Dual-level Adaptive Self-Labeling for Novel Class Discovery in Point Cloud Segmentation

Ruijie Xu[1,*], Chuyu Zhang[1,*], Hui Ren[1], and Xuming He[1,3]

[1] ShanghaiTech University, Shanghai, China
[2] Shanghai Engineering Research Center of Intelligent Vision and Imaging, Shanghai, China
{xurj2022,zhangchy2,renhui,hexm}@shanghaitech.edu.cn

Abstract. We tackle the novel class discovery in point cloud segmentation, which discovers novel classes based on the semantic knowledge of seen classes. Existing work proposes an online point-wise clustering method with a simplified equal class-size constraint on the novel classes to avoid degenerate solutions. However, the inherent imbalanced distribution of novel classes in point clouds typically violates the equal class-size constraint. Moreover, point-wise clustering ignores the rich spatial context information of objects, which results in less expressive representation for semantic segmentation. To address the above challenges, we propose a novel self-labeling strategy that adaptively generates high-quality pseudo-labels for imbalanced classes during model training. In addition, we develop a dual-level representation that incorporates regional consistency into the point-level classifier learning, reducing noise in generated segmentation. Finally, we conduct extensive experiments on two widely used datasets, SemanticKITTI and SemanticPOSS, and the results show our method outperforms the state of the art by a large margin.

Keywords: Novel class discovery · Point clouds semantic segmentation · Long-tailed learning

1 Introduction

Point cloud segmentation is a core problem in 3D perception [19] and potentially useful for a wide range of applications, such as autonomous driving and intelligent robotics [21, 29]. Recently, there has been tremendous progress in semantic segmentation of point clouds due to the utilization of deep learning techniques [16, 17]. However, current segmentation methods primarily focus on a closed-world setting where all the semantic classes are known beforehand. As such it has difficulty in coping with open-world scenarios where both known and novel classes coexist, which are commonly seen in real-world applications.

For open-world perception, a desirable capability is to automatically acquire new concepts based on existing knowledge [15]. While there has been much effort into addressing the problem of novel class discovery for 2D or RGBD

* Both authors contributed equally. Code is available at Github.

images [10, 14, 25, 41], few works have explored the corresponding task for 3D point clouds. Only recently, Riz et al. [28] propose an online point-wise clustering method for discovering novel classes in 3D point cloud segmentation. To avoid degenerate solutions, their method relies on an equal class-size constraint on the novel classes. Despite its promising results, such a simplified assumption faces two key challenges: First, the distribution of novel classes in point clouds is inherently imbalanced due to the different physical sizes of objects and the density of points. Imposing the equal-size constraint can be restrictive, causing the splitting of large classes or the merging of smaller ones. In addition, point-wise clustering tends to ignore the rich spatial context information of objects, which leads to less expressive representation for semantic segmentation.

To tackle the above challenges, we propose a dual-level adaptive self-labeling framework for novel class discovery in point cloud segmentation. The key idea of our approach is two-fold: 1) We design a novel self-labeling strategy that adaptively generates high-quality imbalanced pseudo-labels for model training, which facilitates clustering novel classes of varying sizes; 2) To incorporate semantic context, we develop a dual-level representation of 3D points by grouping points into regions and jointly learns the representations of novel classes at both the point and region levels. Such a dual-level representation imposes additional constraints on grouping the points likely belonging to the same category. This helps in mitigating the noise in the generated segmentation.

Specifically, our framework employs an encoder to extract point features for the input point cloud and average pooling to compute representations of pre-computed regions. Both types of features are fed into a prototype-based classifier to generate predictions across both known and novel categories for each point and region. To learn the feature encoder and class prototypes, we introduce a self-labeling-based learning procedure that iterates between pseudo-label generation for the novel classes and the full model training with cross-entropy losses on points and regions. Here the key step is to generate imbalanced pseudo labels, which is formulated as a semi-relaxed Optimal Transport (OT) problem with adaptive regularization on class distribution. Along with the training, we employ a data-dependent annealing scheme to adjust the regularization strength. Such a design prevents discovering degenerate solutions and meanwhile enhances the model flexibility in learning the imbalanced data distributions.

To demonstrate the effectiveness of our approach, we conduct extensive experiments on two widely-used datasets: SemanticKITTI [3] and SemanticPOSS [26]. The experimental results show that our method outperforms the state-of-the-art approaches by a large margin. Additionally, we conduct comprehensive ablation studies to evaluate the significance of the different components of our method. The contributions of our method are summarized as follows:

1. We propose a novel adaptive self-labeling framework for novel class discovery in point cloud segmentation, better modeling imbalanced novel classes.
2. We develop a dual-level representation for learning novel classes in point cloud data, which incorporates semantic context via augmenting the point prediction with regional consistency.

3. Our method achieves significant performance improvement on the SemanticPOSS and SemanticKITTI datasets across nearly all the experiment settings.

2 Related Work

Point cloud semantic segmentation. Point cloud semantic segmentation has attracted much attention in recent years [7, 22, 38, 42]. While previous methods have made significant progress, their primary focus is on closed-world scenarios that heavily rely on annotations for each class and cannot address open-world challenges. In contrast, we aim to develop a model to discover novel classes in 3D open-world scenarios. In the context of point cloud representation learning, incorporating spatial context is pivotal for enhancing representation learning. Several works [24, 39] introduce a hierarchical representation learning strategy that leverages regions as intermediaries to connect points and semantic clusters. Unlike them, we develop a dual-level learning strategy that concurrently learns to map points and regions to semantic classes. Thanks to the learning of region-level representation, our method is less sensitive to the local noises in point clouds. Moreover, we cluster regions into semantic classes by an imbalance-aware self-labeling algorithm instead of simple K-Means.

Novel class discovery. The majority of research on Novel Class Discovery (NCD) has focused on learning novel visual concepts in the 2D image domain via designing a variety of unsupervised losses on novel class data or regularization strategies [10, 13, 15, 34, 37, 40]. Among them, EUMS [41] addresses novel class discovery in semantic segmentation, employing a saliency model for clustering novel classes, along with entropy ranking and dynamic reassignment for clean pseudo labels. More relevantly, Zhang et al. [36] consider the NCD task in long-tailed classification scenarios, and develop a bi-level optimization strategy for model learning. It adopts a fixed regularization to prevent degeneracy, imposing strong restrictions on learned representations, and a complex dual-loop iterative optimization procedure. In contrast, we propose an adaptive regularization strategy, which is critical for the success of our self-labeling algorithm. Moreover, our formulation leads to a convex pseudo-label generation problem, efficiently solvable by a fast scaling algorithm [6, 8] (see Appendix A for detailed comparisons). Perhaps most closely related to our work is [28], which explored the NCD problem for the task of point cloud semantic segmentation. Assuming a uniform distribution of novel classes, they develop an optimal-transport-based self-labeling algorithm to cluster novel classes. However, the method neglects intrinsic class imbalance and spatial context in point cloud data, often leading to sub-optimal clustering results.

