Building Footprint Extraction in Dense Area from LiDAR Data using Mask R-CNN

Sayed A. Mohamed*, Amira S. Mahmoud, Marwa S. Moustafa, Ashraf K. Helmy, Ayman H. Nasr
Data Reception, Analysis and Receiving Station Division, National Authority for Remote
Sensing and Space Science, Cairo, Egypt

***Abstract*—Building footprint extraction is an essential process for various geospatial applications. The city management is entrusted with eliminating slums, which are increasing in rural areas. Compared with more traditional methods, several recent research investigations have revealed that creating footprints in dense areas is challenging and has a limited supply. Deep learning algorithms provide a significant improvement in the accuracy of the automated building footprint extraction using remote sensing data. The mask R-CNN object detection framework used to effectively extract building in dense areas sometimes fails to provide an adequate building boundary result due to urban edge intersections and unstructured buildings. Thus, we introduced a modified workflow to train ensemble of the mask R-CNN using two backbones ResNet (34, 101). Furthermore, the results were stacked to fine-grain the structure of building boundaries. The proposed workflow includes data preprocessing and deep learning, for instance, segmentation was introduced and applied to a light detecting and ranging (LiDAR) point cloud in a dense rural area. The outperformance of the proposed method produced better-regularized polygons that obtained results with an overall accuracy of 94.63%.**

Keywords—Deep learning; object detection; mask R-CNN; point cloud; light detecting and ranging (LiDAR)

I. Introduction

The identification and extraction of urban footprint has become an important research topic and tool in city planning, transportation planning, urban simulation, 3D city modelling, and building change detection [1-3]. Automatic building footprint extraction is needed to meet the rising demand for precise city building outlines. LiDAR data creates digital terrain and surface models [4]. Despite the benefits of using LiDAR to extract vegetation, essential infrastructure, and hydrography, building footprint extraction is desired for estimating population, energy demand, and quality of life [5].Several techniques have been introduced to extract building footprints using optical sensors. These techniques include image-based, LiDAR-based, and data fusion-based [6]. For instance, Image-based technique use spectral properties. Spectral ambiguities and shadow occlusions can lead to inaccurate building footprints. [7]. Nemours approaches used LiDAR intensity, echo, and geometric attributes, but fusing LiDAR and high-resolution images improves performance and robustness.

Deep learning uses multilayer neural networks in many applications [8, 9] such as: object detection [10], image classification, image denoising [11], medical image segmentation [12], image super-resolution [13-15], and depth prediction in stereo and monocular images [16]. Recently, several researches have investigated deep learning algorithms to improve building footprint extraction [17-19] either using CNN or a fully convolutional neural network.

CNN-based object detectors are single and two-stage. Fast R-CNN, faster R-CNN [20], and mask R-CNN are widely identified as two-stage detectors. Fast R-CNN doesn't allow end-to-end training since it uses a selective search to extract region proposals, which reduces the performance. Faster R-CNN replaces Region Proposal Network (RPN) selections, allowing end-to-end training. However, multiscale and small objects are a challenge. Despite their high inference speed [21], YOLO [22], YOLOV2, YOLOV3, and Single Shot Detector SSD [23] are single-stage networks with low detection accuracy in dense and tiny objects. Building footprint extraction requires accuracy; hence a two-stage neural network is used.

Mask R-CNN combines object detection and segmentation to improve overall accuracy and detect small and multiscale objects. But the detection speed is hardly real-time.

Class imbalance is an issue in remote sensing. This occurs when one or more classes are underrepresented in a dataset [7]. Traditional learning algorithms assume a balanced training set, which leads to a bias toward the majority classes. Consequently, the built model predicts poorly since all objects are in the dominating class regardless of the feature vector value [24]. The majority class classification bias is worse for high-dimensional data when variables exceed samples. The problem of skewed class distribution caused by uneven data was ignored. Class imbalance techniques are divided into data and algorithmic techniques. Data level approaches include data sampling, random over sampling, random under sampling, and a hybrid between them and feature selection. Algorithmic approaches are cost-sensitive and hybrid/ensemble. these approaches perform better.

In imbalanced datasets, ensemble classifiers improve single classifiers by merging them. Ensemble learning algorithms improve imbalanced data classification more than data sampling strategies. Due to precision-focused ensemble construction methods, the minority class is unrecognized. Developing ensemble learning algorithms must address class imbalances. Several approaches using ensemble learning and imbalanced learning have been reported [4]. Integrating ensemble-based techniques into an imbalanced dataset reduces overfitting and improves classification accuracy.

*Corresponding Author.

This paper used an ensemble of mask R-CNNs to effectively extract the building footprint using the LiDAR dataset in dense rural areas. The dense area in Maghagha city contains a skew distribution between structure and unstructured buildings. Datasets of GIS buildings were integrated with the collected LiDAR dataset to improve the building extraction results. The main contributions are summarized as follows:

- The mask R-CNN framework was used to effectively extract building footprints in dense areas.
- Different core mask R-CNN networks were adopted to benefit from transfer learning and different strategies (data augmentation and postprocessing).
- Class imbalance was handled in building types using a weighted voting ensemble approach.

The remainder of this work is structured as follows: Section 2 provides relevant work and a summary of point cloud classification methods. Section 3 introduces the proposed LiDAR building footprint extraction method from the point cloud. Section 4 summaries the results of several tests done to evaluate the efficiency of the proposed LiDAR classification method on real data for Maghagha area. Section 5 concluded the findings.

