference time. We report per-class results on CD, EMD, and BCD for a clear comparison across methods and a view of the different properties of these metrics.

Implementation Details. All the models are trained using the Adam optimizer [10] with the learning rate initialized at $1e^{-4}$ and decayed by 0.7 every 40 epochs. We use a batch size of 32 and a total epoch of 80. Our work is implemented with PyTorch and is run on a Tesla V100 GPU. We set $\alpha = 1000$ for BCD, and $\beta = 5, \gamma = -2$ for our approach in the main experiments.

5.1 Point Cloud Completion on MVP Dataset

Networks are trained with the CD loss except when specified. The main experimental results are reported in Table 2. We first take a look at the traditional CD and EMD, and the results of them do not exhibit a clear positive correlation among these methods: PCN [31] and TopNet [21] have a relatively high loss for both metrics; PCN++ notably lowers down EMD while marginally raising CD for sacrifice; MSN [13] benefits from the multi-surface design and obtains a low EMD result, while its CD performance is not that satisfying; VRC [16] outperforms the other methods under the metric of CD, yet we observe that it takes advantage of the accuracy-balance trade-off in Sec. 4.1, suffering from an imbalanced point distribution and sub-optimal visual quality. Though not reflected by CD, the high EMD and BCD loss expose the problem. Furthermore, VRC trained with EMD loss (VRC-EMD) unsurprisingly achieves a lower EMD yet higher CD than being trained with CD, since it generates a more balanced shape while introducing more local noises, which even disturbs the observed parts (Fig. 4). Experimental results above reveal that since CD and EMD focus on different aspects of the point cloud, inconsistent behaviors may exist between them that confuse the similarity measurement, which prevents either of them from being a comprehensive metric.

As for BCD, we find it to be a more reliable and consistent metric through the observations below by controlling the variables: take all the CD-EMD-BCD tuples (for each category or for the averaged outcomes) in Table 2, for those that have a similar value of CD, BCD would be dominated by their

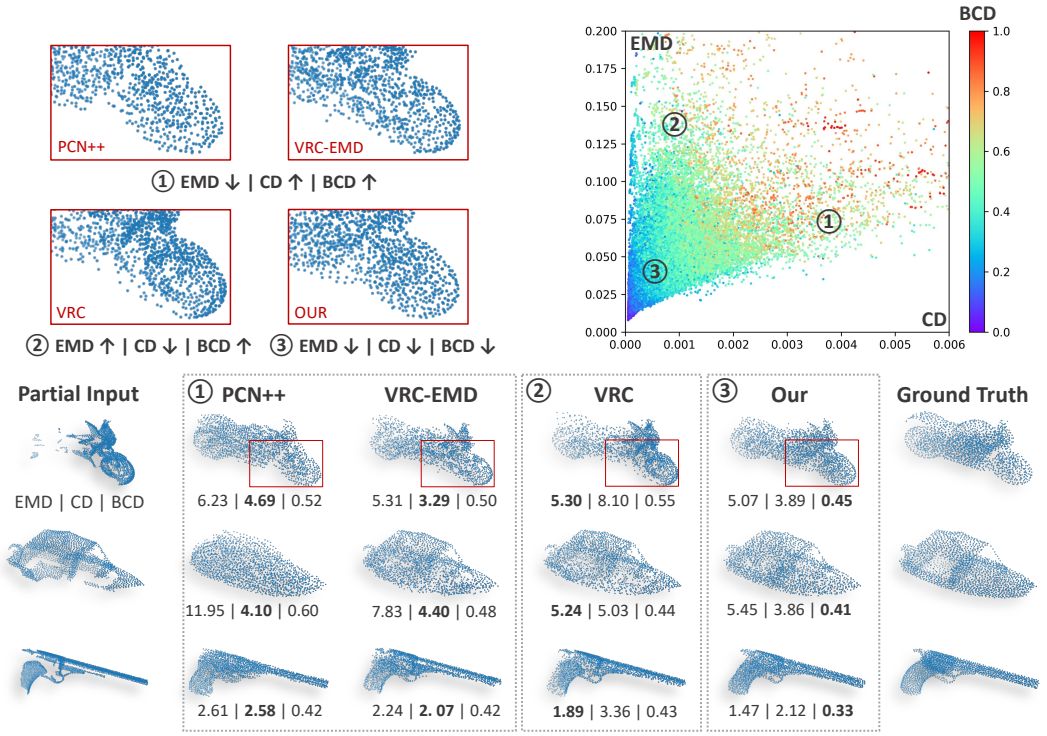

Figure 4: Comparison of CD, EMD, and BCD through examples and statistics. **Below:** examples from different methods that represent three typical cases: 1) EMD is low while CD is high; 2) CD is low while EMD is high; and 3) both CD and EMD are low. BCD is only lowered down in the third case. **Upper left:** A close shot at the cases above. **Upper right:** visualization of the three metrics on the test set, we can observe that 1) the positive correlation between CD and EMD is weak and 2) BCD considers both of them, hence becomes lower towards the original point of the fan-shaped scatter points.