Optimal transport for pseudo labeling. Unlike naive pseudo labeling [20], Optimal Transport (OT) [27, 33]-based methods allow us to incorporate prior class distribution into pseudo-labels generation. Therefore, it has been used as a pseudo-labels generation strategy for a wide range of machine learning tasks,

including semi-supervised learning [18, 30, 31], clustering [1, 4, 35], and domain adaptation [5, 11, 23]. However, most of these works assume the prior class distribution is either known or simply the uniform distribution, which is restrictive for NCD. By contrast, we consider a more practical scenario, where the novel class distribution is unknown and imbalanced, and design a semi-relaxed OT formulation with a novel adaptive regularization.

3 Method

In this section, we first introduce the problem setup of novel class discovery for point cloud segmentation and an overview of our method in Sec.3.1. We then describe our network architecture, including dual-level representation of point clouds in Sec.3.2. Subsequently, we present in detail our adaptive self-labeling framework for model learning that discovers the novel classes in Sec.3.3. Finally, we introduce our strategy to estimating the number of novel classes in Sec.3.4.

3.1 Problem Setup and Overview

For the task of point cloud segmentation, the novel class discovery problem aims to learn to classify 3D points of a scene into known and novel semantic classes from a dataset consisting of annotated points from the known classes and unlabeled points from novel ones.

Formally, we consider a training set of 3D scenes, where each scene comprises two parts: 1) an annotated part of the scene $\{(x_n^s, y_n^s)\}_{n=1}^N$, which belongs to the known classes C^s and consists of original point clouds along with the corresponding labels for each point; 2) an unknown part of the scene $\{(x_m^u)\}_{m=1}^M$, which belongs to the novel classes C^u and does not contain any label information. These two sets C^s and C^u are mutually exclusive, i.e., $C^s \cap C^u = \emptyset$. Our goal is to learn a point cloud segmentation network that can accurately segment new scenes in a test set, each of which includes both known and novel classes.

To tackle the challenge of discovering novel classes in point clouds, we introduce a dual-level adaptive self-labeling framework to learn a segmentation network for both known and novel classes. The key idea of our method includes two aspects: 1) utilizing the spatial smooth prior of point clouds to generate regions and developing a dual-level representation that incorporates regional consistency into the point-level classifier learning; 2) generating imbalance pseudo-labels with a novel adaptive regularization. An overview of our framework is depicted in Fig.1.

3.2 Model Architecture

We adopt a generic segmentation model architecture consisting of a feature encoder for the input point cloud and a classifier head to generate the point-wise class label prediction. Note that to capture both known and novel classes, our feature encoder is shared by all the classes $C^s \cup C^u$ and the output space of our classifier head also includes known and novel classes. Below we first introduce our feature representation and encoder, followed by the classifier head.

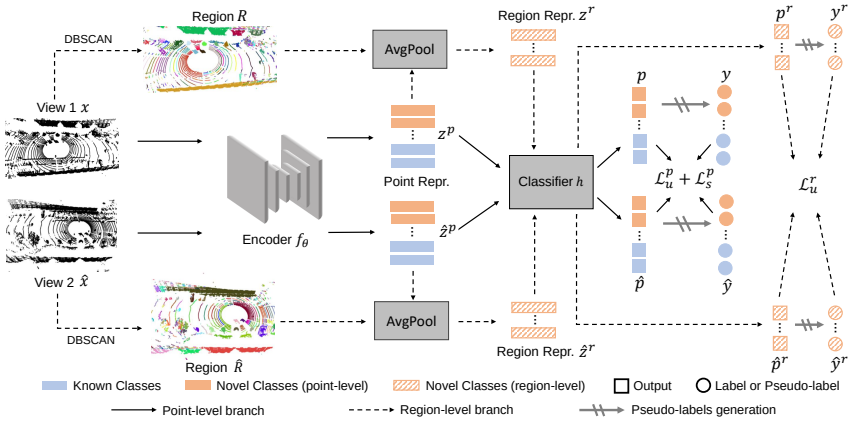

Fig. 1: Our method starts with two views of the same point cloud (x and $\hat{x}$) and clusters the points into corresponding regions. Then, we extract individual point features via a forward pass and calculate regional representations by averaging the point features within each region. Next, we make predictions p by the classifier h, and generate pseudo labels y for unlabeled points and regions using our novel adaptive self-labeling algorithm (generating pseudo-labels does not involve gradients). Lastly, we exchange the pseudo labels between the two views and update the model accordingly.

Dual-level Representation. Instead of treating each point independently, we exploit the spatial smoothness prior to 3D objects in our representation learning. To this end, we adopt a dual-level representation of point clouds that describes the input scene at different granularity. Specifically, given an input point cloud $\mathbf{X}$, we first use a backbone network f_θ to compute a point-wise feature $\mathbf{Z}^p = \{\mathbf{z}_i^p\}$, where $\mathbf{z}_i^p \in \mathbb{R}^{D\times 1}$. In this work, we employ MinkowskiUNet [7] for the backbone. In addition, we cluster points into regions based on their coordinates and then compute regional features by average pooling of point features. Concretely, during training, we first utilize DBSCAN [9] to generate K_i regions, $\mathcal{R} = \{r_k\}_{k=1}^{K_i}$, for unlabeled point in sample i, and calculate the regional features as follows,

$$\{r_k\}_{k=1}^{K_i} \leftarrow \text{DBSCAN}(\{x_i^u\}_{i=1}^{M}), \quad \mathbf{z}_k^r = \text{AvgPool}\{\mathbf{z}_i^p | \mathbf{z}_i^p = f_\theta(x_i^u),\ x_i^u \in r_k\}, \tag{1}$$

where $\mathbf{z}_k^r$ is the feature of region r_k. Such a dual-level representation allows us to enforce regional consistency in representation learning.

Prototype-based Classifier. We adopt a prototype-based classifier design for generating the point-wise predictions. Specifically, we introduce a set of prototypes for known and novel classes, denoted as $h = [h^s, h^u] \in \mathbb{R}^{D\times(|C^s|+|C^u|)}$, and D denotes the dimension of the last-layer feature. For each point or region, we compute the cosine similarity between its feature and the prototypes, followed by Softmax to predict the class probabilities. Here we use the same set of prototypes for the points and regions, which enforces a consistency constraint within each region and results in a more compact representation for each class.

