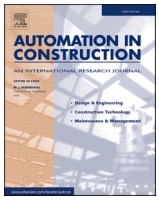
Automation in Construction 159 (2024) 105258

Contents lists available at ScienceDirect

Automation in Construction

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



Computer vision based first floor elevation estimation from mobile LiDAR data

Jiahao Xia, Jie Gong *

Department of Civil and Environmental Engineering, Rutgers, the State University of New Jersey, NJ, USA

ARTICLE INFO

Keywords:
First floor elevation
Computer vision
Mobile LiDAR
Flood risk and mitigation

ABSTRACT

First Floor Elevation (FFE) of a house is crucial information for flood management and for accurately assessing the flood exposure risk of a property. However, the lack of reliable FFE data on a large geographic scale significantly limits efforts to mitigate flood risk, such as decision on elevating a property. The traditional method of collecting elevation data of a house relies on time-consuming and labor-intensive on-site inspections conducted by licensed surveyors or engineers. In this paper, we propose an automated and scalable method for extracting FFE from mobile LiDAR point cloud data. The fine-tuned yolov5 model is employed to detect doors, windows, and garage doors on the intensity-based projection of the point cloud, achieving an mAP@0.5:0.95 of 0.689. Subsequently, FFE is estimated using detected objects. We evaluated the Median Absolute Error (MAE) metric for the estimated FFE in Manville, Ventnor, and Longport, which resulted in values of 0.2 ft, 0.27 ft, and 0.24 ft, respectively. The availability of FFE data has the potential to provide valuable guidance for setting flood insurance premiums and facilitating benefit-cost analyses of buyout programs targeting residential buildings with a high flood risk.

1. Introduction

It is widely recognized that flooding is among the most prevalent natural disasters, and it is not only the deadliest but also the most costly disaster on Earth [1]. Due to the climate change, the water levels around the world's coasts is raising and the flood risk will inevitably increase in these areas [2]. Moreover, the changes in Land Use and Land Cover (LULC), infrastructure and population demographics are also among the leading factors that damaging floods are observed increasing in severity, duration and frequency in recent decades [3]. From 2000 to 2015, the total population in the inundation zone observed using satellite data grew by 58–86 million, and the proportion of the population exposed to floods is expected to increase further based on the climate change projection for 2030 [4]. Research indicates that the absolute damage loss from floods could rise by a factor of 20 by the end of the century along the global socioeconomic development if in the absence of risk mitigating measures [5]. The compounded effect of these driving forces poses a challenge to understanding the causes, consequences, and mitigation strategies associated with flooding events.

To diminish the effect of flood hazards on infrastructure assets and enhance flood resilience, it is necessary to implement flood management actions, particularly in communities that are vulnerable to flooding and hurricanes [6]. Flood management can be divided into four phases: mitigation, preparedness, response, and recovery [7]. The measures used in flood management can be classified into two categories: structural and non-structural ones. For example, the structural measures employed during the mitigation stage encompass activities such as elevating flood-prone properties, implementing buyout programs, or facilitating the relocation of affected communities. Flood insurance, specifically the National Flood Insurance Program (NFIP) managed by the Federal Emergency Management Administration (FEMA) in the United States, is a vital non-structural method aimed at reducing the socio-economic impact of floods during the preparedness stage. Lowest Floor Elevation (LFE) of a building is an essential information for the measures utilized in the full phases of the flood management. According to the description of FEMA, the lowest floor refers to the lowest enclosed living area (including basement) other than building access, parking, or storage [8]. In the newly implemented Risk Rating 2.0 program, among which the risk-based premiums are introduced and can yield a positive societal benefit, First Floor Elevation (FFE) is used as one of the factors to calculate the insurance rates [9,10]. The flood vulnerability analysis of individual buildings can be conducted through comparing the FFE

* Corresponding author at: Department of Civil and Environmental Engineering, Rutgers, the State University of New Jersey, NJ 08854, USA.
E-mail address: jg931@soe.rutgers.edu (J. Gong).

https://doi.org/10.1016/j.autcon.2023.105258
Received 21 July 2023; Received in revised form 13 December 2023; Accepted 27 December 2023
Available online 13 January 2024

with the Base Flood Elevation (BFE, i.e., the elevation of surface water resulting from a flood that has a 1% chance of equaling or exceeding that level in any given year) which can be obtained from the Flood Insurance Rate Map (FIRM).

