

Channel Attention and Normal-Based Local Feature Aggregation Network (CNLNet): A Deep Learning Method for Predisaster Large-Scale Outdoor Lidar Semantic Segmentation

Chang Liu^{id}, *Graduate Student Member, IEEE*, Linlin Ge, *Member, IEEE*, Wei Xiang^{id}, *Senior Member, IEEE*, Zheyuan Du^{id}, *Member, IEEE*, and Qi Zhang

Abstract—Predisaster information storage is crucial for effective disaster response. The discussion regarding deep learning-based light detection and ranging (Lidar) semantic segmentation technology for indoor small items has been ongoing in recent years. However, the methods applicable to large-scale outdoor Lidar datasets for predisaster information storage remain limited. This study aims to propose a novel deep learning-based network for city-scale Lidar semantic segmentation to support predisaster information storage, called channel attention and normal-based local feature aggregation network (CNLNet). This network is designed to segment common urban land cover objects, including buildings and vegetation. This network incorporates surface normal information and the channel attention (CA) mechanism into the RandLA-Net backbone. Ablation studies have been devised to assess the performance of these two features. During the preprocessing step, color information from optical images is fused with Lidar data. The findings demonstrate that CNLNet can enhance the accuracy of the RandLA-Net backbone by improving mean intersection over union (mIoU) by at least 1%–2%. Including one of these two features also contributes to the backbone's improved accuracy. Notably, CNLNet outperforms other well-known networks in terms of accuracy with the test of the public Sematic3D dataset. This study further reveals that the proposed network excels in building segmentation, a crucial facet of predisaster information storage. Moreover, the results show that spatial resolution, whether at 0.5 or 10 m per pixel for optical images, has limited influence on outcomes. One theoretical contribution of this study is the demonstration of the advantages of integrating either surface normal information or a CA mechanism to enhance large-scale

outdoor Lidar semantic segmentation. Labeled Lidar datasets have been created for training. The practical contribution is that it can optimize disaster response by efficiently facilitating predisaster information storage.

Index Terms—Deep learning, light detection and ranging (Lidar), predisaster, semantic segmentation.

I. INTRODUCTION

THE frequency of destructive natural disasters is on the rise due to the increasing occurrence of extreme weather events attributed to climate change. Natural disaster management has gained significant global attention recently [1]. To avoid the disastrous and chaotic aftermath, pre-emptive measures are valuable before the impact of a disaster. Predisaster information storage allows postdisaster decision-makers to strategize rescue routes and determine suitable locations for temporary housing, thus enabling swift disaster response.

As a component of predisaster information storage, the retention of predisaster urban land cover visualization data is invaluable for disaster analysis reconnaissance [2]. These data should be stored and periodically updated to expedite postdisaster analysis and management processes. However, conventional in situ data collection methods have several issues, including being labor-intensive, time-consuming, costly, and potentially dangerous. Remote sensing technology offers a swift and efficient alternative for urban land cover visualization data collection due to its capacity to acquire extensive data on a large scale with relative ease.

Light detection and ranging (Lidar) has recently gained significant attention in remote sensing because of its 3-D information and higher vertical accuracy with better penetration than conventional photogrammetry. Compared with conventional in situ urban data collection methods, Lidar usually spends less time, which helps operators save time and labor costs [3]. Due to the rapid development of deep learning, there has been a burgeoning interest in its application to remote sensing-based Lidar semantic segmentation in recent years [4]. Therefore, a deep-learning-based Lidar semantic segmentation could solve rapid predisaster land cover visualization data collection and storage.

Unlike 2-D imagery, Lidar point cloud data belong to non-Euclidean geometry data. Therefore, semantic

Manuscript received 11 October 2023; accepted 18 November 2023. Date of publication 5 December 2023; date of current version 19 December 2023. This work was supported by SmartSat Cooperative Research Centre under Grant P2.30s, whose activities are funded by the Australian Government's CRC Program. (*Corresponding author: Linlin Ge.*)

Chang Liu and Linlin Ge are with the School of Civil and Environmental Engineering, Faculty of Engineering, University of New South Wales, Sydney, NSW 2052, Australia (e-mail: chang.liu17@unsw.edu.au; l.ge@unsw.edu.au).

Wei Xiang is with the School of Computing, Engineering and Mathematical Sciences, La Trobe University, Melbourne, VIC 3086, Australia (e-mail: w.xiang@latrobe.edu.au).

Zheyuan Du is with the School of Civil and Environmental Engineering, Faculty of Engineering, University of New South Wales, Sydney, NSW 2052, Australia, and also with the Minerals, Energy and Groundwater Division, Geoscience Australia, Canberra, ACT 2601, Australia (e-mail: z.du@unsw.edu.au).

Qi Zhang is with the School of Engineering and Design, Technical University of Munich, 80333 Munich, Germany (e-mail: rachelqi.zhang@tum.de).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TGRS.2023.3339475>, provided by the authors.

Digital Object Identifier 10.1109/TGRS.2023.3339475

segmentation methods for 3-D data cannot simply be decreased to 2-D segmentation. Reducing the dimensionality from 3-D to 2-D inevitably results in the loss of information. To design deep learning methods suitable for 3-D semantic segmentation while retaining the inherent 3-D data, the development of point-based networks began in 2017.

In 2017, PointNet directly took points as its input, which was the first point-based network. It learns features with a shared multilayer perceptron (MLP) [5]. Nevertheless, the local structures and the mutual interactions between features cannot be extracted by a shared MLP in PointNet [5]. To learn richer local geometry in point clouds and capture a broader context for each point, several methods have been introduced to develop PointNet, such as neighboring feature pooling. In particular, PointNet++ was proposed soon after the generation of PointNet to categorize points hierarchically and progressively learn from larger local regions. It achieved better results than PointNet according to the conducted experiments [6]. Following PointNet++, Jiang et al. [7] introduced a PointSIFT module to stack and encode the point information from eight spatial orientations using a three-stage-ordered convolution process.

Given the rapid advancements in point-based deep learning methods for 3-D semantic segmentation, certain scholars have commenced discourse on the topic of large-scale outdoor Lidar semantic segmentation (LOLSS). For instance, RandLA-Net was proposed for LOLSS as a lightweight network for saving processing time [8]. It applies random point downsampling to attain a high level of efficiency in memory and computation. A local feature aggregation (LFA) unit was further proposed to capture and retain geometric features. However, there is still a lack of enough studies for large-scale scenarios. Most advanced networks are still only designed for small or indoor scenes [5], [6].

Moreover, there remains a significant gap in the full-fledged development of semantic segmentation techniques aimed at storing predisaster land cover information. Specifically, several possible methods have not been fully discussed for disaster-related research, and deep learning networks have not been extensively trained to account for the potential occurrence of natural disasters in the geographical areas highlighted in the selected datasets. Therefore, there is a lack of efficient and accurate Lidar semantic segmentation methods that can classify predisaster large-scale land cover classification.

In order to solve these problems, this study aims to provide a deep learning LOLSS network by creating a dataset tailored to the targeted task to store the 3-D information of predisaster large-scale outdoor land cover objects.

II. METHODOLOGY

A. Data and Study Extents

This study chose four own labeled places and one public dataset to test the proposed network. The own labeled places include Kapiti Coast, Tasman, Nelson, New Zealand, and Kumamoto, Japan. These four places were chosen because they are both tectonically active urban areas near the sea. These sites present challenges for in situ observations and pose

TABLE I
ORIGINAL LABELED CLASSES OF THE DATA

Lidar dataset	Number of labeled classes	Labeled classes
Kapiti Coast, Tasman, Nelson	5	Ground, low vegetation, medium vegetation, high vegetation, buildings
Kumamoto (pre-earthquake)	4	Ground, low vegetation, medium vegetation, high vegetation

multiple natural hazard risks [9]. Moreover, three of them have already been the sites of significant natural disasters. Continuous heavy rain caused severe landslips and flooding in Tasman and Nelson in August 2022 [10], and a severe earthquake occurred on April 16, 2016, in Kumamoto [11]. In this study, we specifically chose datasets collected in proximity to the timing of these floods and before the earthquake event.

