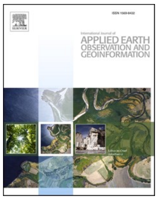
International Journal of Applied Earth Observations and Geoinformation 115 (2022) 103129

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag



WSPointNet: A multi-branch weakly supervised learning network for semantic segmentation of large-scale mobile laser scanning point clouds

Xiangda Lei [a], Haiyan Guan [a,*], Lingfei Ma [b,*], Yongtao Yu [c], Zhen Dong [d], Kyle Gao [e], Mahmoud Reza Delavar [f], Jonathan Li [e]

[a] School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China
[b] School of Statistics and Mathematics, Central University of Finance and Economics, Beijing 102206, China
[c] Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huaian 223003, China
[d] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
[e] Department of Geography and Environmental Management and Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[f] College of Engineering, University of Tehran (UT), Tehran 1439951154, Iran

ARTICLE INFO

Keywords:
MLS point cloud
Semantic segmentation
Weakly supervised learning
Ensemble prediction

ABSTRACT

Semantic segmentation of large-scale mobile laser scanning (MLS) point clouds is essential for urban scene understanding. However, most of the existing semantic segmentation methods require a large quantity of labeled data, which are labor-intensive and time-consuming. To this end, we propose a multi-branch weakly supervised learning network (WSPointNet) to solve this challenge. Our method includes a basic weakly supervised framework and a multi-branch weakly supervised module. With input point clouds and few labels, the basic weakly supervised framework outputs the prediction values of the input point clouds and the underlying supervised signals of the whole network. Next, the multi-branch weakly supervised module explores the potential information of the unlabeled and labeled points while preventing model over-fitting. Concretely, the module includes an ensemble prediction constraint branch, a contrast-guided entropy regularization branch, and an adaptive pseudo-label learning branch. The ensemble prediction constraint branch aims to enhance the prediction stability of the point cloud. The contrast-guided entropy regularization branch is proposed to prevent model over-fitting by comparing the ensemble prediction labels with the current prediction labels. The adaptive pseudo-label learning branch provides efficient and adaptive supervised signals for model training by the consistency cost and ensemble prediction. Extensive experiments conducted on two MLS benchmarks showed that our WSPointNet achieved a promising semantic segmentation performance with sparse annotated points. For the public Toronto3D dataset, with only 0.1% labeled points, our WSPointNet obtained an overall accuracy of 96.76% and a mIoU of 78.96%, which outperformed most of comparative fully supervised methods.

1. Introduction

Mobile laser scanning (MLS) point clouds, due to the characteristics of large scale, high density, accurate geo-reference, and fine-grained three-dimensional (3D) view of objects and surroundings, have been employed in a wide range of urban applications, such as high-definition (HD) mapping, transportation infrastructure management and inventory, 3D model reconstruction, and autonomous driving (Han et al., 2021; Tao et al., 2020; Luo et al., 2020). Although many efforts have been made to efficiently and effectively process discrete, irregularly distributed, and severely occluded MLS point clouds, it is still a challenge to accurately identify typical objects in complex road environments and assign a semantic label to each point (Han et al., 2021; Liang et al., 2021).

In recent years, deep learning, which has received much attention in photogrammetry and remote sensing, has shown great potential for 3D point cloud data processing (Guo et al., 2021; Luo et al., 2022). For example, PointNet, initially proposed by Qi et al. (2017a), directly performed semantic segmentation on point clouds. Subsequently, inspired by the PointNet model, more relevant frameworks were developed for point cloud semantic segmentation (Qi et al., 2017b; Thomas et al., 2019; Wang et al., 2019a; Hu et al., 2020). Although the

* Corresponding authors.
E-mail addresses: leixd@nuist.edu.cn (X. Lei), guanhy.nj@nuist.edu.cn (H. Guan), l53ma@cufe.edu.cn (L. Ma), allennessy@hyit.edu.cn (Y. Yu), dongzhenwhu@whu.edu.cn (Z. Dong), y56gao@uwaterloo.ca (K. Gao), mdelavar@ut.ac.ir (M. Reza Delavar), junli@uwaterloo.ca (J. Li).

https://doi.org/10.1016/j.jag.2022.103129
Received 26 September 2022; Received in revised form 10 November 2022; Accepted 22 November 2022
Available online 28 November 2022

aforementioned approaches obtained state-of-the-art performances on different public benchmark datasets, they all relied on a large quantity of qualified annotated data for training. Notably, labeling voluminous MLS point clouds in complicated scenarios is very time-consuming and labor-intensive (Luo et al., 2018).

Weakly supervised learning (WSL) methods have been increasingly developed to effectively classify MLS point clouds into different classes of interest with fewer labeled samples (Zhang et al., 2021b). Xu and Lee (2020) proposed a multi-branch WSL method, which, by using 10 % labeled points for training, achieved a comparable point cloud semantic segmentation performance to most fully supervised methods. However, labeling 10 % points of a point cloud is still a heavy workload for MLS point clouds. To implicitly increase the amount of available supervised signals, Hu et al. (2021) proposed a semantic query network (SQN), which investigated sparse labels and semantic similarities between neighboring points. Although the SQN method achieved competitive point cloud semantic segmentation accuracies with only 0.1 % points labeled from the entire point cloud, it still inadequately explored the informative features of all unlabeled points. Wang and Yao (2022) investigated the features of unlabeled points by exploiting ensemble prediction constraint, entropy regularization, and online soft pseudo-labeling. Although this method achieved desirable semantic segmentation performance on airborne laser scanning (ALS) datasets with 0.1 % labeled points, it suffered from the following issues: 1) the overlaps of predicted classes were penalized by minimizing the entropy values of all unlabeled points, which led to overconfidence predictions, i.e. model over-fitting, and poor point cloud semantic segmentation performance; 2) the weight of a pseudo-label was quantified by an entropy value, which only indicated the degree of uncertainty in the current predictions, and was unable to solve the model over-fitting issue.

To fully utilize the potential information of unlabeled points while avoiding the aforementioned problems, we propose a multi-branch weakly supervised module, consisting of an ensemble prediction constraint (EPC) branch, a contrast-guided entropy regularization (CG-ER) branch, and an adaptive pseudo-label learning (A-PL) branch. The EPC (Wang and Yao, 2022) branch aims to improve the prediction stability of point clouds by generating ensemble predictions (i.e., comparison samples) and minimizing consistency costs (i.e. the variances between the ensemble prediction distributions and the current prediction distributions). The CG-ER branch divides unlabeled points into confidence and non-confidence prediction point sets by contrasting the ensemble prediction labels with the current prediction labels. Correspondingly, entropy maximization is performed on the non-confidence prediction point sets to prevent from model over-fitting. Finally, the A-PL branch calculates the pseudo-label weights of the unlabeled points via the consistency cost, ensuring that the network focuses more on the confidence prediction points with high pseudo-label weights and ultimately augments more supervised signals for model training while preventing from model over-fitting. In addition, the RandLA-Net is taken as the basic weakly supervised framework, on which feature extraction and point cloud prediction are performed to provide underlying supervised signals for model training. Therefore, we integrate the basic weakly supervised framework with the multi-branch weakly supervised module for MLS point cloud semantic segmentation, terming the network as WSPointNet in this study.

Our major contributions can be summarized below:

- The WSPointNet, a network that integrates a RandLA-Net framework and a multi-branch weakly supervised module, is proposed for large-scale MLS point cloud semantic segmentation tasks with sparse and randomly-sampled labeled points, achieving a competitive point cloud semantic segmentation accuracy over supervised methods.
- A multi-branch weakly supervised module, which consists of an ensemble prediction constraint approach, a contrast-guided entropy regularization approach, and an adaptive pseudo-label learning approach, is proposed for fully exploring the potential information of unlabeled points while preventing from model over-fitting, and can be designed as a plug-in for flexible integration into other mainstream frameworks.

The remainder of the paper is organized as follows. In Section 2, we systematically review fully supervised and weakly supervised deep learning methods for point cloud semantic segmentation. The proposed method is completely explained in Section 3. Section 4 presents two MLS datasets, related experiments, and experimental analyses to validate the effectiveness of the proposed method. In Section 5, the concluding remarks are summarized and the suggested future works are presented.

2. Related work

2.1. Semantic segmentation of point clouds

In contrast of traditional point cloud semantic segmentation methods that relied mainly on manually designed features, deep learning-based semantic segmentation methods have made an increasing number of advancements (Guo et al., 2021). Early deep learning networks required a regular input data format, such as images (Feng et al., 2018; Boulch et al., 2018) or voxels (Wang et al., 2018; Meng et al., 2019). Therefore, data conversion was an essential prerequisite for processing irregular, discrete, and unstructured raw point clouds, resulting in a loss of 3D features and an increase of memory redundancy.

PointNet, a pioneering network, was designed to directly process point clouds (Qi et al., 2017a). The PointNet used multi-layer perceptron (MLP) and max-pooling operations to extract point-wise and global features for point cloud semantic segmentation. Although the PointNet achieved high computational efficiency, it was incapable of capturing local features of a point cloud, leading to ineffective semantic segmentation in complex scenarios. PointNet++ (Qi et al., 2017b), which was then developed as an improved PointNet framework, encoded multi-scale features by a hierarchical network structure and extracted local features from neighboring points. Inspired by the PointNet and PointNet++, different local feature extraction strategies were adopted to improve point cloud semantic segmentation accuracies, including neighborhood feature pooling in SoNet (Li et al., 2018) and PointWeb (Zhao et al., 2019), graph convolutions in DGCNN (Wang et al., 2019b) and GAN (Wang et al., 2019a), kernel function in KPConv (Thomas et al., 2019), PointConv(Wu et al., 2019), and A-CNN (Komarichev et al., 2019).

Although the abovementioned approaches have achieved great successes in point cloud semantic segmentation and object recognition, most of them were capable of dealing with only small-scale point clouds. However, large-scale point clouds were often segmented into a set of small-scale point blocks via blocking operations, leading to an increase of point cloud pre-processing time and damages in geometrical completeness (Du et al., 2021). RandLA-Net, proposed by Hu et al. (2020), employed a computation and memory efficient random sampling strategy for point cloud down-sampling and a point cloud local feature aggregation (LFA) module for increasing the point receptive field, which contributed to the interpretation of large-scale point clouds. Although these methods achieved desirable semantic segmentation performances on different publicly-available benchmarks, they all relied heavily on a large quantity of high-quality labels for training. However, a point cloud semantic annotation task is very laborious and time-consuming, especially for a large data volume of MLS point clouds in complex scenarios. Therefore, to overcome the deficiency of labeling point clouds, weakly supervised semantic learning methods that rely on only a small quantity of annotated points have gained much attention.

