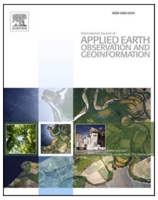
International Journal of Applied Earth Observations and Geoinformation 108 (2022) 102723

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

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



AGFP-Net: Attentive geometric feature pyramid network for land cover classification using airborne multispectral LiDAR data

Dilong Li[a], Xin Shen[b], Haiyan Guan[c,*], Yongtao Yu[d], Hanyun Wang[e], Guo Zhang[b], Jonathan Li[f], Deren Li[b]

[a] College of Computer Science and Technology, Huaqiao University, Xiamen, FJ 361021, China
[b] State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, HB 430072, China
[c] School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing, JS 210044, China
[d] Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huaian, JS 223003, China
[e] School of Surveying and Mapping, Information Engineering University, Zhengzhou, HN 45000, China
[f] Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada

ARTICLE INFO

Keywords:
Land cover classification
Airborne multispectral LiDAR point clouds
Attention mechanism
Feature pyramid
Graph geometric moments convolution

ABSTRACT

Accurate land cover (LC) classification plays an important role in ecosystem protection, climate changes, and urban planning. The airborne multispectral LiDAR data are increasingly used for high-resolution and accurate LC classification tasks. However, most of the existing methods lack of the comprehensive extraction of the spatial geometric structure features, and ignore the fusion of multi-scale extracted features. In this paper, a point-wise deep learning-based method is proposed for LC classification based on airborne multispectral LiDAR data. We present a novel convolution operator to efficiently extract the spatial geometric structure features, called attentive graph geometric moments convolution (AGGM Convolution). Besides, to fuse the extracted multi-scale features, we propose a feature up-sampling module and construct a feature pyramid to integrate the features with different scales. The proposed method was evaluated using multispectral LiDAR data acquired with an airborne Teledyne Optech Titan system. In comparison with the previously developed state-of-the-art point cloud segmentation models, the proposed method behaves superiorly with an overall accuracy of 96.9% and a Kappa index of 0.950 on the test scenes. The quantitative assessments demonstrate that the proposed method performs effectively and efficiently in land cover classification tasks.

1. Introduction

Land cover (LC) classification is an important means to monitor the change of the Earth surface, global ecosystem (Lunetta et al., 2002), Earth radiation balancing (Hanna, 2007), and climate (Feddema et al., 2005). In the early study of LC classification, the multispectral image data were utilized as the main data source to acquire the Earth surface information. Since the different LC types show different spectral reflectivity in various wavelengths, the land cover could be classified by the spectral information extracted from the multispectral image data (Wilkinson, 2005). As the rapid development of the society, the demands of surface change monitoring are being more and more precise, which requires higher resolution and accuracy LC classification products. Theoretically, the higher resolution of the multispectral image, the higher accuracy of the LC classification products. However, according to the research of (Wilkinson, 2005), the accuracy of LC classification did not significantly improve with the resolution of the multispectral image in the past decades. The main reason is that the separability of land covers is reduced by the between-class spectral confusing and within-class spectral diversity in the multispectral image (Yan et al., 2012). Besides, the shadow and shaded areas are also the factors (Dare, 2005; Zhou et al., 2009). Therefore, many researchers think the accuracy of LC classification is limited to the only data source input, and the other data sources could be introduced as the supplement to further enhance the accuracy of LC classification (Yan et al., 2012).

In the past two decades, light detection and ranging (LiDAR) data have gradually treated as a widely used remote sensing data source for Earth observations and analyses owing to its unique property of

* Corresponding author.
E-mail addresses: scholar.dll@hqu.edu.cn (D. Li), xinshen@whu.edu.cn (X. Shen), guanhy.nj@nuist.edu.cn (H. Guan), allennessy@hyit.edu.cn (Y. Yu), guozhang@whu.edu.cn (G. Zhang), junli@uwaterloo.ca (J. Li), drli@whu.edu.cn (D. Li).

https://doi.org/10.1016/j.jag.2022.102723
Received 24 November 2021; Received in revised form 13 January 2022; Accepted 12 February 2022
Available online 24 February 2022

containing high precision three-dimensional spatial information (Glennie et al., 2013). Compared with the two-dimensional image data, LiDAR data can obtain the more precise terrain and ground surface information, and cannot be affected by the cloud coverage, weather condition, and relief displacement (Glennie et al., 2013). In the research of (Antonarakis et al., 2008; Lodha et al., 2007; Mallet et al., 2008), the feasibility of classifying land covers by using the airborne LiDAR data was analyzed and validated. However, the regular LiDAR data only contains single-wavelength spectral information, which severely limits its land cover separability, especially for the complex scenes with similar shapes. To overcome this drawback, many researches integrated the spectral information and spatial information to achieve better classification performance.

Compared with the use of only data source, the combination of multispectral image and regular LiDAR data can achieve better classification results indeed. For instance, (Kim and Kim, 2014) fused the WorldView-2 image data and airborne regular LiDAR data, and achieved higher LC classification accuracy than the previous works. For multimodal data fusion, (Hong et al., 2021a) presented a general multimodal deep learning framework, which considered different fusion strategies and used for the models with CNNs. However, the heterogeneous data collected from different sensors have different data formats, projections, resolutions, and acquisition time. It inevitably produces errors during the data fusion process. The additional data preprocessing and data calibration works are essential to narrow the error (Yan et al., 2012), but it is still an open problem to perfectly fuse the heterogeneous data.

The emergence of multispectral LiDAR technologies avoids the above problems caused by the fusion of multi-source data and attracts many researchers to utilize multispectral LiDAR data in this research field. Firstly, (Wichmann et al., 2015; Gong et al., 2015) assessed the potential and feasibility of applying multispectral LiDAR data into LC classification. Then, (Bakuła et al., 2016; Morsy et al., 2017a; Teo and Wu, 2017) validated that using the multispectral LiDAR data could achieve better performance than the fusion of multispectral image and regular LiDAR data.

In regard of the methods using in this field, classical machine learning-based methods are still the mainstream. The paradigmatic architectures first extract the handcrafted features, which are usually called “feature extraction” or “feature representation”. Typical features include height (Teo and Wu, 2017; Matikainen et al., 2016), spectral signature (Teo and Wu, 2017; Matikainen et al., 2016), texture, and normalized difference vegetation index (NDVI) (Teo and Wu, 2017; Matikainen et al., 2016). Then, the classical machine learning methods are utilized as the classifier to recognize the different land cover types. Typical techniques include maximum likelihood (Bakuła et al., 2016; Morsy et al., 2017a; Fernandez-Diaz et al., 2016), random forest (Matikainen et al., 2016; Matikainen et al., 2017a; Matikainen et al., 2017b), and support vector machine (SVM) (Teo and Wu, 2017; Ekhtari et al., 2018). Nevertheless, in terms of the performance, these methods are greatly impacted by parameter settings and feature selection.

Recently, the success of deep learning-based methods applying to image processing has motivated the data-driven approaches to apply in this field. In current study (Pan et al., 2020), the traditional used machine learning-based classifiers were replaced by CNNs. Unsurprisingly, CNNs, as the more powerful classifier, achieve better performance than the other classical classifiers. However, due to the unstructured nature of point clouds, (Pan et al., 2020) needs to convert the raw point clouds into images, which inevitably causes information loss. To deeply mine the spectral features, (Hong et al., 2021b; Hong et al., 2021c) utilized deep learning techniques (GCNs and Transformers) to extract spectral features for classification tasks, and achieved a promising classification performance.

To further improve the performance of the LC classification results, we design a point-wise deep learning-based method to directly classify the raw multispectral LiDAR data into land covers of interest. The main contributions of this paper are listed as follows:

1. We propose a novel convolution operator, called AGGM Convolution, which combines the attention mechanisms and graph geometric moments convolution to extract and aggregate the local geometric features effectively. The attention mechanisms can integrate the learned features with the learnable weights, which achieves significant improvement than the max-pooling or average-pooling operation.
2. We propose a feature up-sampling module and construct a feature pyramid to integrate the features with different scales. The feature up-sampling module can convert the extracted features with different scales and sizes into the specified form. The feature pyramid not only comprises the features in the encoder layers, but also comprises the features in the decoder layers, which contains more details from different scales.
3. We validate the potential of multispectral airborne LiDAR data and the effectiveness of the proposed method for LC classification applications, which provides the positive reference for the further research in this field.

The rest of this paper is organized as follows. Section 2 presents the related work. Section 3 introduces the study area and the data used in this paper. Section 4 details the proposed method. Section 5 presents the experimental results. Section 6 provides the concluding remarks.

2. Related work

Since the early works usually converted the multispectral LiDAR data into images, in terms of input data, the approaches of LC classification by using multispectral LiDAR data can be divided into two branches: the image-based methods and point-based methods. Besides, we also review the related point-wise deep learning-based methods.

2.1. *Image-based LC methods*

The image-based methods need to extract the features images firstly. Then, various classification methods are utilized to classify these feature images. According to the classification strategy, the image-based methods can be further categorized into two types: pixel-based and object-based (or segment-based) methods.

2.1.1. *Pixel-based LC methods*

Based on the pixel values contained in the feature images, the pixel-based methods usually directly apply the classifiers or classification strategies (like threshold) to globally classify the land cover types of each pixel. (Bakuła et al., 2016) classified the multispectral LiDAR data into six classes by using Maximum Likelihood Classification (MLC), and analyzed the impact of different combinations of inputs. The best attempt achieved an overall accuracy of 91%. Similarly, (Morsy et al., 2017a) integrated the three rasterized intensity images (extracted from the three channel wavelength LiDAR data) with the Digital Surface Model (DSM) rasterized image, and achieved an overall accuracy of 89.9% by using the MLC. Through experiments, (Fernandez-Diaz et al., 2016) observed that when the intensity images of channel 2 and channel 3 were used as the input, the MLC algorithm can obtain the optimal overall accuracy of 90.2%. On the contrary, adding the intensity image of Channel 1 would lead to the decline of the classification accuracy.