The absence of reliable FFE data on a large geographic scale, such as boroughs or states, poses significant limitations for flood management. FEMA actively promotes the use of Elevation Certificates by communities to document elevation information for properties and demonstrate compliance with floodplain management ordinances. The Elevation Certificates can serve as a valuable reference for property owners when applying for flood insurance. The traditional method of collecting elevation information for Elevation Certificates relies on on-site inspection, which is typically conducted by a licensed surveyor or engineer. According to FEMA, the vertical accuracy of the Elevation Certificates surveyed by GPS can be accurate up to 0.5 ft. with a 95% confidence level [11]. However, the manual inspection process for collecting elevation data is both time-consuming and labor-intensive. According to the Elevation Certificate and Instructions provided by FEMA, it is estimated that an average of 3.75 h is required per elevation certificate to gather all the necessary elevation information. Consequently, scaling this process for large areas poses significant challenges. Addressing this data gap is crucial for advancing flood management practices and reducing the potential impacts of flooding.

Extracting FFE information from public geospatial data sets is a promising approach in addressing the gap in building structural elevation data. For instance, several studies have focused on extracting FFE information from the publicly accessible Google Street View (GSV) images with computer vision methods [12,13]. However, the spatial accuracy of the GSV images is limited [14], particularly when compared to elevation certificates. Another concern with GSV images is that they often become outdated in floodplain areas due to constant rebuilding and renovation activities. Other public geospatial data sets such as airborne LiDAR have also been utilized to estimate FFE information [15]. However, due to the angles of observations, airborne LiDAR cannot capture information about building facades. Consequently, estimating FFE from airborne LiDAR relies on proxy measures, such as ground elevations, which can only provide FFE estimates with limited accuracy. Researchers have also experimented with capturing FFE data with drone-based mapping and infrared thermography [16,17]. Nevertheless, these methods cannot be scaled up to encompass a large number of buildings due to either the requirement of manual interpretation or the very limited sample sizes. It is possible to capture survey-grade LiDAR data with UAV LiDAR. But due to the constraints of batteries, LiDAR systems used on drones tend to have lower spatial resolutions than those used on vehicular platforms. Additionally, battery and fly zone restrictions often limit the area that can be covered by drone-based survey grade mapping systems. It is also important to note that most current studies on extracting FFE information from pubic geospatial data sets have confined their study areas to the block or street level, and their proposed methods have been validated primarily on residential buildings with homogeneous architectural styles [13,17,18].

Mobile LiDAR technology provides a scalable solution to rapidly scan building facades at the street level, providing more detailed building façade information than what can be obtained with airborne and UAV LiDAR system. With a tactile inertia navigation system and high accuracy laser scanners, mobile LiDAR can achieve survey-grade spatial mapping accuracy even at driving speeds. The resulting point cloud data can be used to generate 3D digital elevation models of buildings across large communities. This provides survey-grade mapping data that has the potential to support the extraction of FFE information on a city-wide scale. In a previously published study, we demonstrated the feasibility of manually extracting FFE information from large mobile LiDAR point cloud data, achieving accuracy comparable to traditional survey methods [19]. However, the substantial volume of point cloud data, often encompassing hundreds of millions or even billions of points containing multiple attributes (e.g., coordinates, intensity, GPS time, and even color), poses challenges for manual FFE information extraction across extensive areas. Additionally, occlusions resulting from line-of-sight issue create gaps in point cloud data coverage, significantly complicating the process of FFE information extraction.

As a sub-field of artificial intelligence, computer vision aims at enabling machines or artificial systems automatically interpret visual inputs (e.g., images, videos and point cloud) and derive meaningful information for decision making [20–25]. Benefiting from the availability of large-scale visual datasets across various computer vision tasks and advancements in hardware, such as Graphics Processing Units (GPUs), deeper neural networks can be trained to achieve state-of-the-art performance, and some of these networks have demonstrated the ability to outperform the human visual system. However, despite these advancements, research on extracting FFE information from mobile LiDAR data with AI methods is still very limited. This scarcity could be attributed to several factors. First, there is a shortage of annotated mobile LiDAR data for segmentation and recognition tasks. Second, zero-shot learning based on large-scale foundational models is predominantly trained on data that differs significantly from mobile LiDAR data. Third, the availability of mobile LiDAR data for extensive flood-related studies is not always guaranteed for large areas. The convergence of these factors has resulted in an underexplored area that nevertheless holds immense potential to enhance floodplain management.

This study focuses on the design and evaluation of computer vision based FFE information extraction from large-scale mobile LiDAR point cloud data. It addresses two key questions:

(1) How can data representations derived from mobile LiDAR data can be optimized to enhance the utilization of foundational computer vision models and reduce the reliance on annotated training data?
(2) How to design and train machine learning models to reliably and accurately extract FFE information in communities characterized by wide variety of building typologies?

