

# SUG: Single-dataset Unified Generalization for 3D Point Cloud Classification

Siyuan Huang<sup>1,2,\*</sup>, Bo Zhang<sup>2,†</sup>, Botian Shi<sup>2</sup>, Peng Gao<sup>2</sup>, Yikang Li<sup>2</sup>, Hongsheng Li<sup>3</sup>

<sup>1</sup>Shanghai Jiaotong University, <sup>2</sup>Shanghai AI Laboratory, <sup>3</sup>CUHK MMLab

siyuan.sjtu@sjtu.edu.cn, bo.zhangzx@gmail.com

## Abstract

*Although Domain Generalization (DG) problem has been fast-growing in the 2D image tasks, its exploration on 3D point cloud data is still insufficient and challenged by more complex and uncertain cross-domain variances with uneven inter-class modality distribution. In this paper, different from previous 2D DG works, we focus on the 3D DG problem and propose a Single-dataset Unified Generalization (SUG) framework that only leverages a single source dataset to alleviate the unforeseen domain differences faced by a well-trained source model. Specifically, we first design a Multi-grained Sub-domain Alignment (MSA) method, which can constrain the learned representations to be domain-agnostic and discriminative, by performing a multi-grained feature alignment process between the splitted sub-domains from the single source dataset. Then, a Sample-level Domain-aware Attention (SDA) strategy is presented, which can selectively enhance easy-to-adapt samples from different sub-domains according to the sample-level inter-domain distance to avoid the negative transfer. Experiments demonstrate that our SUG can boost the generalization ability for unseen target domains, even outperforming the existing unsupervised domain adaptation methods that have to access extensive target domain data. Our code is available at <https://github.com/SiyuanHuang95/SUG>.*

## 1. Introduction

Recently, point clouds-based vision tasks [28] have achieved remarkable progress on the public benchmarks [4, 6, 34], which largely owes to the fact that the collected point clouds are carefully annotated and sufficiently large. But in the real world, acquiring such data from a new target domain and manually labeling these extensive 3D data are highly dependent on professionals in this field.

<sup>†</sup> Project leader.

<sup>\*</sup> This work was done when Siyuan Huang was an intern at Shanghai AI Laboratory.

One effective solution to transfer the model from a fully-labeled source domain to a new domain without extra human labor is Unsupervised Domain Adaptation (UDA) [7, 27, 39, 46], whose purpose is to learn a more generalizable representation between the labeled source domain and unlabeled target domain, such that the model can be adapted to the data distribution of the target domain. For example, when point cloud data distribution from the target domain undergoes serious geometric variances [27], performing a correct source-to-target correspondence can boost the model’s adaptability. Besides, GAST [46] learns a domain-shared representation for different semantic categories, while a voting reweighting method is designed [7] that can assign reliable target domain pseudo labels. However, these UDA-based techniques are highly dependent on the accessibility of the target domain data, which is a strong assumption and prerequisite for the models running in unprecedented circumstances, such as autonomous driving systems and medical scenarios. Thus, it is meaningful and vital to investigate the model’s cross-domain generalization ability under the zero-shot target domain constraint, which derives the task of **Domain Generalization (DG)** for 3D scenarios.

However, achieving such zero-shot domain adaptation, *i.e.*, DG, is more challenging in 3D scenarios, mainly due to the following reasons. **(1) Unknown Domain-variance Challenge:** 3D point cloud data collected from different sensors or geospatial regions with different data distributions often present serious domain discrepancies. Due to the inaccessibility of the target domain data, modeling of source-to-target domain variance is intangible. **(2) Uneven Domain Adaptation Challenge:** Considering that our goal is to learn a transferable representation that can be generalized to multiple target domains, a robust model needs to perform an even domain adaptation rather than lean to fit the data distribution on one of the multiple target domains. But for 3D point cloud data with more complex sample-level modality variances, ensuring an even model adaptation under the zero-shot target domains setting remains challenging.

To tackle the above challenges, we study the typical

DG problem in the 3D scenario and introduce a Single-dataset Unified Generalization (SUG) framework to address the 3D point cloud generalization problem. We study a one-to-many domain generalization problem, where the model can be trained on only a single 3D dataset and is required to be *simultaneously generalized* to **multiple target datasets**. Different from previous DG works in 2D scenarios [5, 19, 26], 3D point cloud data have more diverse data distribution within a single dataset, which **provides the possibility to exploit the modality variations across different sub-domains** without accessing any target-domain datasets. Our SUG framework consists of a Multi-grained Sub-domain Alignment (MSA) method and a Sample-level Domain-aware Attention (SDA) strategy. To address the unknown domain-variance challenge, the SUG first splits the selected single dataset into different sub-domains with a domain split module. Then, based on these sub-domains, the baseline model is constrained to simulate as many domain variances as possible from multi-grained features so that the baseline model can learn multi-grained and multi-domains agnostic representations. The SDA is developed to solve the uneven domain adaptation challenge, which assumes that the instances from different sub-domains often present different adaptation difficulties. Thus, we add sample-level constraints to the whole sub-domain alignment process according to the dynamically changing sample-level inter-domain distance, leading to an even inter-domain adaptation process.

We conduct extensive experiments on several common benchmarks [23] under the single-dataset DG setting, which includes: 1) **ModelNet-10**→**ShapeNet-10/ScanNet-10**, meaning that the model is only trained on ModelNet-10 and directly evaluated on **both** ShapeNet-10 and ScanNet-10; 2) **ShapeNet-10**→**ModelNet-10/ScanNet-10**; 3) **ScanNet-10**→**ModelNet-10/ShapeNet-10**. Experimental results demonstrate the effectiveness of the SUG framework in learning generalizable features of 3D point clouds, and it can also significantly boost the DG ability for many selected baseline models.

Our contributions can be summarized as follows:

- 1) From a new perspective of one-to-many 3D DG, we explore the possibilities of adapting a model to multiple unseen domains and study how to leverage the feature’s multi-modal information residing in a single dataset.
- 2) We propose a SUG to tackle the one-to-many 3D DG problem. The SUG consists of a designed MSA method to learn the domain-agnostic and discriminative features during the source-domain training phase and an SDA strategy to calculate the sample-level inter-domain distance and balance the adaptation degree of different sub-domains with different inter-domain distances.

## 2. Related Works

### 2.1. 2D Domain Adaptation and Generalization

Recent Domain Adaptation (DA) works can be roughly categorized into two types: 1) Adversarial learning-based methods [8, 10, 16, 32] that focus on leveraging a domain label discriminator to reduce the inter-domain discrepancy; 2) Moment matching-based methods [14, 15, 30] that refer to aligning the first-order or second-order moments of feature distribution. But under the Domain Generalization (DG) setting where the target domain is unavailable, the above DA methods cannot be directly applied to address the DG problem.

For this reason, some researchers [5, 19, 26] start to explore how to adapt the pre-trained model from its source domain to out-of-distribution domain only using source data. For example, some works [43, 45] try to boost the model generalization ability using mix-up domains, which generates novel data distribution from the mixtures of multi-domains. Besides, self-supervised learning (SSL) [3, 35] also is applied to DG problems to enhance transferable features by leveraging the designed pretext tasks. Although these DG methods have been extensively studied in 2D image tasks, the research on the DG problem in 3D point cloud scenarios still remains under-explored, which motivates us to investigate the zero-shot generalization ability of the existing 3D point cloud models.

### 2.2. 3D Point Cloud Classification

The existing 3D point cloud classification methods can be divided into 1) Projection-based and 2) Point-based methods. The projection-based methods first covert irregular points into structured representations, such as multi-view images [29, 41], voxels [25], and spherical [24]. And then, a 2D or 3D neural network is utilized to extract dense features of the structured representations. In contrast, point-based methods [21, 22, 36] directly learn features from the irregular point clouds. This kind of method can effectively explore the point-wise relations using the designed network such as PointNet [21], which is the first work that directly takes original point clouds as the input and achieves permutation invariance with a symmetric module. Further, considering that point clouds have a variable density at different areas, PointNet++ [22] learns 3D features from multiple semantic levels according to the set abstraction. However, these data-driven point cloud models still face substantial recognition accuracy drops when deployed to an unknown domain.

