An End-to-End Geometric Characterization-aware Semantic Instance Segmentation Network for ALS Point Clouds

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

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

Keywords: Point clouds, Semantic instance segmentation, Airborne Laser Scanning (ALS), Deep learning, Geometric characterization.

Abstract

Semantic instance segmentation from scenes, serving as a crucial role for 3D modelling and scene understanding. Conducting semantic segmentation before grouping instances is adopted by the existing state-of-the-art methods. However, without additional refinement, semantic errors will fully propagate into the grouping stage, resulting in low overlap with the ground truth instance. Furthermore, the proposed methods focused on indoor level scenes, which are limited when directly applied to large-scale outdoor Airborne Laser Scanning (ALS) point clouds. Numerous instances, significant object density and scale variations make ALS point clouds distinct from indoor data. In order to address the problems, we proposed a geometric characterization-aware semantic instance segmentation network, which utilized both semantic and objectness score to select potential points for grouping. And in point cloud feature learning stage, hand-craft geometry features are taken as input for geometric characterization awareness. Moreover, to address errors propagated from previous modules after grouping, we have additionally designed a per-instance refinement module. To assess semantic instance segmentation, we conducted experiments on an open-source dataset. Additionally, we performed semantic segmentation experiments to evaluate the performance of our proposed point cloud feature learning method.

1. Introduction

ALS (Airborne Laser Scanning) point cloud refers to a collection of 3D coordinate points obtained through the use of airborne LiDAR (Light Detection and Ranging) technology, which can represent the 3D structure of the terrain and objects on the Earth’s surface(Polewski and Yao, 2019). Instance segmentation on the ALS point clouds, meanwhile, serving as a crucial role for 3D modelling and scene understanding with a variety of applications like autonomous driving, augmented reality and robot navigation.

The development of instance segmentation have significant progress in recent years, which are driven by advancements in deep learning, computer vision algorithms and sensor technology. 3D instance segmentation from outdoor scene is a challenge task. Firstly, the label of instances have no fixed annotation like semantic class, making it hard to directly predict. Secondly, each scene contains different number of instance.

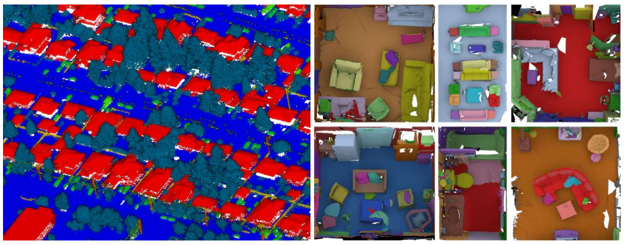

Figure 1. **Comparison between outdoor ALS point cloud and indoor scanning.** Numerous instances, significant object density and scale variations make ALS point clouds distinct from indoor data.

Recent advancements in 3D instance segmentation, as demonstrated by state-of-the-art methods such as 3D-SIS Hou et al. (2019), and SoftGroup Vu et al. (2022), have yielded significant progress. These methods employ two primary strategies: top-down and bottom-up. The top-down approach is well-suited for rapidly processing scenes, whereas the bottom-up approach excels in achieving high-precision segmentation for complex scenes.

In terms of data, the previous works primarily utilized indoor scanning data as input. However, to the best of our knowledge, there are still no study focused on ALS point cloud semantic instance segmentation. ALS point clouds typically encompass a wide range of outdoor scenes with complex instances. Unlike indoor RGB-D data, ALS point clouds often contain obstructed areas with sparse or no points due to scanning positions. Nevertheless, the presence of numerous instances, significant object density, and scale variations make it distinct (Figure 1). Furthermore, the boundaries between different categories in ALS point clouds are ambiguous and irregular. And in our instance segmentation task, the majority of input points are classified as background, indicating that only a few points should be grouped as instances.

In this study, we present an semantic instance segmentation network, which particularly considered the geometry characteristics of ALS point clouds and can be trained in an end-to-end manner. To ensure the quality of our results, we adopt a bottom-up strategy. Semantic predictions are utilized for instance mask proposals, and an another per-instance refinement module is employed for background points segmentation in each proposed instance. As a result, we place particular emphasis on the performance of semantic segmentation. The core idea of our network architecture is that we designed a geometric characterization-aware method for input points’ feature learning, which leads to better performance in semantic segmentation for distinguishing instance category and background. To address the issue of significant variations in object scale, our

https://doi.org/10.5194/isprs-archives-XLVIII-2-2024-435-2024 | © Author(s) 2024. CC BY 4.0 License.

grouping parameters are specially designed based on the average number of points per instance category. Moreover, due to the end-to-end training process, accumulated errors can be avoided. In summary, the key objectives of our work are as follows:

(1) We first introduce a semantic instance segmentation network for ALS point clouds, which can be trained in an end-to-end manner. Both semantic and objectness score are utilized to select potential points for grouping, followed by per-instance refinement module.
(2) We design a geometric characterization-aware feature learning network GFLN. The method leverages the geometric characteristics of point clouds to generate high-level feature representations, resulting in high-quality semantic and instance segmentation for ALS point clouds.
(3) We have conduct evaluation experiments to show that our work outperforms state-of-the-art semantic instance segmentation methods on labeled open-source dataset DALES Object-Singer and Asari (2021).