The original labeled classes from these own labeled datasets are listed in Table I. All unlabeled point clouds were ignored during experiments. The original Lidar point clouds of these places do not include color information, so this study needs to add corresponding colors to Lidar data in the preprocessing step. The color information of red, green, and blue (RGB) bands from optical images is a viable choice for finishing this task. Therefore, this study collected optical images from the same places of the Lidar datasets to fuse 3-D Lidar and 2-D images. The detailed preprocessing steps for each place are introduced in Section II-B.

Lidar data and optical images with RGB bands of these four places are shown in Table II. Sentinel-2 (S2) images of all places were collected. KOMPSAT-3 (K3) images for the 2016 Kumamoto pre-earthquake Lidar data were also collected to test the influence of image resolution on the performance of the proposed network (refer to Section II-D). Since the Lidar data of the datasets were collected by different organizations, their parameters are different. Considering this, the Kapiti Coast, Tasman, and Nelson datasets were applied for both deep learning training (and validation) and testing stages, while the Kumamoto dataset was only utilized in the testing stage to test the generalizability of the networks trained with the other datasets.

Semantic3D is a large-scale open-source dataset. It was chosen to compare the accuracy of the proposed method and other well-known deep learning networks for Lidar semantic segmentation.

Section II-A1 introduces detailed information on the three datasets from New Zealand. Section II-A2 introduces the Kumamoto dataset collected before the 2016 Kumamoto Earthquake. Section II-A3 introduces the Semantic3D that is applied in this study.

1) *Kapiti Coast, Tasman, and Nelson in New Zealand*: The 49 selected patches of point cloud data from New Zealand were selected in this study, as shown in Fig. 1. The 26, 7, and 16 are from Kapiti Coast, Tasman, and Nelson, respectively. The number of Lidar data was chosen because of considering

TABLE II
DATA INFORMATION

Location	Potential disaster	Point cloud			Optical image		
		Data	Point density	Acquired date	Satellite	SR, m/pxl	Acquired date
Kapiti Coast	Earthquakes, storms, floods, landslides	Lidar	27.95 pts/m ²	March 13 to 15, 2021	Sentinel-2	10	July 24, 2021
Tasman	Floods	Lidar	15.05 pts/m ²	August 23, 2022, to September 6, 2022	Sentinel-2	10	September 15, 2022
Nelson							
Kumamoto pre-earthquake	Earthquakes, storms, floods, landslides	Lidar	2.94 pts/m ²	April 15, 2016	KOMPSAT-3	0.5	April 15, 2016
					Sentinel-2	10	March 03, 2016

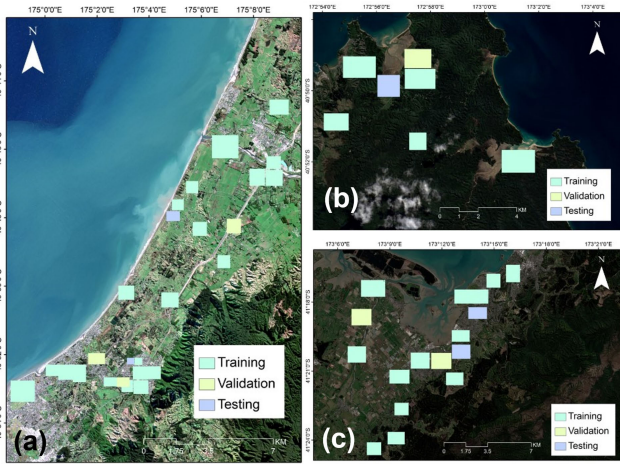


Fig. 1. Locations of datasets in New Zealand: (a) Kapiti Coast; (b) Tasman; and (c) Nelson.

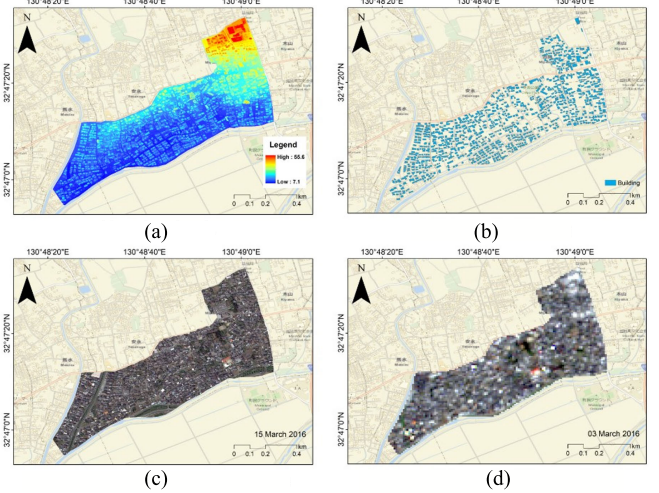


Fig. 2. Data in the Kumamoto pre-earthquake dataset: (a) point clouds; (b) building outlines; (c) K3 image shown in RGB bands; and (d) S2 image shown in RGB bands.

the number of Semantic 3-D data [12] applied in RandLA-Net [8] since this study is developed from RandLA-Net. The training, validation, and testing data are shown in indicolite green, olivine yellow, and sugilite sky colors in Fig. 1. The dataset contains five classes labeled by experts from the data provider, as shown in Table I, including ground, low vegetation, medium vegetation, high vegetation, and buildings [13], [14]. All these labels are kept in this study as the information of all these classes is necessary for recovery plans.

This study chose S2 images for color fusion because it is free and easy to access. After checking all S2 data with the date near the dates of Lidar collection, the dates of S2 images were chosen, as shown in Table II. The images of other dates either contain several clouds or are in the dark.

2) *Kumamoto Pre-Earthquake Dataset*: A mainshock of the 7.0-M_W Kumamoto earthquake struck on April 16, 2016. Four types of pre-earthquake data in Kumamoto were utilized in this study, as shown in Fig. 2, including a Lidar point cloud [Fig. 2(a)], a building outline shapefile [Fig. 2(b)], an optical image from K3 satellite [Fig. 2(c)], and an optical image from S2 satellite [Fig. 2(d)]. The color from blue to red shown in Fig. 2(a) represents the increase in elevation. The original labeled classes in the Lidar point clouds were ground, low vegetation, medium vegetation, and high vegetation. The

building class is an integral part of predisaster information collection, but the original Lidar dataset did not have this class, so the building footprint information was added to the Kumamoto dataset during the preprocessing, which will be introduced in Section II-B2. Optical images from two satellites with different resolutions to test if the image spatial resolution will influence the accuracy of the proposed network. K3 is 0.5 m per pixel, and S2 is 10 m/pixel.

3) *Semantic3D Dataset*: Semantic3D dataset is one of the most popular open-source point cloud datasets for deep learning semantic segmentation. Eight labeled classes from this dataset were chosen in this study, including natural terrain, high vegetation, low vegetation, buildings, hardscape, scanning artifacts, and cars. Four point clouds were selected for the network test according to the design of the RandLA-Net backbone.

