

Hyperspherical Embedding for Point Cloud Completion

Junming Zhang, Haomeng Zhang, Ram Vasudevan
 University of Michigan
 {junming, haomeng, ramv}@umich.edu

Matthew Johnson-Roberson
 Carnegie Mellon University
 mkj@andrew.cmu.edu

Abstract

Most real-world 3D measurements from depth sensors are incomplete, and to address this issue the point cloud completion task aims to predict the complete shapes of objects from partial observations. Previous works often adapt an encoder-decoder architecture, where the encoder is trained to extract embeddings that are used as inputs to generate predictions from the decoder. However, the learned embeddings have sparse distribution in the feature space, which leads to worse generalization results during testing. To address these problems, this paper proposes a hyperspherical module, which transforms and normalizes embeddings from the encoder to be on a unit hypersphere. With the proposed module, the magnitude and direction of the output hyperspherical embedding are decoupled and only the directional information is optimized. We theoretically analyze the hyperspherical embedding and show that it enables more stable training with a wider range of learning rates and more compact embedding distributions. Experiment results show consistent improvement of point cloud completion in both single-task and multi-task learning, which demonstrates the effectiveness of the proposed method.

1. Introduction

The continual improvement of 3D sensors has made point clouds much more accessible, which drives the development of algorithms to analyze them. Thanks to deep learning techniques, state of the art algorithms for point cloud analysis have achieved incredible performance [9, 20–22, 25] by effectively learning representations from large 3D datasets [3, 6, 26] and have many applications in robotics, autonomous driving, and 3D modeling. However, point clouds in the real-world are often incomplete and sparse due to many reasons, such as occlusions, low resolution, and the limited view of 3D sensors. So it is critical to have an algorithm that is capable of predicting complete shapes of objects from partial observations.

Given the importance of point cloud completion, it is

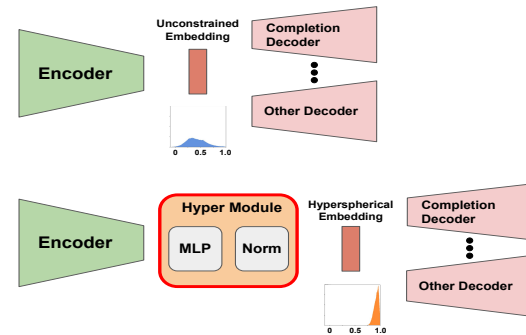


Figure 1. An illustration of the architecture proposed in this paper. The upper subfigure shows the general point cloud analysis structure, where the embedding is directly output from the encoder without constraints. The lower subfigure shows the structure of the model with the proposed hyperspherical module. The figures under the embeddings illustrate the cosine similarity distribution between embeddings, which indicates a more compact embedding distribution achieved by the proposed method and improves point cloud completion.

unsurprising that various methods have been proposed to address this challenge [18, 27, 31, 34, 37, 41]. Most existing methods adapt encoder-decoder structures, in which the encoder takes a partial point cloud as input and outputs an embedding vector, and then it is taken by the decoder which predicts a complete point cloud. The embedding space is designed to be high-dimensional as it must have large enough capacity to contain all information needed for downstream tasks. However, the learned high-dimensional embeddings, as shown in this paper, tend to have a sparse distribution in the embedding space, which increases the possibility that unseen features at testing are not captured by the representation learned at training and leads to worse generalizability of models.

Usually, one real-world application requires predictions from multiple different tasks. For example, to grasp an object in space the robot arm would need the information about the shape, category, and orientation of the target object. In contrast to training all tasks individually from scratch, a more numerically efficient approach would be to train all relevant tasks jointly by sharing parts of networks between different tasks [11, 13, 28]. However, existing point cloud completion methods lack the analysis of accomplish-

ing point cloud completion jointly with other tasks. We show that training existing point cloud completion methods with other semantic tasks together leads to worse performance when compared to learning each individually.

To address the above limitations, this paper proposes a hyperspherical module which outputs hyperspherical embeddings for point cloud completion. The proposed hyperspherical module can be integrated into existing approaches with encoder-decoder structures as shown in Figure 1. Specifically, the hyperspherical module transforms and constrains the output embedding onto the surface of a hypersphere by normalizing the embedding’s magnitude to unit, so only the directional information is kept for later use. We theoretically investigate the effects of hyperspherical embeddings and show that it improves the point cloud completion models by more stable training with large learning rate and more generalizability by learning more compact embedding distributions. We also demonstrate the proposed hyperspherical embedding in multi-task learning, where it helps reconcile the learning conflicts between point cloud completion and other semantic tasks at training. The reported improvements of the existing state-of-the-art approaches on several public datasets illustrate the effectiveness of the proposed method. The main contributions of this paper are summarized as follows:

- We propose a hyperspherical module that outputs hyperspherical embeddings, which improves the performance of point cloud completion.
- We theoretically investigate the effects of hyperspherical embeddings and demonstrate that the point cloud completion benefits from them by stable training and learning a compact embedding distribution.
- We analyze training point cloud completion with other tasks and observe conflicts between them, which can be reconciled by the hyperspherical embedding.

2. Related Work

Point Cloud Completion. Point clouds in most 3D datasets [2, 3, 6, 10, 26] are incomplete and sparse due to reasons, such as occlusions, low resolution, and the limited view of 3D sensors. To address this problem, many works have been proposed to perform point cloud completion, in which people seek to recover the full shape of objects based on the partial observations [27, 31, 34, 37, 41, 43]. Most of them adapt a pipeline of encoding the partial observations into an embedding feature and then decode it to complete point clouds. Since the structures of encoders have been successfully explored by other 3D tasks [17, 21, 22, 32], most completion approaches focus on designing different completion decoders [27, 39, 41]. However, those one-stage decoders are shown to predict point clouds unevenly distributed over the surface of objects and fail to preserve de-

tailed structures in the inputs. Later approaches address these issues by using a coarse-to-fine decoding strategy, in which the decoding process generates several complete shapes at different resolutions, and the partial point clouds are used along with other intermediate decoding results to maximally preserve structures in the inputs [31, 34, 35, 37]. Instead of designing a structure of model, the method developed in the paper focuses on effectively learning representation, which is a more universal problem since the developed method can be integrated into existing structures. We show the improvement of existing point cloud completion by applying the proposed method on both one-stage and coarse-to-fine decoders.