2. Related work

2.1 Point cloud feature learning

2.1.1 Hand-craft feature When processing point clouds for segmentation task, previous works Yao et al. (2012) and Amiri et al. (2017) focused on hand-craft features based on mathematical principles. Generally, it can effectively express the characteristics of points in a certain domain or condition. While the fixed explanation of algorithm leads to it heavily relies on computational parameters. As a result, the methods are difficult to apply in complex and ever-changing environments.

2.1.2 Deep learning Thanks to the development of deep learning technology, recent deep learning methods are capable to learn features directly from points. Based on the network architecture employed for feature learning at each point, representative methods can be categorized into three flows: (1) point-wise multi-layer perceptron (MLP), (2) convolution-based, and (3) graph-based. PointNetQi et al. (2017), as a pioneering work, proposed to conduct point-wise feature learning through shared MLPs and extract global features with a max-pooling layer. In terms of convolution-based methods, Thomas et al. (2019) proposed Kernel Point Convolution (KPConv) with both rigid and deformable kernel strategies, which achieved impressive results in the task of point cloud semantic segmentation. Graph-based works such as ECC Simonovsky and Komodakis (2017) and DGCNN Phan et al. (2018) considered point cloud as a graph with vertexes. Edges are generated based on the neighbors of each point. Then, feature learning will be conducted in the domain of vertexes and edges.

2.2 Point cloud instance segmentation

Point cloud instance segmentation form scene can be generally divided into three main methods: (1) proposal-based and (2) grouping-based and (3) dynamic convolution-based.

2.2.1 Proposal-based Proposal based methods typically involve generating proposals for where objects might be located in the point cloud and then refining these proposals to accurately segment the individual instances. And it's considered to be a top-down strategy. For instance, Yi et al. (2019) introduced a generative model for shape proposal. Hou et al. (2019) proposed 3D-SIS, which uses a 3D convolutional neural network to generate feature-rich embeddings for voxelized instance mask prediction. However, when processing complex scene with densely arranged objects, the strategy may hard to locate objects from instance proposal prediction.

2.2.2 Grouping-based Grouping-based methods is a bottom-up strategy that focus on clustering or grouping points belong to the same object instance. Unlike proposal-based methods that generate proposals, grouping-based methods directly segment the point cloud into clusters, each representing an individual instance. These methods often leverage per-points geometric or semantic predictions to perform the segmentation. Proposed by Pham et al. (2019) JSIS3D performs semantic segmentation and instance segmentation jointly. It uses a multi-value conditional random field (CRF) to enforce consistency between semantic and instance labels, that effectively grouped points into instances. However, simply apply the bottom-up strategy may leads to high objectness loss. Vu et al. (2022) introduced SoftGroup, which performs a bottom-up soft grouping followed by a top-down refinement. Semantic segmentation and instance offset prediction are conducted simultaneously. When performing semantic segmentation before grouping, the method allows the point to be associated the multiple class soft predictions to alleviate the propagation of errors to the subsequent processing. In summary, for the grouping based bottom-up strategy, utilize per-point predictions will make instance predictions more precise and finally refine themselves.

2.2.3 Dynamic convolution-based Dynamic convolution is a technique in convolutional neural networks that allows the shape and size of the convolutional kernel to change dynamically during the forward pass of the network. In 3D semantic instance segmentation, this strategy allows the point-wise convolutional kernel's shape to be adjusted, making the kernel instance-aware. For instance, techniques such as DyCo3D He et al. (2021) can effectively address the inevitable variation in the instance scales by generating instance-aware dynamic convolution kernels in the stage of point cloud feature learning.

Through out these works, they focused on indoor scene, which contain objects of similar size with less occlusion compared with ALS point clouds. Simply using the proposed methods on ALS point clouds is still far from satisfactory Han et al. (2024). Therefore, our method pay more attention to the characteristics of ALS data. On the one hand, in the stage of point cloud feature learning, we take some hand-craft features as input to enhance its geometric awareness. On the other hand, we developed a grouping-based network that specifically tailored the grouping parameters based on the average number of points per instance category. Moreover, as most of the points are background, we designed a semantic segmentation refinement module to enhance the background classification performance for each grouped instance proposal.