EMD performance for most of the time; the pattern holds when we fix EMD and regard CD and BCD. Although the law is not strictly held or theoretically proved, statistical results in Fig. 4 (upper right) support the practical usage of it. In this way, BCD serves as a consistent, stable, and comprehensive metric even when CD and EMD encounter disparity. Further, by incorporating the balanced design into the network architecture, our method reports the lowest BCD loss, second-lowest CD loss, and second-lowest EMD loss on the MVP testing set. Qualitative results (Fig. 4) show that our results are apparently superior in both the global balanced point distributions and local structures, which further indicates that BCD is a more faithful metric to visual quality.

5.2 Metric Comparison

Time Complexity. EMD relies on solving the linear assignment problem in an iterative approximation manner with a practical time complexity between $O(n^2)$ [12] and $O(n^3)$ [17] and usually $O(n^2)$ memory footprints. We adopt an $O(n)$ memory-efficient implementation by [13] and with the error rate $\epsilon = 0.004$ and an iteration of 3000. The most computationally expensive part for CD and BCD is the nearest neighbour selection, which is usually $O(n)$ for time complexity and can be accelerated by special data structures like KD-tree. We observe that both CD and BCD are significantly efficient to be computed compared with EMD; the running time of EMD is also affected by the distributions of the two point sets, which decides the difficulty of finding the optimal assignment. As shown in Table 4, with the same setting as in Fig. 1, a higher mismatched distribution results in obviously heavier time consumption, and the noise intensity also influences the results.

Bounded Distance. One advantage of BCD is its bounded value range, which promotes an average consideration for all samples in the dataset at the evaluation stage. In contrast, CD and EMD are dominated by the worst cases in the dataset due to their unbounded nature, as shown in Fig. 5. Specifically, for CD, 80% of the normalized loss accumulation is contributed by only 50% of the whole set and 50% contributed by the top 25%; a similar case is also observed for EMD. This property

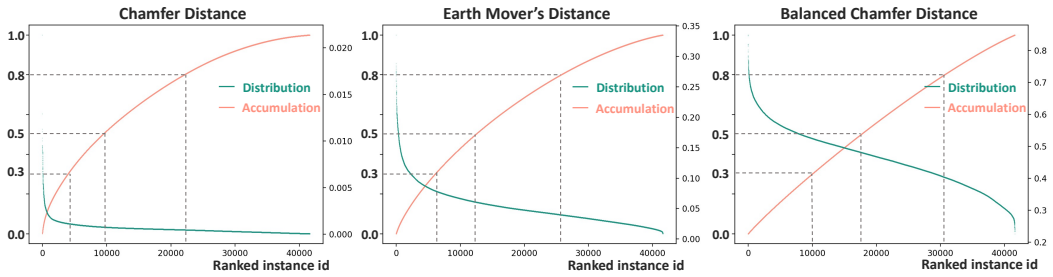

Figure 5: Distance values ranked in a descending order and the accumulation (normalized to 1).

Table 3: Ablation Study

Sampling	Filter	Metric		
		CD	EMD	BCD
		6.09	5.27	0.462
	✓	**5.94**	5.39	0.464
✓		6.91	**3.78**	**0.425**
✓	✓	6.60	3.83	**0.424**

Table 4: Time consumption evaluation.

imbalance	256	512	1024	1536	2048
CD	0.006	0.007	0.008	0.010	0.012
EMD	0.362	0.327	0.267	0.241	0.239
BCD	0.013	0.013	0.013	0.013	0.013
noise	0	0.005	0.01	0.02	0.04
CD	0.008	0.008	0.008	0.008	0.008
EMD	0.272	0.271	0.269	0.264	0.256
BCD	0.013	0.013	0.013	0.013	0.013

makes them fail to comprehensively evaluate the quality of all samples in the dataset. Our BCD overcomes this problem by assuring a clear [0,1] boundary for the distance value, enabling more comprehensive and stable statistical results over the whole test set.

