6. Supplementary

This document provides additional proof, technical details, quantitative results, and qualitative visualization to the main paper.

6.1. Proof of Proposition

Suppose the normalization process follows Equation 1, and the loss at optimization is denoted by L, and the gradient to embedding f follows Equation 2. Based on them, we show the orthogonality between an embedding and its gradient by computing the their inner product:

$$\begin{aligned}\langle f, \frac{\partial L}{\partial f}\rangle &= \frac{\langle f, \frac{\partial L}{\partial \hat{f}}\rangle - \langle f, \hat{f}\rangle\langle \frac{\partial L}{\partial \hat{f}}, \hat{f}\rangle}{\|f\|_2} \\ &= \frac{\langle f, \frac{\partial L}{\partial \hat{f}}\rangle - \langle \hat{f}, \hat{f}\rangle\langle \frac{\partial L}{\partial \hat{f}}, f\rangle}{\|f\|_2} \\ &= \frac{\langle f, \frac{\partial L}{\partial \hat{f}}\rangle - \langle \frac{\partial L}{\partial \hat{f}}, f\rangle}{\|f\|_2} \\ &= 0\end{aligned} \tag{4}$$

6.2. Description of Dataset

ModelNet40 ModelNet40 dataset contains 12,311 shapes from 40 object categories, and they are split into 9,843 for training and 2,468 for testing. Since the dataset does not provide partial point clouds, we evaluate our proposed method on performing point cloud reconstruction and shape classification. We generate input point clouds by evenly sampling 1024 points from the surface of objects and normalize them within a unit sphere, and no data augmentation is used at training.

ShapeNet ShapeNet part dataset [36] dataset contains 16,881 shapes from 16 object categories with 50 parts. Each point cloud contains 2048 points which are generated by evenly sampling from the surface of objects, and we follow the same set splitting as in [22]. Since the dataset does not provide partial point clouds, we evaluate our proposed method on performing point cloud reconstruction and part segmentation.

MVP MVP dataset [18] contains pairs of partial and complete point clouds from 16 categories. Partial point clouds are generated by back-projecting 2.5D depth images into 3D space and complete point clouds are used as ground truth. In the experiments, we apply the set splitting given by the dataset and no data augmentation is used.

GraspNet Most point cloud completion approaches reported results on synthetic datasets, since collecting a real-world dataset with annotation of complete shapes is expensive. Unfortunately, incomplete measurements from real-world 3D sensors differ from those point clouds synthesized in a simulated environment, and approaches trained on synthetic dataset struggle when tasked to perform completion in real-world scenarios. More recently, the GraspNet [5] dataset was released and it contains the groundtruth complete shapes of objects, which helps evaluate point cloud completion. GraspNet contains 190 cluttered and complex scenes captured by RGBD cameras, bringing 97,280 images in total. For each image, the accurate 6D pose and the dense grasp poses are annotated for each object. There are in total 88 objects with provided CAD 3D models, and we use them to generate complete shape groundtruth with 1024 points.

Model	Seg Acc.	CD
PointNet-Folding	92.06	50.08
PointNet-Folding (H)	92.01	**34.75**
PointNet-PCN	92.06	43.61
PointNet-PCN (H)	92.01	**38.18**
PointNet-TopNet	92.06	37.4
PointNet-TopNet (H)	92.01	**35.50**
DGCNN-Folding	92.50	49.21
DGCNN-Folding (H)	92.39	**33.88**
DGCNN-PCN	92.50	42.42
DGCNN-PCN (H)	92.39	**37.11**
DGCNN-TopNet	92.50	36.80
DGCNN-TopNet (H)	92.38	**35.10**

Table 5. Single-task learning on ShapeNet. Overall point segmentation accuracy (Seg Acc.) is reported for part segmentation, and Chamfer Distance (CD) is reported for point cloud reconstruction, multiplied by 10^4. The first column describes the encoders and decoders used in the model, and "H" indicates using the proposed hyperspherical module.

Model	Cls Acc.	CD
PointNet-Folding	87.33	75.86
PointNet-Folding (H)	87.36	**48.88**
PointNet-PCN	87.33	48.17
PointNet-PCN (H)	87.36	**43.55**
PointNet-TopNet	87.33	55.04
PointNet-TopNet (H)	87.36	**49.65**
DGCNN-Folding	89.22	70.32
DGCNN-Folding (H)	89.47	**45.37**
DGCNN-PCN	89.22	46.54
DGCNN-PCN (H)	89.47	**42.70**
DGCNN-TopNet	89.22	55.87
DGCNN-TopNet (H)	89.47	**48.75**

Table 6. Single-task learning on ModelNet40. Overall classification accuracy (Cls Acc.) is reported for shape classification, and Chamfer distance (CD) is reported for point cloud reconstruction, multiplied by 10^4. The first column describes the encoders and decoders used in the model, and "H" indicates using the proposed hyperspherical module.