3. Method

The goal of our work is to take ALS point clouds as input and segment instances. Thus, we propose this end-to-end semantic instance segmentation network. Moreover, for ALS point clouds with special geometric features, we also introduce a novel strategy for 3D point feature learning. The overall architecture is illustrated in Figure 3, which consists of three main parts: semantic

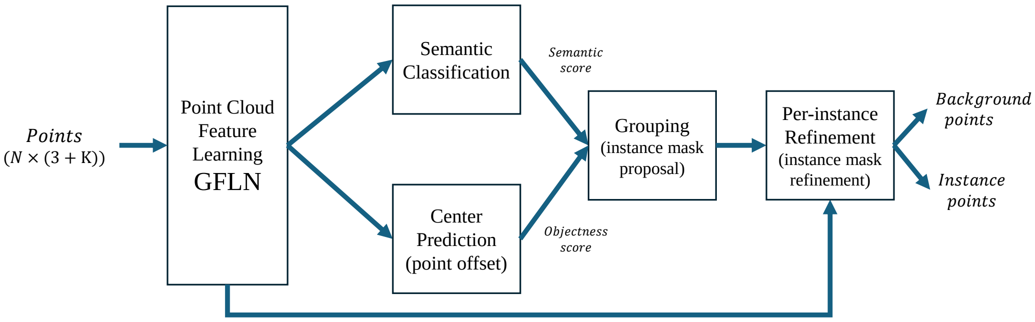
Points
(N × (3 + K))
Point Cloud Feature Learning
GFLN
Semantic Classification
Center Prediction (point offset)
Semantic score
Objectness score
Grouping (instance mask proposal)
Per-instance Refinement (instance mask refinement)
Background points
Instance points

Figure 2. **Overview of the proposed method.** Our network is consisted of the instance center prediction, semantic segmentation, and per-instance refinement modules. Taking ALS point cloud P with K extend dim features as input, instance proposals are obtained by semantic and center prediction modules. Features of each grouped instance are taken into per-instance refinement module for final instance outputs.

segmentation, instance center prediction, and per-instance refinement modules.

Specifically, given an input point cloud $P(x|f)$ with N points and extended by K-dim features. First, point-wise hand craft geometric characterizations are calculated for point cloud feature learning. Then, semantic segmentation and instance center prediction are conducted simultaneously for grouping preliminary instance proposals. Finally, per-instance refinement module is used to re-segment background points of grouped instances.

3.1 Geometric characterization learning

Numerous instances, significant object density and scale variations make the geometric characterization of ALS point clouds distinct. Moreover, compare with the indoor data, outdoor ALS point clouds including much more background points. But for our task of point cloud instance segmentation, the classification of background points becomes essential. For optimal performance, special attention should be paid to the geometric relationship at both local and global scales. Thus, we propose GFLN, a geometric characterization-aware feature learning network (illustrated in Figure 3) which is inspired from Li et al. (2020). Geometric characterization of a point can describe the shape, structure and topological properties. While it’s generic and low-level, which leads limitation of its ability to represent complex scenes. Thus, in GFLN, we take geometric characterizations as prior knowledge with a weight matrix as multi-layer perception (MLP) for learning and generating high-level features. Due to the task, we first get the point and its spherical neighbor area for following analysis. Normal vector N, and the first three eigenvalues $E(\lambda_1 > \lambda_2 > \lambda_3)$ of covariance matrix C in the area are chosen to define the input low-level feature $g_l[N, E]$. Simultaneously, rigid KPConv is adopted as backbone for feature learning of input original points according its impressive results on several open datasets. Figure 4 illustrates the comparison between KPConv (as baseline) and GFLN in the task of semantic segmentation. According to the results, the use of GFLN can result in a more precise division of boundaries between different categories, which can be highly beneficial for the subsequent task of instance grouping. Specifically, our method of point-wise feature learning is a convolution-based U-Net. To enhance both local and global geometric understanding, the radius of neighborhood area will expand after skip connection layer. In this work, GFLN is used for initial point cloud feature learning.

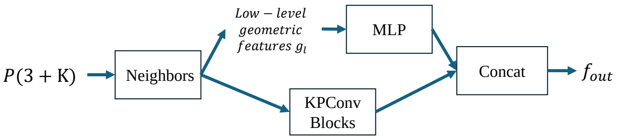

Figure 3. **Geometric characterization-aware feature learning network (GFLN).** For input point P within a spherical range. Low-level geometric features g_l are first be calculated, then it will go through a MLP to generate high-level feature g_h.

3.2 Semantic prediction branch

For all of the input N points’ semantic label prediction, we leverage a softmax layer to obtain the score vector $S = s_1, s_2, ..., s_n \in R^{N \times C}$ where C represents the number of semantic class. The predicted semantic score are supervised by weighted cross entropy loss and illustrated as

$$L_{sem} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j} t_{ij}\log(y_{ij}) \tag{1}$$

Where t_{ij} represents the true label of class j for sample i, and y_{ij} denotes the probability that model predicts sample i as class j.

3.3 Center prediction network

Inspired from VoteNet Qi et al. (2019), we learn the 3D offset from object center for each point. However, for ALS point clouds, the scale of objects from different categories vary significantly. To address this issue, our approach utilizes a 6-layer MLP with a pooling layer to enhance awareness of both local and global context features of points. The output offset vector $O = o_1, o_2, ..., o_n \in R^{N \times 3}$, that represents the x, y, z offsets from point to the geometric center of corresponding object. Shifted points are obtained according to the offsets prediction. Specially, for the background points, the ground truth offset is 0. Furthermore, the features of offset points will be leveraged to obtain the objectness score in the subsequent task. To evaluate the 3D offset o_i, we compare the predicted and ground truth center $y_i = x_i + o_i$ and $g_i = x_i + og_i$ to obtain whether the