B. Data Preprocessing: Data Fusion of Lidar Data With Satellite RGB Data

The main task of this step was to incorporate color information into Lidar data. Although the Lidar coordinate systems varied among different datasets, this study disregarded these

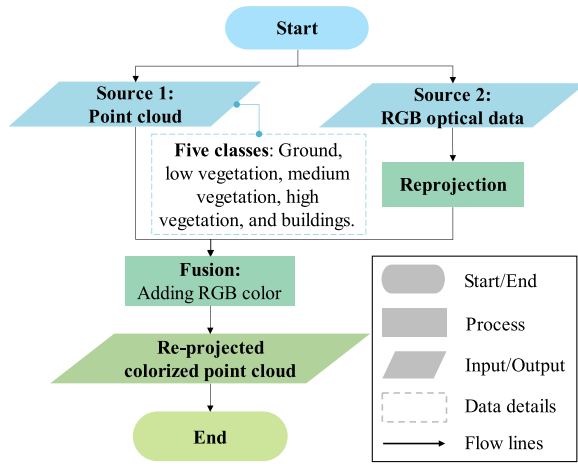


Fig. 3. Workflow of Kapiti Coast data fusion with adding RGB information.

differences during training. However, it is essential to ensure that the coordinate systems of satellite images and Lidar data within the same dataset are consistent. Therefore, all other data, regardless of whether they were in projected or geographic coordinate systems, were transformed to match the coordinate system of the Lidar data.

1) *Preprocessing of the Three Datasets in New Zealand:* Some preprocessing steps were applied before training the deep learning network, as shown in Fig. 3. Since the original Lidar point data do not have colors, this study fused 2-D optical images and 3-D point clouds to obtain the color point cloud using feature manipulation engine [15]. First, the point clouds were loaded. Second, the color optical data were reprojected to the same coordinate system as the Lidar. The optical image was collected from S2, and only RGB bands were applied in this study. Then, the point cloud and the RGB image were fused to obtain the colored point cloud.

2) *Preprocessing of Kumamoto Pre-Earthquake Data:* The original noncolor Lidar dataset contains only four classes without the building class. Since information about the building class and colors is essential for this study, both color bands and building outlines were fused into Lidar point clouds in feature manipulation engine, as illustrated in Fig. 4.

The first fusion is adding building outlines in Lidar. The coordinate system of building outline polygons was reprojected to the coordinate system of the Lidar data, and then, point clouds in building outlines were classified as “building.”

The second fusion is adding color. K3 and S2 images were reprojected to match the system of Lidar. Then, the reprojected RGB bands from clipped optical images were fused with Lidar. The fusion results were the reprojected colored Lidar point cloud dataset with the five classes.

C. Channel Attention and Normal-Based LFA Network

This study proposed a deep learning method called channel attention and normal-based local feature aggregation network (CNLNet) for LOLSS. The main improvements include adding normal information and the channel attention (CA) mechanism in the backbone.

CNLNet adds these two possible helpful approaches in the backbone to increase the accuracy of LOLSS. Surface normal

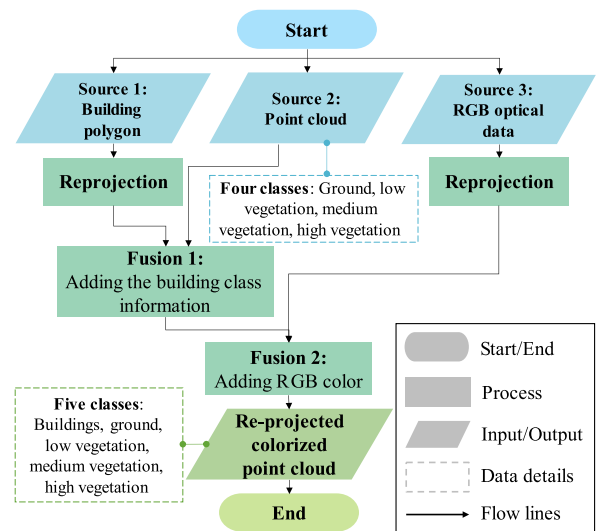


Fig. 4. Workflow of Kumamoto pre-earthquake data fusion with adding the building class and RGB information.

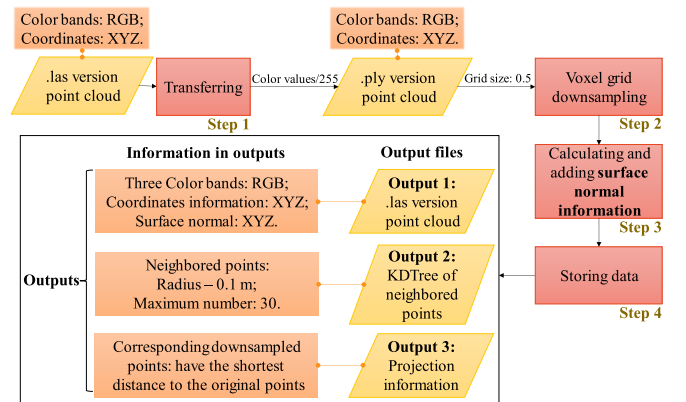


Fig. 5. Workflow of data preparation with adding surface information.

information is important in several point cloud applications. The attention mechanism is widely confirmed effective in 2-D or small-scale 3-D deep learning networks. However, they are not always applied in large-scale predisaster scenarios. Therefore, this study added these two to enhance the backbone and developed a module in it.

1) *Surface Normal Information Addition and Data Preparation:* This section introduces how to calculate the surface normal during data preparation. The collected, revised colored Lidar data (refer to Section II-B) required further processing for data preparation before the training stage.

Surface normal information is one of the essential properties of a geometric surface, and it finds applications in various research areas. For instance, in computer graphics, light rendering depends on normal information to generate shadings and other visual effects to look more realistic. Therefore, this study evaluated the impact of surface normal information on enhancing the accuracy of semantically segmenting large-scale point clouds. The process of adding normal information involves four main steps: data format transformation, voxel grid downsampling, computation of normal estimation, and data storage, as depicted in Fig. 5. Voxel grid downsampling

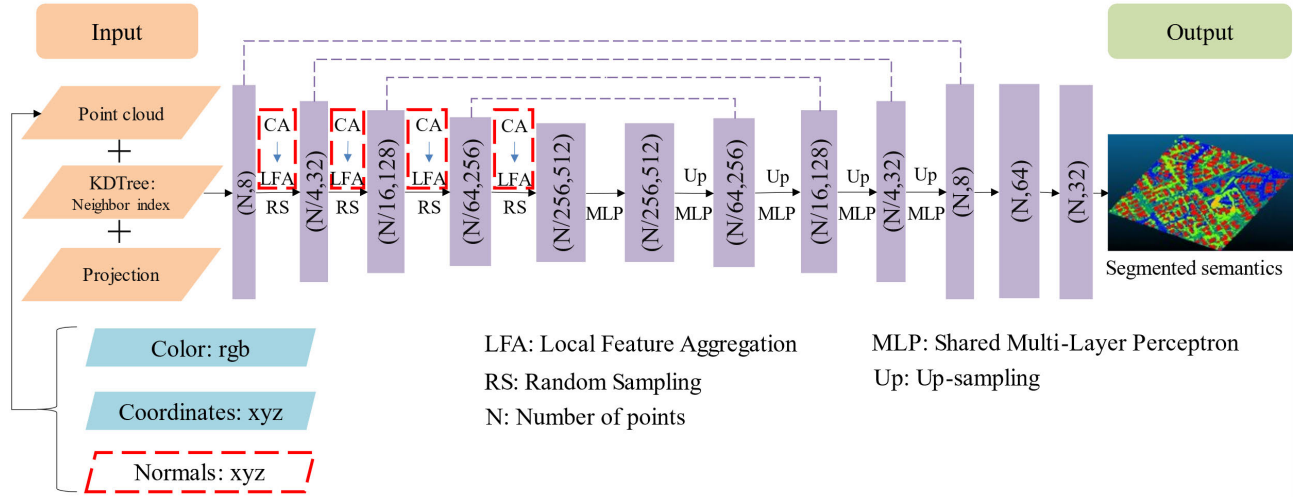


Fig. 6. Architecture of the proposed with the amalgamation of the surface normal information and CA.

is necessary because using the original massive point cloud data as inputs in the computer is impractical.

