Overview

We provide additional details and results to complement the main paper. Specifically, this document includes the following materials:

- Broader impact (Appendix A).
- Details of evaluation metrics (Appendix B).
- Details of baseline methods (Appendix C).
- Details of training losses (Appendix D).
- Additional training and implementation details (Appendix E).
- Additional case study evaluations (Appendix F).
- License information (Appendix G).
- Dataset accessibility and long-term preservation plan (Appendix H).
- Structured metadata (Appendix I).
- Dataset identifier (Appendix J).
- Author statement of responsibility (Appendix K).
- Datasheet for dataset (Appendix L).

A Broader Impact

Fracture reassembly is an important task in the real world, e.g. recovering a shattered artifact or a broken item of kitchenware. In recent years, machine learning algorithms trained on large-scale datasets have shown great advance on this task. We believe Breaking Bad benchmark is an important step towards teaching machines to reassemble physically plausible fractured objects. Our dataset will facilitate future study in this field, and eventually enable robots to free human on the part assembly task. We believe this research benefits both the economy and society.

Potential negative societal impacts. We do not see significant risks of human rights violations or security threats in our dataset and its potential applications. However, since Breaking Bad contributes to the entire field of fracture reassembly, it might trigger further concerns regarding the assembly algorithms. Even with advanced learning methods, human trust in AI is still a problem. For example, since some object fractures have very sharp edges, the imperfect assembly result could harm the users. Therefore, the trained algorithms should be used under supervision and cannot fully replace human. Finally, the fracture simulation code used for dataset generation could be adopted to break down sculptures of humans, which would potentially cause harm to people. Overall, the technical outcomes of this paper need to cooperate with humans to avoid negative societal impacts.

B Evaluation Metrics

Shape chamfer distance. The chamfer distance $\text{CD}(P, Q)$ between two point clouds P and Q is defined as

$$\text{CD}(P, Q) = \sum_{x \in P} \min_{y \in Q} \|x - y\|_2^2 + \sum_{y \in Q} \min_{x \in P} \|y - x\|_2^2. \tag{4}$$

Shape chamfer distance $\text{CD}(S, S^*)$ is computed between the predicted assembly S and the ground-truth assembly S^*.

Part accuracy. Part accuracy measures the percentage of parts whose chamfer distance to ground-truth is less than a threshold τ and is defined as

$$\text{PA} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\Big(\text{CD}\Big(q_i(P_i),\ q_i^*(P_i)\Big) < \tau\Big), \tag{5}$$

where we set $\tau = 0.01$ following [20].

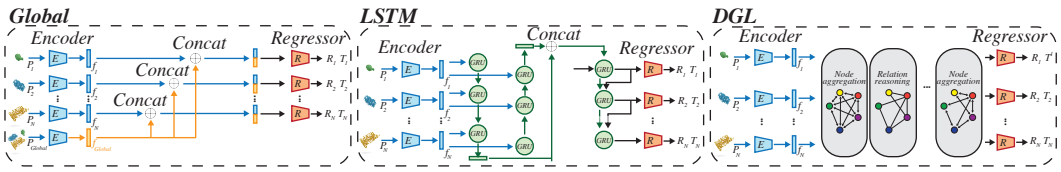

Figure 6: **Baseline model architectures.** (*Left*) Global. (*Middle*) LSTM. (*Right*) DGL.

C Details of Baseline Methods

Figure 6 shows the architecture of the three baseline methods, i.e., Global, LSTM and DGL.

Global. We first extract the part feature for each input point cloud and the global feature using PointNet [44]. Then, we concatenate the global feature with each part feature and apply a shared-weight MLP network to regress the SE(3) pose for each input point cloud.

LSTM. To better learn cross-piece relationship, we develop a bi-directional LSTM [49] module that takes as input part features and sequentially predicts the SE(3) pose for each input point cloud. This resembles the process of sequential decision making when humans perform shape assembly.

