Hard-Soft Pseudo Labels Guided Semi-Supervised Learning for Point Cloud Classification

Yuan He, Guyue Hu, and Shan Yu

***Abstract*—Point clouds are widely applied in 3D visual sensing and perception. However, manually annotating point clouds is much more tedious and time-consuming than that for 2D images. Fortunately, semi-supervised learning can leverage massive unlabeled data to alleviate this issue, which is becoming a promising technique nowadays. In this letter, we propose a novel semi-supervised learning (SSL) framework for point cloud classification, named HPSSL. Its unsupervised learning branch performs both the representation embedding and pseudo-classification tasks. Specifically, both hard and soft pseudo labels of unlabeled samples are generated from a shared classifier to guide the class-aware contrastive learning in our SSL framework. Besides, a prediction consistency strategy is proposed to enhance the discrimination of feature representation and the exactness of pseudo labels. Furthermore, we force the supervised learning branch to interact with the unsupervised learning branch via distribution alignment, thus achieving representation consistency. Extensive experiments on three 3D shape recognition benchmarks demonstrate the effectiveness of the proposed approach.**

***Index Terms*—Point cloud, 3D vision, semi-supervised learning, contrastive learning, pseudo label.**

I. Introduction

With the rapid development of 3D visual scanners and technologies, 3D point cloud has become one of the most important data formats for 3D signal analysis, applied in a variety of applications including autonomous driving [1], [2], robotics [3], [4], [5], virtual/augmented reality [6], [7], etc. Recently, deep neural network-based methods have achieved impressive performance on point cloud analysis [8], [9], [10], [11], [12]. However, most of these approaches require extensive annotated information, which is particularly challenging for 3D point cloud data as its annotation is more time-consuming and costly than 2D images. Thus, it's desirable to explore semi-supervised learning [13], [14], [15] to alleviate the annotation burden of point cloud data and enhance model performance.

Semi-supervised learning (SSL) has been an effective scheme for leveraging massive of unlabeled data along with limited labeled data. The commonly employed approaches are pseudo-labeling and consistency regularization [14], [15]. In addition, extending contrastive learning into semi-supervised algorithms has yielded significant improvements in the field of image processing [16], [17], [18]. Although many SSL works have been proposed for images, SSL is rather under-explored for 3D point clouds. There are a few SSL works [19], [20], [21] utilizing pseudo-labeling and point-level contrastive loss for point cloud semantic segmentation. However, these works mainly focused on learning the representation encoder while neglecting to exploit the semantic classifier.

In addition, some previous works have explored contrastive learning on point clouds [22], [23], [24], and they learn representations from unlabeled data by conducting instance discrimination tasks. The instance discrimination task only clusters samples of the same instance while distinguishing different instances of the same class. However, instances belonging to the same class should be clustered together in the feature space for the point cloud classification task. Unfortunately, there is a heavy disagreement between these two objectives, which will prevent learning more discriminative representations. In summary, these existing works on semi-supervised point cloud classification methods have two heavy limitations: 1) they rely on instance discrimination to learn representations which lacks enough semantic information for classification. 2) they focused on learning the representation encoder while neglecting the significance of the semantic classifier. To move beyond such limitations, we design a novel semi-supervised framework for point cloud classification. Our main contributions are summarized as follows: 1) Both hard and soft pseudo labels are generated to guide the unsupervised representation learning. The hard pseudo-labels provide explicit but probably unreliable class information while the soft pseudo-labels give continuous but probably redundant probability information. The proposed class-aware contrastive loss successfully combines the advantages of these two types of pseudo labels by utilizing hard-soft pseudo labels to guide representation learning. 2) A prediction consistency strategy is designed to further improve the discrimination of feature representation in the encoder and the exactness of pseudo-labels generation in the classifier. 3) We successfully integrate the point cloud encoder, projector, classifier, momentum encoder, and momentum projector into a whole framework to realize hard-soft

Manuscript received 8 February 2024; accepted 28 March 2024. Date of publication 8 April 2024; date of current version 17 April 2024. This work was supported in part by the STI 2030-Major Project under Grant 2021ZD0200402, in part by the Chinese Academy of Sciences (CAS) Project for Young Scientists in Basic Research under Grant YSBR-041, and in part by CAS International Partnership Program under Grant 173211KYSB20200021. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Aykut Koc. *(Corresponding author: Guyue Hu.)*

Yuan He is with the Laboratory of Brain Atlas and Brain-inspired Intelligence, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: heyuan2017@ia.ac.cn).

