EXPLORING LABEL INITIALIZATION FOR WEAKLY SUPERVISED ALS POINT CLOUD SEMANTIC SEGMENTATION

Puzuo Wang[1], Wei Yao[1,2*]

[1] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong
[2] The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen, China

Commission II, WG II/3

KEY WORDS: Airborne Laser Scanning, Point clouds, Semantic segmentation, Weakly supervised learning, Data annotation, Feature extraction

ABSTRACT:

Although a number of emerging point-cloud semantic segmentation methods achieve state-of-the-art results, acquiring fully interpreted training data is a time-consuming and labor-intensive task. To reduce the burden of data annotation in training, semi- and weakly supervised methods are proposed to address the situation of limited supervisory sources, achieving competitive results compared to full supervision schemes. However, given a fixed budget, the effective annotation of a few points is typically ignored, which is referred to as weak-label initialization in this study. In practice, random selection is typically adopted by default. Because weakly supervised methods largely rely on semantic information supplied by initial weak labels, this studies explores the influence of different weak-label initialization strategies. In addition to random initialization, we propose a feature-constrained framework to guide the selection of initial weak labels. A feature space of point clouds is first constructed by feature extraction and embedding. Then, we develop a density-biased strategy to annotate points by locating highly dense clustered regions, as significant information distinguishing semantic classes is often concentrated in such areas. Our method outperforms random initialization on ISPRS Vaihingen 3D data when only using sparse weak labels, achieving an overall accuracy of 78.06% using 1‰ of labels. However, only a minor increase is observed on the LASDU dataset. Additionally, the results show that initialization with category-wise uniformly distributed weak labels is more effective when incorporated using a weakly supervised method.

1. INTRODUCTION

Airborne laser scanning (ALS) data depict the 3D structures of large-scale out-door scenes, which are used in a variety of remote sensing applications. To comprehensively interpret ALS data, an indispensable solution is to acquire category information, which is referred to as semantic segmentation or classification. Recently, an increasing number of deep learning methods have been developed for point-cloud semantic segmentation tasks, achieving state-of-the-art results (Thomas et al., 2019; Hu et al., 2020). However, most of these rely on a large number of precise annotations, which are typically associated with heavy workloads. In point-cloud labeling, the occlusions caused by the scan pattern of ALS systems and discrete data structure of point clouds in 3D space further increase the difficulty of visual interpretation. Additionally, with advances in light detection and ranging (LiDAR) technology, massive point clouds can be easily acquired from diverse platforms. Thus, finely annotating newly acquired point-cloud data is impractical.

To reduce labeling and alleviate data hungry issues, an intuitive solution is to annotate a small part of the entire data. Semi- and weakly supervised methods are proposed to address situations in which the number of labels is low. By exploiting extra information other than the original labels, these methods achieve competitive results compared with full supervision schemes. Several studies have recently emerged in point-cloud semantic segmentation (Xu and Lee, 2020; Hu et al., 2021). An important finding from these studies is that a huge redundancy exists in annotation information of fully labeled data. For example, only a minor accuracy degradation is observed by

comparing classification results using 10% of sparsely distributed labels with that using full labels. This means annotating every point is unnecessary during labeling work, also indicating the importance of weakly supervised learning. Although these methods achieve competitive results using limited labels, the issue of weak-label initialization is often ignored. In some studies, active learning was applied to guide the selection of weak labels (Polewski et al., 2016; Lin et al., 2020). However, the interaction process adds to the workload. To choose weak labels directly before classification, a commonly used strategy is to randomly select a fixed number/ratio of data to assign labels. However, the randomness of weak labels leads to unstable classification results. Additionally, because of the extremely imbalanced category distribution of ALS data, sampled points may be unclassified under a fairly sparse weak label configuration. Another common strategy is to evenly sample points for each class, which ensures sufficient samples across categories. Both strategies initialize weak labels using a random generator, which lacks guidance in selecting representative labels.

This study explores an effective weak-label initialization strategy. The feature space of the point cloud was constructed to guide weak-label selection. Because supervisory sources are absent in unlabeled points, handcrafted features are extracted to form a point-cloud description. Subsequently, a manifold learning approach is incorporated to eliminate feature correlation and construct a more effective feature space. A density-biased selection strategy is proposed to sample more points in the highly dense region in the feature space, where data are often considered to contain useful information. We use a clustering algorithm in advance and sample points in each cluster to avoid overly concentrated samples. After acquiring weak la-

---

* Corresponding author.

bels, classifiers were adopted to evaluate the effectiveness of the different weak-label initialization strategies. Moreover, we conducted experiments using a weakly supervised method for analysis.

2. RELATED WORK

2.1 Point cloud feature description

The feature description depicts discriminative information for each point. Local descriptors, such as FPFH (Rusu et al., 2009) and SHOT (Tombari et al., 2010) were successfully used in applications, such as registration and object classification. However, the results are unsatisfactory when adopting local descriptors in semantic segmentation tasks, as contextual information is essential for achieving accurate results. By contrast, geometric features, which represent 3D surface characteristics of each point, are popular in point cloud classification. A comprehensive study (Weinmann et al., 2015) was proposed to summarize a set of point-cloud geometric features. Self-supervised learning (SSL) has recently become a popular research topic in computer vision. By creating self-supervisory sources, such as data augmentation (Huang et al., 2021) and data completion (Wang et al., 2021), discriminative information is extracted for each sample using the trained model. Nonetheless, this remains a challenge when incorporating SSL into complex ALS data. Thus, this study uses handcrafted features.

