

Supplementary: Elemental Composite Prototypical Network: Few-Shot Object Detection on Outdoor 3D Point Cloud Scenes

A. Class-wise Experiment Results of O-FS3D on nuScenes Validation Set

We report the class-wise mean average precision (mAP) for each class (including base classes and novel classes) and official overall mAP and NDS metrics for all the baselines and our proposed model in Table 1, 2, 3, 4. We observe that the performance of our proposed model is superior to the baselines in all of the cases corresponding to the NDS score. According to mAP, our model is superior in 2 of the cases with comparable results concerning the baseline models. It is also worth noting that our model is performing well in a class-wise scenario for the novel classes and the base classes in most of the settings.

B. Performance Analysis on Low Training Data

To assess the efficiency of our proposed method, we compare our method with the baselines in a scenario where the number of annotations available for training data for novel classes is scarce. For this set of experiments, we choose the Novel-Split 3 of the nuScenes dataset (as mentioned in Section 4.1 of the main paper) and create 4 different episodic training schemes. In each of these episodic training schemes, all the baselines and our model use all available annotation data from base classes but only have access to a few annotations in the novel classes from where they can sample during each episode, thereby effectively making it more difficult for the models to train on novel classes as we gradually decrease the number of available annotations for the novel classes. This experiment is conducted to understand our model’s ability to adapt to the scarcity of annotation data compared to existing baselines. For the first experiment, we allow the models to sample from all available annotations of the novel classes *viz.* The available number of annotations are Bus - 12286, Construction Vehicle - 11050, Bicycle - 8185, and Traffic Cone - 62964. Hence, the models can sample from 23621 annotations per novel class on average. The results of average mAP on the novel classes are shown in Table 5 and the results on official mAP and NDS for all the classes (base and novel) for Novel-Split 3 are shown in Table 6.

From Tables 5 and 6, we observe that in all training schemes the performance drops as we lower the number of

available training annotation data for the novel classes. We also observe that the performance of our proposed method is consistently superior to the baselines in each of these experiments, thereby proving the effectiveness of our method under low-data scenarios.

C. Visualization on nuScenes Validation Set

We show the performance of our proposed **ECPN (PT + El-Proto + FSD)** on different scenes of the nuScenes validation dataset in Figure 1. The first row corresponds to results on VoxelNext trained in the episodic manner (**VoxelNext + EL**). The second row corresponds to VoxelNext’s state-of-the-art fully supervised model and the third row corresponds to our proposed ECPN coupled with elemental prototypes, query reweighting, and feature-similarity-discrimination loss *i.e.*, **ECPN (PT + El-Proto + FSD)**. Compared with the VoxelNext SOTA performance, ECPN performs fairly well for both base classes (green bounding boxes) and novel classes (blue boxes) in few-shot settings.

D. Visualization on In-house Outdoor dataset

We build an in-house outdoor 3D dataset by collecting 3D scenes using a 16 beam synced Velodyne PUCK Hi-Res LiDAR sensor¹. This dataset significantly differs from NuScenes on which our proposed model has been trained and analyzed. Key differences include over-crowded scenarios, and a low-resolution LiDAR sensor used to capture data different from what was used to capture NuScenes data (NuScenes uses 32-beam synced LiDAR sensor). Our dataset contains 300 frames per scene.

We use our proposed model which is trained on NuScenes in a few-shot episodic manner and run inference on our in-house outdoor dataset without any extra fine-tuning. We show qualitative results in Figure 2 that have been achieved by our proposed **ECPN (PT + El-Proto + FSD)** model.

In the figure, we show 4 frames each belonging to 3 different outdoor scene scans (in Row 1, 3, 5), and their corresponding outputs (in Row 2, 4, 6) for qualitative visualization. Each output contains one or multiple object bounding boxes with their color code as the classification ID

¹LiDAR Specification: <https://ouster.com/products/hardware/vlp-16>

Method	Novel Split - 1 (1st Random Split)										mAP	NDS
	Novel Classes				Base Classes							
	Trailer	C.V	Byc	T.C	Car	Truck	Bus	Ped	Mot	Barrier		
VoxelNext + EL	0.40	0.00	0.00	10.40	0.00	0.00	3.20	0.00	6.60	0.00	2.05	7.09
VoxelNext + PT + EL	1.50	14.50	22.10	32.80	0.00	30.90	62.5	34.00	58.10	21.40	27.78	42.33
VoxelNext + PT + Avg-Proto + EL	7.96	6.70	6.09	41.56	75.77	49.36	47.93	76.15	54.76	55.98	42.23	42.95
ECPN (PT + El-Proto + FSD)	16.61	7.36	21.03	40.40	72.96	49.83	61.96	77.27	47.66	52.09	44.72	47.25