II. Related Work

LiDAR is an effective remote sensing technology for precisely describing terrain geometry. Thus, it is a viable solution for mapping dense urban areas to support infrastructural reconstruction, maintenance, and visibility. LiDAR technology provides very precise spatial resolution and height information [25]. Many studies have outlined the benefits of applying LiDAR data in characterizing urban structures [26, 27]. LiDAR point cloud data segmentation for automatic building extraction improves building detection and surface extraction in urban scenes [28]. A point feature based on normal vector variance is presented to extract buildings from LiDAR point cloud data by merging point- and grid-based features [29]. Building footprints using LiDAR data were needed to build a dataset for the open data portal and evaluate the minimal acceptable criteria for accurate building extraction [30]. Airborne laser scanning is a good choice in urban planning because of its capacity to determine building height, mobility, and rapid data acquisition [19, 31].

Morphology utilizes a filtering window to create an identical output image. A morphological operation compares a point's value to its neighbors. Dilation and erosion are morphological procedures that extend or diminish structures. Dilation adds border points, but erosion subtracts. The image structure dictates how many points are added or removed. The structuring element is a set of coordinates that determines the performance [1, 32].

Recently, machine learning algorithms in LiDAR have been generalized. Insufficient, complicated structure and large size limit the machine learning performance. In [4], The authors presented an effective method for combining point cloud and optical data. The method extracted points and super voxel features. The TraAdaboost algorithm with multi classes was utilised to improve LiDAR-based point cloud classification. The results demonstrated the improved classification performance compared with nonregistered LiDAR points. In [5] Integrated spectral signatures from diverse sensors into LiDAR point cloud classification utilizing multiple feature spaces using machine learning algorithms.

Furthermore, several efforts were conducted [1, 4, 5] to assess LiDAR data in building footprint extraction, varying between semiautomatic to automatic. Extracting a building footprint can be divided into three phases: isolating nonground points, segmenting building points, and extracting the building outline from the building footprint segmentation, as shown in Fig. 1.

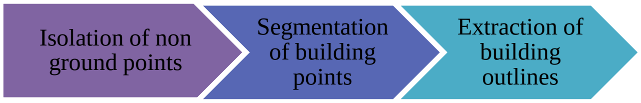

Isolation of non ground points
Segmentation of building points
Extraction of building outlines

Fig. 1. Typical Building Footprint Extraction from LiDAR Data.

Different levels of filters, such as morphological filters [5, 24, 33], progressive densification [34], surface-based filters [35], and segmentation-based filtering [36], can be applied in the isolation of nonground points phase. These filtering methods work well on flat ground but poorly on undulating terrain [37]. Another common approach was to employ a nonhierarchical classification to separate land-use types. Here, point cloud intensity was used to improve classification. However, roadways and parking lots have the same intensity values as building rooftops; including them does not improve results [38].

In the segmentation phase, trees, utilities, buildings, etc. are used as nonground points. A traditional method for separating these objects uses thresholds on a digital surface model (DSM) or a normalized DSM (difference between Digital Terrain Models DTM and DSM) [39]. This isn't always successful, as trees and buildings are often comparable heights and close together [40]. Other approaches for separating trees from building points include morphological filters, texture analysis, and plane fitting [41], and hierarchical object-oriented classification [42]. Notably, the acuracy depends on study area complexity [43]. SSD is considered simple compared with other approaches that use object proposals. SSD encapsulated proposal creation and feature resampling in a single network to simplify training [23]. Mask R-CNN can distinguish the adjacent objects and extract the outline of an object [44]. Finally, an extraction technique to construct a polygon or footprint from noisy and irregular boundaries is required. Examples of most common approaches include the least squares technique [45], nonlinear least squares [46], angle histogram of boundary points [47], weighted line segmentation [5], and invariant parameters using known roof types [48]. The quality of the building footprint depends on the various factors as point density, geometry, and building density [48]. In a fully convolutional network, a Spatial Residual Inception module termed (SRI-Net) was proposed to collect and combine multiscale multilevel features. SRI-Net can detect large buildings easily while preserving global and local details [49]. Due to land-cover changes and delayed geospatial data updates, some building annotations may be missing in the ground truth building mask., thereby leading to confusion in

CNN. To address this issue, the building footprints extraction problem was formulated as a long-tailed classification. Then, a three-term joint loss function was proposed: 1) logit adjusted cross-entropy, 2) weighted dice loss, and 3) boundary alignment loss. The obtained results indicate that the proposed loss function preserves the fine-grained structure of building boundaries, effectively discriminates between building and background pixels, and increases F1-scores [50].

III. Proposed Method

An ensemble method for building footprint extraction was introduced that combines two mask R-CNNs working in tandem, followed by a postprocessing phase to enhance building footprint prediction. Two backbone architectures were adopted ResNet (101, 34). The input layer accepts images of dimensions 256 × 256 and 128 × 128 pixels, respectively. Furthermore, different augmentation approaches were adopted to enhance the results. Fig. 2 shows a graphical representation of the proposed approach.

The mask R-CNN is a versatile model used in different fields [21] and comprises two phases: region proposals generation and classification [51]. This paper adopted the mask R-CNN as the benchmark model for detecting the footprints of rural buildings in dense areas. In the following subsection, data preparation, training, and detection phases were discussed in detail.

A. Data Preparation

Six-stage workflow in extracting the building footprint was adopted a. In the first stage, DSM for the study area was generated a using LiDAR ENVI software. To automatically define the building footprint, only points designated as buildings was filtered. The deep learning framework for the ArcGIS Pro software was incorporated in the preparation and labeling of the second stage. Thus, 150 sample buildings were chosen to serve as training data for the proposed neural network ensemble using "Label Object for Deep Learning." In the third stage, to contain image chips and labels, the sample data were converted into training data using the "Export Training data for Deep Learning" tool. In the fourth stage, two mask R-CNN ResNet backbones (34, 101) were trained and generated models for each of them. In the fifth stage, the "Detecting Objects using Deep Learning" tool was used for testing and data inferencing. Finally, the regularization of the building footprint use was done in the sixth stage by simply removing artifacts and correcting distortions in the building footprint polygons generated using "Feature Extraction."