Feature embedding is often proposed for effectively combining extracted features. It corresponds to dimension reduction, which projects the original features onto a lower-dimensional space. A classical method is principal component analysis (PCA), which produces new uncorrelated variables that successively maximize the variance. However, the fine structure cannot be well preserved using this linear transformation. In contrast, manifold learning can preserve the local structure after feature projection. t-SNE (van der Maaten and Hinton, 2008) and uniform manifold approximation and projection (UMAP)(McInnes et al., 2018) are two representative manifold learning methods. In point-cloud processing, Huang et al. (2020) classified ALS data using multi-scale local feature extraction with manifold embedding.

2.2 Weakly supervised learning

Weakly supervised point-cloud semantic segmentation has gradually attracted the attention of researchers. Under limited labels, studies have exploited potential information in unlabeled data. Xu and Lee (2020) proposed several strategies, including a siamese, inexact supervision, and smooth branch. An approximate result of fully supervised learning was obtained using 10% of the labels. However, the weak labels used were spatial aggregations of the downsampled full-scene labels, signifying a high labeling workload. Liu et al. (2021) constructed a supervoxel of point clouds to create pseudo-labels and applied an iterative training mode to improve the classification results under weak supervision. In Hu et al. (2021), a semantic query network was proposed to share sparse weak-label information in the spatial domain by interpolating features from neighboring points. Wang and Yao (2021b) proposed a pseudo-label-assisted approach for point-cloud semantic segmentation using limited annotations. This was enhanced in Wang and Yao (2021a), in which a plug-and-play weakly supervised framework was introduced, comprising entropy regularization, an ensemble prediction constraint, and online pseudo-labeling. Because of the

flexibility and competitive results achieved using only 1‰ of labels, it was used in this study to explore the effectiveness of weak-label initialization strategies.

3. METHODOLOGY

This study explores the effective weak labels that represent the most discriminative information. Given that the input points $\mathcal{P} \in \mathbb{R}^{N \times D}$ comprise N points with D dimensional features, the M(M≪N) points are assigned labels. Several weak-label initialization strategies were compared in this study. Random selection was the most practical method. Points were selectable either directly from the entire dataset or evenly from each category. A density-biased selection based on the feature space is proposed to provide effective information for weak-label initialization. The handcrafted features of a single point are extracted and projected onto the feature space using manifold learning. The points were then divided into clusters, and weak labels were sampled from each cluster. Subsequently, the weakly supervised method was integrated to evaluate the classification results using different weak-label initialization strategies.

3.1 Random initialization of weak labels

An intuitive strategy for initializing weak labels is to randomly select a certain number of points for annotation. In this study, we analyze two random initialization strategies, directly select points from the entire data set, and assign labels to the same number of points for each category. From a practical perspective, the former strategy is more effective because certain points to be labeled can be determined in advance using a random function. In contrast, to maintain an even category distribution, the operators must select points during the labeling process. However, the category-based strategy ensures an adequate number of points for each class under a fixed abundance, which avoids producing imbalanced weak labels. In the case of an equal total number of labeled points, the workload of the two strategies is almost the same.

3.2 Feature constrained weak-label initialization

Uncertainty in the random initialization strategy leads to diverse weak-label sets, which produce unrobust classification results. Thus, developing a targeted approach is essential for selecting more representative weak labels to achieve a higher accuracy using the same number of labels. Fig. 1 illustrates the framework of the proposed strategy. This study proposes unsupervised feature embedding to explore the relationships between points, guiding label initialization. Several handcrafted features are used to extract point-cloud discriminative information. We then incorporated manifold learning to analyze the correlations between the features and construct a point-cloud feature space. Because significant information often corresponds to highly dense regions in this space, a density-biased selection strategy is proposed for the selection of initial weak labels.

3.2.1 Unsupervised feature extraction Feature description plays a crucial role in point-cloud processing by supplying discriminative information for a variety of applications. Coordinate information is commonly used to explore point-cloud characteristics, and a set of geometric features is developed to construct point-cloud descriptors. Additionally, inherent physical attributes are directly combined with the geometric features.

This study focused on extracting single-point features, and a local neighborhood region was constructed for each point.
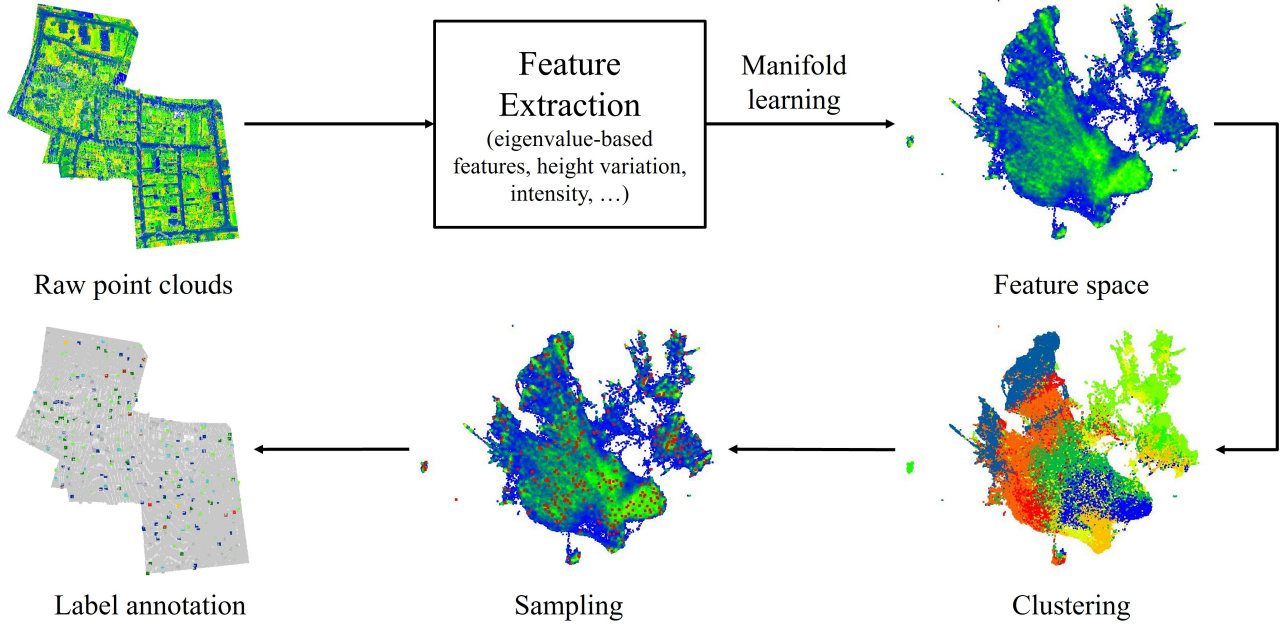
Figure 1. Framework of the feature-constrained weak-label initialization. Handcrafted features were developed to extract point-wise discriminative information. Subsequently, a manifold learning method was proposed to combine extracted features by projecting onto a more informative feature space. To ensure that the entire feature space was cross-sampled, the data were divided into clusters based on which weak labels are acquired using a density-biased selection strategy.