Algorithm 1: Semi-relaxed Optimal Transport Algorithm

Function `Self-Labeling`$(-\log \mathbf{P},\ \gamma,\ \epsilon)$
  $\mathbf{K} = \exp(\log \mathbf{P}/\epsilon), \quad f \leftarrow \frac{\gamma}{\gamma+\epsilon}$
  $\mu, \nu \leftarrow \frac{1}{M}\mathbf{1}_M,\ \frac{1}{|C^u|}\mathbf{1}_{|C^u|}$ `//Marginal distribution`
  $\mathbf{b}_0 \leftarrow \mathbf{1}_{|C^u|}$ `//Initialize b`
  while $\|\mathbf{b}^{t+1} - \mathbf{b}^t\| < 1e-4$ **do**
    $\mathbf{a} \leftarrow \frac{\mu}{\mathbf{K}\mathbf{b}_t}$
    $\mathbf{b}_{t+1} \leftarrow (\frac{\nu}{\mathbf{K}^\top \mathbf{a}})^f$
  end
  $\mathbf{Q} \leftarrow M\mathrm{diag}(\mathbf{a})\mathbf{K}\mathrm{diag}(\mathbf{b})$
  return Q
end

3.3 Adaptive Self-labeling Framework

To handle class-imbalanced data, we propose an adaptive self-labeling framework that dynamically generates imbalanced pseudo-labels. To this end, we adopt the following loss function for the known and novel classes,

$$\mathcal{L} = \mathcal{L}_s + \alpha \mathcal{L}_u^p + \beta \mathcal{L}_u^r, \tag{2}$$

where $\mathcal{L}_s$ is the cross-entropy loss for known classes, $\mathcal{L}_u^p$ is point-level loss and $\mathcal{L}_u^r$ is region-level loss for novel classes. α and β are weight parameters. For the novel classes, we first generate pseudo-labels for points and regions by solving a semi-relaxed Optimal Transport problem and then adopt the cross-entropy loss with the generated labels. The pseudo-code of our algorithm is shown in Appendix B and below we will focus on our novel pseudo-label generation process.

Imbalanced Pseudo Label Generation. The pseudo-labels generation for balanced classes can be formulated as an optimal transport problem as follows [1,36]:

$$\min_{\mathbf{Q}} \frac{1}{M}\langle \mathbf{Q}, -\log \mathbf{P}^u \rangle_F, \quad \text{s.t. } \mathbf{Q}\mathbf{1}_{|C^u|} = \mathbf{1}_M, \mathbf{Q}^\top \mathbf{1}_M = \frac{M}{|C^u|}\mathbf{1}_{|C^u|}, \tag{3}$$

where $\mathbf{Q} \in \mathbb{R}^{M \times |C^u|}$ are the pseudo labels of unlabeled data, $<,>_F$ is Frobenius inner product and $\mathbf{P}^u$ are the output probabilities of the model. For imbalanced point cloud data, we relax the second constraint on the class sizes in Eq. (3), which leads to a parameterized semi-relaxed optimal transport problem as below:

$$\begin{aligned} \min_{\mathbf{Q}} \mathcal{F}_u(\mathbf{Q}, \gamma) = \frac{1}{M}\langle \mathbf{Q}, -\log \mathbf{P}^u \rangle_F + \gamma KL(\frac{1}{M}\mathbf{Q}^\top \mathbf{1}_M, \frac{1}{|C^u|}\mathbf{1}_{|C^u|}) \\ \text{s.t. } \mathbf{Q} \in \{\mathbf{Q} \in \mathbb{R}^{M \times |C^u|} | \mathbf{Q}\mathbf{1}_{|C^u|} = \mathbf{1}_M\}, \end{aligned} \tag{4}$$

where γ is a weight coefficient for balancing the constraint on cluster size distribution in the second term. We further add an entropy term $-\epsilon\mathcal{H}(\frac{1}{M}\mathbf{Q})$ to Eq. (4)

and for any given γ, this entropic semi-relaxed OT problem can be efficiently solved by fast scaling algorithms [6, 8]. Algorithm 1 outlines the optimization process, and further details are provided in Appendix A.

In this work, we propose a novel adaptive regularization strategy that adjusts the weight γ according to the progress of model learning, significantly improving pseudo-label quality. Details of our strategy will be illustrated subsequently.

Adaptive Regularization Strategy. The objective Eq. (4) aims to strike a balance between the distribution represented by model prediction $\mathbf{P}^u$ and the uniform prior distribution. A large γ tends to prevent the model from learning a degenerate solution, e.g. assigning all the samples into a single novel class, but it also restricts the model's capacity to learn the imbalanced data. One of our key insights is that the imbalanced NCD learning requires an adaptive strategy for setting the value of γ during the training. Intuitively, in the early training stage where the model performance is relatively poor, a larger constraint on $\mathbf{Q}^\top \mathbf{1}_M$ is needed to prevent degenerate solutions. As the training progresses, the model gradually learns meaningful clusters for novel classes, and the constraint should be relaxed to increase the flexibility of pseudo-label generation.

To achieve that, we develop an annealing-like strategy for adjusting γ, inspired by the ReduceLROnPlateau method that reduces the learning rate when the loss does not decrease. Here we employ the KL term in Eq. (4) as a guide for decreasing γ, as the value of the KL term reflects the relationship between the distribution of pseudo labels and the uniform distribution. Specifically, our formulation for the adaptive regularization factor is as follows:

$$\gamma_{t+1} = \lambda \gamma_t, \text{ if } KL(\frac{1}{M}\mathbf{Q}^\top \mathbf{1}_M, \frac{1}{|C^u|}\mathbf{1}_{|C^u|}) \leq \rho \text{ consecutively for } T \text{ iter.} \quad (5)$$

where ρ, λ, T and γ_0 are hyperparameters. Compared to typical step decay and cosine decay strategies, our adaptive strategy is aware of the model learning process and allows for more flexible control of γ based on the characteristics of the input itself.

Hyperparameter Search. To search the values of our hyperparameters, we design an indicator score that can be computed on the training dataset. Specifically, our indicator regularizes the total loss in Eq. (2) with a KL term that measures the distance between the distribution of novel classes and the uniform distribution. Formally, the indicator is defined as follows:

$$\mathcal{I} = \mathcal{L} + \gamma KL(\frac{1}{M}\mathbf{Q}^\top \mathbf{1}_M, \frac{1}{|C^u|}\mathbf{1}_{|C^u|}), \quad (6)$$

where γ is obtained by Eq. (5). Empirically, this indicator score provides a balanced evaluation of the model's performance in the known and novel classes.

