MetaSets: Meta-Learning on Point Sets for Generalizable Representations

Chao Huang*, Zhangjie Cao*, Yunbo Wang*, Jianmin Wang, Mingsheng Long (✉)

School of Software, BNRist, Tsinghua University, China

{microhhh9,caozhangjie14,yunbo.thu}@gmail.com, {jimwang,mingsheng}@tsinghua.edu.cn

Abstract

Deep learning techniques for point clouds have achieved strong performance on a range of 3D vision tasks. However, it is costly to annotate large-scale point sets, making it critical to learn generalizable representations that can transfer well across different point sets. In this paper, we study a new problem of 3D Domain Generalization (3DDG) with the goal to generalize the model to other unseen domains of point clouds without any access to them in the training process. It is a challenging problem due to the substantial geometry shift from simulated to real data, such that most existing 3D models underperform due to overfitting the complete geometries in the source domain. We propose to tackle this problem via MetaSets, which meta-learns point cloud representations from a group of classification tasks on carefully-designed transformed point sets containing specific geometry priors. The learned representations are more generalizable to various unseen domains of different geometries. We design two benchmarks for Sim-to-Real transfer of 3D point clouds. Experimental results show that MetaSets outperforms existing 3D deep learning methods by large margins.

1. Introduction

Understanding and reasoning about 3D objects are crucial for AI robots to understand the real world. Recently, with the success of deep learning in 2D vision tasks [8, 25], 3D deep learning techniques have also obtained state-of-the-art classification/detection results in both simulated and real-world datasets of point clouds [20, 22, 38, 19]. However, the high performance of 3D deep learning is mostly achieved when the training data and testing data come from the same domain. But in practical applications, the assumption is always violated. For example, we can easily use the well-annotated CAD models collected from the Internet as training data [35, 4], but when we directly use the learned deep networks to classify real objects, we find that the performance decreases dramatically. Since scanning real objects
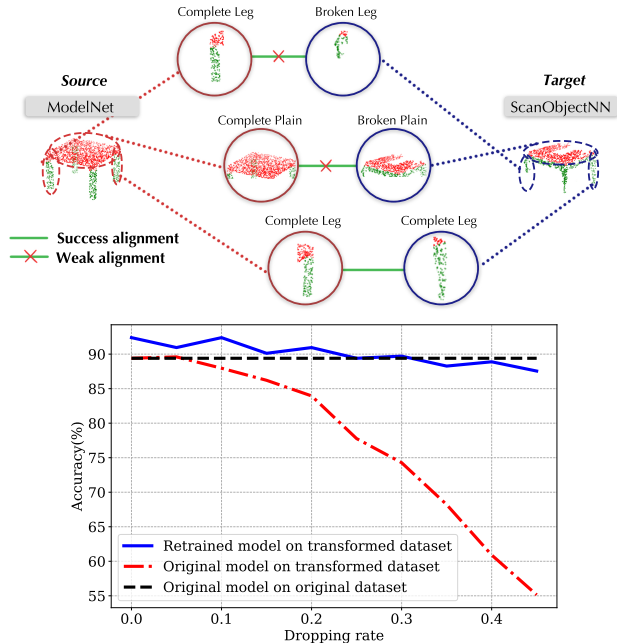
---

*Equal contribution

Figure 1. The geometric domain shift is the key challenge of 3DDG. **Top:** Domain misalignment can be easily caused by broken assemblies or missing parts. **Bottom:** The existing deep networks learned in the complete point sets underperform in the transformed point sets with broken assemblies or missing parts. The X-axis represents the ratio of randomly discarded points for each point set. Further details of transformed datasets are included in Section 3.2.

using depth cameras or LiDAR sensors are extremely costly, it is worthwhile to study the Sim-to-Real transfer learning problem for 3D point clouds.

Under these circumstances, in this paper, we present a new problem of 3D Domain Generalization (3DDG), which aims to learn a classification model in source point sets that can achieve high performance in target point sets that are not accessible during training. The training and testing data are from different domains with a large domain shift. In the case of 3DDG, it is mainly caused by various geometries of shapes from distant domains, namely the *geometry shift*. Figure 1 (top) shows an example of geometry shift between the simulated ModelNet dataset [35] and the real ScanObjectNN [29] dataset. There exist some misalignments of learned

features caused by broken assemblies or missing parts of the entire point cloud. Existing point cloud classification methods [1, 15, 32, 36, 20, 22] perform sub-optimally under such a domain shift [29]. In an early experiment of this paper, we empirically analyze their performance in target domains with incomplete point clouds. The results of PointNet [20] are shown in Figure 1 (bottom). Note that only by retraining it on the transformed target set with missing parts, can the model perform well on it. By contrast, if we directly evaluate a model trained on a full source dataset, the performance will degrade significantly, which can be caused by *overfitting the geometry of training data*.

To mitigate the geometric domain shift, we propose a meta-learning framework named MetaSets that has two contributions to 3DDG. First and foremost, MetaSets improves the previous gradient-based meta-learning algorithms by applying a learned soft-sampling strategy to the meta-tasks. It is partly inspired by the idea that the robustness of the model to out-of-distribution test samples can be improved by penalizing pre-defined data groups with weights that are inversely proportional to the sample size of the groups [26]. Specifically, we split the source dataset into a meta-training set and a meta-validation set, and perform meta-training and meta-validation on corresponding meta-tasks iteratively. In each meta-training phase, a batch of tasks is sampled with learned probabilities that are decided by the corresponding meta-validation error. This sampling strategy dynamically adjusts the training workload on all tasks and encourages the model to focus more on the difficult ones. In this way, the proposed meta-learning algorithm can effectively balance the importance of different tasks.