Table 1. Class-wise results of O-FS3D experiments on novel split 1 of nuScenes object detection validation dataset. We report class-wise mAP and official overall mAP and NDS. EL: Episodic Learning; PT: Pre-training; Avg-Proto: Average Prototypes; El-Proto: Elemental Prototypes; FSD: Feature-Similarity-Discrimination Loss

Method	Novel Split - 2 (2nd Random Split)										mAP	NDS
	Novel Classes				Base Classes							
	C.V	Barrier	Mot	Ped	Car	Truck	Bus	Trailer	Byc	T.C		
VoxelNext + EL	0.00	0.00	9.04	0.00	0.00	0.00	1.05	0.00	0.55	8.51	1.92	8.10
VoxelNext + PT + EL	0.00	5.22	38.04	38.79	0.00	36.21	45.53	14.98	31.13	54.57	26.45	38.35
VoxelNext + PT + Avg-Proto + EL	7.03	22.01	42.98	56.41	74.62	47.82	62.62	23.66	20.03	59.08	41.63	42.90
ECPN (PT + El-Proto + FSD)	6.74	21.09	41.61	67.14	70.31	42.92	63.14	27.37	33.20	60.23	43.38	45.88

Table 2. Class-wise results of O-FS3D experiments on novel split 2 of nuScenes object detection validation dataset. We report class-wise mAP and official overall mAP and NDS. EL: Episodic Learning; PT: Pre-training; Avg-Proto: Average Prototypes; El-Proto: Elemental Prototypes; FSD: Feature-Similarity-Discrimination Loss

Method	Novel Split - 3 (Least examples in each group)										mAP	NDS
	Novel Classes				Base Classes							
	Bus	C.V	Byc	T.C	Car	Truck	Trailer	Ped	Mot	Barrier		
VoxelNext + EL	5.05	0.00	0.00	15.32	0.00	0.00	0.00	0.00	4.70	0.00	2.51	6.14
VoxelNext + PT + EL	34.47	0.44	16.28	22.35	0.00	12.22	18.64	51.15	48.58	8.98	21.31	37.49
VoxelNext + PT + Avg-Proto + EL	41.19	6.64	14.40	33.37	77.20	46.06	27.97	77.71	53.89	53.64	43.21	42.50
ECPN (PT + El-Proto + FSD)	40.92	7.36	18.55	37.14	76.85	43.61	20.84	75.07	51.05	45.40	41.68	44.68

Table 3. Class-wise results of O-FS3D experiments on novel split 3 of nuScenes object detection validation dataset. We report class-wise mAP and official overall mAP and NDS. EL: Episodic Learning; PT: Pre-training; Avg-Proto: Average Prototypes; El-Proto: Elemental Prototypes; FSD: Feature-Similarity-Discrimination Loss

Method	Novel Split - 4 (Least examples in the dataset)										mAP	NDS
	Novel Classes				Base Classes							
	Bus	C.V	Mot	Byc	Car	Barrier	Trailer	Ped	Truck	T.C		
VoxelNext + EL	5.52	0.00	7.59	0.12	0.00	0.00	0.00	0.00	0.00	12.23	2.55	8.82
VoxelNext + PT + EL	20.2	0.81	30.83	16.21	0.00	20.87	26.06	45.96	26.90	54.90	24.29	37.99
VoxelNext + PT + Avg-Proto + EL	41.86	6.86	37.78	17.72	77.30	53.44	26.29	75.38	42.29	65.37	44.44	42.88
ECPN (PT + El-Proto + FSD)	45.18	7.29	38.24	16.13	76.00	53.76	21.61	80.40	43.72	59.90	44.23	46.44

Table 4. Class-wise results of O-FS3D experiments on novel split 4 of nuScenes object detection validation dataset. We report class-wise mAP and official overall mAP and NDS. EL: Episodic Learning; PT: Pre-training; Avg-Proto: Average Prototypes; El-Proto: Elemental Prototypes; FSD: Feature-Similarity-Discrimination Loss