More specifically, this paper outlines a method that integrates building façade detection, point cloud projection, and deep learning based computer vison methods to extract first floor elevation from massive mobile LiDAR data sets. The method underwent rigorous testing and evaluation on mobile LiDAR data collected from three communities. The main contributions of this work include:

- Generation of effective data representations from mobile LiDAR data to facilitate deep learning based FFE information extraction
- Design of an automated and scalable approach for extracting FFE from mobile LiDAR point cloud data, demonstrating the capability to achieve sub-feet accuracy in mean absolute error
- Creation of ground truth FFE data for three representative communities, each facing diverse flooding threats and featuring different types of building typologies. This ground truth data was derived from elevation certificates and manual annotation, providing valuable insights into the uncertainty associated with the proposed method.

The rest of the paper is organized as follows. Section 2 introduces the relevant studies on extracting elevation information of buildings, point cloud representation and 2D object detection. Section 3 covers the study area introduction, datasets, and methodology. Section 4 shows the results of the proposed methodology and analysis of the estimated FFE datasets.

2. Literature review

2.1. Studies on the extraction of building elevation information

The methodology of estimating building elevation information can be classified into two paradigms: (1) regression-based method and (2)

detection-based method. Gordon [18] developed statistical regression model between FFH (First Floor Height) and selected building attributes (i.e., foundation type, DEM value, difference in grade, year built and flood zone), and then calculated the FFE (First Floor Elevation) through combining predicted FFH and LAG (Lowest Adjacent Grade which means the elevation of the ground next to the building). As the building attributes data can be easily obtained from the tax assessor database for most buildings, the regression model can fill the gap where elevation certificate is missing. Of particular concern is the generalization of the statistical model. The relationship between the FFH and building attributes can differ depending on geographical features and architectural styles, for example, the houses are typically built at an elevated height in coastal communities, whereas it is uncommon in inland communities. In the work of Gordon [18], only tabular data is used to build the regression model. Visual data which contains structure appearance information of the buildings has the potential to enhance the model's ability to generalize. The detection-based methods utilize remote sensing data in different modalities, such as 2D images, infrared thermal images and 3D point cloud, to detect components related to the building's elevation. Ning, et al. [12] trained an object detection model to detect door from the Google Street View (GSV) imagery and then calculate the bottom elevation of the door (i.e., First Floor Elevation, FFE) using the roadway elevation and the height of the camera. Needham [13] estimated FFE by using Google Earth and Google Street View, and they converted the vertical pixel distance between the ground and first floor to the real length and then calculate the actual real-world elevation. Diaz, et al. [17] reconstructed the 3D model of the residential communities using the images collected by drones and georeferenced the model with ground control points (GCPs), and finally manually label the elevation (i.e., FFE, LAG) on the 3D model. Point cloud can provide accurate geometry information and Haghighatgou, et al. [26] detect buildings' lowest openings in rural areas from the point cloud collected by Mobile Laser Scanner (MLS). The limitations of current buildings' elevation related studies can be summarized as follows: (1) limited to small area, such as street or block level; (2) most studies only utilized 2D images (i.e., GSV, Google satellite images, UAV images) which have limited spatial measurement accuracy [14]; (3) dependent on manual interpretation and not possible to scale up the process of extracting elevation information for large areas.

2.2. 2D object detection

2D object detection focuses on classifying and localizing objects in 2D images, and it serves as a basis for many higher-level computer vision tasks (i.e., object tracking, instance segmentation, and image captioning) [27]. As a mature technique, object detection has greatly benefited from the advancements in deep learning and has found widespread application in the real world, such as face recognition, robot vision, and video surveillance. Horizontal Bounding Box (HBB) which describes each object with a horizontal rectangle is the most used object representation in 2D object detection, while Oriented Bounding Box (OBB) which can precisely localize oriented object is more appropriate for images taken from the spaceborne or airborne platforms [28]. In the early days prior to the widespread use of deep learning, traditional object detection methods typically involved the following three steps [29]: (1) select informative regions which can be generated with multiscale sliding window; (2) extract visual features which are usually hand-crafted features, such as Scale-Invariant Feature Transform (SIFT) [30] and Histograms of Oriented Gradients (HOG) [31]; (3) leverage a classifier, such as Supported Vector Machine (SVM) [32], to recognize the object. Benefited from the advancement of computation hardware (e.g., GPUs) and the availability of large annotated datasets, such as Microsoft COCO [33], Deep Neural Networks (DNNs) with strong generalization ability have been trained successfully to improve all the computer vision tasks. The DNN-based object detection methods can be divided into two groups: (1) one-stage based method and (2) two-stage based method. The two-stage detectors typically have two stages, and the first generates Region of Interest (ROI) and then estimate the object class and bounding box. The representative two-stage detectors are R-CNN series, such as Fast-RCNN, Faster-RCNN, Mask R-CNN, and they can obtain highly accurate detection results [34–37]. However, the high latency limits the two-stage detectors in time-sensitive applications. One-stage detectors directly learn the object class and bounding box in a single pass, resulting in faster inference speeds compared to two-stage detectors. You Only Look Once (YOLO) series detectors are the most widely used one-stage object detection method, and the dedicated designed network architectures, training sample assignment, and loss functions improve the detection accuracy of the one-stage detectors [38,39]. In recent years, the Transformer architecture which is built on attention mechanisms shows its power in Nature Language Process (NLP), and it's also transferred to Computer Vision areas [40]. The Transformer-based detector, such as DETR [41], can realize the object detection in end-to-end way without using non-maximum suppression (NMS) as a post-processing.