Recently, some researchers utilized the deep learning-based methods to replace the classical machine learning methods as the classifier, which obtained promising performances. (Pan et al., 2020) used the intensity and elevation images of each channel as the input, and then used the convolutional neural networks (CNNs) as the classifier. Compared with their previous work, which used the same inputs and adopted the classical machine learning methods as the classifier, (Pan et al., 2020) achieved a significant improvement on the classification accuracy.

2.1.2. Object-based LC methods

With the development of image semantic segmentation technology, many researchers first pre-segment the input feature images to obtain the roughly-segmented or over-segmented result, then the classifiers are applied to further classify the land cover types based on that. (Teo and Wu, 2017) extracted the spectral, elevation, and textural features from multispectral LiDAR data firstly, then segmented the feature images by using eCognition software. Finally, the SVM algorithm was utilized to classify the pre-segmented images into five classes. Similarly, (Matikainen et al., 2016) extracted 22 classes of features based on spectral information, elevation information, and the NDVI, and segmented these features by the eCognition software. The random forest (RF) algorithm was utilized to obtain the classification results of six types of land covers. The following studies (Matikainen et al., 2017a; Matikainen et al., 2017b) refined the RF-based LC classification methods. (Zou et al., 2016) used several different scales to generate multi-resolution intensity images from multispectral LiDAR data, and segmented these images by using eCognition software. Then, they chose the decision tree as the classifier to obtain the classification results of nine types of land covers. However, the problem of the decision tree method is that it is easy to cause overfitting, especially when the decision tree is deep. (Ghaseminik et al., 2021) presented a segment-based classification scheme for land cover mapping. They firstly employed a mutual segmentation based on intensity and height feature images, then classified the multispectral LiDAR data into seven classes by using the RF classifier. The best attempt achieved an overall accuracy of 94.83%.

2.2. Point-based LC methods

The point-based methods usually separate the ground and non-ground points by elevation information firstly. Then, various classifiers are utilized to classify the point clouds by the other features. Most of the existing point-based methods classify the point cloud by adopting the “point-by-point” classification strategy. To our best knowledge, there are still no point-based methods adopting the “segment-based” classification strategy.

As a pioneering work, Wichmann et al. (2015) proposed a point-based multi-algorithm and multi-phase method. The method applied a hybrid approach of progressive TIN densification to separate the ground points, and classified the rest points into building and vegetation classes by using a RANSAC-based segmentation algorithm. Morsy et al. (2017a) also adopted the multi-algorithm and multi-phase strategy. They used the skewness balancing algorithm to divide the point clouds into ground and non-ground points firstly. Then, they utilized the Jenks natural breaks optimization method to define the threshold of the NDVI values and grouped the non-ground points and ground points into specific classes. To further improve the classification accuracy, in their following work (Morsy et al., 2017b), they replaced the skewness balancing algorithm with the MLC algorithm and obtained a higher overall accuracy. Although these multi-algorithm and multi-phase methods could achieve a decent accuracy, it is difficult to be widely used. Because these methods are usually designed for specific multispectral LiDAR data characteristics, the data characteristics (such as data type, content, and target category) would significantly affect the classification performance.

Ekhtari et al. (2018) proposed a point-based LC classification method, which used the SVM algorithm as the classifier. The method combined the spectral and elevation features of each point as the input, and obtained the category of each point by using the SVM algorithm. Besides, they compared their method with the image-based method, and found that the results of the point-based method show higher accuracy with the same inputs and classifier. Wang and Gu (2020) proposed a 3-D LC classification method based on the tensor representation. They used the second-order tensor to represent the point clouds, which combines the spatial and spectral information. Then, they developed a tensor manifold discriminant embedding (TMDE) algorithm to extract features and classified the input point clouds with these extracted features by the SVM algorithm.

2.3. Point-wise deep learning-based methods

As the pioneering work in this field, PointNet (Qi et al., 2017a) utilized the Multi-Layer Perceptrons (MLPs) to directly extract features from raw point clouds and handle the permutation invariance issue by using symmetric operations. However, PointNet only learns the features from the coordinates of each point, which ignore the relationship between the point and its neighbors. As the extension of PointNet, PointNet++ (Qi et al., 2017b) considered the local semantic relationship of points and extracted the local features by implementing PointNet iteratively. Besides, PointNet++ applied multi-scale feature aggregation strategy to achieve better robustness. To better represent the local semantic relationship, RS-CNN (Liu et al., 2019) considered the more complicated relations between a point and its neighbors, and learned the high-level relationships from low-level relationships through MLPs. Recently, RandLA-Net (Hu et al., 2020) used random sampling to replace the farthest point sampling (FPS), which dramatically reduced the computational consumption and enhanced the handling ability for large-scale scenes. To compensate the uncertainty of random sampling, RandLA-Net (Hu et al., 2020) adopted a local feature aggregation module to increase the receptive field and enhance the local structure learning.

Since the Graph Neural Network (GNN) was developed by (Scarselli et al., 2009), it has been widely investigated for mining unstructured data. DGCNN (Wang et al., 2019a) built the directed graph in both the Euclidean space and the feature space, and dynamically updated the features layer-by-layer. GACNet (Wang et al., 2019b) introduced the attention mechanism into the graph-based methods, and learned the attention weights from local directed graph to achieve better representation of the local feature. GACNN (Wen et al., 2020) considered more comprehensive attentions to refine the feature representation, which included edge attention, density attention, and graph global attention.

In conclusion, most of the existing LC classification methods for airborne multispectral LiDAR data are classical machine learning-based methods. Previous works have proved the superiority of deep learning-based methods and point-based methods.

3. Study areas and datasets

As shown in Fig. 1, the study region we selected in this paper is situated in Whitchurch-Stouffville, Ontario, Canada. The location of the middle position is 43°58′00″ and 79°15′00″, respectively, with regard to the latitude and longitude. The study area covers a total area about 3.2 km^2. We also choose the same 13 typical scenes as in (Li et al., 2020b).

As for the datasets we used in this paper, we select the area_6 and area_7 from the 13 typical scenes as the test scenes, and the rest scenes as the training scenes. To meet the requirement of the LC classification task, we relabel the selected scenes into four classes: tree, building, grass, and road.

As shown in Fig. 2, the number of points collected by different channels is varying sharply. On the one hand, the wavelengths of the three channels show different reflections in land cover water. According to the official description released by Teledyne Optech company, water is best penetrated by the wavelength of channel 2 (532 nm), and completely absorbed by the other two wavelengths of channels (1064 nm and 1550 nm). On the other hand, aquatic plants are common in water area, which causes the extra interferential points collection in water area. Therefore, land cover water is not considered in this study.

Since the multispectral LiDAR data are mainly collected from residential areas, there are rarely exposed soils in these areas. We did some preliminary tests of adding these soil points, and found that the imbalanced distribution of the training data would severely impact the classification performance (Jing et al., 2021). Moreover, the soil points are

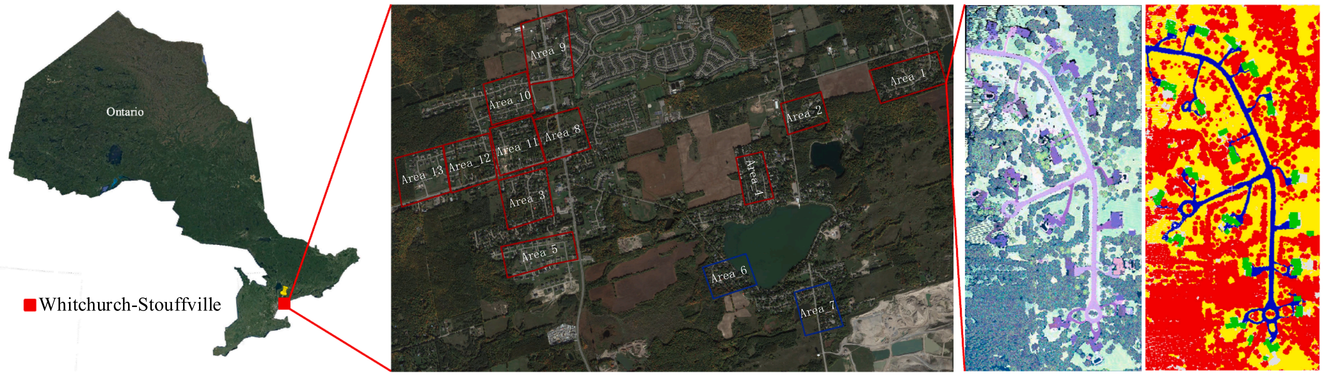


Fig. 1. Study area, selected scenes, preprocessed data, and corresponding labeled data.

