A Performance Analysis of Point CNN and Mask R-CNN for Building Extraction from Multispectral LiDAR Data

Asmaa A. Mandouh[1], Mahmoud El Nokrashy O. Ali[2], Mostafa H.A. Mohamed[3],
Lamyaa Gamal EL-Deen Taha[4], Sayed A. Mohamed[5]
National Authority for Remote Sensing and Space Sciences, Cairo, Egypt[1, 4, 5]
Faculty of Engineering-Al–Azhar University, Cairo, Egypt[2, 3]

***Abstract*—The extraction of buildings from multispectral Light Detection and Ranging (LiDAR) data holds significance in various domains such as urban planning, disaster response, and environmental monitoring. State-of-the-art deep learning models, including Point Convolutional Neural Network (Point CNN) and Mask Region-based Convolutional Neural Network (Mask R-CNN), have effectively addressed this particular task. Data and application characteristics affect model performance. This research compares multispectral LiDAR building extraction models, Point CNN and Mask R-CNN. Models are tested for accuracy, efficiency, and capacity to handle irregularly spaced point clouds using multispectral LiDAR data. Point CNN extracts buildings from multispectral LiDAR data more accurately and efficiently than Mask R-CNN. CNN-based point cloud feature extraction avoids preprocessing like voxelization, improving accuracy and processing speed over Mask R-CNN. CNNs can handle LiDAR point clouds with variable spacing. Mask R-CNN outperforms Point CNN in some cases. Mask R-CNN uses image-like data instead of point clouds, making it better at detecting and categorizing objects from different angles. The study emphasizes selecting the right deep learning model for building extraction from multispectral LiDAR data. Point CNN or Mask R-CNN for accurate building extraction depends on the application. For building extraction from multispectral LiDAR data, two approaches were compared utilizing precision, recall, and F1 score. The point-CNN model outperformed Mask R-CNN. The point-CNN model had 93.40% precision, 92.34% recall, and 92.72% F1 score. Mask R-CNN has moderate precision, recall, and F1.**

Keywords—Multispectral LiDAR; Mask R-CNN; Point CNN; deep learning; building extraction

I. INTRODUCTION

The escalating urbanization of the global population necessitates the development of accurate and efficient techniques for extracting buildings from remote sensing data. The extraction of buildings from remotely sensed data is a crucial procedure with wide-ranging applications, including but not limited to three-dimensional (3D) building modeling, urban planning, disaster assessment, and the maintenance of digital maps and Geographic Information System (GIS) databases [1].

The task of accurately and efficiently identifying buildings from remote sensing data presents several challenges, due to data availability issues, poor data quality, and obstructions caused by nearby objects like trees, automobiles, and mountains [2]. Despite these difficulties, advancement has been made significant in the recent development of building extraction techniques. Building extraction accuracy and effectiveness are projected to increase over time as deep learning algorithms advance and more high-quality remote sensing data become accessible [3].

The multi-spectral Light Detection and Ranging (LiDAR) provides a field for obtaining different spectral responses from different features and collecting various data about the surface and terrain of the land and water [4]. Due to this rationale, the utilization of multi-spectral LiDAR has significantly advanced the field of remote sensing data due to its vast quantity of high-resolution multispectral and spatial data [5]. However, this abundance of data may present a challenge in terms of the human capacity to accurately extract and classify features from the point cloud. Consequently, the rapid development of computer technology and the emergence of artificial intelligence, including machine learning and deep learning, have made it possible to reduce the time and human effort required for precise feature extraction from LiDAR sensors´ point clouds [6, 7]. Automatic extraction of buildings from multispectral LiDAR data is a challenging task, but one that has the potential to be extremely useful for a wide range of applications [7].

The objective of this study is to compare and contrast two distinct methodologies for extracting buildings from multispectral LiDAR data. The first utilizes the deep learning algorithm Mask R-CNN, while the second utilizes Point CNN. Both methods utilize three multispectral LiDAR channels to optimize building extraction.

The paper conforms to this structure. Section II describes the related works. Section III describes the significance of the research. The data and the study area are in Section IV. The Section V describes the methodology. Section VI discusses accuracy assessment mathematically. Section VII provides qualitative and quantitative evaluations of the findings. The discussion and summary concluded in Section VIII.

II. RELATED WORK

The utilization of LiDAR technology offers a significant advantage in terms of three-dimensional spatial accuracy [8], rendering it an optimal choice for various remote sensing applications, particularly in the mapping of densely populated urban regions. Multiple scientific studies have provided evidence supporting the utilization of LiDAR data for the extraction of buildings in urban environments [9, 10].

Building extraction from LiDAR data has been extensively researched, resulting in the development of numerous algorithms in recent years. Nevertheless, the majority of these techniques rely on LiDAR data with a single wavelength. There exists a limited body of research pertaining to the utilization of multispectral LiDAR data for the purpose of building extraction. These studies demonstrate that deep learning can be used to accurately extract buildings from single-wavelength LiDAR data. It is essential to note, however, that the efficacy of deep learning models for building extraction can vary depending on the scene's complexity and the quality of the LiDAR data.