2.2. Weakly-supervised semantic segmentation

In terms of annotation type, weakly supervised learning approaches are generally grouped into two categories: limited indirect labeling and limited point-level labeling. The former converts a point cloud into images (Wang et al., 2020), sub-clouds (Wei et al., 2020), or segments (Tao et al., 2020) before performing labeling operations. Specifically, Wang et al. (2020) transformed a point cloud into 2D images at different viewpoints, and then labeled and input the images into a deep graph convolutional network for point cloud semantic segmentation. Wei et al. (2020) first annotated subcloud-level labels, and then input them into a multi-path region mining module (MPRM) to yield pseudo-labels for point cloud semantic segmentation. Tao et al. (2020) generated pseudo-labels by using a Segmented Grouping (SegGroup) network and a limited segment-level point cloud labeling method. All of these methods required a pre-processing procedure for data annotation. Although these data annotation methods reduce the labeling workload without noticeably deteriorating point cloud semantic segmentation accuracies, they are still time-consuming and costly for large-scale MLS data.

The limited point-level labeling approaches train the models with only a limited number of semantically labeled points. Zhang et al. (2021a) generated pseudo-labels by pre-training the model in Lab color space and then fine-tuned the pre-trained model by sparsely labeled points. Xie et al. (2020) utilized sparsely labeled points to fine-tune the pre-trained model generated by a PointInfoNCE loss, which encouraged consistently distributed positive point-pairs and penalized negative point-pairs for self-supervised pre-training. Additionally, Hou et al. (2021) first employed contrastive scene context to obtain multiple regions of contrast loss for model training, and actively fine-tuned the model with sparsely labeled points. Although the aforementioned methods performed point cloud semantic segmentation tasks without data conversion, they all required point clouds for pre-training. However, the point cloud pre-training requires sufficient data sources and extra training time, which negatively affects the applicability of these semantic segmentation methods.

Hu et al. (2021) proposed an SQN network to increase the amount of available supervised signals for point cloud semantic segmentation with only sparsely labeled points. However, this method under-utilized the features of all unlabeled points, resulting in a less accurate characterization of all the class features. To fully exploit the features of all unlabeled points, multi-branch weakly supervised methods were developed (Jiang et al. 2021; Wang and Yao, 2022; Xu and Lee, 2020; Zhang et al., 2021b). Xu and Lee (2020) first employed incomplete supervision with sparsely labeled points to provide the underlying supervised signals for model training, and fully leveraged the features of unlabeled points through inaccurate supervised, self-supervised, and smoothing constrained branches. Jiang et al. (2021) proposed a contrast loss guided by unlabeled point information to improve model generalization and feature representation in a weakly supervised manner. Among them, the self-supervised branch was built on the augmented samples generated by a rigid random transformation of training points. To generate more effectively augmented samples, Zhang et al. (2021b) used a perturbation branch, including rigid transformation and feature attention, to enforce the consistency constraint. In addition, an incomplete supervision module and a context-aware module were developed to train the model with the features of labeled and unlabeled points. However, the consistency constraint by augmenting raw training points reduced network training efficiency and increased memory burden. Subsequently, Wang and Yao (2022) proposed an ensemble prediction constraint to improve the prediction stability of point clouds by constraining the variances between the ensemble prediction values and the current prediction values. The process of generating ensemble predictions was not involved in network training, and the two predictions, i.e., the ensemble prediction and the current prediction, were generated in one forward propagation, which resulted in higher training efficiency, compared with the augmented sample approaches. In addition, entropy minimization was used to punish the predicted class overlaps for improving the confidence of point predictions. To address pseudo-label noise and low training efficiency in the traditional pseudo-label learning methods (Tao et al., 2020; Wei et al., 2020; Zhang et al., 2021a), Wang and Yao (2022) proposed an online soft pseudo-label strategy, which extended supervised sources in an efficient and non-parametric manner. Although the method achieved the desired semantic segmentation performance on ALS datasets, it still suffered from model over-fitting caused by minimizing the entropy values of all unlabeled points and using only the current predicted entropy values as pseudo-label weights.

3. Method

The WSPointNet is proposed to fully exploit the informative features of both sparsely labeled points and unlabeled points for semantic segmentation of large-scale MLS point clouds. Fig. 1 shows the pipeline of our WSPointNet, which consists of two main components: a basic weakly supervised framework, consisting of the backbone network and an incomplete supervision branch, and a multi-branch weakly supervised module, consisting of an ensemble prediction constraint branch, a contrast-guided entropy regularization branch, and an adaptive pseudo-label learning branch.

As shown in Fig. 1, the input data is first fed into the backbone network to extract the features of points and obtain their predicted probability distributions, referred to as the prediction distributions in this study. Then, ensemble predictions, i.e., the comparison samples obtained by recording the point cloud prediction distributions during training, are then input into different branches along with the current predictions to calculate the total loss. Specifically, the current prediction distributions of labeled points are fed into the incomplete supervision branch to calculate cross-entropy loss, which provides our network with underlying supervised signals. From unlabeled points, the EPC branch is used to improve the prediction stability of point clouds. The CG-ER branch guides the entropy regularization by contrasting the ensemble prediction labels with the current prediction labels. Points with inconsistent results are entropy-maximized to increase their uncertainties and prevent from model over-fitting. The A-PL branch provides additional and adaptive supervised signals for model training by calculating the pseudo-label weights of unlabeled points regarding the consistency cost. Finally, the losses of the aforementioned four branches are combined to obtain a total loss function for network training. The proposed WSPointNet will be detailed in the following sub-sections.

3.1. Basic weakly supervised framework

The basic weakly supervised framework, including the backbone network and an incomplete supervised branch, aims to provide salient features and underlying supervised signals. To obtain the salient and abstractive features, we use the RandLA-Net as our backbone network. Fig. 2 shows the RandLA-Net, which is a typical encoder-decoder architecture (Hu et al., 2020). The encoder contains random sampling (RS) and local feature aggregation (LFA) modules. The decoder is composed of up-sampling and MLP layers. Finally, the fully connected layers are used for point cloud classification. The RandLA-Net network obtains the effective features and the predicted distributions of input points. To provide the underlying supervised signals as well as reducing the effect of category imbalance, we apply a smoother square-root weighted cross-entropy loss function to calculate the loss of the labeled points, i.e., the incomplete supervised loss L_{se}, which is calculated as follows:

$$W_{sqrt} = \frac{1}{\sqrt{N_c \sum_{i=1}^{K} \frac{1}{N_i}}} \tag{1}$$

$$L_{se} = -\frac{1}{|P_l|} \sum_{i}^{|P_l|} W_{sqrt,i} \sum_{c}^{K} y_{ic} \log p_{ic} \tag{2}$$

where W_{sqrt} is the class weight. N_c is the number of points labeled as class c. P_l is the set of labeled points. $|\cdot|$ is the set cardinality. K is the number of classes. p_{ic} and y_{ic} are the prediction label and ground-truth label of

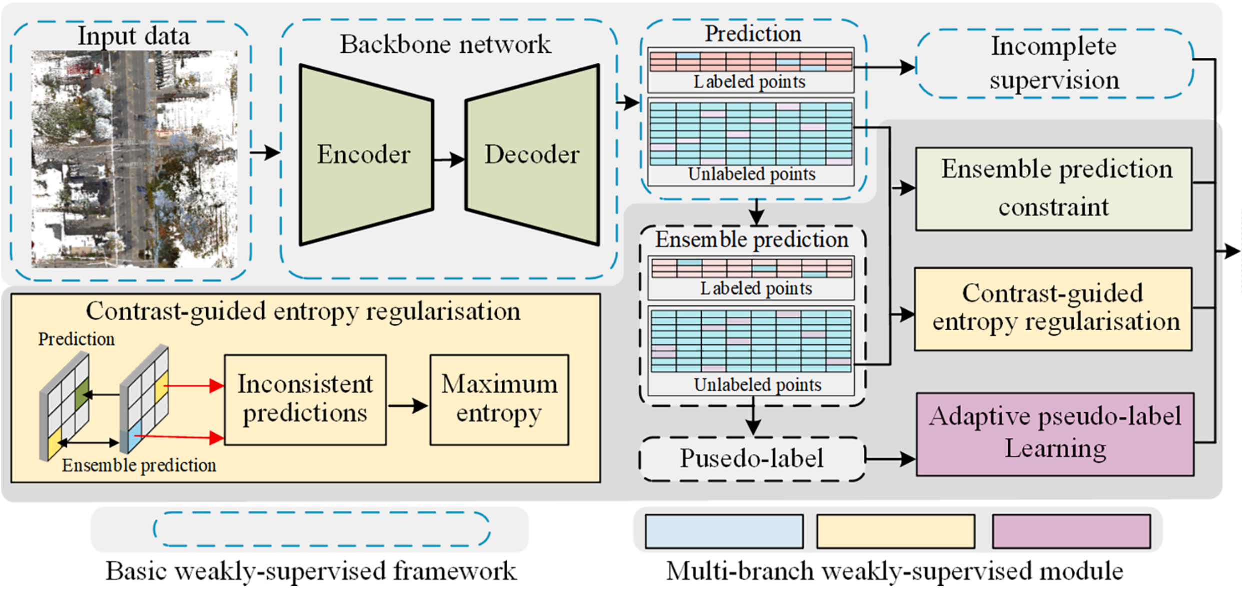


Fig. 1. The pipeline of the WSPointNet.

