MULTISPECTRAL AIRBORNE LASER SCANNING POINT-CLOUDS FOR LAND COVER CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORKS

Lingfei Ma [1], Zhuo Chen [1], Ying Li [1], Dedong Zhang[1], Jonathan Li[1, 2, 3] *, Michael A. Chapman [4]

[1] Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada – (l53ma, zhuo.chen, y242li, junli)@uwaterloo.ca
[2] Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada - junli@uwaterloo.ca
[3] Fujian Key Laboratory of Sensing and Computing for Smart Cities, School of Information Science and Engineering, Xiamen University, Xiamen, Fujian 361005, China – junli@xmu.edu.cn
[4] Department of Civil Engineering, Ryerson University, Toronto, ON M5B 2K3, Canada - mchapman@ryerson.ca

KEY WORDS: Land cover classification, multispectral airborne laser scanning, point cloud, CNN, deep learning

ABSTRACT:

This paper presents an automated workflow for pixel-wise land cover (LC) classification from multispectral airborne laser scanning (ALS) data using deep learning methods. It mainly contains three procedures: data pre-processing, land cover classification, and accuracy assessment. First, a total of nine raster images with different information were generated from the pre-processed point clouds. These images were assembled into six input data combinations. Meanwhile, the labelled dataset was created using the orthophotos as the ground truth. Also, three deep learning networks were established. Then, each input data combination was used to train and validate each network, which developed eighteen LC classification models with different parameters to predict LC types for pixels. Finally, accuracy assessments and comparisons were done for the eighteen classification results to determine an optimal scheme. The proposed method was tested on six input datasets with three deep learning classification networks (i.e., 1D CNN, 2D CNN, and 3D CNN). The highest overall classification accuracy of 97.2% has been achieved using the proposed 3D CNN. The overall accuracy (OA) of the 2D and 3D CNNs was, on average, 8.4% higher than that of the 1D CNN. Although the OA of the 2D CNN was at most 0.3% lower than that of the 3D CNN, the runtime of the 3D CNN was five times longer than the 2D CNN. Thus, the 2D CNN was the best choice for the multispectral ALS LC classification when considering efficiency. The results demonstrated the proposed methods can successfully classify land covers from multispectral ALS data.

1. INTRODUCTION

Defined as the physical composition and features of objects at the surface of the Earth (Costa et al., 2018), land cover (LC) as a crucial parameter is used to supervise the changing world. According to the report released by the United Nations (2015), global urbanization will increase to 66% by 2050. LC change significantly affects the evaporation, transpiration, and heat flux on the ground surface, which further impacts the radiation balance on the Earth. Although the rapid global urbanization increases social and economic opportunities, it affects stability and sustainability of the environment, accelerates the variation of land cover (LC), and consequentially brings challenges to the supervision of LC.

According to Mariano et al. (2018), accurate LC maps are required for the monitoring of the ecosystem and the study of ecosystem processes such as the functions of wetland, the suitability of habitat, and the potential of soil erosion and sedimentation. Inadequate analysis and supervision of LC can lead to many problems for the ecosystem such as the loss, destruction, and degradation of the habitat for various species (Guida-Johnson & Zuleta, 2013). With respect to the biogeophysical aspect, the change of LC directly impacts the physical composition and features of the Earth, which thereby affects the energy availability at the Earth's surface. Thus, it demonstrated that precise and efficient mapping of LC is essential to ensure an accurate representation of LC change, to protect the Earth and to ensure sustainable human-environment development (Zhong et al., 2017).

Traditionally, multispectral images are used to capture information on the surface of the Earth. However, the main problems are that the separability among different LC features can be degraded by the between-class spectral confusion and within-class spectral variation (Yan et al., 2015). Additionally, aerial photos and satellite images are often affected by cloud coverage and weather conditions. Accordingly, airborne mapping light detection and ranging (LiDAR), also known as airborne laser scanning (ALS), has become one of the primary remote sensing technologies for analysing the surface of the Earth due to its good capability of three-dimensional (3D) information acquisition (Glennie et al., 2013). LiDAR is a gauging technique that surveys distance to an object, which can record a set of points that describe the target object. Compared with two-dimension (2D) images, the LiDAR data have the advantages of acquiring more accurate topographic information from the Earth's surface without problems resulting from cloud coverage, weather conditions, and relief displacement (Ma et al., 2018; Ye et al., 2019). Previous studies have well demonstrated the capability of ALS data in LC classification (e.g. Ene et al., 2018; Sun et al., 2018). Using its 3D spatial information, the ALS data can separate objects that have similar spectral signatures such as parking lots and buildings (Glennie et al., 2013). Nevertheless, most of the LiDAR sensors record only one channel of pulses. Thus, the fact that single-wavelength ALS data lack spectral information limits its accuracy for classifying similarly shaped objects in complicated environments. To overcome these limitations, the 3D data obtained by ALS are often integrated with spectral information provided by multispectral images. However, datasets acquired by different

---

* Corresponding author

systems often have different data formats, projections, spatial resolutions, and collection times, which can introduce errors to the data fusion process.

To solve the problems of data fusions, multispectral LiDAR techniques, which can acquire LiDAR data with multiple channels simultaneously, have been recently developed. The Teledyne Optech Titan, which contains three active imaging channels at different wavelengths, is the first commercial multispectral airborne active imaging LIDAR sensor in the world (Karila et al., 2018). Even though only a few related studies have been conducted (e.g. Chen et al., 2018; Zou et al., 2016), the potential of using multispectral ALS technique to map the Earth's surface has been identified. The multispectral ALS data has been proven to be superior to both traditional multispectral optical imagery and typical single-wavelength ALS data for LC classification (Bakuła et al., 2016; Teo and Wu, 2017; Morsy et al., 2017). Thus, it is necessary to seek optimal classification methods for taking full advantages of this new technique.

