

Supplementary material for “Few-shot Class-incremental Learning for 3D Point Cloud Objects”

Townim Chowdhury¹, Ali Cheraghian^{2,3}, Sameera Ramasinghe⁴,
Sahar Ahmadi⁵, Morteza Saberi⁶, and Shafin Rahman^{1*}

¹ Dept. of Electrical and Computer Engineering, North South University, Bangladesh

² School of Engineering, Australian National University, Australia

³ Data61, Commonwealth Scientific and Industrial Research Organisation, Australia

⁴ Australian Institute for Machine Learning, University of Adelaide, Australia

⁵ Business school, The University of New South Wales, Australia

⁶ School of Computer Science and DSI, University of Technology Sydney, Australia

{townim.faisal, shafin.rahman}@northsouth.edu,

ali.cheraghian@anu.edu.au, sameera.ramasinghe@adelaide.edu.au,

sahar.ahmadi@unsw.edu.au, morteza.saberi@uts.edu.au

Abstract. This supplementary material provides additional details in support of the contribution presented in the main paper.

- Section 1: Architecture Details (additional discussion in support of Section 3.1 of the main paper).
- Section 2: K-Means Visualization (additional discussion in support of Section 3.2 of the main paper).
- Section 3: Details of Experimental Setup (additional discussion in support of Section 4 of the main paper).
- Section 4: 3D Object Recognition (additional discussion in support of Section 4.3 of the main paper).

1 Architecture Details

Here, we provide more details on the backbone network, F and the relation module, R .

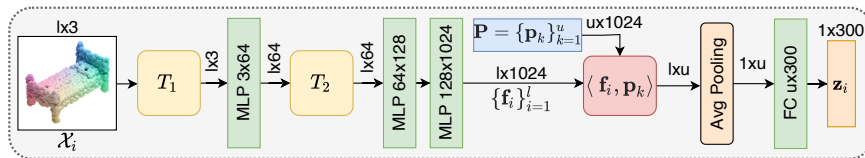


Fig. 1: Detailed network architecture the backbone. T_1 and T_2 are transformation networks for inputs of l points

* Corresponding author

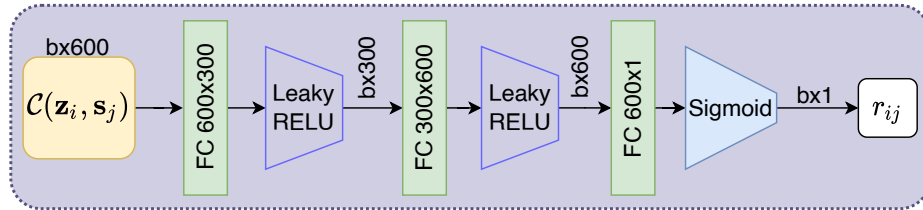


Fig. 2: Detailed architecture for Relation module R . ‘b’ means batch size and r_{ij} is the relation score.



Fig. 3: tSNE visualization for points in different 3D point cloud objects

Backbone network. A mini transformation network T_1 [?] takes raw input point clouds of n points, and the output passes into a shared multi-layer perceptron network of output size 64. This output matrix passes into another feature transformation network T_2 , and the transformed output matrix then passes into a shared multi-layer perceptron network with layer output sizes 64, 128, 1024. This extracted features of n points, \mathbf{f}_i is forwarded with Microshape basis, \mathbf{p}_k in an inner product function, $\langle \mathbf{f}_i, \mathbf{p}_k \rangle$. Then, we calculate the average similarity vector from the output with average pooling. The similarity vector is passed into a fully connected layer (of 300 dimensions), generating the features \mathbf{z}_i . Here, all layers include ReLU and batch normalization.

Relation Module architecture. The input of relation module is generated by a concatenation function $\mathcal{C}(\mathbf{z}_i, \mathbf{s}_j)$, that takes feature of point cloud object, \mathbf{z}_i and the semantic embedding of the task \mathbf{s}_j . This generated input is passed into three fully-connected layers (300,600,1). Except for the output layer, which is Sigmoid and generates relation scores, r_{ij} , all fully-connected layers are associated with LeakyReLU.