One of the earliest studies on building extraction using multispectral LiDAR data was conducted by [11]. They proposed a Graph Geometric Moments Convolutional Neural Network (GGMCNN) model for extracting buildings from airborne multi-spectral LiDAR point clouds. The GGMCNN model is a deep learning model that is specifically designed for processing point cloud data. It takes as input a set of features that are extracted from the point cloud, including the point's elevation, intensity, and spectral information. The GGMCNN model is trained to classify each point cloud as either building or non-building. This study has shown that deep learning can be used to extract buildings from multispectral LiDAR data with high accuracy. However, research is needed to develop and evaluate different deep learning models that deal with point clouds and raster format for this task on more accurate datasets such as multispectral LiDAR data.

Several studies have employed deep learning models to extract buildings from mono-wavelength LiDAR data in a raster format, as evidenced by the works of [3, 12-15]. In contrast, some researchers have studied the extraction of buildings from LiDAR data in the form of point clouds [11]. Nonetheless, a lack of research persists regarding the evaluation of deep learning models that are appropriate for building extraction from multispectral LiDAR data. This is an important area of research, and we hope our study can help fill this gap.

The advent of various deep learning networks specifically designed for processing raw LiDAR data [16-18] has facilitated the direct extraction of buildings from LiDAR point clouds. This is in contrast to previous methods that necessitated data rasterization prior to extracting features from the raster format [19]. Convolutional Neural Networks (CNNs) and their respective lineages are extensively employed and favored networks within the domain of deep learning [20]. One notable advantage of CNNs over previous models is their capacity to autonomously detect significant features without human intervention, rendering them more pragmatic [21]. The deep learning algorithm known as Point Convolutional Neural Network (Point CNN) [19] is distinguished by its capability to directly process raw cloud points, eliminating the requirement for the rasterization process. In a study conducted by [22], a comparative analysis was performed to evaluate the performance of the deep learning algorithm Point CNN in classifying land points in agricultural areas. The findings of the study demonstrated that Point CNN outperformed traditional methods in terms of accuracy. Furthermore, it has been demonstrated that Mask Region-based Convolutional Neural Network (Mask RCNN), a deep learning algorithm belonging to the same category as CNN, has exhibited notable efficacy in the extraction of buildings from LiDAR data. This is achieved through the conversion of cloud points into raster images [23, 24].

However, to the best of our knowledge, there is a lack of scientific investigations on the automatic extraction of buildings using multispectral LiDAR points based on the Point CNN algorithm, and comparing its results with the Mask R-CNN algorithm. Previous studies have not adequately compared these two methodologies. Currently, there exists a significant need for the automated and precise categorization of multispectral LiDAR points across various applications. Also, applying this approach to multispectral LiDAR data is an important step to increase the accuracy of building extraction in complex urban environments and facilitate the emergence of novel applications in subsequent endeavors.

In this study, the main contributions are:

1) Compare and contrast two distinct methodologies for extracting buildings from multispectral LiDAR data: Mask R-CNN and Point CNN. Both methods utilize three multispectral LiDAR channels to optimize building extraction.

2) Investigate the importance of using multispectral LiDAR data for building extraction. Using multispectral LiDAR data can considerably improve the accuracy of building extraction compared to using single-wavelength LiDAR data, according to the study's findings.

Overall, the research on building extraction using deep learning with multispectral LiDAR data is still in its early stages. However, the results from existing studies are promising and suggest that deep learning has the potential to improve the accuracy and efficiency of building extraction in urban environments.

III. Research Significance

This paper presents a significant analysis of two distinct methodologies employed for building extraction from multispectral LiDAR data. The methodologies involve either utilizing point clouds directly or converting them into a raster format. This study demonstrates the effectiveness of deep learning algorithms for building extraction. Furthermore, offers significant insights regarding the utilization of multispectral LiDAR data in the context of building extraction. This study's findings may have a wide variety of practical applications. Urban planners and emergency administrators could use them to automatically generate building footprints, for instance. They could also be used to improve the accuracy of 3D city models.

IV. Study Area and Dataset

The study area represents a complimentary dataset provided by the National Center for Airborne Laser Mapping (NCALM), encompassing an urban region located in Houston in the southeastern sector of Texas, United States of America. The study area, as illustrated in Fig. 1, encompasses the vicinity of the Houston University campus and its immediate surroundings. The study area encompasses an estimated area of 550m². The study area was selected based on its diverse range of building types, encompassing regular residential structures, edifices surrounded by a canopy, and small-scale constructions with haphazard layouts.

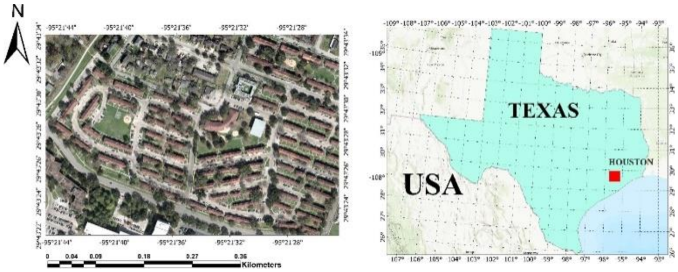


Fig. 1. Study area.

The research area was scanned in February 2017 using a Teledyne Optech Titan multi-spectral Airborne Laser Scanning (ALS) system [25]. The ALS dataset included multispectral data from 14 flight lines. All collected points were separately recorded in 42 LAS files, each corresponding to a distinct channel and strip. Every LAS file contained data on the point source ID, scan angle rank, flight line edge, scan direction flag, returns, GPS time, and intensity values. Detailed information about the dataset is shown in Table I; the Titan sensor was put in an Optech aircraft. The flight plan and equipment parameters are shown in the Table II.