3.4 Estimate the number of novel classes

To deal with realistic scenarios, where the number of novel classes (C^u) is unknown, we extend the classic estimation method [32] in NCD to point clouds

semantic segmentation for estimating C^u. Specifically, we extract representation from a known-class pre-trained model for training data, define a range of possible total class counts ($|C^s| < |C_{all}| <$ max classes), and apply Kmeans to cluster the labeled and unlabeled point clouds across different $|C_{all}|$. Then, we evaluate the clustering performance of known classes under different $|C_{all}|$, and select $|C_{all}|$ with the highest clustering performance as the estimated $|C_{all}|$.

4 Experiments

4.1 Experimental setup

Dataset. We perform evaluation on the widely-used SemanticKITTI [2, 3, 12] and SemanticPOSS [26] datasets. The SemanticKITTI dataset consists of 19 semantic classes, while the SemanticPOSS dataset contains 13 semantic classes. Both datasets have intrinsic class imbalances. For a fair comparison with existing works [28], we divide the dataset into 4 splits and select one split as novel classes, while treating the remaining splits as the known classes. Additionally, to assess the effectiveness of our method under more challenging conditions, we further split the SemanticPOSS dataset into two parts, selecting one part as novel classes. The dataset details are provided in Appendix C.

Evaluation Metric. Following the official guidelines in SemanticKITTI and SemanticPOSS, we conduct evaluations on sequences 08 and 03, respectively. These sequences contain both known and novel classes. For the known classes, we report the IoU for each class. Regarding the novel classes, we employ the Hungarian algorithm to initially match cluster labels with their corresponding ground truth labels. Subsequently, we present the IoU values for each of these novel classes. Additionally, we calculate the mean of columns across all known and novel classes.

Implementation Details. We follow [28] to adopt the MinkowskiUNet-34C [7] network as our backbone. For the parameters in DBSCAN, we set the min_samples to a reasonable value of 2, and select an epsilon value of 0.5, ensuring that 95% of the point clouds are included in the region branch learning process. A detailed analysis of DBSCAN is included in Appendix J. For the input point clouds, we set the voxel size as 0.05 and utilize the scale and rotation augmentation to generate two views. The scale range is from 0.95 to 1.05, and the rotation range is from -$\pi/20$ to $\pi/20$ for three axes. We train 10 epochs and set batch size as 4 for all experiments. The optimizer is Adamw, and the initial learning rate is 1e-3, which decreases to 1e-5 by a cosine schedule. For the hyperparameters, we set $\alpha = \beta = 1$ and fix λ at 0.5. We choose $T = 10$ and $\rho = 0.005$ based on the indicator mentioned in Sec. 3.3 and analyze them in the ablation study. Both the point- and region-level self-labeling algorithms employ the same parameters. All experiments are conducted on a single NVIDIA A100.

Table 1: The novel class discovery results on SemanticPOSS dataset. 'Number' denotes the number of points. 'Full' denotes the results obtained by supervised learning. The gray values are the novel classes in each split.

Split	Method	bike	build.	car	cone.	fence	grou.	pers.	plants	pole	rider	traf.	trashc.	trunk	Novel	Known	All
	Full	45.0	83.3	52.0	36.5	46.7	77.6	68.2	77.7	36.0	58.9	30.3	4.2	14.4	-	-	48.5
0	EUMS	25.7	4.0	0.6	16.4	29.4	36.8	43.8	28.5	13.1	26.8	18.2	3.3	16.9	17.4	21.5	20.3
	NOPS	35.5	30.4	1.2	13.5	24.1	69.1	44.7	42.1	19.2	47.7	24.4	8.2	21.8	35.7	26.6	29.4
	Ours	46.3	**51.5**	**6.0**	35.7	48.5	**83.0**	67.9	**53.1**	35.5	59.3	31.0	2.8	15.5	**48.4**	38.0	41.2
1	EUMS	15.2	68.0	28.0	24.0	11.9	75.1	36.0	74.5	26.9	48.6	26.0	5.6	23.1	21.0	40.0	35.6
	NOPS	29.4	71.4	28.7	12.2	3.9	78.2	**56.8**	74.2	18.3	38.9	23.3	13.7	23.5	30.0	38.2	36.4
	Ours	**31.5**	83.2	48.7	25.4	**23.9**	77.3	53.1	77.1	32.5	57.3	35.0	9.3	18.0	**36.2**	46.4	44.0
2	EUMS	40.1	69.5	27.7	13.5	34.9	76.0	54.7	75.6	5.3	39.2	7.8	8.5	11.9	8.3	44.0	35.7
	NOPS	37.2	71.8	29.7	14.6	28.4	77.5	52.1	73.0	**11.5**	47.1	0.5	10.2	**14.8**	9.0	44.2	36.0
	Ours	45.3	82.8	49.8	28.4	46.3	76.7	66.2	77.2	10.9	58.4	**18.6**	7.3	8.2	**12.6**	53.8	44.3
3	EUMS	41.2	70.7	28.1	**4.3**	38.3	76.7	38.3	75.4	25.8	34.3	28.3	0.4	24.4	13.0	44.7	37.4
	NOPS	38.6	70.4	30.9	0.0	29.4	76.5	56.0	71.8	17.0	31.9	26.2	1.0	22.6	10.9	43.9	36.3
	Ours	45.5	82.9	47.7	0.0	45.1	77.8	66.3	77.7	34.3	**49.1**	35.6	**4.0**	15.3	**17.7**	52.8	44.7

Table 2: Results on splits of SemanticPOSS dataset with more severe imbalance. The gray values are the novel classes in each split. NOPS is based on its released code.