Method	Avg mAP on Number of Annos. Per Novel Class			
	23621 (Avg.)	8000	5000	1000
VoxelNext + PT + EL	18.39	13.74	13.15	9.1
VoxelNext + PT + Avg-Proto + EL	23.90	18.91	15.78	11.63
ECPN (PT + El-Proto + FSD)	25.99	19.41	17.93	12.11

Table 5. Results of O-FS3D experiments on Novel-Split 3 of nuScenes validation dataset on a low-data training scheme. The average mAP of all the novel classes in the Novel-Split 3 is reported. EL: Episodic Learning; PT: Pre-training; Avg-Proto: Average Prototypes; El-Proto: Elemental Prototypes; FSD: Feature-Similarity-Discrimination Loss

Method	Overall mAP and NDS on Number of Annos. Per Novel Class							
	23621 (Avg.)		8000		5000		1000	
	mAP	NDS	mAP	NDS	mAP	NDS	mAP	NDS
VoxelNext + PT + EL	21.31	37.49	23.73	40.73	19.81	39.33	17.63	37.08
VoxelNext + PT + Avg-Proto + EL	43.21	42.50	40.39	43.67	38.68	41.56	37.65	41.74
ECPN (PT + El-Proto + FSD)	41.86	44.68	39.60	41.75	39.57	44.60	37.19	42.31

Table 6. Results of O-FS3D experiments on Novel-Split 3 of nuScenes validation dataset on a low-data training scheme. The official mAP and NDS of all classes (base and novel) in the Novel-Split 3 are reported here. EL: Episodic Learning; PT: Pre-training; Avg-Proto: Average Prototypes; El-Proto: Elemental Prototypes; FSD: Feature-Similarity-Discrimination Loss