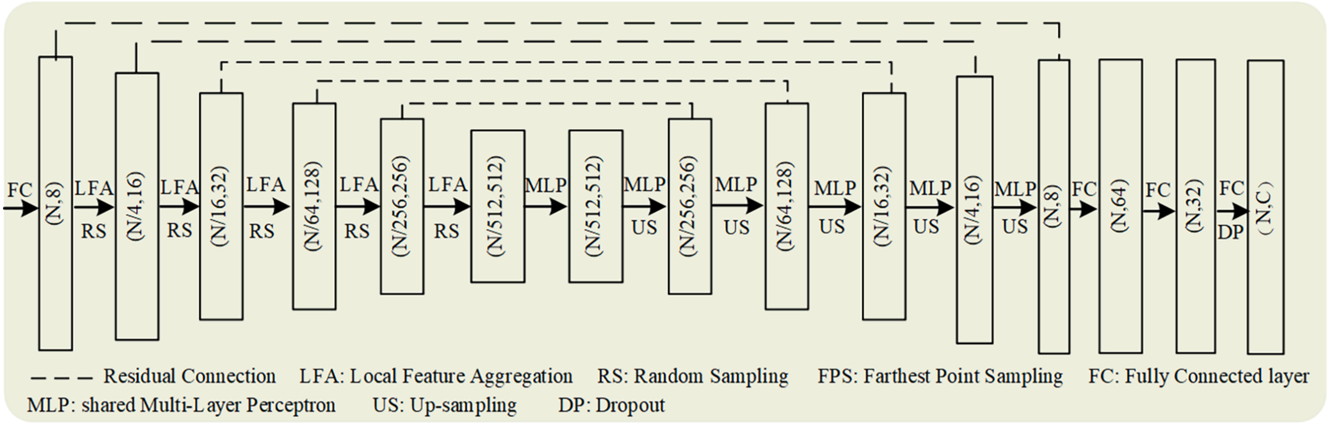


Fig. 2. Illustration of the RandLA-Net.

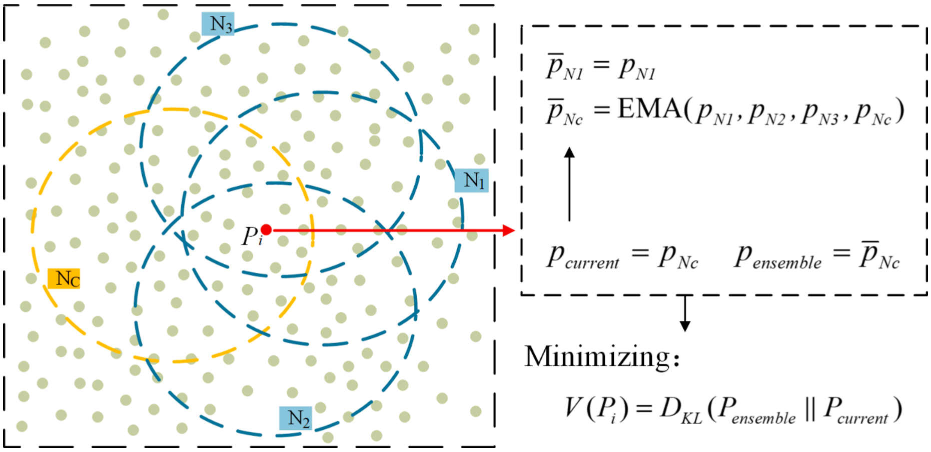

N3
N1
Nc
N2
Pi
$\overline{p}_{N1} = p_{N1}$
$\overline{p}_{Nc} = \mathrm{EMA}(p_{N1}, p_{N2}, p_{N3}, p_{Nc})$
$p_{current} = p_{Nc}$
$p_{ensemble} = \overline{p}_{Nc}$
Minimizing:
$V(P_i) = D_{KL}(P_{ensemble} \| P_{current})$

Fig. 3. Illustration of the EPC branch. ($p_{current}$ is the current prediction distributions. $p_{ensemble}$ is the ensemble prediction distributions. N_1, N_2, N_3, and N_c are the 1st, 2nd, 3rd, and current input data. P_i is an overlap point of the four input data.).

point P_i, respectively. $W_{sqrt,i}$ is the class weight of point P_i.

3.2. *Multi-branch weakly supervised module*

The proposed multi-branch weakly supervised module aims to fully utilize the informative features of unlabeled points while preventing the model from over-fitting. This module includes three branches, i.e., an ensemble prediction constraint branch, a contrast-guided entropy regularization branch, and an adaptive pseudo-label learning branch, each of which will be explained in detail.

3.2.1. *Ensemble prediction constraint branch*

To improve the prediction stability, the ensemble prediction constraint (EPC) branch is employed to obtain comparison predictions and minimize the variances between the current prediction distributions and the comparison prediction distributions. Since the inputs of the RandLA-Net are random and largely overlapped, an exponential moving average (EMA) method (Laine and Aila, 2017) is adopted to record each prediction of the overlapped points and obtain comparison samples, which is called ensemble prediction. As shown in Fig. 3, assuming that there are four input data, (N_1, N_2, $N_{3,}$ and N_c, where N_c represents the current input data). An overlap point P_i of the four inputs (rendered by the red color in Fig. 3) generates four prediction distributions with different global features. After assigning the first prediction distribution p_{N1} of point P_i as the ensemble prediction distribution $\overline{p}_{N1}$, an EMA method is used to record the prediction distributions of the subsequent inputs, i.e., p_{N2}, p_{N3}, and p_{Nc}, as $\overline{p}_N$. The ensemble prediction distribution $\overline{p}_N$ for the N-th update is calculated as follows:

$$\overline{p}_N = \begin{cases} p_N & ,N=1 \\ \alpha\overline{p}_{N-1}+p_N & ,N>1 \end{cases} \tag{3}$$

where α is the updating weight, which is set to 0.9 in this study with reference to Wang and Yao (2022). p_N is the current prediction distributions. $\overline{p}_{N-1}$ is the (N-1)-th updated ensemble prediction distributions.

Compared with the current prediction value, the ensemble prediction value of point P_i aggregates the predictions from the global features of several different input data, and thus it is more representative. In addition, the ensemble prediction value estimated by the EMA method is not involved in network training, and both the current and ensemble prediction values are obtained in one forward propagation. Therefore, the ensemble prediction provides an efficient and accurate comparison data for the proposed weakly supervised module.

As shown in Fig. 3, a Kullback-leibler (KL) divergence D_{KL} is used for describing the variance between the ensemble prediction distribution $\overline{p}_i$ and the current prediction distribution p_i. Then the EPC branch minimizes the variance to improve the prediction stability. The KL divergence of point P_i is referred to as the consistency cost, denoted as $V(P_i)$. The consistency cost $V(P_i)$ and the consistency loss L_{epc} are given by

$$V(P_i) = D_{KL}(\overline{p}_i \| p_i) = \sum_c^K \overline{p}_{ic}\log(\frac{\overline{p}_{ic}}{p_{ic}}) \tag{4}$$

$$L_{epc} = \frac{1}{P_u}\sum_i^{|P_u|} V(P_i) \tag{5}$$

where $\overline{p}_{ic}$ is the ensemble prediction posterior probability of point P_i being predicted as category c. p_{ic} is the current posterior probability of point P_i predicted as category c. Pu is the set of unlabeled points.

3.2.2. *Contrast-guided entropy regularization branch*

To overcome the issue of model over-fitting caused by the lack of sufficient labeling information, the contrast-guided entropy regularization (CG-ER) branch aims to guide the entropy regularization based on the comparison results between the current and ensemble prediction labels. Generally, the ensemble prediction result is considered to be a more accurate and robust value, and is used to estimate the current prediction results of unlabeled points. As shown in Fig. 1, for the points with inconsistent comparison results, their prediction results are considered to be inaccurate. Hence, the entropy maximization is used to prevent the model from over-fitting by encouraging high uncertainty in the predictions. We adopt a maximum entropy proxy approach, which encourages the prediction distributions to be closely uniform distributions (i.e., the maximum entropy state) (Larrazabal et al. 2021). The contrast-guided entropy regularization loss L_{er} is calculated as follows:

$$L_{er} = \frac{1}{|P_n|}\sum_i^{|P_n|}\sum_c^K p_{ic}\log(\frac{p_{ic}}{p_k}) \tag{6}$$

where p_{ic} is the current prediction posterior probability of point P_i being predicted as category c. K is the number of labeled categories. P_n is a point set with the inconsistent comparison results. p_k is the uniform distribution calculated from the reciprocal of the class number. $|\cdot|$ is the set cardinality.

3.2.3. *Adaptive pseudo-label learning branch*

To further exploit the informative features of unlabeled points, we propose the adaptive pseudo-label learning (A-PL) branch, which employs the ensemble prediction labels of unlabeled points as pseudo-labels and calculates pseudo-label weights by the consistency cost V (P_i). For the unlabeled points, the consistency cost describes the variance between the current and ensemble prediction distributions. It is commonly considered that the greater the variance, the higher the probability of the incorrect prediction value (Zheng and Yang, 2021). Therefore, we employ the weight calculation method (i.e., exp(-$V(P_i)$ as the pseudo-label weight of P_i), proposed by Kendall and Gal (2017), to calculate the pseudo-label weights of unlabeled points. This method allows the network to focus more on the pseudo-label points with low consistency costs and ignore the points with high consistency costs. The adaptive pseudo-label learning loss L_{ap} is defined as follows:

$$L_{ap} = -\frac{1}{|P_u|}\sum_i^{|P_u|}\exp(-V(P_i))\sum_c^K \overline{y}_{ic}\log p_{ic} \tag{7}$$

where P_u is the set of unlabeled points. $|\cdot|$ is the set cardinality. p_{ic} is the current prediction posterior probability of point P_i being predicted as category c. $\overline{y}_{ic}$ is the pseudo-label obtained by the argmax function on the ensemble prediction distribution.

3.3. *Training*

The network training included two stages. In the first training stage, the three proposed branches, i.e., the incomplete supervised branch, the EPC branch, and the CG-ER branch, are used for unlabeled points to obtain more accurate ensemble prediction values as the pseudo-labels. In the second training stage, all branches are used for model training to obtain the final model. The total loss L_{all} can be expressed by

$$L_{all} = L_{se} + \lambda_{er}L_{er} + \lambda_{epc}L_{epc} + \lambda_{ap}L_{ap} \tag{8}$$

where λ_{er}, λ_{epc}, and λ_{ap} are the weighting factors of L_{er}, L_{epc}, and L_{ap}, respectively. Specifically, $\lambda_{epc} = \lambda_{er} = 1$ are used during the whole training process. $\lambda_{ap} = 0$ is configured in the first stage. In the second stage, λ_{ap} is set to 1. Both training stages last for 50 epochs, respectively.

4. Experiments and analyses

4.1. *Datasets*

To examine the effectiveness of the WSPointNet, we evaluated it on two large-scale MLS datasets, i.e., the Toronto3D dataset (Tan et al., 2020) and the WHU-MLS dataset (Yang et al., 2021).