Guyue Hu is with the Information Materials and Intelligent Sensing Laboratory of Anhui Province, Anhui Provincial Key Laboratory of Multimodal Cognitive Computation, School of Artificial Intelligence, Anhui University, Hefei 230601, China (e-mail: guyue.hu@ahu.edu.cn).

Shan Yu is with the Laboratory of Brain Atlas and Brain-inspired Intelligence, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and with the Key Laboratory of Brain Cognition and Brain-inspired Intelligence Technology, Chinese Academy of Sciences, Beijing 100049, China, and also with the School of Future Technology, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: shan.yu@nlpr.ia.ac.cn).

Digital Object Identifier 10.1109/LSP.2024.3386115

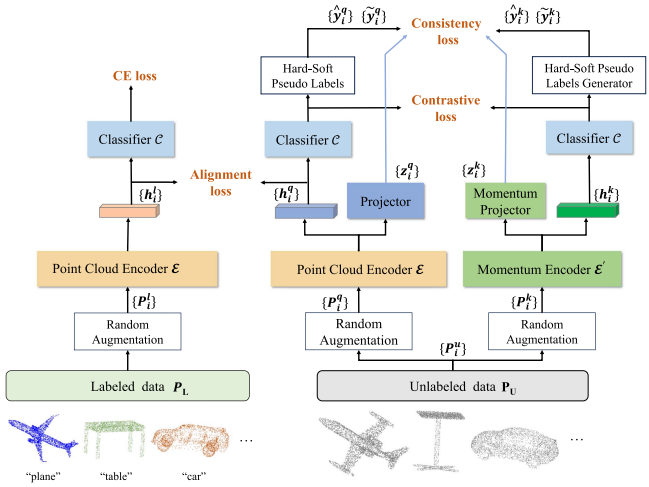

Fig. 1. Overview of the proposed HPSSL framework. The unlabeled data $\mathcal{P}_\mathcal{U}$ is independently augmented to P_i^q and P_i^k. They are passed through the point cloud encoder $\mathcal{E}(\cdot)$ and momentum encoder $\mathcal{E}'(\cdot)$ to generate their representations $\{h_i^q\}$ and $\{h_i^k\}$. Pseudo-labels are obtained through the shared classifier. The cross-entropy (CE) loss, contrastive loss, consistency loss, and alignment loss are integrated to constrain the network training.

labels guided semi-supervised learning, in which the supervised and unsupervised branches adaptively interact with each other via representation distribution alignment.

II. Method

A. Overview

Let $\mathcal{P}$ be a training set of point clouds and divide it into a labeled set and an unlabeled set. Denote $\mathcal{P}_\mathcal{L} = \{P_1^l, \ldots, P_L^l\} \in \mathcal{P}$ as labeled training set and $\{y_1, \ldots, y_L\}$ as their corresponding labels. Similarly, unlabeled training set is denoted as $\mathcal{P}_\mathcal{U} = \{P_1^u, \ldots, P_U^u\} \in \mathcal{P}$. Note that L is usually assumed much smaller than U, namely, $L \ll U$. To leverage the labeled but limited data and the numerous but unlabeled data, HPSSL learns features through a supervised learning branch, an unsupervised learning branch, and their interaction. Fig. 1 shows the overall architecture of HPSSL. The left branch is a standard supervised branch, and the right branch realizes the proposed unsupervised learning which is used to learn representations from massive unlabeled data. Besides, the two branches interact with each other during training.