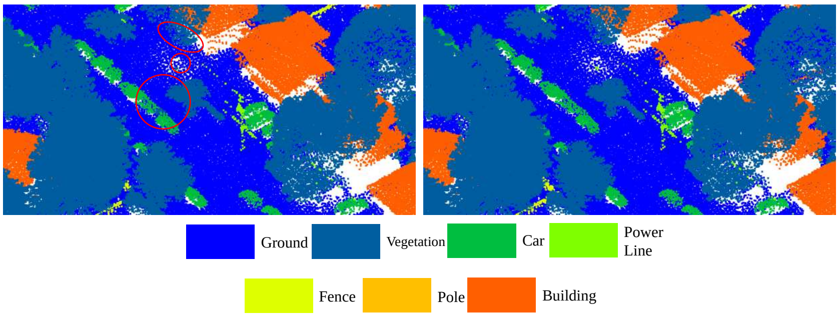

Figure 4. **Semantic segmentation comparison between KPConv (left) and GFLN (right).** GFLN have better performance at boundaries between different categories. KPConv made obvious incorrect predictions within the red circled area.

shifted points is on the object surface. Thus, the 3D offset can be supervised by a regression loss, which is denoted as:

$$L_{shift} = \frac{\sum_i min||Y - g_i|| * M}{M_{sum}} \tag{2}$$

Where Y is the vector of predicted instance center. M represents the ground truth mask of instance points. If point p_i belongs to an instance, $M_i = 1$, otherwise $M_i = 0$. M_{sum} is the ground truth total number of instance points.

3.4 Grouping instance

After point-wise semantic and center prediction, the results are used in the instance grouping stage. Initially, the features of offset points are filtered based on their semantic predictions to obtain subsets where all points within each subset belong to the same class. Then, features of offset points are utilized to generate class-wise objectness scores $S_{obj} = s_1, s_2, ..., s_m \in R^{M\times 1}$ where M represents the number of filtered subset points. Considering that the score represents if points belong to the instance classes, we opt sigmoid as activation function. Then, the points with object scores above a certain threshold t are regarded as positive prediction of object (potential points), which will subsequently perform DBSCAN grouping to get instance proposals. The operation of selecting potential points can improve the semantic precision of the grouping points. As a result, it largely prevents previous semantic errors propagate into the grouping stage. Considering significant variation in the object scale, for each category of instance, we apply different grouping parameters, which depends on the mean instance point number. The loss of objectness scores is calculated by mean squared error (MSE) of instant predictions, which is denoted as:

$$L_{obj} = \frac{1}{N}\sum_{i=1}^{N}(y_i - t_i)^2 \tag{3}$$

Where y and t represents prediction and ground truth labels respectively.

3.5 Per-instance refinement

The per-instance refinement stage reclassify and refines the instance proposals from the previous bottom-up grouping stage. To reduce the error propagated from the previous modules, an additional semantic prediction is conducted. It can be understand as a binary classification to classify background and object points, which take GFLN output features as input, then fed into 3 MLP layers. The output semantic score vector is $S_{refine} = s_1, s_2, ..., s_n \in R^{N\times 2}$. For loss computation in this stage, we adopt the same approach as initial point-wise semantic segmentation in section 3.2.

3.6 Loss and training process

The entire network can be trained end-to-end, with the loss propagated at each stage. The general loss computation is illustrated as:

$$L = \lambda_1 \times L_{shift} + \lambda_2 \times L_{obj} + \lambda_3 \times L_{sem} + \lambda_4 \times L_{refine} \tag{4}$$

Where vector λ donates the corresponding weights. Specifically, we set $\lambda_1 = 10$, $\lambda_2 = 4$, $\lambda_3 = 3$ and $\lambda_4 = 4$.

4. Experiments

4.1 Experiments dataset and preprocessing

In order to verify our work, we conduct experiments on a labeled open source dataset: DALES Object.

4.1.1 DALES Object dataset The DALES Object dataset is a large-scale aerial LiDAR point cloud dataset designed for semantic and instance segmentation tasks. It provides detailed annotations for various natural and man-made objects in urban and suburban environments, which include both semantic and instance-based labeling. The dataset includes over half a billion accurately labeled points covering an area of approximately 10 square kilometers.

We consider originally labeled 7 classes: ground, vegetation, car, power line, fence, pole and building in the experiments. For the task of semantic instance segmentation, we merge the classes of ground, power line and fence as background of instance class that will be ignored when processing. And the format of the utilized features was x, y, z.

4.2 Implementation details

4.2.1 Point cloud feature learning backbone In line with the KPConv method, our implementation includes encoder and decoder blocks (Figure 5). To mitigate gradient vanishing, we employ skip connections through feature concatenation. Within each block, as the neighborhood area radius of the points increases or decreases, down-sampling or up-sampling of points occurs to enhance the understanding of local and global knowledge. Additionally, batch normalization is utilized to improve training speed and stability.