Recognized by the Massachusetts Institute of Technology as one of the ten breakthrough technologies of 2013, deep learning has a powerful capability of learning (Wen et al., 2018). Recently, it has been widely applied in the fields of artificial intelligence and remote sensing because of the notably reduced cost of computing hardware, improved chip processing capability, and the significant development of the learning algorithms (Lecun et al., 2015). Deep learning classification methods, especially for convolutional neural networks (CNN), can acquire higher accuracy than other conventional classification approaches such as the support vector machine (SVM) (Zhang et al., 2018). However, few published research has attempted to use deep learning methods and multispectral ALS data in combination to improve LC classification accuracy.

Therefore, both multispectral ALS techniques and deep learning networks have shown their superiority in LC classification. However, to the best of author's knowledge, since there is no similar research, it is very challenging to build an eligible workflow to train, validate and test deep learning networks using multispectral ALS data with an appropriate data structure. Thus, this study mainly aims to establish a workflow for automated pixel-wise classification using multispectral ALS data with a compatible data structure as input using deep learning based classification networks. The main contributions of this study are: (1) to establish and implement deep learning networks that are appropriate for multispectral ALS data classification; (2) to analyse how different information extracted from the multispectral ALS data impacts classification results; and (3) to assess how different deep learning networks can affect the classification results of multispectral ALS data.
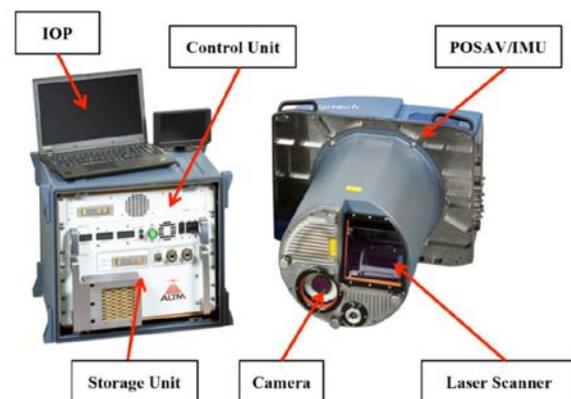
2. DATA AND STUDY AREA

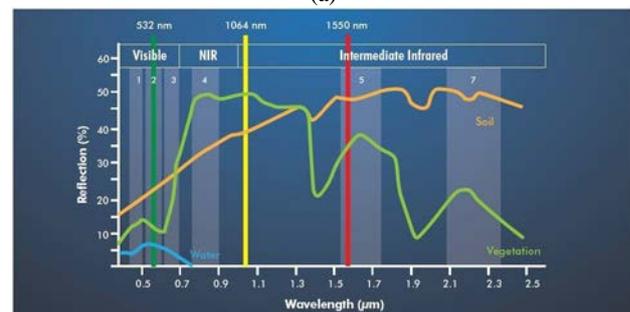### 2.1 Optech Titan Multispectral ALS System

The first commercial multispectral airborne active imaging LIDAR sensor in the world is the Teledyne Optech Titan multispectral ALS system. The components of the Teledyne Optech Titan multispectral ALS system are shown in Figure 1(a), which contains a flight management system, an operator laptop, a digital camera, a laser scanner assembly, a Global Navigation Satellite System (GNSS), an Inertial Measurement Unit (IMU), and a control and data recording unit are essential parts of a multispectral ALS system. It contains three active imaging channels of different wavelengths: 1550 nm (shortwave infrared, SWIR), 1064 nm (near infrared, NIR), and 532 nm (green, G), respectively. The three channels generate laser pulses with

Table 1. Specifications of Optech Titan multispectral ALS system.

Parameter	Specification
Wavelength	Channel 1: 1550 nm (shortwave infrared, SWIR) Channel 2: 1064 nm (near infrared, NIR) Channel 3: 532 nm (green, G)
Forward angles	Channel 1: 3.5° Channel 2: 0° Channel 3: 7°
Pulse repetition frequency	Programmable: 50 -300 KHz per channel; 900 KHz in total
Scan frequency	Programmable: 0 -210 KHz
Point density	Bathymetric: > 15 pts/m^2 Topographic: > 45 pts/m^2
Accuracy	Horizontal: 1/7, 500 × altitude, 1σ Vertical: < 5 – 10 cm, 1σ
Laser range precision	5 < 0.008 m, 1 σ

(a)

(b)

Figure 1. (a): The Teledyne Optech Titan system; (b) Titan laser channels with spectral signatures for selected objects (Teledyne Optech Titan, 2015).

separate forward angles to produce independent scan lines. As shown in Figure 1(b), green vegetation is strongly reflective in the NIR spectrum and slightly reflective in the visible G spectrum. Soil tends to reflect most at the SWIR band but lowest at the green band. Electromagnetic waves are mostly absorbed at the water surface in the NIR and SWIR spectrum. Thus, the three scanning frequencies provided by the Teledyne Optech Titan make it possible to acquire various spectral responses of different materials and to obtain diverse information about the surface of the Earth (Karila et al., 2018). Detailed specifications of the Teledyne Optech Titan are listed in Table 1.
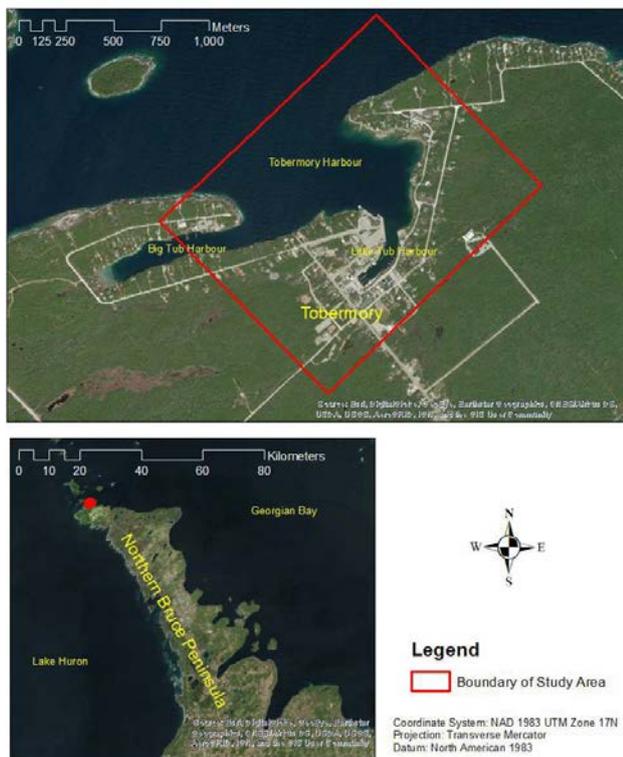
Figure 2. The study area in this research.