### 2.3. 3D Domain Adaptation and Generalization

To investigate how to equip a 3D point cloud model with good domain generalization capability, we have reviewed existing domain transfer-based [1, 18, 23, 27, 39]

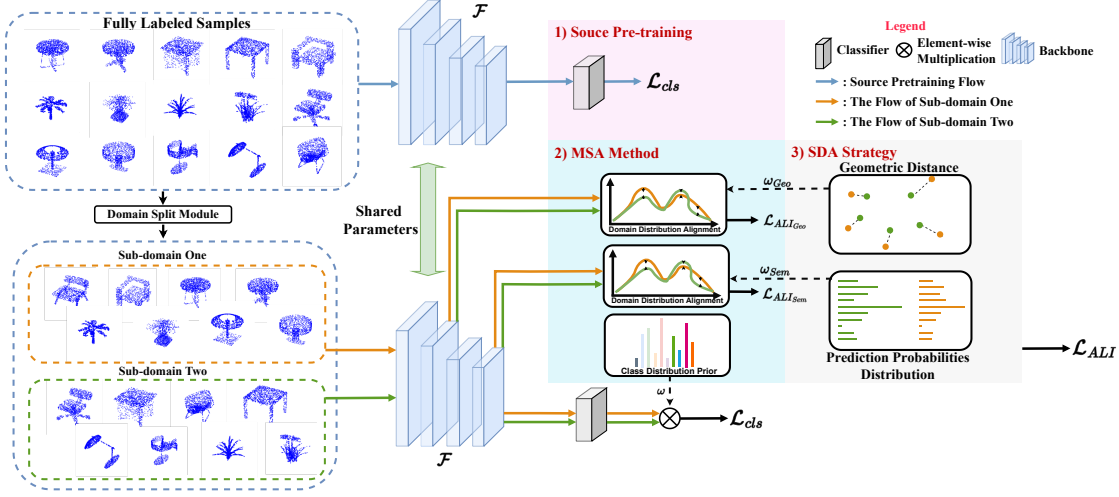


Figure 1. SUG framework, consisting of MSA and SDA to tackle the one-to-many DG problem.

or transfer learning-based 3D point cloud works [40], and find that most of them mainly focus on DA study and fail to generalize to **unseen target domains**. For example, some researchers use self-supervised adaptation methods [1, 18, 27, 39], and design a pretext task to address the common geometric deformations caused by the variances in scanning point clouds. Besides, by deforming a region shape of points and reconstructing the original regions of the shape, DefRec [1] can achieve a good domain adaptation result under different domain shift scenarios. Recently, a geometry-aware DA method [27] is proposed, which employs the underlying geometric information from points. Besides, PDG [37] achieves domain generalization by building a common feature space of part templates and aligning the part-level features. Specifically, PointDAN [23] proposes a Self-Adaptive (SA) node learning with node-level attention to present geometric shape information for points.

However, when performing the domain transfer, the DA methods need to collect extensive target samples in advance to support the adaptation process, which is infeasible for real applications where the target domain is inaccessible or even unknown before deploying the pre-trained model.

### 3. The Proposed Method

The overall SUG framework is illustrated in Fig. 1. For easy understanding, we first give the problem definition of Domain Generalization (DG) for 3D point cloud classification and then introduce the SUG framework, including Multi-grained Sub-domain Alignment (MSA) and Sample-level Domain-aware Attention (SDA) modules. Finally, the overall loss function and DG strategy are described.

#### 3.1. Preliminaries

**Problem Definition.** Suppose that a domain is defined by a joint distribution  $P_{XY}$  on  $\mathcal{X} \times \mathcal{Y}$ , where  $\mathcal{X}$  and  $\mathcal{Y}$  stand for the input data and label space, respectively. In the scope of **DG**,  $K$  source domains  $\mathcal{S} = \{S_k = \{(\mathbf{x}^{(k)}, y^{(k)})\}\}_{k=1}^K$  are available for the training process, where each distinct domain is associated with one distribution  $P_{XY}^k$ . And the goal of DG is to obtain a model  $f: \mathcal{X} \rightarrow \mathcal{Y}$ , trained on the source domain(s), which would have overall minimized prediction errors on the unseen target domain(s).

Point cloud is a set of unordered 3D points  $\mathbf{x} = \{p_i | i = 1, \dots, n\}$ , where each point  $p_i$  is generally represented by its 3D coordinate  $(x_p, y_p, z_p)$  and  $n$  is the number of sampling points of one 3D object. We use  $(\mathbf{x}, y)$  to denote one training sample pair, and  $y$  is its label.

**Single-dataset DG.** In the 3D point-based single-dataset DG setting, the training model *can only access one labeled dataset*  $\mathcal{S}$ , and is required to be evaluated on  $M$  unseen target datasets  $\mathcal{T}$  (usually  $M > 1$ ). The corresponding joint distribution could be described with  $\mathcal{T} = \{T_m = \{(\mathbf{x}^{(m)}, y^{(m)})\}\}_{m=1}^M$ . Also,  $P_{XY}^m \neq P_{XY}^{(k)}, \forall k \in \{1, \dots, K\}, \forall m \in \{1, \dots, M\}$ . In our problem setting,  $\mathcal{Y}_S$  and  $\mathcal{Y}_T$  share the same label space. The goal of 3D DG is to improve the performance of source-trained model  $f$  on the unseen target domain(s) with the following objectives:

$$\min \mathbb{E}_{(\mathbf{x}, y) \in \mathcal{T}} \epsilon(f(\mathbf{x}), y), \quad (1)$$

where  $\epsilon$  is the cross-entropy error in our classification task, which can be further defined as:

$$\mathbb{E}_{\mathcal{T}} [-\log p(\hat{y} = c | \mathbf{x})], \quad (2)$$

where the prediction can be obtained with:

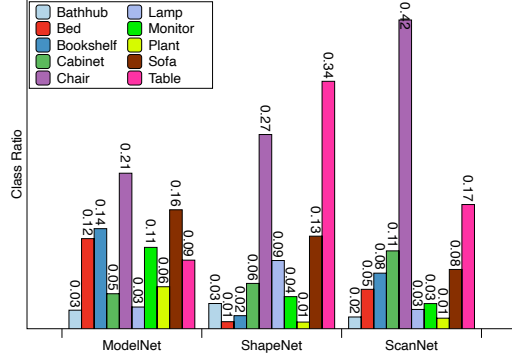


Figure 2. Class distribution shifting across datasets in PointDAN.

$$p(\hat{y} = c | x) = \text{softmax}(\mathcal{C}_\theta(\mathcal{F}_\phi(\mathbf{x}))), \quad (3)$$

where  $\mathbf{x}$  is the input point cloud instance,  $\hat{y}$  is the predicted label. The  $\mathcal{F}$  is the embedding network parameterized by  $\phi$ , and  $\mathcal{C}$  is the classifier parameterized by  $\theta$ .

### 3.2. SUG: A Single-dataset Unified Generalization Framework

To overcome the two challenges discussed in Sec. 1, we introduce a SUG framework consisting of two novel plug-and-play modules, e.g., **Multi-grained Sub-domain Alignment (MSA)** and **Sample-level Domain-aware Attention (SDA)**, which can be inserted into existing 3D backbones to learn more domain-agnostic representations, to be elaborated in Sec. 3.2.1 and 3.2.2, respectively.

First, the single source dataset is fed into a *designed split module* to get multiple sub-domains of the source dataset based on pre-defined heuristics. Then, the embedding network  $\mathcal{F}$  takes all the split sub-domains as the network input and converts the point cloud instance  $\mathbf{x}$  into multi-level feature vectors  $f_l = \mathcal{F}_{\phi,l}(\mathbf{x})$  and  $f_h = \mathcal{F}_{\phi,h}(f_l)$ ,  $f \in \mathbb{R}^{1 \times d}$  where  $f_l$  and  $f_h$  denote the learned low-level and high-level representations. To handle feature discrepancies from different sub-domains, the MSA module is applied to align the multi-grained features, both low- and high-level, which can constrain the network to focus on the domain-agnostic representations. Meanwhile, the SDA module is used to selectively enhance the alignment constraints rising from the easy-to-transfer samples to ensure an even adaptation across different sub-domains.

#### 3.2.1 Multi-grained Sub-domain Alignment

**Class Distribution Alignment.** The 3D point clouds have been deployed in plenty of application scenarios where the objects' distribution shifts significantly, resulting in different distribution patterns residing in different objects, as shown in Fig. 2. To handle such a cross-dataset class-imbalance issue, we incorporate the class-wise sample

weighting  $\alpha$  with the original classification loss (refer to Eq. 2), and the complete weighted classification loss can be written as follow:

$$\mathcal{L}_{cls}(\mathcal{B}) = - \sum_{\mathbf{x} \in \mathcal{B}} \alpha(y) L(\theta; \mathbf{x}), \quad (4)$$

where  $\mathcal{B}$  denotes a data batch. The weighting vector  $\alpha$  could be set following different heuristics, like FocalLoss [13] and DLSA [38], etc. Here, we follow the definition in DLSA [38], where samples are weighted by:

$$\alpha(i) = \frac{m_i^{-q}}{\sum_j m_j^{-q}}, \quad (5)$$

where  $m_i$  is the number of training samples of the class  $i$ , and  $q$  is a positive number controlling the weight distribution. The optimization objective of previous methods, such as FocalLoss [13] and DLSA [38], is to tackle the class imbalance problem within a single dataset, while the optimization function of our method is to tackle the **cross-dataset** class-wise imbalance issue, which is illustrated in Fig. 2. Note that different 3D datasets present an inconsistent class distribution, which motivates us to use Eq. 5 to learn a uniform and even class distribution by re-weighting class distribution for each dataset. Such a way is beneficial to learn more generalizable representations that can avoid overfitting the class distribution of the source dataset.