DGL. Graph neural networks (GNNs) encode part features via edge relation reasoning and node aggregation modules. We remove the node aggregation operation designed for handling geometrically-equivalent parts in DGL, since every piece in our dataset has a unique shape geometry.

In all three models, the point cloud encoder is implemented as PointNet [44] and the pose regressor is implemented as MLP with ReLU non-linearity. The output rotation is parametrized using quaternion representation.

In our experiments, we adopt the official implementations of Global, LSTM and DGL from `https://github.com/hyperplane-lab/Generative-3D-Part-Assembly`. We also include our benchmark code in the supplementary material and will make them publicly available upon acceptance.

D Training Losses

Following previous methods [20, 28, 48, 62], all models are trained with the pose regression loss $\mathcal{L}_{\text{pose}}$, the chamfer distance loss $\mathcal{L}_{\text{chamfer}}$ and the point-to-point mean square error (MSE) loss $\mathcal{L}_{\text{point}}$. We use the pose regression loss $\mathcal{L}_{\text{pose}}$ [7], the chamfer distance loss $\mathcal{L}_{\text{chamfer}}$ [20] and the point-to-point MSE loss $\mathcal{L}_{\text{point}}$ to train each of the baseline models. Denote the ground-truth SE(3) pose as $q_i^* = \{(R_i^*,\ T_i^*)\}$.

Pose regression loss $\mathcal{L}_{\text{pose}}$. The pose regression loss $\mathcal{L}_{\text{pose}}$ is defined as

$$\mathcal{L}_{\text{pose}} = \sum_{i=1}^{N} \|T_i - T_i^*\|_2^2 \ + \ \lambda_{\text{rot}} \|R_i^\top R_i^* - I\|_2^2, \tag{6}$$

where I is the identity matrix.

Chamfer distance loss $\mathcal{L}_{\text{chamfer}}$. We also minimize the chamfer distance between the predicted pose-transformed point clouds and the ground truths, as well as the predicted assembly and the ground truth. The chamfer distance loss $\mathcal{L}_{\text{chamfer}}$ is defined as

$$\mathcal{L}_{\text{chamfer}} = \sum_{i=1}^{N} \text{CD}(R_i P_i,\ R_i^* P_i) + \lambda_{\text{shape}} \text{CD}(S,\ S^*). \tag{7}$$

Point-to-point MSE loss $\mathcal{L}_{\text{point}}$. We empirically observe that adding a point-to-point MSE loss helps improve rotation prediction. Denote P_i^j as the j-th point in the i-th fracture piece point cloud. We minimize the ℓ_2 distance between point clouds transformed by the predicted rotation and by the

Table 7: **Evaluation on fracture reassembly.** We report the results of three learning-based shape assembly models on the `everyday` object subset. The results are averaged over all 20 categories.

Method	RMSE (R) ↓	MAE (R) ↓	RMSE (T) ↓	MAE (T) ↓	CD ↓	PA ↑
	degree	degree	$\times 10^{-2}$	$\times 10^{-2}$	$\times 10^{-3}$	%
Global	80.7	68.0	15.1	12.0	14.6	24.6
LSTM	84.2	72.4	16.2	12.6	15.8	22.7
DGL	79.4	66.5	15.0	11.9	14.3	31.0

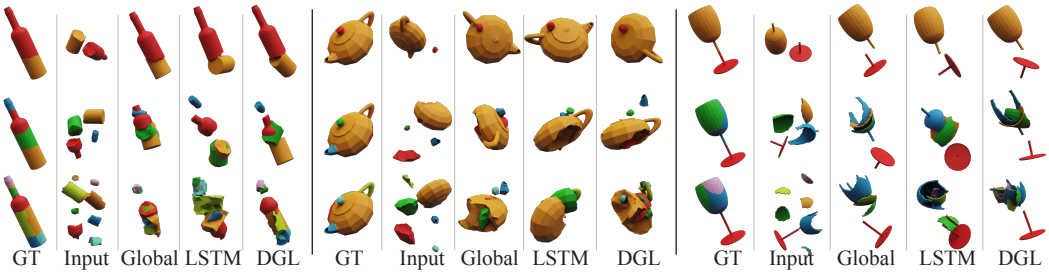

Figure 7: **Visual results of fracture reassembly.** We present visual results and comparisons between methods on the everyday object subset.