B. Unsupervised Learning Branch.

1) Structure: As illustrated in Fig. 1, a batch of point clouds $\{P_i^u\}$ $(i = 1, 2, \ldots, B)$ are sampled from the unlabeled set $\mathcal{P}_\mathcal{U}$. Each sample P_i^u is transformed to a query sample P_i^q and a key sample P_i^k by performing different random augmentation. Then $\{P_i^q\}$ and $\{P_i^k\}$ are passed through the point cloud encoder $\mathcal{E}(\cdot)$ and momentum encoder $\mathcal{E}'(\cdot)$ separately to generate their representations $\{h_i^q\}$ and $\{h_i^k\}$. On the one hand, these representations are projected to a metric space with a linear projector $\mathcal{H}(\cdot)$ and the corresponding momentum projector $\mathcal{H}'(\cdot)$, respectively. The projected and normalized representations $\{z_i^q\}$ and $\{z_i^k\}$ are used for the following contrastive learning. On the other hand, the representations $\{h_i^q\}$ and $\{h_i^k\}$ are also fed to a shared classifier $\mathcal{C}$. The classifier mapped a representation to a C-dimensional classification vector, which describes the probabilities of C classes.

2) Hard-Soft Pseudo Labels: With the classifier's outputs, we can obtain pseudo labels of the unlabeled point cloud samples. Both hard pseudo labels and soft pseudo labels are generated. The hard pseudo labels are given by

$$\begin{cases} \hat{y}_i^q = \text{argmax}(\mathcal{C}(\mathcal{E}(P_i^q))) \\ \hat{y}_i^k = \text{argmax}(\mathcal{C}(\mathcal{E}'(P_i^k))) \end{cases} \tag{1}$$

where the argmax function converts a classification vector to a classification index. Hard pseudo labels are used in the subsequent consistency loss. In addition, the soft pseudo labels are generated by

$$\begin{cases} \widetilde{y}_i^q = \dfrac{topk(\mathcal{C}(\mathcal{E}(P_i^q)))}{\sum_c topk(\mathcal{C}(\mathcal{E}(P_i^q)))_c} \\ \widetilde{y}_i^k = \dfrac{topk(\mathcal{C}(\mathcal{E}'(P_i^k)))}{\sum_c topk(\mathcal{C}(\mathcal{E}'(P_i^k)))_c} \end{cases} \tag{2}$$

where the $topk$ function remains the k largest elements and sets the other elements to zero so that the soft pseudo label can focus on the high-confident part of the classification result. c is the index of class. The top k probabilities are normalized so that their sum is one. Empirically, k is set as 4, which is much smaller than C. The soft pseudo labels are used in the following contrastive learning, which provides class awareness of the similarity measurement in the loss calculation.

C. Class-Aware Contrastive Learning

Contrastive learning is employed to learn features from unlabeled data by reducing the distance between positive sample pairs while enlarging the distance between negative sample pairs in the feature space. InfoNCE loss [25] is widely used in contrastive learning. However, it only considers the contrast between instances but ignores class information. We propose two strategies to fully utilize class information of pseudo labels to guide class-aware contrastive learning.

1) Hard Pseudo Labels Guided Contrastive (HPLC) Loss: With the hard pseudo labels obtained by (1), for each query sample i we can select the key samples with the same hard pseudo label from $\mathcal{D}$, to form a positive set $\mathcal{A}_i = \{a | \hat{y}_i^q = \hat{y}_a^k\}$. Note that the key samples whose hard pseudo label's confidence is lower than a threshold β are excluded from the positive set. The HPLC loss is given by

$$\mathcal{L}_{\text{HPLC}} = -\frac{1}{B}\sum_{i=1}^{B} \log \frac{\sum_{a \in A_i} \exp(\boldsymbol{z}_i^q \cdot \boldsymbol{z}_a^k / \tau)}{\sum_{n=1}^{K} \exp(\boldsymbol{z}_i^q \cdot \boldsymbol{z}_n^k / \tau)} \tag{3}$$