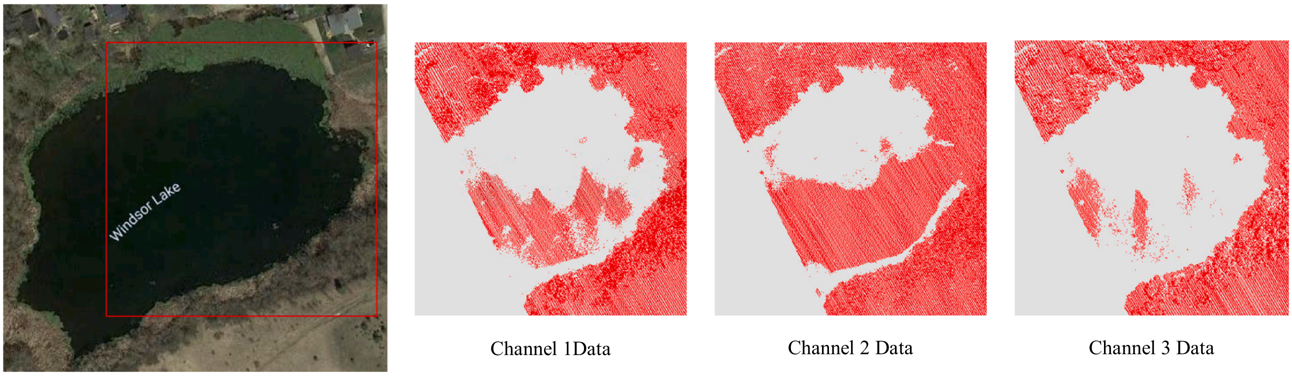


Fig. 2. Illustration of the water point clouds of three channels, respectively.

usually mixed with the grass points, which is difficult to manually label these two classes. Therefore, compared with (Pan et al., 2020), which evaluates the LC classification method with the same study area as ours, we only set four types of land covers here.

4. Methodology

4.1. Workflow overview

As shown in Fig. 3, our proposed LC classification workflow consists of several stages. The raw multispectral LiDAR data were acquired using the airborne Titan multispectral LiDAR system, which include three different channels. To fuse the point clouds individually collected by different channels, we implement the data preprocessing at the very first stage. Here, we use the same approach as that used in (Li et al., 2020b). Besides, we remove the outliers with abnormal heights, such as the points collected from the flying birds and deep holes. As the reasons we mentioned before, we also remove the points of water areas in this stage.

After data preprocessing, we could obtain the usable point cloud data, each point of which integrates three different intensity values from the three channels. To classify the point clouds, we manually label the fused data before feeding them into the model. Here, we label the point clouds into four classes: tree, grass, road, and building.

With the labeled data, we choose parts of the scenes as the training scenes, and the rest of the scenes as the test scenes. To guarantee the training effect of the model, we set the proportion of the training and test scenes close to quadruple. For each training and test scenes, we used the FPS-KNN sample generation method, which we proposed in (Li et al., 2020b), to generate the samples for meeting the input requirements of the point-wise deep learning-based model.

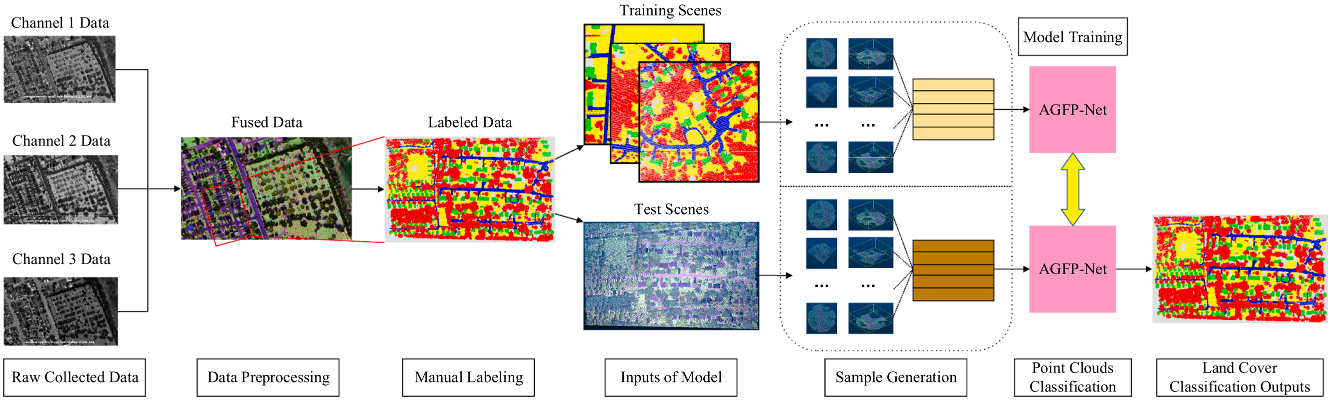


Fig. 3. Workflow of LC Classification.

Subsequently, we first train the model with the labeled training samples, and then classify the unlabeled test samples with the trained model. Here, we propose the Attentive Geometric Feature Pyramid Network (AGFP-Net) as the model used in the workflow. The details of the AGFP-Net are described in section 4.2.

Finally, the point-wise LC classification results could be obtained through the trained model, which means each point of the test scenes could be classified with an estimated label by the trained model.

4.2. Attentive geometric feature pyramid network

4.2.1. Attentive graph geometric moments convolution

1. The graph geometric moments representation of point clouds

In mathematics and statistics, moments are a set of measurements of the distribution and morphological characteristics of variables. Moments are commonly used to describe the image feature during image processing. The geometric moments are a kind of moments-based feature descriptor.

For a two-dimensional density distribution function $f(x,y)$, the moments function φ_{pq} with $p+q$ orders could be defined as (Ming-Kuei, 1962):

$$\phi_{pq} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \psi_{pg}(x,y)f(x,y)dxdy \tag{1}$$

where $\psi_{pq}(x,y)$ is the basis function.

When $\psi_{pq}(x,y) = x^p y^q$, there is the definition of geometric moments:

$$m_{pq} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^p y^q f(x,y)dxdy \tag{2}$$

where $p,q = 0,1,2,\ldots$.

Accordingly, in three-dimensional space, the $p+q+r$ orders geometric moments could be defined as (Yokoya and Levine, 1989):

$$m_{pqr} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^p y^q z^r f(x,y,z)dxdydz \tag{3}$$

where $p,q,r = 0,1,2,\ldots$.

For homogeneous objects in three-dimensional space, the discrete form of the $p+q+r$ orders geometric moments could be defined as (Liu and Tsai, 1990):

$$m_{pqr} = \sum_{\mathbb{R}^3} x^p y^q z^r f(x,y,z) \tag{4}$$

where $\mathbb{R}^3$ is the three-dimensional region in Euclidean space.

A point cloud is a set of points in space, and each point has exact position (coordinates), but without the volume or size. Referring to (Joseph-Rivlin et al., 2018), we define the geometric moments representation of point clouds as the set of $x^p y^q z^r$. For example, the 1 order geometric moments of a point cloud is

$$M_1 = [m_{100}\ m_{010}\ m_{001}] = [x^1y^0z^0\ x^0y^1z^0\ x^0y^0z^1] = [x\,y\,z] \tag{5}$$

To consider the relationship between a point and its neighbors, we construct the local directed graph like the previous graph-based methods DGCNN (Wang et al., 2019a). Coincidently, the directed edge from the nearest neighbors to the central point also has its geometric moments representation, which is the central geometric moments. The central geometric moments μ_{pqr} could be defined as (Tuceryan, 1994):

$$\mu_{pqr} = \sum_{R^3}(x-\bar{x})^p(y-\bar{y})^q(z-\bar{z})^r f(x,y,z) \tag{6}$$

where $(\bar{x},\bar{y},\bar{z})$ is the centroid of the object, which can be obtained from the lower order moments

$$\bar{x} = \frac{m_{100}}{m_{000}}\ \bar{y} = \frac{m_{100}}{m_{000}}\ \bar{z} = \frac{m_{100}}{m_{000}} \tag{7}$$

The different orders geometric moments of point clouds describe the geometry feature from different aspects, such as the 1 order geometric moments of point clouds describes the centroid of each point, which represents its center coordinates. Combining multiple orders geometric moments of point clouds can provide the more comprehensive input information, which could help the model to learn better geometry feature representations theoretically. Considering the objects in outdoor scenes having complex shapes, here, we adopt the combination of the first three orders geometric moments as the input of the proposed model.

2. Attention mechanisms

Attention mechanisms are the signal processing mechanisms that were discovered by scientists in the 1990 s while studying human vision. The researchers in the field of artificial intelligence introduced these mechanisms into the neural networks and achieved promising rewards. Recently, attention mechanisms have been one of the most popular modules widely applied in the deep learning field. Attention mechanisms, as an approach to improve neural networks, have achieved successes in the field of image processing (Fu et al., 2019), natural language processing (Lin et al., 2017a, 2017b), and graph network representation (Chen et al., 2019). The previous studies demonstrated that the attention mechanisms could enhance the feature representation ability of neural network, which inspires us to introduce the attention mechanisms into our previously proposed model.

The attention function is defined as follows (Vaswani et al., 2017):

$$Attention(Q,K,V) = soft\,max(\frac{QK^T}{\sqrt{d_k}})v \tag{8}$$

where matrices Q, K, V indicate a set of queries, keys, and values respectively, and $1/\sqrt{d_k}$indicates the scaling factor. Since the essence of attention mechanisms lies in learning the corresponding “attention” weight from the adjacent channels or features of the current feature to optimize the semantic representation of the current feature, many neural network models utilize the attention mechanisms as a strategy for feature aggregation. In our previous work (Li et al., 2020a), we adopted the average-pooling operation to aggregate the k nearest local neighbouring features into the central feature. Although we had demonstrated that the average-pooling operation is a better choice than the max-pooling operation, there still exists the issue of information loss caused by the average-pooling operation. Therefore, we introduce the attention mechanisms to learn the attention weight matrices from the k nearest local neighbouring features, and sum the weighted features to obtain the feature with the better semantic representation. The details of this process are described as follows:

Step 1: calculating the attention weight matrices. Given a set of local features $F_i = \left\{f_i^1 \cdots f_i^k \cdots f_i^K\right\}$, the corresponding attention weight of each feature is calculated by the function g. Generally, the function g consists of a Multi-Layer Perceptrons (MLP) and a softmax, which could be formulized as follows:

$$s_i^k = g(f_i^k, W) \tag{9}$$

where W is the weights contained in the MLP.