TABLE I. SPECIFICATIONS OF OPTECH TITAN MULTI-SPECTRAL ALS LIDAR (TELEDYNE OPTECH TITAN, 2015)

Parameters	Channel 1	Channel2	Channel 3
Wavelength	1550 nm MIR	1064 nm NIR	532 nm Green
Beam divergence	0.35 mrad(1/e)	0.35 mrad(1/e)	0.70 mrad(1/e)
Look angle	3.5 ° forward	nadir	7.0 ° forward
Effective PRF	50–300 kHz	50–300 kHz	50–300 kHz

TABLE II. THE FLIGHT PLAN AND EQUIPMENT PARAMETERS

Flight Parameter	
Sensor ID	The Optech Titan MW (14SEN/CON340) LiDAR sensor
flying height	460 m AGL
swath width	445 m
overlap	50%
line spacing	225 m
Equipment Parameters	
PRF	175 kHz per channel (525 kHz total)
scan frequency	25 Hz
scan angle	±26° and ±2°cut-off at processing

V. METHODOLOGY

Two strategies for building extraction from multispectral LiDAR data were employed to accomplish the research objective. The first method extracts buildings from the raster data, whereas the second method extracts buildings from the point data.

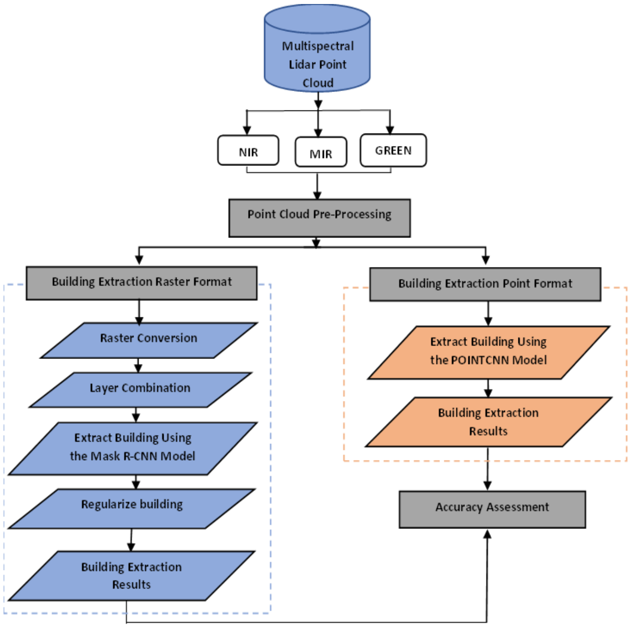


Fig. 2. Pre-processing flow Chart.

The data obtained from Multispectral LiDAR is presented in the LAS format, comprising point cloud values for x, y, z, and intensity. Fig. 2 depicts the flow chart for two distinct approaches to extracting buildings from multispectral LiDAR point cloud data. The building extraction strategy from the raster format is represented in blue, while the building extraction strategy from the point format is represented in orange. The point cloud pre-processing step was executed in two distinct approaches, followed by their respective application to two distinct deep learning algorithms. One algorithm is designed to handle data in raster format, while the other algorithm is specifically designed to process raw data in a point cloud format.

A. *Pre-Processing Point Cloud*

The pre-processing of the point cloud is the same for the two methods and it consists of removing the noise points, integrating the point clouds of the three different channels, segmenting the point cloud according to the boundary of the study area, and finally separating the ground and off-ground point clouds. The non-terrain point cloud is of interest in this paper because it contains the features of the buildings. The following will be elaborated upon in a comprehensive manner below:

1) Data cleaning: A statistical out-linear removal (SOR) algorithm was utilized to eliminate isolated points or any points that fell outside the intensity range [26]. The logarithmic function operates by computing the spatial separation between a given point and its six adjacent neighbors. In cases where the mean distance between a given point and its corresponding exceeds the established minimum threshold, the point is eliminated.

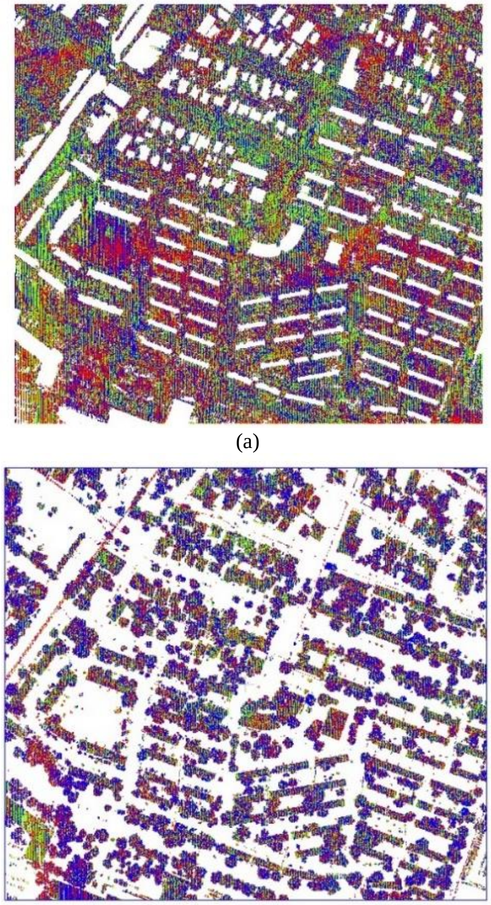
2) Merging and segmenting points: The absence of control points or reference points in our case has compromised the statistical reporting of geometric quality. In order to evaluate our building extraction methodology, a study area was chosen from a single strip to address the problem at hand. Before performing the step of merging the three different wavelengths into a single LAS file, each channel was cut in three directions based on the borders of the chosen study area. In order to obtain the highest efficiency of the multi-spectral LiDAR points, a merger of the three channels was made, as each channel collects data from a different angle of view and with different intensities, in addition to improving the density of the cloud points. The present study involves the integration of three separate point clouds through the utilization of 3-D spatial join methodology. Each individual point cloud of a specific wavelength serves as the reference point among the three. The reference point cloud employs the closest neighbor searching algorithm to identify neighboring points within the other two wavelengths of point clouds. Subsequently, the segmentation algorithm, which is available via the Cloud Compare software, was employed to clip the multispectral LiDAR points according to the selected boundaries of the chosen study area.