**Geometric Shifting Alignment.** Due to the objects' geometric variances in different scenarios and inconsistent data acquisition procedures, the objects from the same class across different datasets present diverse geometric appearances, as illustrated in Fig. 3(a) across different rows. Meanwhile, the geometric appearance of objects from a specific class varies significantly within a single dataset, as illustrated in Fig. 3 across different columns, which offers the potential to use the geometric variances within a single dataset to simulate the ones between different datasets effectively.

To be more specific, we take the low-level feature vector  $f_l$  from the shallow layer of the embedding module  $\mathcal{F}$ , and minimize the alignment loss  $L_{ALI}$  to constraint the geometric features from different sub-domains. We use Maximum Mean Discrepancy (MMD)[2] [17] loss by default:

$$L_{ALI_{Geo}} := L_{MMD_{Geo}} = \frac{1}{n_s n_s} \sum_{i,j=1}^{n_s} \kappa(f_{l_i}^s, f_{l_j}^s) + \frac{1}{n_s n_t} \sum_{i,j=1}^{n_s, n_t} \kappa(f_{l_i}^s, f_{l_j}^t) + \frac{1}{n_t n_t} \sum_{i,j=1}^{n_t} \kappa(f_{l_i}^t, f_{l_j}^t), \quad (6)$$

where  $\kappa$  is the kernel function, and its superscript  $t$  and  $s$  denote two different sub-domains sampled from a single dataset. We use the Radial Basis Function (RBF) kernel in our SUG, consistent with previous work [23].



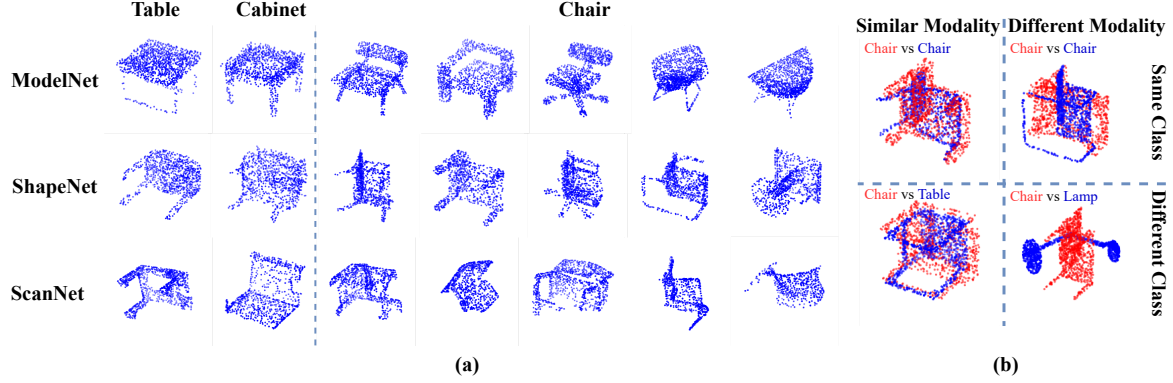


Figure 3. Illustration of distinct characteristics of data in 3D datasets. (a) Geometric and semantic-level domain variances within and between datasets. (b) Geometric similarity comparisons within and between classes.

**Semantic Variance Alignment.** After the high-level features  $f_h$  from  $\mathcal{F}$  are obtained, the semantic variance alignment is applied to minimize the semantic-level discrepancy between features across different sub-domains before feeding into the classifier. The intuition of the semantic alignment arises from the observation that samples from different classes could have similar geometric appearances. As illustrated in Fig. 3(b), the class Table and Cabinet resemble some samples in the Chair class as they are all four-legged items. And by conducting semantic variance alignment, the model will learn less single-domain geometric bias yet discriminative representations. The semantic alignment constraints  $L_{ALI_{sem}}$  can be easily calculated by employing the  $f_{h_j}^t$  and  $f_{h_j}^s$  as the input in Eq. 6.

### 3.2.2 Sample-level Domain-aware Attention

The MSA module, as mentioned above, guides the model to learn more domain-agnostic representations. However, the features inside one mini-batch from different sub-domains do not contribute equally to the sub-domain alignment process since they could contain distinct feature distributions. Ignoring such diversity and imposing equal importance on different samples would result in hard-to-transfer samples deteriorating the generalization procedure. Meanwhile, the designed domain split module in the SUG framework inevitably introduces randomness to different sub-domains with different domain variances, which could also hurt the model generalization performance. Towards safer transfer, we propose the SDA module to enhance the alignment constraints from easy-to-transfer samples. Specifically, we add sample-level weights  $\omega$  to alignment constraints, inversely proportional to the domain distance  $d$ , expressed as:

$$L_{ALI_{weighted}} = \omega * L_{ALI} = \frac{1}{d} * L_{ALI}, \quad (7)$$

where  $d$  could be realized by using either Eq. 8 or Eq. 9. As for the **geometric shifting alignment**, we use the 3D re-

construction metric as the distance function. In our implementation, Chamfer Distance (CD) is used, which can be formulated as follows:

$$d_{CD}(\mathbf{X}, \mathbf{Y}) = \sum_{x \in \mathbf{X}} \min_{y \in \mathbf{Y}} \|x - y\|_2^2 + \sum_{y \in \mathbf{Y}} \min_{x \in \mathbf{X}} \|x - y\|_2^2, \quad (8)$$

where  $\mathbf{X}$  and  $\mathbf{Y}$  are two point cloud instances. The geometric weights  $d_{CD}$  focus on the geometric consistency explicitly, as shown in the first column of Fig. 3(a), where the samples with geometric similarity have relative small CD distance even if they could come from different classes. While for the samples with distinct geometric appearances, the CD distance is higher, and the corresponding alignment constraints would be relaxed.

For the **semantic variance alignment**, we adopt the Jensen Shannon (JS) divergence as our metric. And for symmetric usage, the JS-distance  $d_{JS}(\mathbf{X}, \mathbf{Y})$  is defined as:

$$d_{JS}(\mathbf{X}, \mathbf{Y}) = \frac{1}{2} D_{KL}(\mathbf{X} \parallel \mathbf{Y}) + \frac{1}{2} D_{KL}(\mathbf{Y} \parallel \mathbf{X}), \quad (9)$$

where  $D_{KL}(\mathbf{X} \parallel \mathbf{Y})$  is the discrete format of KL divergence, represented as:

$$D_{KL}(X \parallel Y) = \sum_{c \in \mathcal{C}} X(c) \log \left( \frac{X(c)}{Y(c)} \right), \quad (10)$$

where  $X(c)$  and  $Y(c)$  are the probability of predicting a sample belonging to the class  $c$ . In contrast to the geometric weighting,  $d_{JS}$  emphasizes more semantic consistency and tends to conduct the alignments among samples of the same class.

### 3.3. Overall Objectives and Domain Generalization Strategy

**Overall Objectives.** With the alignment constraints introduced in Sec. 3.2.1 and the corresponding weights stated in Sec. 3.2.2, the complete alignment loss could be defined as

$$L_{ALI} = \omega_{Geo} * L_{ALI_{Geo}} + \omega_{Sem} * L_{ALI_{Sem}}. \quad (11)$$

The overall training loss consists of the classification loss as described in Eq. 4 and the alignment loss in Eq. 11, which can be written as follows:

$$L = L_{cls} + \lambda * L_{ALI}, \quad (12)$$

where  $\lambda$  is the weighting factor to balance the classification task and the alignment process.

**Domain Generalization Strategy.** We train our model in an end-to-end manner, and the training procedure consists of two steps. **Step 1:** Firstly, the model is trained using classification loss as defined in Eq. 4, which can ensure that the model learns discriminative features for the subsequent domain transfer. We train the model using the fully labeled dataset. **Step 2:** Secondly, to learn a robust representation that can be generalized to different target datasets, we train the baseline model with the complete loss  $L$  as defined in Eq. 12, aiming to constraint the learned representations to be domain-agnostic and discriminative. In this step, the whole dataset is split into multiple subsets.