Step 2: summing the weighted features. The learned attention weights can be seen as a kind of filters or masks, which help the model to recognize more important or useful features. The attention weights s_i^kcalculated by the previous step are multiplied with the features f_i^k, and the aggregated feature is the sum of that, which could be formulized as follows:

$$\widetilde{f}_i = \sum_{k=1}^{k}(f_i^k \cdot s_i^k) \tag{10}$$

where $\widetilde{f}_i$is the aggregated feature.

4.2.2. Network architecture

Fig. 4 shows the detailed architecture of the AGFP-Net. (N,D) represent the number of points and the feature dimension, respectively. The schematic diagram of the input clouds is randomly selected from the test samples, and drew with the fake colors, which utilizes the normalized intensity values of the three channels as the RGB values. The AGGM represents the AGGM Convolution, which will be detailed in section 4.2.3. The FPS and FP represent the Farthest Point Sampling method and Feature Propagation operation, respectively. The MLP represents the Multi-Layer Perceptrons. The (N,1) in the last rectangle means that the model directly outputs the predicted label for each point, as shown in the schematic diagram of the output. The FU represents the feature up-sampling module, and the addition symbol means the addition operation.

The details of feature up-sampling module are described as follows. Firstly, we take the extracted features from different layers and the original input point cloud as the input to the module. These extracted features are attached to the corresponding point set, and the size of these point sets are smaller than the original input point cloud. Then, for each point in the original input point cloud, we find its three nearest neighbors in the corresponding point set of the input features. According to the distance between the point and its nearest neighbors, the weights of three nearest features can be calculated. The weighted three nearest features are summed into the points of the original input point cloud. Thus, we can obtain a feature having the same dimension as the input features and the same size as the original input point cloud. Finally, through an MLP and LeakyReLU layer, the dimension of feature is updated to the same as the feature output from the last decoder layer.

The features extracted from different layers represent the different-scale feature representations of the input point clouds, which contain the details in different scales. Therefore, we construct the feature pyramid with these multiple scales features. Unlike the most of feature pyramid architectures that only consider the features output from encoder layers, we also take the decoder layers into account to contain more comprehensive details. With the feature up-sampling module, these size-varying multiple scale features can be processed into the same size and same dimension. Finally, referring to (Lin et al., 2017a, 2017b), we utilize the addition operation to merge the up-sampled features and the feature output from the last decoder layer.

4.2.3. Architecture of the AGGM Convolution

Fig. 5 shows the detailed architecture of the AGGM Convolution. Given a (3 + d)-dimensional point cloud with N points as the input point features, the first three dimensions are the spatial coordinates of the points, and the next d dimensions are the additional features, such as the color, surface normal, and spectral value. The 3D coordinates and the additional features are represented as green and yellow rectangles in Fig. 5, respectively. Since the multispectral LiDAR data we used in this paper have three channels, for the very first inputs of the AGGM Convolution, the additional features are spectral intensity values of the three channels, the dimension d equals 3.

The following dotted arrow indicates splitting one point from the input N points, p_i represents the spatial coordinates of the current point, and f_i represents the corresponding additional features. For each point of the input N points, the local directed graph is generated based on the k-nearest neighbors (KNN) method by its spatial coordinates. Subsequently, the generated graph structure data would be split into three branches for further processing.

For the top branch in Fig. 5, the stacked K green rectangles represent the edges between the central node and its k-nearest neighbors. For the geometric moments representation calculation stage, we calculate the first three order geometric moments to obtain more detailed geometric structure features. The calculated first three order geometric moments of the edges have nineteen dimensions. Then, a shared MLP is implemented to extract the local geometric features, which sets the output dimension with $d_{out/2}$.

For the middle branch in Fig. 5, to meet the data format requirement of the following addition operation, the current point, which is also called central point, is duplicated K times to formulate the same shape as the edges in the top branch. Then, the geometric moments representation of the duplicated current point is calculated like the edges. Similarly, a shared MLP is also used to extract the geometric features, and the output dimension of the shared MLP is also set as $d_{out/2}$.

For the bottom branch in Fig. 5, the stacked K yellow rectangles are the additional features of the current point and its k-nearest neighboring points. The set of additional features f_i^K is fed into a shared MLP to obtain the features with a higher dimension, which has a more powerful semantic representation. The dimension of the output features of the shared MLP is also set as $d_{out/2}$.

The two extracted geometric features are aggregated by the addition operation, which is indicated with the addition symbol in Fig. 5. To fuse the aggregated geometric features and additional features, we utilize the concatenation operation to output the fused features with d_{out} dimensions. Using the concatenated features as the input to the attention module could sufficiently consider the attention influences of both the geometric features and the additional features, which is superior to individually calculate the attention weights. On the other hand, this strategy can also decrease the computational consumption.

As we mentioned above, there are two steps in the attention module. In the first step, the concatenated features are fed into a fully connected layer, and then normalized by a softmax layer. The output of the first step is the attention weight matrix, which is indicated with the K stacked blue rectangles in Fig. 5. In the second step, the attention weight matrix is multiplied the concatenated features. Then, the K weighted features with d_{out} dimensions are summed into one feature with d_{out} dimensions, which is the output of the attention module.

Finally, we append a shared MLP at the end of the AGGM Convolution to increase the robustness of the whole module. Then, the aggre-

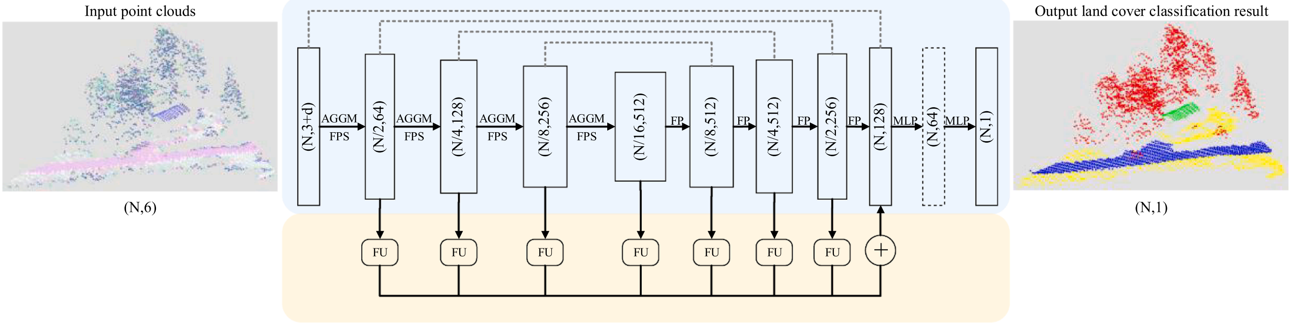


Fig. 4. Architecture of AGFP-Net.

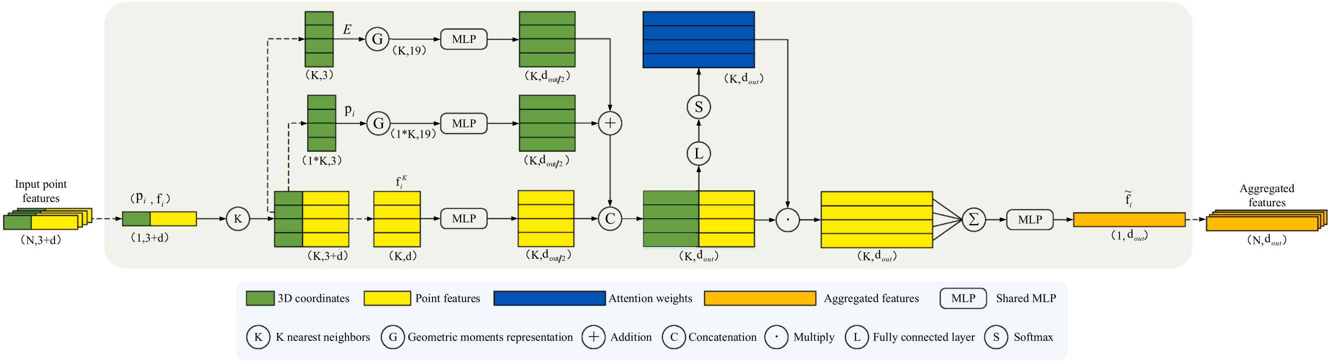


Fig. 5. Architecture of the AGGM Convolution.

gated feature $\widetilde{f}_i$, which is indicated with the orange rectangle in Fig. 5, is obtained by the AGGM Convolution module. With the same operation for each point in the input points, the AGGM Convolution can output the same N features with the specified dimension of d$_{out}$.

5. Results and discussions

5.1. Implementation details

We applied the same training strategy as in (Wang et al., 2019a). The loss function we used is the cross entropy. We chose the stochastic gradient descent (SGD) as the optimizer, whose initial learning rate was set with 0.01, and the learning rate declined thirty percent after each fifty iterations. The training iteration was set with 200. The batch size was set with 12. The momentum was set with 0.9. The LeakyReLU was adopted as the activation function, and the negative slope was set with 0.2. The batch normalization strategy was applied to each MLP layer. When the model was trained, we picked up the one with the best performance from the saved networks, and validated the test data with it. A NIVIDIA 2080 TI GPU was used to train the proposed model.