Eigenvalue-based features were used to represent the surface structure of a local region. Given a local neighbor set $\mathcal{P}_l \in \mathbb{R}^{k \times 3}$ of point p_i, the corresponding derived normalized eigenvalues e_λ with $\lambda \in \{1, 2, 3\}$ are calculated from the 3D covariance matrix of coordinates $\{X, Y, Z\}$. Subsequently, several 3D properties were acquired from the formulas of e_i. We used the following properties: omnivariance O_i, eigenentropy E_i, anisotropy A_i, linearity L_i, planarity P_i, scattering S_i, sum of eigenvalues Σ_i, and local curvature LC_i.

Height-based features are essential in ALS point-cloud feature description. For each point p_i, a local cylinder area is constructed to explore the height features. In this area, the height differences between p_i and the highest point, p_i and the lowest point, and the highest point and lowest point is extracted and denoted as Z_{max-i}, Z_{i-min}, and $Z_{max-min}$, respectively. Multiscale neighborhoods are used to supply sufficient height information in complex outdoor scenes.

Apart from the two types of geometric features, the density D_i and verticality V_i are combined. Density refers to the number of points within a local area, whereas the verticality relates to the normal vector of the $z-axis$, denoted as N_z. Additionally, the inherent physical attributes, such as reflectance R_i and color C_i, of the point cloud are integrated with the above features. Table 1 summarizes a combination of feature descriptions.

3.2.2 Manifold embedding Although the extracted features represent the characteristics of each point from different aspects, the correlations between features should be further analyzed. To reduce the redundancy of extracted features, parsing handcrafted features and projecting them onto a new set of descriptors is essential. Manifold learning, which can preserve local structure better than linear techniques such as PCA, is commonly used in feature embedding. Recently, UMAP(McInnes et al., 2018) was developed as a novel non-linear dimensionality reduction method. Local manifold ap-

Table 1. Feature description of a point cloud

Sort	Features	Definition
Eigenvalue	Omnivariance	$O_i = \sqrt[3]{e_1 \cdot e_2 \cdot e_3}$
	Eigenentropy	$E_i = -\sum_{i=1}^{3} e_i \cdot \ln e_i$
	Anisotropy	$A_i = (e_1 - e_3)/e_1$
	Linearity	$L_i = (e_1 - e_2)/e_1$
	Planarity	$P_i = (e_2 - e_3)/e_1$
	Scattering	$S_i = e_3/e_1$
	Sum	$\Sigma_i = e_1 + e_2 + e_3$
	Local curvature	$LC_i = e_3/(e_1 + e_2 + e_3)$
Height	Height difference	$\Delta H_{max-i} = Z_{max} - Z_i$
		$\Delta H_{i-min} = Z_{max} - Z_i$
		$\Delta H_{max-min} = Z_{max} - Z_{min}$
Other geometrics	Density	D_i
	Verticality	$V_i = 1 - N_z$
Physical attributes	Reflectance	R_i
	Color	C_i

proximations and local fuzzy simplicial set representations were used to construct topological representations of high-dimensional data in the UMAP. Given a set of data $\mathbf{x} \in \mathbb{R}^{N \times d}$ with d dimensional features, UMAP projects $\mathbf{x}$ onto a new feature space $\mathbf{y} \in \mathbb{R}^{N \times d'}(d' < d)$, which can maintain the local structure and, arguably, preserve more of the global structure. UMAP first builds a local fuzzy simplicial set $fs - set[x]$ from the original feature sets $\mathbf{x}$, which is based on k-nearest neighbors. Subsequently, it constructs the relevant weighted graph $\mathbf{A}$ and corresponding degree matrix $\mathbf{D}$, and a spectral embedding $\mathbf{y}$ is initialized from the sorted eigenvectors of the Laplacian matrix $\mathbf{L}$ of $\mathbf{A}$ and $\mathbf{D}$, denoted as