3) Filtering points: This research uses deep learning models to extract buildings from multispectral LiDAR data. To simplify this procedure, the Cloth Simulation Filter (CSF) provided by [15] was used to separate ground points and non-ground points as shown in Fig. 3, and concentrate exclusively on the latter because they include buildings. The CSF filter operates through the inversion of the point cloud and drops a simulated cloth model onto the designated points. The cloth undergoes a settling process as it contends with the opposing forces of gravity and internal cloth tension, which occurs over numerous iterations. The filter necessitates the provision of four input parameters, with the first parameter denoting the terrain type. In this case, the area being examined is characterized by flat terrain. Subsequently, a cloth resolution of 2.0, regulates the texture coarseness or smoothness of the cloth's simulation. The terrain simulation's maximum iteration is 500. Ultimately, a classification threshold of 0.5 was established in order to differentiate between terrestrial and non-terrestrial point clouds, utilizing point distances as the determining factor.

B. Building Extraction of Raster Data

This method utilizes Mask R-CNN and polygon regularization to accomplish its building extraction goals. Mask R-CNN can produce preliminary building polygons from an input image. Then, the basic polygons are transformed into regularized polygons using the polygon regularization technique. The pre-processing stage of the 3D point cloud was conducted to enable the point cloud to be suitable for direct extraction of buildings using the deep learning model Point-CNN. On the other hand, the deep learning model Mask R-CNN requires several steps to process the multi-spectral LiDAR data, with the most crucial of these steps being the conversion of the three-dimensional point cloud into a horizontal image plane through the process of rasterization.

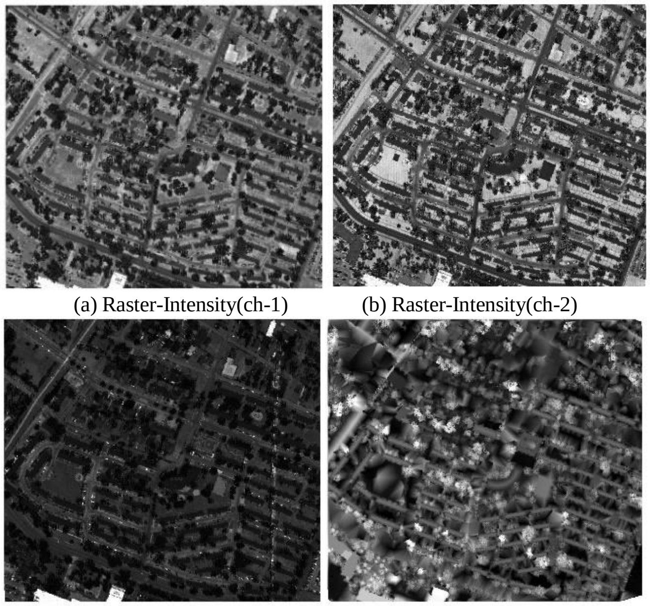
1) Raster conversation: The 3D point cloud was rasterized to a 2D raster with intensity and height information preserved. Cloud intensities at three distinct wavelengths were converted into three distinct intensity images for each channel separately as shown in Fig. 4(a, b, c). The intensity cloud rasterization process was executed by defining a specific set of properties. The cell size parameter was established as 0.1, and the binning interpolation technique was employed to compute the mean intensity value within cells that lack data points. The height raster as shown in Fig. 4(d) was generated from multispectral LiDAR data after merging the three channels and generating a DSM, as opposed to the intensity images for each channel separately. The raster height step was subjected to identical intensity rasterization settings parameters, including interpolation method and cell size, to ensure congruence between the raster resolution and applying the composite process on the four-raster dataset.



(a)

(b)

Fig. 3. Separation of ground and off-ground multispectral LiDAR points using CSF algorithm, (a) representing ground points and (b) representing above-ground points.



(a) Raster-Intensity(ch-1) (b) Raster-Intensity(ch-2)

(c) Raster-Intensity(ch-3) (d) Raster-height

Fig. 4. Rasterization multispectral LiDAR point cloud.

2) Layer combination: To streamline the training and validation process of the Mask R-CNN deep learning model, it is necessary to compress the four images within a multi-data image into a new multi-band raster data set. The input dataset comprised a series of four raster images (Raster Intensity Ch-1, Raster Intensity Ch-2, Raster Intensity Ch-3, and Raster height of non-terrain features). All four images possess identical dimensions in terms of length, width, and cell size. The dimensions of the analyzed space were identical to those of the study area, while the depth of the analysis was determined by the quantity of raster images within the input dataset. Specifically, a depth of four was utilized.