Each batch consists of several spherical areas. Specifically, for the hand-crafted features of points, we adjust the radius of the

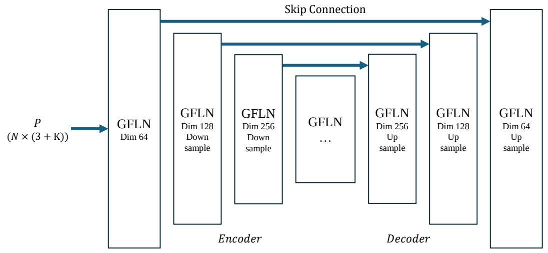

Figure 5. **Illustration of our point cloud feature learning backbone: GFLN’s architecture.** In the forward propagation process, feature dimension of points are transformed. Points of each layer experienced sampling operation. And skip connections are employed to mitigate the issue of vanishing gradients.

neighborhood area based on the radius of the batch spherical areas. Considering the point density, we have set the sampling resolution to 0.5m for the DALES dataset.

4.2.2 Instance grouping During the grouping stage, we employ DBSCAN grouping. Given the significant variation in the average number of object points, we set the grouping parameters based on the object size, as outlined in Table 1. Here, ”eps” denotes the parameter used to define the neighborhood radius, specifying the maximum distance threshold at which two samples are considered neighbors. Specifically, for the filtered subset points fewer than n_p, they will be grouped with $n_p = 6$.

Mean point number	eps	min point number
$n_{obj} >= 2000$	$r * 4$	$n_p/100$
$n_{obj} > 500$	$r * 6$	$n_p/35$
$n_{obj} <= 500$	$r * 8$	$n_p/20$

Table 1. 3D DBSCAN grouping parameters.

4.2.3 Instance merging For trained model validation, a problem arises due to the batch outputs consisting of spherical regions that may not cover the entire scene, resulting in one instance being segmented into parts across different spherical regions. To obtain the complete semantic instance segmentation result, these instances need to be merged. Our solution is as follows:
Let’s assume there is a predicted instance vector p_{ins} in a spherical region s_i of the batch inputs, and the previously predicted instances are stored in vector V_{ins}. First, we calculate the intersection of the two vectors. If there is a sufficient intersection with a stored instance, the two instances will be treated as one. Although the grouping strategy for each instance is based on semantic segmentation, the merge operation may lead to different semantic predictions within a single instance. Therefore, for classification consistency in predicted instances, we filter each instance based on the class with the most occurrences. The remaining points will be converted to background points (not part of any instance).

4.3 Evaluation metrics

4.3.1 Semantic segmentation evaluation For the task of point cloud semantic segmentation we adopt overall accuracy (OA) and F1 scores to evaluate the performance of our method (Equ. 5). Where OA represents is a measure of the proportion of correctly classified points, which provides a general assessment of the model’s performance across all classes. And F1 score is the harmonic mean of precision and recall, which is a single metric that takes into account both false positives and false negatives, making it a useful measure for imbalanced class distributions.

$$\begin{aligned} OA &= \frac{TP}{TP+FP+FN} \\ Precision &= \frac{TP}{TP+FP} \\ Recall &= \frac{TP}{TP+FN} \\ F1 &= 2 \times \frac{Precision \times Recall}{Precision + Recall} \end{aligned} \tag{5}$$

Where TP, FP and FN represents the number of true positive, false positive and false negative predicted points respectively.

4.3.2 Semantic instance segmentation evaluation For the task of semantic instance segmentation, we evaluate the mean class coverage $mCov$ (Equ. 6) and mean class-weighted coverage $mwCov$, which represent the average instance-wise intersection over union (IoU). In order to conduct comprehensive evaluation for the task performance, we test the predicted instances which obtain IoU more than $thre_{iou}$ from the scene (in our experiments, we set $thre_{iou} = 0.1$). Moreover, mean precision and recall of the predicted instances are also calculated in our work.

$$\begin{aligned} mCov &= \sum_{i=1} \frac{1}{M} max_j IoU(p_i, g_j) \\ mwCov &= \sum_{i=1} \frac{1}{M} w_i max_j IoU(p_i, g_j) \\ w_i &= \frac{n_i}{\sum_{j=1} n_j} \end{aligned} \tag{6}$$

Where $IoU(.,.)$ means the IoU between two point sets. p_i and g_i donate the predicted and ground truth instance point clouds. $max_j IoU(p_i, g_j)$ represents the highest IoU of ground truth instance point cloud g_j. M represents the number of instance prediction. And n_i is the point number of ground truth instance i.

5. Results and discussion

5.1 Semantic segmentation results

5.1.1 DALES Object dataset We compared the performance of our proposed point cloud feature learning backbone GFLN with the baseline method KPConv. Figure 7 illustrates results of the two methods. And the accuracy evaluation is shown in Table 3.

The results show that our proposed method for semantic segmentation reached the highest OA of 98.20%. Particularly, when dealing with limited training and testing samples, such as for car and pole classes in this dataset, the improvement is even more significant. Additionally, the results of semantic instance segmentation (GFLN SIS) indicate that the workflow enhanced the classification performance for classes that were previously challenging to classify. For instance, in the case of poles with sparse geometry distribution, employing GFLN as the backbone for point feature learning resulted in an increase in the F1 score from 0 to 0.054. Furthermore, with the constraints of the instance segmentation task, the F1 score further increased from 0.054 to 0.14.