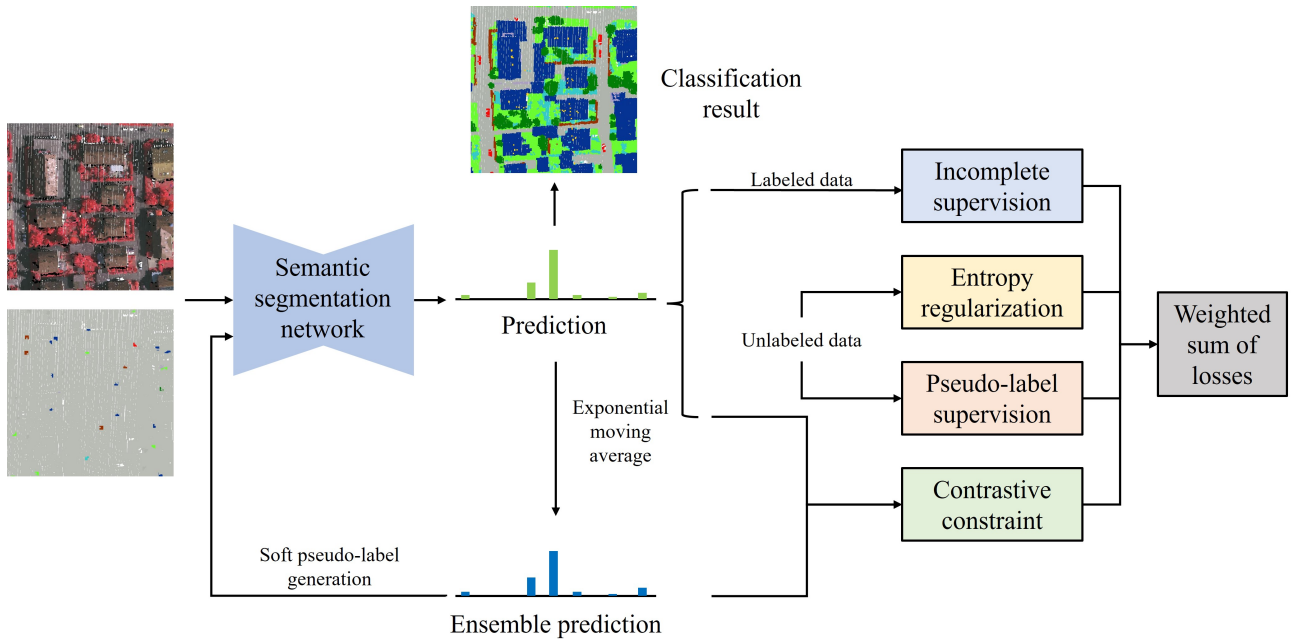$$\mathbf{L} = \mathbf{D}^{1/2}(\mathbf{D} - \mathbf{A})\mathbf{D}^{1/2} \qquad (1)$$

Figure 2. Schematic of the weakly supervised strategy for ALS point-cloud semantic segmentation. Apart from the incomplete supervision of labeled points, three modules are proposed: entropy regularization, contrastive constraint, and soft pseudo-labeling.

UMAP optimizes $\mathbf{y}$ with respect to $fs - set[x]$ cross entropy using stochastic gradient descent, and the optimized embedding $\mathbf{y}$ is the projected results. The target dimension was set to 10.

3.2.3 Density-biased selection We used information from the constructed feature space to initialize weak labels. In the feature space, points in the highly dense region are near the cluster center, which represents the most discriminative information. Thus, we sampled more points in this area to maintain the original density distribution. Specifically, the density is calculated using the number of neighbors in the radius search, and a weighted random selection is proposed. Note that directly choosing M highest-density points is ineffective, as those points are close, which means they represent similar information. Owing to the imbalanced label distribution in ALS data, a considerable discrepancy exists between the density values in different high-density areas. Thus, a clustering method was proposed, in which the points are proportionally sampled from each cluster. The Gaussian mixture model (GMM) (Reynolds, 2009) was integrated to divide the points into clusters. GMM uses the mixture of Gaussians:

$$p(\mathbf{x}|\lambda) = \sum_i^K \alpha_i g_i(\mathbf{x}|\lambda_i) \qquad (2)$$

where λ is the set of parameters $\{\alpha, p\}$, and $\sum_i^K \alpha_i = 1$ and each g_i is a Gaussian density function parameterized by α_i. Then, given a set of data $\mathbf{x}$, the objective is to find λ such that $p(\mathbf{x}|\lambda)$ is a maximum. The number of clusters was empirically set to be the same as that of dataset category. Subsequently, for each cluster $\mathcal{C}$, weighted random sampling was proposed, and the weight of each point was its density in the feature space. We sampled $M \cdot (|\mathcal{C}|/N)$ points to assign weak labels for each $\mathcal{C}$. Thus, the selected points contained useful information from each cluster, and the subset maintained a similar density distribution to that of the full set.

3.3 Weakly supervised learning

After acquiring weak labels, supervised classifiers evaluated the performance. We first analyzed the classification results using only weak labels. A classical machine learning method, random forest (RF) (Breiman, 2001), and a deep-learning network, KPConv (Thomas et al., 2019), were used as classifiers. For the RF, only labeled data were used for training. For KPConv, the entire data was fed to the training model, but the loss was is simply calculated from the labeled points. Owing to the imbalanced weak-label distribution, weighted cross-entropy was used to calculate the loss, which is defined as

$$\mathcal{L} = -\frac{1}{m} \sum_i^m w_i \sum_c^K y_i^c \log p_i^c \qquad (3)$$

where m is the number of labeled points in a training step. p_i^c is the predicted probability of point p_i and y_i^c represents the cth value of the one-hot vector of the label l_i. w_i relates to the category of p_i. Given $l_i = c$, w_i is calculated as $w_i = \frac{1}{N_c/M+0.02}$. N_c is the number of category c in weak labels.