5.2. Accuracy evaluation metrics

To better evaluate the proposed AGFP-Net, we used four kinds of quantitative evaluation metrics, which are commonly used in LC classification tasks. The quantitative evaluation metrics include overall accuracy (OA), Kappa index, producer accuracy (PA), and user accuracy (UA) (Congalton, 1991; Foody, 2010).

Since the proposed model directly outputs the class label of each point in the test scenes, we validated the results by the point-based evaluation strategy, which directly used the number of the correctly and falsely classified points as the inputs for metrics calculation. Compared with the traditional pixel-based and object-based evaluations, the point-based evaluation is more precise and strict, which is more suitable for the proposed point-wise LC classification method.

5.3. Parameter sensitivity analysis

5.3.1. Ablation study

The GGM-Net (Li et al., 2020a) first proposed the graph geometric moments-based convolution operator. Since (Li et al., 2020a) already did the ablation study about the graph geometric moments convolution, batch normalization, and dropout technique, here, we directly adopted the GGM-Net as the baseline (model A in Table 1) of our ablation study. In Table 1, “#points” indicates sample size (i.e., the number of input points), “HA” indicates hierarchical architecture, “AM” indicates attention mechanism, “FP” indicates feature pyramid (including feature up-sampling module).

As shown in Table 1, the baseline only achieved the OA and Kappa index of 89.0% and 0.819. With the hierarchical architecture, model B was sharply improved to higher accuracies with an OA of 94.6% and a Kappa index of 0.913. By introducing the attention mechanism, the core convolution module was enhanced, model C was improved with respect to the OA and Kappa index by 1.4% and 0.023, respectively. Finally, with the feature up-sampling module, model D constructed a feature pyramid with details from multiple scale features and achieved the best OA and Kappa index of 96.9% and 0.950, respectively. Except the main metrics OA and Kappa, the PA and UA of the four land cover types also showed the similar tendency.

We found that the proposed feature up-sampling module and multiple scale feature pyramid construction could be applied as a universal operation. To further validate the function of that, we designed a comparison experiment by applying with and without feature pyramid operation. We picked up the classical hierarchical model, PointNet++,

Table 2
Results of PointNet++ with and without feature pyramid construction.

Model		Road	Grass	Tree	Building	OA (%)	Kappa
PointNet++	PA	74.4	86.9	**94.2**	66.7	88.3	0.811
	UA	**77.0**	91.1	93.5	51.1		
PointNet++ with feature pyramid	PA	**83.8**	**87.6**	93.9	**74.2**	**89.9**	**0.834**
	UA	70.4	**92.8**	**96.0**	**58.0**		

Table 1
Ablation study of AGFP-Net.

Model	#points	HA	AM	FP		Building	Tree	Grass	Road	OA(%)	Kappa
A	4096				PA	76.0	91.6	89.2	79.3	89.0	0.819
					UA	49.0	96.7	89.4	74.9		
B	4096	√			PA	89.6	98.3	90.0	91.1	94.6	0.913
					UA	88.2	96.7	96.3	82.7		
C	4096	√	√		PA	93.6	98.5	93.2	93.9	96.0	0.936
					UA	90.3	98.5	97.3	83.5		
D	4096	√	√	√	PA	**96.9**	**98.8**	**94.4**	**95.3**	**96.9**	**0.950**
					UA	**92.2**	**99.4**	**97.8**	**84.8**		

as the model used for comparison. The training sample size was set as 4096. As shown in Table 2, the model "PointNet++ with feature pyramid" achieved better performance in all metrics, which further validated the effectiveness of the proposed feature up-sampling module and multiple scale feature pyramid construction operation.

5.3.2. Spectral information

Since the different types of land covers have different reflectance intensities with the specific wavelengths, the spectral information collected from different wavelengths (or channels) shows different effects for classifying different types of land covers. To fully assess the function of the spectral information collected from different channels, we designed five comparison experiments with different spectral inputs. Since the spatial coordinates are essential for the proposed model, as shown in Table 3, the inputs of the five designed comparison experiments include the spatial coordinates, spatial coordinates and channel 1 spectral values, spatial coordinates and channel 2 spectral values, spatial coordinates and channel 3 spectral values, spatial coordinates and spectral values from all channels.

The training sample size was set as 4096. Two test scenes, area_6 and area_7, were generated into 257 and 474 samples by the FPS-KNN sampling method, respectively. During the sample generation stage, some points might be allocated in two or three different samples. For these points in the overlapped part, we counted the predicted labels from every involved samples, and chose the most counted predicted label as its final predicted label. Finally, the four accuracy metrics were calculated with the point-wise classification results.

As shown in Table 3, compared with group 1, which only used the spatial coordinates as the input, with the additional spectral information, the LC classification accuracies were improved in different degrees. Using the spatial coordinates and all spectral values as the input achieved the highest OA and Kappa index.

Specifically, by adding the channel 1 spectral values (group 2), there was an effective improvement on the classification accuracy for Road, the PA and UA of Road were improved by 11.9 and 6.1 percentage points, respectively. For the other metrics, there was only a slight improvement. By adding the channel 2 spectral values (group 3), except the Tree and Building, the PA and UA of all the other land cover types had a significant improvement, the OA and Kappa index were also sharply improved by 7.2 and 0.12, respectively. By adding the channel 3 spectral values (group 4), there was a similar improvement trend as group 3, but the more moderate one, the OA and Kappa index were only improved by 3.9 and 0.066, respectively.

Through the comparison experiments, it is verified that adding the additional spectral information can enhance the accuracy of poiny-wise LC classification, and the function of the spectral information collected from different channels show varying effects on different land cover types.

Table 3
Results of AGFP-Net by training with different spectral inputs.

Input		Road	Grass	Tree	Building	OA (%)	Kappa
Spatial coordinates	PA	61.5	78.2	98.6	96.5	89.2	0.822
	UA	25.8	93.1	99.5	89.7		
Spatial coordinates and channel 1 spectral values	PA	73.4	79.8	98.6	**97.3**	90.3	0.840
	UA	31.9	95.3	99.2	89.4		
Spatial coordinates and channel 2 spectral values	PA	94.6	93.9	98.6	94.6	96.4	0.942
	UA	**85.2**	97.3	98.8	91.0		
Spatial coordinates and channel 3 spectral values	PA	87.5	85.7	98.8	94.4	93.1	0.888
	UA	60.1	95.9	99.0	85.9		
Spatial coordinates and all spectral values	PA	**95.3**	**94.4**	**98.8**	96.9	**96.9**	**0.950**
	UA	84.8	**97.8**	**99.4**	**92.2**		

5.3.3. Sample size

The point-wise deep learning-based methods require the fixed sample size, therefore, we tested the performance of the model with different training sample sizes, and determined the sample size according to the comparison results. The different size training samples provide different semantic information and geometric continuity of objects in the scene, which are both critical for the deep learning-based methods. Subconsciously, in terms of sample size, we think the larger the better. Duo to GPU memory limitations, we set the maximum sample size to 4096 and the other two comparisons to 2048 and 1024. All the comparison experiments used the same input features.

As shown in Table 4, the OA and Kappa index were gradually increasing with the enlargement of the sample size. The PA and UA showed the same trend on the four land cover types. For the main evaluation metrics, OA and Kappa index, the accuracies of 4096 sample size was higher than those of 2048 and 1024 sample size by 1.0 and 0.017, 1.8 and 0.028, respectively. As for the PA and UA, the overall trend was also gradually increasing with the enlargement of the sample size.

The results of comparison experiments validated our speculation, that is the larger sample size contributes to the better LC classification accuracies. Besides the FPS sample method, we also considered the random sample method, which used in RandLA-Net (Hu et al., 2020), to reduce the memory consumption. We found that the increase of the sample size did not contribute to the improvement of the classification accuracies obtained by the proposed model and RandLA-Net. The main reasons might be the following aspects. (1) The random sample method might not guarantee the sampling performance as that of the FPS sample method. (2) The data used in the RandLA-Net (Hu et al., 2020) were the point clouds collected from mobile laser scanning (MLS) or terrestrial laser scanning (TLS) systems, which have different coverage and characteristics from airborne multispectral LiDAR data. (3) Compared with the general semantic segmentation task, the land cover classification task places emphasis on classifying the land cover types within a certain area.

5.4. Comparative studies

Since there is still no existing point-wise deep learning-based LC classification method for airborne multispectral LiDAR data, to better assess the performance of our AGFP-Net, we chose the popularly used state-of-the-art networks, which designed for semantic segmentation on point clouds, as the comparison methods. The comparison methods include PointNet (Qi et al., 2017a), PointNet++ (Qi et al., 2017b), DGCNN (Wang et al., 2019a), RS-CNN (Liu et al., 2019), GACNet (Wang et al., 2019b), and RandLA-Net (Hu et al., 2020).

As seen in Table 5, the AGFP-Net achieved the highest accuracies on all metrics and all classes. For the main evaluation metrics, OA and Kappa index, the proposed AGFP-Net achieved even 2.1 and 3.4 percentage points higher than those of the second-highest method, RS-CNN, no mention the other comparison methods. For the other metrics, PA

Table 4
Results of AGFP-Net by training with different sample sizes.

Sample size		Road	Grass	Tree	Building	OA(%)	Kappa
1024	PA	94.7	91.2	98.3	91.0	95.1	0.922
	UA	80.5	97.0	97.6	89.7		
2048	PA	95.3	92.1	98.5	95.4	95.9	0.933
	UA	80.5	97.4	98.7	91.0		
4096	PA	**95.3**	**94.4**	**98.8**	**96.9**	**96.9**	**0.950**
	UA	**84.8**	**97.8**	**99.4**	**92.2**		

Table 5
Results of comparison methods.