6.3. More Experiments

Since ModelNet40 and ShapeNet do not provide pairs of partial and complete point clouds, we report results of point cloud reconstruction along with shape classification on ModelNet40 and part segmentation on ShapeNet.

ModelNet40 Quantitative results on Modelnet40 [36]

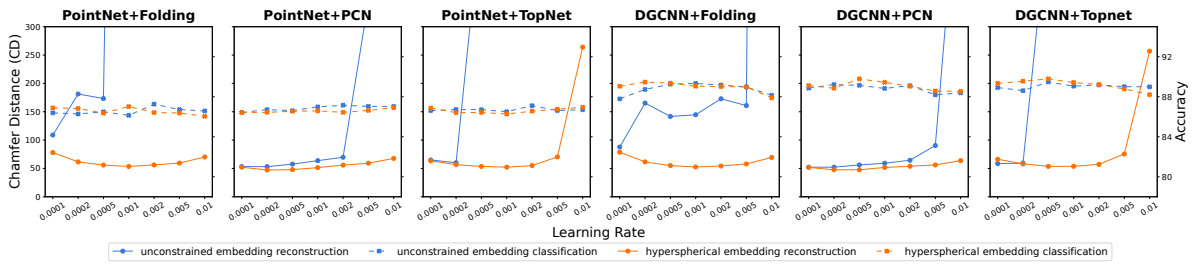

Figure 10. Performance of multi-task learning of point cloud reconstruction and classification on ModelNet40 with different learning rates.

are shown in Table 6. To make a fair comparison, we report results of the combination of two popular encoders, PointNet [21] without T-Net and DGCNN [32], and three different point cloud decoders, Folding [39], PCN [41] and TopNet [27]. The baseline models are compared to their variants added with our proposed hyperspherical modules denoted with (H). As shown by the last column in Table 6, our proposed hyperspherical module helps baseline approaches gain noticeable decrease of Chamfer Distance in all cases. We also test the proposed module in shape classification by removing point cloud decoders and adding fully connected layers. As shown in the second column, the proposed method module leads to slightly better performance of shape classification. Multi-task learning results on shape reconstruction and classification are shown in Figure 10. By comparing the results of models with unconstrained embeddings, the proposed hyperspherical module have little effect on the performance of semantic tasks. However, models with the proposed module have more stable performance when using large learning rates than those with unconstrained embeddings, since the same setting tend to cause the training with unconstrained embeddings unconverged. In terms of converged results, models with our method still outperform their baselines with noticeable improvement.

ShapeNet We report the results on ShapeNet [36] in Table 5. Similar to models constructed for experiments on ModelNet40, we experiment on a combination of different encoders and decoders. As shown by the last column in Table 5, the proposed hyperspherical module improves point cloud reconstruction consistently in all cases. When tasked on part segmentation, the point cloud decoders are removed from the models, and the embeddings concatenated with lifted point-wise features are processed by fully connected layers to predict part category. From the second column in Table 5, part segmentation performance is not affected by the proposed module. Multi-task learning results on shape reconstruction and part segmentation are shown in Figure 11. By comparing the results of models with unconstrained embeddings, the proposed hyperspherical module have little effect on the performance of semantic tasks. However, models with the proposed module have more stable performance when using large learning rates than those with unconstrained embeddings, since the same setting tend to cause the training with unconstrained embeddings unconverged. In terms of converged results, models with our method still outperform their baselines with noticeable improvement.

6.4. More visualization

We show the angular distribution of embeddings by computing the pairwise cosine similarity obtained from the test set in MVP dataset. More visualizations of different classes as described in the plot titles are shown in Figure 12, and the distribution of overall classes is shown in Figure 5

More qualitative results of 3D object detection, pose estimation, and point cloud completion can be found in Figure 13.