Another contribution of MetaSets is the transformation approaches that are specifically designed for point cloud and can be used to build different meta-tasks by expanding the source dataset. The transformed point sets simulate the cases of occlusions, missing parts, and the changes in scanning density in real environments. They can best facilitate the Sim-to-Real transfer learning by providing various geometric priors, and thus prevent the model from overfitting the source dataset. In this way, the proposed algorithm can learn both domain-invariant latent factors that greatly improve the generalization ability of the model, and domain-specific geometric priors that enhance the discriminability.

We design two Sim-to-Real benchmarks based on standard point cloud datasets, and observe that the proposed MetaSets remarkably outperforms existing approaches for 3DDG. We further validate the effectiveness of the meta-learning framework, the soft-sampling strategy with learned probabilities, as well as each transformed point set.

2. Problem Setup

This work studies the problem of 3D Domain Generalization (3DDG), which has been missing a full study at present. In 3DDG, we have a source domain consisting of point clouds and labels $\mathcal{D}_s = \{(\mathbf{P}_s, y_s)\}$, and the goal is to train a highly generalizable model f to achieve low classification errors $\mathcal{E} = \mathbb{E}_{(\mathbf{P}, y) \sim D_t}[f(\mathbf{P}) \neq y]$ on an unseen target domain $\mathcal{D}_t = \{(\mathbf{P}_t, y_t)\}$. Only the data and labels in the source domain are available during training, while those in the target domain are only used for evaluation. The technical challenge of this problem is the variations of data distribution between $\mathcal{D}_s$ and $\mathcal{D}_t$, which usually exist in simulation-to-reality (Sim-to-Real) transfer learning scenarios due to geometry deformation or the change of the point cloud density, and may violate the i.i.d. assumption of the existing deep learning approaches for point cloud classification.

3. MetaSets

In this section, we first introduce the overall framework of a new meta-learning approach for 3D domain generalization named MetaSets. Then we describe the specific meta-tasks in the framework, from which the model can learn meta-level geometric priors and generalize the features to other unseen domains. Figure 2 shows a schematic of our approach.

3.1. The MetaSets Framework

A possible solution to 3DDG is to mitigate the potential geometry shift of a variety of 3D point sets by learning the domain invariant geometric structures. To this end, we propose the MetaSets framework, as shown in Alg. 1, which follows the basic idea of Model-Agnostic Meta-Learning (MAML) [6], and contributes to ensuring the balance of various meta-tasks in the learning process so that features can be broadly generalizable to other unseen domains. We use MAML because it is a typical gradient-based meta-learning approach, and MetaSets is supposed to be a general framework that can be combined with more advanced meta-learning algorithms.

Meta-tasks.   Since in practice we have only one source domain $\mathcal{D}_s$ of point sets that can be used for training, we need to augment $\mathcal{D}_s$ with reasonable data transformations to build multiple meta-tasks. Assuming that there has been a set of transformation functions $\{F_n\}_{n=1}^N$ that we can use off-the-shelf, we may construct n classification tasks $\{\mathcal{T}_n\}_{n=1}^N$ for the transformed point sets, and divide each of them into a meta-training set and a meta-validation set. We will discuss the details of F_n in Section 3.2.

Meta-training.   An inductive bias of MetaSets is that effective *meta-learners* are supposed to perform well on all meta-tasks. For this purpose, in MetaSets, we improve the meta-training process of MAML [6] by explicitly balancing the contributions of all meta-tasks to the meta-learner. As shown in **Line 6** in Alg. 1, at each meta-training step, we dynamically sample K transformation functions denoted by $\{F'_k\}_{k=1}^K$ from the function set, instead of pre-progressing

Algorithm 1 Training process of MetaSets

Input: Source dataset $\mathcal{D}_s = (\mathcal{D}_s^{\text{train}}, \mathcal{D}_s^{\text{val}})$, data minibatch size B, transformation functions $\{F_n\}_{n=1}^N$, number of sampled functions K, validation error bound ϵ, learning rates η and β

1: **Initialize** $\theta \leftarrow \theta_0$
2: $\{p_n\}_{n=1}^N = \frac{1}{N}$ ▷ Initialize the task sample probability
3: **while** $\mathcal{L}_n^{\text{val}} < \epsilon$ for $n = 1, \ldots, N$ **do**
4:     **repeat**   ▷ Meta-training phase
5:         $\{(\mathbf{P}_s^i, y_s^i)\}_{i=1}^B \sim \mathcal{D}_s^{\text{train}}$
6:         $\{F_k'\}_{k=1}^K \sim \text{multinomial}(\{F_n\}_{n=1}^N; K)$ w.r.t. $\{p_n\}_{n=1}^N$ ▷ Sample K functions
7:         **for** $k = 1, \ldots, K$ **do**
8:            $\{\mathbf{P}_k^i\}_{i=1}^B = \{F_k'(\mathbf{P}_s^i)\}_{i=1}^B$ ▷ Obtain transformed point sets
9:            $\mathcal{L}_k \leftarrow \{f_\theta(\mathbf{P}_k^i)\}_{i=1}^B; \ \theta_k' = \theta - \eta\nabla_\theta\mathcal{L}_k$ ▷ According to Eqn (1)
10:        $\mathcal{L}^{\text{cls}} \leftarrow \{f_{\theta_k'}(\mathbf{P}_k^i)\}_{i=1,k=1}^{B,K}; \ \theta \leftarrow \theta - \beta\nabla_\theta\mathcal{L}^{\text{cls}}$ ▷ According to Eqn (2)
11:     **until** end of $\mathcal{D}_s^{\text{train}}$
12:     **repeat**   ▷ Meta-validation phase
13:         $\{(\mathbf{P}_s^i, y_s^i)\}_{i=1}^B \sim \mathcal{D}_s^{\text{val}}$
14:         **for** $n = 1, \ldots, N$ **do**
15:            $\{\mathbf{P}_n^i\}_{i=1}^B = \{F_n(\mathbf{P}_s^i)\}_{i=1}^B$ ▷ Obtain transformed point sets
16:            $\mathcal{L}_n^{\text{val}} \leftarrow \{f_\theta(\mathbf{P}_n^i)\}_{i=1}^B$ ▷ According to Eqn (3)
17:         $\{p_n\}_{n=1}^N \leftarrow \{\text{softmax}(\{\mathcal{L}_n^{\text{val}}\}_{n=1}^N)\}_{n=1}^N$ ▷ According to Eqn (4)
18:     **until** end of $\mathcal{D}_s^{\text{val}}$
19: **return** θ