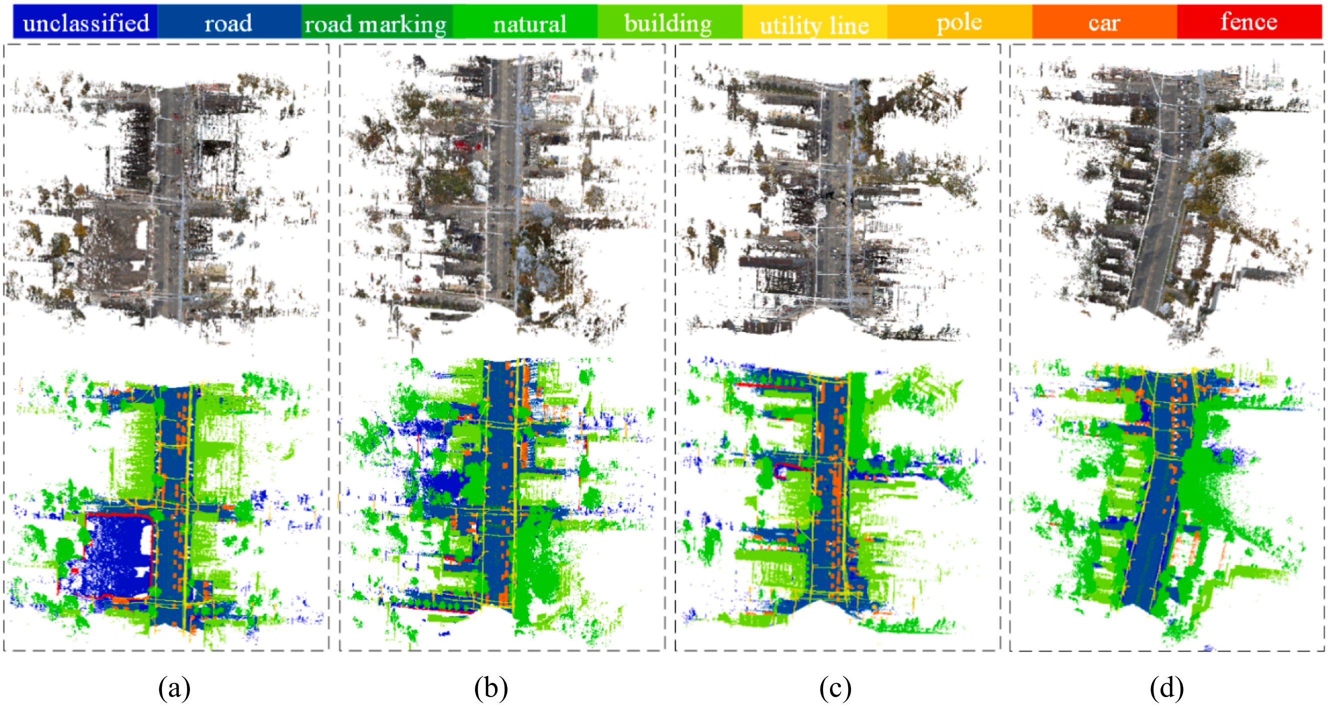


Fig. 4. Toronto3D dataset. (a) Section L001, (b) section L002, (c) section L003, (d) section L004 (from the top to the bottom, the point clouds rendered by RGB information and class labels, respectively).

Toronto3D dataset, a large-scale public MLS benchmark dataset, was collected in the urban areas of Toronto, Canada. The dataset contained approximately 78.3 million points and covered a 1 km^2 urban scene, which was further split into four sections (named as L001, L002, L003, and L004). As shown in Fig. 4, the Toronto3D dataset contains most of the MLS measurement ranges and rich attribute information (e.g., RGB information). The Toronto3D dataset was labeled as eight semantic classes, including road, natural, road marking (rd. m.), building (build.), utility line (util. l.), pole, car, and fence. Every point contained 3D coordinates (i.e., X, Y, and Z coordinates), RGB information, intensity, scan angle rank, and global navigation satellite system (GNSS) time. In this paper, only 3D coordinates and RGB information were used

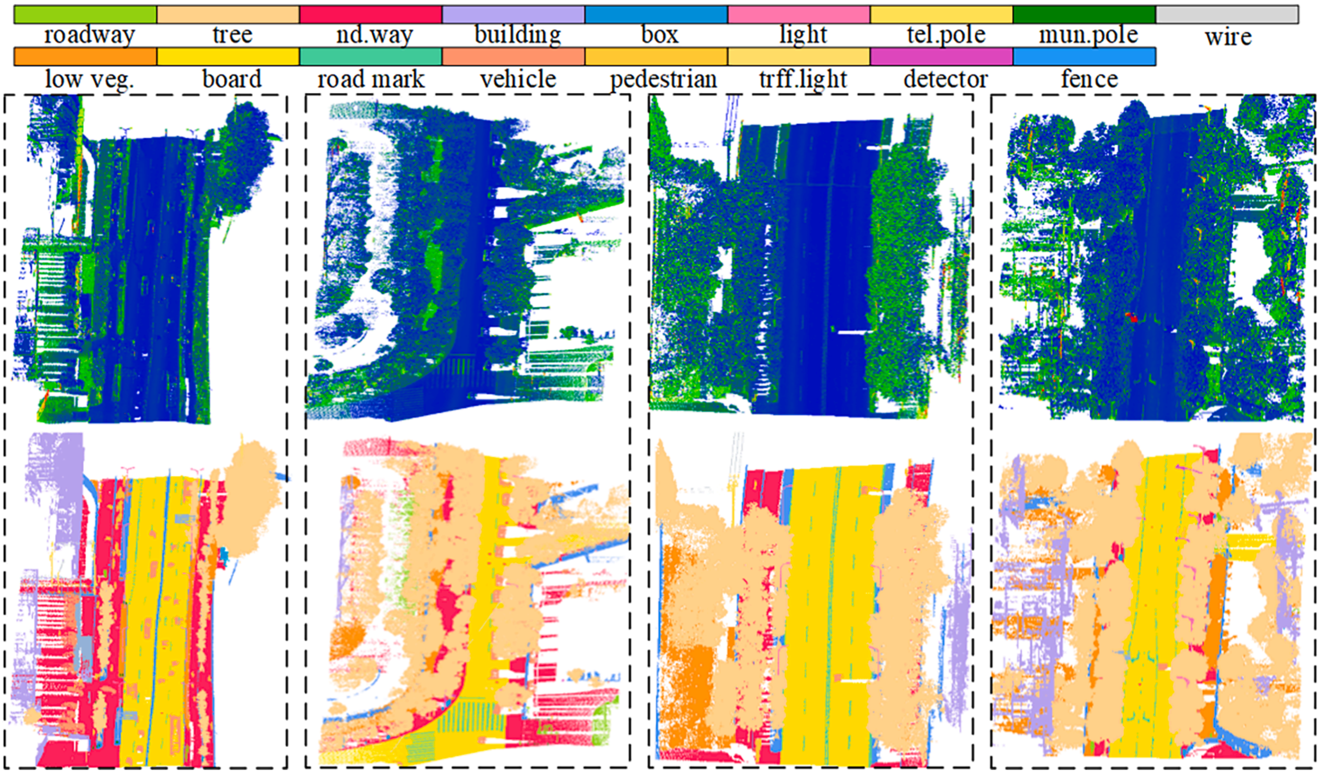


Fig. 5. Four scenes of the WHU-MLS dataset (from top to bottom: point clouds rendered by intensity information and class labels, respectively).

as the input data. Specifically, the L002 section was used for testing, and the others were used for training.

WHU-MLS dataset contains a total of 40 urban scenes with over 300 million points. Fig. 5 illustrates four scenes of the WHU-MLS dataset. As shown in Fig. 5, the WHU-MLS dataset covers typical and complex urban scenes with multiple-type objects. Specifically, the dataset was labeled as 17 classes, including roadway, non-drive way (nd. way), road mark (rd. mark.), building, fence, tree, low vegetation (low veg.), pedestrian, vehicle, light, telegraph pole (tel. pole), traffic light (trff. light), municipal pole (mun. pole), detector, box, board, and wire. Each point contained 3D coordinates, intensity, and the number of returns. In this paper, the 3D coordinates and intensity information were used as the input data. Among the 40 scenes, 10 scenes were used for testing and the others were used for training.

4.2. Implementation details

The Adam optimizer was used with an initial learning rate of 0.01 and a momentum of 0.95 to train 100 epochs. The input number of points was set to 65,536 and the batch size was 4 according to the memory limit of the graphics card. At each epoch, Toronto3D dataset and WHU-MLS dataset were trained for 500 steps and 800 steps, respectively, depending on the number of point clouds. As indicated in Table 1, according to the point cloud density of the dataset, we set the subsampling grid size to 0.04 and 0.08 m for the Toronto3D dataset and WUH-MLS dataset, respectively. For labeled point selection strategy, we applied the sample selection method of SQN: 0.1 % labeled points were randomly selected after sub-sampling pre-processing, and unannotated points were assigned to unclassified class. The detailed values are shown in Table 1. The annotation strategy was more suitable than the other weakly supervised annotation methods (Wang and Yao, 2022; Zhang et al., 2021b) for practical application scenarios. Compared to fully annotated datasets, randomly annotating 0.1 % points greatly reduced annotation cost and effort (Hu et al., 2021).

All experiments were carried out on a Personal Computer with an Intel CoreTM i9-9820X processor, an NVIDIA RTX 2080Ti GPU with 11-GB of memory, and a 64G RAM. We evaluated performance using three metrics: overall accuracy (OA), intersection over union (IoU), and mean intersection over union (mIoU).