5.3 Ablation Study for our Balanced Design

We study the effectiveness of the two components in our balanced design separately, as shown in Table 3. Applying point filtering via the discriminator can further reduce CD loss from the state-of-the-art VRC [16] approach while slightly hurting the performance of EMD and BCD. The result is reasonable since the removed points more likely come from the predictions for the unseen part since they can be noisier, thus increasing the global imbalance between the seen and unseen. The up-and-down sampling operation naturally promotes a more balanced distribution and benefits the EMD metric, while it hurts CD by including more outliers. This side effect can be partially mitigated by the filtering operation. Several examples in Fig. 6 show that unreasonable scattered points are eliminated from the prediction, resulting in better CD results.

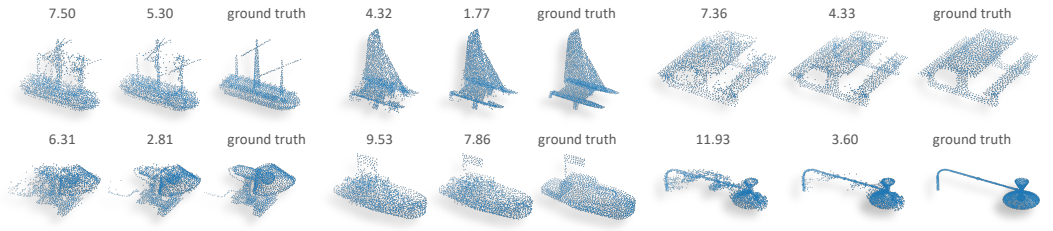

Figure 6: Examples of point filtering (from left to right) based on point discriminator in terms of CD $\times 10^4$.

6 Conclusion

In this work, we propose a new similarity measure for point clouds named Balanced Chamfer Distance (BCD). It is bounded in value, effective in computation, and faithful to visual quality by considering both the density distribution and detailed structures. Our method achieves noticeable improvements under BCD and superior visual quality compared with previous works.

References

[1] Panos Achlioptas, Olga Diamanti, Ioannis Mitliagkas, and Leonidas Guibas. Learning representations and generative models for 3d point clouds. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pages 40–49. PMLR, 10–15 Jul 2018. 1, 2, 3, 5

[2] Xuelin Chen, Baoquan Chen, and Niloy J Mitra. Unpaired point cloud completion on real scans using adversarial training. *The International Conference on Learning Representations*, 2019. 3

[3] Angela Dai, Charles Ruizhongtai Qi, and Matthias Nießner. Shape completion using 3d-encoder-predictor cnns and shape synthesis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5868–5877, 2017. 3

[4] Haowen Deng, Tolga Birdal, and Slobodan Ilic. 3d local features for direct pairwise registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3244–3253, 2019. 1, 3

[5] Haoqiang Fan, Hao Su, and Leonidas J Guibas. A point set generation network for 3d object reconstruction from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 605–613, 2017. 2, 5

[6] Xiaoguang Han, Zhen Li, Haibin Huang, Evangelos Kalogerakis, and Yizhou Yu. High-resolution shape completion using deep neural networks for global structure and local geometry inference. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 85–93, 2017. 3

[7] Pedro Hermosilla, Tobias Ritschel, and Timo Ropinski. Total denoising: Unsupervised learning of 3d point cloud cleaning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 52–60, 2019. 1

[8] Zitian Huang, Yikuan Yu, Jiawen Xu, Feng Ni, and Xinyi Le. Pf-net: Point fractal network for 3d point cloud completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7662–7670, 2020. 3

[9] Daniel P Huttenlocher, Gregory A. Klanderman, and William J Rucklidge. Comparing images using the hausdorff distance. *IEEE Transactions on pattern analysis and machine intelligence*, 15(9):850–863, 1993. 3

[10] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 7

[11] Ruihui Li, Xianzhi Li, Chi-Wing Fu, Daniel Cohen-Or, and Pheng-Ann Heng. Pu-gan: a point cloud upsampling adversarial network. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7203–7212, 2019. 1

[12] Tianyi Lin, Nhat Ho, and Michael Jordan. On efficient optimal transport: An analysis of greedy and accelerated mirror descent algorithms. In *International Conference on Machine Learning*, pages 3982–3991. PMLR, 2019. 3, 8

[13] Minghua Liu, Lu Sheng, Sheng Yang, Jing Shao, and Shi-Min Hu. Morphing and sampling network for dense point cloud completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 11596–11603, 2020. 1, 3, 5, 7, 8