2.3. Point cloud representations

Point cloud, as an important 3D data structure, can describe the accurate geometric information of the real world. Different from the image which has a regular format and can be represented using a matrix, point cloud is a set of coordinates and point attributes (i.e., color, intensity, GPS time). The unordered nature of the data poses a challenge to processing point cloud using DNNs, as the architecture represented by Convolutional Neural Network (CNN) is primarily designed to operate on structured grid-like data. Different representation formats of point cloud have been developed to process collections of 3D points for tasks, such as 3D shape classification, 3D object detection, multi-object tracking and segmentation [42]. The representation formats of point cloud can be divided into three groups: (1) projection; (2) voxel and (3) point-set [43]. Projection-based methods offer an intuitive approach to process point clouds by converting them into view projections (such as bird's eye view, front view) from a specific perspective. This transformation allows leveraging off-the-shelf 2D computer vision algorithms for further analysis and processing. By projecting the 3D points onto a 2D plane, these methods enable the utilization of well-established techniques developed for image-based analysis. Li, et al. [44] project the 3D point cloud into a 2D projection map and detect vehicles using fully convolutional network. While some studies observe the point cloud from a top-down perspective, for example Saleh, et al. [45] detect vehicles on the bird's eye view generated from point cloud. Chen, et al. [46] combines multiple views generated from point cloud for 3D object detection. One limitation of the projection-based methods is that 3D—2D projection will result in information loss (i.e., depth information in front view and height information in bird's eye view). Unlike projection-based methods that convert 3D points to 2D images, voxel-based methods divide 3D space into regular 3D voxels. 3D CNN is used to learn features from the predefined voxels and subsequently perform the downstream tasks, such as 3D object detection [47]. The voxelization process, which involves merging multiple points within the same voxel, can lead to information loss. Additionally, when dividing a large 3D space into smaller voxels, the computation and memory requirements can become burdensome. To utilize all the information contained in the 3D point cloud, many algorithms were developed to directly consume point cloud data. They focus on dealing with the unordered nature of the point cloud, for example PointNet uses a symmetry function for the unordered inputs, while some new model architectures are particularly appropriate for processing unordered data (i.e., graph CNN, self-attention operator in Transformer) [48–50].

3. Material and methods

3.1. Study area and datasets

Ninety percent of natural disasters within the United States involve flooding and both the frequency and damage cost of floods are rising according to the reports from the National Oceanic and Atmospheric Administration (NOAA) [51]. As illustrated in Fig. 1, one inland borough and two coastal cities are selected as study areas in this paper, and all the study areas are in or near the flood zones defined by FEMA. Manville is an inland borough with a total area of 2.45 mile^2 in Somerset County, in the U.S. state of New Jersey (NJ), and estimated population is 10,875 in 2022 according to the United States Census Bureau. Manville, an inland town, experiences frequent riverine flooding, with approximately 490 structures within the 1% annual chance of exceeding the floodplain. Both Ventnor and Longport are coastal cities located in Atlantic County, New Jersey, along the Jersey shore, and they are susceptible to tidal and hurricane storms. According to data from the United States Census Bureau, Ventnor spans a total area of 3.52 mile^2, while Longport covers an area of 1.56 mile^2. In terms of population, the estimated number of residents in Ventnor was 9246 in 2022. Similarly, Longport had an estimated population of 884 in the same year. The building stocks in these communities are varied to a great extent. Manville is a working-class town where almost all homes are primary dwellings, Longport, on the other hand, has been completely rebuilt with million-dollar homes after Hurricane Sandy and has most secondary/vacation homes. Ventnor has a mix of primary and secondary residences. Together they provide a comprehensive mix of building types susceptible to floods.