Split	Method	bike	build.	car	cone.	fence	grou.	pers.	plants	pole	rider	traf.	trashc.	trunk	Novel	Known	All
	Full	45.0	83.3	52.0	36.5	46.7	77.6	68.2	77.7	36.0	58.9	30.3	4.2	14.4	-	-	48.5
0	NOPS	37.4	22.9	8.1	0.0	30.3	78.9	4.8	72.9	**1.0**	42.9	25.8	9.2	**9.6**	7.7	42.5	26.5
	Ours	46.0	**26.1**	**27.5**	**2.8**	46.9	77.6	**35.0**	77.8	0.2	58.6	30.5	3.2	0.0	**15.3**	48.7	33.2
1	NOPS	6.1	71.3	35.6	21.2	3.1	**42.9**	44.5	26.0	24.4	**0.7**	0.6	**0.1**	24.8	11.4	37.0	23.2
	Ours	**26.3**	82.0	51.4	18.0	**10.4**	40.0	67.5	**32.5**	31.2	0.0	**6.3**	0.0	11.7	**16.5**	44.5	29.4

4.2 Results

SemanticPOSS Dataset. As presented in Tab. 1, our approach exhibits significant improvements in novel classes over the previous method across all four splits. Specifically, we achieve an increase of **12.7%** and **6.2%** in split 0 and 1, respectively. It is worth noting that the fully supervised upper bounds for novel classes in split 0 and 1 are 72.7% and 53.3%, respectively, and the performance gaps have been significantly reduced. In the more challenging split 2 and split 3, we observe gains of **3.6%** and **4.7%**, respectively. The corresponding upper bounds for these splits are 26.9% and 33.2%, indicating their increased difficulty compared to splits 0 and 1. On average, we achieve an IoU of 30.2% for novel classes across all four splits, outperforming NOPS (22.5%) by **7.7%**. In addition, we provide a detailed comparison with NOPS on head, medium, and tail classes in Appendix D, as well as under a more comparable setting that applies our training strategy to NOPS in Appendix E.

To further verify that our method can alleviate the imbalanced problem, we divided the SemanticPOSS dataset into two splits, creating a more severe imbalance scenario that poses a greater challenge for clustering novel classes. As shown in Tab. 2, on novel classes, our method outperforms NOPS significantly on both splits, with a margin of **7.6%** on split 0 and **5.1%** on split 1. In particular, for the novel classes, we observe that our improvement mainly stems from the medium classes, such as person and bike. It is worth noting that NOPS employs extra training techniques, such as multihead and overclustering, whereas we use a simpler pipeline without needing them, further demonstrating our effectiveness.

SemanticKITTI Dataset. The results in Tab. 3 demonstrate our superior performance compared to previous methods on different splits. Specifically, we achieve significant improvements of **8.6%**, **3.3%**, and **3.6%** on splits 0, 1, and 2, respectively, for novel classes. The supervised upper bounds for these splits are 82.0%, 42.4%, and 39.6%, respectively. In split 3, our results are slightly higher than NOPS by 0.2%, possibly due to the scarce presence of these novel classes in split 3. On average across all four splits, our approach achieves an IoU of 27.5%, surpassing NOPS (23.4%) by **4.1%** on novel classes.

Visualization Analysis. Additionally, in Fig. 2, we perform visual comparisons on the results between NOPS and our method, and it is evident that our method shows significant improvements compared to NOPS. Specifically, as shown in the first row of Fig. 2, NOPS produces noisy predictions due to uniform constraints, mixing medium classes (e.g., building) and tail classes (e.g., car). In the second and third rows of Fig. 2, NOPS often confuses between medium and head classes, such as building and plants, as well as parking and car. In contrast, our method achieves better results for both datasets due to adaptive regularization and dual-level representation learning, generating high-quality imbalanced pseudo labels. More visual comparisons for additional splits are provided in the Appendix K.

4.3 Ablation Study

Component Analysis. To analyze the effectiveness of each component, we conduct extensive experiments on split 0 of the SemanticPOSS dataset. Here we provide ablation on three components, including Imbalanced Self-Labeling (ISL), Adaptive Regularization (AR), and Region-Level Branch (Region). As shown in Tab. 4, compared to baseline which employs equal-size constraints, imbalanced self-labeling improves performance by **4.2%**. The confusion matrix in Fig. 3 indicates that except for the highly-accurate class “ground", there is a significant improvement in the head and medium classes. This phenomenon is clearly depicted in Fig. 4, where the predictions of the baseline exhibit noticeable noise. From the second and third rows of Tab. 4, the adaptive regularization leads to a significant improvement of **8.2%** in split0 and **4.5%** in overall splits. As shown in Fig. 3, adaptive regularization enhances the quality of pseudo-labels for each class, especially for the head class (plants). We also visualize the class distribution of pseudo-labels in Appendix F, which shows adaptive regularization

Table 3: The novel class discovery results on the SemanticKITTI dataset. 'Full' denotes the results obtained by supervised learning. The four groups represent the four splits in turn, and the gray values are the novel classes in each split.

Method	bi.cle	b.clst	build.	car	fence	mt.cle	m.clst	oth-g.	oth-v.	park.	pers.	pole	road	side2.	terra.	traff.	truck	trunk	veget.	Novel	Known	All
Full	2.9	55.4	89.5	93.5	27.9	27.4	0.0	0.9	19.9	35.8	31.2	60.0	93.5	77.8	62.0	39.8	50.8	53.9	87.0	-	-	47.9
EUMS	5.3	40.0	15.8	79.2	9.0	16.9	2.5	0.1	11.4	14.4	12.7	29.2	42.6	**26.1**	0.1	10.3	47.4	37.9	38.4	24.6	21.1	23.1
NOPS	5.6	47.8	52.7	82.6	13.8	25.6	1.4	1.7	14.5	19.8	25.9	32.1	**56.7**	8.1	**23.8**	14.3	49.4	36.2	44.2	37.1	26.5	29.3
Ours	5.5	51.1	**74.6**	92.3	29.8	22.8	0.0	0.0	23.3	24.8	27.7	59.7	41.4	22.5	23.6	39.3	43.6	51.1	**66.4**	**45.7**	33.7	36.8
EUMS	7.5	42.4	80.0	**76.8**	8.6	19.6	1.4	**0.6**	12.0	14.1	14.0	40.7	86.3	66.5	56.3	12.0	44.8	20.9	72.4	24.2	37.1	35.6
NOPS	7.4	51.2	84.5	50.9	7.3	28.9	1.8	0.0	22.2	19.4	30.4	37.6	90.1	72.2	60.8	16.8	57.3	**49.3**	85.1	25.4	46.2	40.7
Ours	3.7	57.4	89.2	56.5	**17.3**	20.3	0.0	0.0	20.0	**30.6**	34.8	60.6	93.2	77.6	62.0	38.7	56.9	39.2	86.7	**28.7**	50.1	44.5
EUMS	8.3	50.8	83.0	88.1	17.9	2.8	2.3	0.2	3.2	25.4	25.0	20.2	88.3	71.0	57.9	8.6	27.2	38.4	77.0	12.4	42.2	36.6
NOPS	6.7	49.2	86.4	90.8	23.7	2.7	0.6	1.9	**15.5**	29.5	27.9	**36.4**	90.3	73.4	61.2	**17.8**	10.3	46.2	84.3	16.5	48.0	39.7
Ours	3.6	54.2	88.9	93.3	28.4	**10.2**	0.0	0.9	9.6	33.4	32.2	36.1	92.7	77.4	62.2	10.7	**34.2**	51.7	86.9	**20.1**	50.4	42.5
EUMS	**4.0**	2.5	80.1	87.2	16.8	14.0	**15.0**	0.3	14.1	20.8	6.8	37.6	86.8	66.5	55.3	16.2	40.6	38.4	76.2	7.1	43.4	35.7
NOPS	2.3	27.8	86.0	89.9	23.1	24.5	2.9	3.1	18.2	30.1	**16.3**	39.9	90.7	73.5	61.0	17.4	49.8	44.0	83.2	12.4	49.0	41.2
Ours	2.6	**32.5**	88.7	93.3	28.1	24	0.1	1.0	23.7	35.6	15.3	59.8	93.2	77.6	61.4	37.8	56.6	52.1	86.7	**12.6**	54.6	45.8