$\mathcal{D}_s$ in advance. The probabilities of sampling $\{p_n\}_{n=1}^N$ are initialized with a uniform distribution across all the tasks and updated at each meta-validation step (**Line 17** in Alg. 1). In other words, MetaSets not only learns to complete the predefined tasks but also learns to balance their impact on the meta-training process. Based on a minibatch of point clouds $\{\mathbf{P}_s^i\}_{i=1}^B$ randomly sampled from the source meta-training set $\mathcal{D}_s^{\text{train}}$, MetaSets maps them to K transformed point sets $\mathcal{D}_k^{\text{train}} = \{(\mathbf{P}_k^i, y_s^i)\}_{i=1}^B$, where $k \in \{1, \ldots, K\}$ (**Line 8** in Alg. 1). On each task, it computes the classification loss:

$$\mathcal{L}_k = \frac{1}{B}\sum_{i=1}^B \ell\left(f_\theta(\mathbf{P}_k^i), y_s^i\right), \qquad (1)$$

where N is the minibatch size, $\ell(\cdot)$ is the cross-entropy loss, and f_θ is a point cloud classification model, *e.g.*, PointNet [20], parameterized by θ. We compute the parameters after one gradient update as $\theta_k' = \theta - \eta\nabla_\theta\mathcal{L}_k$, then use the following meta-objective to minimize the classification error over all sampled tasks with updated parameters θ_k':

$$\mathcal{L}^{\text{cls}} = \sum_{k=1}^K \frac{1}{B}\sum_{i=1}^B \ell\left(f_{\theta_k'}(\mathbf{P}_k^i), y_s^i\right). \qquad (2)$$

Minimizing the meta-objective achieves high performance on each sampled task, such that we can update the classification model by $\theta \leftarrow \theta - \beta\nabla_\theta\mathcal{L}^{\text{cls}}$.

Meta-validation. After one training epoch over $\mathcal{D}_s^{\text{train}}$, MetaSets performs the mata-validation process to evaluate the classification loss under parameter θ on $\mathcal{D}_s^{\text{val}}$ for the total of N tasks. Concretely, we first sample B point clouds from $\mathcal{D}_s^{\text{val}}$ and transform them into $\mathcal{D}_n^{\text{val}}$ for all N transformation functions $\{F_n\}_{n=1}^N$, and then for each task $\mathcal{T}_n$, we have

$$\mathcal{L}_n^{\text{val}} = \mathbb{E}_{(\mathbf{P},y)\sim\mathcal{D}_n^{\text{val}}} \ell\left(f_\theta(\mathbf{P}), y\right). \qquad (3)$$

The validation loss $\mathcal{L}_n^{\text{val}}$ serves two purposes. First, it can be used as a signal of convergence for the classification model. If $\mathcal{L}_n^{\text{val}}$ is smaller than a small threshold ϵ (*i.e.*, validation error bound) for all tasks $n \in \{1, \ldots, N\}$, the model is considered to be converged.

Second and more importantly, at the end of each meta-validation process, we use $\{\mathcal{L}_n^{\text{val}}\}_{n=1}^N$ to compute the probability of sampling each task for the next meta-training phase, as shown in Figure 2. We propose to increase the probabilities of tasks with higher validation losses to be selected in the next meta-training process so that MetaSets can focus more on these challenging tasks. Specifically, in **Line 17** in Alg. 1, we update the probability of each task by performing softmax over $\{\mathcal{L}_n^{\text{val}}\}_{n=1}^N$:

$$p_n = \frac{\exp(\mathcal{L}_n^{\text{val}})}{\sum_{n=1}^N \exp(\mathcal{L}_n^{\text{val}})}. \qquad (4)$$
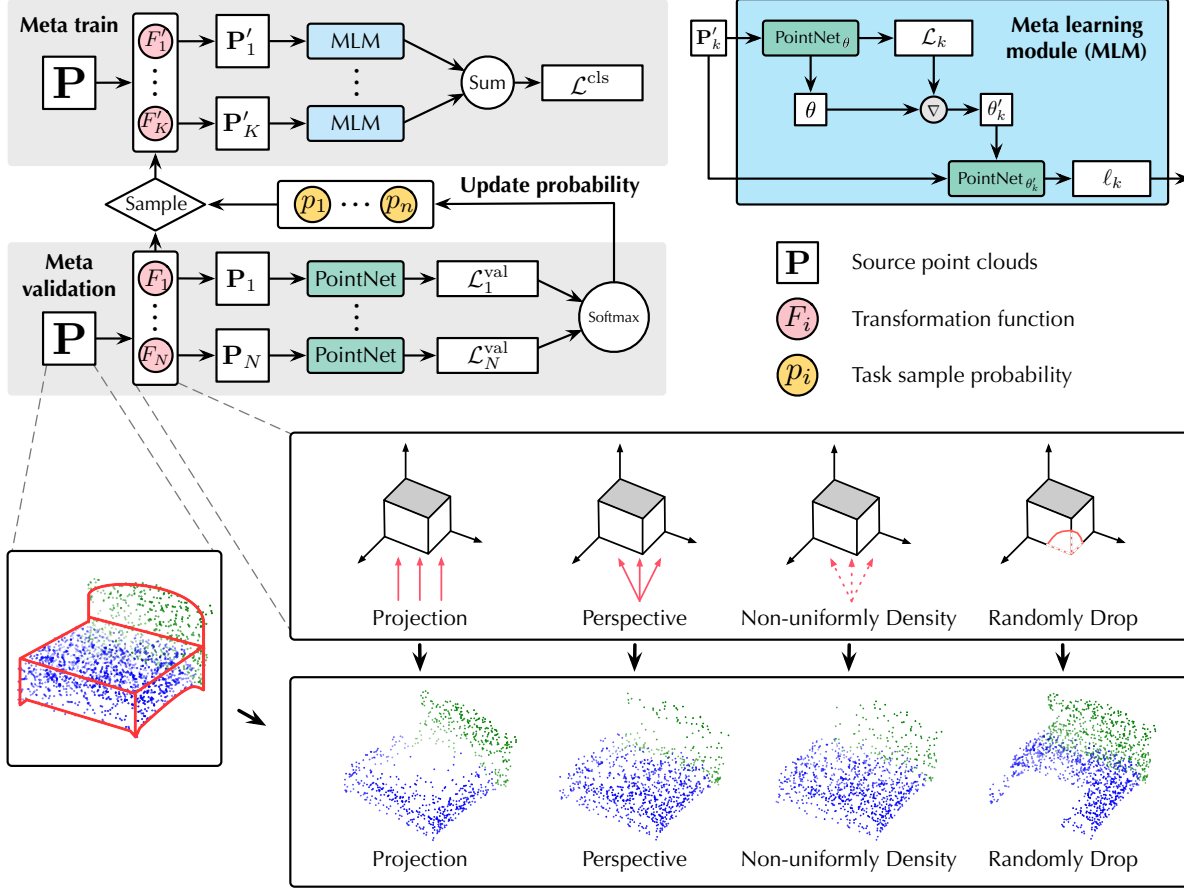
Figure 2. A schematic of the meta-learning framework of MetaSets.