A Mobile Mapping System (MMS) was employed to collect point

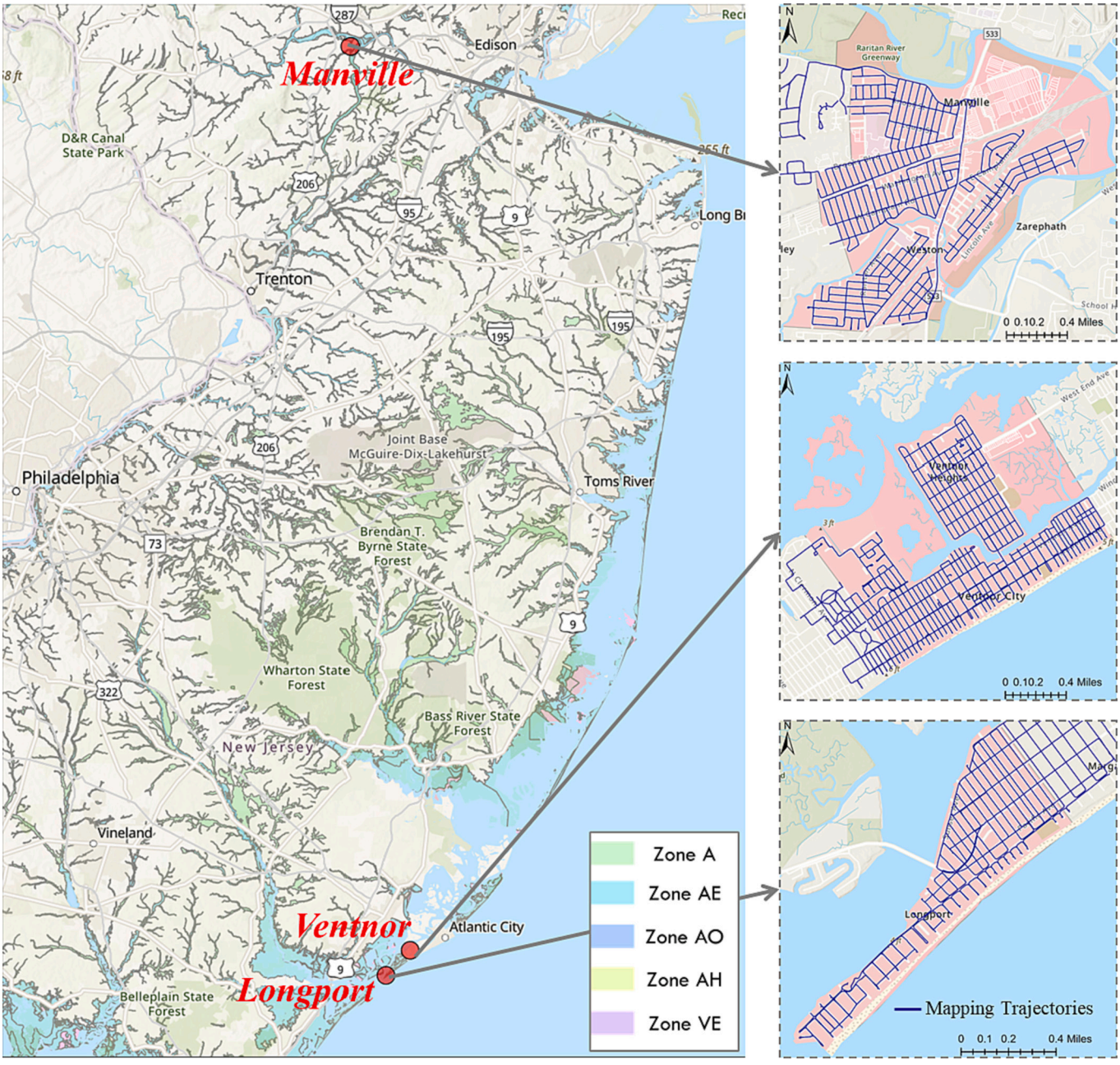


Fig. 1. Study area. Flood zones come from the Flood Insurance Rate Map (FIRM) of FEMA. Zone A, AE, AO, AH and VE are all identified as flood hazard areas with high risk.