Model		Road	Grass	Tree	Building	OA (%)	Kappa
PointNet (Qi et al., 2017a)	PA	74.2	79.4	90.7	63.8	84.3	0.741
	UA	58.0	89.3	92.1	39.6		
PointNet++ (Qi et al., 2017b)	PA	74.4	86.9	94.2	66.7	88.3	0.811
	UA	77.0	91.1	93.5	51.1		
DGCNN (Wang et al., 2019a)	PA	88.3	89.1	94.5	83.8	91.6	0.862
	UA	74.2	94.0	97.2	62.9		
RS-CNN (Liu et al., 2019)	PA	91.5	91.4	97.6	93.0	94.7	0.914
	UA	81.0	96.7	97.9	81.5		
GACNet (Wang et al., 2019b)	PA	83.4	83.4	91.8	71.0	87.4	0.792
	UA	62.3	92.3	94.9	44.5		
RandLA-Net (Hu et al., 2020)	PA	86.0	90.5	96.1	82.7	92.5	0.878
	UA	80.9	94.0	96.2	75.0		
AGFP-Net	PA	**95.3**	**94.4**	**98.8**	**96.9**	**96.9**	**0.950**
	UA	**84.8**	**97.8**	**99.4**	**92.2**		

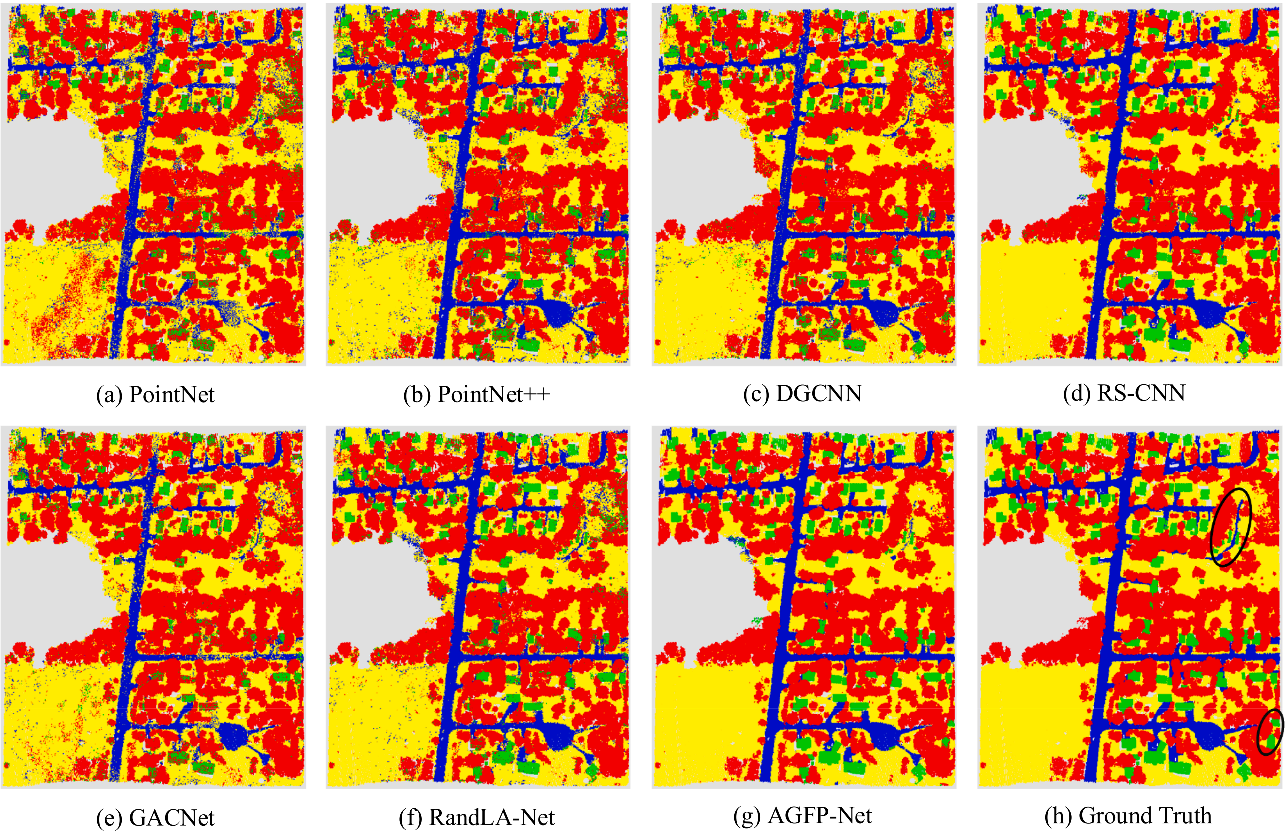
and UA, the AGFP-Net achieved a dramatic improvement on land covers Road and Building, which are 4.0 and 2.7, 2.6 and 11.7 percentage points higher than those of the second-highest method respectively. Some comparison methods misclassified some Road points as the Grass ones, because the two land covers have the similar geometric shapes. The Building points have the similar heights with the Tree points. There are the reasons why many comparison methods achieved low accuracies on Road and Building. The proposed AGFP-Net well overcomes these shortages, and shows the powerful ability of geometric feature extraction. The results of the comparison experiments demonstrated the superiority of the proposed AGFP-Net.

To have a better visualization effect, we chose the area_7 from the two test scenes as the visualization scene, which has more points and well-distributed land cover distribution than area_6. The four types of land covers, Road, Grass, Tree, and Building, are represented with blue, yellow, red, and green in Figs. 6, 7, and 8, respectively. Fig. 6 illustrates the visualization of the classification results of the comparison methods and the ground truth, the two black circles in (h) indicate the region of detailed visualization in Figs. 7 and 8 respectively.

As shown in Fig. 6, the result of the proposed AGFP-Net shows high consistency with the Ground Truth, which achieved less misclassification outlier points and better completeness of the LC classification. The RS-CNN has the most closing visualization effect with the proposed model among all the comparison methods. To better compare the classification performance, we picked two typical sub-scenes in area_7 to illustrate the differences among different methods.

As seen in Fig. 7, our AGFP-Net misclassified part of the road points as the grass points, because the road points in this region have the similar heights to the surrounding grass points. Although some of the comparison methods recognized more road points, their classification accuracies of the other points in this region are much lower than the proposed model, and their misclassified points are disorganized. Relatively, the proposed AGFP-Net and RS-CNN achieved better classification results on the whole, which have the neater classification result and higher consistency with the Ground Truth. By the close-up observation, it could be found that, compared with the RS-CNN, the proposed model obtained better classification results at the edge points between different land cover types. For example, the RS-CNN misclassified the edge points of several buildings as the tree points in this region, which demonstrated the features extracted by the proposed model have better interclass separability.

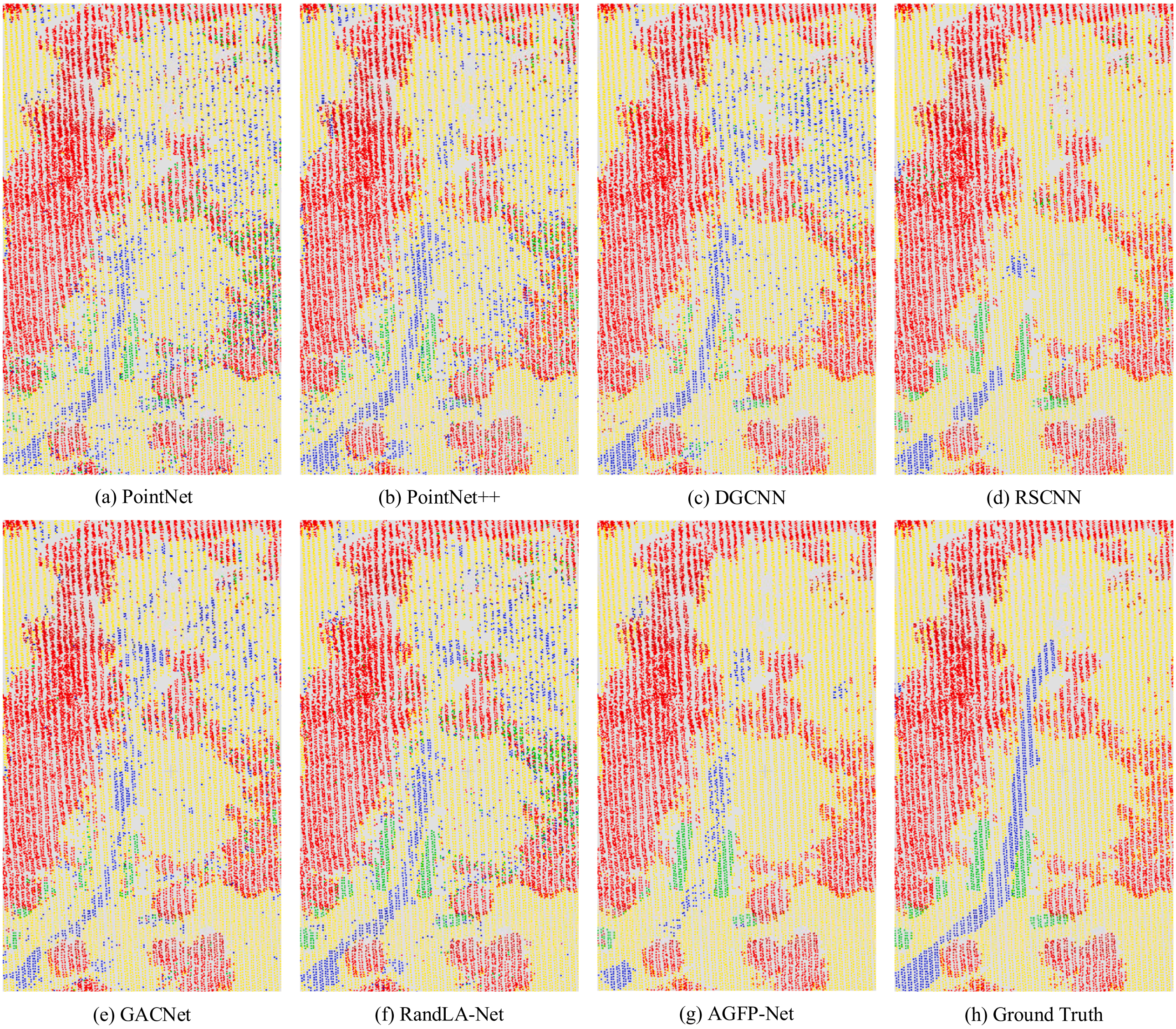
As shown in Fig. 8, with the side view, it could be observed that there are three buildings surrounded by the tall trees in this sub-scene, and some part of the buildings and trees are sheltered or overlapped. For this complex environment, it is difficult to classify the land cover correctly. Among the comparison methods, only DGCNN correctly recognized the part of the building points, but still misclassified more than half of the building points. For the rest methods, there were only few fragmentary