The updated probabilities are then applied to the sampling process of the transformation functions and balance the importance of N meta-tasks in the next training epoch. In 3DDG scenarios, learning $\{p_n\}_{n=1}^{N}$ dynamically prevents the classification model from overfitting the source domain, and encourages it to generalize across a wide range of possible variations of the source point clouds, such as geometric distortion and missing parts. Unlike the MAML algorithm [6] that mainly focuses on encoding domain-invariant knowledge in the meta-leaner, the proposed MetaSets performs *soft-sampling* on meta-tasks to further address the imbalance of their impact, and thus enables the model to learn both domain-invariant and domain-specific knowledge from the expanded source domains.

For the convergence of MetaSets, since the standard meta-learning paradigm has been proved to be able to converge theoretically [5] and empirically [6] even for diverse tasks with a large number of data points and MetaSets follows a similar training paradigm to standard meta-learning, so MetaSets can also converge soundly. For the time cost, the proposed MetaSets only requires an extra meta-validation step compared to standard meta-learning. However, the meta-validation step costs much less time than the meta-

training step (about $1 : 30$ on average in experiments). So the time cost of MetaSets is comparable to standard meta-learning. In experiments, we show that the convergence speed of MetaSets is similar to standard meta-learning and comparable to standard point cloud classification models.

3.2. Transformed Point Sets

The quality of the meta-tasks is the key to the effectiveness of MetaSets, as is the case in any other MAML variant. To improve the generalization ability of the learned model, the tasks are supposed to cover a wide range of possible variations of point clouds from the source domain. As mentioned above, we expand the source domain by applying a set of transformation functions $\{F_n\}_{n=1}^{N}$ to the original point clouds. At the same time, the diversity and representativeness of the transformed point sets are considered. Due to the work of Wu *et al.* [33], the common geometry shift for point clouds is mainly caused by occlusions, density changes, and scanning noises. Accordingly, we design the following three transformation approaches, as shown in Figure 2, that can be used to generate a variety of meta-tasks by changing the hyperparameters or using different combinations of them.

Non-uniform density (P_1, g). When scanning an object with a LiDAR, the closer it is, the denser the point cloud will be. We simulate the non-uniform density by firstly setting an anchor position P_1 outside the point cloud and calculating its distances to each point, where P_1 is randomly selected from a unit sphere. We then discard the points with probabilities (*i.e.*, dropping rate) proportional to the distances. To construct different point sets, we first normalize the distance from P_1 to every point in the point cloud within the range of $[0, 1]$, which is used as the basic drop rate. Then we multiply the basic drop rate with a multiplier $g > 1$, which is called the gate parameter and controls the density of the point cloud, where a larger g means more dropping points and thus a sparser point cloud. The point distribution and the density of the transformed point cloud can be controlled by P_1 and g respectively.

Dropping $(P_2, x\%)$. Inspired by the idea of dropout that keeps deep networks from overfitting, we propose to randomly drop parts of point cloud structures. We randomly select a position P_2 in the point cloud and drop the nearest $x\%$ points, where P_2 and $x\%$ are controllable hyperparameters. This transformation method has two benefits. First, it produces meta-tasks that can prevent the model from overfitting the source domain. Second, it simulate the cases of broken assemblies or missing parts in real environments, and thus can best facilitate the Sim-to-Real generalization.

Self-occlusion $(\vec{v}, W)$. Self-occlusion occurs when the shape is viewed from a certain angle that the back surface of the point clouds is invisible. As shown in Figure 2, we simulate occlude by specifically designing a parallel projection method for point clouds, and take it as a basic operation to construct the meta-tasks. To be specific, we first choose a plain outside the 3D object and project the points cloud to the plain along its normal vector $\vec{v}$. All optional planes are supposed to have the same minimum distance to the point cloud. We then divide the plane into grids of equal size. Within each grid, we keep the closest point along the normal vector and remove the others. We can control the grid size of W to obtain different occlusion patterns. As it approaches infinity, it retains only one point, and when it is small enough, the entire point cloud will be preserved and no occlusion occurs. We can also control the projection angle, *i.e.*, the normal vector of the plane. The transformed point sets with self-occlusion can greatly benefit the transfer learning problem of 3D point clouds in Sim-to-Real scenarios as the inner structure of real-world data is not visible.