In the first step, in order to have the same data format for all point clouds, the clouds with the “.las” format were transferred to the “.ply” format. This is because the proposed network was designed for processing point clouds in the “.ply” version. The values of each color band in the “.las” format were divided by 255 before transferring to the “.ply” clouds. To calculate the normal on a point, the local surface must be estimated to represent itself and its neighbors. Hence, the coordinate values of each point were necessary. Since color information was also needed in this study, both the coordinate values and RGB color information were stored for the next step.

In the second step, voxel grid downsampling was applied to all points. The volume of the originally collected point cloud data is exceptionally large in most situations. Thus, the volume is always reduced by downsampling without affecting the characteristics of a point cloud. This operation can help to save processing time and avoid out-of-memory during training networks. The grid size was 0.5 m in this study.

Third, surface normal information was calculated. To add surface normal information, this study applied Open3D, an open-source Python library, to generate normals. This is because Open3D has already encapsulated the function. The built-in function “estimate_normals” finds K -nearest neighbor (KNN) points within a radius and calculates the principal axis of the adjacent points using covariance analysis [16]. The function chooses a point and its KNNs (i.e., $1 + K$ points in total) to estimate a plane using the least-squares method and then makes a vertical line of the plane through that point, which is its normal vector. Specifically, the problem of estimating the surface normal of a point is simplified as an analysis of eigenvectors and eigenvalues of the covariance matrix calculated from the nearest neighbor of the point. In this study, the search radius was 0.1 m, and the maximum nearest neighbor was 30 using KDTree search for neighborhood search, which are default numbers in Open3D. Choosing

default numbers because these parameters are not the focus of this study.

The normal orientation problem of surface normal calculation should be noted. Two normal candidates with opposite directions are produced from the covariance analysis algorithm. Without knowing the global structure of geometry, both can be correct, which could cause problems. Therefore, Open3D tried to orient the normal to align with the original normal if it existed. Otherwise, Open3D made a random guess. Then, normal values were added to point cloud data.

In the fourth step, three types of outputs were stored, as shown in the black rectangle of Fig. 5. In detail, after producing the point cloud data from the third step, KDTree files and projection files are also generated and stored for each point cloud. Each KDTree file was named “XX_KDTree.pkl.” KDTree files have the information of the nearest N points around each downsampled point. Projection files have stored the number of the downsampled points with the shortest distance from each original point. The original points are the points before the second step—downsampling. These numbers were stored in files named “XX_proj.pkl.” Projection files are necessary because point clouds need to be restored to the original size after semantic segmentation for the downsampled ones in the proposed network. The restoration needs these numbers for nearest neighbor interpolation.

Following the above steps, the final outputs include colored point cloud data in the “.ply” format, KDTree files, and projection files. The information in point clouds contains the RGB bands, three values of coordinate systems, and three values of the corresponding normals.

2) *Architecture of the Proposed CNLNet*: The architecture of the proposed CNLNet is shown in Fig. 6. It is a conventional encoder and decoder architecture with skip connections. The inputs contain three types of files, including point clouds, KDTree, and projected numbers of point clouds. The architecture has four encoding and decoding layers. As shown in those four encoding layers, only a quarter of the point features are retained with the increased feature dimension after each layer for downsampling. Random point sampling is applied for high

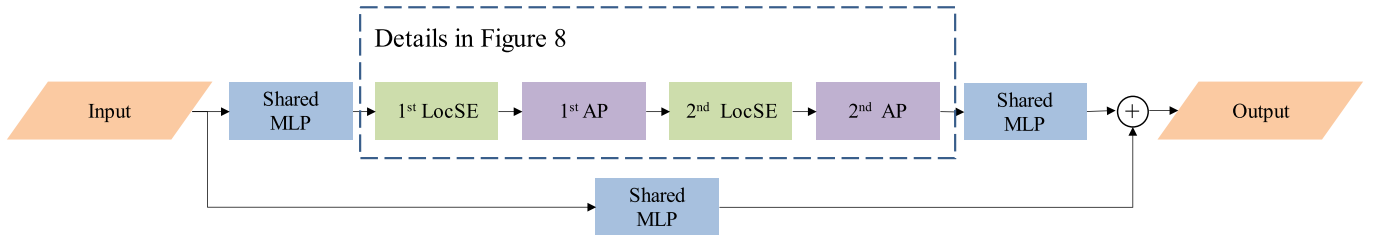


Fig. 7. Architecture of the LFA module.

efficiency of memory and computation, as its computational complexity is only $O(1)$. After that, the point features are upsampled gradually through a nearest-neighbor interpolation in the four decoding layers [8]. The final output is obtained through shared fully connected layers. The final output is the predicted class of each point. To be noticed, this study adds contents in red rectangles, including normal information and CA in the LFA module. The details of the backbone including LFA and CA are introduced as follows.

Datasets in New Zealand were applied for training, validation, and testing with the number of point clouds 38, 6, and 5, respectively. Kumamoto data were only applied for the tests because this area is too small to separate it into three parts for training, validation, and testing.

3) *RandLA-Net Backbone*: The backbone of the proposed network is RandLA-Net, that is, “random sampling and an effective local feature aggregator” [8]. Although several networks showed promising results for small point cloud semantic segmentation, most cannot directly scale up to large scenarios. This is because of their high memory and computational costs. The benefit of RandLA-Net is that it was designed for large-scale point cloud semantic segmentation with less memory and computation, which is suitable for predisaster tasks. Therefore, this study chose it as the backbone.

RandLA-Net is a lightweight pointwise MLP network. Point-based deep learning methods for semantic segmentation can be roughly divided into pointwise MLP, point convolution, recurrent neural network (RNN)-based, and graph-based methods [4]. MLP is a supplement of a feed-forward neural network, including the input layer, the output layer, and the hidden layers.

RandLA-Net designed an LFA module with shared MLP preserving local geometric structures and other useful local features. The LFA module has two key units: local spatial encoding (LocSE) and attentive pooling (AP), as shown in Fig. 7. The LocSE unit is applied for local geometric structures, and the AP unit is applied for saving those useful local features. Their details are shown in Fig. 8. In the LocSE unit, the KNN algorithm is utilized to find neighbor points based on the pointwise Euclidean distances. K represents the number of neighbor points. K is 16 in this study. After finding the neighbor points, MLP is applied to encode the relative point positions between every center point and its neighboring points. Hence, the local geometric structures are encoded for every center point to augment neighboring point features by LocSE. After that, the AP unit is applied to aggregate the neighboring point features. This unit applied shared MLP

TABLE III
DESIGNED NETWORKS OF ABLATION STUDY

Network 1 (CNLNet)	Network 2	Network 3	Network 4
Backbone + Normals + CA	Backbone + Normals	Backbone + CA	Backbone

followed by SoftMax function to learn a unique attention score for every feature. Then, the features are weighted and summed. RandLA-Net stacks multiple LocSE and AP units with a skip connection as a dilated residual block. In order to avoid overfitting during the training stage and keep computation efficiency, only two sets of LocSE and AP are stacked [8].

4) *CA in CNLNet*: CA was added to the proposed network to examine its effect on the final output. Adapting the previous work of a multibranch network [17], CA applied this study that deduces the channel number to 1 and then recovers back to the original number. With this operation, the relevance between each channel and key information in channels can be more obvious and easier for the computer to learn. The benefit of this attention mechanism is that it can usually help to achieve significant improvement in accuracy in terms of 2-D semantic segmentation [18]. Moreover, since the accuracy needs to increase and LFA does not contain CA, this study applied CA in 3-D semantic segmentation to test its effects.

The novel proposed network consists of LFA and CA with surface normals. CA is added into the LocSE unit of the LFA module as shown in red rectangles of Fig. 8. The details are explained in Fig. 9. K and d represent the number of neighbor points and the feature dimension, respectively. First, the matrix is transposed from (K, d) to (d, K) . Then, the transposed is multiplied by the original matrix. The dimension of the multiplication result is squeezed to 1 by max pooling and restored to d by copying and subtracting an activation function. The multiplication of the original matrix and the restored one is operated after that. Last, the attentive result is the sum.