ground-truth rotation, respectively, as

$$\mathcal{L}_{\text{point}} = \sum_{i=1}^{N} \sum_{j} ||R_i P_i^j - R_i^* P_i^j||_2^2. \tag{8}$$

Total loss $\mathcal{L}$. Each baseline model is trained by optimizing the following objective function with hyperparameters $\lambda_{\text{chamfer}}, \lambda_{\text{point}}$ balancing the three loss terms

$$\mathcal{L} = \mathcal{L}_{\text{pose}} + \lambda_{\text{chamfer}} \mathcal{L}_{\text{chamfer}} + \lambda_{\text{point}} \mathcal{L}_{\text{point}}. \tag{9}$$

E Training and Implementation Details

We implement all baseline models using PyTorch [41]. We adopt the Adam optimizer [26] for training. The learning rate is initialized at 1×10^{-3} and decayed to 1×10^{-5} using a cosine schedule [31]. We report the performance averaged over three runs for each baseline method. We train all the baseline models on one NVIDIA RTX6000 GPU for 200 epochs with a batch size of 32. We perform early stopping when the model achieves best performance on the validation set. The hyperparameters are set to $\lambda_{\text{rot}} = 0.2$, $\lambda_{\text{shape}} = 1$, $\lambda_{\text{chamfer}} = 10$, and $\lambda_{\text{point}} = 1$. We choose them on the Global model via cross-validation and fix them for all baseline methods in all experiments.

F Additional Case Study Evaluations

F.1 Additional Results of Fracture Reassembly

Table 7 and Figure 7 present additional quantitative and qualitive results of fracture reassembly on the `everyday` object subset, respectively.

F.2 Additional Results of Ablation Study

Table 8 reports additional quantitative results of ablation study on the `everyday` object subset.

F.3 Additional Results of Model Pre-training and Fine-tuning

Table 9 reports additional quantitative results of model pre-training and fine-tuning.

Table 8: **Ablation study: Number of fracture pieces.** We train DGL on the `everyday` object subset in three settings and report performance evaluated on fractured objects of different numbers of fracture pieces. The results are averaged over all 20 categories.

Test set	RMSE (R) ↓	MAE (R) ↓	RMSE (T) ↓	MAE (T) ↓	CD ↓	PA ↑
pieces	degree	degree	$\times 10^{-2}$	$\times 10^{-2}$	$\times 10^{-3}$	%
Results of training on fractured objects with 2 to 20 fracture pieces						
2-20	79.4	66.5	15.0	11.9	14.3	31.0
21-50	84.4	72.7	20.1	16.4	15.0	7.5
51-100	85.1	73.7	21.3	17.3	23.0	4.8
Results of training on fractured objects with 2 to 50 fracture pieces						
2-20	79.9	67.0	14.8	11.7	14.0	29.9
21-50	84.5	72.9	19.6	16.0	14.0	7.7
51-100	84.8	73.5	20.5	16.8	18.1	4.7
Results of training on fractured objects with 2 to 100 fracture pieces						
2-20	79.8	67.0	14.4	11.4	14.0	29.4
21-50	84.3	72.8	19.2	15.7	14.5	7.4
51-100	85.3	74.1	20.0	16.4	13.9	4.8

Table 9: **Analysis of model pre-training and fine-tuning.** We report the results of three learning-based shape assembly models on the `artifact` subset.