3) Mask R-CNN: The buildings in this method were extracted using the workflow of the Mask R-CNN deep learning model, as illustrated in Fig. 5. Initially, the multi-spectral LiDAR data underwent a qualification process for its inclusion in the deep learning model, as previously stated [27]. This data is comprised of four images that possess identical length, width, and cell size, and have been compressed accordingly. Subsequently, the training data intended for the deep learning model was exported. To this end, the study area was partitioned into training, validation, and test data sets. Subsequently, the training and validation sets, along with each input image, were utilized to produce the training data set. The input images were partitioned into image tiles of size 256 × 256 pixels, with a 50% overlap step-shift, in order to conform to the input requirements of the Mask R-CNN architecture and to guarantee that all buildings are represented in at least one image tile. The training dataset comprised 184 image slices and 239 features. Following that, the Mask R-CNN model is trained through the utilization of a designated training dataset. Table III. presents the parameters that were utilized to train the model. The Mask R-CNN architecture comprises several components: including a backbone, a Region Proposal Network (RPN), a Region of an Interest alignment layer (RoI Align), a Bounding box regressor, and a mask generation head [28] . To predict segmentation masks on each Region of Interest (ROI), the Mask R-CNN builds on the Faster R-CNN by adding a network branch to the original (ROI) [24]. To each ROI, the tiny FCN was applied that predicts a pixel-wise segmentation mask for building and nonbuilding regions. The first building polygon is found by following a region's boundary [29]. ArcGIS API for Python, with the Tensorflow and Keras libraries, was used to create and implement the Mask R-CNNs. It's based on a Region Proposal Network (RPN) and ResNet50 backbone.

TABLE III. MASK R-CNN TRAINS THE MODEL PARAMETERS

Parameter	Description
Backbone model	ResNet-50
Training and Test set was split into	70/20%
Validation	10%
Processing type	Nvidia GeForce RTX 2060 GPU
Batch size	4
Learning rate strategy	Stop when the model stop improve
Learning rate	0.0001

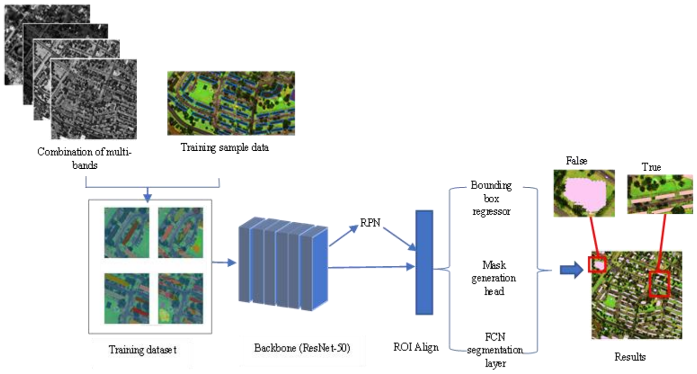


Fig. 5. The Mask R-CNN procedure for building extractions using Multi-Spectral LiDAR Dataset.

4) Regularize building footprint: After extracting the buildings with Mask R-CNN, it produces irregular and distorted polygons that don't have straight lines and right angles for edges due to the pixel-labeling location performed by Mask R-CNN [30]. In order to get rid of this randomness in building polygons, the regularized building footprint step was used. A polyline compression algorithm was used to reduce these distortions of building polygons. Table IV shows the parameter used in this algorithm to obtain a much cleaner and closer footprint to buildings than the results that we got from the deep learning algorithm Mask-RCNN.

TABLE IV. REGULARIZE BUILDING FOOTPRINT PARAMETERS

Parameter	Description
Method	Right angle
Tolerance	1
Precision	0.25
Diagonal penalty	1.5
Minimum radius	0.1
Maximum radius	1000000
Processor type	GPU

C. Building Extraction of Point Cloud Data

The Point Convolutional Neural Network (Point CNN) algorithm is utilized in this research to extract buildings directly from the raw multispectral point cloud without a rasterizing step. The Point CNN architecture consists of an encoder network and a decoder network as shown in Fig. 6, both of which contain X-Conv layers. The cipher network consists of four collective abstraction units to iteratively extract multiscale features of scale (1/256, 1/256,1/512, 1/1024) concerning the entry point cloud with point number N. Concerning entry points, the entry point merged of intensities of three channels, and adding the elevation point cloud was DSM. K is set from 8 to 16 in this study, where K is the neighboring points around the representative points and N is the number of the point in the previous layer. In this setting, the final point has a 1:0 receptive field since it sees all points from the previous layer, and its features contribute to an accurate semantic interpretation of the shape. In the second X-Conv layer, a dilation rate of D = 2 was used, and then gradually increased in the third and fourth X-Conv layers to ensure that all remaining representative points see the entire figure and are all suitable for making predictions. In this way, the last X-Conv is more thoroughly trained layers, as more connections are involved in the network. The decoder network comprises three feature propagation units, which gradually restore a robust feature mean representation to produce a high-quality classifier point cloud. The output data size of the encoder is N/256, N/512, and N/1024. Finally, a fully connected layer is added on top of the last output of the X-Conv layer, followed by a loss, to train the network.