of the object. The outdoor data contains **car** as base-class

with **Red** bounding boxes. **Pedestrian**, and **Motor-cycle** are

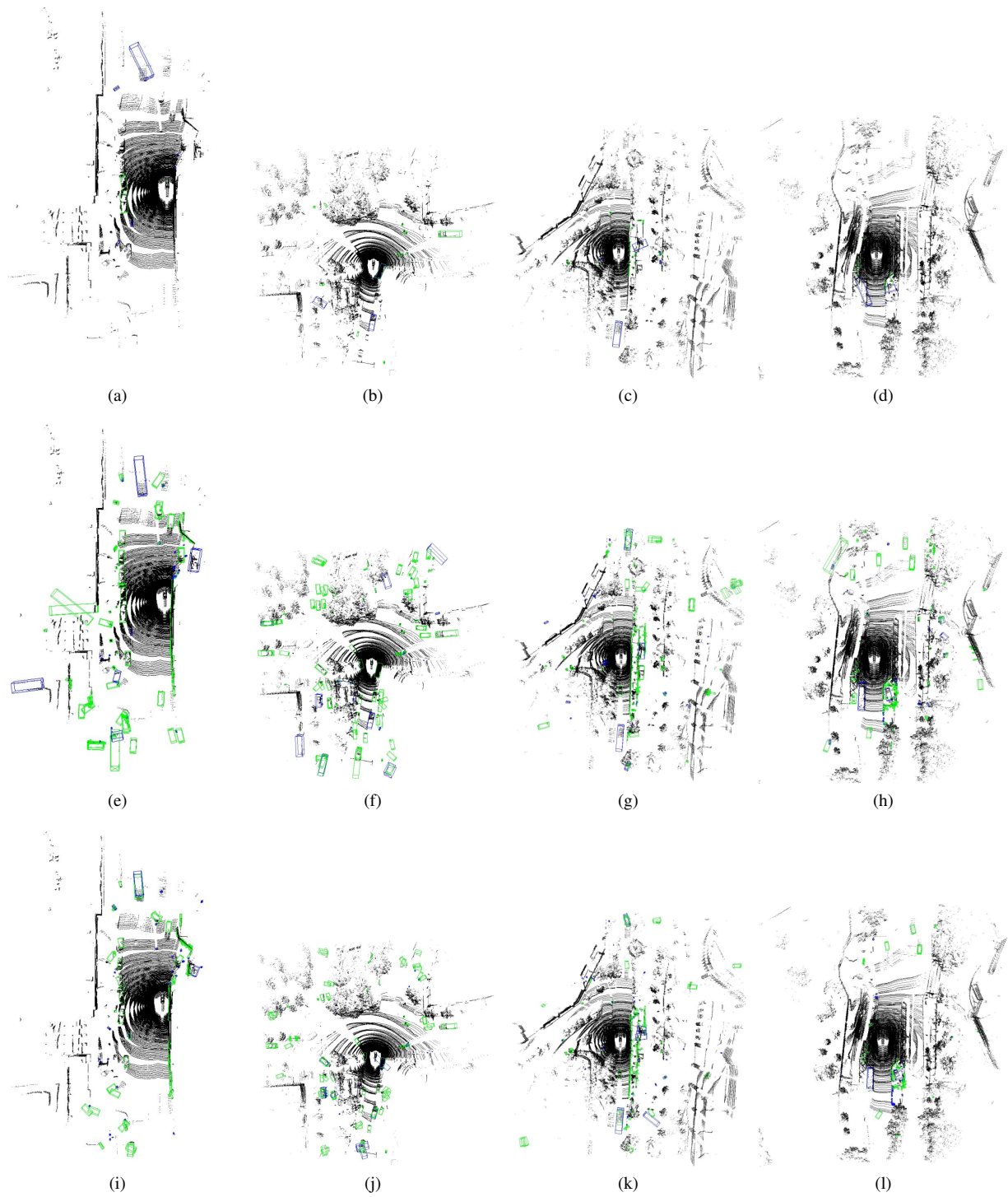


Figure 1. Qualitative Results on the nuScene validation dataset. The first row corresponds to the results of VoxelNext + EL, the second row corresponds to the results of Fully Supervised VoxelNext SOTA as per and the third row corresponds to the results of ECPN (PT + El-Proto + FSD). We use green for the base class and blue for the novel class bounding boxes. PT: Pre-Training; El-Proto: Elemental Prototypes; FSD: Feature-Similarity-Discrimination Loss