Method	RMSE (R) ↓	MAE (R) ↓	RMSE (T) ↓	MAE (T) ↓	CD ↓	PA ↑
	degree	degree	$\times 10^{-2}$	$\times 10^{-2}$	$\times 10^{-3}$	%
Results of training the model from scratch						
Global	84.8	73.0	16.7	14.0	19.0	12.7
LSTM	85.2	73.6	17.2	14.3	23.5	6.6
DGL	85.8	74.2	16.8	13.9	19.4	12.8
Results of fine-tuning from the model in Table 3						
Global	83.8	71.8	16.6	13.8	19.0	13.3
LSTM	84.6	73.1	16.8	14.0	21.5	11.7
DGL	81.7	69.7	16.6	13.8	17.3	19.4

F.4 Additional Results of Generalization to Unseen Objects

Table 10 reports additional quantitative results of generalization to unseen objects.

F.5 Evaluation of Models Trained on Each Category

Table 11 top block reports the results of the three baseline models trained on each cateogry in the `everyday` object subset independently. DGL achieves the best performance in most categories, demonstrating the importance of using GNNs for relation reasoning between fracture pieces.

F.6 Evaluation of Models Trained on 20 Categories All Together

Table 11 bottom block presents the results of the three baseline models trained on 20 categories in the `everyday` object subset altogether. Compared to models trained on each respective categories, all three models achieve similar results in terms of translation and rotation, but have worse performance in chamfer distance and part accuracy. We hypothesize that this is because when trained with objects of a single category, the model is able to learn shape priors for this category, guiding the model to correctly predict an SE(3) pose for each fracture piece. The comparison also shows that training a model that generalizes to different object categories is an open challenge.

Table 10: **Generalization to unseen objects.** We report the results of three learning-based shape assembly models on the `other` subset.

Method	RMSE (R) ↓	MAE (R) ↓	RMSE (T) ↓	MAE (T) ↓	CD ↓	PA ↑
	degree	degree	$\times 10^{-2}$	$\times 10^{-2}$	$\times 10^{-3}$	%
Results of testing the model in Table 3						
Global	86.4	72.9	19.4	16.3	42.2	6.0
LSTM	84.9	73.1	18.7	15.5	45.3	4.8
DGL	86.6	73.5	20.1	16.6	38.5	7.5
Results of testing the model in the bottom block of Table 5						
Global	83.9	71.9	18.8	15.5	39.2	6.7
LSTM	82.9	70.3	17.9	14.9	40.3	5.5
DGL	81.3	69.9	17.2	14.5	36.6	8.3

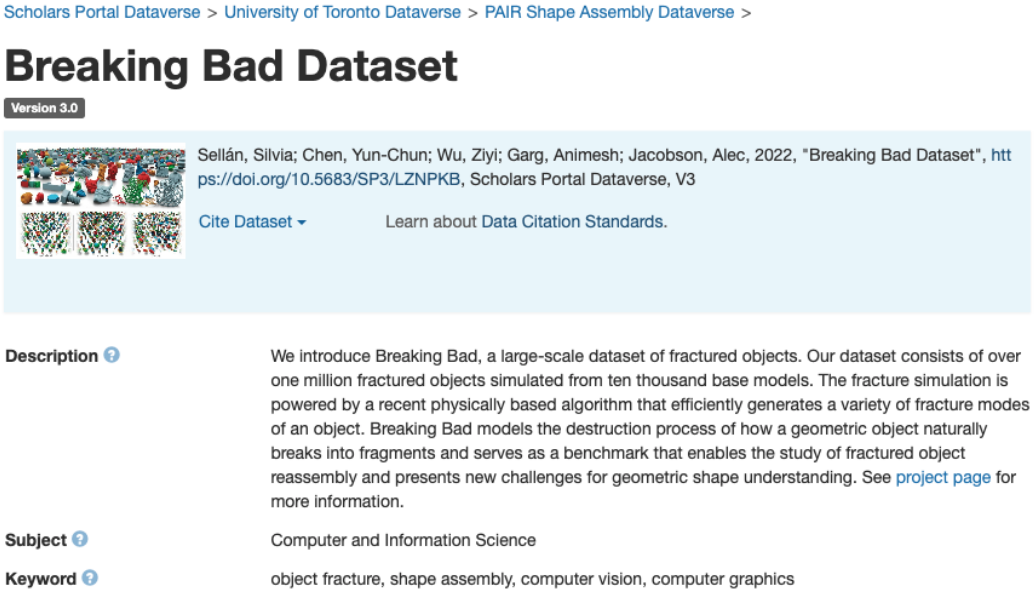

Figure 8: **Screenshot of the Dataverse platform.** The Breaking Bad dataset is hosted on the Dataset plaform and can be accessed at `https://doi.org/10.5683/SP3/LZNPKB`.