## 4. Experiments

### 4.1. Datasets and Implementation Details

**Datasets.** To conduct the experimental evaluation for domain adaptation setting, PointDAN [23] extracts point cloud samples of 10 shared classes from ModelNet40 [34], ShapeNet [4], and ScanNet [6]. We follow the work [23] and select the same datasets to verify the effectiveness of the proposed method. **ModelNet-10** ( $M$ ) contains a total of 4183 training samples and 856 test samples of 10 classes, which are collected using a 3D CAD model. **ShapeNet-10** ( $S$ ) has 17378 frames for training and 2492 frames for testing, and these frames are produced using a 3D CAD model. **ScanNet-10** ( $S^*$ ) includes a total of 7879 samples that are re-scanned from real-world scene.

**Implementation Details.** For our SUG framework, we employ the PointNet [21] and DGCNN [36] as the feature embedding network while the classifier  $C_\theta$  is constructed with a Multi-Layer Perceptron (MLP) using a three-layer fully-connected network, which is consistent with previous UDA works [27, 46]. To further validate our SUG using more advanced backbones, we also conduct experiments with PointTransformer[44], KPConv[31]. We initialize a dataloader for each sampled sub-domain, and two sub-domains are sampled by default. The sample weighting control  $q$  in Eq. 5 and the hyper-parameters of  $\lambda$  are set to be 0.2 and 0.5, respectively. During the training phase, we use the common data augmentations described in the work [23]. The Adam optimizer [12] is utilized using an initial learning rate of 0.001 and 0.0001, weight decay of 0.00005 and 0.0001 for DGCNN and PointNet backbones,

Table 1. Results on PointDAN plugged with domain split module under the **one-to-many** Domain Generalization (DG) setting. **Avg** denotes the mean adaptation accuracy across all target domains. Using the naive domain split module can boost the zero-shot performance compared with Source-only model.

Method	Setting	Backbone	$M$ as Source Domain		Avg.
			$M \rightarrow S$	$M \rightarrow S^*$	
w/o Adapt	Source-Only	PointNet	42.5	22.3	32.4
<b>w/Splitting</b>	DG	PointNet	54.5	36.3	45.4
<b>w/Splitting</b>		DGCNN	80.8	53.2	67.0

respectively. When testing the generalization performance, to make a fair comparison with the works [27, 46], we align each object along x and y axes for the DGCNN backbone, and no alignment procedure is applied for experiments on the PointNet backbone. During the DG adaptation process, we mainly judge whether the model adaptation state reaches optimal by the designed cross-domain alignment loss. The adaptation process ends when the change of the alignment loss tends to be stable and has less fluctuation. We report the mean value over the three runs for all our experiments. And we use  $\mathcal{F}$ 's third layer and  $C_\theta$ 's second layer as the low- and high-level features, respectively.

### 4.2. How to Split: Domain Split Module Descriptions

This section describes the prior knowledge-based domain split module, which is the prerequisite to enable a given UDA framework to tackle the DG problem. Note that our splitting procedure is conducted class-wise within a source dataset to ensure that each sub-domain contains all dataset categories. According to the prior knowledge source, that module could be *Random Splitting*, *Geometric Splitting*, *Feature Clustering Splitting*, etc.

In our SUG implementation, we use *Random Splitting* as the default domain split module. Specifically, we conduct random sampling and split a single source dataset into different sub-domains with the same sample size, where the domain characteristic of each sub-domain is identical to that of the original one. Please refer to Appendix A for more discussions about the hand-designed domain split module.

### 4.3. How to Align: DG Baseline Implementation

In this part, we study how to use the off-the-shelf UDA technique to achieve unseen domain generalization. First, we use the domain split module to generate different sub-domain data. Then, when a source dataset is clustered into  $K$  sub-domains, 3D UDA methods such as PointDAN [23] can be used to perform a sub-domain adaptation within a single dataset. In our baseline practice, we directly utilize the implementation from PointDAN [23] without any further modification to align the feature gaps between different sub-domains. It can be seen from Table 1 that, by leveraging

Table 2. Results on PointDA-10 under the **one-to-many** Domain Generalization (DG) setting. **Note that** our SUG can be *simultaneously* generalized to multiple target domains without accessing any target samples. In contrast, UDA methods can only be adapted to a single target domain. For example, the GAST model adapts from the domain  $M$  to another domain  $S$ , but the adapted model cannot perform well in a new domain  $S^*$ .

Method	Setting	Backbone	$M$ as Source Domain			$S$ as Source Domain			$S^*$ as Source Domain			Avg.
			$M \rightarrow S$	$M \rightarrow S^*$	Avg.	$S \rightarrow M$	$S \rightarrow S^*$	Avg.	$S^* \rightarrow M$	$S^* \rightarrow S$	Avg.	
w/o Adapt	Source-Only	PointNet DGCNN	42.5	22.3	32.4	39.9	23.5	31.7	34.2	46.9	40.6	34.8
			83.3	43.8	63.6	75.5	42.5	59.0	63.8	64.2	64.0	62.2
PointDAN (NeurIPS'19)	UDA	PointNet	64.2	33.0	48.6	47.6	33.9	40.8	49.1	64.1	56.6	48.7
GAST (ICCV'21)	UDA	DGCNN	84.8	59.8	72.3	80.8	56.7	68.8	81.1	74.9	78.0	73.0
SLT (CVPR'22)	UDA	DGCNN	86.2	58.6	72.4	81.4	56.9	69.2	81.5	74.4	77.9	73.2
PDG (NeurIPS'22)	DG	DGCNN	85.6	57.9	71.8	73.1	50.0	61.6	70.3	66.3	68.3	67.2
<b>our SUG</b>	DG	PointNet DGCNN	64.3	40.7	52.5	44.0	36.2	40.1	44.5	54.7	49.6	47.4
			82.8	57.2	70.0	74.8	52.2	63.5	73.1	69.5	71.3	68.3

Table 3. Results on down-sampling the whole source dataset using different methods where the model is trained on ModelNet-10. PointNet is used as the backbone.

Down-sampling Methods	Diversity	Size	$M \rightarrow S$	$M \rightarrow S^*$	Avg.
Full Dataset	High	4183	64.3	40.7	52.5
Split & Select A	Low	1015	48.9	33.6	41.3
Split & Select B	Low	975	53.7	45.0	49.4
Random Sampling	High	1044	55.4	45.2	50.3

the above domain split modules to split a single dataset into different sub-domains, the baseline model [23] can simultaneously boost the model generalization ability for multiple unseen datasets. It also can be concluded that **multi-modal distribution exists within a single-source dataset**. As a result, a hand-designed domain split method coupled with an off-the-shelf UDA baseline can enhance unseen domain generalization.

Besides, we also observe that the classification accuracy of the model in the target domain is related to the selected network structure. This is intuitive since different network structures have different capacities to learn features with different sensitivities to the source-to-target feature variations.

#### 4.4. Experimental Results using SUG

**SUG Implementation.** Although a naive UDA baseline coupled with our designed domain split modules can enhance the model’s zero-shot recognition ability, it is still important for one-to-many adaptation to exploit multi-modal feature variations across different sub-domains and further learn as many domain variances as possible. Table 2 shows the experiments using the designed MSA and SDA, where *Random Splitting* is applied to obtain sub-domains, and MMD is used for the alignment constraint by default. First, our results show that the state-of-the-art 3D-based UDA methods [27, 46] cannot work well under the one-to-many generalization scenario. For example, GAST [46] can obtain a relatively high result (84.8%) under the  $M \rightarrow S$  setting. Still, the adapted model has a severe accuracy drop

under another target domain  $M \rightarrow S^*$ . This is mainly because these methods often try to perform the explicit cross-domain alignment between the source domain and a specific target domain, which is hard to ensure that the adapted model has an even generalization toward different domains. In contrast, our SUG achieves higher one-to-many zero-shot generalization results for different target domains (*e.g.* 82.8% for  $S$  and 57.2% for  $S^*$ ).

**SUG Limitation.** Our SUG framework assumes that the source domain dataset presents multi-modal feature distributions, which can be implicitly exploited to model the feature distribution differences in the multi-modal distributions. In the 3D scenario, our assumption holds since the 3D point cloud samples for each class often have diverse appearances and geometric shapes, as shown in Fig. 3. Here, we further discuss the limitation cases of our SUG from: **the diversity of source domain distribution gradually decreases**.

To this end, we first split the given single dataset into  $M$  sub-domains and then select one of the sub-domains from the splitting results (1 out of 4 splits) as the training set, which is described in Sec. 4.2 and denoted as **Split & Select**. Specifically, **Split & Select A** is obtained with *Feature Clustering Splitting* while **Split & Select B** is with *Geometric Splitting*. For comparison, we also randomly sample from the complete set of the given single dataset, denoted as **Random Sampling**. The most significant difference between the above down-sampling methods is that, a single split (sub-domain) has much fewer data diversity and domain variances inside than the randomly sampled one with a distribution status similar to the original dataset. We observe from Table 3, that when the data distribution within the sampled sub-domain becomes more undiversified, the zero-shot generalization ability of the model from a source domain to multi-target domains will drop. Moreover, by comparing the first row with the fourth row, it can be concluded that the training sample size also matters for the DG performance.

