VISLab@WPI

Graduate School of Information Sciences
Tohoku University

Loss Distillation via Gradient Matching for Point Cloud Completion with Weighted Chamfer Distance

Fangzhou Lin, Haotian Liu*, Haoying Zhou*, Songlin Hou*, Kazunori Yamada, Gregory Fischer, Yanhua Li, Haichong Zhang, and Ziming Zhang*

IROS '24
ABU DHABI

Introduction

VISLab@WPI

- **Problems Statements:**
  - Chamfer Distance (CD) and variants, are sensitivity to outliers and need time-consuming parameter tuning.
  - Can we define a weighted CD to boost the performance of the vanilla CD?
  - If we can reproduce the exact gradients in training, we can then reproduce the performance of a certain loss (HyperCD in this case).

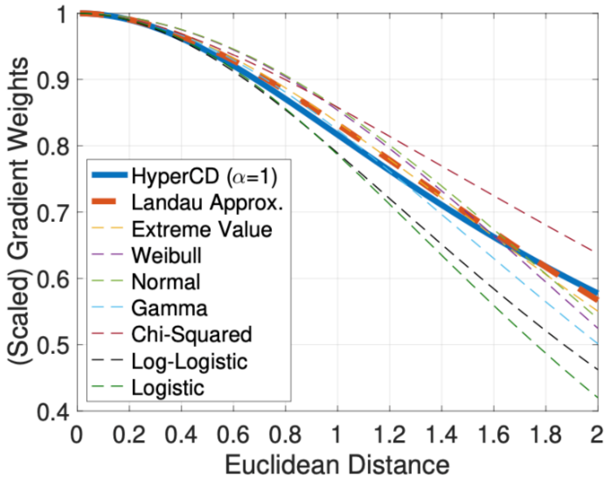


Fig 1. Illustration of distributions that are similar to the gradient weighting distribution from HyperCD and can used as weighting functions in weighted CD.

- **Motivation:**
  - We aim to explore functional spaces to search for good functions.
  - The distances in the metric as a training loss should be weighted in some form rather than uniform.
  - We borrow the idea from network distillation, rather than networks, we aim to learn losses functions instead.

- **Contribution:**
  - We propose a gradient-matching method for loss distillation.
  - We show strong performance for point cloud completion based on weighted CD
  - Our method can be simply done by simulated data with mathematical derivations.

Background

- **Chamfer Distance:** CD serves as a popular training loss in point cloud completion for training neural networks.
  - **In our work:**

$$D_{CD}(x_i, y_i) = \frac{1}{|x_i|}\sum_{j=1}^{|x_i|}\min_k d(x_{ij}, y_{ik}) + \frac{1}{|y_i|}\sum_{k=1}^{|y_i|}\min_j d(x_{ij}, y_{ik})$$

$$d(x_{ij}, y_{ik}) = \begin{cases} \|x_{ij} - y_{ik}\| & \text{as } \textit{L1-distance} \\ \|x_{ij} - y_{ik}\|^2 & \text{as } \textit{L2-distance} \end{cases}$$

- **Hyper CD:** Defines as follow equation

$$d(x_{ij}, y_{ik}) = arccosh\left(1 + \alpha\|x_{ij} - y_{ik}\|^2\right), \alpha > 0.$$