Method	mCov	mwCov	mPre	mRec	mF1	mIoU
SoftGroup	0.575	0.519	**0.927**	0.544	0.686	0.522
Ours	**0.645**	**0.656**	0.905	**0.697**	**0.788**	**0.65**

Table 2. Accuracy evaluation of point cloud semantic instance segmentation on DALES Object dataset.

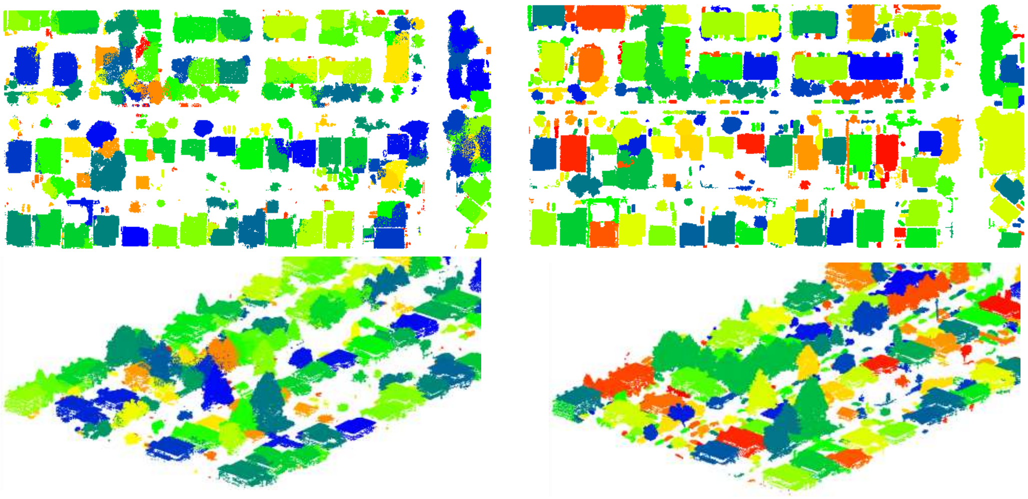

Figure 6. **Point cloud semantic instance segmentation performance on DALES Object dataset.** Ours (left), ground truth (right), where different instances are shown in different colors.

5.2 Semantic instance segmentation results

5.2.1 DALES Object dataset We conduct semantic instance segmentation by our proposed method. And for comparison, we choose SoftGroup Vu et al. (2022) as baseline, which utilized a bottom-up strategy to generate soft grouping proposals and then refines the results with a top-down per-instance refinement module. For each instance, although grouping strategy is based on semantic segmentation, the merge operation may leads to different semantic prediction in one instance. The segment results are depicted in Figure 6. Table 2 provides the general accuracy evaluation result. And Table 5 shows the class-wise evaluation results.

Upon analyzing the results, our proposed method demonstrates superior overall performance compared to the baseline, particularly for large-scale building and vegetation. However, the F1 score for cars was the lowest, despite high precision. This is likely due to the small number of points for each car object, leading to errors in predicting the instance center offset. During the grouping stage, some shifted points were disregarded, while others were grouped into different objects (Figure 8). We attribute this to subsampling, which resulted in low geometric resolution and ambiguity for small-sized instances. In the evaluation of vegetation objects, high precision but low recall was observed. Upon reviewing the ground truth label, we believe this is due to the subjective definition of ground truth vegetation objects (Figure 9), resulting in over-segmentation, particularly in low vegetation areas.

5.3 Ablation study

5.3.1 Per-instance refinement We compared the results of two models (on DALES Object dataset), one of which utilized per-instance refinement, while the other did not. Our experiments demonstrate that the per-instance refinement module has a positive impact, increasing mean class coverage and mmIoU in accuracy evaluation. We provide the comparison result in Table 4.

5.4 Discussions

5.4.1 Runtime analysis The training and validation are conducted in a same GTX 1080Ti GPU. Based on our testing on DALES Object dataset, for the task of point cloud semantic instance segmentation the average time for one step in a epoch is 8 seconds (with per-instance refinement module) and 5 seconds (without per-instance refinement module). And for the task of semantic segmentation the average time are decreased to 0.5 seconds.

5.4.2 Downstream work challenge: semantic instance completion Since our work follows a bottom-up grouping-based approach, after grouping the instance points, we can proceed with other downstream tasks such as instance completion.
Instance completion refers predicting missing part of 3D instances from incomplete or occluded 3D data. Methods for example, Yuan et al. (2018) proposed the first learning-based architecture PCN, which leveraged global feature from incomplete input point cloud to generate coarse result, and then predict detailed output via folding operation. In this case, we followed PCN, and tried to train an end-to-end point cloud semantic instance completion network.But during the implementation, we found it is still a challenge task.
(1)When working with outdoor data, it is not feasible to input the entire scene in a single batch. Consequently, objects near the boundary will be truncated, resulting in unavoidable structural deficiencies.
(2)During the end-to-end training process, the input points for the completion sub-network module consist of the output of the previous module, containing numerous error predictions that propagate into the subsequent completion network.
(3)Some scenes do not contain target instances that can be fed into the completion module, resulting in the inability to calculate loss for that batch, leading to gradient anomalies.
To ensure model convergence, we propose a potential solution. Inspired from Wang and Yao (2022), a prediction with a high posterior probability is typically more likely to be correct. There-