Table 4. Ablation studies of class-wise classification accuracy, where the model is trained on ModelNet-10 and directly evaluated on ScanNet-10 ( $M \rightarrow S^*$ ). PointNet is used as the backbone. Avg. is the per-class average result.

Methods	CD Align.	GS Align.	SV Align.	SDA	Bathtub	Bed	Bookshelf	Cabinet	Chair	Lamp	Monitor	Plant	Sofa	Table	Avg.
Supervised					88.9	88.6	47.8	88.0	96.6	90.9	93.7	57.1	92.7	91.1	83.5
w/o Adapt					59.4	1.0	18.4	7.4	55.7	43.5	84.8	60.0	3.4	39.7	37.3
PointDAN					84.7	1.6	19.0	1.3	81.9	63.3	90.5	82.3	2.2	82.9	51.0
our SUG	✓		✓	✓	68.1	2.6	20	0.0	49.0	53.9	95.3	86.4	0.3	79.5	45.5
	✓	✓		✓	64.5	5.6	17.0	0.7	75.2	61.4	90.5	78.6	0.2	88.4	48.2
	✓	✓	✓	✓	64.1	0.0	17.8	1.43	75.9	55.7	92.0	90.0	0.0	85.0	48.2
	✓	✓	✓		65.9	5.0	34.0	4.0	74.1	58.4	91.7	77.3	0.0	86.7	49.7
	✓			✓	80.9	0.0	19.1	0.0	73.3	63.8	93.7	72.5	0.0	80.9	48.4
	✓	✓	✓	✓	76.9	2.0	25.0	2.0	81.5	57.6	89.7	88.2	0.4	85.0	<b>50.8</b>

Table 5. Results on PointDA-10 under the **one-to-many** Domain Generalization (DG) setting with additional backbones *e.g.* KP-Conv(KP) and Point Transformer (PT).

Method	Backbone	$M$ as Source Domain		Avg.
		$M \rightarrow S$	$M \rightarrow S^*$	
w/o Adapt	KP	81.8	46.1	63.9
	PT	84.1	54.8	69.5
our SUG	KP	81.1	47.7	<b>64.4</b>
	PT	83.4	58.4	<b>70.9</b>

#### 4.5. Further Analyses

The proposed SUG is a unified framework where each designed module can be extended with other advanced designs.

**Additional Backbones.** To further validate that SUG can be generalized to different point cloud backbones, we select another two state-of-the-art 3D point-cloud backbone networks, *e.g.*, Point Transformer [44], and KPConv [31] to conduct the one-to-many DG experiments. It can be seen from Table 5 that by coupling with our method, both networks can achieve a better DG classification performance gain. For more implementation details and deeper analyses, please refer to Appendix C. And it can be concluded that the SUG can be well extended with different feature backbones.

**Additional Alignment Constraints.** Contrastive Loss (CL) is also known for its capability of constraining learned features. We conduct experiments to compare the performance between CL and MMD loss used in the SUG framework. Specifically, for the implementation of Contrastive Loss, we directly use the PyTorch Implementation [20], which is a variation of the [9] with cosine distance. As shown in Table 6, SUG with CL or MMD alignment can boost the DG performance. For more implementation details and analyses, please refer to Appendix E.

**Ablation Studies.** In Table 4, we conduct the ablation studies from the following two aspects: 1) MSA method that consists of Class Distribution (CD Align.), Geometric Shifting (GS Align.), and Semantic Variance (SV Align.) alignments; 2) SDA strategy. First, MSA learns the domain-agnostic features from various granularities, in-

Table 6. Average results of unseen domains  $S$  and  $S^*$  using the CL and MMD alignment designs, and we employ the PointNet as the backbone. G and S stand for geometric and semantic alignments, respectively. M and C stand for MMD and CL constraints, respectively.

Alignment	G-M	G-C	S-M	S-C	Avg.
Pure MMD	✓		✓		52.5
Pure CL		✓		✓	46.0
MIX	✓			✓	52.3
MIX		✓	✓		47.3

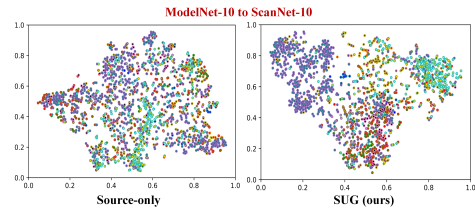


Figure 4. tSNE results (PointNet). Different colors denote different classes. For more visualization results, please see the figures in the Appendix.

cluding class, geometry, and semantic levels. Table 4 shows that each newly-added alignment constraint can bring accuracy gains. Besides, we also conducted experiments on removing the SDA to investigate the effectiveness of the designed SDA. The results shown in Table 4 demonstrate that, by enhancing some easy-to-adapt instances to keep an even adaptation, SDA significantly boosts the generalization accuracy from 48.4% to 50.8%.

**Hyper-parameters Analyses.** Further quantitative analyses on the hyper-parameters setting, including alignment layer selection, batch size, and weight factor  $\lambda$ , can be found in Appendix F.

**tSNE Results.** We visualize features from the source-only model and our SUG in Fig. 4. The visualizations show that features learned by SUG can improve the model discriminability of different classes' features from unseen domains.

## 5. Conclusion

We proposed a SUG framework to study the one-to-many DG in 3D scenarios. SUG consists of an MSA method to exploit the data diversity residing in a given



source dataset and further learn domain-agnostic and discriminative representations, an SDA strategy to selectively increase the domain adaptation degree for easy-to-adapt instances. Equipped with the SUG, the existing 3D baseline models can perform a domain generalization process well and recognize many unseen classes and instances. Extensive experiments verify that the SUG framework is general and effective in tackling the 3D DG problem.

## References

- [1] Idan Achituve, Haggai Maron, and Gal Chechik. Self-supervised learning for domain adaptation on point clouds. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 123–133, 2021. 2, 3
- [2] Karsten M Borgwardt, Arthur Gretton, Malte J Rasch, Hans-Peter Kriegel, Bernhard Schölkopf, and Alex J Smola. Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics*, 22(14):e49–e57, 2006. 4
- [3] Fabio M Carlucci, Antonio D’Innocente, Silvia Bucci, Barbara Caputo, and Tatiana Tommasi. Domain generalization by solving jigsaw puzzles. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2229–2238, 2019. 2
- [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. 1, 6
- [5] Wuyang Chen, Zhiding Yu, Shalini De Mello, Sifei Liu, Jose M Alvarez, Zhangyang Wang, and Anima Anandkumar. Contrastive syn-to-real generalization. In *ICLR*, 2021. 2
- [6] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5828–5839, 2017. 1, 6
- [7] Hehe Fan, Xiaojun Chang, Wanyue Zhang, Yi Cheng, Ying Sun, and Mohan Kankanhalli. Self-supervised global-local structure modeling for point cloud domain adaptation with reliable voted pseudo labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6377–6386, 2022. 1
- [8] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International Conference on Machine Learning*, 2014. 2
- [9] Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*, volume 2, pages 1735–1742. IEEE, 2006. 8
- [10] Guoliang Kang, Lu Jiang, Yunchao Wei, Yi Yang, and Alexander G Hauptmann. Contrastive adaptation network for single-and multi-source domain adaptation. *IEEE transactions on pattern analysis and machine intelligence*, 2020. 2
- [11] Nitish Shirish Keskar, Dheevatsa Mudigere, Jorge Nocedal, Mikhail Smelyanskiy, and Ping Tak Peter Tang. On large-batch training for deep learning: Generalization gap and sharp minima. *arXiv preprint arXiv:1609.04836*, 2016. 14
- [12] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014. 6
- [13] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017. 4
- [14] Mingsheng Long, Yue Cao, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Transferable representation learning with deep adaptation networks. *IEEE transactions on pattern analysis and machine intelligence*, 41(12):3071–3085, 2018. 2
- [15] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael I Jordan. Learning transferable features with deep adaptation networks. In *International Conference on Machine Learning*, 2015. 2
- [16] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. In *Advances in Neural Information Processing Systems*, pages 1640–1650, 2018. 2
- [17] Mingsheng Long, Jianmin Wang, Guiguang Ding, Jianguang Sun, and Philip S Yu. Transfer feature learning with joint distribution adaptation. In *Proceedings of the IEEE international conference on computer vision*, pages 2200–2207, 2013. 4
- [18] Xiaoyuan Luo, Shaolei Liu, Kexue Fu, Manning Wang, and Zhijian Song. A learnable self-supervised task for unsupervised domain adaptation on point clouds. *arXiv preprint arXiv:2104.05164*, 2021. 2, 3
- [19] Vihari Piratla, Praneeth Netrapalli, and Sunita Sarawagi. Efficient domain generalization via common-specific low-rank decomposition. In *International Conference on Machine Learning*, pages 7728–7738. PMLR, 2020. 2
- [20] Pytorch. Pytorch cosine embedding loss, 2022. 8
- [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. 2, 6, 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. *Advances in neural information processing systems*, 30, 2017. 2
- [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. *Advances in Neural Information Processing Systems*, 32, 2019. 2, 3, 4, 6, 7, 11
- [24] Yongming Rao, Jiwen Lu, and Jie Zhou. Spherical fractal convolutional neural networks for point cloud recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 452–460, 2019. 2
- [25] Gernot Riegler, Ali Osman Ulusoy, and Andreas Geiger. Octnet: Learning deep 3d representations at high resolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3577–3586, 2017. 2
- [26] Shiv Shankar, Vihari Piratla, Soumen Chakrabarti, Siddhartha Chaudhuri, Preethi Jyothi, and Sunita Sarawagi. Generalizing across domains via cross-gradient training. In