- **Weighting Functions:** Distributions as weighting functions in WeightedCD

Distribution	Params	Mode m	PDF
Chi-Squared	k	$\max(k-2, 0)$	$\frac{1}{2^{k/2}\Gamma(k/2)}x^{k/2-1}e^{-x/2}$
Extreme Value	β	0	$\frac{1}{\beta}e^{-(z+e^{-z})}, z = \frac{x}{\beta}$
Weibull	k, λ	$\begin{cases}\lambda\left(\frac{k-1}{k}\right)^{1/k}, & k > 1, \\ 0, & k \le 1.\end{cases}$	$\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1}e^{-(x/\lambda)^k}$
Log-Logistic	α, β	$\begin{cases}\alpha\left(\frac{\beta-1}{\beta+1}\right)^{1/\beta}, & \text{if } \beta > 1, \\ 0, & \text{otherwise}\end{cases}$	$\frac{\beta}{x}\left(1+\left(\frac{x}{\alpha}\right)^{-\beta}\right)^{-1-\beta}$
Gamma	α, β	$\begin{cases}\frac{\alpha-1}{\beta}, & \text{for } \alpha \ge 1, \\ 0, & \text{for } \alpha < 1\end{cases}$	$\frac{\beta^\alpha x^{\alpha-1}e^{-\beta x}}{\Gamma(\alpha)}$
Logistic	σ	0	$\frac{e^{-x/\sigma}}{\sigma(1+e^{-x/\sigma})^2}$
Normal	σ	0	$\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\sigma^2}\right)$
Landau Approx.	-	0	$\frac{1}{\sqrt{2\pi}}\exp\left\{\left(-\frac{x+e^{-x}}{2}\right)\right\}$

Method

VISLab@WPI

Weighted CD

$$D_W(x_i, y_i) = \frac{1}{|x_i|}\sum_{j=1}^{|x_i|} f(\tilde{d}_{ijk})\tilde{d}_{ijk} + \frac{1}{|y_i|}\sum_{k=1}^{|y_i|} f(\tilde{d}_{ikj})\tilde{d}_{ikj} \quad (1)$$

$$\text{s.t. } \tilde{d}_{ijk} = \min_k d(x_{ij}, y_{ik}), \tilde{d}_{ikj} = \min_j d(x_{ij}, y_{ik}). \quad (2)$$

Loss Distillation via Gradient Matching:

Mimic the learning behavior of the target teacher CD (HyperCD in this case).

$$\min_{f\in\mathcal{F}} \mathbb{E}_{\tilde{d}\sim\tilde{\mathcal{D}}} \left\| z^{(H)}(\tilde{d}) - z^{(W)}(\tilde{d}) \right\| \approx \min_{f\in\mathcal{F}} \sum_{\tilde{d}} p(\tilde{d}) \left\| z^{(H)}(\tilde{d}) - z^{(W)}(\tilde{d}) \right\| \quad (3)$$

$$z^{(H)}(\tilde{d}) = \frac{2\alpha\tilde{d}}{\sqrt{(1+\alpha\tilde{d}^2)^2 - 1}}, z^{(W)}(\tilde{d}) = f'(\tilde{d})\tilde{d} + f(\tilde{d}) \quad (4)$$

Algorithm 1 Loss Distillation via Gradient Matching

Input : a PDF f with parameters $\mathcal{A}$, $z^{(H)}$, $\{(\tilde{d}, p(\tilde{d}))\}$
Output: $\mathcal{A}$

Discretize the parameter space into $\{\mathcal{A}_i\}$ for grid search;
Compute the mode m_i for each $\mathcal{A}_i$ used in $z^{(W)}$;
$\mathcal{A}^* = argmin_{\{\mathcal{A}_i\}} \sum_{\tilde{d}} p(\tilde{d}) \left\| z^{(H)}(\tilde{d}) - z^{(W)}(\tilde{d}) \right\|$;
return $\mathcal{A} \leftarrow \mathcal{A}^*$