B. Backbone Initialization

Image patches of size (256 × 256 and 128 × 128) were fed to ResNet-101 and ResNet-34 backbones in mask R-CNN, respectively to extract features. Table I shows a detailed description of both backbones. The residual family converges faster and achieves better training results compared with the shallow network. The image patches were fed to the backbone architecture to extract feature maps using transfer learning. These feature maps served as input for the next layer, after which the RPN was applied. This forecasts whether an object is present in that region. Here the regions obtained from RPN, which the model predicts, contain some objects and take various shapes. Hence, a pooling layer was applied to make the shape of all regions uniform. Next, these regions were fed to a fully connected network to forecast class labels and bounding boxes.

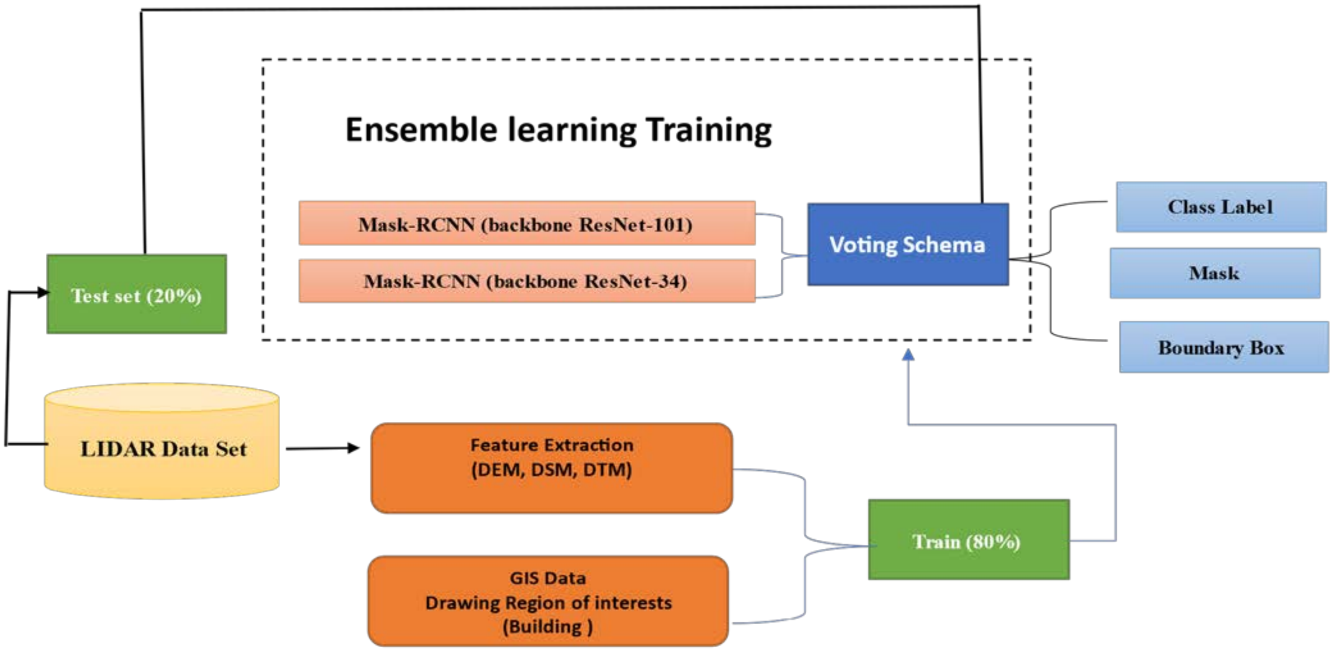


Fig. 2. Workflow of Mask R-CNN Ensemble Learning by using Two different Backbones Resnet 101 and 34.

C. The Ensemble Voting Schema

Dense urban in the Maghagha study area is considered a mix of structured and unstructured building units. However, the skew class distribution between most of the unstructured and structured buildings introduced a bias in favor of the majority class. To address this problem, a hard-weighted voting scheme of the selected models was set ranging from 0 to 1. For each roof type, each model has a class likelihood score. The scores were multiplied by the CNN weights. Then, the products for the weighting step were summed. The weights were chosen at random using the Bayesian optimization process. Finally, the output class was decided using the maximum probability index. With default parameter settings, the Bayesian optimization method was run 100 times. The weighted voting is given in Eq (1).

$$\text{weighted voting} = \mathrm{w}_1 * \mathrm{Model}_1 + \cdots . + \mathrm{w}_\mathrm{n} * \mathrm{Model}_\mathrm{n} \qquad (1)$$

where n denotes the number of models and w_1 and $Model_1$ represent the weight and probability score of the selected Mask R-CNN model, respectively. Two models are considered in this work (n=2).

TABLE I. RSNET-34 AND 101 ARCHITECTURES

Layer name	Output size	ResNet-34	ResNet-101
Conv1	112 × 112	7 × 7,64, stride 2 3 × 3 max pool, stride 2	
Conv2.X	56 × 56	$\begin{bmatrix}3\times3,64\\3\times3,64\end{bmatrix}\times 3$	$\begin{bmatrix}1\times1,64\\3\times3,64\\1\times1,256\end{bmatrix}\times 3$
Conv3.X	28 × 28	$\begin{bmatrix}3\times3,128\\3\times3,128\end{bmatrix}\times 4$	$\begin{bmatrix}1\times1,128\\3\times3,128\\1\times1,512\end{bmatrix}\times 4$
Conv4.X	14 × 14	$\begin{bmatrix}3\times3,256\\3\times3,256\end{bmatrix}\times 6$	$\begin{bmatrix}1\times1,256\\3\times3,256\\1\times1,1024\end{bmatrix}\times 23$
Conv5.X	7 × 7	$\begin{bmatrix}3\times3,512\\3\times3,512\end{bmatrix}\times 3$	$\begin{bmatrix}1\times1,512\\3\times3,512\\1\times1,2048\end{bmatrix}\times 4$
	1 × 1	Average pool, SoftMax	
FLOPs		3.6 × 109	7.6 × 109