- ICLR*, 2018. 2
- [27] Yuefan Shen, Yanchao Yang, Mi Yan, He Wang, Youyi Zheng, and Leonidas J Guibas. Domain adaptation on point clouds via geometry-aware implicits. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7223–7232, 2022. 1, 2, 3, 6, 7
- [28] Shaoshuai Shi, Chaoxu Guo, Li Jiang, Zhe Wang, Jianping Shi, Xiaogang Wang, and Hongsheng Li. Pv-rcnn: Point-voxel feature set abstraction for 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10529–10538, 2020. 1
- [29] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 945–953, 2015. 2
- [30] Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *European conference on computer vision*, pages 443–450. Springer, 2016. 2
- [31] Hugues Thomas, Charles R Qi, Jean-Emmanuel Deschaud, Beatriz Marcotequi, François Goulette, and Leonidas J Guibas. Kpconv: Flexible and deformable convolution for point clouds. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6411–6420, 2019. 6, 8, 12
- [32] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7167–7176, 2017. 2
- [33] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008. 11
- [34] Kashi Venkatesh Vishwanath, Diwaker Gupta, Amin Vahdat, and Ken Yocum. Modelnet: Towards a datacenter emulation environment. In *2009 IEEE Ninth International Conference on Peer-to-Peer Computing*, pages 81–82. IEEE, 2009. 1, 6
- [35] Shujun Wang, Lequan Yu, Caizi Li, Chi-Wing Fu, and Pheng-Ann Heng. Learning from extrinsic and intrinsic supervisions for domain generalization. In *European Conference on Computer Vision*, pages 159–176. Springer, 2020. 2
- [36] 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):1–12, 2019. 2, 6, 12
- [37] Xin Wei, Xiang Gu, and Jian Sun. Learning generalizable part-based feature representation for 3d point clouds. In *Advances in Neural Information Processing Systems*. 3
- [38] Yue Xu, Yong-Lu Li, Jiefeng Li, and Cewu Lu. Constructing balance from imbalance for long-tailed image recognition. *arXiv preprint arXiv:2208.02567*, 2022. 4
- [39] Jihan Yang, Shaoshuai Shi, Zhe Wang, Hongsheng Li, and Xiaojuan Qi. St3d: Self-training for unsupervised domain adaptation on 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10368–10378, 2021. 1, 2, 3
- [40] Chuanguan Ye, Hongyuan Zhu, Yongbin Liao, Yanggang Zhang, Tao Chen, and Jiayuan Fan. What makes for effective few-shot point cloud classification? In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1829–1838, 2022. 3
- [41] Tan Yu, Jingjing Meng, and Junsong Yuan. Multi-view harmonized bilinear network for 3d object recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 186–194, 2018. 2
- [42] Chongzhi Zhang, Mingyuan Zhang, Shanghang Zhang, Daisheng Jin, Qiang Zhou, Zhongang Cai, Haiyu Zhao, Xi-anlong Liu, and Ziwei Liu. Delving deep into the generalization of vision transformers under distribution shifts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7277–7286, 2022. 12
- [43] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017. 2
- [44] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip HS Torr, and Vladlen Koltun. Point transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16259–16268, 2021. 6, 8, 12
- [45] Wei Zhu, Le Lu, Jing Xiao, Mei Han, Jiebo Luo, and Adam P Harrison. Localized adversarial domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7108–7118, 2022. 2
- [46] Longkun Zou, Hui Tang, Ke Chen, and Kui Jia. Geometry-aware self-training for unsupervised domain adaptation on object point clouds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6403–6412, 2021. 1, 6, 7

Table 7. Results on different domain split methods under the **one-to-many** Domain Generalization (DG) setting. **Avg** denotes the mean adaptation accuracy across all target domains. The results of Random Splitting are averaged over three runs, and we report the mean values over the three runs.

Domain Split Method	Setting	Backbone	M as Source Domain			S as Source Domain			S* as Source Domain		
			$M \rightarrow S$	$M \rightarrow S^*$	Avg.	$S \rightarrow M$	$S \rightarrow S^*$	Avg.	$S^* \rightarrow M$	$S^* \rightarrow S$	Avg.
w/o Adapt	Source-Only	PointNet	42.5	22.3	32.4	39.9	23.5	31.7	34.2	46.9	40.6
PointDAN (NeurIPS'19)	UDA	PointNet	64.2	33.0	48.6	47.6	33.9	40.8	49.1	64.1	56.6
<b>Random Splitting</b>	DG	PointNet	54.5	36.3	45.4	37.8	31.7	34.8	45.0	53.0	<b>49.0</b>
<b>Geometric Splitting</b>			57.4	41.7	<b>49.6</b>	30.3	31.6	31.0	38.3	44.2	41.3
<b>Entropy Splitting</b>			55.4	42.5	49.0	36.5	27.7	32.1	41.7	49.9	45.8
<b>Feature Clustering Splitting</b>			60.4	36.1	48.3	45.4	31.7	<b>38.6</b>	37.6	45.6	41.6
<b>Random Splitting</b>	DG	DGCNN	80.8	53.2	<b>67.0</b>	69.4	49.5	59.5	61.4	57.6	59.5
<b>Geometric Splitting</b>			79.3	49.9	64.6	56.7	53.2	55.0	40.5	64.4	52.5
<b>Entropy Splitting</b>			73.6	49.3	61.5	72.8	50.3	<b>61.6</b>	42.9	60.9	51.9
<b>Feature Clustering Splitting</b>			77.8	52.9	65.4	71.0	47.6	59.3	63.0	59.3	<b>61.2</b>

## A. Discussion of the Hand-designed Domain Split Module

This section gives more details regarding the prior knowledge-based domain split module. We implement different splitting methods for the domain split module, and the potentials for other advanced splitting designs exist.

**Random Splitting.** We conduct random sampling and split a single source dataset into different sub-domains with the same sample size, where the domain characteristic of each sub-domain is identical to that of the original one.

**Geometric Splitting.** In our practice, we randomly select one sample as the anchor sample of a specific class, then compute the geometrical distance between other samples of the same class and the selected anchor sample. After getting all registration scores of all samples, the current class is clustered into  $K$  sub-domains according to the calculated score. We use Iterative Closest Point (ICP) registration score by default. And we discuss the choice of geometric metric with ICP against Chamfer Distance (CD) in Appendix B.

**Entropy Splitting.** We quantify the uncertainty of the classifier's predictions with the entropy criterion  $H(\mathbf{g}) = -\sum_{c \in \mathcal{C}} X(c) \log X(c)$ . Note that the classifier used here is a pre-trained model on the source domain. For the domain split module, the single dataset is clustered into  $K$  sub-domains based on the entropy scores of all samples.

**Feature Clustering Splitting.** We infer the whole dataset with the pre-trained model on the source dataset and save the feature maps before feeding them to the classifier. After that, we use Principal Component Analysis (PCA) with t-SNE [33] to get the dimensional reduced representations and apply  $K$ -means to get  $K$  sub-domain clusters.

We conduct extensive experiments to show the Domain Generalization (DG) results for different domain split module choices. Specifically, the domain split module is directly plugged with the classic UDA framework PointDAN [23],

where that split module splits the source dataset into two subsets, each taken as the source and target domain by PointDAN, respectively. The experiments result is shown in Table 7.

However, we observe that these DG results achieved by the domain split module are unstable for different cross-domain settings. Here, we give **two main reasons** for such instability as follows.