Embedding in Point Cloud. Most methods for point cloud analysis use a max pooling layer as the last layer to address the permutation issue contained in the unordered point set [17, 21, 22, 32]. Embeddings output from those encoders are learned from large datasets using an end-to-end training. However, the embeddings learned in this way are not imposed with constrain and tend to have a sparse distribution, which increases the possibility of testing inputs accidentally falling into the unseen regions during training due to gaps or holes in the embedding space [15, 29]. The expressiveness of the embeddings can be improved by other techniques, such as adversarial training [1, 33], probabilistic modeling [18, 43]. In multi-task learning, point cloud embeddings are shared by different decoders to improve efficiency of models [5, 20, 25]. Unfortunately, those decoders learned by different task losses may require different embedding distributions, which may lead to optimization conflicts during training, and an unconstrained embedding space makes such issues occur more frequently [24, 40]. In this paper, we propose to use a simple but effective hyperspherical embedding for point cloud completion and demonstrate the success of the proposed method in both single-task learning and multi-task learning.

Hyperspherical Embedding. Effectively obtaining an embedding that is needed for the corresponding learning task from raw input data is very important. Compared to the embeddings learned without constraints, some works proposed to use the hyperspherical embedding by normalizing it onto a unit hypersphere and have shown success in many fields, such as representation learning, metric learning, and face verification [14–16, 19, 29, 30, 42, 44]. All those works suggested unit hypersphere is a nice feature space, and some of them tried to explain the improvement by showing the empirical observation of stabler training and better distribution of embeddings when applying hyperspherical embeddings [14, 30, 44]. But all of them lack a comprehensive analysis of why the hyperspherical embedding leads to a better distribution of embeddings and the effects in multi-task learning. Thus, this paper delves into the hyperspherical embedding and we apply it to point cloud completion.

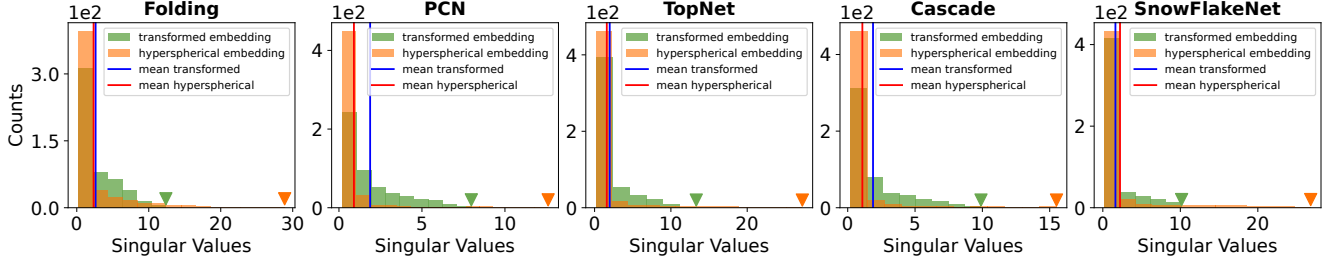


Figure 2. An illustration of distributions of singular values. We compute singular values of weights in the layer right before the embedding, obtained from the point cloud completion models on MVP dataset with different architectures described in the plot title. The mean of singular values are denoted by vertical lines, and the inverted triangle denotes the largest singular value. This figure shows that the hyperspherical embedding leads to learning more poorly conditioned weights, while the means of singular values with different embeddings are similar.

We uncover that the hyperspherical embedding drives the model to learn complexity reduction of high-dimensional data by poorly conditioned weights and thus leads to a compact embedding distribution.

3. Hyperspherical Embedding for Learning Point Cloud

This section describes the proposed hyperspherical module and investigates effects of the hyperspherical embedding.

3.1. Proposed Hyperspherical Module

To address the sparse embedding distribution, we propose a new hyperspherical module and its structure is shown in Figure 1. The proposed module contains two layers, a multi-layer perceptron (MLP) layer and a normalization layer. The outputs from the encoder are first transformed by the MLP layer and then the normalization layer constrains the features onto the surface of hypersphere by l_2 normalization,

$$\hat{f} = \frac{f}{\|f\|_2} \quad (1)$$

where the $\|f\|_2 = \sqrt{\sum_i f_i^2}$, and \hat{f} denotes the l_2 normalized embedding of f .

3.2. Effects of Hyperspherical Embedding

In this part, we investigate the effects of optimizing the l_2 normalized embedding and we draw several conclusions: 1) the gradient of the embedding before normalization is orthogonal to itself; 2) the magnitude of the embedding before normalization increases at each update during training; 3) the increased magnitude enables more stable training with a wider range of learning rates compared to unconstrained embeddings; 4) the embedding distribution is compact in the angular space, resulting better point cloud completion performance in both single-task and multi-task learning.

Proposition 1. *During optimizing l_2 normalized embedding \hat{f} , the computed gradient to the embedding before nor-*

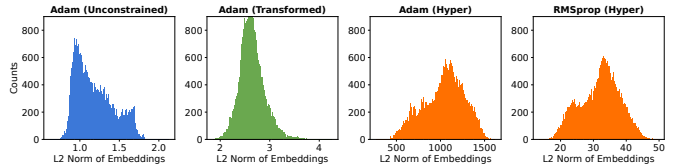


Figure 3. Norm distribution of embeddings. The embeddings are derived from the MVP test set on point cloud completion, obtained by using different embeddings or different optimizers, as described in the plot titles. Unconstrained embeddings (Unconstrained) are from models without the proposed module. Transformed embeddings (Transformed) are from models with the proposed module but removing the normalization layer. Hyperspherical embeddings (Hyper) are from models using the proposed module, and the norm shown in this figure is computed before the normalization.

malization denoted by f is orthogonal to itself, $\langle f, \frac{\partial L}{\partial f} \rangle = 0$.