IV. EXPERIMENTAL RESULTS

A. Dataset

Maghagha city is located in the north of El-Minya Governorate, Egypt. It is positioned between longitudes (30° 30′: 31° E), and latitudes (28° 30′: 29° N), as shown in Fig. 3, and covers an area of approximately 2,700 km2 [52]. Trimble® AX60, a high-performance, adaptable, and fully integrated airborne LiDAR solution designed to fulfill most aerial survey needs were utilized. The AX60 is a complete system that provides optimum quality, operational flexibility and efficiency, and in-service reliability [53]. The dataset was acquired using an Airborne Beechcraft B200. Furthermore, another dataset of high-resolution optical data, Nikon IC65+ and 2D-RGB imagery, from the same aircraft with sensors being rigidly fixed to the same platform used. Table II shows the parameters of the system used. The collected dataset study area comprises 10 LAS files, each approximately 2.2 km in width and 18.5 km in length. Additionally, we collected 580 TIF RGB images measuring 1.6 km in length and 1.2 km in width. Because of limited computation power, one LAS file was used to generate DSM and Digital Elevation Model DEM for the data object segmentation process. As a sample training set, only two roof types were considered: structured and unstructured, as shown in Fig. 4.

B. Evaluation Matrics

The proposed building footprint extraction workflow performance was evaluated using the overall accuracy (OA), precision, recall, and F-score. The precision computed by Eq. (2) shows the average of images that are correctly identified to the total number of structured and unstructured buildings that are correctly and non-correctly identified with the reference input.

$$\text{Precision (P)} = \frac{\mathrm{T_p}}{\mathrm{T_p}+\mathrm{F_p}} \qquad (2)$$

where $\mathrm{T_p}$ and $\mathrm{F_p}$ represent the true and false positives, respectively.

TABLE II. THE PARAMETERS OF THE USED SYSTEM

Trimble® AX60 System	
LiDAR point clouds	
Sensor model	Trimble AC IQ180
Laser wavelength	Near-infrared
Laser pulse repetition rate (PRR)	100–400 kHz
Scanning mechanism	Rotating polygon mirror
Scan frequency (max.)	200 Hz
Operating flight altitude	50–4700 m (164–15,500 ft) AGL
Range measurement accuracy	2 cm
Intensity capture	16-bit dynamic range for each echo
Digital aerial camera	
Model	Nikon IC 65+
Array size	80 MP
Channels	Three (RGB)
Shutter type	Electronically controlled leaf shutter
Ground sample distance	>5 cm
Calibration	Geometrically and radiometrically

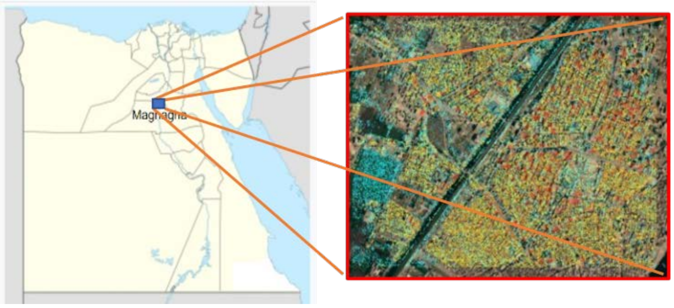


Fig. 3. Location Map of the Study Area (Maghagha, El-Minya Governorate, Egypt).

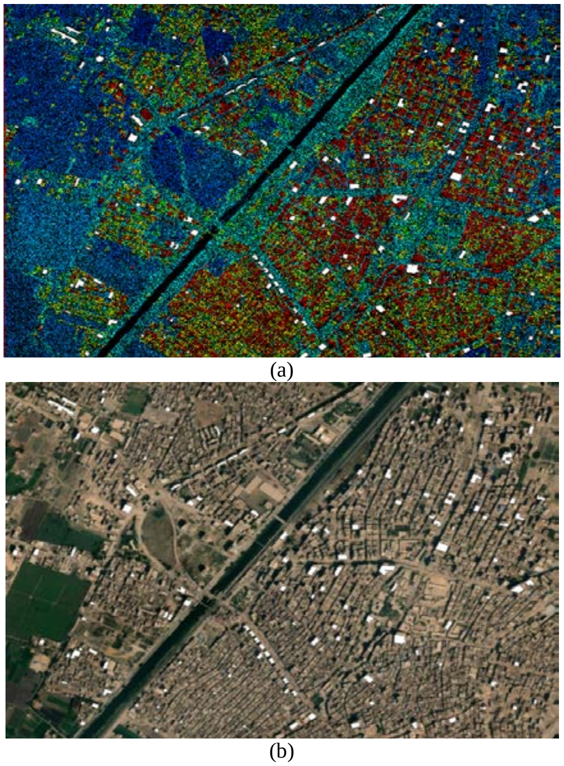


Fig. 4. Samples of Training Set Roofs, Structured and Nonstructured Covered (a) Point Clouds and (b) Very High-resolution RGB Image.

Recall, Eq. (3) defines the average number of structured and nonstructured buildings that are correctly identified of the total number of buildings that are correctly and non-correctly identified.

$$\text{Recall (R)} = \frac{T_p}{T_p+F_N} \quad (3)$$

where F_N represents the false negative.