1) The distribution shift patterns across datasets are quite different. ModelNet-10 and ShapeNet-10 are both CAD-generated datasets. As a result, they contain similar geometric characteristics; at least both are without occlusions and follow a similar appearance. In this way, using *Feature Clustering* to emphasize the semantic discrepancy and alignment would bring more gains (As reported in Table 7 for  $M \rightarrow S$  and  $S \rightarrow M$  experiments). In contrast, ScanNet-10 is obtained from the real world and was initially designed for segmentation tasks. In other words, it is pretty different in semantic and geometric views, as shown in Fig. 3 in the main text. In such a situation, emphasizing solely geometric or semantic discrepancy is not optimal. At the same time, random splitting is a strong baseline to conduct the alignment operation. This phenomenon is consistent with our experiment results in  $S^* \rightarrow M$ ,  $S^* \rightarrow S$ , and  $S \rightarrow S^*$  cross-domain settings.

2) The split results achieved by the different domain splitting module choices are quite imbalanced along the sample size. Take *Entropy Clustering* as an example. Since the pre-trained model with source domain-related distribution characteristics is used, the model will be pretty confident in predicting the source-domain samples, resulting in a quite imbalanced clustering result. For example, the PointNet backbone on the ScanNet-10 dataset will get 3504 samples vs. 2606 samples for each sub-domain. But it will get worse on the easier dataset like ModelNet-10, where 2542 samples vs. 1641 samples for each sub-domain. Such imbalance is harmful to the model training since that will bring

the bias from the source dataset characteristic to the training procedure.

Table 8. The number of parameters of the backbone networks employed by SUG.

Network	Parameters
PointNet [21]	3.5M
DGCNN [36]	1.8M
KPConv [31]	5.3M
Point Transformer [44]	9.6M

## B. Discussion on Metrics for Geometric Splitting

Chamfer Distance (CD) and ICP are the most widely used metrics for comparing the similarity of two point clouds. To further validate how the choices of metrics would affect the geometric splitting and final DG performance, we conduct a comparison for ICP and CD-based geometric splitting experiments. We summarize the following empirical findings based on the experimental results in Table 9.

1) The CD could also be one comparable metric for the geometric splitting procedure. Moreover, the final DG average performance is slightly worse than the ICP-based methods.

2) Both ICP and CD metrics could describe the geometric similarity between two point clouds. However, since ICP is in the iterative manner, where the Rotation Matrix and translation vector are optimized during the registration process, in such a way, the ICP score could take the view differences (while with similar geometric appearances) into consideration. Moreover, that situation (similar appearances, different views) is quite common in the point cloud datasets.

3) Since ICP is conducted sequentially and iteratively and is hard to be optimized for parallel computing and thus takes much more time than Chamfer Distance, the CD would be a good choice for a large-scale dataset.

## C. Combination with Point Transformer and KPConv

To further verify the superiority of our SUG in boosting baseline models, we select another two state-of-the-art 3D point-cloud backbone networks, *e.g.*, Point Transformers [44] and KPConv [31] to conduct the one-to-many DG study. The corresponding experimental results are shown in Table 5. And we summarize the following two main empirical findings.

1) By coupling with our method, the Point Transformer [44] can achieve a better one-to-many DG classification performance gain, such average 1.4% for  $M \rightarrow S$ ,  $M \rightarrow$

$S^*$  settings. But it should be pointed out that the accuracy gain of Point Transformer is relatively slight compared with that of the DGCNN backbone. This is mainly because the transformer-based methods could learn many discriminative features during the model training phase, consistent with the observations in [42]. However, Point Transformer takes much more time in model training and hyper-parameter tuning since it is a much heavier network as reported in Table 8.

2) The DG classification performance gain on KPConv [31] is relatively minor since the dataset-related parameter settings, like query radius, are sensitive to different target domains. Besides, we observe that, during the inference process, the points selected by the kernel of KPConv for ModelNet-10 are generally more than 100 points (the first layer) but less than 80 points if we did not change the parameters when used for ScanNet-10. The cross-domain feature alignment process brings more negative effects toward source-similar ModelNet-10 than positive gains toward source-dissimilar ScanNet-10, which results in a lower average classification accuracy across different datasets.

## D. Qualitative analyses

**More tSNE results between source-only model and our SUG.** The main text shows the tSNE visualization results of high-level features learned by the source-only model and our SUG, respectively. In this part, we give more tSNE visualization results for cross-domain settings such as the adaptation from ShapeNet-10 to ModelNet-10, ShapeNet-10 to ScanNet-10, *etc.* As illustrated in Fig. 5 to Fig. 7, these visualization results demonstrate that the features from an unseen target domain (*e.g.*, ModelNet-10) have distinct feature discrimination for different classes, further verifying that the learned features are domain-agnostic and discriminative for unseen domains.

**More tSNE results of the domain alignment process.** We split the training dataset (source domain) into two sub-domains. And then use the model without alignment process and the proposed SUG to train on those two sub-domains. After the training, we visualize and compare the extracted features of those two models by t-SNE. The visualization results are shown in Fig. 8.

**More tSNE results of sub-domains characteristics using the random splitting module.** Moreover, to validate the consistency of the distribution from sub-domains characteristics with the Random Splitting module, we split a single source dataset into different sub-domains using the random sampling strategy. Then we use the pre-trained model to extract features from each sub-domain, and tSNE is applied to compare the features. The visualization results are shown in Fig. 9.



Table 9. Comparison between ICP Score and Chamfer Distance used for dataset splitting.

Split Method	Metric	Setting	Backbone	$M$ as Source Domain		
				$M \rightarrow S$	$M \rightarrow S^*$	Avg.
Geometric Splitting	ICP Score	DG	PointNet	57.4	41.7	49.6
	Chamfer Distance			55.0	41.4	48.2

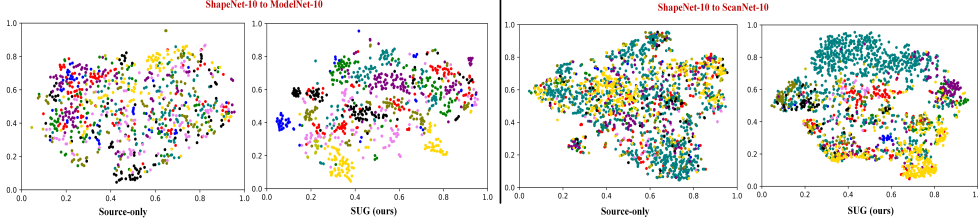


Figure 5. tSNE results of ModelNet-10-10 and ScanNet-10 datasets, where the model is trained on the ShapeNet-10 dataset, and different colors denote different classes.

## E. Discussion on the Alignment Constraints

**Additional Alignment Constraints with Contrastive Loss.** In practice, we have to first explicitly define the positive and negative pairs for CL, which is quite complex in our setting. Positive pairs are the samples with similar geometric appearances for geometric alignments, regardless of whether they are from the same class. Meanwhile, the positives are always from the same class for semantic alignments. For simplicity, we directly take the geometric features as negative pairs when they come from different classes under the CL constraint. Experimentally, we use ModelNet-10 as the source domain and evaluate on ShapeNet-10 and ModelNet-10. We report the average results on these two datasets. Note that we have not yet to tune the parameter for CL loss much. Based on the above experimental results, we summarize the following empirical findings.

1) As we can see from Table 6 when we replace the MMD loss with CL loss for semantic-level alignment, the final results are still competitive since both CL and MMD can make learned features to be domain-invariant. However, the results for CL loss for geometric-level alignment are much worse. The main reason behind those accuracy differences is that CL focuses on capturing the high-level feature variances. At the same time, it tends to ignore some low-level information for describing domain shifts.

2) Based on the experiments in Table 6, we are delighted that the SUG has the potential to be a unified framework where the sub-domain alignment module could be replaced using other recently-proposed alignment loss functions such as Contrastive Loss.

**Discussion on the MMD Constraints Design.** Generally speaking, the alignment should be conducted between

Table 10. Average results of unseen domains  $S$  and  $S^*$  among different MMD-based alignment methods, and we employ the  $M$  as the source domain. PointNet is used as the backbone.

MMD Alignment	Avg. Results
Soft-MMD	52.5
Hard-MMD	51.8
Max-Hard-MMD	52.3

different modalities. And as shown in Fig. 3(b) in the main text, similar modalities exist across classes, while different modalities exist within a single class. As a result, we are expected to exploit multiple modality information from both intra- and inter-classes fully and thus do not perform a hard class-wise MMD-based alignment. Specifically, to avoid losing the label information, we first turn the class label into a scaled one-hot vector and then concatenate it with the feature maps before conducting the MMD alignment, termed the *Soft-MMD*.