4.3. Overall performance

4.3.1. Toronto3D dataset

Fig. 6 shows the results obtained by the WSPointNet on the Toronto3D dataset. As shown in Fig. 6, compared with the true labels, visual inspection indicated that our WSPointNet was capable of correctly classifying the majority of points despite using only 0.1 % labeled points. To further demonstrate the semantic segmentation results, the misclassified results were rendered by the red color, as shown in Fig. 6(c). Most red points were presented in the areas of trees and utility lines, as well as in areas far away from roads. Fig. 7 shows the three close-view examples of the Toronto3D dataset. The first example in the top row mainly illustrated several utility poles and a tree with a large and leafy canopy. The second example in the second row was a typical street scene, containing a building, poles (including an advertising board and utility poles), and utility lines. The third example in the bottom row mainly illustrated buildings, cars, the ground, and trees. First, the WSPointNet misclassified some utility line points and partial points of the utility pole close to the trees (see the red box on the top row). Second, the WSPointNet misclassified some pole line points, particularly the points of the billboard close to the building facades (see the red box in the bottom row). Third, the WSPointNet misclassified some road points close to cars, and the building points close to trees. The reasons for the above phenomena might be caused by: (1) their geometric and spectral similarities, (2) our method relying on sparsely labeled points that incompletely describe the geometrical shapes of the objects, thereby leading to the presentence of the semantic segmentation errors, (3) some misclassified objects were far away from the roads and received object occlusion. Such points were too sparse to describe the geometrical shapes of the objects, which also led to the incorrect semantic segmentation results.

Table 1
Pre-processing results of two datasets.

Dataset	Grid size	Raw pts	Grid sampling pts	Training pts	Anno. pts (0.1 %)
Toronto3D	0.04 m	78.3 M	24.3 M	18.4 M	18,387
WHU-MLS	0.08 m	324.9 M	70.8 M	56.4 M	56,409

Several recently published semantic segmentation methods were selected for performance comparison. Our comparative experiments were divided into two groups: supervised and weakly supervised, as shown in Tables 2 and 3, respectively.

In the first group of the comparative experiments, the input data only contained three coordinates of the MLS point clouds (w/o RGB). The eight supervised methods, i.e., PointNet++ (Qi et al., 2017b), PointNet++ (MSG) (Qi et al., 2017b), DGCNN (Wang et al., 2019b), MS-PCNN (Ma et al., 2019), RandLA-Net (Hu et al., 2020), MS-TGNet (Tan et al., 2020), TGNet (Li et al., 2019), and KPFCNN (Thomas et al., 2019), were selected for the comparison. Among these supervised methods, the KPFCNN, RandLA-Net, and MS-TGNet achieved over 95 % overall accuracy (OA). In this study, our WSPointNet achieved an OA of 95.26 % and a mIoU of 70.42 %, respectively, when only 0.1 % labeled points were used for training. Despite being a weakly-supervised method, our results were comparable to those of the top three supervised methods, i.e., KPFCNN, RandLA-Net, and MS-TGNet. Due to the large errors in the semantic segmentation of road markings, the WSPointNet obtained a relatively lower mIoU value than that of the MS-TGNet and the RandLA-Net. This might be the fully supervised methods can learn the geometrical relations of the dense road marking points in a certain neighborhood, thus obtaining their discriminative features. In contrast, the road marking points after 0.1 % down-sampling are too sparse to describe the geometrical completeness of the road markings, leading to the WSPointNet failed to provide their salient feature representation. To test the impact of the RGB information on the point cloud semantic segmentation, we combined the coordinates and the RGB information as the input data for model training and testing (w/RGB). As shown in Table 2, compared with the RandLA-Net using the same input data, the WSPointNet obtained similar accuracy and the best IoU scores in the road and road marking categories. Compared with the method using coordinates, the WSPointNet with the RGB information achieved the best OA and mIoU. The OA and mIoU values of the WSPointNet surpassed those of the best supervised method, RandLA-Net, by 1.13 % and 1.24 %, respectively. From the comparative results, we concluded that our WSPointNet with only sparsely labeled points outperformed most supervised semantic segmentation methods when dealing with large-scale MLS point clouds. Moreover, the performance of our WSPointNet can be further improved by using both 3D coordinates and RGB information.

Given the fact that fewer weakly supervised methods were available for a fair comparison, we compared the WSPointNet with the SQN (Hu et al., 2021) and the baseline method on the Toronto3D dataset. Note that the baseline was the RandLA-Net using only 0.1 % labeled points for model training. The SQN method obtained the multi-level encoded features by fully leveraging the features of the labeled points via a query network.

As shown in Table 3, the WSPointNet obtained the best OA and mIoU scores. Specifically, as only the coordinates were used as input, our method improved OA and mIoU by 2.42 % and 1.07 %, respectively, when compared to the SQN. Also, the WSPointNet outperformed the baseline by an OA improvement of 8.74 % and a mIoU increase of 4.99

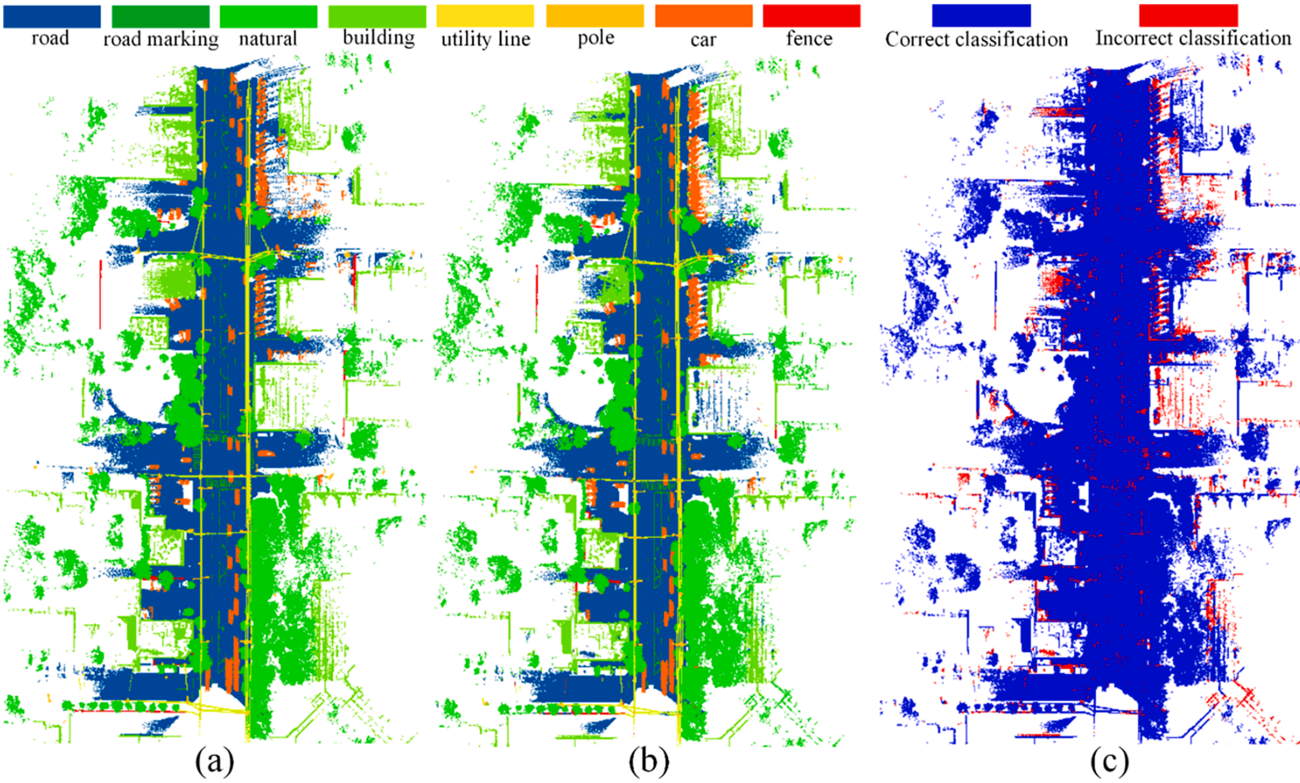


Fig. 6. Qualitative results obtained by the proposed WSPointNet on the Toronto3D dataset: (a) true labels, (b) semantic segmentation results, (c) misclassified results.

%, respectively. For specific classes, the WSPointNet gained at least 2.5 % IoU improvement of the poles and fences, while poorly performed on the road markings due to a lack of effective feature representation. We considered that the WSPointNet might be sensitive to spatial features. For the poles and fences that are distinct, structured geometries in 3D point clouds, our ensemble prediction module learns more spatial features of the unlabeled points at different stages, which helps describe the poles and fences in a comprehensive way. In addition, the proposed pseudo-labelling strategy provides a similar fully-supervised means to collect the salient contextual features of the poles and fences. Thus, with the 3D structural distinctiveness of the objects (e.g., poles and fences) and the use of the supervised learning, the WSPointNet obtained a better identification of these 3D objects.

In contrast, road markings are normally attached to roads, and can be considered as a part of the roads. It is hard to correctly extract or classify the road markings from only LiDAR data. For the Toronto3D dataset without spectral information, the labelled road marking training samples provide extremely limited geometrical and contextual features because the geometrical features of the road markings are similar to those of the roads in most cases. Thus, without the distinct characteristics of the road markings in 3D point clouds, our ensemble prediction is incapable of mining the useful road marking features from the unlabeled points. Particularly, the subsequent GC-ER and A-PL strategies based on the ensemble prediction might further worsen the geometrical representation of the road markings. However, we noted that the road markings were painted on the roads with high reflective materials, the RGB information was also investigated in the three models. As seen in Table 3 and Fig. 7, with the RGB information, all models outperformed their counterparts, improving the semantic segmentation accuracies of all classes, particularly the road marking. The IoU value of the road markings dramatically increased by up to 66.99 %.

4.3.2. WHU-MLS dataset

To further evaluate the robustness of our WSPointNet, we tested it on the WHU-MLS dataset. As we mentioned above, the WSPointNet achieved better semantic segmentation performance when using 0.1 % labeled points for training. Thereby, we used 0.1 % labeled points for training on the WHU-MLS dataset. Fig. 8 shows the semantic segmentation results of various objects (17 classes) in three urban areas with complex environments. As seen in Fig. 8(c), visual inspection indicated that our WSPointNet correctly classified most of the categories, but it also misclassified the points with similar characteristics, such as non-driveways and roadways, road marking edges and roadways, and low vegetation and trees.