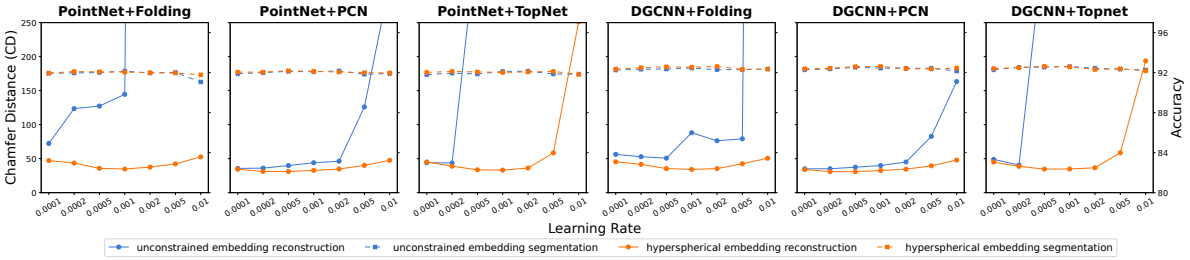

Figure 11. Performance of multi-task learning of point cloud reconstruction and part segmentation on ShapeNet with different learning rates.

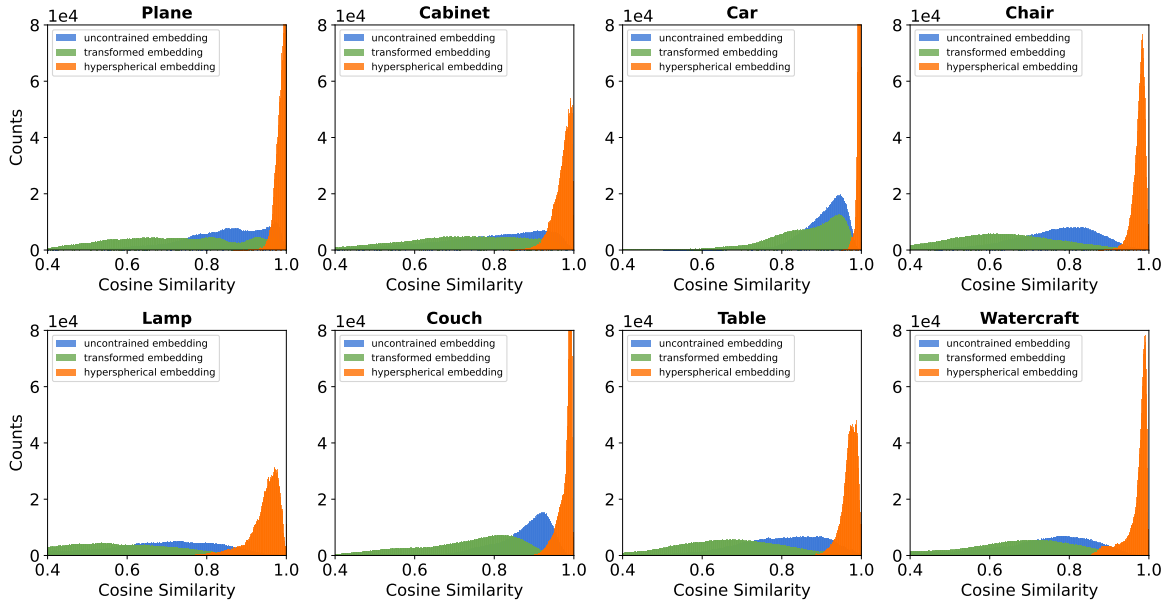

Figure 12. Cosine similarity distribution of embeddings. We compute pairwise cosine distance between embeddings obtained from the test set in MVP dataset. We visualize the distribution of different classes as described in the plot titles. Hyperspherical embeddings have more compact angular distribution.

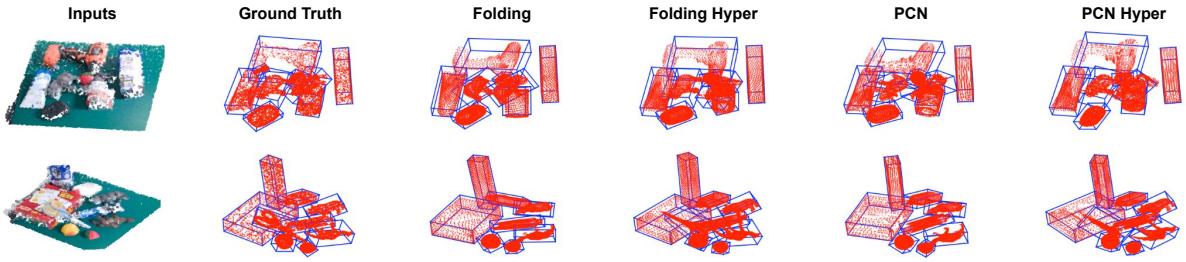

Figure 13. More qualitative 3D detection, pose estimation, and point cloud completion results on GraspNet test set.