2.2 Data and Study Area

The study area is around 1.99 km^2 and is located at the northern tip of Tobermory, Ontario, Canada (see Figure 2). Tobermory is a town that is located at the northern tip of the municipality of Northern Bruce Peninsula and can be found between Lake Huron and Georgian Bay. Visual inspection of the study area reveals that there are six main LC classes: Water (WAT), Trees (TRE), Bare Land (BAL), Roads (ROD), Buildings (BUD), and Other Impervious Surfaces (OIS) such as parking lots and concrete docks.

Two kinds of datasets were used in the study, a multispectral ALS dataset and an orthophoto dataset. The multispectral ALS point clouds were collected by Teledyne Optech Titan multispectral ALS system in April of 2015. The Titan sensor was installed in an Optech's aircraft that flew at an altitude of about 457 m above ground level during the data collection. The total number of points is over 94 million with the average point spacing of 0.28 m and average point density of 47.4 pts/m^2, respectively. After collection, the data were first pre-processed by Optech. All points were calibrated, and all three channels were automatically aligned using the Optech's Lidar Mapping Suite software.

The orthophotos were provided by the Southwestern Ontario Orthophotography Project (SWOOP). The SWOOP images were collected using a Leica ADS100 airborne digital sensor between April 12 and May 23, 2015. The SWOOP images were collected at 2,377 m above mean terrain to produce 20 cm-resolution orthophotos with four channels (i.e. red, green, blue, and near-infrared). Because of its high spatial resolution and multi-wavelength, this dataset can clearly and precisely depict the Earth's surface of the study area. Thus, the orthophotos were considered as the ground truth for labelling.
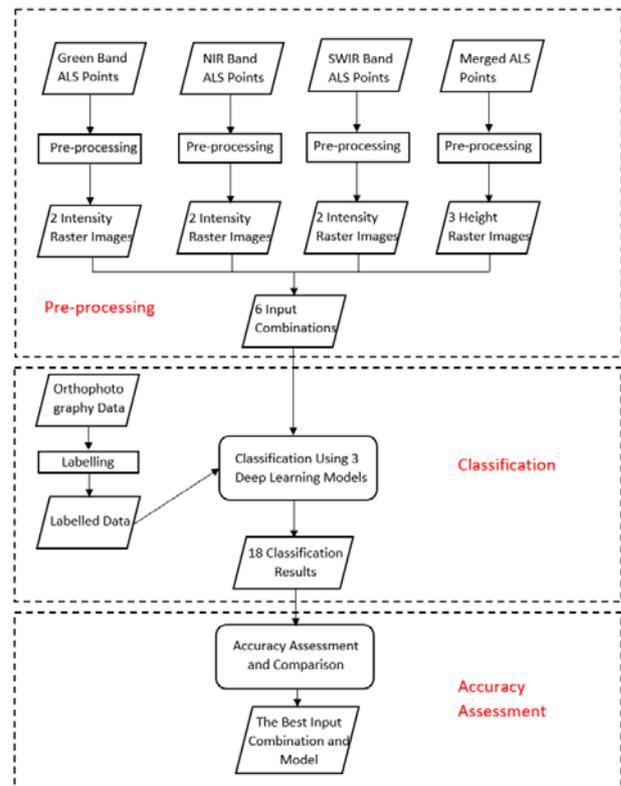

Figure 3. Workflow of the propose method.

3. METHOD

Figure 3 shows the workflow of the proposed method which contains three main parts: data pre-processing, classification, and accuracy assessment. The multispectral ALS point clouds were pre-processed at first. A total of nine raster images with different information were generated from the pre-processed point clouds. These images were assembled into six input data combinations. Meanwhile, the labelled dataset was created using the orthophotos as the ground truth. Also, three CNN-based networks established. Then, each input data combination was used to train and validate each network. This step developed eighteen LC classification models with different parameters to predict LC types for pixels. Therefore, a total of eighteen classification results were produced. Finally, accuracy assessments and comparisons were done for the eighteen classification results to seek an optimal scheme. Details of each part are described in the following sections.

3.1 Multispectral ALS Data Pre-processing

The data was firstly pre-processed by the Optech. All points were calibrated through geoid correction, and all three channels were automatically aligned using the Optech's Lidar Mapping Suite software. In this study, the original survey strips were cropped according to the boundary of the study area. Then data pre-processing was conducted for each of the thirty LAS files separately before merging them together. Moreover, a statistical outliner removal (SOR) filter, provided by CloudCompare v2.6.2 software, was utilized to eliminate isolated outliners which were away from all other points. After the outlier removal and intensity normalization steps, the thirty LAS files were merged into three point clouds based on the channel that they belonged to. Then, the three point clouds at different laser wavelength bands were projected to the 2D horizontal plane and rasterized into three intensity images. According to prior knowledge, a reasonable size is two to four times the point spacing (ArcGIS Desktop Help,
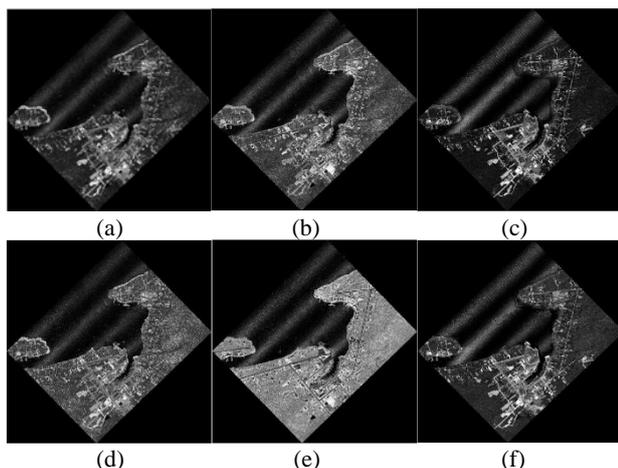
(a)   (b)   (c)