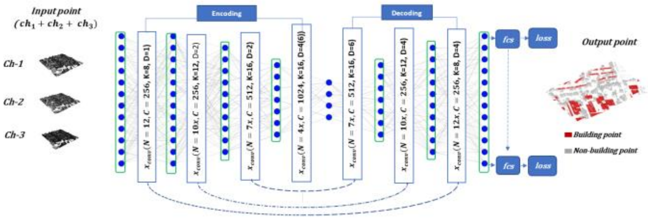

Input point
$(ch_1 + ch_2 + ch_3)$
Ch-1
Ch-2
Ch-3
Encoding
Decoding
$x_{conv}(N = 12, C = 256, K=8, D=1)$
$x_{conv}(N = 10x, C = 256, K=12, D=2)$
$x_{conv}(N = 7x, C = 512, K=16, D=2)$
$x_{conv}(N = 4x, C = 1024, K=16, D=4(6))$
$x_{conv}(N = 7x, C = 512, K=16, D=6)$
$x_{conv}(N = 10x, C = 256, K=12, D=4)$
$x_{conv}(N = 12x, C = 256, K=8, D=4)$
fcs
loss
fcs
loss
Output point
Building point
Non-building point

Fig. 6. Point CNN architecture for building extraction from multispectral LiDAR where, $\boldsymbol{I_1}$: Intensity of channel-1, $\boldsymbol{I_2}$: Intensity of channel -2, $\boldsymbol{I_3}$: Intensity of channel-3, N: Number of the point in the previous layer.

VI. ACCURACY ASSESSMENT

The following measures were employed for assessing the effectiveness of the proposed approaches, all of them are standard for any semantic segmentation and classification work.

$$Precision = \frac{T_p}{T_P + F_p} * 100\% \quad (1)$$

where the term "precision" refers to the proportion of accurately labeled data points relative to the total number of data points that were labeled, the percentage of data points that were successfully classified relative to the total number of data points that were expected to be classified with this value is the recall. F1-score is the arithmetic mean of the precision and recall values are given as follows:

$$Recall = \frac{T_p}{T_p + F_n} * 100\% \quad (2)$$

$$F1 - score = \frac{2 * precision * recall}{precision + recall} \quad (3)$$

Where: T_p is a number of point clouds that are classified as true positive building extraction, T_n are truly negative, F_p is false positive and F_n is a false negative.

VII. RESULTS AND DISCUSSION

The study area was subjected to the training of two deep learning models, namely Mask R-CNN and Point-CNN for the purpose of extracting buildings from a Multispectral LiDAR point cloud. The training dataset and validation for both models were selected using the same buildings to facilitate comparison. The evaluation of building extraction results was conducted using a confusion matrix approach. The reference points for ground truth were obtained from orthorectified aerial photographs captured by the same multispectral LiDAR system in the same survey.

A. Building Extraction Result of Mask R-CNN

As shown in Fig. 7(a), the outcomes of utilizing the mask R-CNN deep learning model on the multi-spectral LiDAR data after converting them into raster format, as this model confirmed its ability to train and extract the footprints of buildings, but in an irregular style. After they were included in the polyline compression algorithm to create the uniformity of the building, the results are shown in Fig. 7(b).

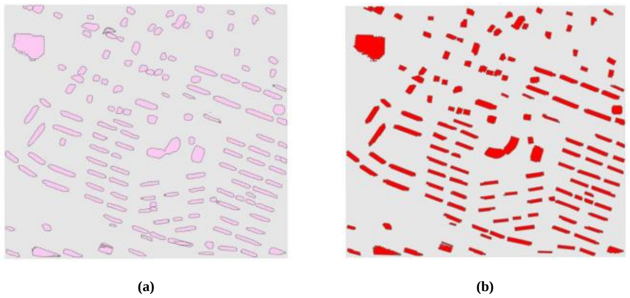


Fig. 7. Building extraction results of mask RCNN: (a) before using regularize, (b) after regularize.

B. Building Extraction Result of Point CNN

The visual outcomes of building extraction through the utilization of the Point CNN deep learning model are presented in Fig. 8. Specifically, Fig. 8(a) displays the raw multispectral LiDAR points prior to building recognition, with all points depicted in grey. On the other hand, Fig. 8(b) shows the outcomes attained by the Point CNN model in building recognition, where buildings are highlighted in red and the background is depicted in grey. The employment of Point CNN in contrast to the use of Mask R-CNN is observed to yield superior outcomes. The aforementioned model demonstrated a capacity to discern points of construction with a mean precision of 0.9340 within the given dataset. Furthermore, the algorithm demonstrated a notable proficiency in distinguishing between the points of buildings and those of trees, particularly in cases where buildings were encompassed by thick clusters of trees. This model was also able to extract the points of small and irregular buildings accurately.

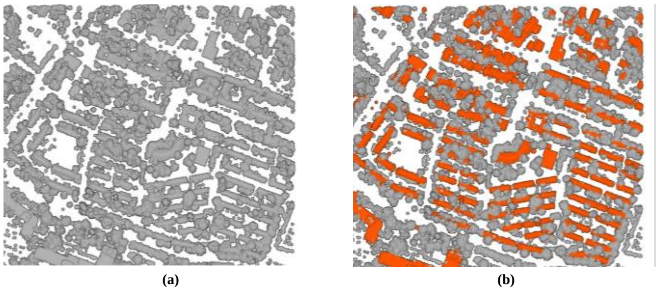


Fig. 8. Building extraction results of Point CNN: (a) before building extraction, (b) after building extraction.