D. Ablation Studies on Four In-House Labeled Datasets

The design of an ablation study with five evaluation metrics for detecting the impact of information and CA on segmentation is introduced in this section.

Four networks were tested in the ablation study, as shown in Table III. They were designed to demonstrate the benefit of adding surface normal information or CA in the backbone. The

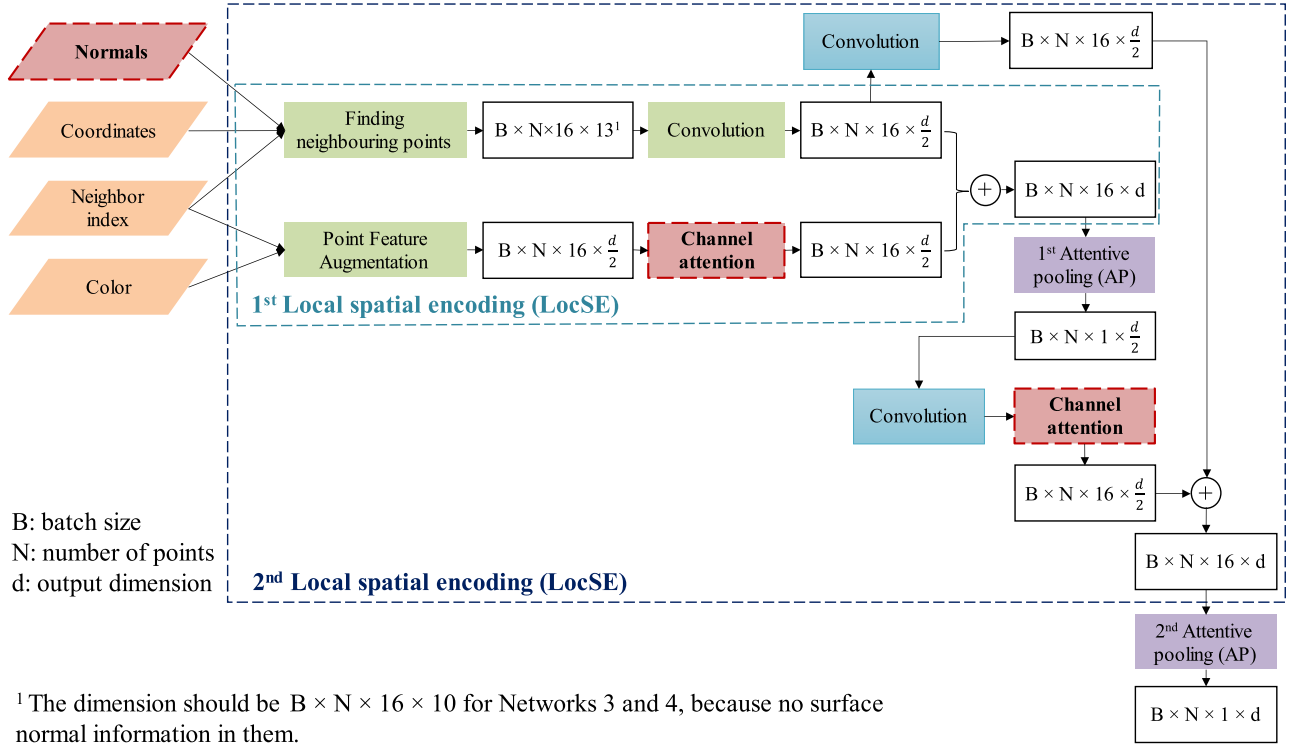


Fig. 8. Structure details of the CA added to the modified LocSE and AP.

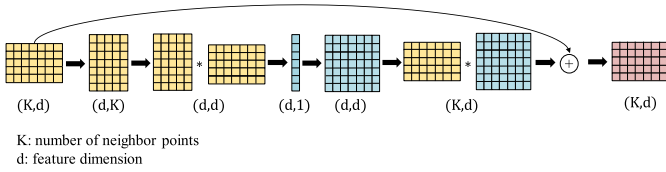


Fig. 9. Details of CA.

backbone added both normal information and the CA block is Network 1. The backbone with only normal information or CA was designed as Networks 2 and 3. The original RandLA-Net backbone network was tested at last, which is Network 4. Each point in data is represented by its coordinates, normal, and color information in Networks 1 and 2. It is represented by the coordinates and colors in Networks 3 and 4.

Five evaluation metrics were chosen. These five metrics were calculated for testing the segmentation performance of each network, including true positive (TP), false negative (FN), false positive (FP), intersection over union (IoU), and semantic segmentation accuracy (SSA). The summation of TP, true negative (TN), FN, and FP is the whole number of points in one point cloud. TP represents a point whose tested label is the same as its true label. TN in each class indicates the points that both its tested and true labels do not belong to that class. In the results of a class, FN refers to the point that its tested label does not belong to this class, but its true label does. On the other hand, FP in results of a particular class means the tested result is in this class but its true label does not.

The IoU shown in (1) is a mathematical way to choose the best network by checking the degree of similarity of the

output produced by the proposed networks with the ground truth. The higher the IoU value, the better performance of the chosen network. After observing initial results, this study predominately discussed IoU rather than the other four metrics (i.e., TP, FN, TP, and SSA). One reason is that IoU contains TP, FN, and FP. Discussing IoU would be more helpful for data analysis than only analyzing a single TP, FN, or FP. Another reason is that several relevant articles utilized IoU as the metric [10], [18].

SSA shown in (2) was not considered as the main metric mainly because a high SSA score does not necessarily indicate favorable result in this study. The number of each point cloud is huge. If TP, FN, and FP were all very low in one class, TN would be very high in this class. In such cases, even if the SSA approaches 1, it would not necessarily mean that the results are optimal. Hence, SSA serves only as a supplementary metric for reference.

Mean IoU (mIoU) was also calculated for multiclass-based semantic segmentation. The mIoU represents the average between the IoU of all segmented classes over all the images of each tested point cloud. All networks were trained ten times, and the network that had the highest mIoU value for validation data was chosen to be applied to the test data. It shows the correctly segmented area over all the areas that the network segmented

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (1)$$

$$\text{Semantic Segmentation Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (2)$$

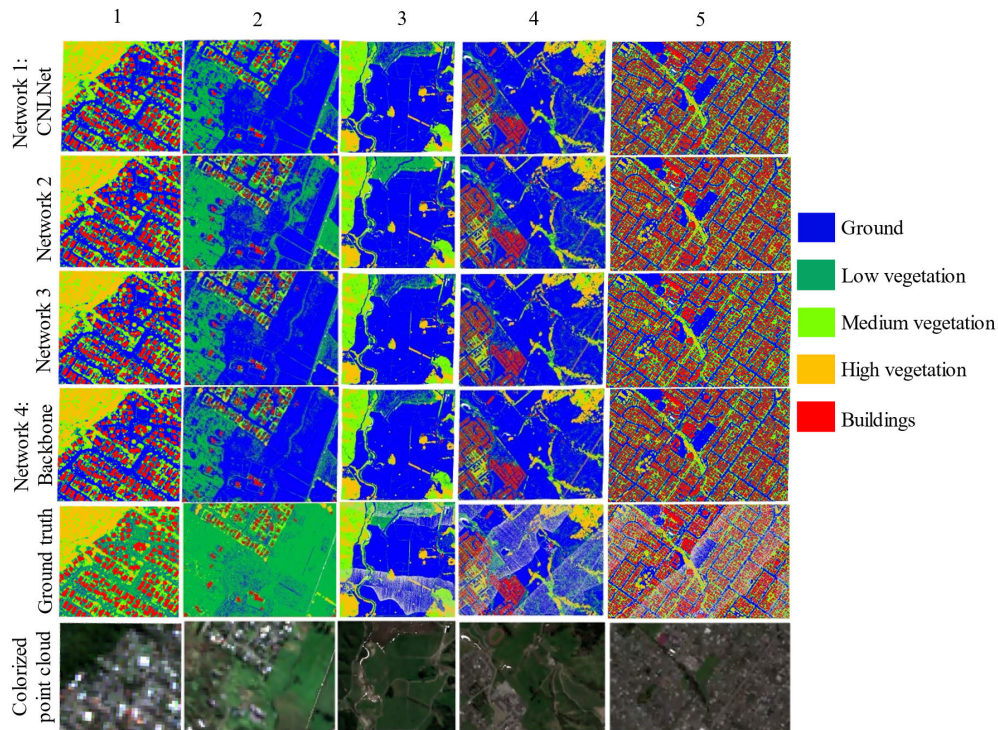


Fig. 10. Visualization results of ablation study for Kapiti Coast, Tasman, and Nelson datasets.