Proof. Suppose the normalization process follows Equation 1 and the loss at optimization is denoted by L . The gradient to embedding f can be formulated as:

$$\frac{\partial L}{\partial f} = \frac{\frac{\partial L}{\partial \hat{f}} - \hat{f} \langle \frac{\partial L}{\partial \hat{f}}, \hat{f} \rangle}{\|f\|_2} \quad (2)$$

Based on it, we can show the orthogonality by computing the inner product between the embedding and its gradient. More details can be found in the supplementary, and similar conclusions are also reported in [29, 42, 44].

Proposition 2. *For standard stochastic gradient descent (SGD), each update of embedding f will monotonically increase its norm, $\|f\|_2$.*

Proof. The orthogonality between the embedding and its gradient from Proposition 1 indicates that applying gradient at each update increases the norm of an embedding, which is validated by showing the distribution of embedding’s norm in Figure 3. Similar observations are reported in [42, 44]. This property is based on the vanilla gradient descent algorithm and does not strictly hold for optimizers that use momentum or separate learning rates for individual parameters. However, we still find the same effect empirically

hold for other SGD-based optimizers, such as Adam [12] and RMSprop [8], as illustrated in Figure 3.

Proposition 3. *The magnitude of the gradient is inversely proportional to the norm of the embedding, $\frac{\partial L}{\partial f} \propto \frac{1}{\|f\|_2}$.*

Proof. Considering the norm $\|f\|_2$ in the denominator in Equation 2, the magnitude of the gradient is inversely proportional to the norm of an embedding. This conclusion is similar to the one in [44]. However, we further note that this effect enables optimizing neural networks with a wider range of learning rates. In particular, the norm of embedding trained with a large learning rate quickly increases shown by Proposition 2 until an appropriate effective learning rate is reached, while the same setting puts the model at risk of overshooting the minima. So it may not be able to converge when using unconstrained embeddings. Normalizing the weights of neural networks has similar effects reported by [23], while in this paper we focus on normalizing the layer of embedding and keep weights untouched, which makes the implementation easier. Figure 9 shows the comparison results with hyperspherical embedding and unconstrained embeddings trained using different learning rates, and more results can be found in Figure 10 and Figure 11 in the supplementary. All of them show that the hyperspherical embedding leads to more stable training with a wider range of learning rates and better performance of point cloud completion.

Proposition 4. *Considering a vector is transformed by a matrix, i.e., $f = Wx$. During optimization, the increased norm of f requires a poorly conditioned matrix W .*

Proof. The matrix W can be decomposed by singular value decomposition (SVD):

$$W = U\Sigma V^T \quad (3)$$

where U and V are orthonormal matrices and Σ is a diagonal matrix. Based on Equation 3, the norm of the transformed output f is only related to singular values contained in Σ , since orthonormal matrices U and V do not modify the magnitude of inputs. In our case, the norm of f increases during training from Proposition 2, so the $\|\Sigma\|_2$ will increase during optimization. However, increasing all singular values in $\|\Sigma\|$ makes the weight large, which adversely affects the performance of models by overfitting the training data [7]. Empirically we do not observe this issue and find that the mean of the singular values trained with normalized embeddings stays similarly to the one without using it, shown in Figure 2. Therefore, the increase of certain singular values in Σ will inevitably lead to decrease of other singular values. In particular, the large singular values get increased and small singular values get decreased to increase the $\|\Sigma\|_2$, which causes the weight W to be poorly conditioned as it is illustrated in Figure 2.

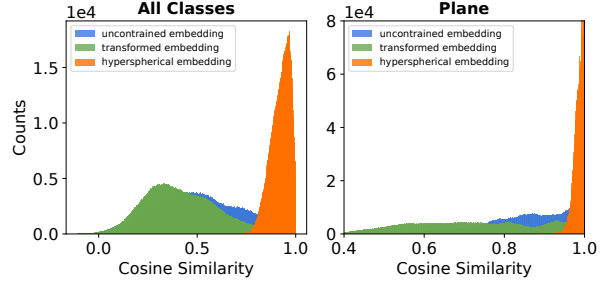


Figure 4. Cosine similarity distribution of embeddings. We compute pairwise cosine distance between embeddings obtained from the test set in MVP dataset. We visualize the distribution in either one class or overall classes as described in the plot titles. It shows that hyperspherical embeddings have more compact angular distribution. More visualizations of different classes can be found in Figure 12 in the supplementary.

One effect of poorly conditioned weights is that it will reduce the complexity in the input high-dimensional data by keeping information on principle axes while ignoring information on other axes. By doing this, transformed vectors by those weights tend to point in a similar direction. Moreover, the hyperspherical embedding follows a l_2 normalization, so the discrepancy of embeddings in magnitude is further removed. We compute pairwise cosine distance of embeddings trained with or without l_2 normalization in the test set and visualize the distribution in Figure 4. It shows that hyperspherical embeddings have a more compact angular distribution, while the unconstrained embedding distribution tends to be sparse. This compact embedding distribution helps the model generalize well on unseen data at testing and increases the generalizability of models.

The learned compact embedding distribution also helps reconcile the learning conflicts in multi-task learning. The resulting compact embedding distribution forces different tasks to learn within the shared space, while the unconstrained embedding space provides tasks the freedom to land on optimal embedding distributions with discrepancy. We use the gradient cosine similarity proposed by [40] to measure the conflicts between different tasks and visualize the training process in Figure 5. The figure shows that the hyperspherical embeddings lead to smaller gradient conflicts at training, as larger cosine similarity indicates smaller gradient conflicts. More discussion about the effects on multi-task learning can be found in Sec 4.3.