In summary, through the different types of transformations introduced above, we can simulate different geometry shifts of self-occlusions, density changes, and missing parts. We control the hyperparameters in $(P_1, g, P_2, x\%, \vec{v}, W)$ to develop a set of functions $\{F_n\}_{n=1}^N$ at the very beginning of the training process. Each of them can be used to con-

struct a transformed point set $\mathcal{D}_n = \{(F_n(\mathbf{P}), y)\}$, where $(\mathbf{P}, y) \in \mathcal{D}_s$. However, it is worth noting that, the imbalance of the task difficulty may make the meta-learning approach less effective for 3DDG. To solve this problem, we introduce the *soft-sampling* technique to dynamically adjust the training workload of the model on each task.

Another benefit of the above transformation methods is that we design tasks at the input level so that they can be easily combined with existing approaches that have specific designs in model architectures [23, 24]. Besides, the overall MetaSets framework is extendable and can be generalized to other target domains by simply defining new tasks with specific geometric priors.

4. Experiments

To evaluate MetaSets in Sim-to-Real scenarios, we build two 3DDG benchmarks upon the synthetic datasets, Model-Net [35] and ShapeNet [4], as well as a real dataset ScanObjectNN [29]. See supplementary materials for more details.

Compared methods. We compare MetaSets with five state-of-the-art point cloud classification models: Point-Net [20], PointNet++ [22], DGCNN [32], ConvPoint [2], LDGCNN [37], and PointCNN [15], covering both PointNet-based, CNN-based and Graph-based methods. Besides, we further compare MetaSets with PointDAN [23], a domain adaptation approach for point clouds based on PointNet.It is worth mentioning that PointDAN [23] requires unlabeled target data for training, while for 3DDG, the distribution of target data is not accessible. Thus, compared with PointDAN [23], MetaSets tries to tackle a more challenging problem.

Implementation details. We divide each source domain into a training set and a validation set at a scale of $5 : 1$, following the experimental protocol of domain generalization [7]. Similar to [6], we implement our meta-learning algorithm with first-order approximation to achieve high computational efficiency. For all the compared methods, we tune hyperparameters by performing cross-validation in the source domain. For MetaSets, we construct three tasks for each transformation method, where the parameters are selected as follows: (1) we monotonically change the parameters, from preserving all points to deleting all points; When we observe that the shape of the resulting point cloud begins to change, we record the parameters as t_1, and when we observe that the shape is almost unrecognizable, we record the parameters as t_2. (2) We randomly sample three parameters within the range of (t_1, t_2) and add them to the task set. At each iteration in the meta-training phase, we sample four tasks from the task set of all transformations. Hyperparameters η and β are tuned by cross-validation, as shown in Alg. 1. Please see supplementary materials for their sensitivity analysis. Unless otherwise specified, we use PointNet [20] as the backbone of MetaSets. All experiments
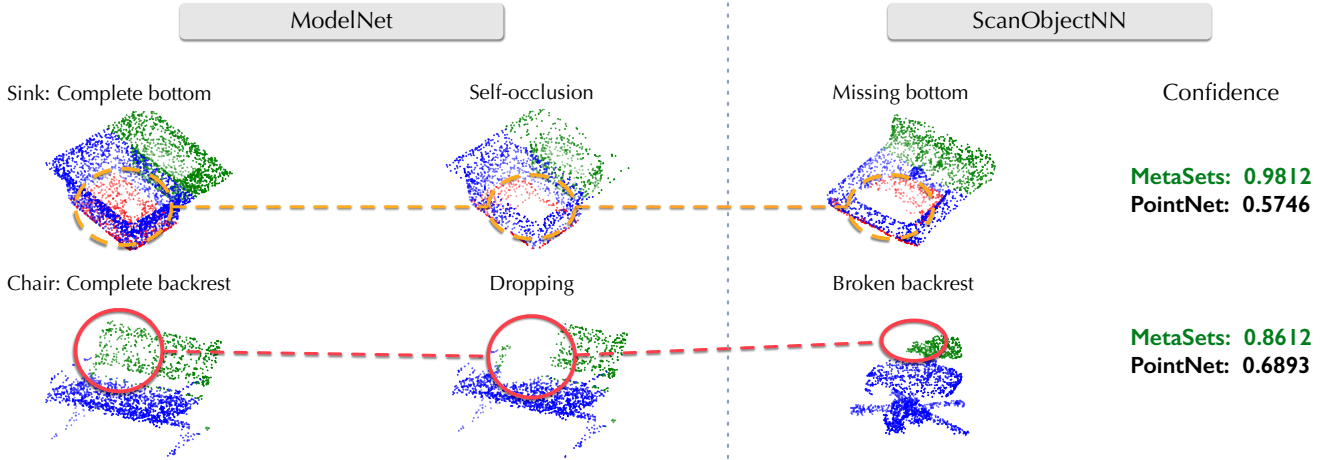
Figure 3. Visualization of point clouds from the source domain (**first column**), the transformed point sets (**second column**), and the target domain (**third column**). The geometries of the transformed point clouds match well with those of the target clouds, which enables the proposed MetaSets to alleviate the influence of geometric domain shift.

are conducted using PyTorch [18]. We perform each experiment three times and report the average and the standard deviation of the results.

4.1. Geometry Overfitting

We first empirically study how our current point cloud classification methods are affected by geometry shift. As shown in Figure 1 (please go back to the first page), we evaluate PointNet [20] on the ModelNet dataset by randomly dropping some parts of the point cloud. As we can see, the classification accuracy becomes lower as more points are deleted. One might argue that this performance degradation is because the remainder of the point cloud does not have sufficient geometric information and is, therefore, less recognizable. Thus, we re-train the PointNet on transformed point clouds and the performance rises back to the model on the original dataset. The results show that the point clouds after transformation are still recognizable, and the reason for the performance degradation is the geometry shift between training and testing, which is a common situation in most Sim-to-Real scenarios. The model overfits the geometry of the complete point cloud shapes and cannot recognize the geometry of transformed shapes.