Method	OA	F1 Score						
		Ground	Vegetation	Car	Power line	Fence	Pole	Building
KPConv (Baseline)	98.00%	0.991	0.965	0.89	0.904	0.796	0	0.988
GFLN (ours)	**98.20%**	**0.992**	**0.969**	**0.908**	**0.911**	**0.862**	0.054	**0.989**
GFLN SIS (ours)	97.37%	0.99	0.957	0.834	0.791	0.744	**0.14**	0.982

Table 3. Accuracy evaluation of point cloud semantic segmentation.

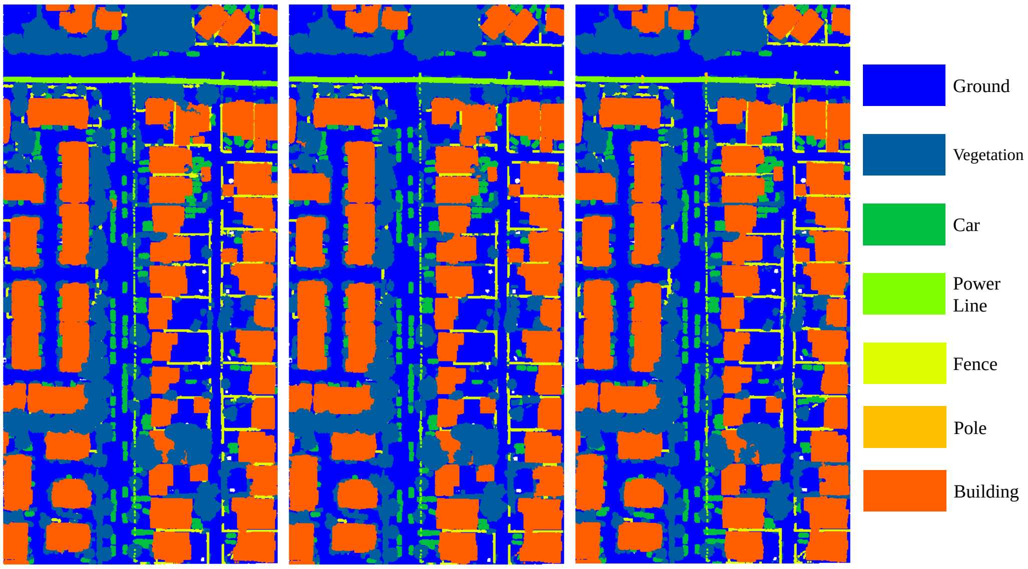

Figure 7. **Comparison of point cloud semantic segmentation performance on DALES Object dataset.** KPConv base line (left), GFLN (middle) and ground truth (right).

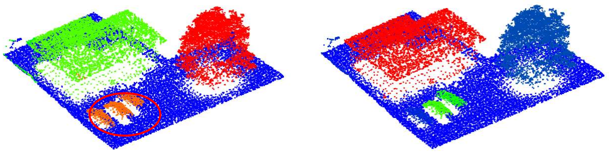

Figure 8. **Point cloud semantic instance segmentation results on DALES Object dataset.** Ours (left), ground truth (right). For relatively small size of instance like car within the red circled area, three of them are grouped as one instance.

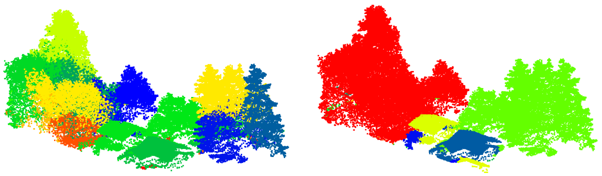

Figure 9. **Point cloud semantic instance segmentation results on DALES Object dataset.** Ours (left), ground truth (right). Our results segment individual trees, whereas the ground truth label combines trees that are close to each other as one instance.

fore, we define a soft instance proposal as an instance proposal with predicted scores exceeding a fixed threshold t. This operation ensures the high precision of the input instance points, enabling the subsequent completion tasks to proceed normally. By adopting the soft instance proposal strategy, the whole training will be divided into two steps. Step 1 aims to train soft instance proposal to feed into the completion network. Step 2 involves the completion training process to generate completed instances with semantic labels.

5.5 Limitations

Our method focuses on ALS point cloud semantic instance segmentation. While the framework achieved segmentation of different categories of objects in outdoor scenes, the overall performance is still relatively lower than in indoor scenes. We believe that future proposed networks can lead to further improvements. Throughout the entire semantic instance training process, the performance of semantic segmentation only showed a partial increase for certain classes compared to training with only the semantic segmentation network. However, we still believe that the overall performance of semantic segmentation will be enhanced by the instance segmentation module. Additionally, in the instance grouping stage, there is a sensitivity to parameters. When changing the scene domain, such as from a suburb to an urban area, the grouping parameters (see Table 1) should be reset, as the object attributes have changed significantly.