(d)   (e)   (f)

Figure 4. Multispectral ALS intensity models on the study area:
(a) $I_{SWIR+All+Avg}$; (b) $I_{NIR+All+Avg}$; (c) $I_{Green+All+Avg}$; (d)$I_{SWIR+First+Avg}$
; (e) $I_{NIR+First+Avg}$; and (f) $I_{Green+First+Avg}$.
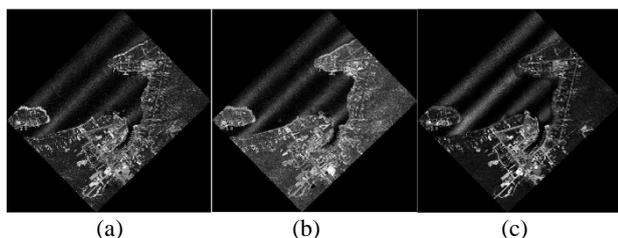

(a)   (b)   (c)

Figure 5. Multispectral ALS height models on the study area:
(a) $H_{First+Max}$; (b) $H_{First+Avg}$; and (c) $H_{All+Min}$.

n.d.). Based on the average point spacing of the dataset (Channel 1: 0.45 m per point; Channel 2: 0.43 m per point; Channel 3: 0.44 m per point), 0.8m, 1m, 1.2m, and 1.5m was tested as the cell size, respectively. After comparing voids of the generated result maps, ground resolution of the raster images was established at 1 m.

In an intensity image, a cell held an intensity value. With regard to a cell that contained more than one points, the cell value was defined as the distance-weighted average intensity value of all points within the cell. The distance-weighted average value was calculated as follows:

$$u(x) = \begin{cases} \frac{\sum_{i=1}^{N} w_i(x)u_i}{\sum_{i=1}^{N} w_i(x)}, & if\ d(x,x_i) \neq 0\ for\ all\ i \\ u_i, & if\ d(x,x_i) = 0\ for\ some\ i \end{cases} \quad (1)$$

$$w_i(x) = \frac{1}{d(x,x_i)} \quad (2)$$

where $u(x)$ is the final value of a given pixel at the central point x based on values of all points within the pixel samples $u_i = u(x_i)$ for $i = 1, 2, ..., N$. d is the given distance from the known point x_i to the unknown point x. w_i is the weight for a point x_i.

For a cell that contained no points, linear interpolation was used because of its high computational efficiency to compute the average of intensity values of eight neighbouring cells. To explore the most appropriate input dataset for multispectral ALS classification using deep learning methods, the steps of generating intensity images were executed twice. In the first execution, all returns of a single pulse from a laser were used to produce three intensity images at different laser wavelength bands (i.e., $I_{Green+All+Avg}$, $I_{NIR+All+Avg}$, and $I_{SWIR+All+Avg}$, respectively). In the second execution, only the first return of each pulse was considered to produce three intensity images (i.e., $I_{Green+First+Avg}$, $I_{NIR+First+Avg}$, and $I_{SWIR+First+Avg}$, respectively). The average intensity values of all returns reflected the content of objects while the average intensity values of the first returns

Table 2. Contents of input combination.

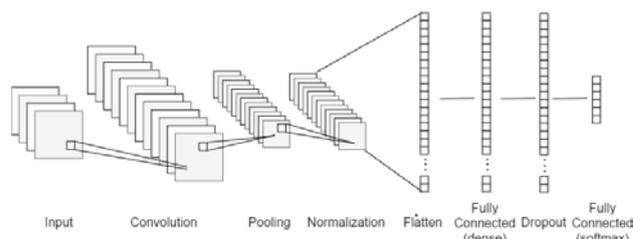| Combination	Content
1	$I_{Green+All+Avg}$ + $I_{NIR+All+Avg}$ + $I_{SWIR+All+Avg}$ + $H_{First+Max}$
2	$I_{Green+All+Avg}$ + $I_{NIR+All+Avg}$ + $I_{SWIR+All+Avg}$
3	$I_{NIR+All+Avg}$ + $H_{First+Max}$
4	$I_{Green+All+Avg}$ + $I_{NIR+All+Avg}$ + $I_{SWIR+All+Avg}$ + $I_{Green+First+Avg}$ + $I_{NIR+First+Avg}$ + $I_{SWIR+First+Avg}$ + $H_{First+Max}$
5	$I_{Green+All+Avg}$ + $I_{NIR+All+Avg}$ + $I_{SWIR+All+Avg}$ + $H_{First+Max}$ + $H_{First+Avg}$ + $H_{All+Min}$
6	$I_{Green+All+Avg}$ + $I_{NIR+All+Avg}$ + $I_{SWIR+All+Avg}$ + $I_{Green+First+Avg}$ + $I_{NIR+First+Avg}$ + $I_{SWIR+First+Avg}$ + $H_{First+Max}$ + $H_{First+Avg}$ + $H_{All+Min}$

Figure 6. Structure of the proposed CNNs.

described the top surface of these objects. Therefore, total six intensity images were generated (see Figure 4). Similarly, to provide more information related to height, the height information was first normalized to reduce the influence of steep terrain fluctuation, the steps for generating a height image were then repeated twice for the average height of first returns and the minimum height of all returns, respectively. Thus, three height images were generated in total, named as $H_{First+Max}$, $H_{First+Avg}$, and $H_{All+Min}$, respectively (see Figure 5).