4. Experiments

Experiments are divided into three parts. We first report results on different datasets to evaluate the effectiveness of the proposed methods in Sec 4.1. Second, we conduct a detailed ablation study to validate the design of our hyperspherical module in Sec 4.2. Finally, we analyze and visualize the effects of hyperspherical embeddings introduced in Sec 4.3. All experiments are conducted on NVIDIA Tesla V100 GPUs, and we use default training settings for all

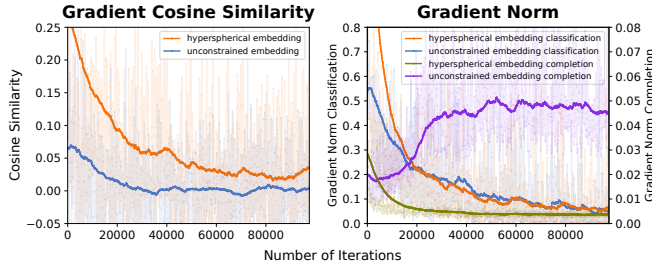


Figure 5. An illustration of gradient conflicts between tasks in multi-task learning during training. We visualize the gradient cosine similarity and gradient magnitude as indicated by the titles of subfigures, obtained by training point cloud completion and shape classification on MVP dataset. Hyperspherical embeddings lead to smaller gradient conflicts between tasks in multi-task learning.

baseline methods. In multi-task learning experiments, we use fully connected layers and cross entropy loss to train semantic tasks in multi-task learning, while decoders with more sophisticated structures are used to generate complete point clouds trained using Chamfer Distance [4].

4.1. Experiments on Different Datasets

Note, a longer description of the considered datasets and experimental results on **ModelNet40** and **ShapeNet** can be found in the supplementary document.

MVP We evaluate point cloud completion on MVP [18] and compare the results of generating complete shape with 2048 points, and qualitative results can be found in Figure 6. Compared to the predictions without our hyperspherical module, the predicted point clouds tend to be a bit more blurry and contain more noise points. We assume it is due to the test embeddings falling into regions not close to features captured at training. More quantitative results are shown in Table 1. We compare the baseline models by adding the hyperspherical module after the encoder and denote them by (H). One exception is SnowFlakeNet, in which the decoding process also consists of encoding modules, so we add hyperspherical modules in its decoder as well. To achieve fair comparison, we train all methods with the best training setting and report their results. Table 1 shows that using our proposed module brings consistent improvements of all existing completion approaches by 3 ~ 9% decrease of average class Chamfer Distance. Multi-task learning on MVP dataset will be discussed in Sec 4.3.

GraspNet To demonstrate on real-world scenarios, in this part we aim to detect objects in 3D space along with predicting their complete shapes on GraspNet dataset [5]. We modified the structures of VoteNet [20], which was designed to detect 3D objects from point clouds. The input point clouds are converted from the depth image captured by RGBD sensors. After extracting the embedding from each proposal, three branches of decoders are followed to generate predictions of 6DoF 3D bounding boxes, se-

mantics and objectness scores, and complete point clouds of objects, respectively. In the evaluation, object detection is measured by mean average precision (mAP), poses of objects are measured by the symmetry metric proposed in [38], and the point cloud completion is measured by Chamfer Distance. We evaluate the effectiveness of the proposed hyperspherical module in this multi-task learning scenario, and the results are shown in Table 2. Since the model structures are the same except for decoders, we distinguish the models by the decoders in the first column. Two-stage completion decoders, such as Cascade and SnowFlakeNet, are demonstrated to have better performance on synthetic dataset, but they refine on perfect partial point clouds located on object’s surfaces, which are inaccessible in this case. As it is shown in Table 2, hyperspherical embeddings help all three metrics with noticeable improvement comparing to unconstrained embeddings. Qualitative results can be found in Figure 7 and Figure 13 in the supplementary.

4.2. Ablation Study

Results of the ablation study are shown in Table 3 and numbers in the table are the average class Chamfer Distance of completion models with folding decoder reported on the MVP test set. The accuracy of the model is improved after applying the proposed hyperspherical module, reducing the Chamfer Distance from 10.39 to 9.47, shown by row one and row four. The structural design of the hyper module is validated by results in the first four rows, and they also indicates that the major improvement is brought by the normalization layer, from 10.12 to 9.47. From row three and row five, changing MLP layers, activation layer or batch normalization layer does not improve the performance of point cloud completion compared to the design used in this paper. To validate the choice of l_2 manifold, we also report the results of applying l_1 (9.99) and l_3 (9.69) normalization. All of them are shown to boost the performance, but the l_2 manifold used in this paper has the best results.

4.3. Experiments on the Effects of Hyperspherical Embedding

In this section, we study the effects of hyperspherical embedding and provide empirical results and visualizations that align with the conclusions claimed in Sec 3.

Increased Magnitude of Embedding From Proposition 2, the magnitude of an embedding gets increased during the optimization, and we visualize the distribution of embedding magnitude from test set on MVP dataset in Figure 3. When optimizing our proposed hyperspherical embedding denoted by (Hyper), the scale of embedding’s magnitude before normalization is significantly larger than those unnormalized denoted by (Unconstrained and Transformed). Furthermore, we observe in two leftmost subfigures that the range of embeddings’ magnitude changes little