6. Conclusion

In this study, we have introduced an end to end geometric characterization-aware semantic instance completion network for ALS point clouds. The network incorporates hand-crafted geometry features into the point feature learning stage, resulting in a better understanding of the geometric relationship between points. Points offset to its corresponding instance center are learned for the task of instance segmentation. Both semantic and offset prediction are utilized to enhance the instance grouping. Moreover, a final per-instance refinement are conducted to

Per-instance refinement	mCov	mwCov	mPre	mRec	mF1	mIoU
No	0.483	0.496	**0.916**	0.521	0.664	0.497
Yes	**0.645**	**0.656**	0.905	**0.697**	**0.788**	**0.65**

Table 4. Ablation study on performance of per-instance refinement module.

Methods	Class	Precision	Recall	IoU	F1 Score
SoftGroup	Vegetation	0.881	0.63	0.574	0.697
	Car	0.887	0.577	0.514	0.641
	Building	0.955	0.669	0.638	0.737
Ours	Vegetation	0.893	0.693	0.642	0.74
	Car	0.904	0.609	0.553	0.68
	Building	0.941	0.793	0.745	0.825

Table 5. Class-wise accuracy evaluation of point cloud semantic instance segmentation on DALES Object dataset.

refine the instance proposal and semantic segmentation results. For future work, we intend to explore how to improve the overall instance accuracy and conduct the down stream task of semantic instance completion. We believe that the performance of semantic segmentation will be enhanced through instance completion. Furthermore, instances after completion are expected to exhibit improved performance in 3D modeling and scene understanding.

Acknowledgements

This work was supported by National Natural Science Foundation of China (Grant No.42171361) and by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. PolyU 15215023).

References

Amiri, N., Polewski, P., Yao, W., Krzystek, P., Skidmore, A., 2017. Detection of single tree stems in forested areas from high density ALS point clouds using 3D shape descriptors. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4, 35–42.

Han, X., Liu, C., Zhou, Y., Tan, K., Dong, Z., Yang, B., 2024. WHU-Urban3D: An urban scene LiDAR point cloud dataset for semantic instance segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 209, 500–513.

He, T., Shen, C., Van Den Hengel, A., 2021. Dyco3d: Robust instance segmentation of 3d point clouds through dynamic convolution. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 354–363.

Hou, J., Dai, A., Nießner, M., 2019. 3d-sis: 3d semantic instance segmentation of rgb-d scans. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4421–4430.

Li, W., Wang, F.-D., Xia, G.-S., 2020. A geometry-attentional network for ALS point cloud classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 164, 26–40.

Pham, Q.-H., Nguyen, T., Hua, B.-S., Roig, G., Yeung, S.-K., 2019. Jsis3d: Joint semantic-instance segmentation of 3d point clouds with multi-task pointwise networks and multi-value conditional random fields. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8827–8836.

Phan, A. V., Le Nguyen, M., Nguyen, Y. L. H., Bui, L. T., 2018. Dgcnn: A convolutional neural network over large-scale labeled graphs. *Neural Networks*, 108, 533–543.

Polewski, P., Yao, W., 2019. Scale invariant line-based coregistration of multimodal aerial data using L1 minimization of spatial and angular deviations. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152, 79-93.

Qi, C. R., Litany, O., He, K., Guibas, L. J., 2019. Deep hough voting for 3d object detection in point clouds. *proceedings of the IEEE/CVF International Conference on Computer Vision*, 9277–9286.

Qi, C. R., Su, H., Mo, K., Guibas, L. J., 2017. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 652–660.

Simonovsky, M., Komodakis, N., 2017. Dynamic edge-conditioned filters in convolutional neural networks on graphs. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3693–3702.

Singer, N. M., Asari, V. K., 2021. DALES Objects: A Large Scale Benchmark Dataset for Instance Segmentation in Aerial Lidar. *IEEE Access*, 9, 97495-97504.

Thomas, H., Qi, C. R., Deschaud, J.-E., Marcotegui, B., Goulette, F., Guibas, L. J., 2019. Kpconv: Flexible and deformable convolution for point clouds. *Proceedings of the IEEE/CVF international conference on computer vision*, 6411–6420.

Vu, T., Kim, K., Luu, T. M., Nguyen, T., Yoo, C. D., 2022. Softgroup for 3d instance segmentation on point clouds. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2708–2717.

Wang, P., Yao, W., 2022. A new weakly supervised approach for ALS point cloud semantic segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 188, 237–254.

Yao, W., Krzystek, P., Heurich, M., 2012. Tree species classification and estimation of stem volume and DBH based on single tree extraction by exploiting airborne full-waveform LiDAR data. *Remote Sensing of Environment*, 123, 368–380.

Yi, L., Zhao, W., Wang, H., Sung, M., Guibas, L. J., 2019. Gspn: Generative shape proposal network for 3d instance segmentation in point cloud. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3947–3956.

Yuan, W., Khot, T., Held, D., Mertz, C., Hebert, M., 2018. Pcn: Point completion network. *2018 international conference on 3D vision (3DV)*, IEEE, 728–737.