Table 4 shows the quantitative results obtained by the WSPointNet and the other comparative methods. We used the experimental results presented in Han et al. (2021) as the benchmark for the WSPointNet. Note that, for this group of benchmarks on the WHU-MLS dataset, the overall accuracy was not presented in Table 3 because this metric was not used for performance assessment in Han et al. (2021). As shown in Table 4, we conclude that the quantitative results were consistent with the visual performance, and our WSPointNet achieved competitive results in comparison with the other methods. The WSPointNet obtained an mIoU of 56.48 % using only 0.1 % labeled points, which remarkably outperformed the three supervised methods, i.e., PointNet++, PointConv (Wu et al., 2019), and Han's method (Hu et al., 2021). The reasons behind this phenomenon include: 1) a block preprocessing method used in the three supervised methods might damage the geometric completeness of the objects; 2) the used baseline itself was able to provide a relatively superior feature representation using only 0.1 % labeled points; 3) our weakly supervised strategy contributed to the augmentation of useful features explored from the unlabeled points. Moreover, we found that the baseline using only 0.1 % labeled points obtained a mIoU of only 50.00 %, almost a degradation of 10 % when comparing with the baseline. However, the WSPointNet still obtained the competitive performance with the baseline when using only 0.1 % samples. Specifically, the WSPointNet achieved the best IoU scores of the buildings, low vegetation, roads, and road markings among all the

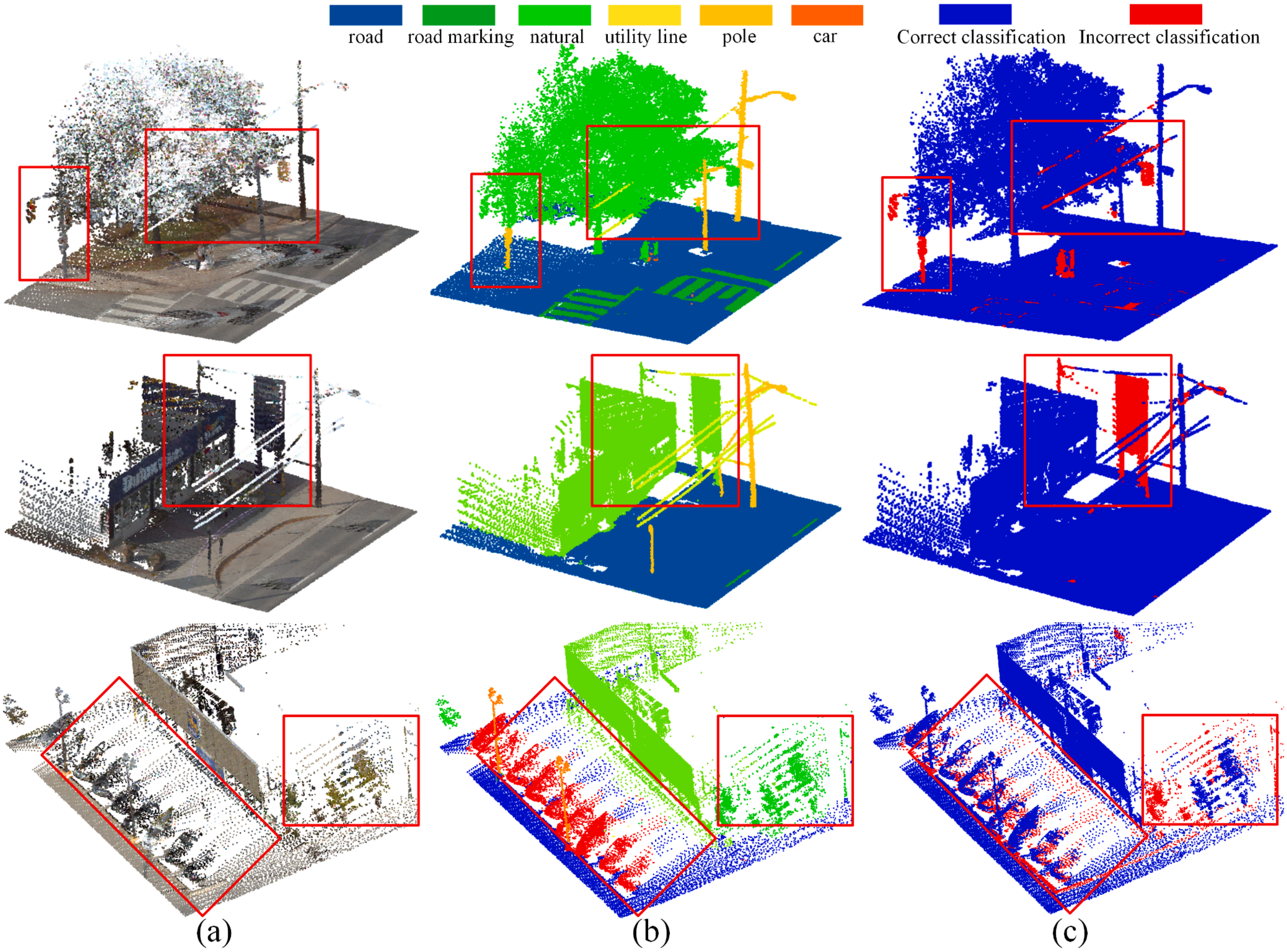


Fig. 7. Close-views of the three examples of the Toronto3D dataset: (a) raw point cloud, (b) semantic segmentation results obtained by our method, (c) misclassified results.

Table 2
Quantitative results obtained by comparing our WSPointNet with the eight fully supervised benchmark methods on the Toronto3D dataset. The scores of the eight comparison methods were obtained from Hu et al. (2021). The underlined represents the best scores in the methods without RGB information, and the bold represents the best scores in all the methods.

Method	OA(%)	mIoU(%)	IoU(%)							
			road	rd m.	natural	build.	util.l.	pole	car	fence
PointNet++	84.88	41.81	89.27	0.00	69.06	54.16	43.78	23.30	52.00	2.95
PointNet++(MSG)	92.56	59.47	92.90	0.00	86.13	82.15	60.96	62.81	76.41	14.43
DGCNN	94.24	61.79	93.88	0.00	91.25	80.39	62.40	62.32	88.26	15.81
KPFCNN	95.39	69.11	94.62	0.06	96.07	91.51	87.68	**81.56**	85.66	15.72
MS-PCNN	90.03	65.89	93.84	3.83	93.46	82.59	67.80	71.95	91.12	22.50
TGNet	94.08	61.34	93.54	0.00	90.83	81.57	65.26	62.98	88.73	7.85
MS-TGNet	95.71	70.50	94.41	17.19	95.72	88.83	76.01	73.97	**94.24**	23.64
RandLA-Net(w/o RGB)	95.63	77.72	94.53	42.44	96.62	93.1	86.56	76.83	92.55	39.14
RandLA-Net (w/ RGB)	**97.15**	**81.88**	96.69	64.10	**96.85**	**94.14**	**88.03**	77.48	93.21	**44.53**
WSPointNet (w/o RGB, 0.1 %)	95.26	70.42	94.48	0.00	95.14	92.84	82.39	70.78	89.1	38.62
WSPointNet (w/ RGB, 0.1 %)	96.76	78.96	**96.70**	**66.99**	94.89	90.79	83.68	75.71	88.37	34.54

comparative methods. The reasons for these behaviors were that the information of unlabeled points was fully exploited by enforcing the multi-branch weakly supervised module, thereby improving the capabilities of the model for classifying objects with spectral and spatial similarities.

4.4. Ablation study

A series of ablation studies were conducted to assess the performance of the proposed methods, including the square-root weighted (SRW) loss function, the EPC branch, the CG-ER branch, and the A-PL branch. These modified models were tested on the WHU-MLS dataset and the Toronto3D dataset with only coordinates. 0.1 % labeled points were used for training the networks. The specific experimental results were shown in Table 5.

Effect of square-root weighted loss function. We adopted the square-root weighted loss function to replace the loss function of the baseline method (i.e., Baseline), and the resultant network was named as

Table 3
Quantitative results obtained by comparing our WSPointNet with Baseline and SQN methods on the Toronto3D dataset (0.1% labeled points). The scores of the SQN methods were obtained from Hu et al. (2021). The underlined represents the best scores in the methods without RGB information, and the bold represents the best scores in all the methods. Baseline scores were obtained for the RandLA-Net using only sparse labeling.

Method	OA (%)	mIoU(%)	IoU (%)							
			road	rd m.	natural	build.	util.l.	pole	car	fence
Baseline(w/o RGB)	86.52	65.43	81.69	9.70	94.02	91.39	79.32	61.9	80.88	24.55
SQN(w/o RGB)	92.84	69.35	93.74	16.83	92.55	89.04	82.50	63.98	88.17	28.01
WSPointNet (w/o RGB)	95.26	70.42	94.48	0.00	**95.14**	**92.84**	82.39	70.78	**89.1**	**38.62**
Baseline(w/ RGB)	95.47	74.46	95.33	57.43	93.57	87.29	77.09	72.97	86.65	25.32
SQN(w/ RGB)	96.67	77.75	96.69	65.67	94.58	91.34	83.36	70.59	88.87	30.91
WSPointNet (w/ RGB)	**96.76**	**78.96**	**96.70**	**66.99**	94.89	90.79	**83.68**	**75.71**	88.37	34.54