G License Information

The licenses of code and data is summarized in Section 4.4. The baseline code will be released under the MIT license.

H Dataset Accessibility and Long-Term Preservation Plan

The instructions to access our dataset is summarized on our project page at `https://breaking-bad-dataset.github.io/`.

The dataset is stored on the Dataverse platform (`https://dataverse.scholarsportal.info/`), which is a multidisciplinary, secure, Canadian research data repository, supported by academic libraries and research institutions across Canada. Dataverse has been running for more than 5 years and hosted over 5,000 datasets. We believe that this platform can ensure the stable accessibility and long-term perservation of our dataset.

A screenshot of our dataset hosted on the Dataverse platform is shown in Figure 8.

I Structured Metadata

Since we are using an existing data repository to host our dataset, the structured metadata is automatically generated by the platform.

J Dataset Identifier

The DOI for our dataset is doi:10.5683/SP3/LZNPKB.

The benchmark code is attached in the supplementary material and will be released on Github upon acceptance.

K Author Statement of Responsibility

The authors confirm all responsibility in case of violation of rights and confirm the licence associated with the dataset and code.

L Datasheets for Dataset

We provide our responce in reference to the Datasheets for Datasets [11] standards.

L.1 Motivation

- **For what purpose was the dataset created?** To study the geometric fracture reassembly task.
- **Who created the dataset and on behalf of which entity?** The authors listed on this paper, which are researchers from the University of Toronto.
- **Who funded the creation of the dataset?** The research is funded in part by NSERC Discovery (RGPIN2017–05235, RGPAS–2017–507938), New Frontiers of Research Fund (NFRFE–201), the Ontario Early Research Award program, the Canada Research Chairs Program, a Sloan Research Fellowship, the DSI Catalyst Grant program and gifts by Adobe Systems.

L.2 Composition

- **What do the instances that comprise the dataset represent?** Each data instance contain several 3D meshes, which are fractures simulated from breaking an object under random external impact.
- **How many instances are there in total?** 1,047,400.
- **Does the dataset contain all possible instances or is it a sample of instances from a larger set?** Our data is generated by a simulation program. Therefore, it is a subset of all possible simulation outcomes.
- **What data does each instance consist of?** It consists of several 3D mesh files which are fractures of an entire object.
- **Is there a label or target associated with each instance?** No.
- **Is any information missing from individual instances?** No.
- **Are relationships between individual instances made explicit?** Yes. All instances are grouped based on the object category.
- **Are there recommended data splits?** Yes. We provide the data splits used for evaluation with the benchmark code.
- **Are there any errors, sources of noise, or redundancies in the dataset?** Probably, but we are not aware of them at the moment. If any are discovered, they will be fixed and versioned.

- **Is the dataset self-contained, or does it link to or otherwise rely on external resources?** It is self-contained.
- **Does the dataset contain data that might be considered confidential?** No.
- **Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety?** No.
- **Does the dataset relate to people?** No.

L.3 Collection Process

- **How was the data associated with each instance acquired? What mechanisms or procedures were used to collect the data? How was it verified?** Object meshes are obtained from previous public datasets and filtered based on certain criterions discussed in the main paper. The object fractures are generated by a simulation algorithm which is conditionally accepted to a peer-reviewed journal ACM TOG. Prior to running simulation on all the object meshes, we test it on a few examples to adjust hyper-parameters and verify their correctness by human eyes.
- **Who was involved in the data collection process and how were they compensated?** The authors listed on this paper.
- **Over what timeframe was the data collected?** All simulated fractures were created during early April 2021.
- **Were any ethical review processes conducted?** Not applicable (no human data collection).
- **Does the dataset relate to people?** No.