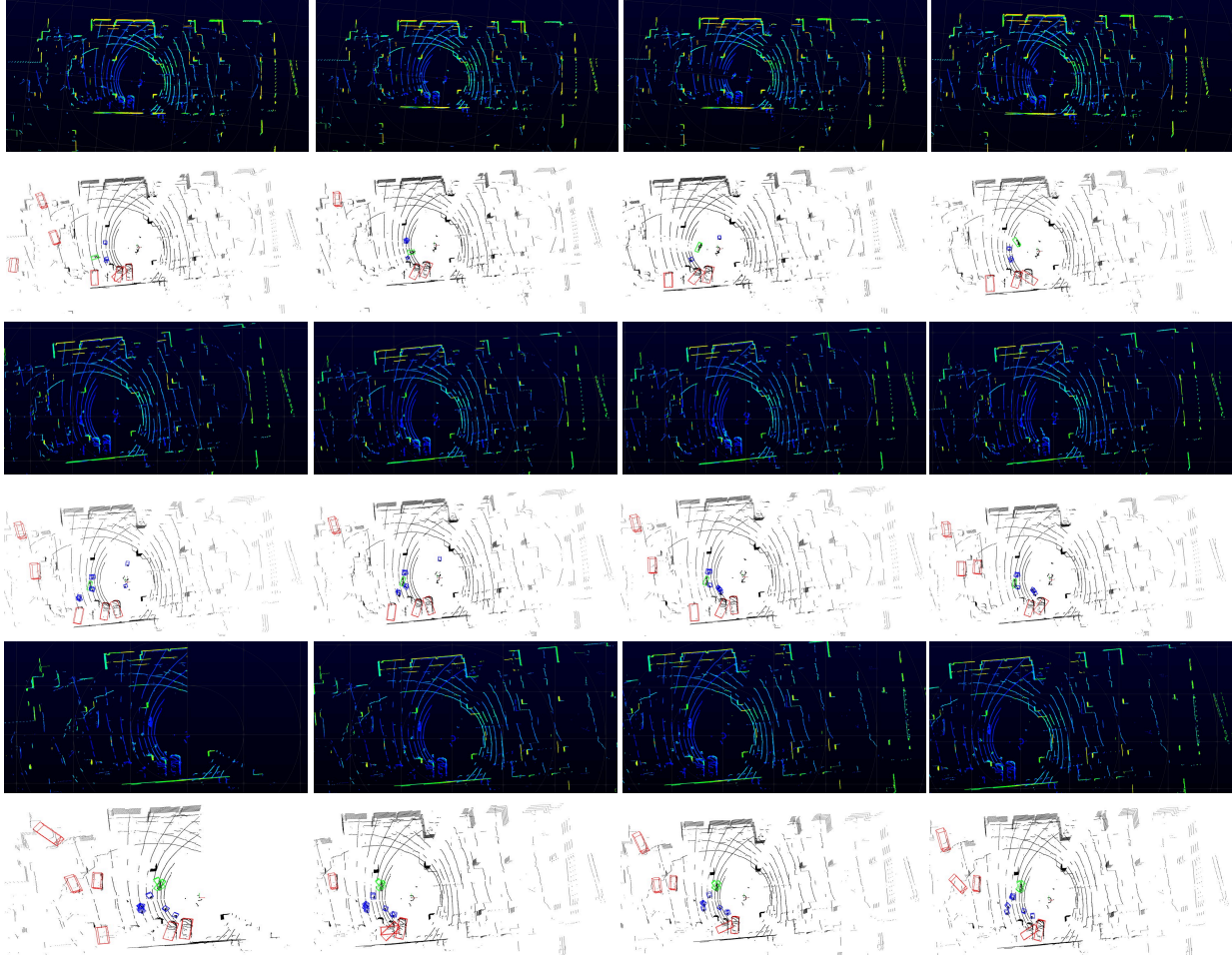


Figure 2. Visualization on Inhouse Outdoor dataset. Row 1, 3, and 5 are input from the LiDAR sensor to the model. Row 2, 4, and 6 are corresponding outputs with colored bounding boxes shown in the prediction. **Red** corresponds to Car, **Blue** corresponds to Pedestrian, and **Green** corresponds to Motor-cycle.

Method	mAP	NDS	Car	Truck	Bus	Trailer	C.V.	Ped	Mot	Byc	T.C.	Bar
VoxelNext [2]	60.0	67.1	85.6	58.4	71.6	38.6	17.9	85.4	59.7	43.4	70.8	68.1

Table 7. Fully supervised results of VoxelNext on nuScenes validation dataset as reported in [2]. C.V.: Construction Vehicle, Ped: Pedestrian, Mot: Motorcycle, Byc: Bicycle, T.C.: Traffic Cone, Bar: Barricade

novel-classes with **Blue**, and **Green** bounding boxes. We also provide a *demo video* of the frames from these scenes being predicted.

E. Inference Time Analysis

We conduct two different inference time analyses on our proposed architecture using NuScenes and our outdoor dataset. We run inference on 5 different scenes for both datasets and report the average inference time for an input 3D scene. For the NuScenes dataset and our outdoor

datasets, the average inference time is **0.2096 seconds**, and **0.1125 seconds** per scene, respectively. As evident from this inference time analysis, the quality of the input 3D scene, density of points, and overall number of points in the scene, etc. have a direct impact on the inference time. Since our data was captured with a 16 beam synced LiDAR and the nuScenes data was captured with a 32 beam synced LiDAR sensor, there are about twice as many points in each scene of the nuScenes dataset compared to ours. Hence, it also takes about twice the time to infer one scene from nuScenes

compared to our dataset.

F. Supervised VoxelNext Results

Since we use VoxelNext [2] in all of our baselines and our proposed method, we add the fully supervised VoxelNext results for the nuScenes dataset for comparison purposes. Table 7 represents the results.

G. Related Works

In this section, we discuss the related works in detail for a comprehensive overview of the field.

3D Point Cloud Object Detection: Current approaches for 3D point cloud object detection can be categorized into three main types: voxel based, point based and combined point-voxel based methods. In voxel-based approach [2, 8, 35], the point cloud is projected onto 2D grids or 3D voxels, allowing the use of CNNs directly. Point-based methods [22, 36, 37], on the other hand, provide raw point cloud as input to feature extraction networks, such as PointNet++ [15], to generate features for individual points before detection. The point-voxel based methods combine both, for example, STD [37] utilizes PointNet++ to extract semantic information from sparse points, which are then voxelized for detailed refinement. Similarly, PV-RCNN [21] integrates 3D sparse convolution with PointNet-like set abstraction to enhance semantic discrimination. Comparing the three subdivisions, voxel-based methods remains the most promising option for outdoor real-time applications [16]. While these fully supervised methods have achieved impressive 3D detection performance, they require large amounts of training data, which can be expensive in many real-world scenarios. To the best of our knowledge, there are no few-shot learning strategies designed for such voxel-based object detection methods.