(a) PointNet (b) PointNet++ (c) DGCNN (d) RS-CNN

(e) GACNet (f) RandLA-Net (g) AGFP-Net (h) Ground Truth

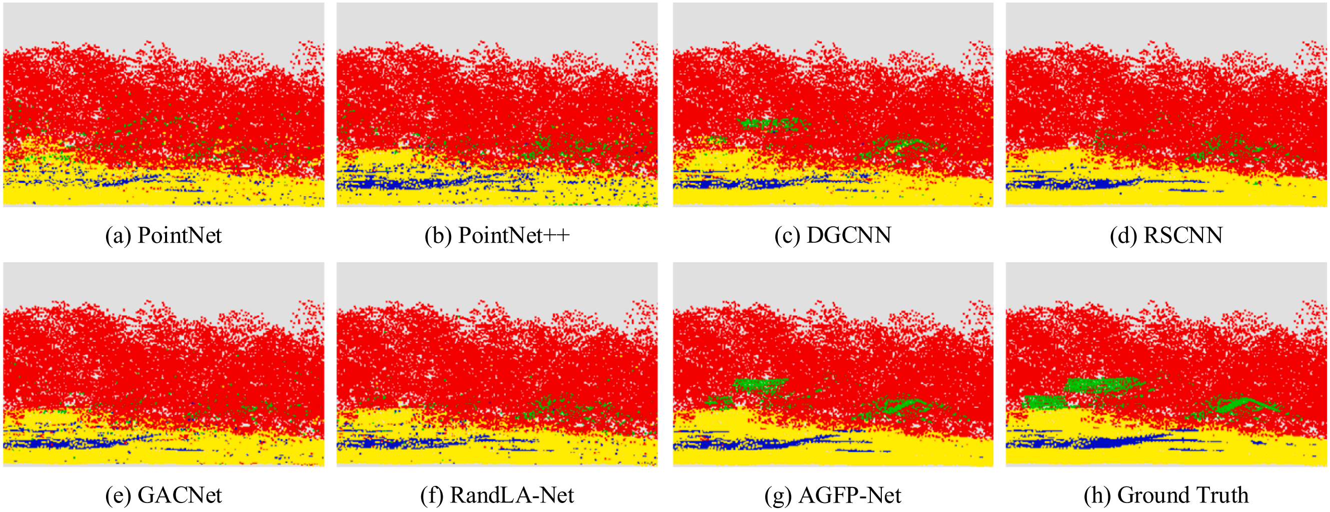
Fig. 6. Visualization of the comparison methods.



(a) PointNet (b) PointNet++ (c) DGCNN (d) RSCNN

(e) GACNet (f) RandLA-Net (g) AGFP-Net (h) Ground Truth

Fig. 7. Detailed visualization of the comparison methods.



(a) PointNet (b) PointNet++ (c) DGCNN (d) RSCNN

(e) GACNet (f) RandLA-Net (g) AGFP-Net (h) Ground Truth

Fig. 8. Detailed visualization of the comparison methods.

building points correctly classified, or even failed to be correctly classified at all. The proposed AGFP-Net correctly classified most of the building points, which achieved obviously better classification result. Since the building has obvious geometric structure characteristics, the detailed visualization of this sub-scene further validated the high efficiency of the proposed model in terms of the spatial geometric structure extraction.

Although the proposed method achieved the promising LC classification accuracies, there is still room for improvement. Firstly, the spectral information could be further mined. In this paper, we mainly focused on the improvement of local geometric feature extraction and feature aggregation, and thus extracted spectral features just by regular linear or MLP operations. Theoretically, the spectral information might be comprehensively explored by a customized feature extractor, contributing for improving LC classification accuracies. Secondly, limited by the sample size, the data processing ability is inadequate for point clouds with a large-scale scene. As we mentioned above, we, to enlarge the sample size, replaced our FPS sample method with the random sample method. However, we found that the sample size used by the random sample method led to a degradation of classification accuracies. Therefore, the more efficient sample method is needed to be developed to trade off the data processing ability of large-scale point clouds and high LC classification accuracy.

6. Conclusion

In this paper, we proposed a point-wise AGFP-Net LC classification method, which only uses the raw multispectral LiDAR point clouds as the input, and directly outputs the predicted label for each point. Correspondingly, we adopt the point-based evaluation instead of the traditionally used pixel-based or object-based evaluation, which is more precise, strict, and suitable for the proposed point-wise method. By introducing the attention mechanisms, we design a novel convolution operator, called AGGM Convolution, which extracts and aggregates the geometry features effectively. Moreover, we propose a feature up-sampling module to up-sample the features extracted from layers, and construct a feature pyramid with multiple scales to merge more comprehensive details. Experiments validated the feasibility and effectiveness of the proposed method. Visual inspections and quantitative evaluations showed that the proposed method is superior for LC classification from airborne multispectral LiDAR point clouds. In addition, to investigate the potential of airborne multispectral LiDAR data, we validated the function of different wavelength spectral values by the comparison experiments of LC classification, which could provide the reference for the related researches and applications. Although the proposed method achieved quite high accuracies on four metrics, an obvious limitation is that the collected multispectral LiDAR data only has three different wavelength channels, which is far less than the channels of hyper-spectral images. In the next future, the registered hyper-spectral imagery can be fused to generate the hyper-spectral point clouds, and the proposed method will be further tested with these datasets. Besides, how to handle the large scale scenes and maintain the high accuracy at the same time is also one of our future works.

CRediT authorship contribution statement

Dilong Li: Conceptualization, Software, Methodology, Writing – original draft, Writing – review & editing. **Xin Shen:** Validation, Formal analysis, Investigation, Supervision. **Haiyan Guan:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Project administration, Funding acquisition. **Yongtao Yu:** Data curation, Writing – review & editing, Funding acquisition. **Hanyun Wang:** Investigation, Writing – review & editing. **Guo Zhang:** Writing – review & editing. **Jonathan Li:** Resources, Supervision, Writing – review & editing. **Deren Li:** Supervision.

Declaration of Competing Interest

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

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grants 62076107 and 41971414, in part by the Natural Science Foundation of Fujian Province under Grant 2021J05059, in part by the Open Research Fund of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University under Grant 21E01, and in part by the Scientific Research Funds of Huaqiao University under Grant 605-50Y21026. The authors would like to acknowledge the anonymous reviewers for their valuable comments in improving the paper.

References

Antonarakis, A.S., Richards, K.S., Brasington, J., Bithell, M., Muller, E., 2008. Retrieval of Vegetative Fluid Resistance Terms for Rigid Stems Using Airborne Lidar. J. Geophys. Res.-Biogeo. 113 (G2), n/a–n/a.

Bakuła, K., Kupidura, P., Jełowicki, Ł., 2016. Testing of Land Cover Classification from Multispectral Airborne Laser Scanning Data. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLI-B7, 161–169.

Chen, Y.P., Rohrbach, M., Yan, Z.C., Yan, S.C., Feng, J.S., Kalantidis, Y., 2019. Graph-based Global Reasoning Networks. Proc. Cvpr Ieee 433–442.

Congalton, R.G., 1991. A Review of Assessing The Accuracy of Classifications of Remotely Sensed Data. Remote Sens. Environ. 37 (1), 35–46.

Dare, P.M., 2005. Shadow Analysis in High-resolution Satellite Imagery of Urban Areas. Photogramm. Eng. Rem. S. 71 (2), 169–177.

Ekhtari, N., Glennie, C., Fernandez-Diaz, J.C., 2018. Classification of Airborne Multispectral Lidar Point Clouds for Land Cover Mapping. Ieee J.-Stars 11 (6), 2068–2078.

Feddema, J.J., Oleson, K.W., Bonan, G.B., Mearns, L.O., Buja, L.E., Meehl, G.A., Washington, W.M., 2005. The Importance of Land-cover Change in Simulating Future Climates. Science 310 (5754), 1674–1678.

Fernandez-Diaz, J.C., Carter, W.E., Glennie, C., Shrestha, R.L., Pan, Z., Ekhtari, N., Singhania, A., Hauser, D., Sartori, M., 2016. Capability Assessment and Performance Metrics for the Titan Multispectral Mapping Lidar. Remote Sens.-Basel 8.

Foody, G.M., 2010. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. Photogram. Rec. 25 (130), 204–205.