Searching Scheme

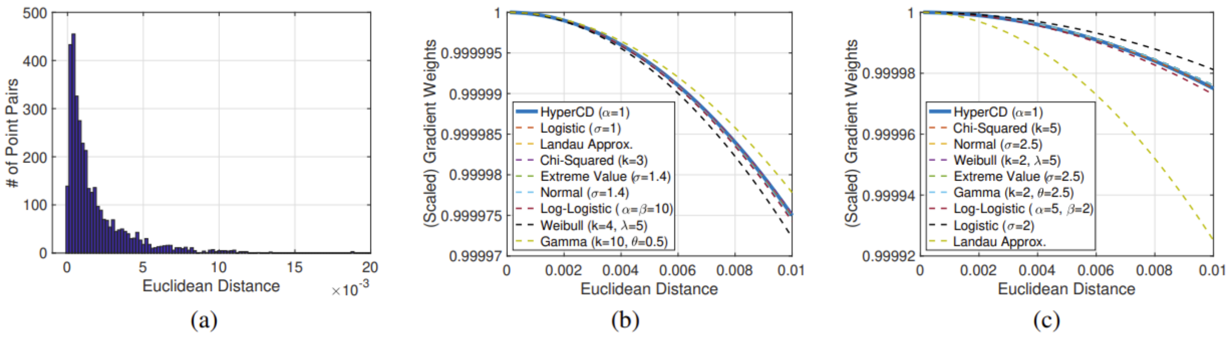


Fig 2. **(a)** Illustrate the distance distribution from HyperCD. In **(b-c)** respectively, where all the curves are rescaled by the maximum values and ordered by the minimum of Loss Distillation equation 3.

Point Cloud Completion with Weighted CD

Algorithm 2 Point Cloud Completion with Weighted CD

Input : a weighting function f, a network h with learnable parameters ω, training data $\{(\tilde{x}_i, y_i)\}$
Output: ω

repeat
Pick a sample $(\tilde{x}_i, y_i)$ uniformly at random;
Compute $\tilde{d}_{ijk}, \forall j$ in $\tilde{x}_i$ and $\tilde{d}_{ikj}, \forall k$ in y_i;
Compute the weighted CD loss based on Eq. 4;
Update the parameters ω using the gradient of the loss;
until *converges*;
return ω

$$\min_\omega \sum_i \left[\frac{1}{|x_i|}\sum_{j=1}^{|x_i|} f(\tilde{d}_{ijk})\tilde{d}_{ijk} + \frac{1}{|y_i|}\sum_{k=1}^{|y_i|} f(\tilde{d}_{ikj})\tilde{d}_{ikj} \right]$$

$$\text{s.t. } x_i = h(\tilde{x}_i; \omega) = \{x_{ij}\}, \forall i,$$
$$\tilde{d}_{ijk} = \min_k d(x_{ij}, y_{ik}), \tilde{d}_{ikj} = \min_j d(x_{ij}, y_{ik}).$$

Results

VISLab@WPI

- **Completion Results:**
  - Evaluated on both synthesis: PCN, ShapeNet-55, and real-word datasets: KITTI. One of the proposed weightedCD, namely LandauCD, consistently outperformed HyperCD and other variants in multiple benchmarks.