F-score is defined in Eq. (4). If the obtained value is 1, the object detection is best and is worst when at 0.

$$\text{F-score} = \frac{2PR}{P+R} \quad (4)$$

Finally, OA represents the ratio of correctly identified structured and nonstructured buildings to the total number of buildings.

C. Experimental Setup

The suggested building footprint detection model was tested using datasets from Maghagha areas. The chosen dataset contains a mix of urban variety, including both structured and unstructured roofs. Various roofing materials, shapes, widths, and heights were used on the buildings. The LiDAR data used were collected on March 25, 2015. The system had a 30° scanning angle and a ±15° camera angle. The LiDAR data have an average point density of 7 points/m2 and a point spacing of 0.38 m. Overall, the minimum and maximum elevations of the operating area were 46.83 and 90.5 m, respectively. In the working area, DSM varied from 47.15 to 104.87 m. The raw LiDAR point clouds were used to create two separate products: DEM and DSM. Furthermore, the laser scanning equipment also captured RGB images along with the point clouds. The orthophotos collected had a spatial resolution of 20 cm.

DSM was created using Inverse Distance Weighting IDW interpolation with a spatial resolution of 0.05 m. Meanwhile, DEM was created using the multiscale curvature classification filtering algorithm in ArcGIS (MCC) [20]. This solution has several advantages, including a built-in function in ArcGIS software that simplifies the deployment and allows integration into an automated processing workflow. With a mini-batch size of 2, the models were trained with 20 epochs and a learning rate of 0.0001. All tests were run on an Intel (R) Core i7 3.40 GHz processor with an NVIDIA GeForce GTX 1080-Ti GPU. These parameters were chosen based on their experimentally high accuracy. Because of the limited computational resources, optimizing the training algorithm settings may enhance performance even further.

D. Results and Discussion

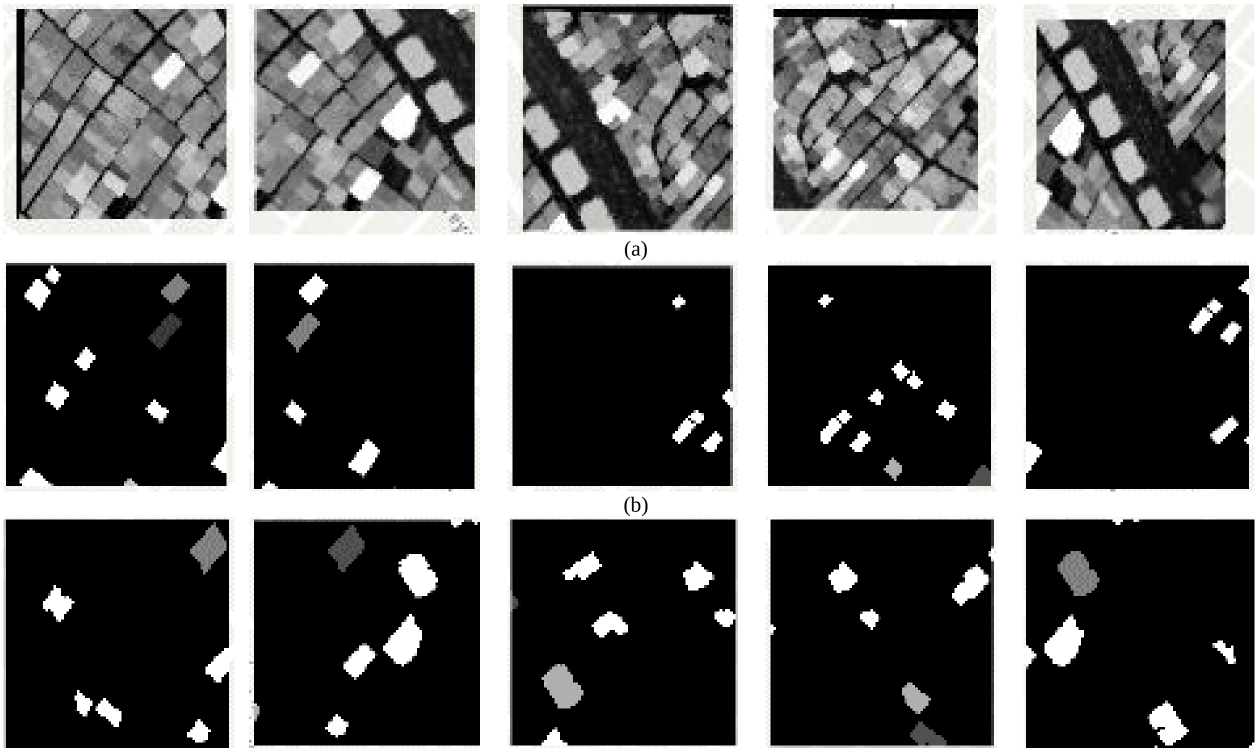
Experiments were conducted in several regions with varying numbers of buildings and roof shapes to demonstrate the detection accuracy of the framework. Fig. 5 shows the visual results. Despite the discontinuous and unclear borders in the DSM pictures, the suggested approach reliably identifies building footprints from highly populated locations. Furthermore, by overcoming the obstacles of location, form, and size, the mask R-CNN approach precisely partitioned the building footprints.

Fig. 6 shows another visual result for building footprints. From the results, the proposed method can accurately localize and segment building footprints under several settings, due to the extraction of a representative set of features by ResNet-34 and the segmentation capabilities of Mask R-CNN. Thus, the localization and segmentation ability may be slightly reduced for samples with large changes in size, particularly in dense regions. Fig. 7 shows a snapshot of results obtained from ResNet101, ResNet 34, and the proposed ensemble. One can observe the outperformance of the proposed ensemble results compared with the other two backbone architectures.