cloud of the selected study area, and the mapping trajectories are displayed on Fig. 1. Most residential buildings in the three cities were covered by the mapping tasks. The rotation speed of the laser scanner (i.e., Z + F PROFILER® 9012) was set at 100 revolutions per second, allowing for the collection of over 1 million points per second. Operating at an approximate speed of 20 miles per hour, the MMS facilitated the efficient acquisition of high-quality and dense 3D data of the surrounding environment. Table 1 provides a summary of the data, indicating that they were collected on four different dates, primarily during the summer months. In the case of Manville, some areas were not covered during the initial mapping right after Hurricane Ida in 2021 due to road accessibility issues. Consequently, an additional scanning was conducted the following year. The total length of mapping trajectories for Manville amounted to 63.4 miles. Longport, being the smallest city among the three study areas, required five trajectories with a combined length of 27.1 miles to cover the community adequately. Ventnor, as the study site with the largest area, necessitated 12 trajectories spanning a total length of 70.6 miles.

3.2. Methodology

The overall framework of the proposed method for the automated and scalable extraction of FFE is depicted in Fig. 2. The process involves several key steps. Firstly, the point cloud is clipped for each building based on parcels data, and subsequently projected onto a Bird's Eye View. Next, building facades are detected and projected onto the front view, which includes an intensity map and an elevation map. To identify specific building components (i.e., window, door, and garage door), a yolov5 model is trained using manual annotations from the intensity view. Finally, the bottom of the front door and the information from elevation map are utilized to calculate the FFE. Further details regarding the framework and its implementation will be elaborated in the subsequent sections of this paper.

3.2.1. Pre-processing of point cloud

Although the point cloud data contains multiple attributes, this study focuses solely on utilizing the coordinates and intensity information. The point cloud data is represented by an unorder set $\{(X_i, Y_i, Z_i, I_i)\}$, where $i = 1, 2, \ldots, N$. Here, X_i, Y_i, Z_i, and I_i denote the coordinates and intensity of the ith point in the set. The parcels data, an important geographic information system (GIS) dataset, enables spatial analysis and visualization. The parcels utilized in this research were developed during the Parcels Normalization Project in 2008–2014 by the NJ Office of GIS (NJOGIS). The coordinates of both point cloud and parcels were transformed into the New Jersey State Plane Coordinate System, NAD83, with units of measure are in feet. The point cloud generated along the mapping trajectory is clipped for each building based on parcels data, and it enables the extraction of building-specific information and facilitates subsequent analysis.

3.2.2. Building facades detection

In most cases, the elevation measurement of the bottom of the front door is used as FFE. To ensure accurate identification of key components and minimize the influence of non-relevant points, a simple yet effective projection-based method is employed to extract building facades. The BEV representation of 3D point cloud is encoded by density. The point cloud is projected onto XY plane and discretized into 2D grid with a resolution of 0.05 ft. Each cell within the BEV projection stores the number of points and a fixed threshold of 30 is applied to generate a binary mask that highlights the facades. To determine the orientation of building facade, a Gaussian Mixture Model is employed to detect lines on the binary mask derived from the BEV density map. The orthogonal direction of the detected line, denoted as θ, indicates the orientation of the building facade. This orientation angle θ is subsequently utilized in the projection of the 3D point cloud onto the front view. For buildings with multiple recorded facades, typically two orthogonal facades, two orientation angles are utilized to generate two distinct front views for the same building.

Table 1
Information of the mapping trajectories.

City	Date	Number of trajectories	Length of Trajectories (mi)
Manville	20,210,906	9	52.9
	20,220,705	3	10.5
Ventnor	20,210,811	12	70.6
Longport	20,210,809	5	27.1

3.2.3. Front view representation

When projecting the 3D point cloud of the detected facade onto the front view, a rotation is applied along the *Z*-axis using the orientation angle θ. This rotation aligns the building facade with the XZ plane. As illustrated in formula (1), the front view representation of 3D point cloud is encoded based on intensity and elevation information with a resolution of 0.05 ft, where $View_{intensity}$, $View_{elevation}$, $N_{i,j}$ are the intensity-based front view, elevation-based front view and number of points within cell (i, j). Each pixel (i, j) is represented by the average intensity and the Z coordinate of all the points within that cell.