Methods	Plane	Cabinet	Car	Chair	Lamp	Couch	Table	Boat	Avg.
FoldingNet [41]	14.31	9.49	15.80	12.61	15.55	16.41	15.97	13.65	14.99
TopNet [45]	7.61	13.31	10.90	13.82	14.44	14.78	11.22	11.12	12.15
AtlasNet [46]	6.37	11.94	10.10	12.06	12.37	12.99	10.33	10.61	10.85
GRNet [47]	6.45	10.37	9.45	9.41	7.96	10.51	8.44	8.04	8.83
CRN [48]	4.79	9.97	8.31	9.49	8.94	10.69	7.81	8.05	8.51
NSFA [49]	4.76	10.18	8.63	8.53	7.03	10.53	7.35	7.48	8.06
FBNet [50]	3.99	9.05	7.90	7.38	5.82	8.85	6.35	6.18	6.94
PCN [36]	5.50	22.70	10.63	8.70	11.00	11.34	11.68	8.59	11.27
FoldingNet [41]	9.49	15.80	12.61	15.55	16.41	15.97	13.65	14.99	14.31
HyperCD + FoldingNet	7.89	12.90	10.67	14.55	13.87	14.09	11.86	10.89	12.09
InfoCD + FoldingNet	7.90	12.68	10.83	14.04	14.05	14.56	11.61	11.45	12.14
Landau CD + FoldingNet	7.30	12.69	10.46	13.00	11.92	13.39	10.86	10.59	11.27
PMP-Net [42]	5.65	11.24	9.64	9.51	6.95	10.83	8.72	7.25	8.73
HyperCD + PMP-Net	5.06	10.67	9.30	9.11	6.83	11.01	8.18	7.03	8.40
InfoCD + PMP-Net	4.67	10.09	8.87	8.59	6.38	10.48	7.51	6.75	7.92
Landau CD + PMP-Net	4.59	10.10	8.90	8.57	6.38	10.47	7.49	6.75	7.92
PoinTr [11]	4.75	10.47	8.68	9.39	7.75	10.93	7.78	7.29	8.38
HyperCD + PoinTr	4.42	9.77	8.22	8.22	6.62	9.62	6.97	6.67	7.56
InfoCD + PoinTr	4.06	9.42	8.11	7.81	6.21	9.38	6.57	6.40	7.24
Landau CD + PoinTr	4.12	9.49	8.07	7.82	6.30	9.28	6.76	6.41	7.28
SnowflakeNet [12]	4.29	9.16	8.08	7.89	6.07	9.23	6.55	6.40	7.21
HyperCD + SnowflakeNet	3.95	9.01	7.88	7.37	5.75	8.94	6.19	6.17	6.91
InfoCD + SnowflakeNet	4.01	8.81	7.62	7.51	5.80	8.91	6.21	6.05	6.86
Landau CD + SnowflakeNet	3.98	8.97	7.78	7.40	5.76	8.86	6.16	6.14	6.88
PointAttN [14]	3.87	9.00	7.63	7.43	5.90	8.68	6.32	6.09	6.86
DCD + PointAttN	4.47	9.65	8.14	8.12	6.75	9.60	6.92	6.67	7.54
HyperCD + PointAttN	3.76	8.93	7.49	7.06	5.61	8.48	6.25	5.92	6.68
InfoCD + PointAttN	3.72	8.87	7.46	7.02	5.60	8.45	6.23	5.92	6.65
Gamma CD + PointAttN	3.83	8.96	7.58	7.15	5.69	8.56	6.34	6.01	6.76
Chi-Squared CD + PointAttN	3.77	8.93	7.49	7.08	5.64	8.50	6.28	5.95	6.70
Log-Logistic CD + PointAttN	3.78	8.92	7.47	7.10	5.63	8.51	6.29	5.94	6.70
Extreme-Value CD + PointAttN	3.73	8.88	7.46	7.03	5.61	8.46	6.25	5.92	6.66
Landau CD + PointAttN	3.72	8.88	7.46	7.04	5.60	8.47	6.24	5.93	6.66
SeedFormer [13]	3.85	9.05	8.06	7.06	5.21	8.85	6.05	5.85	6.74
DCD + SeedFormer	16.42	26.23	21.08	20.06	18.30	26.51	18.23	18.22	24.52
HyperCD + SeedFormer	3.72	8.71	7.79	6.83	5.11	8.61	5.82	5.76	6.54
InfoCD + SeedFormer	3.69	8.72	7.68	6.84	5.08	8.61	5.83	5.75	6.52
Log-Logistic CD + SeedFormer	3.86	9.07	7.79	6.89	5.15	8.64	5.87	5.78	6.63
Gamma CD + SeedFormer	3.84	9.01	7.82	6.89	5.13	8.63	5.88	5.75	6.61
Chi-Squared CD + SeedFormer	3.75	8.90	7.71	6.80	5.11	8.48	5.77	5.68	6.53
Extreme-Value CD + SeedFormer	3.73	8.88	7.70	6.80	5.08	8.48	5.75	5.65	6.51
Landau CD + SeedFormer	3.65	8.68	7.64	6.80	5.04	8.57	5.79	5.71	6.49

Table. 1 Comparison on PCN dataset in terms of per-point L1-CD x 1000 (Lower the better).