2 K-Means Visualization

We plot a tSNE visualization for all points in 10000 random different 3D point cloud objects in Fig. 3. We notice that some clusters have been formed from where we calculated the Microshapes.

3 Details of Experimental Setup

We experiment on two synthetic datasets i.e. ModelNet40 [5], ShapeNet [1] and three real-scanned dataset i.e. ScanObjectNN [4], Common Objects in 3D (CO3D) [3] with our proposed two different experimental setups.

3.1 Within-dataset Experiment

We design within-dataset experimental setups by ordering all classes in descending order for a dataset based on sample frequency. It assists us in distinguishing between base and novel classes since base classes have more instances than novel classes. Rare objects have fewer samples in the actual world. So, following this order, we create a realistic experimental setting. Therefore, all within-dataset experiments follow long-tail distribution, shown in Fig. 4. The base and incremental classes are treated as the head and tail classes of this data distribution, respectively.

(1) *ModelNet40*: It comprises 12,311 3D point cloud objects from 40 categories. We select 20 classes as base classes with 7438 training instances and 1958 test instances. The rest of the 20 classes are used for four incremental tasks consisting of 510 test instances.

(2) *ShapeNet*: It has 50604 shapes from 55 categories. We select 25 classes as base classes with the topmost training instances and a total of 36791 training and 9356 test samples. Then we choose the rest of the 30 classes as few-shot incremental classes with 887 test instances.

(3) *CO3D*: It is composed of 50 MS-COCO types of 3D point clouds. We choose 25 base classes, with 12493 training and 1325 test instances. The remaining 25 classes with 407 test instances are utilized for incremental training.

3.2 Cross-dataset Experiment

For cross-dataset experiments, we choose synthetic dataset as base class and real-scanned dataset as novel class. Table 1 shows the detailed data distribution for three experimental setups.

(1) *ModelNet40* \rightarrow *ScanObjectNN*: We follow the selection of classes from [2]. Here, we select 26 base classes from ModelNet40. On the other hand, ScanObjectNN has 15 classes with 2902 3D point cloud objects, but we choose non-overlapped 11 classes from ScanObjectNN for incremental tasks as novel classes.

(2) *ShapeNet* \rightarrow *ScanObjectNN*: We select 44 disjoint classes from ShapeNet