Besides, we have implemented different MMD-based alignment methods by changing the class-label information constraint, such as *Hard-MMD*, which means that only samples from the same class are aligned, and *Max-Hard MMD*, which means that we first re-order the samples from different domains to let them have most class overlapping, and then conduct the Hard-MMD. Our experiments showed that Soft-MMD outperforms other MMD-based alignment designs, as shown in Table 10.

## F. Discussion on Hyper-parameters in SUG

The experiments in this part employ PointNet as the backbone and ModelNet-10 as the source domain. We

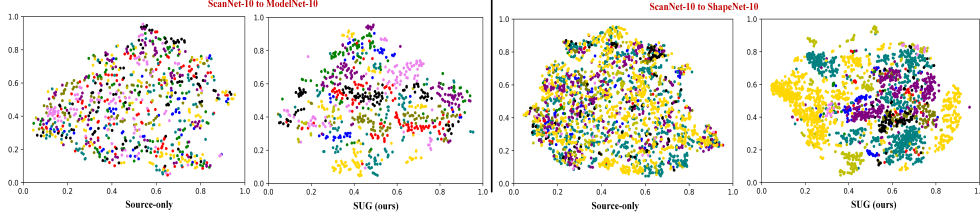


Figure 6. tSNE results of ModelNet-10 and ShapeNet-10 datasets, where the model is trained on the ScanNet-10 dataset, and different colors denote different classes.

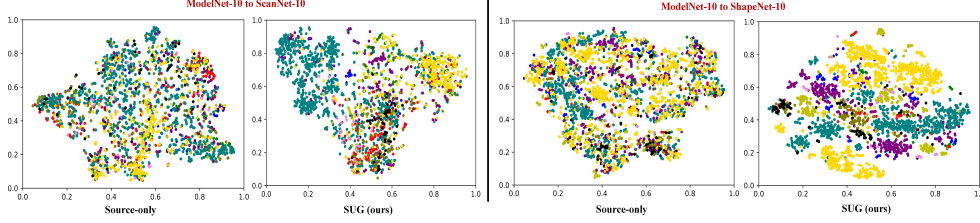


Figure 7. tSNE results of ScanNet-10 and ShapeNet-10 datasets, where the model is trained on the ModelNet-10 dataset, and different colors denote different classes.

Table 11. Average results of unseen domains  $S$  and  $S^*$  trained with different layer selection settings, and we employ the  $M$  as the source domain. PointNet is used as the backbone. **D** denotes the default choice used in SUG.

Embedding $\mathcal{F}$	Avg.	Header $\mathcal{C}_\theta$	Avg.
Layer-1	48.8	Layer-1	49.9
Layer-2	49.7	Layer-2( <b>D</b> )	52.5
Layer-3( <b>D</b> )	52.5	Layer-3	48.2
Layer-4	49.4	-	-
Layer-5	48.2	-	-

report the average prediction results on ShapeNet-10 and ScanNet-10.

**Layer Selection for Low-level and High-level Features.** In the default SUG setting, we use the features from  $\mathcal{F}$ ’s third layer and  $\mathcal{C}_\theta$ ’s second layer as the low and high-level features, respectively. To further explore how the layer selection for features would affect the SUG performance, we change the selection choices of the layers. Specifically, to validate the choice for geometric features, we use the features from  $\{1, 2, 3, 4, 5\}$ -layer of the embedding module as the geometric features while keeping the second layer of the classification module as default. For semantic features experiments, we used the features from  $\{1, 2, 3\}$ -layer of the classification module while keeping the third layer of the embedding module as default. Based on the experimental results in Table 11, we summarize the following empirical findings.

1) For Embedding Module Layer Selection: The features from too shallow layers (e.g., Layer-1) contain much less

information and would be sensitive to noise. In contrast, if we choose the features from too deeper layers, the geometric and fine-grained information would be overtaken by the deep semantic information. At the same time, when we choose that deeper features, the geometric alignment would be much similar to semantic alignment and thus lose its discriminability.

2) For Classification Module Layer Selection: The features from the shallow layer (e.g., Layer-1) are similar to the geometric ones and would lose semantic alignment ability. At the same time, the last layer’s features are too high-level and lose a lot of semantic information.

**Weight Factor  $\lambda$ .** The  $\lambda$  in Eq. 12 achieves a trade-off between the classification task and the alignment process. We conduct ablation studies to investigate the sensitivity of the  $\lambda$  value setting on our SUG performance. The corresponding results are shown in Table 12.

**Batch Size  $\mathcal{B}$ .** We conduct experiments to change the batch-size value, where we keep other settings as default. The results are shown in Table 13.

According to the experimental results shown in Table 13, our method can achieve good generalization across different batch-size settings. The mini-batch data could not contain enough information related to the domain distribution for the batch-size setting with a small value. As a result, the SUG could not learn the domain-invariant features well. In contrast, it can be observed that the degradation of generalization’s ability when we continuously enlarge the batch size, mainly because the large-batch training procedure tends to converge to sharp local minimizers [11].

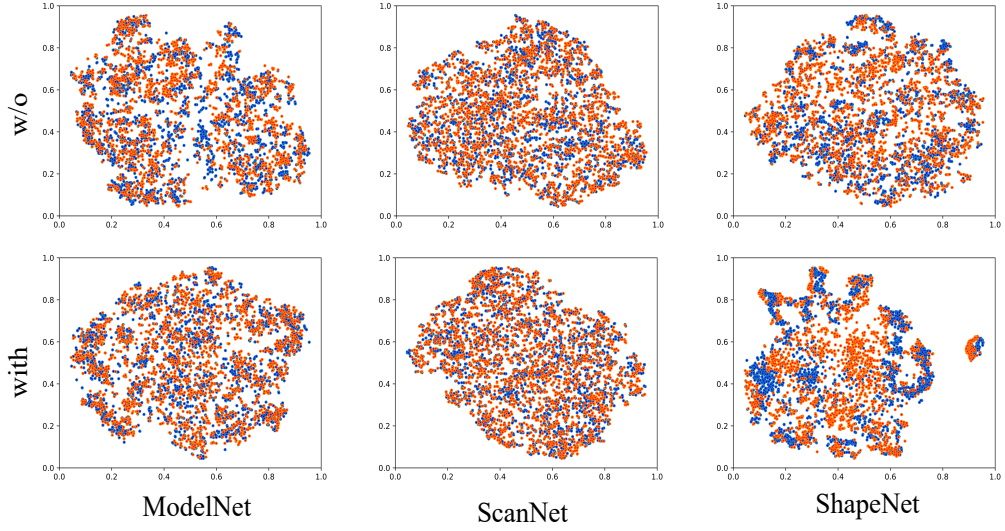


Figure 8. tSNE results of sub-domains without and with alignment module. The first and second rows show the learned features without or with the feature alignment process, respectively. Different colors denote features from different sub-domains.

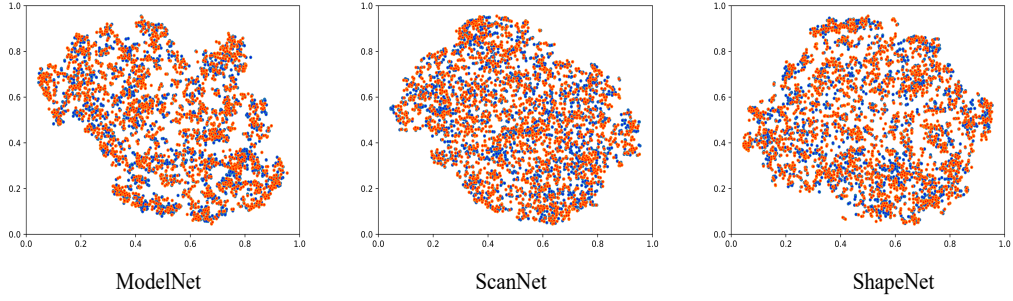


Figure 9. tSNE results of different sub-domains divided by Random Splitting module without using domain alignment. Different colors denote features from different sub-domains.

Table 12. Average results of unseen domains  $S$  and  $S^*$  using different  $\lambda$  values in Eq. 12, and we employ the  $M$  as the source domain. PointNet is the backbone. **D** denotes the default choice used in the SUG.

$\lambda$	Avg. Results
0.25	44.7
0.50( <b>D</b> )	52.5
0.75	51.2
1.0	47.8
2.0	49.1
3.0	48.5
4.0	45.9
5.0	44.7

Table 13. Average results of unseen domains  $S$  and  $S^*$  trained with different batch sizes, and we employ the  $M$  as the source domain.

Batch Size	Avg. Results
16	51.35
32	52.86
64( <b>D</b> )	52.45
128	50.45
256	50.52
512	47.51