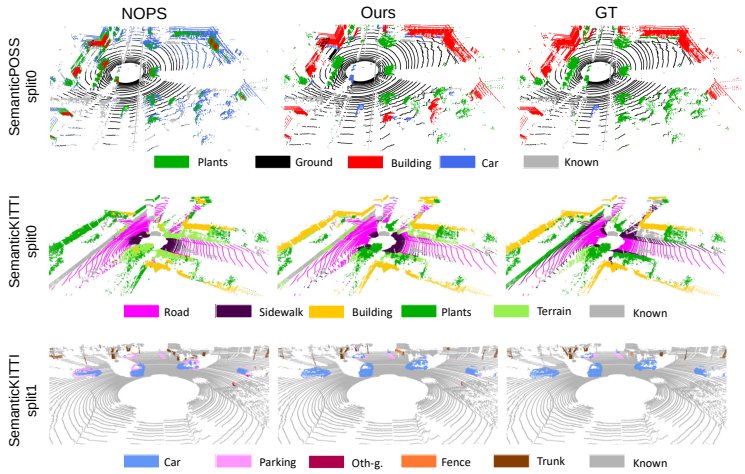

Fig. 2: Visualization comparison between Our method and NOPS on the SemanticPOSS and SemanticKITTI datasets. In the first and second rows, compared to NOPS, our method achieves much better segmentation for the 'Building', significantly reducing confusion with tail classes (such as 'Car') or medium classes (like 'Plants'). In the third row, our approach generates better segmentation for 'Parking' and 'Car'.

provides greater flexibility than fixed regularization term. According to the third, fourth and last rows of Tab. 4, the inclusion of the region-level branch leads to a **9.1%** improvement and an additional **4.2%** improvement built upon the AR. In addition, more experiments and analysis on prototype learning are included in Appendix G. In Fig. 3, there's a significant improvement in pseudo-labels for each category, particularly for the tail class (car) and the head class (plants). From Fig. 4, it is evident that the region-level branch can correct cases where a

Table 4: Ablation study on SemanticPOSS, focusing on novel classes. **ISL** and **AR** denote imbalanced self-labeling and regularization. **Region** denotes region-level learning. The last two columns represent the average mIoU for split 0 and across all splits.

ISL	AR	Region	Split0 Building	Car	Ground	Plants	Avg	Overall Avg
			21.6	2.7	76.6	26.1	31.8	20.9
✓			27.6	3.1	81.2	32.1	36.0	23.9
✓	✓		53.1	5.3	81.1	37.4	44.2	28.4
✓		✓	41.9	9.3	83.6	45.6	45.1	26.9
✓	✓	✓	51.5	6.0	83.0	53.1	48.4	30.2

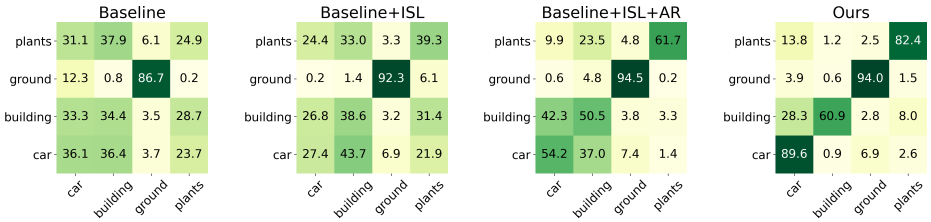

Fig. 3: Confusion Matrix, GT on the y-axis, Pseudo Label on the x-axis. (i, j) represents the % of GT in class j assigned pseudo label i. We categorize 'plants' and 'ground' as head classes, 'building' as medium, and 'car' as tail classes.

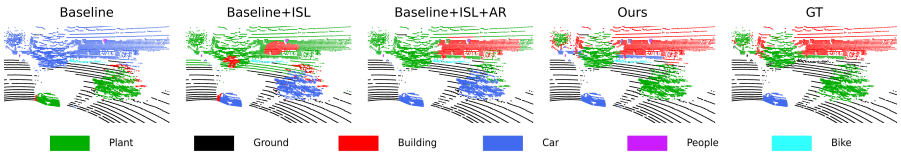

Fig. 4: Visualization analysis. The introduction of ISL notably reduces the misclassification between 'Plant' and 'Car'. Then, the integration of AR further mitigates the confusion between 'Plant' and 'Building'. Ultimately, the incorporation of Region component (Ours) effectively minimizes the mix-up between 'Plant', 'Car', and 'Building'.

Table 5: Analysis of adaptive regularization on SemanticPOSS dataset. GT denotes we directly assign the ground truth distribution of cluster size.

γ	0.01	0.05	0.1	0.5	1	5	$+\infty$	GT	Adaptive
Split0	10.1	33.8	33.3	36.0	32.2	33.8	31.8	32.5	44.2

single object is mistakenly labeled as multiple categories. Due to the utilization of spatial priors, where closely-located points are highly likely to belong to the same category, our region-level branch can correct misclassifications by considering context from neighboring points, preventing splitting a single object into multiple entities. Those experiments validate the effectiveness of each component in our method.

Table 6: Comparison between Ours and NOPS with the estimated number of novel classes in Split 0 of SemanticPOSS. The estimated C^u is 3, and the ground truth is 4.