Few-Shot Learning: Recent few-shot learning (FSL) methods predominantly rely on meta-learning. They fall into three main categories: metric-based, optimization-based and model-based. Metric-based methods [6, 11, 23, 25, 28, 39] aim to learn a function for embedding tasks and predicting labels based on distances. For instance, the Prototypical Network [23] computes average embedding vectors (prototypes) for each class, determining the class with the closest prototype distance for a new image. The optimization-based methods [4, 9, 12, 17] aim to improve model’s ability to adapt quickly to new tasks by focusing on learning optimization states like model initialization [4] or step sizes [12]. The model-based methods [13, 14, 18, 27] use specialized architectures [13, 14] and memory mechanisms [18, 27] to rapidly infer parameters. The aforementioned works primarily concentrate on 2D image understanding. Recently, several few-shot learning methods for point cloud understanding have been introduced [20, 42]. For example, Sharma et al. [20] pro-

posed self-supervised pre-training tasks for few shot learning that uses tree-based hierarchical partitioning. However, there has been less research on few-shot 3D point cloud object detection task.

Few-Shot Object Detection (FSOD): FSOD methods can be broadly categorized into four groups: data augmentation, transfer-learning, metric learning and meta-learning-based methods. Data augmentation based methods [24, 33, 41] utilize prior knowledge to increase data variance for novel categories. Transfer learning methods, as demonstrated in [29, 44], uses a simpler two-phase approach, involving initial training on base categories and subsequent fine-tuning on both base and novel categories with balanced data. Further, metric learning based methods [3, 5, 10, 19] embeds samples in a lower-dimensional space and facilitate effective training with fewer instances by classifying test samples based on their closest embedded training samples. RepMet [19] uses Gaussian mixture models and an embedding loss to maintain a margin between query features and class representatives. Fan et al. [3] introduce an attention-based RPN that enhances proposal generation and uses a multi-relation detector for measuring similarity between RoI features of query and support objects. Finally, meta-learning based methods [7, 30] quickly adapt to new tasks using a meta-learner trained on diverse tasks. MetaYOLO [7] improves query features with weighting coefficients generated during meta-learning, while Meta R-CNN [34] reweighs only ROI features for improved detection performance. Although, FSOD methodologies have gained traction in image related tasks, their adoption in 3D point clouds has been limited due to its unordered and irregular nature.

Self-supervised and Unsupervised methods for 3D Point Cloud Object Detection: Recently, self-supervised and unsupervised learning techniques are also gaining significant attention as they reduce the reliance on large-scale annotated datasets. For these approaches, similar to few-shot learning, the goal is to perform well with limited annotated samples. Some of these methods leverage unlabeled point cloud sequences [1], or generate pseudo-label using predictions from other collaborative units equipped with accurate detector [38] or other modalities [31]. Unsupervised approaches like Oyster [40] uses DBSCAN clustering on LiDAR point clouds to initialize pseudo ground truth followed by confidence-based forward-and-reverse tracking without ego-motion compensation, while CPD [32] refines pseudo-labels using commonsense prototypes to address sparsity challenge.

3D Point Cloud Few-shot Object Detection in Indoor Scene: As a pioneer work, Prototypical VoteNet [43] uses episodic training to learn class-agnostic geometric prototypes to enhance the local features of novel samples and class-specific prototypes to refine the object features. Another recent work, Prototypical Variational Autoencoder [26]

learns a probabilistic multi-center Gaussian Mixture Model (GMM)-like posterior, with each distribution centering at a prototype. However, both these works solve the problem of 3D-FSOD in indoor scenario. To the best of our knowledge, our paper presents the first attempt to investigate few-shot prototypical learning in outdoor Lidar 3D object detection.