Then, a multi-tiered raster-based architecture was used to contain rich information. Because raster-based images could retain relative position information of raster cells to the maximum extent, which would benefit the LC classification. Moreover, a raster image could store a type of information extracted from the original point clouds. Therefore, the input dataset was a stack of several raster images with same length, width, and cell size. The length and width were the same as those of the study area. The depth depended on the number of raster images that the input dataset had. These layers were assembled into a variety of combinations as listed in Table 2.

3.2 Convolutional Neural Network Classification
Before training the networks and predicting LC classes, a labelled dataset that explained the LC type for each pixel within the study area was required. In the labelling process, first, a blank raster image whose length, width, cell size, and coordinate were exactly the same as those of previously generated intensity images and height images was created. Then, each cell of the blank raster was manually assigned an LC class according to the reference map, which was based on the orthophotos of the study area. Finally, six major types were selected: WAT (i.e., open water, harbours, and small lakes), TRE (i.e., single and multiple), BAL (i.e., commercial and residential buildings), ROD (i.e., straight road and crossroad), BUD (i.e., sand, rocky area, and grass), and OIS (i.e., parking lot, concrete area, and boats).

After the previous labelling process, three CNN-based networks were designed based on the number of the dimension of the convolutional layers. They are 1D CNNs, 2D CNNs, and 3D CNNs, respectively. These CNNs were established using the scripts shown in Figure 6. In this research, each proposed CNN

Table 3. Hyperparameters involved in the designed CNNs.

Hyperparameter	Value		
	1D-CNN	2D-CNN	3D-CNN
Shape of input unit	(1, 1, depth)	(7, 7, depth)	(7, 7, depth)
Number of kernels	256	512	512
Size of kernels	3	(3, 3)	(3, 3, 3)
Size of pooling windows	2	(2, 2)	(2, 2, 2)
Units of dense	1024	2048	2048
Rate of training, validation, and testing data	(60% : 10% : 30%)		
Learning rate	0.005	0.005	0.005

has seven hidden layers including a convolutional layer, a pooling layer, two fully connected layers, and three other functional layers (see Figure 6). Although the framework of the proposed neural network is simple, it has been proven to be efficient and robust. During the training process, batch normalization was used after each convolutional operation. Then, the rectified linear unit (ReLU) activation function was employed. Furthermore, dropout layer, which randomly sets a fraction rate of input units to 0 to prevent complex co-adaptations on training data, was applied to prevent the overfitting. After feature extraction, 3D feature maps were flattened to 1D feature vectors. Then, the vectors were feed into fully connected (FC) layers. In this study, each hyperparameter was tested separately, keeping all other hyper-parameters constant. As listed in Table 3, there were five key hyper-parameters involved in the establishment of each CNN according to multiple experiments. To better train a model and avoid overfitting caused by insufficient training data, training data account for more than half of the entire dataset. Validation data often account for only 10% of the dataset. Thus, different rates of training, validation and testing data were tested for each model. Finally, based on the overall accuracy evaluation, training, validation, and testing data was ascertained as 60%, 10%, and 30%, respectively. Moreover, the learning rate was set as 0.005 for the whole training process based on prior knowledge.

3.3 Accuracy Assessment

To validate the accuracy of labelled data, a 10 x10 pixels window was created to randomly capture 1000 sampling areas after labelling. As a result, a total of 100,000 pixels within these sampling areas were relabelled and compared with the previous labelling results. The proportion of the pixels that were relabelled uniformly to the quantity of total tested pixels was calculated. In order to evaluate the proposed methods, the confusion matrix, commission errors (CE), user's accuracy (UA), omission errors (OE), producer's accuracy (PA), overall accuracy (OA), and kappa coefficient were calculated.

4. RESULTS AND DISCUSSION

4.1 Results of Labelling

Each 1m pixel within the study area was labelled as one of the six LC types based on the rules and steps proposed in Section 3.2. The manually labelled image is presented in Figure 7. It can be seen that the boundary of each class has been smoothly and clearly labelled. Additionally, details in the study area such as a few BAL pixels along the ROD have been accurately labelled. There is a total of 1,990,682 pixels in the study, around 80% of pixels of the study area belong to WAT or TRE while BUD and ROD only occupy less than 3% of pixels, respectively. Moreover, only 7% of the pixels are BAL pixels. Excessive imbalance of area of each LC type may negatively influence classification results since the number of pixels of a specific class may be too

Table 4. Confusion matrix of labelled dataset.

LC type		First labelled dataset							
		WAT	TRE	ROD	BAL	BUD	OIS	Total	Accuracy
Relabelled dataset	WAT	38764	1	0	0	1	0	38766	99.99
	TRE	0	35286	1	1	2	0	35290	99.99
	ROD	0	1	3991	0	0	0	3992	99.97
	BAL	0	0	0	8038	0	3	8041	99.96
	BUD	1	0	0	1	3926	0	3928	99.95
	OIS	0	0	0	2	0	9981	9983	99.98
	Total	38765	35288	3992	8042	3929	9984	100000	N\A
	Accuracy	99.99	99.99	99.97	99.95	99.92	99.97	N\A	N\A

Table 5. Accuracy evaluation.

Input data combination	1D-CNN		2D-CNN		3D-CNN	
	OA (%)	Kappa	OA (%)	Kappa	OA (%)	Kappa
1	90.3	0.85	96.5	0.95	96.8	0.95
2	79.2	0.68	96.5	0.95	96.6	0.95
3	82.5	0.73	90.8	0.87	91.0	0.87
4	91.2	0.87	97.0	0.96	97.2	0.96
5	90.3	0.85	96.6	0.95	96.8	0.95
6	91.0	0.86	97.0	0.96	97.0	0.96

Figure 7. Labelled LC map of the study area.

small to be learned. As mentioned in section 2.2, 100,000 pixels were relabelled to validate the labelled dataset. As indicated in Table 4, a confusion matrix is created to clearly display the validation result. As a result, the accuracy of each class is higher than 99.9% no matter which labelled dataset is used as the reference dataset. Furthermore, OA of statistics listed in Table 4.3 is 99.99%, which means 99.99% pixels have the same labels in the first-labelled dataset and the relabelled dataset. Thus, it can be concluded that the labelled dataset is reliable.