Besides ablation studies of the proposed network, this study tested the influence of 2-D image resolution on segmentation results. The Kumamoto pre-earthquake dataset was applied for this test. This was because this dataset has two optical images with different resolutions.

E. Comparison of the Public Semantic3D Dataset

To detect the performance of the proposed network, these four networks were trained and tested on the public dataset Semantic3D [12]. Only coordinates and RGB information with eight labeled classes from the dataset were used to train and test different methods. Some well-known and state-of-the-art networks were also tested for comparison, including PointNet [5], PointNet++ [6], and ShellNet [19]. The tested point clouds were chosen according to the selected test datasets provided by Hu et al. [8], which include four point clouds.

III. RESULTS

This section presents the semantic segmentation results of the ablation studies, which contain results of the four networks using five metrics with the test data.

As mentioned in Section II, this study set four datasets as test data. Five classes were tested, including buildings, ground, low vegetation, medium vegetation, and high vegetation. Visualization results and quantitative results are stated in this section. Five point clouds were tested. The first two tested point clouds are from Kapiti Coast. The third is from the Tasman dataset, and the last two are in the Nelson dataset.

A. Hardware and Environment

In this study, one Nvidia RTX 2080Ti GPU card, CUDA 11.3, Python 3.6, and TensorFlow 1.15 were applied.

Since TensorFlow versions 1 and 2 have huge differences, it would be more convenient to use TensorFlow version 1 to fit its version in the original RandLA-Net backbone. “Batch size during training” is 2, and “Number of steps per epoch” is 1000. It took 8 h to run in the GPU version with 100 epochs.

B. Results on New Zealand Datasets

Five patches of point clouds were chosen as the test data from the three New Zealand datasets. Their visualization results are shown in Fig. 10. Red, blue, dark green, bright green, and orange represent buildings, ground, low, medium, and high vegetation, respectively. Based on the visual observation, compared with the ground truths of the point clouds, most buildings were recognized correctly, but most medium and high-vegetation points were mistakenly recognized as ground and low-vegetation points.

Results for each class with the five metrics are shown in A.TABLE I of supplemental materials. The first class is the bare-ground class. The proposed Network 1 performed best for the ground segmentation according to IoU results. Networks 2 and 3 are nearly the same as the backbone.

The following three classes are the three vegetation classes, including low, medium, and high. Medium-vegetation segmentation results performed best among these three classes in all networks according to their IoUs. In all low-vegetation results, Network 4 performed best among the four networks with the highest IoUs. The highest IoUs of medium- and high-vegetation results were also the results of Network 1. IoUs of low vegetation are always the lowest among the three vegetation classes.

The fifth class is the building class. Network 1 always performs best among all networks, but Networks 2 and 3 do

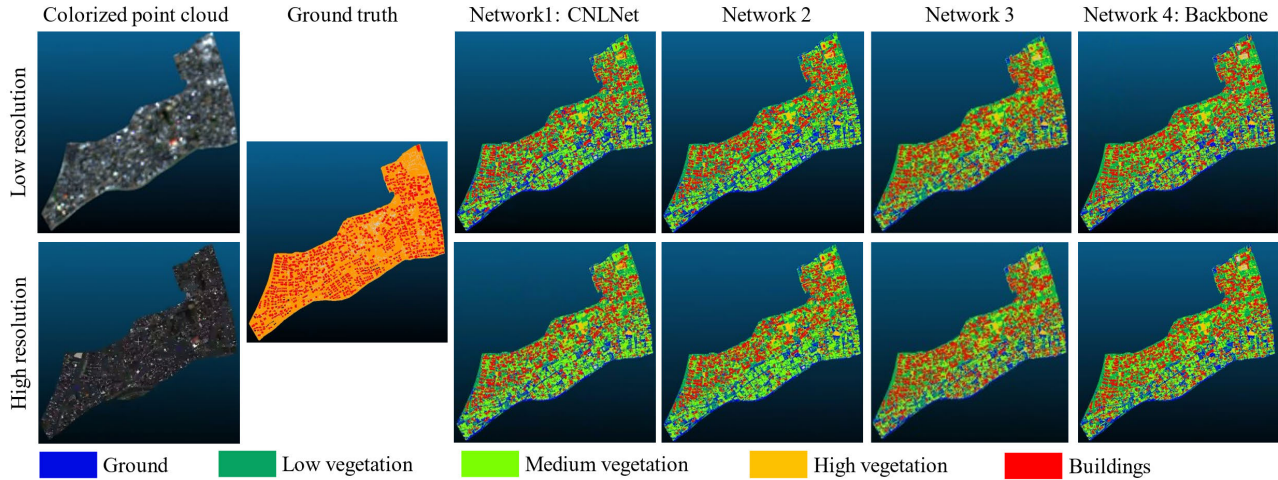


Fig. 11. Visualization results of ablation study for 2016 Kumamoto pre-earthquake data.

not show any significant advantage over the backbone of Network 4. IoUs of the tested point clouds from the Kapiti Coast dataset in CNLNet are very high, which are 0.95 and 0.92, respectively. They are the top two highest among all tested point clouds with the top two highest TP values. A probable reason for this is that the number of point clouds belonging to the building class in the Kapiti Coast dataset accounts for a significant proportion of the total number of building points.

After comparing the performance of each network for every class, the result differences between the five classes should also be mentioned. Compared with the other classes, the building class always has the highest IoU among the five classes in the results of each network, which most are higher than 0.90 in some test results. It is convinced that the RandLA-Net backbone is suitable for building detection. Segmentation of ground and low vegetation performed worst in results according to IoUs. The likely reason is data imbalance. The numbers of Lidar points in these classes are lower than those of others.

C. Results of 2016 Kumamoto Pre-Earthquake Data

The 2016 Kumamoto pre-earthquake point cloud dataset with both high and low resolutions of optical satellite images was tested. The number of total points in this point cloud is 1438042. As mentioned in Table II, the resolution of the high-resolution image is 0.5 m/pixel, and that of the low-resolution image is 10 m/pixel. Fig. 11 shows their visualization results. Five classes were segmented. Red, blue, light green, bright green, and orange represent buildings, ground, low vegetation, medium vegetation, and high vegetation, respectively. It can be easily found that most high-vegetation points were mistakenly segmented as other classes, such as low and medium vegetation.

Quantitative results of the 2016 Kumamoto pre-earthquake data are shown in A. TABLE II of supplemental materials. The results of all the five classes are listed in it.

The first test class is the ground class. All networks performed not so well for this class no matter with high or low resolution. The number of FP points is too high, no matter which network. The second class is the low-vegetation class. Similar to segmentation results for the ground class, IoUs were nearly zero for all networks. Thousands of points were detected as low vegetation wrongly. In other words, the FP values of these two classes are high. Moreover, nearly no TP points have been detected, as shown in the results of the ground and the low-vegetation classes. The first probable reason is that the information difference of segmentation labels between the training data and these test data is large, which are from different datasets. The second possible reason is that the points from those two classes are too few to be detected in these test data.