4.2. ModelNet to ScanObjectNN

On this benchmark, we select the 11 categories shared by the ModelNet40 and ScanObjectNN datasets. Models are trained on ModelNet and evaluated on ScanObjectNN. We report the class-wise results in the supplementary materials.

Visualization of transformed point sets. Figure 3 illustrates the differences between geometries in the synthetic source domain and those in the real target domain, and how the transformed point sets contribute to learning more generalizable features. Due to self-occlusion or missing parts,

Method	Object	Object & Background
PointNet [20]	55.90±1.47	49.48±2.28
PointNet++ [22]	47.30±0.53	40.42±1.17
ConvPoint [2]	57.40±0.44	55.44±0.32
DGCNN [32]	61.68±1.26	57.61±0.44
PointCNN [15]	50.32±0.43	46.11±0.43
LDGCNN [37]	62.29±0.22	58.83±0.43
PointDAN [23] on PointNet	63.32±0.85	55.13±0.97
MetaSets on PointNet	68.28±0.79	57.19±1.23
MetaSets on ConvPoint	65.05±0.56	61.33±0.32
MetaSets on DGCNN	**72.42**±0.21	**65.66**±1.06

Table 1. Accuracy (%) on the ModelNet→ScanObjNN benchmark.

these point clouds that are randomly sampled from the real domain suffer from "missing bottom" and "broken backrest", which leads to the decrease of the classification confidence of the original PointNet. In contrast, the transformed point sets proposed in this work can help bridge the geometry gap between complete and imperfect point clouds, enabling the meta-learning model to learn representations that can be easily generalized to the geometries in the target domain.

Comparison with the single-domain state of the art. As we can see from Table 1, MetaSets outperforms previous point cloud classification methods, including the state-of-the-art LDGCNN, under the 3DDG setting. The sub-optimal performance of previous single-domain approaches, which are also trained on the source domain only, is caused by overfitting the geometry of complete point clouds in the training set and thus lacking the ability to generalize to the real domains with imperfect shapes. MetaSets obtains remarkable and consistent improvement, indicating that it is model-agnostic and performs better when the backbone network gets stronger.

Method	None	Density	Dropping	Self-occlusion	All tasks
MetaSets on PointNet	**58.95**	**65.47**	**64.00**	**63.58**	**68.28**
MetaSets w/o meta-learning	55.90	62.58	62.95	61.47	63.16

Table 2. The ablation study of using different meta-tasks (**columns**) and different optimization algorithms (**rows**) on ModelNet→ScanObjNN.

Meta-training method	Density		Dropping		Self-occlusion		Avg	
	Acc (%)	Loss	Acc (%)	Loss	Acc (%)	Loss	Acc (%)	Loss
Density	59.58	1.49	57.26	1.61	62.53	1.42	59.79	1.51
Dropping	56.63	1.68	58.30	1.69	62.31	1.55	59.08	1.64
Self-occlusion	51.16	1.85	53.73	1.93	62.74	1.68	55.88	1.82
Final: MetaSets	**68.42**	**1.13**	**65.05**	**1.20**	**63.16**	**1.26**	**65.54**	**1.20**
MetaSets w/o soft-sampling	67.09	1.35	60.07	1.51	60.35	1.49	62.50	1.45

Table 3. Generalization analysis across meta-tasks on the ModelNet→ScanObjNN benchmark. See text for more details.

Comparison with PointDAN [23]. PointDAN is a domain adaptation model that incorporates target unlabeled data into training. However, as shown in Table 1, it underperforms MetaSets by 63.32% vs. 68.28%. Note that (1) MetaSets and PointDAN use the same backbone: PointNet, and (2) MetaSets does not require any target data throughout the training. Thanks to the geometry priors learned from various "imagined" point clouds from the source domain, MetaSets tries to solve a more challenging problem while turns out to achieve superior results. Therefore, we can consider MetaSets as a good solution to the 3DDG problem.

Ablation study on meta-tasks. Table 2 compares the effectiveness of each individual meta-task. We respectively train the model on the source dataset without any transformed data (the "None" column), the expanded source dataset with data transformation, and the one with all types of transformed data (the "All tasks" column). The first row is the proposed MetaSets that integrates geometrical transformations with the meta-learning approach based on soft-sampling. The second row, "w/o meta-learning", means learning directly on the above training dataset rather than using any meta-learning algorithms. It is worth noting that the soft-sampling technique is only effective if the number of tasks in the meta-learning process is greater than 2. Particularly, for the "None" column, the difference between the two models is that MetaSets is trained with the MAML algorithm on the raw dataset while the "w/o meta-learning" baseline model is trained using the naïve SGD, as opposed to bilevel optimization (L9-10 in Alg. 1). For both compared models, we can observe that each individual transformation method (as the meta-task or data augmentation) independently brings in performance gains upon the "None" column, and the best performance is achieved when using all of them, indicating that these transformation methods are complementary to each other for inferring different geometry priors.

Generalization across meta-tasks. We further evaluate the generalization ability of MetaSets even across different transformed sub-datasets. For each of the first three rows in Table 3, we only use one type of transformation method to build the meta-task, and then evaluate the model on point sets produced by different types of transformation methods. For the last two rows, we use the full set of meta-tasks in the training process. Each column indicates the task on which meta-validation is performed, where we report the meta-validation accuracy (*Acc*) and cross-entropy loss (*Loss*). The results show that the model learned from the point clouds after non-uniform density transformation can be better generalized to other types of data with different geometry variations. The final results are further improved by using all three transformation techniques to build meta-tasks.

Ablation study on model components. From Table 2, MetaSets trained with the proposed meta-learning algorithm significantly outperforms the "w/o meta-learning" baseline model that uses the transformed point sets to augment the source dataset. The results validate the effectiveness of meta-learning. Further, Table 3 compares the final MetaSets with a baseline model that is trained with all meta-tasks but without soft-sampling. It can be observed that soft-sampling reduces the cross-entropy loss of meta-validation remarkably. In fact, the classification difficulty is not the same on all transformed point sets, and the proposed soft-sampling method allows the training process to focus more on the difficult ones, so as to reduce the meta-validation loss of all tasks more effectively.