$$View_{intensity}(i,j) = \frac{1}{N_{i,j}} \sum_{k=1}^{N_{i,j}} I_k \tag{1}$$

$$View_{elevation}(i,j) = \frac{1}{N_{i,j}} \sum_{k=1}^{N_{i,j}} Z_k$$

3.2.4. Detection of building components

A fine-tuned yolov5 model is utilized to detect key building components (i.e., window, door and garage door) related to the FFE on the intensity-based front view [52]. Yolov5 is a state-of-the-art one-stage 2D object detector known for its high accuracy and efficiency in detecting target objects. The choice of model size and image size are important considerations for achieving optimal training results. The yolov5 model comes in different sizes, namely Nano (n), Small (s), Medium (m), Large (l), and extra-large (xl). Larger models generally yield better detection results but require more GPU memory during training and lead to slower inference speeds. Training at a higher resolution benefits the detection of small objects. For this study, all training is based on models pre-trained on the COCO dataset, with the selected input image sizes of 640 and 1280. The objects in the COCO dataset have a large pixel coverage, while the target objects in this paper, such as windows and doors, are relatively small compared to the entire building facade. To address this, a slicing-aided hyper inference and fine-tuning approach is employed to enhance the performance of the yolov5 detector for small objects [53]. The intensity-based front view images are sliced into 256*256 patches for fine-tuning and inference. The tool LabelImg is used for manual annotation of selected objects on the intensity-based front view, and all images are randomly selected from Ventnor city for labeling [54]. Examples of manually annotated building components are illustrated in Fig. 3. The labeled bounding boxes exclude window and door frames, and we treat separate windows as distinct instances. In total, 1216 intensity images were labeled, with 80% were used as the training set, 10% for validation and the remaining 10% for testing. Four metrics, namely Precision (P), Recall (R), mAP@0.5, and mAP@0.5 : 0.95, are utilized to evaluate the performance of the detector. Precision and Recall are calculated using formula (2), where TP, FP, and FN represent True Positives, False Positives and False Negatives, respectively. The reported precision and recall are derived through maximizing F1 score. The mAP metrics are used by COCO, and mAP@0.5 represents the mean average precision at Intersection over Union (IoU) of 0.5 over all categories, while mAP@0.5 : 0.95 denotes averaged precision at 10 IoU thresholds, from 0.5 to 0.95 with an interval of 0.05, over all categories.

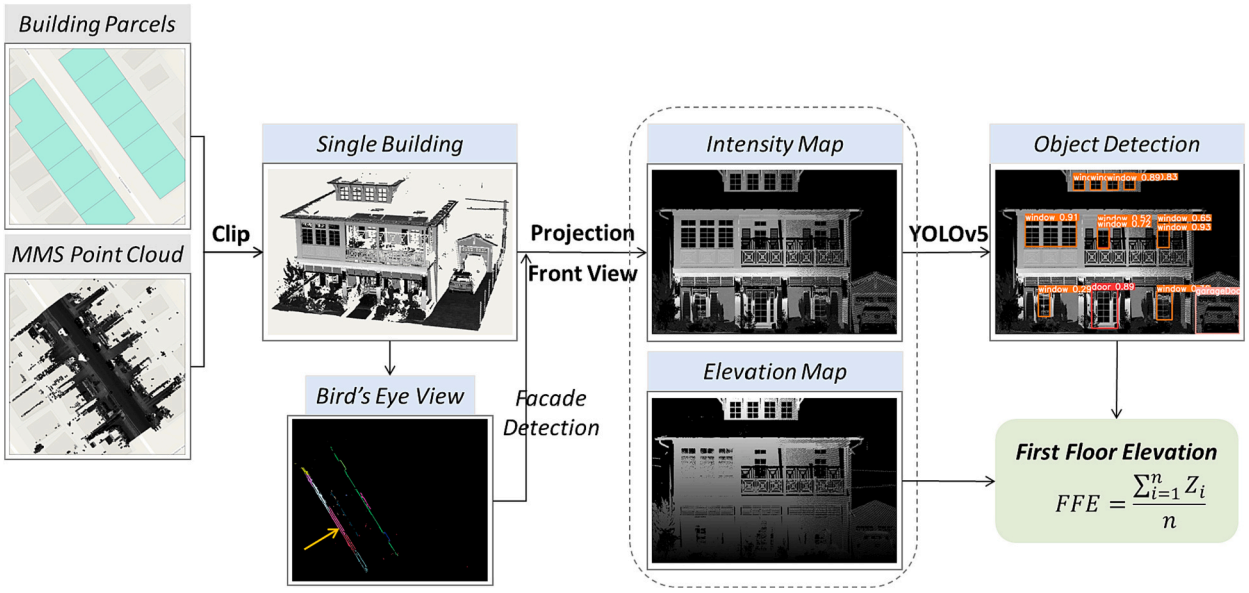


Fig. 2. The framework of the proposed method for the automated and scalable extraction of FFE based on mobile point cloud data.