We conducted experiments further using a weakly supervised method. This study adopted our previous study (Wang and Yao, 2021a), which integrates deep-learning networks. Fig. 2 illustrates the workflow. Under weak supervision, potential information in the unlabeled data was exploited to enhance the classification result. Entropy regularization (ER) was adopted to minimize class overlap and generate predictions with high confidence. Moreover, an ensemble prediction constraint (EPC) was developed to enhance the robustness of the trained model by comparing the prediction at the current training step to the ensemble value. In addition, the proposed method developed an online soft pseudo-labeling (OSPL) strategy to further improve the performance of the model. KPConv was the backbone network in our weakly supervised method.
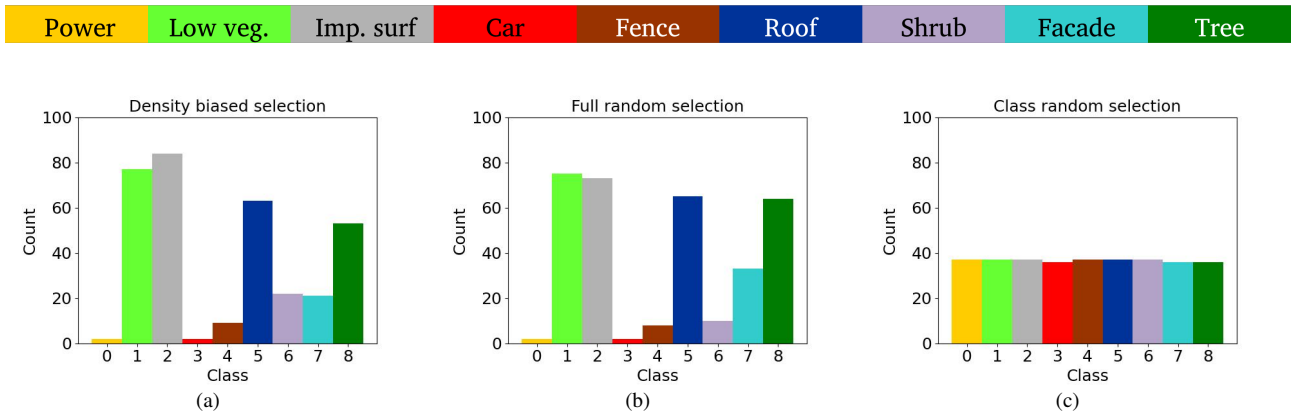
Power	Low veg.	Imp. surf	Car	Fence	Roof	Shrub	Facade	Tree

Figure 3. Comparison of class distribution of weak-label initialization.

4. EXPERIMENT AND RESULTS

4.1 Dataset

Experiments were conducted on the ISPRS Vaihingen 3D data (ISPRS) (Cramer, 2010; Rottensteiner et al., 2012) and large-scale ALS data for semantic labeling in dense urban areas (LASDU) (Ye et al., 2020; Li et al., 2013).

ISPRS dataset The dataset contains ALS data, and aerial images were obtained from the Stuttgart region of Germany. The dataset was divided into two parts: training and testing. The dataset included 9 categories. As multiple scan data were contained in the dataset, we set the subsampling grid size to d = 0.4 m to remove redundant points in overlapping ALS strips and maintain an even point density. The classification result of the deleted points is based on the nearest neighbor point during testing. The format of the used features is {X, Y, Z, Intensity, IR, R, G}.

LASDU dataset The study area is a valley along the Heihe River in northwest China, which covers an urban area of approximately 1 km². The area was divided into four connected sections, two of which were used as the training set, and the remaining two as the test set. Five categories were predefined in the dataset. Considering the even distribution and relatively low density of point clouds in the LASDU dataset, raw data were directly used for training and testing. The format of the used features is {X, Y, Z, Intensity}.

4.2 Implementation

In feature extraction, a spherical neighborhood was used to calculate eigenvalue-based features, and the radius was set to 2 m. For height-based features, we used three-scale cylindrical neighborhoods to explore height variation, and radii of 2, 4, and 6 m.

For the experiments using RF the classifier, we used 500 decision trees to acquire robust results using limited weak labels. For the experiments using KPConv, we sliced blocks with a radius of 20 m. The model was implemented in the PyTorch framework and trained on a GeForce GTX 1080Ti 11 GB GPU.

4.3 Evaluation metrics

The overall accuracy (OA) and F1 score were used to evaluate the performance of our method. OA is the percentage of predictions correctly classified, and the F1 score is the harmonic mean of the precision and recall, presented as:

$$precision = \frac{tp}{tp + fp},$$
$$recall = \frac{tp}{tp + fn}, \qquad (4)$$
$$F1 = 2 \times \frac{precision \times recall}{precision + recall},$$

where tp, fp, and fn are the true positives, false positives, and false negatives, respectively.

4.4 Results of ISPRS dataset

We analyzed the classification result using 1‰ of labels, corresponding to 330 points. Fig. 3 illustrates the category distribution of the weak-label initialization strategies. Compared to full random selection (FRS), our density-biased method (DBS) produces imbalanced weak labels. In contrast, class random selection (CRS) was used to produce samples with a uniform distribution of categories. Classification tasks were then conducted using produced weak labels. We first evaluated the effectiveness using only weak labels for training. A weakly supervised method was then incorporated for further comparison.

4.4.1 RF classification result Table 2 presents the comparison of the classification results. DBS achieved highest OA: 72.21%. Additionally, compared to FRS, DBS performed better in OA and Avg. F1. Although CRS achieved the highest Avg. F1 score, its OA value was considerably lower than that of the other two. This is because the number of weak labels of the dominant categories in CRS is much smaller, affecting the accuracy of these classes. Typically, the accuracy of dominant categories has a significant impact on the OA.