4.2 Results of LC Classification

There are six input data combinations (i.e. Combination 1-6, defined in Section 3.1) and three CNNs used in this thesis for LC

Water ▪ Trees ▪ Roads ▪ Bare Land ▪ Buildings ▪ Other Impervious Surface

(a) Combination 1

(b) Combination 2

(c) Combination 3

(d) Combination 4
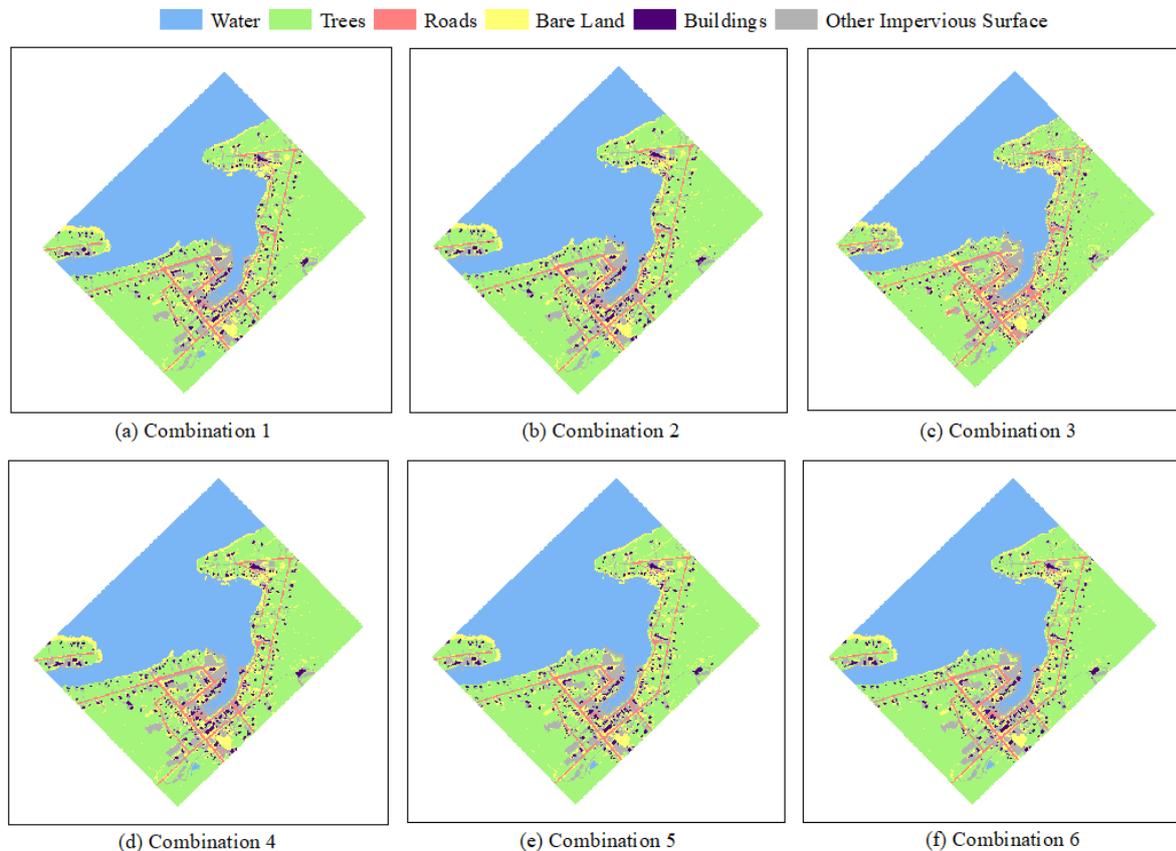
(e) Combination 5

(f) Combination 6

Figure 8. Predicted maps of different CNNs with Combination 4.

classification. Therefore, there are totally eighteen trained and validated models. To ensure the significance of the classification accuracy, each model was run 10 times. The averaged OA and kappa coefficient of each model are listed in Table 5. It can be seen that the highest overall classification accuracy of 97.2%, with a kappa index of 0.96, can be achieved using the proposed 3D CNN and input data Combination 4. The OA of the three input datasets designed in this study, Combinations 4, 5 and 6, is on average 3.8% point higher than that of the classic input datasets, Combination 1, 2 and 3. Combination 4 achieves the best OA and kappa coefficient no matter which CNN is applied. Combination 3 on average performs worst compared to others, especially when the 2D or 3D CNNs is applied. Combination 2 is the most sensitive to the alteration of CNNs; when using this input data, OA of 2D or 3D CNN is at least 17% higher than OA of 1D CNN. Furthermore, the 3D CNN obtains the highest OA and kappa coefficient, indicating that it has a high success for pixel-wise LC classification and performs significantly better than random. OA and kappa coefficient of the 2D CNN are only slightly lower than them of the 3D CNN, suggesting that the 2D CNN also performs well in pixel-wise LC classification. Conversely, the 1D CNN achieves the lowest classification accuracy with relatively low OA and kappa coefficient. The OA of the 2D and 3D CNNs was on average 8.4% higher than that of the 1D CNN. Thus, this study achieves an admirable classification result, which is better than most of the published multispectral ALS data classification results. Accordingly, the predicted maps of different input data combinations using 3D CNN are shown in Figure 8, which visualizes the classification results of the best-trained models. Furthermore, the prediction time for 1D CNN, 2D CNN, and 3D CNN are 214 s, 534 s, and 2,669 s, respectively, which has been proven to be efficient for

Table 6. Comparative study and results.