The proposed approach can precisely identify varied shapes of the building footprint with an average accuracy of 0.9463 on the dataset. Furthermore, by overcoming the differences in position, size, and shape, the suggested method can precisely segment regular and nonregular roofs. Evaluation measures (OA, precision, recall, and F-score) were applied to better understand the performance of proposed strategy. Table III shows the outcomes of the proposed approach. The results obtained show average overall accuracy, precision, recall, and F-score of 94.63%, 82%, 97.60%, and 88.46%, respectively.

TABLE III. OBTAINED ACCURACY, PRECISION, RECALL, AND F1- SCORE OF DIFFERENT BACKBONE COMPARED WITH THE PROPOSED

	OA	**Precision**	**Recall**	**F-score**
ResNet34	81%	32.6%	33.18%	31.4%
ResNet101	88.75%	72.19%	70.6%	71.4%
Proposed ensemble	94.63%	82%	97.60%	88.46%

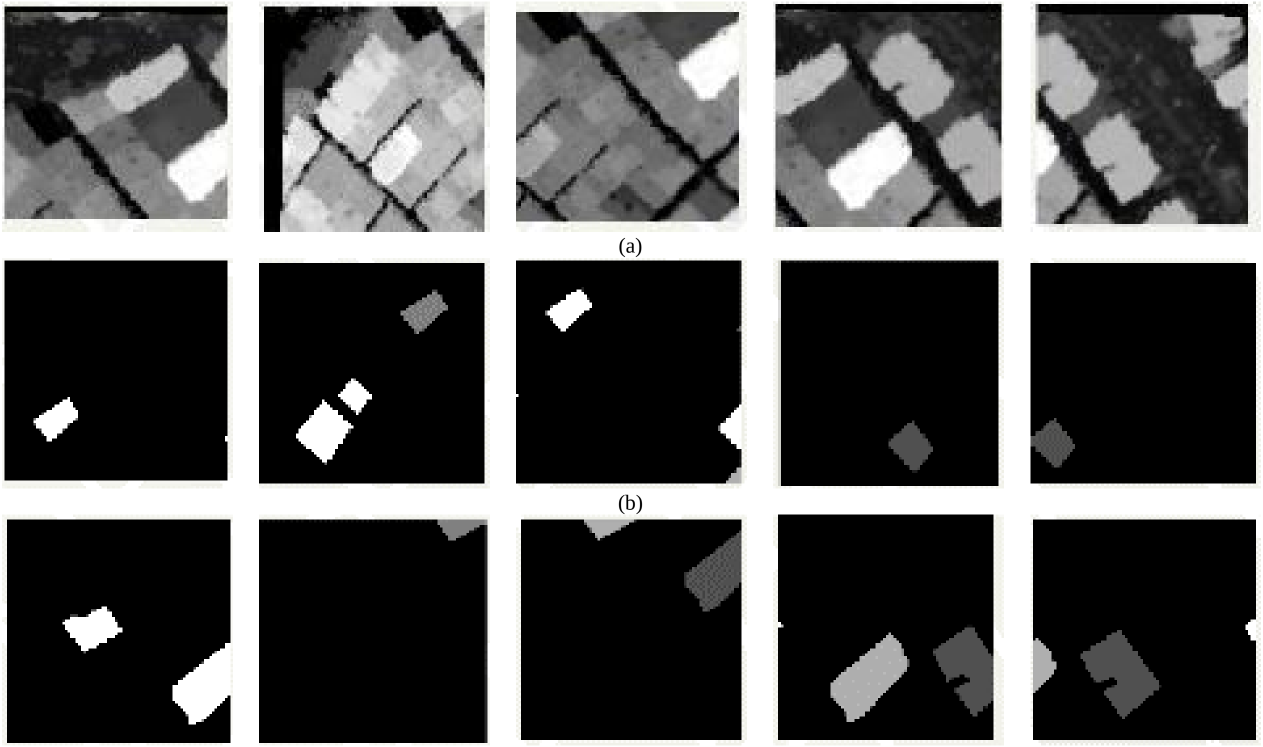


(a)

(b)

(c)

Fig. 5. Visual Results via ResNet-101 Building Footprints Extraction. (a) Input Images. (b) Ground Truth Mask Images. (c) Mask Output Images.

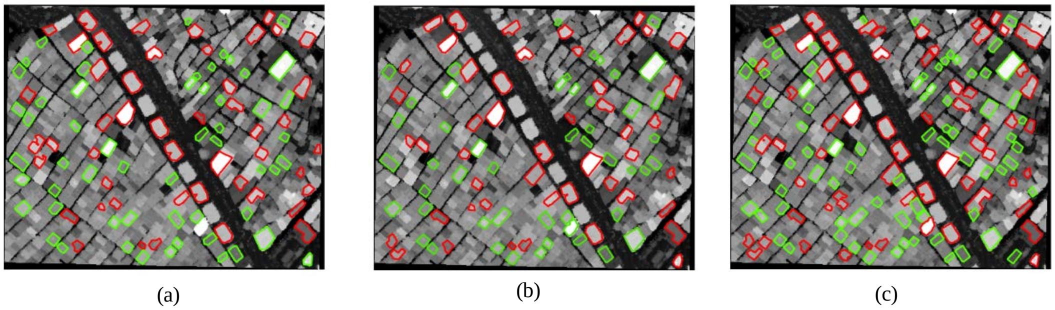


(a)

(b)

(c)

Fig. 6. Visualization Results via ResNet-34 Building Footprint Extraction. (a) Input Images. (b) Ground Truth Mask Images. (c) Mask Output.



(a) (b) (c)

Fig. 7. Visualization Results for Building Footprint Extraction via a) ResNet-101, (b) ResNet-34, (c) Proposed Ensemble (Green and Red Colors Indicate Regular and Nonregular Roofs, Respectively).