4.3. ShapeNet to ScanObjetNN

To ensure that the performance gain is not due to the bias of our method to a particular benchmark, we conduct experiments on ShapeNet to ScanObjectNN. In this benchmark, we train on the ShapeNet and evaluate on the ScanObjectNN. We select the 9 categories shared by ShapeNet and ScanObjectNN. As shown in Table 4, our MetaSets still outperforms all the other methods testing on objects either without or with a background. In particular, MetaSets achieves a larger margin on objects with background, which has a large domain shift from the source domain. Note that with a more advanced backbone (DGCNN vs. PointNet), the margin

Method	Object	Object & Background
PointNet [20]	54.00±0.32	45.50±0.99
PointNet++ [22]	45.50±0.64	43.25±1.23
ConvPoint [2]	52.58±0.58	50.67±0.88
DGCNN [32]	57.42±1.01	54.42±0.80
PointCNN [15]	49.42±0.29	43.92±0.63
LDGCNN [37]	57.92±0.63	52.50±0.25
PointDAN [23] on PointNet	54.95±0.87	43.00±0.95
MetaSets on PointNet	55.25±0.35	49.50±0.43
MetaSets on DGCNN	**60.92**±0.76	**59.08**±1.01

Table 4. Accuracy (%) on the ShapeNet→ScanObjNN benchmark.

between MetaSets and the backbone is larger, which indicates an advanced backbone better utilizes the features from MetaSets and MetaSets has the potential to achieve higher performance with a better backbone in the future.

5. Related Work

Point cloud classification. Deep learning methods for 3D shape data were originally designed for 3D meshes [35], multi-view images [28, 21, 3], and voxels [16]. More recently, the emergence of point clouds has encouraged a growing number of deep networks specifically designed for this simple 3D data structure. PointNet addresses the permutation invariance of point clouds by leveraging the max pooling [20]. PointNet++ improves PointNet by encoding point clouds hierarchically at multiple scales [22]. Another focus of recent approaches is to develop the convolution operation applied to point clouds [15, 32, 36, 34]. However, all of these methods assume that both the training and testing data are from the same domain. When generalized across domains, these methods tend to overfit the geometry of the source domain and significantly degrade performance in the target domain. Therefore, a new method is needed to address the domain generalization problem for 3D point clouds.

Learning generalizable features. The domain generalization (DG) problem is to train a model that can be generalized to unseen target domains, which has been studied extensively in the field of image classification [17]. Most existing methods focus on learning domain invariant representations [17, 7, 14, 12, 13]. Khosla *et al.* [9] and Li *et al.* [10] used a hierarchy of parameters composed of domain-agnostic and domain-specific parts. Shankar *et al.* [27] and Volpi *et al.* [31] augmented the training domain with adversarial perturbations. Li *et al.* [11] proposed to split source domains into meta-train and meta-test splits and conduct standard MAML across splits without data transformations. However, all of these methods are only validated on image datasets but may fail on the 3D data due to the large gap between 2D images and 3D point clouds. Previous work for image classification has shown that some heavy data augmentation techniques, such as color jittering and rotation, in some cases can be

more helpful than gradient-based adversarial perturbations, when the goal is to improve domain generalization performance [30]. Similarly, the jittering method has also been used for point clouds by PointNet [20] for the general data augmentation purpose, but this approach does not explicitly consider any possible data discrepancy across domains. Different from the prior work, we propose three transformation approaches which are complementary to each other and particularly designed for 3DDG. These approaches show better performance than straightforward 3D data augmentation.

Cross-domain point cloud classification. As the deep networks achieve remarkable progress in recognizing synthetic point clouds, the research focus in this field is gradually shifting to real-world data by improving the transferability of deep models. But, as shown in [29], the performance degrades due to the existence of a simulation-to-reality gap, such as geometry deformation or the change of the point cloud density. In order to reduce the domain gap, PointDAN minimizes the Maximum Mean Discrepancy across domains [23]. However, it can only be used as a solution to the problem of domain adaptation with known target domain data distribution. Unlike the above methods, our work provides an early study of 3DDG by performing meta-learning over the *transformed point sets* that contain a variety of geometry priors of point clouds for generalizable representations.

6. Conclusion

This paper presented a new research problem of 3DDG, which aims to learn generalizable models that can be transferred to unseen target domains. Due to the geometry shift, the previous deep models tend to overfit the geometry of the synthetic point clouds and deteriorate on the target real-world data. In this work, we proposed MetaSets. First, it improves MAML by introducing a soft-sampling technique for the meta-tasks, which dynamically adjusts the training workload on each task, and enables the model to focus on the more difficult tasks. Second, we defined a set of basic data transformation methods to construct the meta-tasks, such that the transformed point sets can cover various point cloud geometries in real domains. We designed two Sim-to-Real 3DDG benchmarks, and demonstrated that MetaSets remarkably outperformed the state-of-the-art point clouds classification models, even including a domain adaptation approach that uses target data for training.

Acknowledgments

References

[1] Yizhak Ben-Shabat, Michael Lindenbaum, and Anath Fischer. 3DmFV: Three-dimensional point cloud classification in real-time using convolutional neural networks. *RA-L*, 3(4):3145–3152, 2018.

[2] Alexandre Boulch. Convpoint: Continuous convolutions for point cloud processing. *Computers & Graphics*, 88:24 – 34, 2020.

[3] Zhangjie Cao, Qixing Huang, and Ramani Karthik. 3d object classification via spherical projections. In *3DV*, pages 566–574. IEEE, 2017.

[4] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. ShapeNet: An information-rich 3D model repository. *arXiv preprint arXiv:1512.03012*, 2015.