L.4 Preprocessing, clearning and labelling

- **Was any preprocessing/cleaning/labeling of the data done?** No. All the data were generated by the simulation program. The labels are automatically obtained from the original shape datasets.

L.5 Uses

- **Has the dataset been used for any tasks already?** Yes. We have benchmarked existing 3D shape assembly methods on our dataset in this paper.
- **Is there a repository that links to any or all papers or systems that use the dataset?** No other papers use the dataset yet.
- **What (other) tasks could the dataset be used for?** Apart from fracture reassembly, the dataset could also be used for training machine learning algorithms to predict fractures of objects under external force.
- **Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?** No.
- **Are there tasks for which the dataset should not be used?** All tasks involving object part-whole study are valid. The dataset may not be suitable for other tasks.

L.6 Distribution

- **Will the dataset be distributed to third parties outside of the entity on behalf of which the dataset was created?** Yes. It will be completely publicly available via a project page based on a Github repo and the links listed thereupon.
- **How will the dataset will be distributed?** The code for data generation will be available on Github. The data will be released on a permanent data hosting platform.
- **When will the dataset be distributed?** Immediately.
- **Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)?** Yes. See the discussion of license in Section 4.4. We also detail it in the corresponding field on the Dataverse platform.

- **Have any third parties imposed IP-based or other restrictions on the data associated with the instances?** No.
- **Do any export controls or other regulatory restrictions apply to the dataset or to individual instances?** No.

L.7 Maintenance

- **Who is supporting/hosting/maintaining the dataset?** Code for data generation is hosted in a publicly-accessible Github repo. The Github account is owned by the PAIR Lab which the authors belong to. The data hosting platform is also publicly-accessible and maintained by the University of Toronto Libraries.
- **How can the owner/curator/manager of the dataset be contacted?** The corresponding author of the paper can be contacted via email listed in the authorlist.
- **Is there an erratum?** Not yet. But there will be in the future if we discover any errors in the dataset. It will be displayed on the project page.
- **Will the dataset be updated?** If any errors of the data are discovered, we will update future versions of the dataset.
- **Will older versions of the dataset continue to be supported/hosted/maintained?** Yes. All data will be versioned.
- **If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so?** Yes. As we provide the code for fracture simulation, users can easily generate new object fractures based on meshes they provide. Also, users can submit errors they discover via Github issues or emails, and discuss potential improvement with authors.

Table 11: **Evaluation of fracture reassembly.** (*Top block*) Results are reported using the models trained on each respective categories in the `everyday` object subset independently. (*Bottom block*) Results are reported using the models trained on all categories in the `everyday` object subset together.