V. Conclusion

Buildings are fundamental for urban planning and are essential in the development of a city. The extraction of precise building footprints from remote sensing data has been a topic of consideration. Recently, it has received much attention. Building data are useful in many geospatial applications, including urban planning, risk assessment, 3D city modeling, environmental sciences, and natural disaster damage assessment. Satellite photographs, aerial shots, radar scans, and laser scanning data can all be used to determine the footprint of a building. LiDAR provides a precise and efficient method of getting elevation data, which can be used to extract ground objects such as buildings. The ability to collect high-density point clouds quicker, great vertical precision, and low cost are all advantages of LiDAR over traditional photogrammetry. However, accurate extraction of buildings in urban dense areas with imprecise boundaries is difficult due to the presence of nearby objects. This paper proposed a building footprint extraction model tested using the LiDAR dataset. The study was chosen because the dense rural areas have a mix of urban elements, including both structured and unstructured roofs. Conclusively, the trained building footprint extraction model can detect all structured and unstructured buildings in the LiDAR data. The detected buildings could be saved as a feature layer and used for various data products to derive business value.

References

[1] Guo, Liang, Xingdong Deng, Yang Liu, Hugui He, Hong Lin, Guangxin Qiu, and Weijun Yang. "Extraction of dense urban buildings from photogrammetric and LiDAR point clouds." IEEE Access 9 2021: 111823-111832.

[2] M. Khoshboresh-Masouleh, F. Alidoost, and H. Arefi, "Multiscale building segmentation based on deep learning for remote sensing RGB images from different sensors," Journal of Applied Remote Sensing, vol. 14, no. 3, 2020, p. 034503.

[3] A. S. Mahmoud, S. A. Mohamed, M. S. Moustafa, R. A. El-Khorib, H. M. Abdelsalam, and I. A. El-Khodary, "Training Compact Change Detection Network for Remote Sensing Imagery," IEEE Access, vol. 9, 2021, pp. 90366-90378.

[4] Khoshboresh-Masouleh M, Saradjian MR. Robust building footprint extraction from big multi-sensor data using deep competition network. arXiv preprint arXiv:2011.02879. 2020 Nov 4.

[5] K. Zhang, J. Yan, and S.-C. Chen, "Automatic construction of building footprints from airborne LIDAR data," IEEE Transactions on Geoscience and Remote Sensing, vol. 44, no. 9,2006, pp. 2523-2533.

[6] J. Zhang and X. Lin, "Advances in fusion of optical imagery and LiDAR point cloud applied to photogrammetry and remote sensing," International Journal of Image and Data Fusion, vol. 8, no. 1, 2017, pp. 1-31.

[7] K. R. Adeline, M. Chen, X. Briottet, S. Pang, and N. Paparoditis, "Shadow detection in very high spatial resolution aerial images: A comparative study," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 80, 2013, pp. 21-38.

[8] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700-4708.

[9] Donahue, Jeff, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell."Decaf: A deep convolutional activation feature for generic visual recognition," in International conference on machine learning, 2014, pp. 647-655: PMLR.

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

[11] Q. Shi, X. Tang, T. Yang, R. Liu, and L. Zhang, "Hyperspectral image denoising using a 3-D attention denoising network," IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 12, 2021, pp. 10348-10363.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention, 2015, pp. 234-241: Springer.

[13] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 4799-4807.

[14] M. S. Moustafa and S. A. Sayed, "Satellite Imagery Super-Resolution Using Squeeze-and-Excitation-Based GAN," International Journal of Aeronautical and Space Sciences, vol. 22, no. 6, 2021, pp. 1481-1492.

[15] S. A. Mohamed, A. S. El-Sherbeny, A. H. Nasr, and A. K. Helmy, "A New Image Super-Resolution Restoration Algorithm," International Journal of Computer Applications, vol. 173, 2017, pp. 5-12.

[16] D. Eigen and R. Fergus, "Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 2650-2658.

[17] W. Li, C. He, J. Fang, J. Zheng, H. Fu, and L. Yu, "Semantic segmentation-based building footprint extraction using very high-resolution satellite images and multi-source GIS data," Remote Sensing, vol. 11, no. 4, 2019, p. 403.

[18] J. Xing, Z. Ruixi, R. Zen, D. M. S. Arsa, I. Khalil, and S. Bressan, "Building extraction from google earth images," in Proceedings of the 21st International Conference on Information Integration and Web-based Applications & Services, 2019, pp. 502-511.

[19] D. He, Q. Shi, X. Liu, Y. Zhong, and L. Zhang, "Generating 2m fine-scale urban tree cover product over 34 metropolises in China based on deep context-aware sub-pixel mapping network," International Journal of Applied Earth Observation and Geoinformation, vol. 106, 2022, p. 102667.

[20] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 6, 2016, pp. 1137-1149.

[21] 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.

[22] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779-788.

[23] Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. "Ssd: Single shot multibox detector," in European conference on computer vision, 2016, pp. 21-37: Springer.

[24] J. Kilian, N. Haala, and M. Englich, "Capture and evaluation of airborne laser scanner data," International Archives of Photogrammetry and Remote Sensing, vol. 31, 1996, pp. 383-388.

[25] 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, 2020, p. 1677.

[26] 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, 2015, pp. 295-310.

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

[28] Awrangjeb, Mohammad, Guojun Lu, and C. Fraser. "Automatic building extraction from LiDAR data covering complex urban scenes." The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 40, no. 3, 2014, pp 25.

[29] S. Du, Y. Zhang, Z. Zou, S. Xu, X. He, and S. Chen, "Automatic building extraction from LiDAR data fusion of point and grid-based features," ISPRS journal of photogrammetry and remote sensing, vol. 130, 2017, pp. 294-307.