References

- [1] Stefan Andreas Baur, Frank Moosmann, and Andreas Geiger. Liso: Lidar-only self-supervised 3d object detection. In *European Conference on Computer Vision*, pages 253–270. Springer, 2025. 5
- [2] Yukang Chen, Jianhui Liu, Xiangyu Zhang, Xiaojuan Qi, and Jiaya Jia. Voxelnex: Fully sparse voxelnet for 3d object detection and tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21674–21683, 2023. 4, 5
- [3] Qi Fan, Wei Zhuo, Chi-Keung Tang, and Yu-Wing Tai. Few-shot object detection with attention-rpn and multi-relation detector. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4013–4022, 2020. 5
- [4] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1126–1135. PMLR, 06–11 Aug 2017. 5
- [5] Guangxing Han, Yicheng He, Shiyuan Huang, Jiawei Ma, and Shih-Fu Chang. Query adaptive few-shot object detection with heterogeneous graph convolutional networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3263–3272, 2021. 5
- [6] Ruibing Hou, Hong Chang, Bingpeng Ma, Shiguang Shan, and Xilin Chen. Cross attention network for few-shot classification. *Advances in neural information processing systems*, 32, 2019. 5
- [7] Bingyi Kang, Zhuang Liu, Xin Wang, Fisher Yu, Jiashi Feng, and Trevor Darrell. Few-shot object detection via feature reweighting. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8420–8429, 2019. 5
- [8] Alex H Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders for object detection from point clouds. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12697–12705, 2019. 5
- [9] Kwonjoon Lee, Subhransu Maji, Avinash Ravichandran, and Stefano Soatto. Meta-learning with differentiable convex optimization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10657–10665, 2019. 5
- [10] Bohao Li, Boyu Yang, Chang Liu, Feng Liu, Rongrong Ji, and Qixiang Ye. Beyond max-margin: Class margin equilibrium for few-shot object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7363–7372, 2021. 5
- [11] Hongyang Li, David Eigen, Samuel Dodge, Matthew Zeiler, and Xiaogang Wang. Finding task-relevant features for few-shot learning by category traversal. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1–10, 2019. 5
- [12] Zhenguo Li, Fengwei Zhou, Fei Chen, and Hang Li. Meta-sgd: Learning to learn quickly for few-shot learning. *arXiv preprint arXiv:1707.09835*, 2017. 5
- [13] Nikhil Mishra, Mostafa Rohaninejad, Xi Chen, and Pieter Abbeel. A simple neural attentive meta-learner. *arXiv preprint arXiv:1707.03141*, 2017. 5
- [14] Tsendsuren Munkhdalai, Xingdi Yuan, Soroush Mehri, and Adam Trischler. Rapid adaptation with conditionally shifted neurons. In *International conference on machine learning*, pages 3664–3673. PMLR, 2018. 5
- [15] 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. 5
- [16] Rui Qian, Xin Lai, and Xirong Li. 3d object detection for autonomous driving: A survey. *Pattern Recognition*, 130:108796, 2022. 5
- [17] Andrei A Rusu, Dushyant Rao, Jakub Sygnowski, Oriol Vinyals, Razvan Pascanu, Simon Osindero, and Raia Hadsell. Meta-learning with latent embedding optimization. *arXiv preprint arXiv:1807.05960*, 2018. 5
- [18] Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, and Timothy Lillicrap. Meta-learning with memory-augmented neural networks. In *International conference on machine learning*, pages 1842–1850. PMLR, 2016. 5
- [19] Eli Schwartz, Leonid Karlinsky, Joseph Shtok, Sivan Harary, Mattias Marder, Sharathchandra Pankanti, Rogerio Feris, Abhishek Kumar, Raja Giries, and Alex M Bronstein. Repmet: Representative-based metric learning for classification and one-shot object detection. *arXiv preprint arXiv:1806.04728*, 4323, 2018. 5
- [20] Charu Sharma and Manohar Kaul. Self-supervised few-shot learning on point clouds. *Advances in Neural Information Processing Systems*, 33:7212–7221, 2020. 5
- [21] 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. 5
- [22] Shaoshuai Shi, Xiaogang Wang, and Hongsheng Li. Point-rcnn: 3d object proposal generation and detection from point cloud. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 770–779, 2019. 5
- [23] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. 5
- [24] Bo Sun, Banghuai Li, Shengcai Cai, Ye Yuan, and Chi Zhang. Fscf: Few-shot object detection via contrastive proposal encoding. In *Proceedings of the IEEE/CVF conference on com-*