Table 2. RF classification results

Strategy	OA	Avg. F1
DBS	72.21±0.80	45.63±2.58
FRS	71.71±0.67	44.42±2.09
CRS	63.10±1.97	46.40±1.38
Full supervision	79.27	60.09

4.4.2 KPConv classification result Compared with the RF classifier, KPConv achieved much better results for every strategy, as presented in Table 3. Evidently, the DBS acquired both the highest OA and Avg. F1. For Avg. F1, the DBS

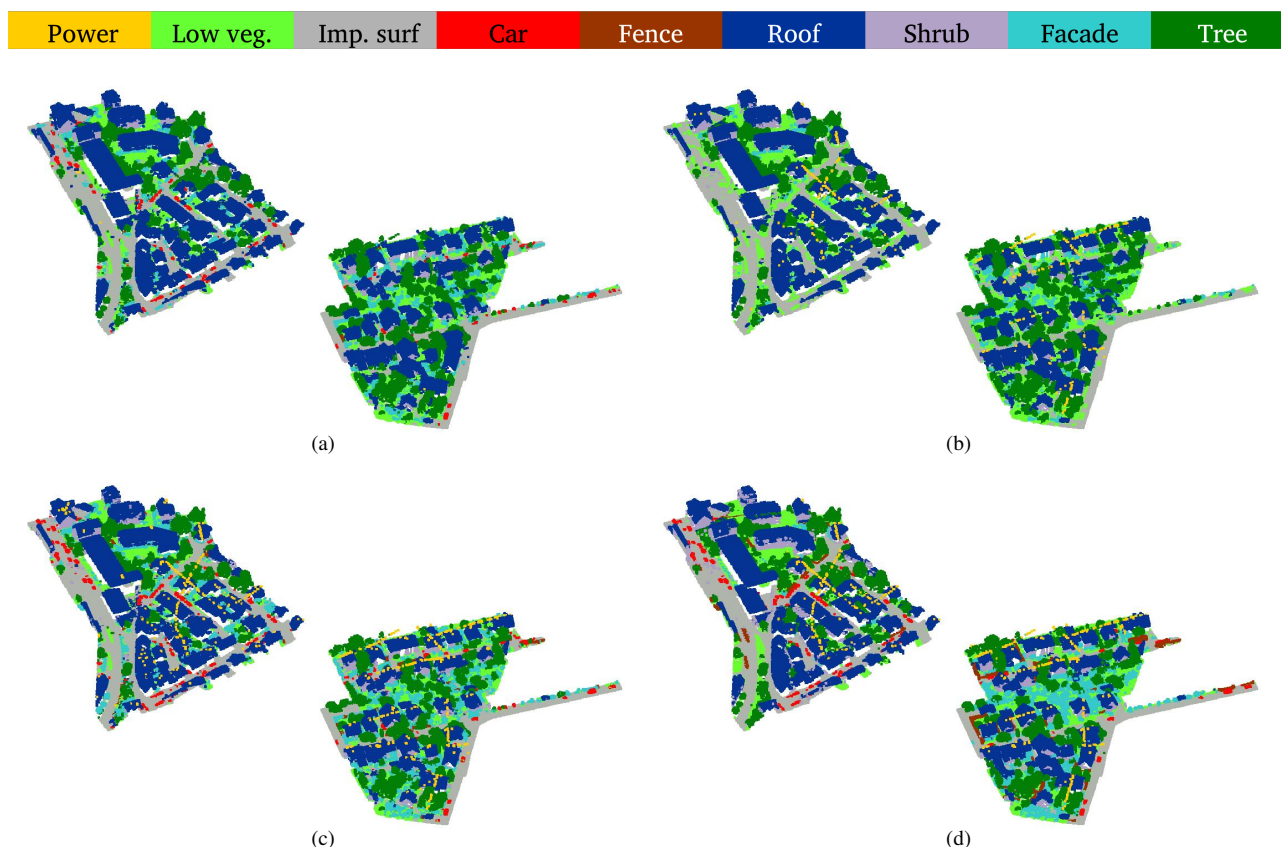| Power | Low veg. | Imp. surf | Car | Fence | Roof | Shrub | Facade | Tree |

Figure 4. Illustration of classification map of the ISPRS dataset. (a), (b), (c), and (d) represent the DBS, FRS, CRS, and ground truth, respectively.

achieved a 3% increase. In contrast, the CRS is far less effective. Comparing the two classifiers, that the DBS performs better when using only weak labels. Moreover, given the same abundance, samples with uniform distribution of categories achieve unsatisfactory results.

Table 3. KPConv classification results

Strategy	OA	Avg. F1
DBS	78.06±1.17	56.39±1.82
FRS	77.74±0.75	53.77±3.19
CRS	70.55±1.29	52.58±0.96
Full supervision	83.97	72.27

4.4.3 Weakly supervised classification result The objective of the weak-label initialization strategy is to provide more effective information and boost the performance of weakly supervised methods under the same abundance. Thus, evaluating the classification results is important when incorporating these methods. Table 4 presents the comparison. Using weakly supervised learning, the OA increased to a similar number for all strategies. However, a significant difference was present in the Avg. F1 score. While the CRS achieved a competitive result, 68%, degradation was observed in the other two strategies. Power lines, cars, and fences were severely misclassified into three categories. From Fig. 3, the number of weak labels in these categories is extremely small. As weakly supervised methods largely rely on initial weak labels, limited supervisory sources can cause confirmation bias, which can hinder performance. Fig. 4 presents the classification maps. For dominant categories, such as impervious surfaces and roofs, all strategies performed well. However, some categories were completely

misclassified in the DBS and FRS. In (a), no point is inferred as a power line, while almost all points classified as car in (b) were misclassified as low vegetation. In contrast, a satisfactory result was achieved for all categories in (c).

4.5 Results of LASDU dataset

The classification results using 1‰ of labels were evaluated, corresponding to 1694 points. Similar to the experimental evaluation of the ISPRS dataset, we compared the performance using FRS, DBS, and CRS.

4.5.1 RF classification result Table 5 presents a comparison of the classification results. DBS achieved the highest OA: 80.61%. Similar to the result from the ISPRS dataset, CRS achieved the highest Avg. F1 score and lowest OA. The results show that in unbalanced ALS data, evenly distributed weak labels lead to considerable performance degradation on OA when using only weak labels.