The precision achieved through the utilization of the Point format approach surpasses that of the Raster format methodology. It is anticipated that the inclusion of supplementary spectral data, such as indices, may enhance the ultimate outcomes of building extraction. Ongoing investigations are being conducted with the aim of enhancing the precision of the building extraction.

C. Comparison between Two Methods

The two methods were compared using evaluation scales such as precision, recall, and F1-score to enhance the comparison. Table V displays the obtained results from using the Point CNN model, which indicates a precision of 93.40%, a recall of 92.34%, and an F1-score of 92.72%. While Mask R-CNN gave less accurate results, it demonstrates a precision of 74.66%, a recall of 67.43%, and an F1-score of 71.17%.

According to the findings of our research, Point CNN performs better than Mask R-CNN when it comes to the extraction of buildings from multispectral LiDAR data in terms of both accuracy and efficiency. In comparison to Mask R-CNN, Point CNN is capable of directly extracting features from point clouds without the need for any pre-processing steps such as voxelization. This results in a greater resolution and significantly faster processing times. In addition to this, Point CNN is able to manage irregular point clouds, which are typical in the case of LiDAR data. Nevertheless, Mask R-CNN continues to have advantages compared with Point CNN in a number of contexts, since it operates on raster data like satellite images or aerial photos rather than point clouds. Also, Mask R-CNN is more suited to identifying and categorizing objects seen from a variety of angles. This makes it an ideal candidate for this task: the Precision, Recall, and F1-score obtained from the two different methods.

TABLE V. THE PRECISION, RECALL, AND F1-SCORE OBTAINED FROM THE TWO DIFFERENT METHODS

Method	Mask R-CNN	Point- CNN
Precision%	74.66	93.40
Recall%	64.43	92.34
F1_Score%	69.17	92.72

TABLE VI. THE TRUE-POSITIVE(T_P), FALSE-POSITIVE(F_P), AND FALSE-NEGATIVE(F_N) FROM THE TWO DIFFERENT METHODS

Method	T_P	F_P	F_N
Mask-RCNN (mask)	112	38	85
Point CNN (points)	897392	26528	91137

According to Tables V, and VI, the results suggest that Point CNN is a more effective method for building extraction from multispectral LiDAR data. It has a higher TP, Precision, and F1-Score than Mask R-CNN. However, Mask R-CNN has a higher Recall, indicating that it is less likely to miss buildings.

Comparing our findings to LiDAR building extraction research [13, 30], our study had 93.40% accuracy, 92.34% recall, and 92.72% F1. Point CNN retrieves features from point clouds without rasterization, which may explain this. Accuracy and processing speed improve. Furthermore, the integration of the three distinct spectra of the multi-spectral LiDAR plays a crucial role in accurately discerning and distinguishing buildings and other features.

According to the study's findings, Point CNN outperformed Mask R-CNN in building extraction from multispectral LiDAR data. This is probable because Point CNN processes point cloud data directly, preserving its structure and characteristics. This enables Point CNN to capture fine-grained geometric details and relationships within the point cloud, which is crucial for accurate building extraction and avoids voxelization, which is sometimes required by Mask R-CNN. This improves efficacy because no data is lost in the conversion process. Maintaining the original spatial distribution of points without voxelization is also essential for accuracy in point clouds with irregular spacing. Due to its architecture, it can manage multispectral LiDAR data from a variety of perspectives. The model captures and uses data from diverse perspectives to increase building extraction accuracy. Mask R-CNN, on the other hand, is well-suited for distinguishing and categorizing objects seen from a variety of angles because it operates on raster data such as satellite images or aerial photographs.

In addition, it is recommended that future research endeavors include evaluating the performance of Point CNN and Mask R-CNN on other datasets, including datasets with different types of scenes (e.g., urban, rural, forested) and datasets with different types of multispectral LiDAR data (e.g., different wavelengths, different point densities, and use of spectral indicators).

VIII. CONCLUSION

A study was conducted to analyze multispectral LiDAR data to extract buildings from a residential area situated near the University of Houston, situated in the state of Texas, United States of America. The present investigation undertook a comparative analysis of two discrete deep learning models that are categorized under the Convolutional Neural Network (CNN) family. The present investigation aimed to assess the effectiveness of the method employed in extracting buildings in two separate scenarios. The study involved the utilization of a genuine dataset of multispectral LiDAR data for experimentation purposes. Before inputting LiDAR points into the Point CNN deep learning model, processing operations were executed. Similarly, operations were conducted to transform cloud points into pixels for input into the mask R-CNN deep learning model. Furthermore, a classification of architectural structures was conducted after their acquisition via mask R-CNN. The standardization of ground truth reference was implemented to facilitate a comparison between the two methods, as this is orthorectified aerial photographs captured by the same multispectral LiDAR system in the same survey. It can be concluded that the use of the CNN point model with the proposed approach, which combines the advantages of the intensity of the three different wavelengths plus the height component of the DSM gives better results for extracting buildings from multispectral LiDAR point data, where the accuracy of the results improved by about 30%.

ACKNOWLEDGMENT

The authors thank the Hyperspectral Image Analysis Lab at the University of Houston for providing the original Optech Titan data.

REFERENCES

[1] G. Chitturi, "Building Detection in Deformed Satellite Images Using Mask R-CNN," ed, 2020.