Classification method	Methods	No. of classes	OA (%)
CNN	In this study	6	97.2
MLC	Bakula et al. (2016)	6	90.9
	Fernandez-Diaz et al., (2016)	5	90.2
	Morsy et al., (2017a)	4	89.9
SVM	Teo and Wu, (2017)	5	96.0
RF	Zou et al. (2016)	6	95.9
	Matikainen et al. (2017)	9	91.6

large-scale LC classification from massive multispectral ALS point clouds.

4.3 Comparison

A comparative study of LC classification methods for multispectral ALS data is carried out. Results achieved by Combination 4 and the three CNNs are used for the comparisons. The classification accuracy of the proposed CNNs is compared with that of three widely used traditional classification methods (i.e. MLC, SVM, and RF). All involved studies are summarized in Table 6.

MLC algorithm, which assigns cells an LC class based on the measure of the highest likelihood, were applied in three studies to map LC classes for multispectral ALS data. Bakuła et al. (2016) used a raster-based MLC to classify a multispectral ALS point cloud into six classes, achieving an overall accuracy of 91% in the best test. In this attempt, WAT, TRE, and BUD were classified accurately; however, BLD and OIS were misclassified

(Bakuła et al. 2016). Furthermore, Morsy et al. (2017a) also applied a raster-based MLC method to classify multispectral ALS data to four classes and obtained an overall accuracy of 89.9%. Similarly, Fernandez-Diaz et al. (2016) implemented a raster-based MLC to categorize a multispectral ALS dataset into five LC classes with the best overall accuracy of 90.2%. To conclude, OA of MLC methods is slightly lower than that of the 1D CNN and at least 6% lower than the 2D and 3D CNNs. The reason might be that the MLC assumes that a training sample is normally distributed, which is often not the case. This incorrect assumption can introduce errors, especially when classifying urban landscapes.

SVM applies optimization algorithms to determine the location of ideal boundaries that can most effectively distinguish between classes (Lamine et al., 2018). An object-based SVM classification method was tested by Teo and Wu (2017) to categorize multispectral ALS data into five classes, achieving an overall accuracy of 96%. Although OA of the SVM is only slightly lower than that of the 2D and 3D CNNs, SVM classification still has a major limitation since the selection of the kernel function and the setting of proper parameter values are decided subjectively by the users and only few studies have been conducted on the determination of the optimal choice of kernel function and proper settings for corresponding parameter (Petropoulos et al., 2012). Moreover, with the increased number of classes to be classified, the classification accuracy will remarkably decrease, which lead to SVM-based classification methods are not suitable for large-scale LC classification. Additionally, SVM needs hand-designed features which significantly impact the classification accuracy. Compared to SVM, the proposed CNN-based methods are more robust and efficient with different scales of input point clouds.

The RF method is a collection of Decision Trees, which are the predictive model that uses a set of binary rules as nodes to acquire the best solution. An object-based decision tree model was implemented to multispectral ALS data by Zou et al. (2016) to accomplish a 9-class LC classification, reaching an overall accuracy of 91.6%. However, the decision tree algorithm tends to over-fit training data, especially when a tree is particularly deep. Matikainen et al. published several articles (2017a; 2017b) on the application of an object-based RF LC classification method to multispectral ALS datasets, which achieved an OA of 95.9% for six classes. This method performed better for BUD, TRE, and OIS, but lead to low correctness for BAL. The OA of the RF method only slightly lower than that of the 2D and 3D CNNs. Also, the large number of trees in the RF method may make classification process slow, especially when applied to a large dataset such as a dense multispectral ALS point cloud in a large area.

To conclude, the 2D and 3D CNNs proposed in the study can achieve higher LC classification accuracy for multispectral ALS data than the traditional classification methods especially the MLC. Although as classic machine learning algorithms, SVM and RF can provide relatively precise and reliable classification results for multispectral ALS data, both of them require hand-designed features which significantly impact the classification accuracy. This characteristic of classic machine learning algorithms makes them highly user-dependent. This limitation of SVM and RF also can be conquered by all deep learning networks like the CNNs proposed in the study.

5. CONCLUSION

In this paper, we have proposed a workflow for an automated pixel-wise LC classification for multispectral ALS data using CNN-based networks. A total of six input datasets are used with multi-tiered architecture and three CNNs are proposed to seek an optimal scheme. An overall classification accuracy of 97.2%, with a kappa index of 0.96, is achieved using the proposed 3D CNN framework with Combination 4. It represents a significant classification accuracy since it is on average 4% higher than the accuracy of the published multispectral ALS LC classification methods. Generally, this study delivers the feasibility of combining, for the first time, multispectral ALS data and deep learning to improve the performance of the automated pixel-wise LC classification.

Furthermore, this study analyzes how different information extracted from the multispectral ALS data impacts the classification accuracy by comparing various input data combinations. It demonstrates that the spectral information is more helpful than the height information for CNN-based LC classification. However, when using the additional information together, the added height information deteriorates classification performance when using the additional spectral information solely. Therefore, the optimal rasterized multispectral ALS dataset for LC classification should consist of height information of the first returns and spectral information of the first returns and all returns. This comparison reveals that the multispectral ALS technique is superior to both traditional multispectral optical imagery and typical single-wavelength ALS data for LC classification.

In conclusion, the results presented in this study show that the LC classification accuracy can be remarkably improved by using the multispectral ALS data and CNNs. The study indicates the potential of multispectral ALS data in LC mapping, which may draw more people' attention to this new technique. Once the multispectral ALS data become widely available, more multispectral ALS data with professional labelling datasets will be published for researchers to investigate, which may further improve LC classification. Consequentially, it may accelerate the development of multispectral ALS techniques. However, labelling work needs rich experience, which has a great impact on the labelling accuracy.

REFERENCES

Bakuła, K., Kupidura, P. and Jełowicki, Ł., 2016. Testing of land cover classification from multispectral airborne laser scanning data. *International Archives*, 41, 161-169.

Chen, X., Ye, C, Li, J. and Chapman, M.A., 2018. Quantifying the Carbon Storage in Urban Trees Using Multispectral ALS Data. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.,* 11(9), 3358-3365.