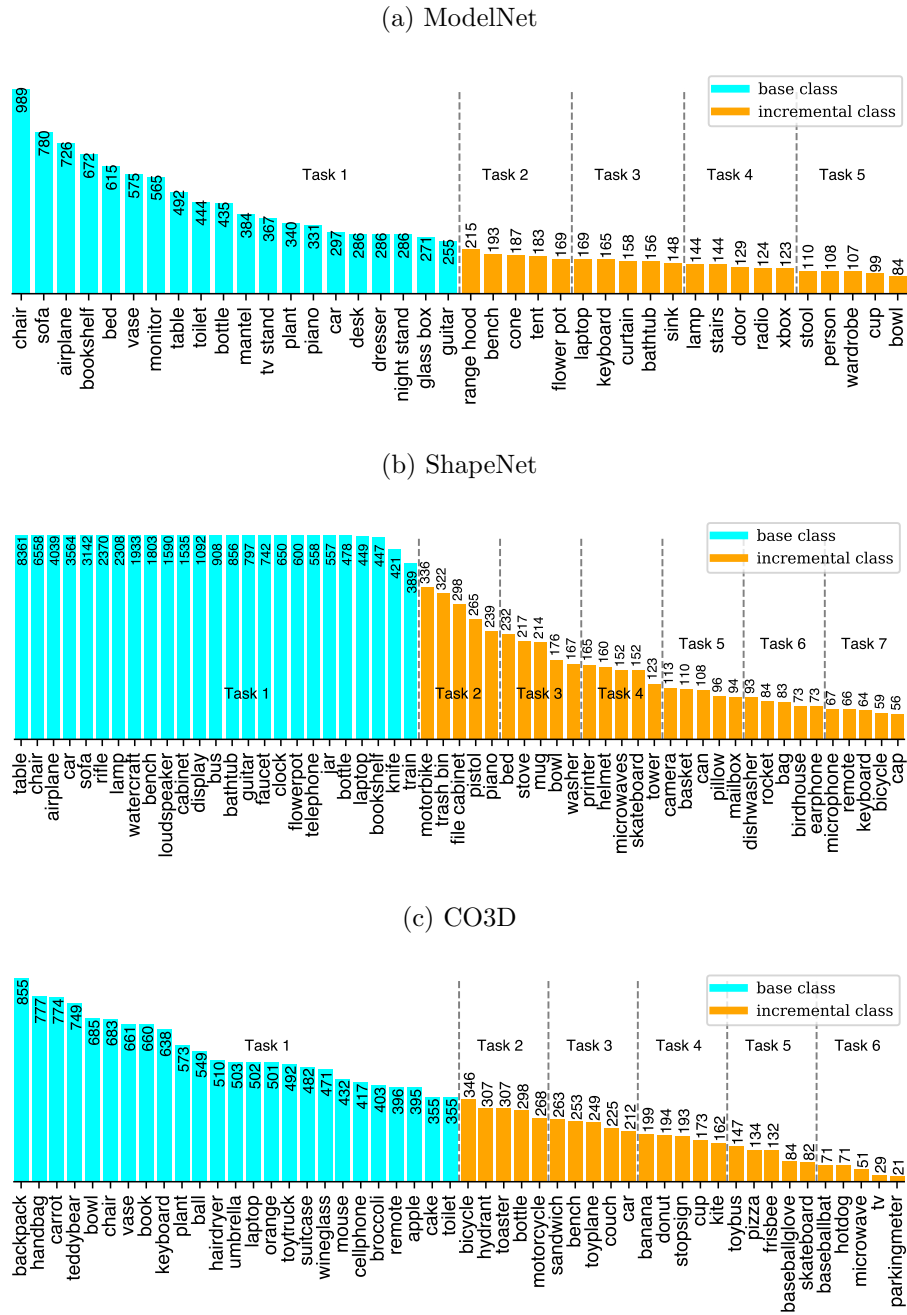


Fig. 4: Data distribution for within-dataset experiments by sorting all classes in descending order based on instance frequency. The plot clearly shows that all three setups follow a long-tail distribution. In the experimental setup of (b) ShapeNet, some long bars have been clipped for better visualization.

Table 1: Details of cross-dataset experimental setup. Task 1 represents the base class, whereas the rest of the tasks represent the novel class

(a) ShapeNet \rightarrow CO3D		
Dataset	Task	Name of class
ShapeNet	1	airplane, trash bin, basket, bathtub, bed, birdhouse, bookshelf, bus, cabinet, camera, can, cap, clock, dishwasher, display, faucet, file cabinet, guitar, helmet, jar, knife, lamp, loudspeaker, mailbox, microphone, mug, piano, pillow, pistol, flowerpot, printer, rifle, rocket, stove, table, tower, train, vessel, washer
	2	apple, backpack, ball, banana, baseballbat
	3	baseballglove, bench, bicycle, book, bottle
	4	bowl, broccoli, cake, car, carrot
	5	cellphone, chair, couch, cup, donut
	6	frisbee, hairdryer, handbag, hotdog, hydrant
	7	keyboard, kite, laptop, microwave, motorcycle
	8	mouse, orange, parking meter, pizza, plant
	9	remote, sandwich, skateboard, stopsign, suitcase
	10	teddybear, toaster, toilet, toybus, toyplane
	11	toytruck, tv, umbrella, vase, wineglass

(b) ShapeNet \rightarrow ScanObjectNN		
Dataset	Task	Name of class
ShapeNet	1	airplane, basket, bathtub, bench, bicycle, birdhouse, bottle, bowl, bus, camera, can, cap, car, clock, keyboard, dishwasher, earphone, faucet, file cabinet, guitar, helmet, jar, knife, lamp, laptop, loudspeaker, microphone, microwaves, motorbike, mug, piano, pistol, flowerpot, printer, remote, rifle, rocket, skateboard, stove, telephone, tower, train, watercraft, washer
	2	bag, bin, box, cabinet, chair
	3	desk, display, door, shelf, table
	4	bed, pillow, sink, sofa, toilet

(c) ModelNet \rightarrow ScanObjectNN		
Dataset	Task	Name of class
ModelNet	1	airplane, bathtub, bottle, bowl, car, cone, cup, curtain, flower pot, glass box, guitar, keyboard, lamp, laptop, mantel, night stand, person, piano, plant, radio, range hood, stairs, tent, tv stand, vase
	2	cabinet, chair, desk, display
	3	door, shelf, table, bed
	4	sink, sofa, toilet

as base classes. For base model training, there are 22318 training and 5845 test instances from ShapeNet. Also, from ScanObjectNN, we select all 15 classes for few-shot incremental model training with 581 test instances.