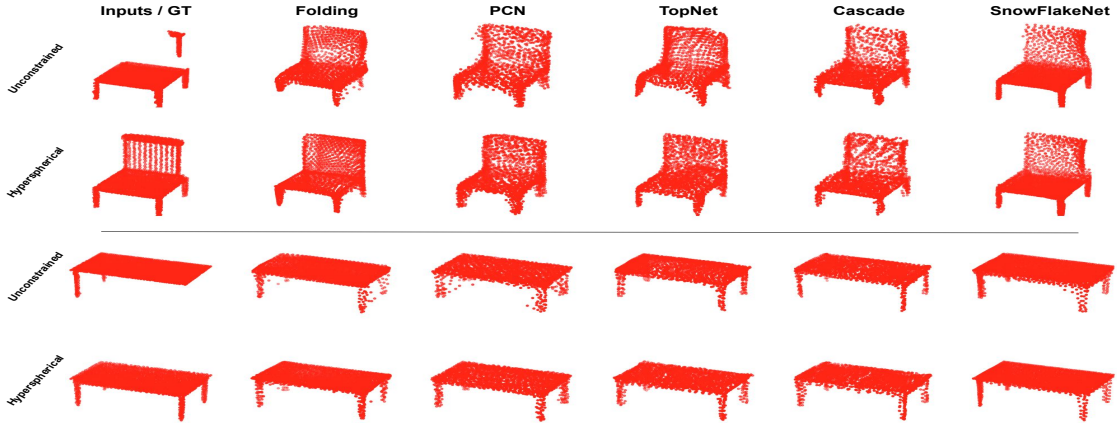


Figure 6. Qualitative results of various state-of-the-arts point cloud completion approaches on MVP test set.

Model	plane	cabinet	car	chair	lamp	sofa	table	wcraft	bed	bench	bshelf	bus	guitar	mbike	pistol	sboard	average
Folding [39]	4.71	9.08	6.81	15.22	23.12	10.28	14.32	9.90	22.02	10.28	14.48	5.24	2.02	6.91	7.21	4.59	10.39
Folding (H)	4.48	8.82	6.68	13.79	21.44	9.66	12.98	8.57	18.93	8.96	13.44	4.95	2.03	6.43	6.22	4.20	9.47
PCN [41]	4.23	9.35	6.73	13.56	20.94	10.51	14.20	9.81	21.32	9.98	15.08	5.45	1.90	6.23	6.23	5.03	10.03
PCN (H)	4.24	9.14	6.49	13.04	22.47	10.04	12.99	8.75	18.95	9.33	13.93	5.06	1.84	6.00	5.92	4.15	9.52
TopNet [27]	4.63	9.23	6.79	14.31	19.50	10.48	14.30	9.65	20.54	10.12	15.53	5.36	2.09	6.77	7.74	4.94	10.12
TopNet (H)	4.07	9.13	6.75	13.08	19.45	10.03	12.85	8.89	19.50	9.63	14.33	5.23	2.03	6.66	6.42	3.92	9.50
Cascade [31]	2.66	8.69	6.02	10.22	13.07	8.76	9.90	6.67	16.44	7.56	11.00	4.97	1.98	4.58	4.54	2.78	7.49
Cascade (H)	2.61	8.52	5.97	9.52	12.03	8.71	9.83	6.46	15.78	7.17	11.15	4.90	1.88	4.50	4.24	2.76	7.25
SnowFlakeNet [37]	1.94	7.61	5.61	6.77	6.82	7.09	7.21	4.65	10.98	4.76	7.54	4.16	1.14	3.78	3.15	2.67	5.37
SnowFlakeNet (H)	1.89	7.26	5.36	6.50	7.59	6.72	6.63	4.67	10.39	4.39	7.37	4.03	0.95	3.60	3.15	2.84	5.21

Table 1. The performance of different completion approaches trained on MVP dataset. The Chamfer Distance is reported, multiplied by 10^4 , on the provided test set. “H” indicates using the proposed hyperspherical module.

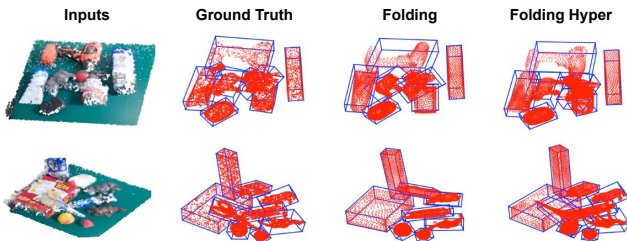


Figure 7. Qualitative 3D detection, pose estimation, and point cloud completion results on GraspNet test set.

Model	mAP (0.25)	CD	Pose Acc.
Folding	70.50	0.18	52.42
Folding (H)	71.21	0.14	54.01
PCN	69.11	0.21	50.33
PCN (H)	70.93	0.15	52.39

Table 2. Performance on GraspNet test set. Average precision with 3D IoU threshold 0.25 (mAP 0.25) is reported for object detection, Chamfer Distance (CD) is reported for point cloud completion, multiplied by 10^4 , and pose accuracy (Pose Acc.) is reported for 6D pose estimation. The first column indicates the structure of decoder used in the model, and “H” indicates using the proposed hyperspherical module.

after adding a MLP layer, which validates that l_2 normalization is the key to increase embedding magnitude. Even

Base	1-layer MLP	2-layer MLP	ReLU&BN	l_1	l_2	l_3	CD
✓							10.39
✓	✓						10.12
✓	✓					✓	9.57
✓	✓		✓			✓	9.47
✓	✓					✓	10.06
✓	✓	✓					10.02
✓	✓					✓	9.69
✓	✓				✓		9.99

Table 3. Results of ablation study. The reported metric is Chamfer Distance, multiplied by 10^4 , of point cloud completion on MVP test set. “Base” indicates not using the proposed module, or unconstrained embeddings; “1-layer MLP” indicates using one MLP layer in the hyperspherical module; “2-layer MLP” indicates using two MLP layers in the hyperspherical module; “ReLU&BN” indicates using ReLU activation and Batch Normalization after each MLP layer; “L*” indicates the type of normalization in the hyperspherical module.

though the effect is based on vanilla gradient descent, we still find it empirically holds for modern optimizers, such as Adam and RMSProp shown in the two rightmost subfigures of Figure 3.