Fu, J., Liu, J., Tian, H.J., Li, Y., Bao, Y.J., Fang, Z.W., Lu, H.Q., 2019. Dual Attention Network for Scene Segmentation. Proc. Cvpr Ieee 3141–3149.

Ghaseminik, F., Aghamohammadi, H., Azadbakht, M., 2021. Land Cover Mapping of Urban Environments Using Multispectral LiDAR Data under Data Imbalance. Remote Sens. Appl.: Soc. Environ. 21, 100449. https://doi.org/10.1016/j.rsase.2020.100449.

Glennie, C.L., Carter, W.E., Shrestha, R.L., Dietrich, W.E., 2013. Geodetic Imaging with Airborne LiDAR: the Earth's surface revealed. Rep. Prog. Phys. 76 (8), 086801. https://doi.org/10.1088/0034-4885/76/8/086801.

Gong, W., Sun, J., Shi, S., Yang, J., Du, L., Zhu, B.o., Song, S., 2015. Investigating the Potential of Using the Spatial and Spectral Information of Multispectral LiDAR for Object Classification. Sensors-Basel 15 (9), 21989–22002.

Hanna, E., 2007. Radiative Forcing of Climate Change: expanding the concept and addressing uncertainties. By the National Research Council (NRC). National Academies Press, Washington DC, USA, 2005. 207 pp. Paperback. Weather 62 (4), 109-109.

Hong, D., Gao, L., Yokoya, N., Yao, J., Chanussot, J., Du, Q., Zhang, B., 2021a. More Diverse Means Better: Multimodal Deep Learning Meets Remote-Sensing Imagery Classification. IEEE Trans. Geosci. Remote Sens. 59, 4340–4354.

Hong, D., Gao, L., Yao, J., Zhang, B., Plaza, A., Chanussot, J., 2021b. Graph Convolutional Networks for Hyperspectral Image Classification. IEEE Trans. Geosci. Remote Sens. 59 (7), 5966–5978.

Hong, D., Han, Z., Yao, J., Gao, L., Zhang, B., Plaza, A., Chanussot, J., 2021c. SpectralFormer: Rethinking Hyperspectral Image Classification with Transformers. IEEE Trans. Geosci. Remote Sens. 1–1.

Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N., Markham, A., 2020. In: RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds, pp. 11105–11114.

Jing, Z., Guan, H., Zhao, P., Li, D., Yu, Y., Zang, Y., Wang, H., Li, J., 2021. Multispectral LiDAR Point Cloud Classification Using SE-PointNet++. Remote Sens. 13 (13), 2516. https://doi.org/10.3390/rs13132516.

Joseph-Rivlin, M., Zvirin, A., Kimmel, R., 2018. Mo-Net: Flavor the Moments in Learning to Classify Shapes.

Kim, Y., Kim, Y., 2014. Improved Classification Accuracy Based on the Output-Level Fusion of High-Resolution Satellite Images and Airborne LiDAR Data in Urban Area. Ieee Geosci. Remote S. 11, 636–640.

Li, D.L., Shen, X., Yu, Y.T., Guan, H.Y., Wang, H.Y., Li, D.R., 2020a. GGM-Net: Graph Geometric Moments Convolution Neural Network for Point Cloud Shape Classification. IEEE Access 8, 124989–124998.

Li, D., Shen, X., Yu, Y., Guan, H., Li, J., Zhang, G., Li, D., 2020b. Building Extraction from Airborne Multi-Spectral LiDAR Point Clouds Based on Graph Geometric Moments Convolutional Neural Networks. Remote Sens.-Basel 12 (19), 3186. https://doi.org/10.3390/rs12193186.

Lin, T., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017a. Feature Pyramid Networks for Object Detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 936–944.

Liu, S.-T., Tsai, W.-H., 1990. Moment-preserving corner detection. Pattern Recognition 23, 441–460.

Lin, Z., Feng, M., Santos, C., Yu, M., Xiang, B., Zhou, B., Bengio, Y., 2017. A Structured Self-attentive Sentence Embedding.

Liu, Y.C., Fan, B., Xiang, S.M., Pan, C.H., 2019. Relation-Shape Convolutional Neural Network for Point Cloud Analysis. In: 2019 Ieee/Cvf Conference on Computer Vision and Pattern Recognition (Cvpr 2019), pp. 8887–8896.

Lodha, S.K., Kreps, E.J., Helmbold, D.P., Fitzpatrick, D.N., 2007. Aerial LiDAR Data Classification Using Support Vector Machines (SVM). International Symposium on 3d Data Processing.

Lunetta, R.S., Ediriwickrema, J., Johnson, D.M., Lyon, J.G., McKerrow, A., 2002. Impacts of Vegetation Dynamics on the Identification of Land-cover Change in a Biologically Complex Community in North Carolina, USA. Remote Sens. Environ. 82 (2-3), 258–270.

Mallet, C., Bretar, F., Soergel, U., 2008. Analysis of Full-Waveform Lidar Data for Classification of Urban Areas. Photogramm. Fernerkun 337–349.

Matikainen, L., Hyyppä, J., Litkey, P., 2016. Multispectral Airborne Laser Scanning for Automated Map Updating. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLI-B3, 323–330.

Matikainen, L., Karila, K., Hyyppä, J., Puttonen, E., Litkey, P., Ahokas, E., 2017a. Feasibility of Multispectral Airborne Laser Scanning for Land Cover Classification, Road Mapping and Map Updating. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLII-3/W3, 119–122.

Matikainen, L., Karila, K., Hyyppä, J., Litkey, P., Puttonen, E., Ahokas, E., 2017b. Object-based Analysis of Multispectral Airborne Laser Scanner Data for Land Cover Classification and Map Updating. Isprs J. Photogramm. 128, 298–313.

Ming-Kuei, H., 1962. Visual Pattern Recognition by Moment Invariants. IRE Trans. Inf. Theory 8 (2), 179–187.

Morsy, S., Shaker, A., El-Rabbany, A., 2017. Multispectral LiDAR Data for Land Cover Classification of Urban Areas. Sensors 17 (5), 958. https://doi.org/10.3390/s17050958.

Morsy, S., Shaker, A., El-Rabbany, A., 2017. Clustering of Multispectral Airborne Laser Scanning Data Using Gaussian Decomposition. In: Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLII-2/W7, pp. 269–276.

Pan, S., Guan, H., Chen, Y., Yu, Y., Nunes Gonçalves, W., Marcato Junior, J., Li, J., 2020. Land-cover Classification of Multispectral LiDAR Data using CNN with Optimized Hyper-parameters. Isprs J. Photogramm. 166, 241–254.

Qi, C.R., Su, H., Mo, K.C., Guibas, L.J., 2017. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. Proc. Cvpr Ieee 77–85.

Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017b. PointNet ++ : Deep Hierarchical Feature Learning on Point Sets in a Metric Space. Adv. Neur. In. 30.

Scarselli, F., Gori, M., Ah Chung Tsoi, Hagenbuchner, M., Monfardini, G., 2009. The Graph Neural Network Model. IEEE Trans. Neural Netw. 20 (1), 61–80.

Teo, T.-A., Wu, H.-M., 2017. Analysis of Land Cover Classification Using Multi-Wavelength LiDAR System. Appl. Sci. 7 (7), 663. https://doi.org/10.3390/app7070663.

Tuceryan, M., 1994. Moment-based texture segmentation. Pattern Recognition Letters 15, 659–668.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention Is All You Need. Adv. Neur. In. 30.

Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M., 2019. Dynamic Graph CNN for Learning on Point Clouds. ACM T. Graph. 38 (5), 1–12.

Wang, L., Huang, Y.C., Hou, Y.L., Zhang, S.M., Shan, J., 2019. Graph Attention Convolution for Point Cloud Semantic Segmentation. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (Cvpr 2019), pp. 10288–10297.

Wang, Q., Gu, Y., 2020. A Discriminative Tensor Representation Model for Feature Extraction and Classification of Multispectral LiDAR Data. Ieee T. Geosci. Remote 58 (3), 1568–1586.

Wen, C., Li, X., Yao, X., Peng, L., Chi, T., 2020. Airborne LiDAR Point Cloud Classification with Graph Attention Convolution Neural Network. ArXiv abs/2004.09057.

Wichmann, V., Bremer, M., Lindenberger, J., Rutzinger, M., Georges, C., Petrini-Monteferri, F., 2015. Evaluating the Potential of Multispectral Airborne LiDAR for Topographic Mapping and Land Cover Classification. ISPRS Ann. Photogramm. Remote Sens. Spatial. Inf. Sci. II-3/W5, 113–119.

Wilkinson, G.G., 2005. Results and Implications of A Study of Fifteen Years of Satellite Image Classification Experiments. Ieee T. Geosci. Remote 43 (3), 433–440.

Yan, W.Y., Shaker, A., Habib, A., Kersting, A.P., 2012. Improving Classification Accuracy of Airborne LiDAR Intensity Data by Geometric Calibration and Radiometric Correction. Isprs J. Photogramm. 67, 35–44.

Yokoya, N., Levine, M.D., 1989. Range image segmentation based on differential geometry: a hybrid approach. IEEE Transactions on Pattern Analysis and Machine Intelligence 11, 643–649.

Zhou, W., Huang, G., Troy, A., Cadenasso, M.L., 2009. Object-based Land Cover Classification of Shaded Areas in High Spatial Resolution Imagery of Urban Areas: A Comparison Study. Remote Sens. Environ. 113 (8), 1769–1777.

Zou, X., Zhao, G., Li, J., Yang, Y., Fang, Y., 2016. 3D Land Cover Classification based on Multispectral LiDAR Point Clouds. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLI-B1, 741–747.