Costa, H., Foody, G.M. and Boyd, D.S., 2018. Supervised methods of image segmentation accuracy assessment in land cover mapping. *Remote Sens. Environ.,* 205, 338-351.

Ene, L.T., Gobakken, T., Andersen, H.E., Næsset, E., Cook, B.D., Morton, D.C., Babcock, C. and Nelson, R., 2018. Large-area hybrid estimation of aboveground biomass in interior Alaska using airborne laser scanning data. *Remote Sens. Environ.,* 204, 741-755.

Fernandez-Diaz, J.C., Carter, W.E., Glennie, C., Shrestha, R.L., Pan, Z., Ekhtari, N., Singhania, A., Hauser, D. and Sartori, M., 2016. Capability assessment and performance metrics for the Titan multispectral mapping LiDAR. *Remote Sens.*, 8(11), 936-969.

Glennie, C.L., Carter, W.E., Shrestha, R.L. and Dietrich, W.E., 2013. Geodetic imaging with airborne LiDAR: The Earth's surface revealed. *Rep. Prog. Phys.*, 76(8), 086801.

Guida-Johnson, B. and Zuleta, G.A., 2013. Land-use land-cover change and ecosystem loss in the Espinal ecoregion, Argentina. *Agric. Ecosyst. Environ.*, 181, 31-40.

Karila, K., Matikainen, L., Litkey, P., Hyyppä, J. and Puttonen, E., 2018. The effect of seasonal variation on automated land cover mapping from multispectral airborne laser scanning data. *Int. J. Remote Sens.*, 1-19.

Lamine, S., Petropoulos, G.P., Singh, S.K., Szabó, S., Bachari, N.E.I., Srivastava, P.K. and Suman, S., 2018. Quantifying land use/land cover spatio-temporal landscape pattern dynamics from Hyperion using SVMs classifier and FRAGSTATS®. *Geocarto Int.*, 33(8), 862-878.

LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *Nature*, 521(7553), 1-436.

Ma, L., Li, Y., Li, J., Wang, C., Wang, R. and Chapman, M., 2018. Mobile laser scanned point-clouds for road object detection and extraction: A review. *Remote Sens.*, 10(10), 1531.

Mariano, D.A., dos Santos, C.A., Wardlow, B.D., Anderson, M.C., Schiltmeyer, A.V., Tadesse, T. and Svoboda, M.D., 2018. Use of remote sensing indicators to assess effects of drought and human-induced land degradation on ecosystem health in Northeastern Brazil. *Remote Sens. Environ.,* 213, 129-143.

Matikainen, L., Karila, K., Hyyppä, J., Litkey, P., Puttonen, E. and Ahokas, E., 2017. Object-based analysis of multispectral airborne laser scanner data for land cover classification and map updating. *ISPRS J. Photogramm. Remote Sens.*, 128, 298-313.

Matikainen, L., Karila, K., Hyyppä, J., Puttonen, E., Litkey, P. and Ahokas, E., 2017. Feasibility of multispectral airborne laser scanning for land cover classification, road mapping and map updating. *ISPRS Archives*, 119-122.

Morsy, S., Shaker, A. and El-Rabbany, A., 2017a. Multispectral LiDAR data for land cover classification of urban areas. *Sensors*, 17(5), 958-979.

Morsy, S., Shaker, A. and El-Rabbany, A., 2017. Clustering of multispectral airborne laser scanning data using gaussian decomposition. *ISPRS Archives*, 42.

Petropoulos, G.P., Kalaitzidis, C. and Vadrevu, K.P., 2012. Support vector machines and object-based classification for obtaining land-use/cover cartography from Hyperion hyperspectral imagery. *Comput. Geosci.*, 41, 99-107.

Sun, Y., Zhang, X., Xin, Q. and Huang, J., 2018. Developing a multi-filter convolutional neural network for semantic segmentation using high-resolution aerial imagery and LiDAR data. *ISPRS J. Photogramm. Remote Sens.*, 143, 3-14.

Teledyne Optech Titan, 2015. Multispectral LiDAR system: high precision environmental mapping, Retrieved from http://www.teledyneoptech.com/wp-content/uploads/Titan-Specsheet-150515-WEB.pdf.

Teo, T.A. and Wu, H.M., 2017. Analysis of Land Cover Classification Using Multi-Wavelength LiDAR System. *Applied Sciences*, 7(7), 663-683.

Wen, C., Sun, X., Li, J., Wang, C., Guo, Y. and Habib, A., 2019. A deep learning framework for road marking extraction, classification and completion from mobile laser scanning point clouds. *ISPRS J. Photogramm. Remote Sens.*, 147, 178-192.

Yan, W.Y., Shaker, A. and El-Ashmawy, N., 2015. Urban land cover classification using airborne LiDAR data: A review. *Remote Sens. Environ*, 158, 295-310.

Ye, C., Li, J., Jiang, H., Zhao, H., Ma, L., & Chapman, M., 2017. Semi-automated Generation of Road Transition Lines Using Mobile Laser Scanning Data. *IEEE Trans. Intell. Transp. Syst.*, 1-14. Doi: 10.1109/TITS.2019.2904735

Zhang, C., Sargent, I., Pan, X., Gardiner, A., Hare, J. and Atkinson, P.M., 2018. VPRS-based regional decision fusion of CNN and MRF classifications for very fine resolution remotely sensed images. *IEEE Trans. Geosci. Remote Sens.*, 56(8), 4507-4521.

Zhong, Z., Li, J., Ma, L., Jiang, H., & Zhao, H., 2017, July. Deep residual networks for hyperspectral image classification. In *2017 IGARSS*, 1824-1827, IEEE.

Zou, X., Zhao, G., Li, J., Yang, Y., & Fang, Y., 2016. 3D land cover classification based on multispectral LiDAR point clouds. *ISPRS Archives*, XLI-B1, 741-747.