Enabling Wider Range of Learning Rates When optimizing a normalized embedding, the gradient at each update can be computed by following the Equation 2, and it indicates that the magnitude of gradient is inversely proportional to the norm of the embedding shown in Proposition 3. One benefit of this finding is that the increased norms of embeddings help neural networks gain robustness to varying

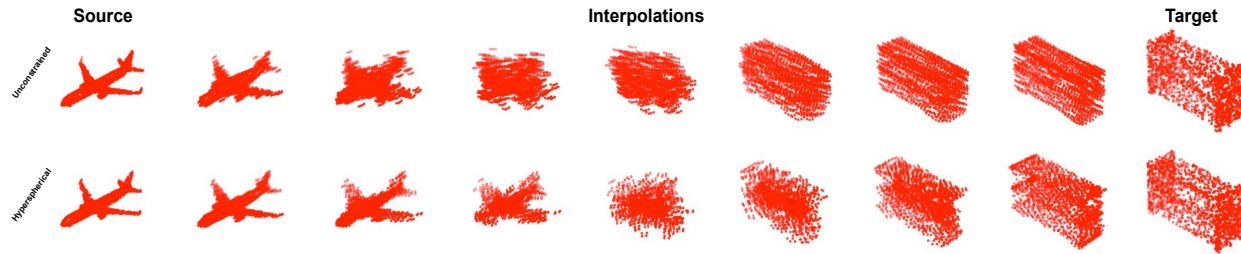


Figure 8. Illustration of point cloud interpolation in the embedding space. The generated point clouds with hyperspherical embeddings have more clear clues from source or target shapes than those with unconstrained embeddings.

values of learning rate. Figure 9 (Figure 10 and Figure 11 are in the supplementary) show the results of models on different tasks using different learning rates. Compared to the point cloud completion results with unconstrained embedding, the models using hyperspherical embeddings perform more stably under a wider range of learning rates with better performance in particular for large learning rates. If the learning rate is large, the norm of embeddings before normalization quickly increases and makes the effective learning rate small enough to stabilize training. When trained with small learning rates, models perform similarly using either hyperspherical embeddings or unconstrained embeddings. This can be explained by the observation that slightly increased or similar norms of embeddings using the hyperspherical embeddings compared to unconstrained embeddings, and a small learning rate would slow down the speed of increase of embeddings' norm.

Compact Embedding Distribution Based on Proposition 2 and Proposition 4, optimizing hyperspherical embedding drives the model to learn complexity reduction by ill-conditioned weights. We empirically verify this by visualizing the singular value distribution of weights trained with point cloud completion task on MVP dataset, and the results are shown in Figure 2. We observe that singular values of weights learned with the hyperspherical embedding have a larger value span than those without normalized embedding, while the majority of them are located near zero.

The resulting poorly conditioned weights make a compact embedding distribution in the angular space. Figure 4 illustrates the angular distribution of embeddings by computing cosine similarity of pairwise embeddings from the MVP test set. Adding a MLP layer to transform the unconstrained embedding does not change the span of angular distribution significantly, while the hyperspherical embeddings have much narrower angular span and are distributed more compactly on both single class and overall classes. Compared to a compact embedding distribution, one disadvantage of the sparse embedding distribution is that it increases the possibility of unseen features falling far away from the seen features at training, which inevitably worsens the generalization of models at testing. To visually demonstrate the degree of sparsity in the embedding space, we use trained

models that perform point cloud reconstruction on ModelNet40 to interpolate embeddings on the embedding space and generate point clouds, and the results are shown in Figure 8. Compared to results from unconstrained embedding space, the point clouds in hyperspherical embeddings space have more clear clues from source or target shapes, because the interpolated hyperspherical embeddings are closer to features captured at training. Thus, it helps with generating more reasonable shapes of objects at testing.

Improvement on Multi-task Learning In addition to single-task learning, the proposed hyperspherical module also improves the performance of point cloud completion in multi-task learning with other semantic tasks. To verify this, we report the results of jointly training point cloud completion and shape classification on MVP dataset in Figure 9. More results of multi-task learning on ModelNet40 and ShapeNet datasets can be found in the supplementary. By comparing the results of models with unconstrained embeddings, the proposed hyperspherical module has little benefit on the performance of semantic tasks. However, the proposed module improves the point cloud completion task with more stable performance when using large learning rates than those with unconstrained embeddings, since the same settings tend to cause the unconstrained models unconverged. In terms of the converged results, models with our method still outperform their baselines.

To make a fair comparison, we also report results of other approaches developed for multi-task learning using different types of embeddings trained on MVP dataset as shown in Table 4. The second to fifth columns show results of models with different training strategies indicated by the column title. Unsurprisingly, models with the proposed hyperspherical modules outperform the baselines on point cloud completion in both single-task and multi-task learnings under all settings. By comparing the completion results in multi-task learning, manually searching the optimal weights between completion task and classification task takes much time, but it achieves the best completion results with little affection on classification performance. Furthermore, the rightmost column (S. vs. M.) presents the percent of changes comparing the best completion results in multi-task learning to those in single-task learning. It shows im-

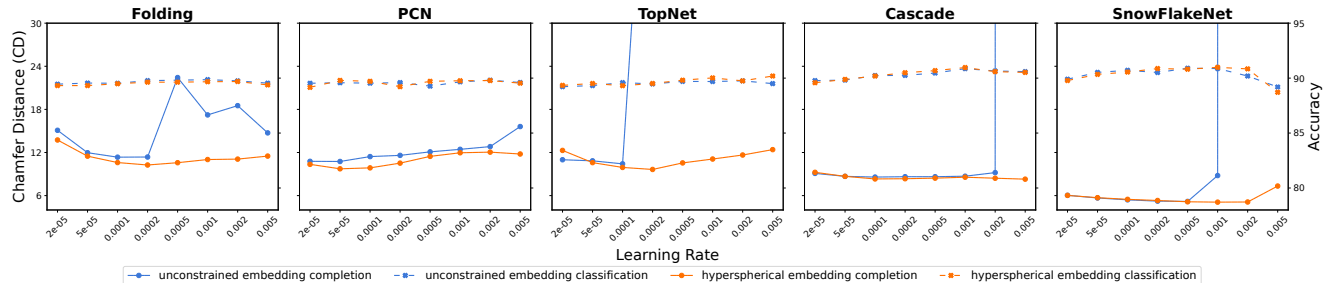


Figure 9. Performance of multi-task learning of point cloud completion and classification on MVP dataset with different learning rates.