2) Hard-Soft Pseudo Labels Guided Contrastive (HSPLC) Loss: Considering the inherent inaccuracies and sensitivity to label noise in hard labels, we propose to use the soft pseudo labels given by (2) to further improve semi-supervised learning. The HSPLC loss is calculated by

$$\mathcal{L}_{\text{HSPLC}} = -\frac{1}{B}\sum_{i=1}^{B} \log \frac{\sum_{a=1}^{K} \widetilde{y}_i^q \cdot \widetilde{y}_a^k \exp(\boldsymbol{z}_i^q \cdot \boldsymbol{z}_a^k / \tau)}{\sum_{n=1}^{K} \exp(\boldsymbol{z}_i^q \cdot \boldsymbol{z}_n^k / \tau)} \tag{4}$$

Algorithm 1: Semi-Supervised Training.

Input: labeled data $\{\mathcal{P}_\mathcal{L}\}$, unlabeled data $\{\mathcal{P}_\mathcal{U}\}$, model module $\{\mathcal{E}, \mathcal{C}, \mathcal{H}, \mathcal{E}^{'}, \mathcal{H}^{'}\}$, sample dictionary $\mathcal{D}$
M: number of optimization steps
for $m = 1$ **to** M **do**
Sample mini-batch of labeled samples $P^l \leftarrow \mathcal{P}_\mathcal{L}$ and unlabeled samples $P^u \leftarrow \mathcal{P}_\mathcal{U}$;
Perform random augmentations twice
$P^q \leftarrow Aug_1(P^u), P^k \leftarrow Aug_2(P^u)$;
Perform forward propagation
$h^l \leftarrow \mathcal{E}(P^l), h^q \leftarrow \mathcal{E}(P^q)$
$z^q \leftarrow \mathcal{H}(\mathcal{E}(P^q))$, $z^k \leftarrow \mathcal{H}^{'}(\mathcal{E}^{'}(P^k))$;
Compute the Cross-entropy loss $\mathcal{L}_{\text{CE}} \leftarrow CE(\mathcal{C}(\mathcal{E}(P^l)), y^l)$;
Computer hard pseudo labels $\hat{y}^q, \hat{y}^k$ by Eq. 1 ;
Computer soft pseudo labels $\widetilde{y}^q, \widetilde{y}^k$ by Eq. 2 ;
Update the dictionary $\mathcal{D} \leftarrow z^k, \hat{y}^k, \widetilde{y}^k$;
Compute the contrastive loss $\mathcal{L}_{\text{HPLC}}$ by Eq. 3 or $\mathcal{L}_{\text{HSPLC}}$ by Eq. 4 ;
Compute the consistency loss $\mathcal{L}_{\text{CONS}}$ by Eq. 5 ;
Compute the alignment loss $\mathcal{L}_{\text{CONS}} \leftarrow \text{MMD}(z^q, z^k)$;
Update $\{\mathcal{E}, \mathcal{C}, \mathcal{H}\}$ using the total loss $\mathcal{L} = \mathcal{L}_{\text{CE}} + \lambda_1 \mathcal{L}_{\text{CONT}} + \lambda_2 \mathcal{L}_{\text{CONS}} + \lambda_3 \mathcal{L}_{\text{MMD}}$;
Update $\{\mathcal{E}^{'}, \mathcal{H}^{'}\}$ by EMA.
end

Unlike the HPLC loss, the HSPLC loss incorporates weights based on the probability of label consistency to regulate the loss. The utilization of soft pseudo labels serves the dual purpose of discarding unreliable labels while enhancing the feature similarity for positive samples.