Method	Building	Car	Ground	Plants	Avg
NOPS	25.54	0.00	68.15	34.12	31.95
Ours	64.05	0.00	82.22	67.63	**53.47**

Estimate the number of novel classes. For computational simplicity, we conduct experiments on splits 0 of the SemanticPOSS dataset and randomly sample 800,000 points from all scenes to estimate $|C^u|$. We set max classes to 50, which is an estimate of the maximum number of new classes that might appear in a typical scene. The estimated $|C^u|$ is 3, which is close to the ground truth value (GT is 4). Finally, we conduct experiments with $|C^u|$ as 3. As Tab. 6 illustrated our method still outperforms NOPS by a large margin.

Adaptive Regularization and Hyperparameters Selection. To analyze the impact of adaptive regularization, we compare it with various fixed regularization factors, as illustrated in Tab. 5. We notice that employing a very small fixed γ, such as 0.05 as indicated in the table, results in a weak prior constraint, and the model tends to learn a degenerate solution where all samples are assigned to a single cluster. When the γ increases to 0.5, the model achieves optimal results, but the increment decreases when the γ further increases. Compared with adaptive γ, the optimal results of fixed γ is nearly **8.2%** lower, demonstrating that the adoption of an adaptive γ not only enhances the model's flexibility but also prevents any performance degradation. Furthermore, we experiment with the setup adopting the GT class distribution and substituting the KL constraint in Eq. (4) with an equality constraint. Surprisingly, the results indicate that the GT class distribution constraint is not the optimal solution for clustering imbalanced novel classes. At last, in Fig. 5, we visualize the γ curves for SemanticPOSS in four splits. Split 0 exhibits the highest rate of change, followed by Split 1, while Splits 2 and 3 remain constant, indicating that our strategy is adaptive to each dataset.

To further validate the effectiveness of adjusting γ based on KL divergence, we also compare it with typical step decay and cosine annealing strategies. For the step decay, we set the initial γ to 1 and decay it by multiplying it with λ every epoch. For the cosine annealing approach, we also set the initial γ to 1 and reduce it to the minimum value (min γ). From the Tab. 7 and Tab. 8, we observe that the results of simple step decay and cosine annealing are nearly **10%** worse than adaptive γ (which is 44.2). We believe that these two typical strategies lack flexibility compared to adaptive γ. They might not facilitate the adaptive control of the γ decay process based on the model learning process.

To choose the hyperparameters ρ and T according to the indicator outlined in Sec. 3.3, we conduct experiments for various values of ρ and T. The results are displayed in Tab. 9 and Tab. 10. Additionally, we plot the indicator's curve for each experiment in Fig. 6 and 7. The plots reveal that when ρ falls within the

Table 7: Step decay results

λ	0.1	0.3	0.5	0.7	0.9
Step decay	34.0	34.2	32.9	34.6	33.3

Table 8: Cosine annealing results

min γ	0.1	0.05	0.01	0.005	0.001
Cosine annealing	32.2	32.5	35.8	36.1	32.0

Table 9: The results for different ρ

ρ	0.05	0.01	0.005	0.001
Split0	10.08	45.84	44.21	33.11

Table 10: The results for different T

T	5	10	20	30
Split0	44.46	44.21	44.12	32.48

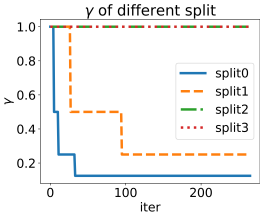

Fig. 5: γ variation

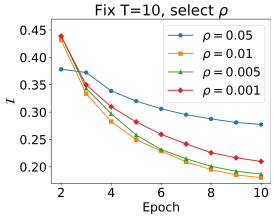

Fig. 6: Selecting ρ

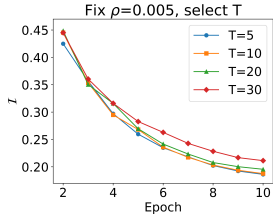

Fig. 7: Selecting T

range of 0.01 to 0.005, and T is set between 5 and 20, the indicator value remains low while achieving a high novel IoU. Those results demonstrate the efficiency of our hyperparameters selection strategy and the robustness of our method.

Limitations. One limitation is our problem setup which follows [28] and only addresses scenarios where unlabelled data constitutes novel classes. In contrast, a more realistic open-world setting necessitates handling situations where both known classes and novel classes lack labels. Nevertheless, we anticipate that our method will establish a robust baseline and stimulate further research aimed at addressing the challenges presented by practical open-world situations.

5 Conclusion

In this paper, we propose a novel dual-level adaptive self-labeling framework for novel class discovery in point cloud segmentation. Our framework formulates the pseudo label generation process as a Semi-relaxed Optimal Transport problem and incorporates a novel data-dependent adaptive regularization factor to gradually relax the constraint of the uniform prior based on the distribution of pseudo labels, thereby generating higher-quality imbalanced pseudo labels for model learning. In addition, we develop a dual-level representation that leverages the spatial prior to generate region representation, which reduces the noise in generated segmentation and enhances point-level classifier learning. Furthermore, we propose a hyperparameters search strategy based on training sets. Extensive experiments on two widely used datasets, SemanticKITTI and SemanticPOSS, demonstrate the effectiveness of each component and the superiority of our method.

Acknowledgments This work was supported by National Science Foundation of China under grant 62350610269, Shanghai Frontiers Science Center of Human-centered Artificial Intelligence, and MoE Key Lab of Intelligent Perception and Human-Machine Collaboration (ShanghaiTech University).