The next two classes are medium and high vegetation. Although their IoUs were also nearly zero, the number of their TP points was much higher than those of ground and low vegetation. Besides IoUs, SSA results for medium vegetation were higher than those for high vegetation, while the numbers of FP points in medium-vegetation results were higher than those in high vegetation for all these three networks.

The last class is the building class. The highest IoUs were the building class results among all five detected classes in all networks. This might be because the number of points labeled as buildings are high in the training dataset. According to IoUs, Network 3 performed the best, which shows that its generalizability for building segmentation is the best of these networks.

Among all segmented classes, the generalizabilities of all tested networks in the ablation study are not ideal except for the building class. There are some possible reasons. Although the two datasets both have these five labeled classes and colors, the labeled information of the classes in the 2016 Kumamoto pre-earthquake dataset is much different from those of the New Zealand datasets. As mentioned in Section II-C, only New Zealand datasets are applied for training due to the small area of 2016 Kumamoto pre-earthquake data.

D. Results on Semantic3D

The results are listed in Table IV. Network 3 has the highest mIoU of the tested eight classes. Networks 1–4 all achieve acceptable results compared with the other networks.

However, Network 1 performed slightly worse than Networks 2 and 3 after adding both normal information and CA in the backbone though it performed best in some class results. The probable reason is that Network 1 was overfitted. Overfitting might exist if the network is too complicated. In total, the results demonstrate that both surface normal information and CA have helped with large-scale outdoor point cloud semantic segmentation based on the RandLA-Net backbone. Each of them can improve mIoU by 1%–2% than the backbone. This might be due to its overly complicated structure. The network’s performance with adding both CA and surface information (Network 1) is not as good as the network with only adding one.

IV. DISCUSSION

This study designed ablation studies to demonstrate the benefits of adding surface normal information and CA mechanism in LOLSS for predisaster information classification and storage.

IoUs of the building class were always the highest in the results in all own labeled datasets among all tested networks. This reflects that these networks are all suitable for segmenting buildings. Besides that, in the test of the Kumamoto dataset, the building segmentation IoUs were significantly higher than the results of other classes. The training and validation steps did not contain Kumamoto data. Hence, the generalizability of the trained network for building segmentation is the highest. Moreover, the results for the Kumamoto point clouds with different resolutions of optical satellite images were very similar. It can be concluded that the optical satellite image resolutions may have little influence on the performance of the proposed model.

In addition to the analysis of the IoU of each class, the overall IoU of all classes should be discussed. As mentioned in Section II-D, the mIoU of each network was calculated to analyze its performance considering the results of all classes. The mIoU results for all classes in the five tested point clouds are shown in Table V. It should be noted that mIoUs of Kumamoto data are not discussed because IoU values of the other four classes are nearly zero except IoU of the building class due to the poor generalizabilities of these four classes. Table V shows that mIoU values of Network 1 are always the highest in these networks for all tested point clouds. Moreover, the mIoUs of Network 2 are higher than those of Network 3, so it shows that adding normal information might be more helpful than CA to semantic segmentation.

Other metrics also demonstrated that the designed network is suitable for segmenting buildings from the background. The TPs of the building class in A. TABLE I and A. TABLE II of supplementary documents are very high. The SSA of the building class in A. TABLE I is nearly 1, and its SSA in A. TABLE II is the highest among SSAs of all classes.

TABLE IV
RESULTS OF DIFFERENT METHODS ON SEMANTIC3D

Network	mIoU
Network 1: Backbone + Normals + CA	0.67
Network 2: Backbone + Normals	0.68
Network 3: Backbone + CA	0.70
Network 4: Backbone	0.67
ShellNet [19]	0.63
PointNet++ [6]	0.42
PointNet [5]	0.41

TABLE V
MEAN IOU OF THE NEW ZEALAND TEST DATASETS

Tested point cloud	mIoU			
	Network 1: Backbone + Normals + CA	Network 2: Backbone + Normals	Network 3: Backbone + CA	Network 4: Backbone
1	0.68	0.65	0.62	0.57
2	0.62	0.55	0.54	0.53
3	0.66	0.64	0.64	0.64
4	0.64	0.61	0.61	0.61
5	0.60	0.55	0.55	0.55

The results for Semnatic3D also demonstrated that the designed network is suitable for predisaster land cover object segmentation from the background.

Based on the abovementioned discussion, it can be concluded that surface normal information and CA can improve segmentation accuracy. The proposed CNLNet can improve mIoU by 1%–11% compared to the backbone in different scenarios. Besides that, in contrast to the RandLA-Net backbone and other well-known networks, each of these two types of feature information (Networks 2 and 3) can help to improve the accuracy of semantic segmentation. The network with only adding surface information (Network 2) is more effective between these two types of networks.

V. CONCLUSION

In this study, a network named CNLNet was proposed to enhance the precision of deep learning-based LOLSS for predisaster land cover information segmentation and preservation. Surface normals and CA were added to this network. A labeled large-scale land cover Lidar dataset was first created in this study with considering potential natural disaster occurrences in the geographical areas highlighted in the selected datasets, including Kapiti Coast, Tasman, and Nelson in New Zealand and Kumamoto in Japan. Optical satellite images were integrated as inputs. Compared with the state-of-the-art RandLA-Net backbone and other renowned networks, the findings demonstrate the benefits of surface normal information and CA applied to LOLSS. Normal information can provide more feature information, and CA can emphasize key information in channels, so they can improve the accuracy of segmentation results. Furthermore, the proposed network exhibits the strongest generalizability for the building class. Interestingly, the network that incorporated either surface normals or CA alone slightly outperformed the one incorporating

both during the test on the open-source Semantic3D dataset. The likely reason is that overfitting might occur if a network is too complex. With the potential to save labor and mitigate in situ risks, the practical implication of this method lies in its applicability for urban land cover segmentation from 3-D Lidar point clouds, particularly for building segmentation. The outcomes can be utilized for predisaster urban visualization data information storage and update, thus expediting postdisaster emergency response efforts. Further research is suggested to find an approach to improve the segmentation accuracy of separating classes other than buildings, such as low, medium, and high vegetation. The normal to the hyperplane is an important vector that separates different classes of points. An incorrect normal vector would result in a poorly performing classifier. Therefore, the setting of the direction of surface normal could also be discussed in future studies.

ACKNOWLEDGMENT

Chang Liu thanks to OpenTopography for light detection and ranging (Lidar) datasets and Geospatial Information Authority of Japan (GSI) for the 2016 Kumamoto land cover shapefiles.