D. Prediction Consistency Regularization

In the unsupervised branch, the model is required to produce consistent predictions when different perturbed versions of the same point cloud are fed into the network. To enforce this consistency, we utilize the hard pseudo labels provided by (1) to form a consistency loss

$$\mathcal{L}_{\text{CONS}} = \frac{1}{B} \sum_{i=1}^{B} \mathbb{1}(\hat{y}_i^k \geq \beta) \mathbf{H}(y_i^k, y_i^q) \tag{5}$$

where $\mathbb{1}$ is the indicator function, $\mathbf{H}$ is the cross-entropy loss. β is a confidence threshold. $y_i^q = \mathcal{C}(\mathcal{E}(P_i^k))$ and $y_i^k = \mathcal{C}(\mathcal{E}'(P_i^k))$ are predicted class distribution of the corresponding query and key sample. This loss guides the model to capture the invariant and consistent features under various changes of a point cloud instance.

E. Representation Alignment

Considering the interaction between the supervised and unsupervised branches, we expect the feature distributions of the two branches to be aligned. However, since labeled and unlabeled data are carried out by supervised and unsupervised optimization objectives respectively in the semi-supervised scheme, there is a misalignment between their feature distributions. Inspired by unsupervised domain adaptation techniques addressing cross-domain distribution discrepancies, we leverage Maximum Mean Discrepancy (MMD) [26], [27] to estimate the distribution discrepancy. By reducing the MMD loss $\mathcal{L}_{\text{MMD}}$, we gradually align the distributions of representations, thereby enhancing the performance of the classification task.

TABLE I
POINT CLOUD CLASSIFICATION ACCURACY (%) ON MODELNET40 WITH 5%, 10%, 20%, 40% LABELS OF TRAINING SET

Method	5%	10%	20%	40%
Sup-only	68.4	77.2	81.9	85.5
Pseudolabels [34]	69.1	78.7	82.6	86.4
S4L [16]	70.6	78.9	82.7	87.1
STL [20]	73.5	80.1	81.7	87.3
FixMatch [35]	74.2	80.9	81.8	87.3
ConMatch [18]	74.7	80.8	82.3	87.4
MarginMatch [36]	75.3	81.6	82.5	87.2
Ours	76.4	81.8	83.2	87.9

F. Total Loss and Training Procedure

The total loss is the weighted sum of the four loss terms,

$$\mathcal{L} = \mathcal{L}_{\text{CE}} + \lambda_1 \mathcal{L}_{\text{CONT}} + \lambda_2 \mathcal{L}_{\text{CONS}} + \lambda_3 \mathcal{L}_{\text{MMD}} \tag{6}$$

where $\mathcal{L}_{\text{CE}}$ denotes cross-entropy loss used in the supervised learning branch. $\mathcal{L}_{\text{CONT}}$ is the contrastive loss that can be implemented with either $\mathcal{L}_{\text{HPLC}}$ or $\mathcal{L}_{\text{HSPLC}}$. λ_1, λ_2 and λ_3 are non-negative weights. The training procedure of the proposed SSL framework is presented in Algorithm 1.

III. EXPERIMENTS

A. Experimental Setup

1) Datasets: Our experiments are performed on three main point cloud classification datasets. ModelNet40 dataset [28] consists of 9843 train objects and 2468 test objects from 40 categories. ShapeNet dataset [29] has 51,190 objects from 55 categories. ModelNet40-C dataset [30] is a comprehensive benchmark dataset on corruption robustness of 3D point cloud recognition, consisting of 15 corruption types with 5 severity levels. Following the conventional practice in SSL, we split the training set into labeled and unlabeled data proportionally.

2) Implementation: All comparisons with semi-supervised methods are made under the same setting to be fair. The encoder can be implemented using PointNet [31], DGCNN [32], and other techniques, such as [33]. On three datasets, we sample 1024 points from each sample for training and testing. Following MoCo [25], the momentum encoder is updated by the exponential moving average of the encoder's weights. During training, we first train the supervised part for 10 epochs and then train the whole framework.

