Unsupervised Learning of 3D Semantic Keypoints with Mutual Reconstruction

Haocheng Yuan[1], Chen Zhao[2], Shichao Fan[1], Jiaxi Jiang[1] and Jiaqi Yang*[1]

[1] Northwestern Polytechnical University, China
{hcyuan, fsc_smile, jshmjjx}@mail.nwpu.edu.cn,
jqyang@nwpu.edu.cn,
[2] École Polytechnique Fédérale de LausanneL, Switzerland
chen.zhao@epfl.ch

1 Additional Experiments

1.1 Encoding Backbone Replacement

We employ various encoding backbones to extract feature from point clouds. Dual alignment score (DAS), mean intersection over union (mIoU) and part correspondence ratio (PCR) are utilized for evaluation. We conduct the experiment on ShapeNet [1] dataset.

As witnessed by the Table 1, both the PointConv [4] and PointNet++ [3] perform well on the three metrics, while the PointNet [2] fails to achieve competitive results. The hierarchical architecture of PointConv [4] and PointNet++ [3] may explain their success. Specifically, we calculate the average time-per-epoch of the tree encoding backbones on different categories in Table 2. The batch-size is set to 6 by default and we train all models with a single NVIDIA GTX 2080Ti GPU. The time cost of PointConv [4] and PointNet++ [3] is greater than PointNet [2].

1.2 Visualization Results

More visualization results of predicted semantic 3D keypoints by our model are displayed in Figure. 1. Our method ensures great semantic consistency on several categories from ShapeNet[1].

1.3 Downstream Task

As is shown in Figure. 2, our keypoints detector can be applied to *semantic shape alignment*.

Table 1: **Three metrics on ShapeNet [1] with different encoding backbones.**

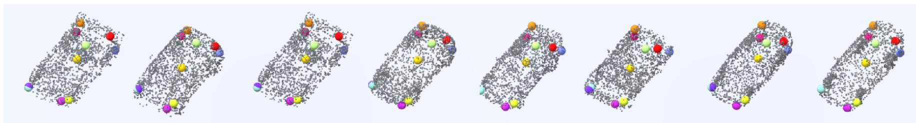
	Airplane			Chair			Table			Mean		
	DAS	mIoU	PCR	DAS	mIoU	PCR	DAS	mIoU	PCR	DAS	mIoU	PCR
PointConv [4]	77.2	**80.0**	**87.3**	76.9	67.4	79.4	78.1	**56.2**	81.9	77.4	**67.8**	82.8
PointNet++ [3]	**81.0**	79.1	81.5	**83.1**	**68.9**	**85.2**	**78.5**	54.1	**85.7**	**80.8**	67.3	**84.1**
PointNet [2]	24.1	19.3	30.4	27.3	19.4	41.2	26.9	27.2	35.3	26.1	21.9	35.6

Table 2: **Average time-per-epoch (min) of different encoding backbones on ShapeNet [1]**

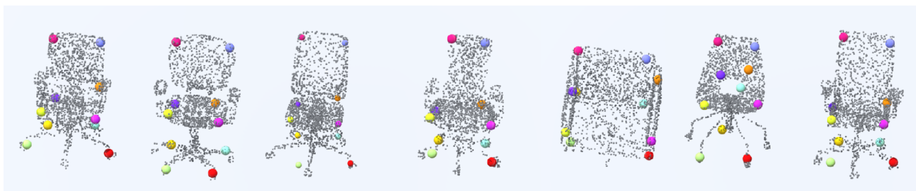
	Airplane	Chair	Table	Mean
PointConv [4]	19.2	26.5	12.1	19.2
PointNet++ [3]	11.7	16.2	7.8	11.9
PointNet [2]	**5.3**	**9.5**	**5.8**	**6.9**

References

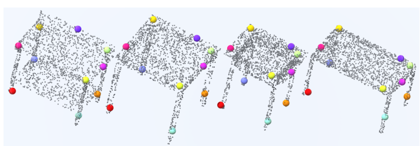
1. Chang, A.X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., et al.: Shapenet: An information-rich 3d model repository. arXiv preprint arXiv:1512.03012 (2015)
2. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 652–660 (2017)
3. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. arXiv preprint arXiv:1706.02413 (2017)
4. Wu, W., Qi, Z., Fuxin, L.: Pointconv: Deep convolutional networks on 3d point clouds. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9621–9630 (2019)



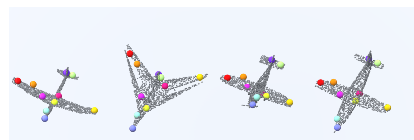
(a) car



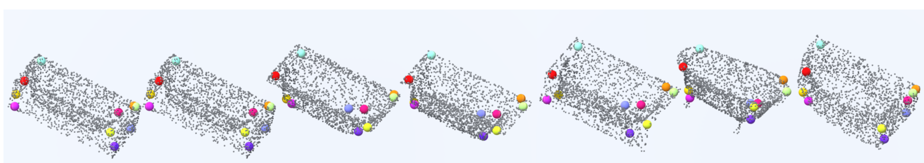
(b) chair



(c) table



(d) airplane



(e) bathtub

Fig. 1: **Visualization Results of More Categories on ShapeNet [1]**

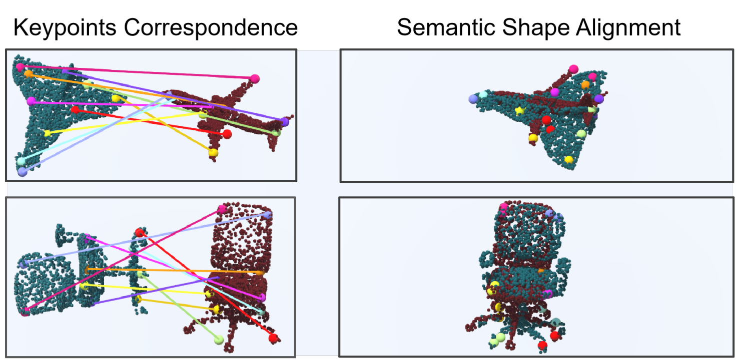


Fig. 2: **Semantic shape alignment with our predicted semantic keypoints.**