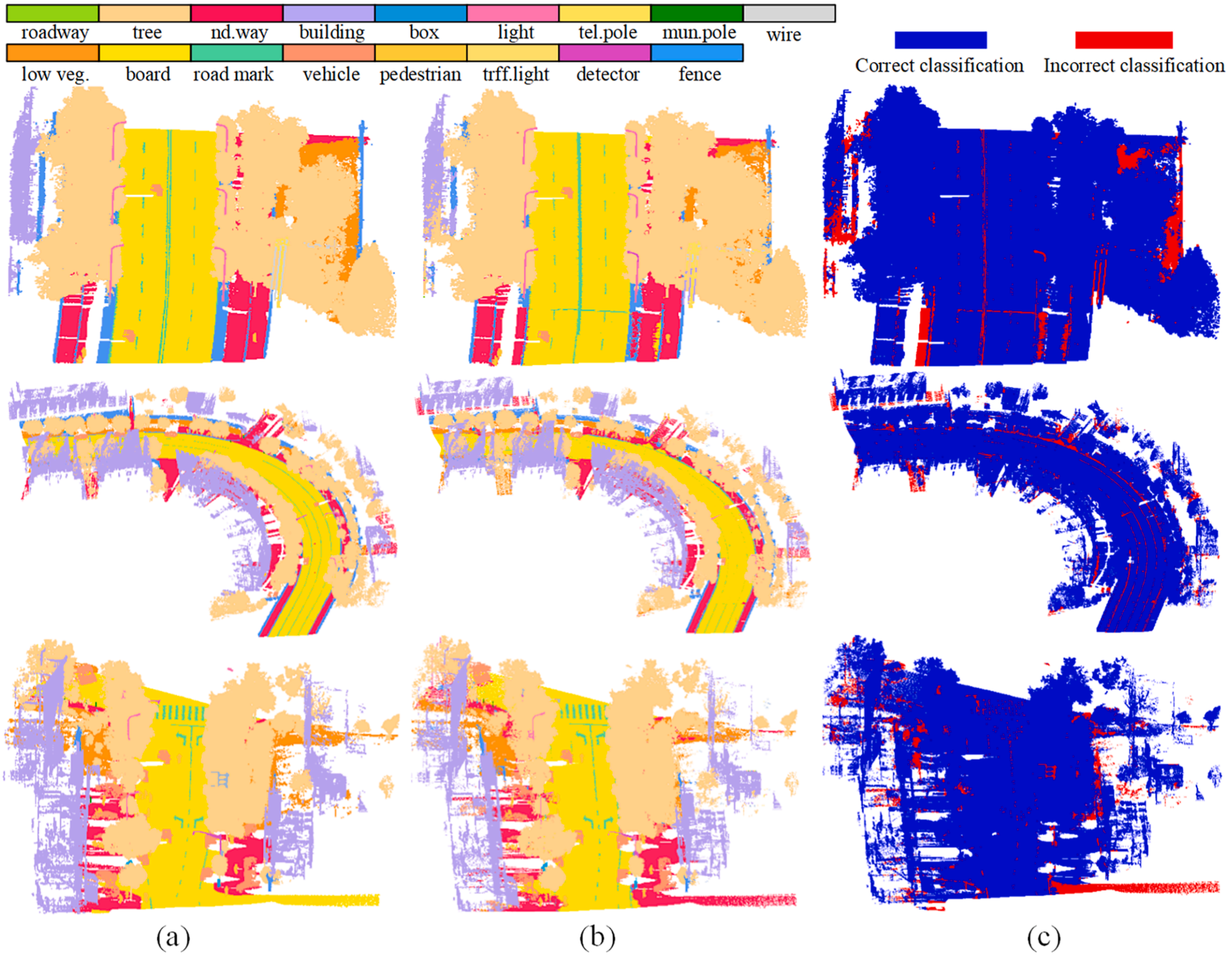


Fig. 8. Qualitative results of three scenes in the WHU-MLS dataset: (a) ground-truth labels, (b) semantic segmentation results, (c) misclassified results.

Model A. As shown in Table 5, for the Toronto3D dataset, model A achieved an improvement of 8.26 % and 2.78 % for OA and mIoU, compared with the Baseline. For the WHU-MLS dataset, model A achieved an increase of 0.41 % and 2.14 % for OA and mIoU. The experimental results showed that the SRW loss function enabled model to focus more effectively on small sample classes compared with baseline method, and improved the accuracy of semantic segmentation for MLS point clouds.

Effect of ensemble prediction constraint. We embedded the EPC branch into the basic framework (i.e., Model A), and the resultant network was named as Model B. As shown in Table 5, model B improved the mIoU by 1.67 % and 3.45 % on the Toronto3D and WHU-MLS datasets, compared with Model A. The experimental results showed that the EPC branch enhanced the consistency of network predictions, and improved the accuracy of semantic segmentation for MLS point clouds.

Effect of contrast-guided entropy regularization. In this experiment, the CG-ER branch was added to Model B, and the resultant network was named as Model C. The experimental results in Table 5 showed that, compared with Model B, Model C achieved a OA improvement of about 0.69 % and 0.19 % on the Toronto3D and WHU-MLS datasets, respectively. We argued that the above results can mainly be attributed to the CG-ER branch, which prevented network from overfitting by guiding unlabeled non-confidence prediction points for

Table 4
Quantitative results obtained by comparing different methods on the WHU-MLS dataset. The scores of the fully supervised comparison methods were obtained from Han et al. (2021). The bold represents the best scores in all methods. Baseline scores were obtained for the RandLA-Net using the training parameters of this paper.

Methods	mIoU (%)	IoU (%)								
		tree	nd.way	building	box	light	tel.pole	mun.pole	low veg.	board
		roadway	rd.mark.	vehicle	pedestrian	trff.light	detector	fence	wire	
PointNet++	41.10	83.30	42.00	72.70	6.60	59.10	30.80	7.80	33.10	13.90
		80.00	29.50	76.70	38.90	25.00	11.00	56.30	32.70	
PointConv	46.40	85.60	48.90	73.50	28.20	59.70	35.70	20.00	32.40	16.00
		82.00	30.60	76.20	53.80	28.70	27.60	52.60	36.50	
Han’s method	52.80	84.50	**58.40**	77.10	45.40	**71.80**	**49.90**	**26.50**	34.10	20.20
		83.60	38.10	79.10	60.80	31.00	31.30	57.90	47.20	
Baseline	**60.54**	**90.45**	51.33	87.29	**55.8**	68.42	46.38	16.4	33.89	**33.6**
		91.26	55.25	**95.21**	**83.99**	**50.25**	**41.4**	**68.77**	**59.46**	
Baseline (0.1 %)	50.00	89.10	52.51	82.62	31.35	49.95	20.58	6.82	33.73	22.76
		91.41	53.83	93.53	72.92	38.78	9.49	64.71	36.01	
WSPointNet (0.1 %)	56.48	90.23	56.43	**88.39**	41.14	56.28	28.65	24.88	**36.32**	26.94
		92.52	**58.24**	94.65	69.97	44.68	29.93	65.7	55.28	

Table 5
A comparison of the point cloud semantic segmentation results with different models. The bold represents the best scores in all models.

Model	SRW	EPC	CG-ER	A-PL	Toronto3D		WHU-MLS	
					OA(%)	mIoU (%)	OA(%)	mIoU (%)
Baseline					86.52	65.43	89.50	50.00
A	√				94.78	68.21	89.91	52.14
B	√	√			94.55	69.88	90.26	55.59
C	√	√	√		95.24	69.93	90.45	55.66
WSPointNet	√	√	√	√	**95.26**	**70.42**	**90.83**	**56.48**

entropy maximization.

Effect of adaptive pseudo-label learning. As shown in Table 5, compared with Model C without the A-PL branch, our WSPointNet obtained a better performance, with a mIoU increase of 0.49 % and 0.82 % on the Toronto3D and WHU-MLS datasets. The reason was that the A-PL branch adaptively adjusts the pseudo-label weights according to the consistency cost, providing effective supervisory signals to improve the semantic segmentation accuracy.

4.5. Analyses on various weakly-supervised strategies

To further validate the effectiveness of the proposed weakly supervised strategies, we also compared it with some recently-published weakly supervised strategies on our underlying framework, and the comparison results were shown in Table 6.

Entropy minimization. In most weakly supervised strategies, the entropy minimization was commonly used to reduce prediction class overlap and obtain more discriminative features. We replaced the proposed GC-ER branch with the entropy minimization in this study. As shown in Table 6, the WSPointNet improved the mIoU value by 1.25 % and 1.05 % on the Toronto3D and WHU-MLS datasets, respectively, compared with the entropy minimization method. This is because the entropy minimization is prone to over-fitting the model, while the GC-ER method can alleviate this over-fitting problem and finally improve the model performance.

Online soft pseudo-labeling. The online soft pseudo-labeling strategy (OS-PL) used entropy values to calculate pseudo-label weights. We replaced the OS-PL method with our A-PL method. The comparative results were shown in Table 6. Compared with the OS-PL method, the A-PL method improved the mIoU value by 1.09 % and 2.08 % on the Toronto3D and WHU-MLS datasets, respectively. Above quantitative results showed that our A-PL method constructed more effective adaptive pseudo-label weights by using the variance between ensemble predictions and current predictions.

The weakly supervised strategies of the Wang’s (Wang and Yao, 2022**).** Wang and Yao (2022) used the mean-square error (MSE) loss, the entropy minimization, and the OS-PL method to calculate the consistency constraint, entropy regularization, and pseudo-labeling losses, respectively. As seen from Table 6, the WSPointNet outperformed the mIoU of Wang’s (2022) method by 2.95 % and 2.32 % on the Toronto3D and WHU-MLS datasets, respectively, which demonstrated that our proposed weakly supervised strategy was considered to obtain a more effective supervised signal than the Wang’s, and finally improved the model performance.

The weakly supervised strategies of the PSD method. The PSD approach generated perturbation samples through a new framework, and thus enlarged supervised signals through consistency constraints as well as contextual modules. Because the PSD method requires a large amount of GPU memory for generating perturbation samples, we hardly used the default parameters for our PSD experiments. Therefore, according to the limits of the graphics card, we performed this group of experiments by adjusting the parameter settings (i.e., the number of

Table 6
A Comparison result of weakly methods with different weakly supervised strategies. “*” represents that our method used the same parameters of the PSD method for the comparative experiments. The bold represents the best scores in all methods.

Datasets	Metrics	WSPointNet	Entropy Minimization	Online soft pseudo-labeling	Wang and Yao (2022)	PSD	WSPointNet*
Toronto3D	OA(%)	**95.26**	95.10	95.12	94.83	90.20	94.09
	mIoU(%)	**70.42**	69.17	69.33	67.47	55.74	61.72
WHU-MLS	OA(%)	**90.83**	90.41	90.42	90.43	89.06	89.24
	mIoU(%)	**56.48**	55.43	54.40	54.16	43.42	44.63

input points was 53,248 and the batch size was 2). As shown in Table 6, the WSPointNet improved the OA and mIoU by 3.89 % and 5.98 % on the Toronto3D dataset, respectively, compared with the PSD method with the same parameters. These results showed that our weakly supervised strategy can reduce the memory consumption while effectively using the unlabeled point features to increase the classification accuracy of the model.

5. Conclusion

This paper presented a multi-branch weakly supervised learning network (i.e., WSPointNet) for semantic segmentation of large-scale MLS point clouds. The network first employed the RandLA-Net to extract informative features of point clouds, and used incomplete supervision with the randomly selected sparsely labeled points to provide underlying supervised signals for model training. Then, a multi-branch weakly supervised module, including the ensemble prediction constraint, contrast-guided entropy regularization, and adaptive pseudo-label learning branches, was performed by employing ensemble prediction to fully exploit the informative features of unlabeled points. Comparison results with the fully supervised methods showed that the WSPointNet using sparse labels achieved competitive semantic segmentation accuracies. On the Toronto3D using 0.1 % labeled data, the WSPointNet achieved an OA of 96.76 % and a mIoU of 78.96 %, surpassing the baseline method by 1.29 % and 4.50 %, respectively. Thus, the WSPointNet can effectively reduce the workload of data annotation, contributing to the applications of MLS point clouds in urban scenarios. In the future, more advanced techniques will be explored to obtain more effectively supervised sources and more contextual information.