Model	Single Task		Equal Weights		PCGrad [40]		Uncert. [11]		Weight Search		S. vs. M.
	Acc	CD	Acc	CD	Acc	CD	Acc	CD	Acc	CD	
Folding	89.68	10.39	89.77	11.37	89.67	11.21	89.81	11.22	89.12	10.45	-0.58
Folding (H)	89.91	9.47	89.63	10.26	89.77	10.13	89.51	10.07	89.43	9.40	0.74
PCN	89.62	10.03	89.58	10.75	89.41	10.75	89.33	10.77	89.26	10.37	-3.39
PCN (H)	89.55	9.52	89.79	9.73	89.56	9.58	89.69	9.58	89.78	9.45	0.74
TopNet	89.49	10.12	89.59	10.42	89.84	10.33	89.58	10.52	89.43	10.24	-1.19
TopNet (H)	89.55	9.50	89.51	9.64	89.80	9.59	89.90	9.48	89.74	8.79	7.47
Cascade	90.91	7.49	90.23	8.58	90.33	8.53	90.27	8.51	90.18	7.50	-0.13
Cascade (H)	90.51	7.25	90.19	8.32	90.02	8.18	90.32	8.32	90.48	7.22	0.41
SnowFlakeNet	90.93	5.37	90.90	5.19	90.99	5.29	90.18	5.27	90.75	5.04	6.15
SnowFlakeNet (H)	90.91	5.21	90.98	5.09	90.95	5.21	90.13	5.11	90.82	5.02	3.65

Table 4. Comparison results of models using different multi-task training strategies on MVP dataset. Results of shape classification (Acc) and point cloud completion (CD) are reported, multiplied by 10^4 . “S. vs. M.” shows the percentage of performance change comparing best completion results in multi-task learning to those in single-task learning.

provement of completion performance trained in multi-task learning when using our method, while the models with unconstrained embeddings struggle in the degradation of completion performance when they are trained with shape classification. One exception is SnowFlakeNet with unconstrained embeddings, which gets improved on completion when trained with classification. We argue that it is due to multi-task learning helps with reducing the overfitting issue observed when training SnowFlakeNet in single-task learning, but our proposed method (5.02) still outperforms its baseline (5.04).

This paper focuses on point cloud completion, but the empirical results show a neutral effect on classification. Table 4 shows slight improvement (Folding and TopNet) and regression (PCN, Cascade and SnowFlakeNet) in single-task classification. Hyperspherical embedding leads to a compact embedding space, which means that the inter-class space is reduced compared to regular embedding space. We suppose that it raises challenges when learning a classifier and explains that hyperspherical embedding is not commonly adopted in single-task learning of classification. However, our method helps classification to be robust in multitasking while we observe noticeable degradation of classification without our method in weight search column.

To delve into the effects of hyperspherical embedding on multi-task learning, we visualize the conflicts between tasks when training them jointly in Figure 5. Specifically, the measure we visualize is the cosine similarity of gradients on the shared encoders with respect to different task

losses, where negative values indicate conflicting gradients, as proposed by [40]. The visualization shows that the gradient cosine similarity with hyperspherical embeddings are almost positive, while those with unconstrained embeddings tend to be more negative, which explains the improvement of point cloud completion in multi-task learning with classification observed in Table 4. Moreover, we find that the classification task dominates the training since its magnitude of gradient is significantly larger than gradients with respect to completion loss, as shown in the right subfigure in Figure 5. This aligns well with the results of weight search experiments, where smaller classification weight or larger completion weight lead to better completion performance.

5. Conclusions

This paper proposes a general module for point cloud completion. In particular, our hyperspherical module transforms and normalizes the output from the encoder onto the surface of a hypersphere before it is processed by the following decoders. We study the effects of the proposed hyperspherical embeddings in both theoretical and experimental ways. Extensive experiments are performed on synthetic and real-world datasets, and the achieved state-of-the-art results in both single-task and multi-task learnings demonstrate the effectiveness of the proposed method.

Acknowledgements. This work is supported by the Ford Motor Company via award N022977 and the National Science Foundation under award 1751093.