[2] W. Nurkarim and A. W. Wijayanto, "Building footprint extraction and counting on very high-resolution satellite imagery using object detection deep learning framework," Earth Science Informatics, vol. 16, no. 1, pp. 515-532, 2023.

[3] A. Gamal et al., "Automatic LIDAR building segmentation based on DGCNN and Euclidean clustering," Journal of Big Data, vol. 7, pp. 1-18, 2020.

[4] K. Bakuła, "Multispectral airborne laser scanning-a new trend in the development of LiDAR technology," Archiwum Fotogrametrii, Kartografii i Teledetekcji, vol. 27, 2015.

[5] O. A. Mahmoud El Nokrashy, L. G. E.-D. Taha, M. H. Mohamed, and A. A. Mandouh, "Generation of digital terrain model from multispectral LiDar using different ground filtering techniques," The Egyptian Journal of Remote Sensing and Space Science, vol. 24, no. 2, pp. 181-189, 2021.

[6] M. M. Taye, "Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions," Computers, vol. 12, no. 5, p. 91, 2023.

[7] S. S. Ojogbane, S. Mansor, B. Kalantar, Z. B. Khuzaimah, H. Z. M. Shafri, and N. Ueda, "Automated building detection from airborne LiDAR and very high-resolution aerial imagery with deep neural network," Remote Sensing, vol. 13, no. 23, p. 4803, 2021.

[8] A. Novo, N. Fariñas-Álvarez, J. Martínez-Sánchez, H. González-Jorge, and H. Lorenzo, "Automatic processing of aerial LiDAR data to detect vegetation continuity in the surroundings of roads," Remote Sensing, vol. 12, no. 10, p. 1677, 2020.

[9] W. Y. Yan, A. Shaker, and N. El-Ashmawy, "Urban land cover classification using airborne LiDAR data: A review," Remote Sensing of Environment, vol. 158, pp. 295-310, 2015.

[10] I. Prieto, J. L. Izkara, and E. Usobiaga, "The application of lidar data for the solar potential analysis based on the urban 3D model," Remote Sensing, vol. 11, no. 20, p. 2348, 2019.

[11] D. Li et al., "Building extraction from airborne multi-spectral LiDAR point clouds based on graph geometric moments convolutional neural networks," Remote Sensing, vol. 12, no. 19, p. 3186, 2020.

[12] E. Maltezos, A. Doulamis, N. Doulamis, and C. Ioannidis, "Building extraction from LiDAR data applying deep convolutional neural networks," IEEE Geoscience and Remote Sensing Letters, vol. 16, no. 1, pp. 155-159, 2018.

[13] S. A. Mohamed, A. S. Mahmoud, M. S. Moustafa, A. K. Helmy, and A. H. Nasr, "Building Footprint Extraction in Dense Area from LiDAR Data using Mask R-CNN," International Journal of Advanced Computer Science and Applications, vol. 13, no. 6, 2022.

[14] F. H. Nahhas, H. Z. Shafri, M. I. Sameen, B. Pradhan, and S. Mansor, "Deep learning approach for building detection using lidar–orthophoto fusion," Journal of sensors, vol. 2018, 2018.

[15] W. Zhang et al., "An easy-to-use airborne LiDAR data filtering method based on cloth simulation," Remote sensing, vol. 8, no. 6, p. 501, 2016.

[16] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3D classification and segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 652-660.

[17] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," Advances in neural information processing systems, vol. 30, 2017.

[18] Z. Jing et al., "Multispectral LiDAR point cloud classification using SE-PointNet++," Remote Sensing, vol. 13, no. 13, p. 2516, 2021.

[19] Y. Li, R. Bu, M. Sun, W. Wu, X. Di, and B. Chen, "Pointcnn: Convolution on x-transformed points," Advances in neural information processing systems, vol. 31, 2018.

[20] A. Dhillon and G. K. Verma, "Convolutional neural network: a review of models, methodologies, and applications to object detection," Progress in Artificial Intelligence, vol. 9, no. 2, pp. 85-112, 2020.

[21] L. Alzubaidi et al., "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," Journal of Big Data, vol. 8, pp. 1-74, 2021.

[22] N. Fareed, J. P. Flores, and A. K. Das, "Analysis of UAS-LiDAR Ground Points Classification in Agricultural Fields Using Traditional Algorithms and PointCNN," Remote Sensing, vol. 15, no. 2, p. 483, 2023.

[23] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," Advances in neural information processing systems, vol. 28, 2015.

[24] A. Mahmoud, S. Mohamed, R. El-Khoribi, and H. Abdel Salam, "Object detection using adaptive mask RCNN in optical remote sensing images," Int. J. Intell. Eng. Syst, vol. 13, no. 1, pp. 65-76, 2020.

[25] T. O. Titan, "Multispectral LiDAR system: high precision environmental mapping," ed, 2015.

[26] A. Carrilho, M. Galo, and R. C. Dos Santos, "STATISTICAL OUTLIER DETECTION METHOD FOR AIRBORNE LIDAR DATA," International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, vol. 42, no. 1, 2018.

[27] K. Yu et al., "Comparison of classical methods and mask R-CNN for automatic tree detection and mapping using UAV imagery," Remote Sensing, vol. 14, no. 2, p. 295, 2022.

[28] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2961-2969.

[29] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431-3440.

[30] K. Zhao, J. Kang, J. Jung, and G. Sohn, "Building extraction from satellite images using mask R-CNN with building boundary regularization," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 247-251.