Method	Beer Bottle	Bowl	Cup	Drinking Utensil	Mug	Plate	Spoon	Tea Cup	Toy	Wine Bottle	Bottle	Cookie	Drink Bottle	Mirror	Pill Bottle	Ring	Statue	Teapot	Vase	Wine Glass	Mean
Models trained only on each respective categories in the `everyday` object subset																					
RMSE (R) (deg.)																					
Global	86.2	67.7	74.4	75.6	76.5	86.4	85.9	80.9	85.1	76.0	78.5	87.0	78.9	85.2	85.5	82.1	85.6	82.6	78.8	75.9	80.7
LSTM	87.0	77.0	82.1	84.1	80.8	86.8	85.7	86.6	86.7	79.7	80.4	87.3	83.4	87.0	85.9	83.2	86.3	84.6	85.5	83.7	84.2
DGL	84.3	67.8	74.2	75.4	75.9	86.8	82.4	71.2	84.5	76.1	78.9	87.1	73.9	86.6	87.7	81.8	82.9	83.0	78.6	68.3	79.4
MAE (R) (deg.)																					
Global	72.9	54.2	61.8	63.2	63.2	70.9	74.1	68.5	73.3	62.7	65.3	74.5	66.4	73.0	73.0	70.2	72.8	71.9	66.0	62.8	68.0
LSTM	74.9	64.7	70.4	72.7	69.6	73.5	74.5	74.6	75.6	67.1	67.7	74.9	71.2	75.3	73.8	71.9	74.3	73.6	74.2	73.0	72.4
DGL	70.4	54.4	61.4	62.7	62.9	71.3	70.7	58.4	72.9	62.6	65.6	73.4	59.9	73.9	75.4	70.1	70.7	72.6	65.6	54.2	66.5
RMSE (T) ($\times 10^{-2}$)																					
Global	10.5	17.6	17.9	18.2	18.4	17.2	18.5	12.5	17.5	6.6	11.3	17.1	9.4	15.7	17.5	21.0	17.3	15.4	16.0	7.2	15.1
LSTM	11.0	18.0	20.0	21.6	20.7	18.3	18.1	14.5	17.5	7.4	11.9	17.4	10.2	16.0	20.1	22.8	19.7	11.9	18.1	8.1	16.2
DGL	6.4	18.0	18.8	15.9	21.5	17.3	19.4	14.3	18.5	4.7	10.5	20.1	6.8	15.7	21.6	19.2	16.2	13.6	15.1	6.0	15.0
MAE (T) ($\times 10^{-2}$)																					
Global	7.2	14.8	15.2	15.5	15.7	13.4	13.3	10.6	14.3	4.7	8.9	12.5	6.8	12.2	14.5	17.3	12.7	12.0	12.7	5.9	12.0
LSTM	7.4	14.9	16.5	18.0	17.2	14.0	13.1	12.3	14.4	5.1	9.1	13.2	7.3	12.4	16.1	17.6	14.3	9.3	13.9	6.6	12.6
DGL	4.6	15.1	15.6	13.5	17.9	13.4	15.3	11.7	15.1	3.4	8.3	15.3	5.1	12.1	17.1	15.4	12.1	10.9	11.9	4.9	11.9
Chamfer Distance ($\times 10^{-3}$)																					
Global	6.4	11.0	20.7	15.0	20.2	6.2	38.8	10.5	22.7	2.3	10.8	8.9	4.3	23.3	17.0	31.6	8.0	11.9	13.6	8.3	14.6
LSTM	4.8	13.2	24.4	26.5	20.2	7.1	55.2	12.1	25.8	1.4	11.3	8.5	3.0	22.7	16.4	22.9	6.0	8.1	17.5	9.6	15.8
DGL	1.5	8.3	15.2	9.6	15.7	4.6	89.2	7.3	23.4	0.7	9.8	5.2	1.3	24.5	15.4	27.7	5.4	8.5	8.8	3.5	14.3
Part Accuracy (%)																					
Global	50.2	16.0	14.5	21.9	11.5	13.8	7.8	15.5	7.2	65.4	44.9	20.0	56.5	10.0	16.3	4.9	17.2	41.0	23.5	34.4	24.6
LSTM	48.5	14.2	8.5	5.7	5.9	11.5	9.9	5.9	5.8	64.2	46.2	21.5	50.0	8.9	17.0	6.3	20.2	54.3	15.3	33.9	22.7