References

- [1] Yingjie Cai, Kwan-Yee Lin, Chao Zhang, Qiang Wang, Xiaogang Wang, and Hongsheng Li. Learning a structured latent space for unsupervised point cloud completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5543–5553, 2022. 2
- [2] Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, and James Hays. ArgoVerse: 3d tracking and forecasting with rich maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 2
- [3] Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*, 2017. 1, 2
- [4] Haoqiang Fan, Hao Su, and Leonidas J Guibas. A point set generation network for 3d object reconstruction from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 605–613, 2017. 5
- [5] Hao-Shu Fang, Chenxi Wang, Minghao Gou, and Cewu Lu. Graspnet-1billion: A large-scale benchmark for general object grasping. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11444–11453, 2020. 2, 5, 11
- [6] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361. IEEE, 2012. 1, 2
- [7] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016. 4
- [8] Geoffrey Hinton, Nitish Srivastava, and Kevin Swersky. Neural networks for machine learning. *Coursera, video lectures*, 264(1):2146–2153, 2012. 4
- [9] Qingyong Hu, Bo Yang, Linhai Xie, Stefano Rosa, Yulan Guo, Zhihua Wang, Niki Trigoni, and Andrew Markham. Learning semantic segmentation of large-scale point clouds with random sampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 1
- [10] Xinyu Huang, Peng Wang, Xinjing Cheng, Dingfu Zhou, Qichuan Geng, and Ruigang Yang. The apolloscape open dataset for autonomous driving and its application. *IEEE transactions on pattern analysis and machine intelligence*, 42(10):2702–2719, 2019. 2
- [11] Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7482–7491, 2018. 1, 8
- [12] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 4
- [13] Shikun Liu, Edward Johns, and Andrew J Davison. End-to-end multi-task learning with attention. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1871–1880, 2019. 1
- [14] Weiyang Liu, Rongmei Lin, Zhen Liu, Li Xiong, Bernhard Schölkopf, and Adrian Weller. Learning with hyperspherical uniformity. In *International Conference On Artificial Intelligence and Statistics*, pages 1180–1188. PMLR, 2021. 2
- [15] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 212–220, 2017. 2
- [16] Weiyang Liu, Yan-Ming Zhang, Xingguo Li, Zhiding Yu, Bo Dai, Tuo Zhao, and Le Song. Deep hyperspherical learning. *Advances in neural information processing systems*, 30, 2017. 2
- [17] Yongcheng Liu, Bin Fan, Shiming Xiang, and Chunhong Pan. Relation-shape convolutional neural network for point cloud analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8895–8904, 2019. 2
- [18] Liang Pan, Xinyi Chen, Zhongang Cai, Junzhe Zhang, Haiyu Zhao, Shuai Yi, and Ziwei Liu. Variational relational point completion network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8524–8533, 2021. 1, 2, 5, 11
- [19] Shi Pu, Kaili Zhao, and Mao Zheng. Alignment-uniformity aware representation learning for zero-shot video classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19968–19977, 2022. 2
- [20] Charles R Qi, Or Litany, Kaiming He, and Leonidas J Guibas. Deep hough voting for 3d object detection in point clouds. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9277–9286, 2019. 1, 2, 5
- [21] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 652–660, 2017. 1, 2, 12
- [22] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in neural information processing systems*, pages 5099–5108, 2017. 1, 2, 11
- [23] Tim Salimans and Durk P Kingma. Weight normalization: A simple reparameterization to accelerate training of deep neural networks. *Advances in neural information processing systems*, 29, 2016. 4
- [24] Ozan Sener and Vladlen Koltun. Multi-task learning as multi-objective optimization. *Advances in neural information processing systems*, 31, 2018. 2
- [25] Shaoshuai Shi, Xiaogang Wang, and Hongsheng Li. Pointcnn: 3d object proposal generation and detection from point cloud. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–779, 2019. 1, 2
- [26] Pei Sun, Henrik Kretschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2446–2454, 2020. 1, 2

- [27] Lyne P Tchappmi, Vineet Kosaraju, Hamid Rezatofighi, Ian Reid, and Silvio Savarese. Topnet: Structural point cloud decoder. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 383–392, 2019. [1](#), [2](#), [6](#), [12](#)
- [28] Jonas Uhrig, Marius Cordts, Uwe Franke, and Thomas Brox. Pixel-level encoding and depth layering for instance-level semantic labeling. In *German conference on pattern recognition*, pages 14–25. Springer, 2016. [1](#)
- [29] Feng Wang, Xiang Xiang, Jian Cheng, and Alan Loddon Yuille. Normface: L2 hypersphere embedding for face verification. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 1041–1049, 2017. [2](#), [3](#)
- [30] Tongzhou Wang and Phillip Isola. Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In *International Conference on Machine Learning*, pages 9929–9939. PMLR, 2020. [2](#)
- [31] Xiaogang Wang, Marcelo H Ang Jr, and Gim Hee Lee. Cascaded refinement network for point cloud completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 790–799, 2020. [1](#), [2](#), [6](#)
- [32] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (TOG)*, 38(5):146, 2019. [2](#), [12](#)
- [33] Cheng Wen, Baosheng Yu, and Dacheng Tao. Learning progressive point embeddings for 3d point cloud generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10266–10275, 2021. [2](#)
- [34] Xin Wen, Tianyang Li, Zhizhong Han, and Yu-Shen Liu. Point cloud completion by skip-attention network with hierarchical folding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1939–1948, 2020. [1](#), [2](#)
- [35] Xin Wen, Peng Xiang, Zhizhong Han, Yan-Pei Cao, Pengfei Wan, Wen Zheng, and Yu-Shen Liu. Pmp-net: Point cloud completion by learning multi-step point moving paths. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7443–7452, 2021. [2](#)
- [36] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1912–1920, 2015. [11](#), [12](#)
- [37] Peng Xiang, Xin Wen, Yu-Shen Liu, Yan-Pei Cao, Pengfei Wan, Wen Zheng, and Zhizhong Han. Snowflakenet: Point cloud completion by snowflake point deconvolution with skip-transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5499–5509, 2021. [1](#), [2](#), [6](#)
- [38] Yu Xiang, Tanner Schmidt, Venkatraman Narayanan, and Dieter Fox. Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes. *arXiv preprint arXiv:1711.00199*, 2017. [5](#)
- [39] Yaoqing Yang, Chen Feng, Yiru Shen, and Dong Tian. Foldingnet: Point cloud auto-encoder via deep grid deformation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 206–215, 2018. [2](#), [6](#), [12](#)
- [40] Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. *Advances in Neural Information Processing Systems*, 33:5824–5836, 2020. [2](#), [4](#), [8](#)
- [41] Wentao Yuan, Tejas Khot, David Held, Christoph Mertz, and Martial Hebert. Pcn: Point completion network. In *2018 International Conference on 3D Vision (3DV)*, pages 728–737. IEEE, 2018. [1](#), [2](#), [6](#), [12](#)
- [42] Dingyi Zhang, Yingming Li, and Zhongfei Zhang. Deep metric learning with spherical embedding. *Advances in Neural Information Processing Systems*, 33:18772–18783, 2020. [2](#), [3](#)
- [43] Junming Zhang, Weijia Chen, Yuping Wang, Ram Vasudevan, and Matthew Johnson-Roberson. Point set voting for partial point cloud analysis. *IEEE Robotics and Automation Letters*, 6(2):596–603, 2021. [2](#)
- [44] Xu Zhang, Felix Xinnan Yu, Svebor Karaman, Wei Zhang, and Shih-Fu Chang. Heated-up softmax embedding. *arXiv preprint arXiv:1809.04157*, 2018. [2](#), [3](#), [4](#)