(3) *ShapeNet \rightarrow CO3D*: Base classes are selected from Shapenet, while for few-shot incremental steps, the classes are chosen from CO3D. This setup represents the most realistic scenario. We select non-overlapped 39 classes from Shapenet for base classes. Also, CO3D has 50 classes with 16557 training and 1732 test instances. Here, we choose all 50 classes for few-shot incremental tasks.

Table 2: 3D recognition on common objects of ModelNet40 (synthetic training) and ScanObjectNN (real-scanned testing)

Method	Accuracy
Baseline (without Microshape)	44.25
Ours (with Microshape)	46.34

4 3D Object Recognition

Microshape based 3D point description can benefit many problems beyond FSCIL. In Table 2, we perform 3D object recognition experiment on 11 common objects of ModelNet40 and ScanObjectNN. Here, the model is trained on the common (synthetic) objects of ModelNet40 and evaluated on the same (real-scanned) objects from ScanObjectNN. There is a clear domain gap from the

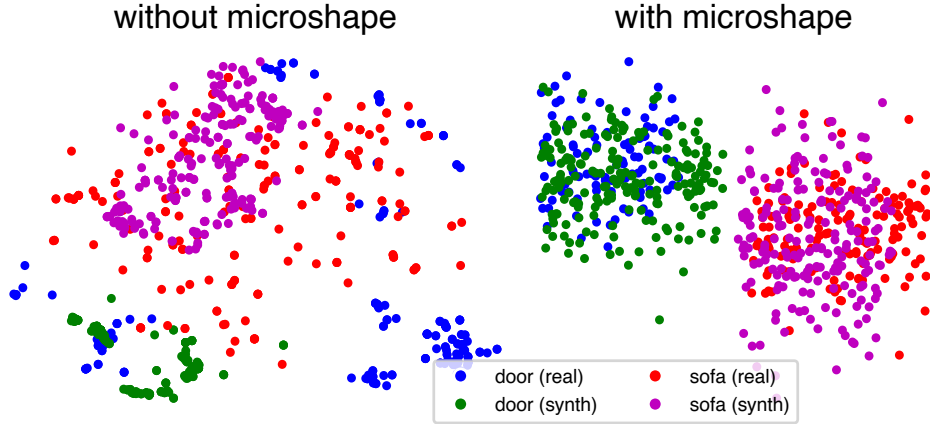


Fig. 5: Effect of using Microshape. Only two classes, ‘door’ and ‘sofa’, are used for visualization. Synthetic and real instances form different clusters on the left because of the domain gap. Microshape-based feature minimizes this gap by mixing both instances in the same cluster

data distribution of training (synthetic) and testing (real-scanned) instances. We attempt to reduce this gap using our Microshape based backbone and relation network (see Fig. 5). Our approach combining of Microshape and relation network with language prototype outperforms the baseline (without Microshape). It is possible because our method helped reducing the domain gap of training and testing data.

References

1. Chang, A.X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., et al.: Shapenet: An information-rich 3d model repository. arXiv preprint arXiv:1512.03012 (2015)
2. Cheraghian, A., Rahman, S., Chowdhury, T.F., Campbell, D., Petersson, L.: Zero-shot learning on 3d point cloud objects and beyond. arXiv preprint arXiv:2104.04980 (2021)
3. Reizenstein, J., Shapovalov, R., Henzler, P., Sbordone, L., Labatut, P., Novotny, D.: Common objects in 3d: Large-scale learning and evaluation of real-life 3d category reconstruction. In: ICCV (2021)
4. Uy, M.A., Pham, Q.H., Hua, B.S., Nguyen, D.T., Yeung, S.K.: Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In: ICCV (2019)
5. Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J.: 3D ShapeNets: A deep representation for volumetric shapes. In: CVPR (2015)