DGL	71.6	17.3	20.3	26.6	12.2	13.2	6.6	25.9	8.4	77.2	50.2	23.1	67.4	11.0	17.2	13.9	23.2	50.2	30.7	54.0	31.0
Models trained on 20 categories altogether in the `everyday` object subset																					
RMSE (R) (deg.)																					
Global	90.2	69.7	74.2	73.9	76.6	78.6	83.3	77.8	88.8	77.8	80.3	89.1	77.5	85.0	89.3	90.0	79.3	82.4	78.8	83.2	81.3
LSTM	86.2	83.1	81.0	79.0	80.1	85.7	85.1	88.4	87.0	82.1	82.8	87.2	82.3	84.7	86.2	90.2	84.5	83.7	82.7	88.5	84.5
DGL	88.4	69.9	75.2	77.0	76.7	74.7	81.1	76.4	86.6	75.8	80.4	90.3	74.4	85.7	87.4	85.5	86.0	78.4	80.5	82.4	80.6
MAE (R) (deg.)																					
Global	76.8	57.4	61.6	61.1	64.3	67.5	71.3	66.4	76.8	64.6	67.8	78.3	64.7	72.9	77.5	78.7	67.0	71.4	66.0	70.8	69.1
LSTM	74.6	71.6	69.2	66.6	68.1	74.4	73.8	76.9	75.4	70.2	71.0	76.2	70.2	73.4	74.8	78.3	73.1	72.6	70.8	76.5	72.9
DGL	74.9	57.1	62.4	63.5	64.4	61.3	68.8	64.0	74.7	62.8	67.1	78.0	60.4	73.9	75.0	73.4	74.0	67.2	67.3	69.3	68.0
RMSE (T) ($\times 10^{-2}$)																					
Global	12.1	17.7	16.8	17.9	18.4	17.5	16.7	13.9	17.5	7.3	12.1	17.9	8.6	16.8	17.1	23.3	15.9	13.9	16.8	10.7	15.4
LSTM	10.4	17.9	17.9	18.6	19.0	17.9	17.0	13.8	18.2	7.7	12.5	18.0	9.2	18.0	18.5	22.5	14.6	13.2	17.3	9.7	15.6
DGL	10.0	18.3	17.1	17.9	19.4	17.4	19.6	14.2	18.7	5.0	11.4	18.8	6.0	16.5	15.8	22.5	14.4	14.4	15.2	6.2	14.9
MAE (T) ($\times 10^{-2}$)																					
Global	8.2	14.6	13.9	15.3	15.6	14.2	12.1	12.1	14.4	5.2	9.4	13.6	6.4	13.3	14.3	17.5	11.6	11.2	13.2	8.6	12.2
LSTM	7.2	14.7	14.8	15.7	16.0	14.1	12.1	11.9	14.8	5.3	9.5	14.4	6.6	14.3	15.3	16.8	10.9	11.2	13.5	7.7	12.3
DGL	6.8	15.0	14.2	15.1	16.2	13.9	15.2	11.9	15.1	3.7	8.9	14.5	4.5	12.8	12.9	17.5	10.9	11.7	12.0	5.1	11.9
Chamfer Distance ($\times 10^{-3}$)																					
Global	4.9	15.9	16.6	16.8	18.5	37.8	46.0	14.2	19.6	2.2	11.7	33.4	2.2	27.8	14.9	61.0	8.4	13.6	15.1	14.6	19.8
LSTM	4.9	21.0	19.0	17.4	18.3	32.1	49.6	16.4	20.2	2.1	12.4	35.6	2.4	33.8	18.2	58.2	6.2	14.0	15.8	13.6	20.6
DGL	3.0	9.1	11.5	8.9	14.4	26.0	87.8	11.5	20.4	1.0	11.1	10.5	1.0	27.3	17.2	45.4	8.1	10.0	9.4	6.2	17.0
Part Accuracy (%)																					
Global	44.9	11.2	17.9	22.5	10.8	1.5	12.0	7.2	12.8	61.9	39.6	1.3	60.0	5.0	15.9	0.1	17.8	34.3	19.4	21.8	20.9
LSTM	50.3	5.8	13.4	18.8	8.6	2.2	8.3	4.1	9.5	61.5	38.4	2.7	55.4	2.5	13.6	1.0	23.2	34.9	19.1	26.4	20.0
DGL	59.9	15.9	22.7	24.5	14.2	9.2	3.5	14.5	11.8	73.9	46.6	18.3	67.7	8.7	23.9	3.1	26.2	41.1	30.1	46.4	28.1