4.5.2 KPConv classification result Table 6 lists the classification results for KPConv. Compared with the FRS, improvement was not evident in the DBS. As the number of weak labels increases in the LASDU dataset, the uncertainty caused by random selection decreases, leading to the ineffectiveness of our density-biased strategy. CRS underperforms in both evaluation metrics.

4.5.3 Weakly supervised classification result Table 7 presents the comparison. From the table, the results are to similar to those from the ISPRS dataset. Although the three

Table 4. Comparison of full and weak supervision schemes on the ISPRS dataset

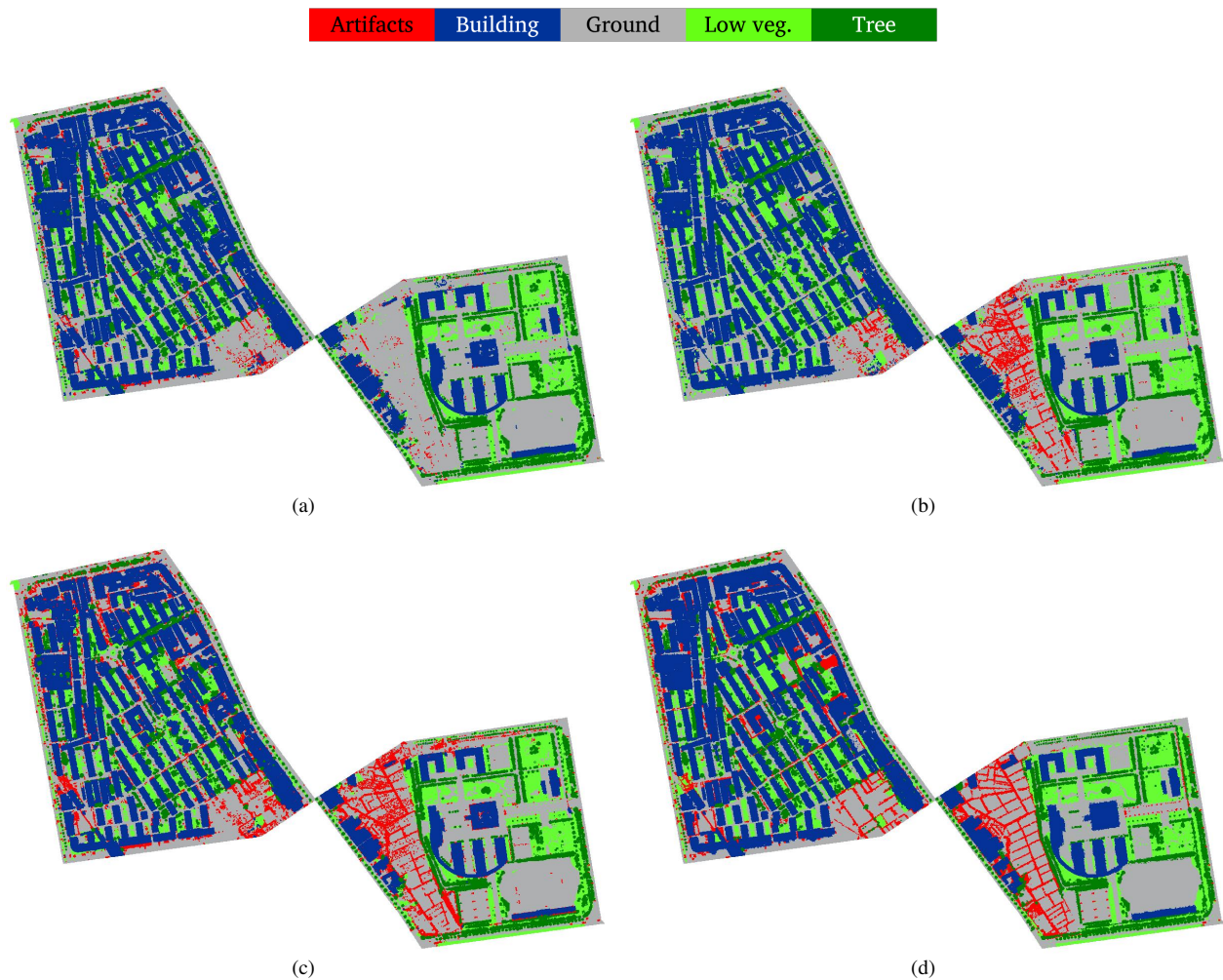| Method	F1 Score									Avg. F1	OA
	Power	Low veg.	Imp. Surf.	Car	Fence	Roof	Facade	Shrub	Tree		
DBS	3.46	77.19	89.59	33.19	5.69	94.75	59.03	14.39	81.25	50.74±3.50	81.90±0.49
FRS	27.10	80.23	90.39	0.21	0.04	95.04	56.26	30.22	80.92	51.16±1.79	82.54±0.16
CRS	74.67	79.07	89.27	78.49	15.43	93.80	61.52	40.74	79.96	68.10±0.76	81.66±0.98
Full supervision	76.25	81.99	91.36	79.57	34.77	94.55	61.77	47.90	82.28	72.27	83.97

Figure 5. Illustration of classification map on the LASDU dataset. (a), (b), (c), and (d) represent the DBS, FRS, CRS, and ground truth, respectively.