[5] Alireza Fallah, Aryan Mokhtari, and Asuman Ozdaglar. On the convergence theory of gradient-based model-agnostic meta-learning algorithms. In *AISTATS*, pages 1082–1092, 2020.

[6] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *ICML*, pages 1126–1135, 2017.

[7] Muhammad Ghifary, W Bastiaan Kleijn, Mengjie Zhang, and David Balduzzi. Domain generalization for object recognition with multi-task autoencoderss. In *ICCV*, pages 2551–2559, 2015.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016.

[9] Aditya Khosla, Tinghui Zhou, Tomasz Malisiewicz, Alexei A Efros, and Antonio Torralba. Undoing the damage of dataset bias. In *ECCV*, pages 158–171. Springer, 2012.

[10] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. Deeper, broader and artier domain generalization. In *ICCV*, pages 5542–5550, 2017.

[11] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. Learning to generalize: Meta-learning for domain generalization. *arXiv preprint arXiv:1710.03463*, 2017.

[12] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. Learning to generalize: Meta-learning for domain generalization. In *AAAI*, 2018.

[13] Da Li, Jianshu Zhang, Yongxin Yang, Cong Liu, Yi-Zhe Song, and Timothy M Hospedales. Episodic training for domain generalization. In *ICCV*, pages 1446–1455, 2019.

[14] Haoliang Li, Sinno Jialin Pan, Shiqi Wang, and Alex C Kot. Domain generalization with adversarial feature learning. In *CVPR*, pages 5400–5409, 2018.

[15] Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. Pointcnn: Convolution on x-transformed points. In *NeurIPS*, pages 820–830, 2018.

[16] Daniel Maturana and Sebastian Scherer. VoxNet: A 3d convolutional neural network for real-time object recognition. In *IROS*, pages 922–928, 2015.

[17] Krikamol Muandet, David Balduzzi, and Bernhard Schölkopf. Domain generalization via invariant feature representation. In *ICML*, pages 10–18, 2013.

[18] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché Buc, E. Fox, and R. Garnett, editors, *NeurIPS*, pages 8024–8035. 2019.

[19] Charles R Qi, Wei Liu, Chenxia Wu, Hao Su, and Leonidas J Guibas. Frustum pointNets for 3D object detection from RGB-D data. In *CVPR*, pages 918–927, 2018.

[20] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. PointNet: Deep learning on point sets for 3D classification and segmentation. In *CVPR*, pages 652–660, 2017.

[21] Charles R Qi, Hao Su, Matthias Nießner, Angela Dai, Mengyuan Yan, and Leonidas J Guibas. Volumetric and multi-view CNNs for object classification on 3D data. In *CVPR*, pages 5648–5656, 2016.

[22] Charles R Qi, Li Yi, Hao Su, and Leonidas J Guibas. PointNet++: Deep hierarchical feature learning on point sets in a metric space. In *NeurIPS*, pages 5099–5108, 2017.

[23] Can Qin, Haoxuan You, Lichen Wang, C-C Jay Kuo, and Yun Fu. PointDAN: A multi-scale 3D domain adaption network for point cloud representation. In *NeurIPS*, pages 7190–7201, 2019.

[24] Yongming Rao, Jiwen Lu, and Jie Zhou. Global-local bidirectional reasoning for unsupervised representation learning of 3d point clouds. In *CVPR*, 2020.

[25] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: towards real-time object detection with region proposal networks. In *NeurIPS*, pages 91–99, 2015.

[26] Shiori Sagawa, Pang Wei Koh, Tatsunori B Hashimoto, and Percy Liang. Distributionally robust neural networks. In *ICLR*, 2019.

[27] Shiv Shankar, Vihari Piratla, Soumen Chakrabarti, Siddhartha Chaudhuri, Preethi Jyothi, and Sunita Sarawagi. Generalizing across domains via cross-gradient training. *arXiv preprint arXiv:1804.10745*, 2018.

[28] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view convolutional neural networks for 3D shape recognition. In *ICCV*, pages 945–953, 2015.

[29] Mikaela Angelina Uy, Quang-Hieu Pham, Binh-Son Hua, Thanh Nguyen, and Sai-Kit Yeung. Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In *ICCV*, pages 1588–1597, 2019.

[30] Riccardo Volpi and Vittorio Murino. Addressing model vulnerability to distributional shifts over image transformation sets. In *ICCV*, pages 7980–7989, 2019.

[31] Riccardo Volpi, Hongseok Namkoong, Ozan Sener, John C Duchi, Vittorio Murino, and Silvio Savarese. Generalizing to unseen domains via adversarial data augmentation. In *NeurIPS*, pages 5334–5344, 2018.

[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. *TOG*, pages 1–12, 2019.

[33] Bichen Wu, Xuanyu Zhou, Sicheng Zhao, Xiangyu Yue, and Kurt Keutzer. SqueezesegV2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a liDAR point cloud. In *ICRA*, pages 4376–4382. IEEE, 2019.

[34] Wenxuan Wu, Zhongang Qi, and Li Fuxin. PointConv: Deep convolutional networks on 3D point clouds. In *CVPR*, pages 9621–9630, 2019.

[35] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3D shapeNets: A deep representation for volumetric shapes. In *CVPR*, pages 1912–1920, 2015.

[36] Yifan Xu, Tianqi Fan, Mingye Xu, Long Zeng, and Yu Qiao. SpiderCNN: Deep learning on point sets with parameterized convolutional filterss. In *ECCV*, pages 87–102, 2018.

[37] Kuangen Zhang, Ming Hao, Jing Wang, Clarence W de Silva, and Chenglong Fu. Linked dynamic graph cnn: Learning on point cloud via linking hierarchical features. *arXiv preprint arXiv:1904.10014*, 2019.

[38] Yin Zhou and Oncel Tuzel. VoxelNet: End-to-end learning for point cloud based 3D object detection. In *CVPR*, pages 4490–4499, 2018.