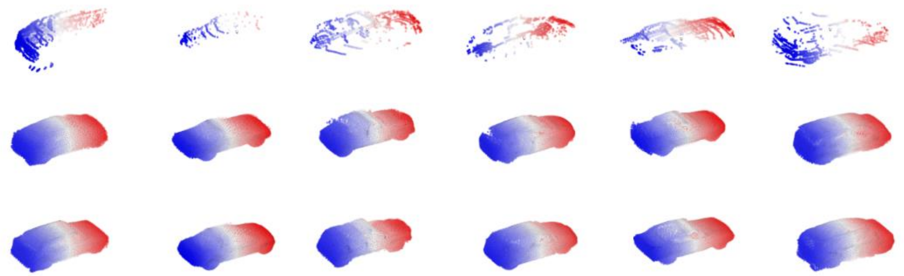


Fig. 3 Visualization of the KITTI benchmark, Row 1: input, Row 2: HyperCD, and Row 3: LandauCD.

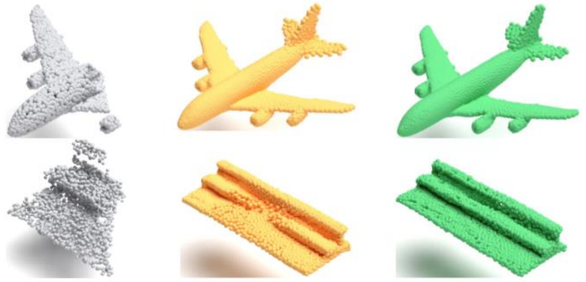


Fig.4 Visualization of ShapeNet-55 benchmark. Gray is the input. Yellow is HyperCD. Green is Landau CD.

Results

- **Training Analysis:**
  - In Fig. 5, our weightedCD losses exhibit a more rapid convergence compared to HyperCD.
  - In Table.2, the training efficiency still consist with original target CD.

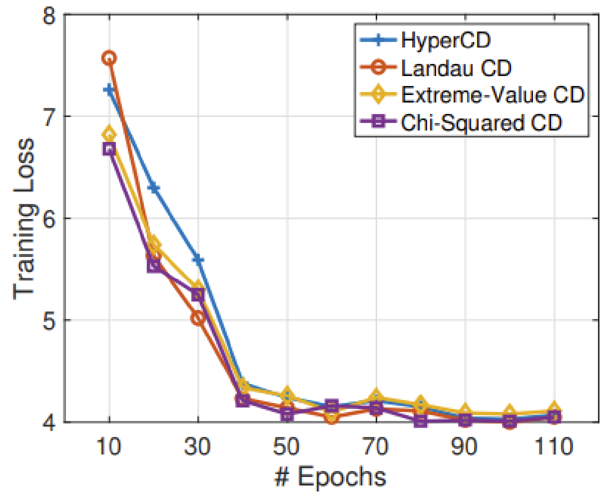


Fig. 5 Convergence comparison

Loss Functions	Computation Time
HyperCD	0.4239±0.0019
InfoCD	0.4298±0.0014
LandauCD	0.4341±0.0026

Table. 2 Training efficiency comparison

Conclusion

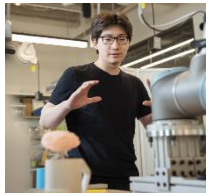
- A novel loss distillation method for point cloud completion by mimicking the learning behavior of HyperCD based on weighted CD.
- An efficient and effective gradient matching algorithm to search for potential weighting functions for weighted CD.
- A bilevel optimization problem to train backbone networks, based on our iterative differentiation algorithm.
- We conduct comprehensive experiments with different datasets to demonstrate the effectiveness of weighted CD losses

Our Team

VISLab@WPI

Fangzhou Lin: PhD student at WPI and Tohoku University

Haotian Liu: Undergraduate student at WPI and research intern at NEU



Haoying Zhou: PhD student at WPI and visiting student at JHU

Songlin Hou: Research engineer at Dell Technology

Dr. Ziming Zhang: Assistant Professor at WPI ECE department