GEOPARD: Geometric Pretraining for Articulation Prediction in 3D Shapes

Supplementary Material

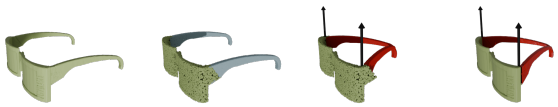

Figure 1. Results using PartSTAD, an off-the-shelf segmentation method. *(1)* input shape, *(2)* PartSTAD segmentation, *(3)* sampled articulated state from predicted parameters. *(4)* same articulation state from GT segmentation and GT parameters.

Model	AE ↓	PE ↓	R-ACC ↑	P-ACC ↑
GEOPARD-u-GT	7.24	0.07	0.96	0.98
GEOPARD-u-PartSTAD	9.34	0.08	0.95	0.93

Table 1. Using off-the-shelf vs ground-truth segmentation.

1. Off-the-shelf segmentation

We evaluated the robustness of our approach using part segmentations produced by PartSTAD [**?**], a few-shot, VLM-based point cloud model. Since VLMs perform surface segmentation, we filter out object classes with interior structures and test on the following classes: *box*, *dishwasher*, *door*, *eyeglasses*, *refrigerator*, and *trash can*.

Table 1 shows the average quantitative results across these classes in the unlabeled setting. Although there is a slight degradation in metrics compared to ground-truth ("GT") segmentation, overall performance remains competitive, indicating that our method can maintain reliable articulation predictions under imperfect segmentation. We illustrate this qualitatively in Figure 1 – our method predicts the correct rotation for test shape even with segmentation noise.

We note that shape segmentation is orthogonal to our contributions; our method may benefit further from advances in this space.

2. Implementation details

Our network stacks 6 layers of cross-attention and 12 layers of self-attention, each with 8 heads. The decoder consists of lightweight 2-layer MLP blocks whose output dimensionality adapts to the target parameter set (pretraining vs. finetuning). We use a batch size of 60 and train on 2 NVIDIA A40 GPUs.