Funding Sources

This research was partially funded by the National Natural Science Foundation of China under Grants No. 41871380, 41971414, 62076107, and 42101451, and also partially supported by the Emerging Interdisciplinary Project of Central University of Finance and Economics in China.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

Boulch, A., Guerry, J., Le Saux, B., Audebert, N., 2018. SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks. Comput. Graphics. 71, 189–198.

Feng, Y., Zhang, Z., Zhao, X., Ji, R., Gao, Y., 2018. GVCNN: group-view convolutional neural networks for 3D shape recognition. In: Proc. CVPR, pp. 264-272. https://doi.org/10.1109/CVPR.2018.00035.

Du, J., Cai, G.R., Wang, Z.Y., Huang, S.F., Su, J.H., Junior, J.M., Smit, J., Li, J., 2021. ResDLPS-Net: Joint residual-dense optimization for large-scale point cloud semantic segmentation. ISPRS J. Photogramm. Remote Sens. 182, 37–51.

Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L.i., Bennamoun, M., 2021. Deep learning for 3D point clouds: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 43 (12), 4338–4364.

Hou, J., Graham, B., Nießner, M., Xie, S., 2021. Exploring data-efficient 3D scene understanding with contrastive scene contexts. In: Proc. CVPR, pp. 15582-15592. https://doi.org/10.1109/CVPR46437.2021.01533.

Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N., Markham, A., 2020. RandLA-Net: Efficient semantic segmentation of large-scale point clouds. In: Proc. CVPR, pp. 11105-11114. https://doi.org/10.1109/CVPR42600.2020.01112.

Hu, Q., Yang, B., Fang, G., Guo, Y., Leonardis, A., Trigoni, N., Markham, A., 2021. SQN: Weakly-supervised semantic segmentation of large-scale 3D point clouds with 1000x fewer labels. In arXiv preprint arXiv: 2104.04891. https://arxiv.org/abs/2104.04891.

Jiang, L., Shi, S., Tian, Z., Lai, X., Liu, S., Fu, C., Jia J., 2021. Guided point contrastive learning for semi-supervised point cloud semantic segmentation. In: Proc. ICCV, pp. 6403-6412. https://doi.org/10.1109/ICCV48922.2021.00636.

Kendall, A., Gal, Y., 2017. What uncertainties do we need in bayesian deep learning for computer vision? In: Proc. NeurIPS, pp. 5580–5590.

Komarichev, A., Zhong, Z., Hua, J., 2019. A-CNN: Annularly convolutional neural networks on point clouds. In: Proc. CVPR, pp. 7413-7422. https://doi.org/10.1109/CVPR.2019.00760.

Laine, S., Aila, T., 2017. Temporal ensembling for semi-supervised learning. In: 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24–26, 2017, Conference Track Proceedings, OpenReview.net.

Larrazabal, A., Martinez1, C., Dolz, J., Ferrante, N., 2021. Maximum entropy on erroneous predictions (MEEP): Improving model calibration for medical image segmentation. In arXiv preprint arXiv: 2112.12218. https://arxiv.org/abs/2112.12218.

Li, J., Chen, B. M., Lee, H. G., 2018. SO-Net: Self-organizing network for point cloud analysis. In: Proc. CVPR, pp. 9397-9406. https://doi.org/10.1109/CVPR.2018.00979.

Han, X., Dong, Z., Yang, B., 2021. A point-based deep learning network for semantic segmentation of MLS point clouds. ISPRS J. Photogramm. Remote Sens. 175, 199–214.

Li, Y., Ma, L., Zhong, Z., Cao, D., Li, J., 2019. TGNet: Geometric graph cnn on 3D point cloud segmentation. IEEE Trans. Geosci. Remote Sens. 58 (5), 3588–3600. https://doi.org/10.1109/TGRS.2019.2958517.

Liang, X., Fu, Z., Sun, C., Hu, Y., 2021. MHIBS-Net: Multiscale hierarchical network for indoor building structure point clouds semantic segmentation. Int. J. Appl. Earth Obs. Geoinf. 102, 102449.

Luo, H., Wang, C., Wen, C., Chen, Z., Zai, D., Yu, Y., Li, J., 2018. Semantic Labeling of Mobile LiDAR Point Clouds via Active Learning and Higher Order MRF. IEEE Geosci. Remote Sens. Lett. 56 (7), 3631–3644.

Luo, H., Khoshelham, K., Fang, L., Murray, R.M., 2020. Unsupervised scene adaptation for semantic segmentation of urban mobile laser scanning point clouds. ISPRS J. Photogramm. Remote Sens. 169, 253–267.

Luo, H., Li, L., Fang, L., Wang, H., Wang, C., Guo, W., Li, J., 2022. Domain Adaptation for Object Classification in Point Clouds via Asymmetrical Siamese and Conditional Adversarial Network. IEEE Geosci. Remote Sens. Lett. 19, 1–5.

Ma, L., Li, Y., Li, J., Tan, W., Yu, Y., Chapman, M., 2019. Multi-scale point-wise convolutional neural networks for 3D object segmentation from LiDAR point clouds in large-scale environments. IEEE Trans. Intell. Transp. Syst. 22 (2), 821–836. https://doi.org/10.1109/TITS.2019.2961060.

Meng, H., Gao, L., Lai, Y., Manocha, D., 2019. VV-Net: Voxel VAE net with group convolutions for point cloud segmentation. In: Proc. ICCV, pp. 8499-8507. https://doi.org/10.1109/ICCV.2019.00859.

Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017a. PointNet: deep learning on point sets for 3D classification and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.Vol. 2017-July, pp. 77-85. https://doi.org/10.1109/CVPR.2017.16.

Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017b. PointnNet++: deep hierarchical feature learning on point sets in a metric space. In: Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2017-December, 5099-5108. http://arxiv.org/abs/1706.02413.

Tan, W., Qin, N., Ma, L., Li, Y., Du, J., Cai, G., Yang, K., Li, J., 2020. Toronto-3D: A large-scale mobile lidar dataset for semantic segmentation of urban roadways, in. In: Proc. CVPR Workshops, pp. 797-806. https://doi.org/10.1109/CVPRW50498.2020.00109.

Tao, A., Duan, Y. Q., Wei, Y., Lu, J., Zhou, J., 2020. Seg-group: Seg-level supervision for 3D instance and semantic segmentation. In arXiv preprint arXiv: 2012.10217. https://arxiv.org/abs/2012.10217.

Thomas, H., Qi, C.R., Deschaud, J.E., Marcotegui, B., et al., 2019. KPConv: Flexible and deformable convolution for point clouds. In: Proce. ICCV, pp. 6411–6420. https://doi.org/10.1109/ICCV.2019.00651.

Wang, L., Huang, Y., Hou, Y., Zhang, S., Shan, J., 2019a. Graph attention convolution for point cloud semantic segmentation. In: Proc. CVPR, pp. 10288-10297. https://doi.org/10.1109/CVPR.2019.01054.

Wang, H. Y., Rong, X. J., Yang, L., Feng, J., Tian, Y., 2020. Weakly supervised semantic segmentation in 3D graph-structured point clouds of wild scenes. arXiv preprint arXiv:2004.12498. https://arxiv.org/abs/2004.12498.

Wang, L., Huang, Y., Shan, J., He, L., 2018. MSNet: multi-scale convolutional network for point cloud classification. Remote Sens. 10 (4), 612.

Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M., 2019. Dynamic graph CNN for learning on point clouds. ACM Trans. Graph. 38 (5), 146. https://doi.org/10.1145/3326362.

Wang, P.Z., Yao, W., 2022. A new weakly supervised approach for ALS point cloud semantic segmentation. ISPRS J Photogramm Remote Sens. 288, 237–254. https://doi.org/10.1016/j.isprsjprs.2022.04.016.

Wei, J.C., Lin, G.S., Yap, K. H., Hung, T., Xie, L., 2020. Multi-path region mining for weakly supervised 3D semantic segmentation on point clouds. In: Proc. CVPR, pp. 4383-4392. https://doi.org/10.1109/CVPR42600.2020.00444.

Wu, W., Qi, Z., Fuxin, L., 2019. Pointconv: Deep convolutional networks on 3D point clouds, In: Proc. CVPR, pp. 9613-9622. https://doi.org/10.1109/CVPR.2019.00985.

Xie, S. N., Gu, J. T., Guo, D. M., Qi, C. R., Guibas, L. J., Litany, O., 2020. PointContrast: Unsupervised pre-training for 3D point cloud understanding. In: Proc. ECCV, pp. 574–591. https://doi.org/10.1007/978-3-030-58580-8_34.

Xu, X., Lee, G. H., 2020. Weakly supervised semantic point cloud segmentation: Towards 10x fewer labels. In: Proc. CVPR, pp. 13703-13712. https://doi.org/10.1109/CVPR42600.2020.01372.

Yang, B., Han, X., Dong, Z., 2021. Point cloud benchmark dataset WHU-TLS and WHU-MLS for deep learning. J. Remote Sens. 25 (1), 231–240.

Zhang, Y. C., Qu, Y. Y., Xie, Y., 2021b. Perturbed self-distillation: Weakly supervised large-scale point cloud semantic segmentation. In: Proc. ICCV, pp. 15500-15508. https://doi.org/10.1109/ICCV48922.2021.01523.

Zhang, Y. C., Li, Z. H., Xie, Y., Qu, Y. Y., Li, C. H., Mei, T., 2021a. Weakly supervised semantic segmentation for large-scale point cloud. In: Proc. AAAI 35(4): 3421-3429.

Zhao H., Jiang L., Fu C. W., Jia J., 2019. PointWeb: Enhancing local neighborhood features for point cloud processing, in: Proc. CVPR, pp. 5560-5568, https://doi.org/10.1109/CVPR.2019.00571.

Zheng, Z.D., Yang, Y., 2021. Rectifying pseudo label learning via uncertainty estimationfor domain adaptive semantic segmentation. Int. J. Comput. Vis. 129 (4), 1106–1120. https://doi.org/10.1007/s11263-020-01395-y.