[30] J.-S. Proulx-Bourque, H. McGrath, D. Bergeron, and C. Fortin, "Extraction of Building Footprints from LiDAR: An Assessment of Classification and Point Density Requirements," in Advances in Remote Sensing for Infrastructure Monitoring: Springer, 2021, pp. 259-271.

[31] T. Tang and L. Dai, "Accuracy test of point-based and object-based urban building feature classification and extraction applying airborne LiDAR data," Geocarto international, vol. 29, no. 7, 2014, pp. 710-730.

[32] S. Zhang, F. Han, and S. M. Bogus, "Building Footprint and Height Information Extraction from Airborne LiDAR and Aerial Imagery," in Construction Research Congress 2020: Computer Applications, 2020, pp. 326-335: American Society of Civil Engineers Reston, VA.

[33] K. Zhang, S.-C. Chen, D. Whitman, M.-L. Shyu, J. Yan, and C. Zhang, "A progressive morphological filter for removing nonground measurements from airborne LIDAR data," IEEE transactions on geoscience and remote sensing, vol. 41, no. 4, 2003, pp. 872-882.

[34] J. Pérez-García, J. Delgado, J. Cardenal, C. Colomo, and M. Ureña, "Progressive densification and region growing methods for LIDAR data classification," International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 39, no. B3, 2012, pp. 155-160.

[35] N. Pfeifer, S. O. Elberink, and S. Filin, "Automatic tie elements detection for laser scanner strip adjustment," International Archives of Photogrammetry and Remote Sensing, vol. 36, no. 3/W3, 2005, pp. 1682-1750.

[36] S. Filin and N. Pfeifer, "Segmentation of airborne laser scanning data using a slope adaptive neighborhood," ISPRS journal of Photogrammetry and Remote Sensing, vol. 60, no. 2, 2006, pp. 71-80.

[37] A. L. Montealegre, M. T. Lamelas, and J. De La Riva, "A comparison of open-source LiDAR filtering algorithms in a Mediterranean forest environment," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 8, no. 8, 2015, pp. 4072-4085.

[38] M. Ghanea, P. Moallem, and M. Momeni, "Automatic building extraction in dense urban areas through GeoEye multispectral imagery," International journal of remote sensing, vol. 35, no. 13, 2014, pp. 5094-5119.

[39] C. Beumier and M. Idrissa, "Digital terrain models derived from digital surface model uniform regions in urban areas," International Journal of Remote Sensing, vol. 37, no. 15, 2016, pp. 3477-3493.

[40] Q.-Y. Zhou and U. Neumann, "Complete residential urban area reconstruction from dense aerial LiDAR point clouds," Graphical Models, vol. 75, no. 3, 2013, pp. 118-125.

[41] M. Awrangjeb and C. S. Fraser, "Automatic segmentation of raw LiDAR data for extraction of building roofs," Remote Sensing, vol. 6, no. 5, 2014, pp. 3716-3751.

[42] M. Awrangjeb, C. S. Fraser, and G. Lu, "BUILDING CHANGE DETECTION FROM LIDAR POINT CLOUD DATA BASED ON CONNECTED COMPONENT ANALYSIS," ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, vol. 2, 2015.

[43] X. Liu, H. Hu, and P. Hu, "Accuracy assessment of LiDAR-derived digital elevation models based on approximation theory," Remote Sensing, vol. 7, no. 6, 2015, pp. 7062-7079.

[44] Fang, Weili, Lieyun Ding, Peter ED Love, Hanbin Luo, Heng Li, Feniosky Pena-Mora, Botao Zhong, and Cheng Zhou. "Computer vision applications in construction safety assurance," Automation in Construction, vol. 110, 2020, p. 103013.

[45] A. Zarea, A. Mohammadzadeh, and M. Valadanzoej, "Extraction and 3D Reconstruction of Buildings Using LiDAR Data and Aerial Image," Journal of Geomatics Science and Technology, vol. 4, no. 3, 2015, pp. 167-186.

[46] I. Lokhat and G. Touya, "Enhancing building footprints with squaring operations," Journal of Spatial Information Science, vol. 2016, no. 13, 2016, pp. 33-60.

[47] S. G. Salve and K. C. Jondhale, "Shape matching and object recognition using shape contexts," in 2010 3rd International Conference on Computer Science and Information Technology, vol. 9, 2010, pp. 471-474: IEEE.

[48] H.-G. Maas and G. Vosselman, "Two algorithms for extracting building models from raw laser altimetry data," ISPRS Journal of photogrammetry and remote sensing, vol. 54, no. 2-3, 1999, pp. 153-163.

[49] Liu, Penghua, Xiaoping Liu, Mengxi Liu, Qian Shi, Jinxing Yang, Xiaocong Xu, and Yuanying Zhang."Building footprint extraction from high-resolution images via spatial residual inception convolutional neural network," Remote Sensing, vol. 11, no. 7, 2019, p. 830.

[50] J. Kang, R. Fernandez-Beltran, X. Sun, J. Ni, and A. Plaza, "Deep Learning-Based Building Footprint Extraction With Missing Annotations," IEEE Geoscience and Remote Sensing Letters, 2021.

[51] Wu, Q., Feng, D., Cao, C., Zeng, X., Feng, Z., Wu, J. and Huang, Z. "Improved Mask R-CNN for Aircraft Detection in Remote Sensing Images," Sensors, vol. 21, no. 8, 2021, p. 2618.

[52] A. Faid and S. Mansour, "Management of Groundwater Reservoir in Maghagh Aquifer System Using Modeling and Remote Sensing Technique (Upper Egypt)," 2006.

[53] E. van Rees, "Trimble's AX60i and AX80," GeoInformatics, vol. 17, no. 5, 2014, p. 36.