- puter vision and pattern recognition, pages 7352–7362, 2021. 5
- [25] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1199–1208, 2018. 5
 - [26] Weiliang Tang, Biqu Yang, Xianzhi Li, Yunhui Liu, Pheng Ann Heng, and Chi-Wing Fu. Prototypical variational autoencoder for 3d few-shot object detection. In *Neural Information Processing Systems*, 2023. 5
 - [27] Kien Tran, Hiroshi Sato, and Masao Kubo. Memory augmented matching networks for few-shot learnings. *International Journal of Machine Learning and Computing*, 9(6), 2019. 5
 - [28] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. *Advances in neural information processing systems*, 29, 2016. 5
 - [29] Xin Wang, Thomas E Huang, Trevor Darrell, Joseph E Gonzalez, and Fisher Yu. Frustratingly simple few-shot object detection. *arXiv preprint arXiv:2003.06957*, 2020. 5
 - [30] Yu-Xiong Wang, Deva Ramanan, and Martial Hebert. Meta-learning to detect rare objects. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9925–9934, 2019. 5
 - [31] Maciej K Wozniak, Hariprasath Govindarajan, Marvin Klingner, Camille Maurice, Ravi Kiran, and Senthil Yogamani. S3pt: Scene semantics and structure guided clustering to boost self-supervised pre-training for autonomous driving. *arXiv preprint arXiv:2410.23085*, 2024. 5
 - [32] Hai Wu, Shijia Zhao, Xun Huang, Chenglu Wen, Xin Li, and Cheng Wang. Commonsense prototype for outdoor unsupervised 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14968–14977, 2024. 5
 - [33] Jiaxi Wu, Songtao Liu, Di Huang, and Yunhong Wang. Multi-scale positive sample refinement for few-shot object detection. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVI 16*, pages 456–472. Springer, 2020. 5
 - [34] Xiaopeng Yan, Ziliang Chen, Anni Xu, Xiaoxi Wang, Xiaodan Liang, and Liang Lin. Meta r-cnn: Towards general solver for instance-level low-shot learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9577–9586, 2019. 5
 - [35] Yan Yan, Yuxing Mao, and Bo Li. Second: Sparsely embedded convolutional detection. *Sensors*, 18(10):3337, 2018. 5
 - [36] Zetong Yang, Yanan Sun, Shu Liu, Xiaoyong Shen, and Jiaya Jia. Ipod: Intensive point-based object detector for point cloud. *arXiv preprint arXiv:1812.05276*, 2018. 5
 - [37] Zetong Yang, Yanan Sun, Shu Liu, Xiaoyong Shen, and Jiaya Jia. Std: Sparse-to-dense 3d object detector for point cloud. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1951–1960, 2019. 5
 - [38] Jinsu Yoo, Zhenyang Feng, Tai-Yu Pan, Yihong Sun, Cheng Perng Phoo, Xiangyu Chen, Mark Campbell, Kilian Q Weinberger, Bharath Hariharan, and Wei-Lun Chao. Learning 3d perception from others’ predictions. *arXiv preprint arXiv:2410.02646*, 2024. 5
 - [39] Sung Whan Yoon, Jun Seo, and Jaekyun Moon. Tapnet: Neural network augmented with task-adaptive projection for few-shot learning. In *International conference on machine learning*, pages 7115–7123. PMLR, 2019. 5
 - [40] Lunjun Zhang, Anqi Joyce Yang, Yuwen Xiong, Sergio Casas, Bin Yang, Mengye Ren, and Raquel Urtasun. Towards unsupervised object detection from lidar point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9317–9328, 2023. 5
 - [41] Weilin Zhang and Yu-Xiong Wang. Hallucination improves few-shot object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13008–13017, 2021. 5
 - [42] Na Zhao, Tat-Seng Chua, and Gim Hee Lee. Few-shot 3d point cloud semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8873–8882, 2021. 5
 - [43] Shizhen Zhao and Xiaojuan Qi. Prototypical votenet for few-shot 3d point cloud object detection. *Advances in neural information processing systems*, 35:13838–13851, 2022. 5
 - [44] Chenchen Zhu, Fangyi Chen, Uzair Ahmed, Zhiqiang Shen, and Marios Savvides. Semantic relation reasoning for shot-stable few-shot object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8782–8791, 2021. 5