B. Experimental Results

1) Shape Recognition: First, we evaluate our semi-supervised framework on the ModelNet40 dataset [28] for point cloud classification. We use PointNet [31] as the backbone network and supervised baseline (*Sup-only*). Table I shows the results of the shape classification. Our method

TABLE II
POINT CLOUD CLASSIFICATION ACCURACY (%) ON SHAPENET WITH 1%, 5%, 10% LABELS OF TRAINING SET

Method	PointNet			DGCNN		
	1%	5%	10%	1%	5%	10%
Sup-only	63.5	72.3	77.4	69.7	78.6	82.1
Pseudolabels [34]	66.2	72.6	78.2	71.4	79.8	82.7
S4L [16]	65.8	72.4	77.9	71.8	79.4	82.5
FixMatch [35]	66.9	73.6	79.3	73.6	79.5	83.2
ConMatch [18]	67.2	74.5	78.8	73.3	79.9	83.5
MarginMatch [36]	67.8	74.2	79.5	73.5	80.3	83.4
Ours	68.4	75.9	79.7	74.1	80.2	83.9

TABLE III
POINT CLOUD CLASSIFICATION ACCURACY (%) ON MODELNET40-C WITH 5%, 10%, 20%, 40% LABELS OF TRAINING SET

Method	5%	10%	20%	40%
Sup-only	50.3	57.6	62.8	66.3
Pseudolabels [34]	53.5	58.3	63.5	67.1
S4L [16]	54.1	59.7	64.3	68.2
STL [20]	57.4	61.8	65.2	67.9
FixMatch [35]	59.2	62.5	65.1	68.3
ConMatch [18]	59.7	63.9	65.6	68.2
MarginMatch [36]	59.4	63.7	66.1	68.4
Ours	59.6	64.2	66.3	68.5

achieves state-of-the-art performance and benefits more when fewer labels are used. Compared to semi-supervised methods (Pseudolabels [34], S4L [16], STL [20], FixMatch [35], ConMatch [18], MarginMatch [36]), HPSSL outperforms these strong baselines with margins of 0.2–6%. It's worth noting that the improvement of our method to other methods is more significant with fewer labels. Further, we conduct experiments of point cloud classification on ShpaeNet [29] which is a larger dataset. The previously described has once again been confirmed. And that, to explore the compatibility of the architectures, we use DGCNN as the encoder except for PointNet. As shown in Table II, HPSSL with PointNet or DGCNN is better than supervised baselines, even superior to other SSL methods described above, demonstrating that our proposed framework can be generalized to various encoders. As expected, the gain obtained by our method decreases as the amount of labels increases. Moreover, to validate its effectiveness on corruption data, we compare HPSSL with other semi-supervised methods on ModelNet40-C [30]. As depicted in Table III, our HPSSL consistently achieves the best performance using 10%, 20% and 40% labels of training point clouds. For example, our HPSSL respectively outperforms classical semi-supervised methods S4L [16] and Pseudolabels [34] with large margins of 5.5% and 6.1% when using only 5% labeled data, which clearly demonstrates the superiority of the proposed HPSSL.

2) Ablation Study: We then investigate the effectiveness of each component in HPSSL: class-aware contrastive learning(*Ccl.*), representation alignment(*Ra.*), and prediction consistency regularization (*Pc.*). In Table IV, the model *Sup.+Csl.* is trained with a cross-entropy loss for labeled data and an InfoNCE loss for unlabeled data. Instead of InfoNCE loss, *Sup.+Ccl.+Hp.* and *Sup.+Ccl.+Hsp.* exploit unlabeled data through HPLC and HSPLC loss, respectively. They improve 2.8% and 3.9% over *Sup.+Csl.*, indicating that hard-soft pseudo labels can boost the performance of our method. *Sup.+Ccl.+Ra.* boosts 1.9% and 1.2% over *Sup.+Ccl.+Hp.* and *Sup.+Ccl.+Hsp.* by applying MMD to align feature distributions of labeled and unlabeled data. What's more, *Pc.* further improves the performance. These experimental results demonstrate that each of these components plays an important role in our semi-supervised learning.