References

1. Asano, Y.M., Rupprecht, C., Vedaldi, A.: Self-labelling via simultaneous clustering and representation learning. In: International Conference on Learning Representations (ICLR) (2020)
2. Behley, J., Garbade, M., Milioto, A., Quenzel, J., Behnke, S., Gall, J., Stachniss, C.: Towards 3d lidar-based semantic scene understanding of 3d point cloud sequences: The semantickitti dataset. The International Journal of Robotics Research **40**(8-9), 959–967 (2021)
3. Behley, J., Garbade, M., Milioto, A., Quenzel, J., Behnke, S., Stachniss, C., Gall, J.: Semantickitti: A dataset for semantic scene understanding of lidar sequences. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 9297–9307 (2019)
4. Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., Joulin, A.: Unsupervised learning of visual features by contrasting cluster assignments. Advances in Neural Information Processing Systems **33**, 9912–9924 (2020)
5. Chang, W., Shi, Y., Tuan, H., Wang, J.: Unified optimal transport framework for universal domain adaptation. Advances in Neural Information Processing Systems **35**, 29512–29524 (2022)
6. Chizat, L., Peyré, G., Schmitzer, B., Vialard, F.X.: Scaling algorithms for unbalanced optimal transport problems. Mathematics of Computation **87**(314), 2563–2609 (2018)
7. Choy, C., Gwak, J., Savarese, S.: 4d spatio-temporal convnets: Minkowski convolutional neural networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3075–3084 (2019)
8. Cuturi, M.: Sinkhorn distances: Lightspeed computation of optimal transport. Advances in neural information processing systems **26** (2013)
9. Ester, M., Kriegel, H.P., Sander, J., Xu, X., et al.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: kdd. vol. 96, pp. 226–231 (1996)
10. Fini, E., Sangineto, E., Lathuilière, S., Zhong, Z., Nabi, M., Ricci, E.: A unified objective for novel class discovery. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9284–9292 (2021)
11. Flamary, R., Courty, N., Tuia, D., Rakotomamonjy, A.: Optimal transport for domain adaptation. IEEE Trans. Pattern Anal. Mach. Intell **1**, 1–40 (2016)
12. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: 2012 IEEE conference on computer vision and pattern recognition. pp. 3354–3361. IEEE (2012)
13. Gu, P., Zhang, C., Xu, R., He, X.: Class-relation knowledge distillation for novel class discovery. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 16428–16437. IEEE Computer Society (2023)
14. Han, K., Rebuffi, S.A., Ehrhardt, S., Vedaldi, A., Zisserman, A.: Autonovel: Automatically discovering and learning novel visual categories. IEEE Transactions on Pattern Analysis and Machine Intelligence (2021)

15. Han, K., Vedaldi, A., Zisserman, A.: Learning to discover novel visual categories via deep transfer clustering. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8401–8409 (2019)
16. Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N., Markham, A.: Randla-net: Efficient semantic segmentation of large-scale point clouds. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11108–11117 (2020)
17. Lai, X., Liu, J., Jiang, L., Wang, L., Zhao, H., Liu, S., Qi, X., Jia, J.: Stratified transformer for 3d point cloud segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8500–8509 (2022)
18. Lai, Z., Wang, C., Cheung, S.c., Chuah, C.N.: Sar: Self-adaptive refinement on pseudo labels for multiclass-imbalanced semi-supervised learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4091–4100 (2022)
19. Landrieu, L., Simonovsky, M.: Large-scale point cloud semantic segmentation with superpoint graphs. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4558–4567 (2018)
20. Lee, D.H., et al.: Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: Workshop on challenges in representation learning, ICML. vol. 3, p. 896. Atlanta (2013)
21. Li, Y., Ma, L., Zhong, Z., Liu, F., Chapman, M.A., Cao, D., Li, J.: Deep learning for lidar point clouds in autonomous driving: A review. IEEE Transactions on Neural Networks and Learning Systems **32**(8), 3412–3432 (2020)
22. Li, Z., Wang, W., Li, H., Xie, E., Sima, C., Lu, T., Qiao, Y., Dai, J.: Bevformer: Learning bird's-eye-view representation from multi-camera images via spatiotemporal transformers. In: European conference on computer vision. pp. 1–18. Springer (2022)
23. Liu, Y., Zhou, Z., Sun, B.: Cot: Unsupervised domain adaptation with clustering and optimal transport. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 19998–20007 (June 2023)
24. Long, F., Yao, Ting abd Qiu, Z., Li, L., Mei, T.: Pointclustering: Unsupervised point cloud pre-training using transformation invariance in clustering. In: CVPR (2023)
25. Nakajima, Y., Kang, B., Saito, H., Kitani, K.: Incremental class discovery for semantic segmentation with rgbd sensing. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 972–981 (2019)
26. Pan, Y., Gao, B., Mei, J., Geng, S., Li, C., Zhao, H.: Semanticposs: A point cloud dataset with large quantity of dynamic instances. In: 2020 IEEE Intelligent Vehicles Symposium (IV). pp. 687–693. IEEE (2020)
27. Phatak, A., Raghvendra, S., Tripathy, C., Zhang, K.: Computing all optimal partial transports. In: International Conference on Learning Representations (2023)
28. Riz, L., Saltori, C., Ricci, E., Poiesi, F.: Novel class discovery for 3d point cloud semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9393–9402 (2023)
29. Roriz, R., Cabral, J., Gomes, T.: Automotive lidar technology: A survey. IEEE Transactions on Intelligent Transportation Systems **23**(7), 6282–6297 (2021)
30. Taherkhani, F., Dabouei, A., Soleymani, S., Dawson, J., Nasrabadi, N.M.: Transporting labels via hierarchical optimal transport for semi-supervised learning. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IV 16. pp. 509–526. Springer (2020)

31. Tai, K.S., Bailis, P.D., Valiant, G.: Sinkhorn label allocation: Semi-supervised classification via annealed self-training. In: International Conference on Machine Learning. pp. 10065–10075. PMLR (2021)
32. Vaze, S., Han, K., Vedaldi, A., Zisserman, A.: Generalized category discovery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7492–7501 (2022)
33. Villani, C., et al.: Optimal transport: old and new, vol. 338. Springer (2009)
34. Yang, M., Zhu, Y., Yu, J., Wu, A., Deng, C.: Divide and conquer: Compositional experts for generalized novel class discovery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14268–14277 (2022)
35. Zhang, C., Ren, H., He, X.: P^2ot: Progressive partial optimal transport for deep imbalanced clustering. In: The Twelfth International Conference on Learning Representations (2023)
36. Zhang, C., Xu, R., He, X.: Novel class discovery for long-tailed recognition. Transactions on Machine Learning Research (2023)
37. Zhang, S., Khan, S., Shen, Z., Naseer, M., Chen, G., Khan, F.S.: Promptcal: Contrastive affinity learning via auxiliary prompts for generalized novel category discovery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3479–3488 (2023)
38. Zhang, Y., Zhou, Z., David, P., Yue, X., Xi, Z., Gong, B., Foroosh, H.: Polarnet: An improved grid representation for online lidar point clouds semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9601–9610 (2020)
39. Zhang, Z., Yang, B., Wang, B., Li, B.: Growsp: Unsupervised semantic segmentation of 3d point clouds. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17619–17629 (2023)
40. Zhao, B., Han, K.: Novel visual category discovery with dual ranking statistics and mutual knowledge distillation. Advances in Neural Information Processing Systems **34** (2021)
41. Zhao, Y., Zhong, Z., Sebe, N., Lee, G.H.: Novel class discovery in semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4340–4349 (2022)
42. Zhu, X., Zhou, H., Wang, T., Hong, F., Ma, Y., Li, W., Li, H., Lin, D.: Cylindrical and asymmetrical 3d convolution networks for lidar segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9939–9948 (2021)