[14] Trung Nguyen, Quang-Hieu Pham, Tam Le, Tung Pham, Nhat Ho, and Binh-Son Hua. Point-set distances for learning representations of 3d point clouds. *arXiv preprint arXiv:2102.04014*, 2021. 3

[15] Yinyu Nie, Yiqun Lin, Xiaoguang Han, Shihui Guo, Jian Chang, Shuguang Cui, and Jian Jun Zhang. Skeleton-bridged point completion: From global inference to local adjustment. *Advances in Neural Information Processing Systems*, 2020. 3

[16] Liang Pan, Xinyi Chen, Zhongang Cai, Junzhe Zhang, Haiyu Zhao, Shuai Yi, and Ziwei Liu. Variational relational point completion network. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021. 3, 5, 7, 9, 12

[17] Ofir Pele and Michael Werman. Fast and robust earth mover's distances. In *2009 IEEE 12th international conference on computer vision*, pages 460–467. IEEE, 2009. 8

[18] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*, 1(2):4, 2017. 5

[19] Edward J Smith and David Meger. Improved adversarial systems for 3d object generation and reconstruction. In *Conference on Robot Learning*, pages 87–96, 2017. 3

[20] Maxim Tatarchenko, Stephan R Richter, René Ranftl, Zhuwen Li, Vladlen Koltun, and Thomas Brox. What do single-view 3d reconstruction networks learn? In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3405–3414, 2019. 2

[21] Lyne P Tchapmi, Vineet Kosaraju, Hamid Rezatofighi, Ian Reid, and Silvio Savarese. Topnet: Structural point cloud decoder. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 383–392, 2019. 3, 7

[22] Dahlia Urbach, Yizhak Ben-Shabat, and Michael Lindenbaum. Dpdist: Comparing point clouds using deep point cloud distance. *The European Conference on Computer Vision (ECCV)*, 2020. 3

[23] Xiaogang Wang, Marcelo H Ang Jr, and Gim Hee Lee. Cascaded refinement network for point cloud completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 790–799, 2020. 3, 5

[24] Xin Wen, Tianyang Li, Zhizhong Han, and Yu-Shen Liu. Point cloud completion by skip-attention network with hierarchical folding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1939–1948, 2020. 3

[25] Rundi Wu, Xuelin Chen, Yixin Zhuang, and Baoquan Chen. Multimodal shape completion via conditional generative adversarial networks. In *The European Conference on Computer Vision (ECCV)*, August 2020. 3

[26] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1912–1920, 2015. 7

[27] Chulin Xie, Chuxin Wang, Bo Zhang, Hao Yang, Dong Chen, and Fang Wen. Style-based point generator with adversarial rendering for point cloud completion. *arXiv preprint arXiv:2103.02535*, 2021. 3

[28] Haozhe Xie, Hongxun Yao, Shangchen Zhou, Jiageng Mao, Shengping Zhang, and Wenxiu Sun. Grnet: Gridding residual network for dense point cloud completion. *arXiv preprint arXiv:2006.03761*, 2020. 3

[29] Guandao Yang, Xun Huang, Zekun Hao, Ming-Yu Liu, Serge Belongie, and Bharath Hariharan. Pointflow: 3d point cloud generation with continuous normalizing flows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4541–4550, 2019. 1

[30] Yaoqing Yang, Chen Feng, Yiru Shen, and Dong Tian. Foldingnet: Point cloud auto-encoder via deep grid deformation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 206–215, 2018. 3

[31] Wentao Yuan, Tejas Khot, David Held, Christoph Mertz, and Martial Hebert. Pcn: Point completion network. In *International Conference on 3D Vision (3DV)*, pages 728–737. IEEE, 2018. 1, 3, 5, 7

[32] Junzhe Zhang, Xinyi Chen, Zhongang Cai, Liang Pan, Haiyu Zhao, Shuai Yi, Chai Kiat Yeo, Bo Dai, and Chen Change Loy. Unsupervised 3d shape completion through gan inversion. In *CVPR*, 2021. 1

[33] Wenxiao Zhang, Qingan Yan, and Chunxia Xiao. Detail preserved point cloud completion via separated feature aggregation. *arXiv preprint arXiv:2007.02374*, 2020. 3

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] See supplementary material.
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No] We will release the code upon the paper accepted.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See Sec. 5.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [No]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] See Sec. 5.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] MVP dataset [16].
   (b) Did you mention the license of the assets? [No]
   (c) Did you include any new assets either in the supplemental material or as a URL? [No]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [No]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [No]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]