REFERENCES

- [1] C. Liu, S. M. E. Sepasgozar, Q. Zhang, and L. Ge, "A novel attention-based deep learning method for post-disaster building damage classification," *Expert Syst. Appl.*, vol. 202, Sep. 2022, Art. no. 117268.
- [2] M. He, Q. Zhu, Z. Du, H. Hu, Y. Ding, and M. Chen, "A 3D shape descriptor based on contour clusters for damaged roof detection using airborne LiDAR point clouds," *Remote Sens.*, vol. 8, no. 3, p. 189, Feb. 2016.
- [3] Q. Zhang, L. Ge, S. Hensley, G. I. Metternicht, C. Liu, and R. Zhang, "PolGAN: A deep-learning-based unsupervised forest height estimation based on the synergy of PolInSAR and LiDAR data," *ISPRS J. Photogramm. Remote Sens.*, vol. 186, pp. 123–139, Apr. 2022.
- [4] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun, "Deep learning for 3D point clouds: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 12, pp. 4338–4364, Dec. 2021.
- [5] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 77–85.
- [6] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–10.
- [7] M. Jiang, Y. Wu, T. Zhao, Z. Zhao, and C. Lu, "PointSIFT: A SIFT-like network module for 3D point cloud semantic segmentation," 2018, *arXiv:1807.00652*.
- [8] Q. Hu et al., "RandLA-Net: Efficient semantic segmentation of large-scale point clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11105–11114.
- [9] Kapiti Coast District Council. *Kapiti's Natural Hazards*. Accessed: Mar. 10, 2022. [Online]. Available: <https://www.kapiticoast.govt.nz/our-district/cdem/kapitis-natural-hazards/>
- [10] Nelson Government. *Weather Event August 2022—Recovery Progress Report March 2023*. Accessed: Apr. 11, 2023. [Online]. Available: <https://www.nelson.govt.nz/services/weather-event-august-2022/>
- [11] M. Yamada, J. Ohmura, and H. Goto, "Wooden building damage analysis in Mashiki town for the 2016 Kumamoto earthquakes on April 14 and 16," *Earthq. Spectra*, vol. 33, no. 4, pp. 1555–1572, Nov. 2017.
- [12] T. Hackel, N. Savinov, L. Ladicky, J. D. Wegner, K. Schindler, and M. Pollefeys, "Semantic3D.Net: A new large-scale point cloud classification benchmark," 2017, *arXiv:1704.03847*.
- [13] OpenTopography. *LiDAR of Kapiti Coast, Wellington, New Zealand 2021*. Accessed: May 18, 2022. [Online]. Available: <https://portal.opentopography.org/lidarDataset?jobId=pc1644298041020>
- [14] OpenTopography. *Top of the South Flood, Nelson and Tasman, New Zealand 2022*. Accessed: Apr. 10, 2023. [Online]. Available: <https://portal.opentopography.org/datasetMetadata?otCollectionID=OT.032023.2193.1>
- [15] Safe Software. *Introduction of FME Desktop*. Accessed: Mar. 20, 2022. [Online]. Available: <https://www.safe.com/fme/fme-desktop/>
- [16] Open3D. *Open3D Geometry Estimating Normal Function*. Accessed: Mar. 20, 2022. [Online]. Available: http://www.open3d.org/docs/0.7.0/python_api/open3d.geometry.estimate_normals.html
- [17] X. Geng, S. Ji, M. Lu, and L. Zhao, "Multi-scale attentive aggregation for LiDAR point cloud segmentation," *Remote Sens.*, vol. 13, no. 4, p. 691, Feb. 2021.
- [18] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141. [Online]. Available: http://openaccess.thecvf.com/content_cvpr_2018/papers/Hu_Squeeze-and-Excitation_Networks_CVPR_2018_paper.pdf
- [19] Z. Zhang, B.-S. Hua, and S.-K. Yeung, "ShellNet: Efficient point cloud convolutional neural networks using concentric shells statistics," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1607–1616.



Chang Liu (Graduate Student Member, IEEE) received the B.E. degree in civil engineering from the University of Science and Technology Beijing, Beijing, China, in 2017, and the M.Phil. degree from the University of New South Wales (UNSW), Sydney, NSW, Australia, in 2020, where she is currently pursuing the Ph.D. degree with the School of Civil and Environmental Engineering.

Her research interests include remote sensing image processing, deep-learning-based algorithm development and its applications in natural disaster management, and Lidar data processing.



Linlin Ge (Member, IEEE) received the B.E. degree (First Class Hons.) in optical engineering from the Wuhan Technical University of Surveying and Mapping, Wuhan, China, in 1985, the M.Sc. degree in crustal deformation from the Institute of Seismology, China Earthquake Administration, Wuhan, and the Ph.D. degree in GPS and remote sensing from the University of New South Wales (UNSW), Sydney, NSW, Australia, in 2001.

From 1997 to 1998, he was a Post-Doctoral Research Fellow with the Meteorological Research Institute, Japan Meteorological Agency, sponsored by the Science and Technology Agency of Japan, Tsukuba, Japan. He is currently a Professor of remote sensing and Earth observation with the School of Civil and Environmental Engineering, UNSW. He has been studying Earth observation techniques since 1985.

Prof. Ge is a Foundation Member of the Surveying and Spatial Sciences Institute. He is a Life Member of the American Geophysical Union (AGU). He is also a Principal Investigator for satellite missions such as ENVISAT, ALOS, COSMO-SkyMed, Radarsat-2, and ALOS-2. Among many other prestigious awards, he and his team's work to support the Sichuan Earthquake rescue effort won them the highly prestigious JK Barrie Award for Overall Excellence at the 2008 Asia-Pacific Spatial Excellence Awards. He was named "NSW Scientist of the Year 2009" in the category Physics, Earth Sciences, Chemistry and Astronomy for his work in near real-time satellite remote sensing.



Wei Xiang (Senior Member, IEEE) was the Foundation Chair and the Head of discipline of Internet of Things (IoT) engineering with James Cook University, Cairns, QLD, Australia. He is currently the Cisco Research Chair of Artificial Intelligence (AI) and IoT and the Director of the Cisco-La Trobe Centre for AI and IoT with La Trobe University, Melbourne, VIC, Australia. Due to his instrumental leadership in establishing Australia's first accredited IoT engineering degree program, he was inducted into Pearcey Foundation's Hall of Fame in October

2018. He has authored over 300 peer-reviewed articles including three books and 220 journal articles. His research interests include IoT, wireless communications, machine learning for IoT data analytics, and computer vision.

Prof. Xiang is a TEDx Speaker and an Elected Fellow of the IET in U.K. and Engineers Australia. He received the TNQ Innovation Award in 2016, Pearcey Entrepreneurship Award in 2017, and Engineers Australia Cairns Engineer of the Year in 2017. He was a co-recipient of four Best Paper Awards at WiSATS'2019, WCSP'2015, IEEE WCNC'2011, and ICWMC'2009. He has been awarded several prestigious fellowship titles. He was named a Queensland International Fellow by the Queensland Government of Australia, from 2010 to 2011, an Endeavor Research Fellow by the Commonwealth Government of Australia, from 2012 to 2013, a Smart Futures Fellow by the Queensland Government of Australia, from 2012 to 2015, and a JSPS Invitational Fellow jointly by the Australian Academy of Science and Japanese Society for Promotion of Science, from 2014 to 2015. He was the Vice-Chair of the IEEE Northern Australia Section from 2016 to 2020. He is currently an Associate Editor of IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE INTERNET OF THINGS JOURNAL, IEEE ACCESS, and *Nature Journal of Scientific Reports*. He has served in a large number of international conferences in the capacity of general co-chair, TPC co-chair, symposium chair, etc.



Zheyuan Du (Member, IEEE) received the B.E. degree in remote sensing from the China University of Geosciences, Wuhan, China, in 2012, the M.Sc. (GIS) degree from The University of Edinburgh, Edinburgh, U.K., in 2013, and the Ph.D. degree in radar remote sensing from the University of New South Wales (UNSW), Sydney, NSW, Australia, in 2018.

From 2018 to 2021, he was a Lecturer with the University of Sydney, Sydney, NSW, and concurrently held the position of a Research Fellow at UNSW. He joined Geoscience Australia, Canberra, ACT, Australia, in 2021, and currently holds a Senior Scientist position. His research interests include interferometric synthetic aperture radar (InSAR) algorithm and application, hydrology, and image classification.



Qi Zhang received the B.E. degree in information engineering and the M.E. degree in electronic and communication engineering from the Beijing Institute of Technology, Beijing, China, in 2014 and 2016, respectively, and the Ph.D. degree in remote sensing from the University of New South Wales (UNSW), Sydney, NSW, Australia, in 2022.

She is currently working as a Post-Doctoral Researcher with the Technical University of Munich, Munich, Germany. Her research interests include deep-learning-based algorithm design and its applications on the Earth observation of synthetic aperture radar and multispectral remote sensing.