

A Appendix

A.1 Data processing

Point cloud generation. We process 3D CAD models to point cloud by sampling 4,096 points for each model, using Kaolin [30]. Then we apply min-max normalization to the sampled point cloud for each model.

Chamfer distance computation. The input of the Chamfer distance is the pre-processed point cloud for each CAD model that are normalized using min-max normalization (e.g. for each dimension, the min and max will be 0 and 1 after the normalization, and the scaling is affine).

Jaccard distance computation. Pairwise IoU (Intersection over Union, Jaccard index) of the objects are calculated first. The input of the IoU is voxel grids generated from the the pre-processed point cloud for each CAD model that are normalized using min-max normalization. The IoU is not normalized for the case when the object is too thin to normalize, and thus the IoU of it versus anything will be zero. Then the Jaccard distance will be $1 - \text{IoU}$ by definition.

Clustering result visualization. As shown in Figure 4, for each baseline method, we select the corresponding cluster which contains the same input CAD model. In this example, all the baseline methods have the capability to group similar objects. To make a comparison among all these baseline methods, we can see that the end-to-end deep clustering methods (DeepCluster, DEC, IIC, SCAN) overall outperform the two-stage clustering methods (AtlasNet, BYOL, MVCNN based methods): the general shapes of the clustered objects for deep clustering methods are more similar than those obtained by two-stage clustering methods.

A.2 Human annotator instruction and payment

As we discussed in Section 3, the human standard to determine the similarity relationship between every two CAD models relies on previous expert knowledge that is difficult to describe. We provided the following instructions to our human annotators, to help them be more consistent. First, the similarity should be determined by the geometry, instead of the functionality. Second, the annotators should assume that the similarity relationship is rotational and translational invariant.

We pay the human annotators 10 dollars per hour. In total, we spent around 6,000 dollars for human annotations.

A.3 Independent benchmark results

Since we have annotation from three annotators, we produce benchmark results by using evaluation matrix with each similarity matrix from three annotators independently and have the average as our final result shown in Figure 3. As shown in Figure 5, we provide each of benchmark results generated from independent annotation of three annotators.











































BYOL	AtlasNet	MVCNN	SCAN	IIC	DEC	DeepCluster
						
						
						
						
						
						

Figure 4: **Clustering result visualization.** We select one CAD model (first row), and visualize all the objects that are grouped with this model for all baseline methods.

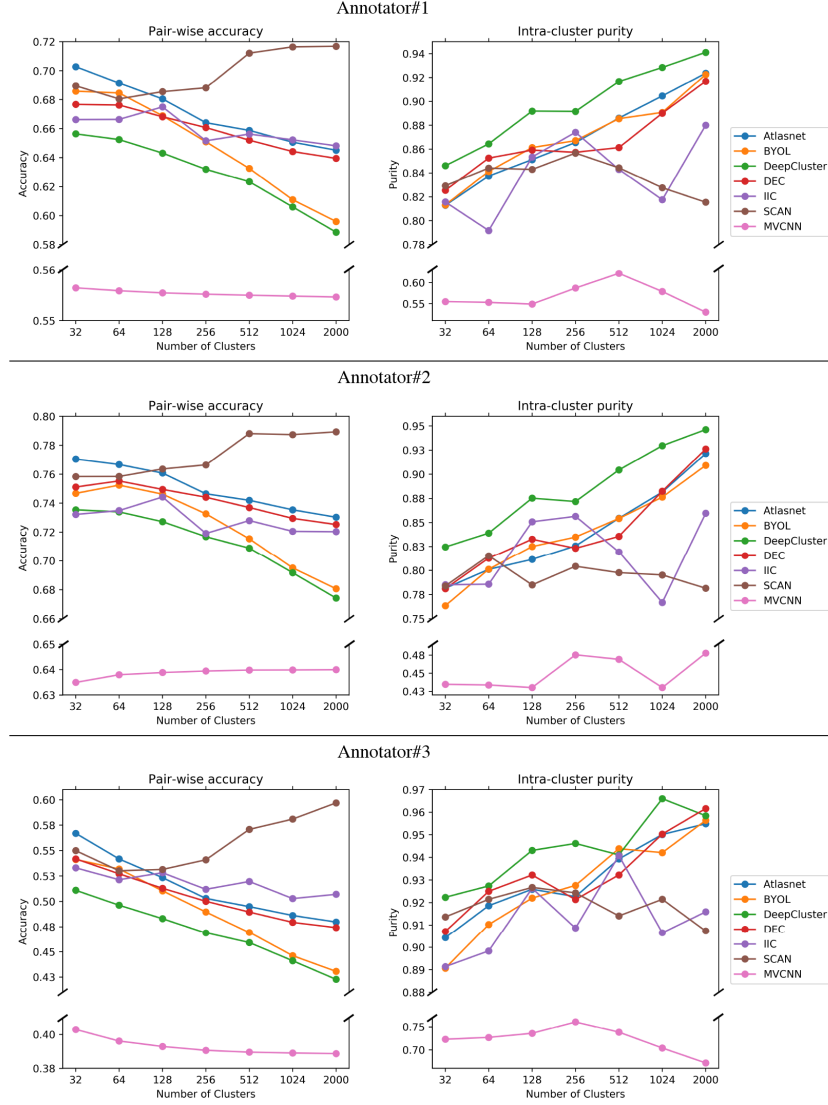


Figure 5: Independent Cluster3D benchmark results.