TABLE IV
ABLATION STUDY ON SEMI-SUPERVISED LEARNING METHODS ON MODELNET40 WITH 5% LABELS

Method	Accuracy
Sup-only	68.4
Sup. + Csl.	70.6
Sup. + Ccl. + Hp.	73.8
Sup. + Ccl. + Hsp.	74.5
Sup. + Ccl. + Hsp. + Ra.	75.7
Sup. + Ccl. + Hsp. + Ra. + Pc.	76.4
Ours	76.4

(a) Sup. (b) Sup.+Ccl. (c) Sup.+Ccl.+Ra.+Pc.(Ours)

Fig. 2. t-SNE visualization of representations learned by *Sup.*, *Sup. + Ccl.*and *Sup. + Ccl. + Ra. + Pc.* (ours) on ModelNet10 dataset. (a) *Sup.* is trained with the supervised objective for the labeled data. (b) *Sup. + Ccl.* is trained through the supervised objective and class-aware contrastive loss. (c) *Sup. + Ccl. + Ra. + Pc.* is our HPSSL method. The squares with black denote the labeled data, and the others are unlabeled data. Note that different colors denote different classes.

3) The T-SNE Visualization: To further explore the mechanism of HPSSL, we visualize the feature distributions of the labeled and unlabeled point clouds on ModelNet10 [28] training dataset with t-SNE in Fig. 2. From Fig. 2, We can find that the model *Sup.* trained with only supervised objects is hard to distinguish the decision boundaries of unlabeled point clouds. Compared to *Sup.*, the model *Sup.+Ccl.* presents tighter distributions, which benefit from class-aware contrastive learning. However, long-tail distributions still exist for unlabeled samples. our HPSSL with *Sup.+Ccl.+Ra.+Pc.* makes decision boundaries more clearer. Overall, the comparison results further validate the effectiveness of HPSSL. Our proposed algorithm, which combines semi-supervised learning with contrastive learning, demonstrates the capability to guide our model in acquiring more discriminative features.

IV. CONCLUSION

In this paper, we propose a novel semi-supervised learning framework for 3D point cloud classification. Our approach is designed to train the model with significantly fewer annotations, alleviating the burden of annotating extensive point cloud data. Through extensive experiments, we validate the efficacy of the proposed pseudo label-guided class-aware contrastive loss, prediction consistency strategy, and representation alignment, demonstrating the superior performance of our method. Notably, we observed a positive correlation between the scarcity of labels and the gains obtained by our method.

REFERENCES

[1] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun, "Deep learning for 3D point clouds: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 12, pp. 4338–4364, Dec. 2021.

[2] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-view 3D object detection network for autonomous driving," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1907–1915.

[3] E. Nezhadarya, E. Taghavi, R. Razani, B. Liu, and J. Luo, "Adaptive hierarchical down-sampling for point cloud classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 12956–12964.

[4] Y. Wang and J. M. Solomon, "Deep closest point: Learning representations for point cloud registration," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 3523–3532.

[5] S. Wang, F. Qin, Y. Tong, X. Shang, and Z. Zhang, "Probabilistic boundary-guided point cloud primitive segmentation network," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 2529413.

[6] Z. Gojcic, C. Zhou, J. D. Wegner, L. J. Guibas, and T. Birdal, "Learning multiview 3D point cloud registration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 1759–1769.

[7] H. Jiang, F. Yan, J. Cai, J. Zheng, and J. Xiao, "End-to-end 3D point cloud instance segmentation without detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 12796–12805.

[8] J. Li, B. M. Chen, and G. H. Lee, "SO-net: Self-organizing network for point cloud analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 9397–9406.

[9] J. Hou, "Permuted sparse representation for 3D point clouds," *IEEE Signal Process. Lett.*, vol. 26, no. 12, pp. 1847–1851, Dec. 2019.

[10] S. Shi, X. Wang, and H. Li, "PointRCNN: 3D object proposal generation and detection from point cloud," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 770–779.