Table 5. RF classification results

Strategy	OA	Avg. F1
DBS	80.61±0.29	62.94±1.07
FRS	80.23±0.32	64.29±1.02
CRS	75.69±0.99	66.09±0.86
Full supervision	82.51	70.17

Table 6. KPConv classification results

Strategy	OA	Avg. F1
DBS	85.77±0.30	74.03±0.61
FRS	85.97±0.18	74.26±0.44
CRS	80.66±0.43	72.37±0.41
Full supervision	88.93	79.12

strategies achieved nearly identical results on OA, CRS significantly surpasses the other two on Avg. F1. In addition, performance degradation was observed for the Avg. F1 of both DBS and FRS. Fig. 5 presents the classification maps. The main classification gap between CRS and the other two lies in the resulting artifacts, and a large number of points that belong to this category are misclassified into other classes.

5. CONCLUSION

This study explores a weak-label initialization strategy for the semantic segmentation of ALS data. Apart from applying random selection on the whole set or evenly selecting for each category, a density-biased strategy was proposed that increases the probability of being selected in the highly dense region of the feature space. Although our method achieves better results using only limited weak labels, evenly distributed weak labels demonstrated more applicability when incorporated with

Table 7. Comparison of full and weak supervision schemes on the LASDU dataset

Method	F1 Score					Avg. F1	OA
	Artifacts	Building	Ground	Low veg.	Tree		
DBS	15.10	95.29	89.80	67.65	87.33	71.04±1.06	86.75±0.24
FRS	22.36	95.60	90.54	68.96	86.50	72.79±1.27	87.20±0.09
CRS	42.43	95.12	90.95	72.73	87.62	77.76±0.56	87.52±0.21
Full supervision	46.89	96.25	92.25	73.35	86.87	79.12	88.93

a weakly supervised method. Thus, the feature information of point clouds will be further explored in future works to improve weak label initialization and achieve robust and accurate classification results.

ACKNOWLEDGEMENTS

REFERENCES

Breiman, L., 2001. Random forests. *Machine learning*, 45(1), 5–32.

Cramer, M., 2010. The DGPF-Test on Digital Airborne Camera Evaluation Overview and Test Design. *Photogrammetrie - Fernerkundung - Geoinformation*, 2010(2), 73-82.

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. *arXiv preprint arXiv:2104.04891*.

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. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11105–11114.

Huang, R., Hong, D., Xu, Y., Yao, W., Stilla, U., 2020. Multi-Scale Local Context Embedding for LiDAR Point Cloud Classification. *IEEE Geoscience and Remote Sensing Letters*, 17(4), 721-725.

Huang, S., Xie, Y., Zhu, S.-C., Zhu, Y., 2021. Spatio-temporal self-supervised representation learning for 3d point clouds. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 6535–6545.

Li, X., Cheng, G., Liu, S., Xiao, Q., Ma, M., Jin, R., Che, T., Liu, Q., Wang, W., Qi, Y. et al., 2013. Heihe watershed allied telemetry experimental research (HiWATER): Scientific objectives and experimental design. *Bulletin of the American Meteorological Society*, 94(8), 1145–1160.

Lin, Y., Vosselman, G., Cao, Y., Yang, M. Y., 2020. Active and incremental learning for semantic ALS point cloud segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 169, 73–92.

Liu, Z., Qi, X., Fu, C.-W., 2021. One thing one click: A self-training approach for weakly supervised 3d semantic segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1726–1736.

McInnes, L., Healy, J., Saul, N., Grossberger, L., 2018. UMAP: Uniform Manifold Approximation and Projection. *The Journal of Open Source Software*, 3(29), 861.

Polewski, P., Yao, W., Heurich, M., Krzystek, P., Stilla, U., 2016. Combining Active and Semisupervised Learning of Remote Sensing Data Within a Renyi Entropy Regularization Framework. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(7), 2910-2922.

Reynolds, D., 2009. *Gaussian Mixture Models*. Springer US, Boston, MA, 659–663.

Rottensteiner, F., Sohn, G., Jung, J., Gerke, M., Baillard, C., Benitez, S., Breitkopf, U., 2012. THE ISPRS BENCHMARK ON URBAN OBJECT CLASSIFICATION AND 3D BUILDING RECONSTRUCTION. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, I-3, 293–298.

Rusu, R. B., Blodow, N., Beetz, M., 2009. Fast point feature histograms (fpfh) for 3d registration. *2009 IEEE International Conference on Robotics and Automation*, 3212–3217.

Thomas, H., Qi, C. R., Deschaud, J.-E., Marcotegui, B., Goulette, F., Guibas, L., 2019. Kpconv: Flexible and deformable convolution for point clouds. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 6410–6419.

Tombari, F., Salti, S., Di Stefano, L., 2010. Unique signatures of histograms for local surface description. K. Daniilidis, P. Maragos, N. Paragios (eds), *Computer Vision – ECCV 2010*, 356–369.

van der Maaten, L., Hinton, G., 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9(86), 2579-2605.

Wang, H., Liu, Q., Yue, X., Lasenby, J., Kusner, M. J., 2021. Unsupervised point cloud pre-training via occlusion completion. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 9782–9792.

Wang, P., Yao, W., 2021a. A new weakly supervised approach for ALS point cloud semantic segmentation. *arXiv preprint arXiv:2110.01462*.

Wang, P., Yao, W., 2021b. WEAKLY SUPERVISED PSEUDO-LABEL ASSISTED LEARNING FOR ALS POINT CLOUD SEMANTIC SEGMENTATION. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-2-2021, 43–50.

Weinmann, M., Jutzi, B., Hinz, S., Mallet, C., 2015. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing*, 105, 286-304.

Xu, X., Lee, G. H., 2020. Weakly supervised semantic point cloud segmentation: Towards 10× fewer labels. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 13703–13712.

Ye, Z., Xu, Y., Huang, R., Tong, X., Li, X., Liu, X., Luan, K., Hoegner, L., Stilla, U., 2020. LASDU: A Large-Scale Aerial LiDAR Dataset for Semantic Labeling in Dense Urban Areas. *ISPRS International Journal of Geo-Information*, 9(7).