[11] S. Liu et al., "Deep learning-based point cloud analysis," in *3D Point Cloud Analysis*. Berlin, Germany: Springer, 2021, pp. 53–86.

[12] S. Zhang, S. Cui, and Z. Ding, "Hypergraph spectral clustering for point cloud segmentation," *IEEE Signal Process. Lett.*, vol. 27, pp. 1655–1659, 2020.

[13] R. Sheikhpour et al., "A survey on semi-supervised feature selection methods," *Pattern Recognit.*, vol. 64, pp. 141–158, 2017.

[14] Z. Zhang, Y. Li, Z. Zhang, C. Zhang, and M. Gao, "Adaptive matrix sketching and clustering for semisupervised incremental learning," *IEEE Signal Process. Lett.*, vol. 25, no. 7, pp. 1069–1073, Jul. 2018.

[15] N. Gu, M. Fan, H. Qiao, and B. Zhang, "Discriminative sparsity preserving projections for semi-supervised dimensionality reduction," *IEEE Signal Process. Lett.*, vol. 19, no. 7, pp. 391–394, Jul. 2012.

[16] X. Zhai, A. Oliver, A. Kolesnikov, and L. Beyer, "S4L: Self-supervised semi-supervised learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 1476–1485.

[17] T. Chen et al., "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1597–1607.

[18] J. Kim et al., "ConMatch: Semi-supervised learning with confidence-guided consistency regularization," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2022, pp. 674–690.

[19] S. Deng, Q. Dong, B. Liu, and Z. Hu, "Superpoint-guided semi-supervised semantic segmentation of 3D point clouds," in *Proc. Int. Conf. Robot. Automat.*, 2022, pp. 9214–9220.

[20] H. Li et al., "Semi-supervised point cloud segmentation using self-training with label confidence prediction," *Neurocomputing*, vol. 437, pp. 227–237, 2021.

[21] L. Jiang et al., "Guided point contrastive learning for semi-supervised point cloud semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 6423–6432.

[22] S. Xie et al., "PointContrast: Unsupervised pre-training for 3D point cloud understanding," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2020, pp. 574–591.

[23] S. Huang, Y. Xie, S.-C. Zhu, and Y. Zhu, "Spatio-temporal self-supervised representation learning for 3D point clouds," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 6535–6545.

[24] M. Afham, I. Dissanayake, D. Dissanayake, A. Dharmasiri, K. Thilakarathna, and R. Rodrigo, "CrossPoint: Self-supervised cross-modal contrastive learning for 3D point cloud understanding," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 9902–9912.

[25] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 9729–9738.

[26] A. Gretton et al., "A kernel two-sample test," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 723–773, 2012.

[27] Y. Zhu, F. Zhuang, and D. Wang, "Aligning domain-specific distribution and classifier for cross-domain classification from multiple sources," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 5989–5996.

[28] Z. Wu et al., "3D ShapeNets: A deep representation for volumetric shapes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1912–1920.

[29] A. X. Chang et al., "ShapeNet: An information-rich 3D model repository," 2015, *arXiv:1512.03012*.

[30] J. Sun et al., "Benchmarking robustness of 3D point cloud recognition against common corruptions," 2022, *arXiv:2201.12296*.

[31] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 77–85.

[32] Y. Wang et al., "Dynamic graph CNN for learning on point clouds," *ACM Trans. Graph.*, vol. 38, no. 5, pp. 1–12, 2019.

[33] K. H. R. Chan et al., "ReduNet: A white-box deep network from the principle of maximizing rate reduction," *J. Mach. Learn. Res.*, vol. 23, 2022, Art. no. 114.

[34] D.-H. Lee et al., "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Proc. Workshop Challenges Representation Learn.*, 2013, no. 896, pp. 1–6.

[35] K. Sohn et al., "FixMatch: Simplifying semi-supervised learning with consistency and confidence," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 596–608.

[36] T. Sosea and C. Caragea, "MarginMatch: Improving semi-supervised learning with pseudo-margins," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 15773–15782.