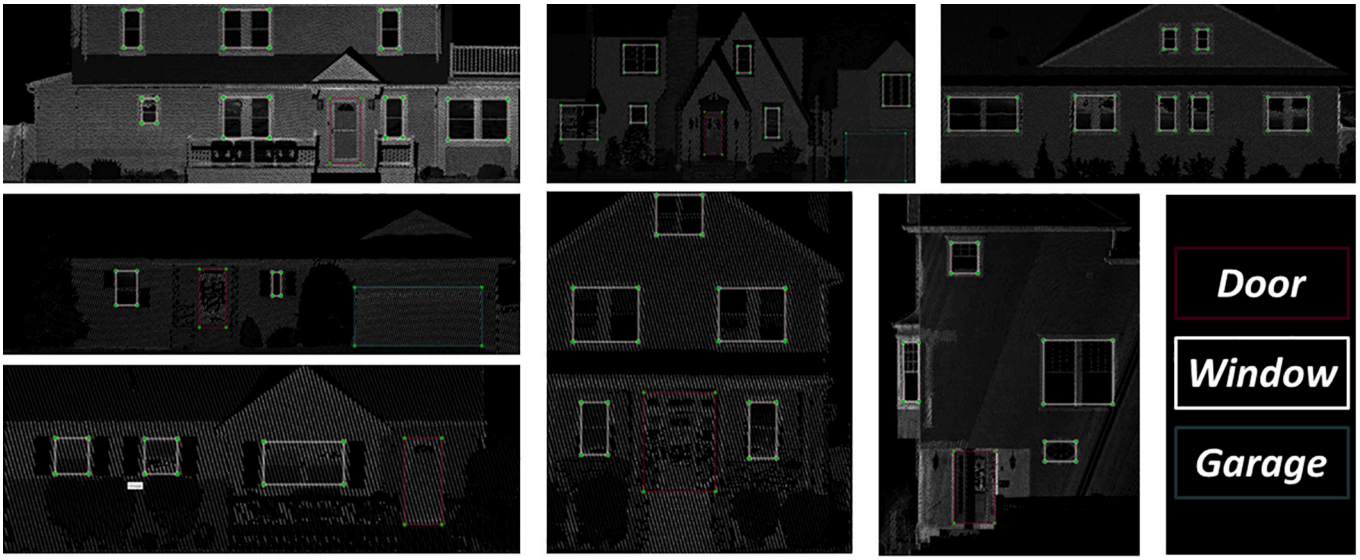


Fig. 3. Examples of manually annotated building components.

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$R = \frac{TP}{TP + FN}$$

$$F1 = 2 \times \frac{P \times R}{P + R}$$

3.2.5. Estimation of FFE

The intensity-based front view and elevation-based front view in this study are derived from the same 3D point cloud data. Consequently, the bounding boxes of the target objects detected in the intensity images can be directly mapped to the corresponding locations in the elevation images. The calculation of the FFE is performed using formula (3), where n represents the number of pixels with valid elevation values along the bottom line of the bounding box of the door, and Z_i represents the elevation value for the ith pixel. To mitigate the potential impact of misdetections of doors on floors higher than the first floor, a rule-based post-processing method is employed. The pseudo code for this post-processing method is provided in Table 2. The variables $Elev_{door}$, $Elev_{window}$, and $Elev_{facade}$ denote the lowest elevation of detected doors, windows and building facade, respectively. These values are calculated using the same formula as the FFE. The threshold value of 9 ft is utilized, which corresponds to the average height of the first story measured from the exterior of residential buildings in Ventnor city. This measurement typically indicates the distance from the bottom of the door to the ceiling based on the point cloud data.

$$FFE = \frac{\sum_{1}^{n} Z_i}{n} \tag{3}$$

Table 2
Pseudo code of rule-based post-processing method.

IF $Elev_{door} - Elev_{window} > 9ft$ AND $Elev_{window} - Elev_{facade} < 9ft$:
$Elev_{door}$ is identified as door on floors higher than the first floor and is removed.

The accuracy of the estimated FFE data is evaluated using two types of ground references: (1) manually annotated FFE data and (2) FFE data obtained from local Elevation Certificates records. The former involves the extraction of FFE data by directly labeling the bottom point of the front door in the 3D point cloud view. Labeling in the 3D space presents greater challenges compared to 2D image annotation. To facilitate this task, a web-based labeling tool is developed based on the Potree library [55]. This tool enables multiple human annotators to work on the point cloud annotation simultaneously. Elevation Certificates, particularly their digital records, are not readily available for most communities. 145 validated Elevation Certificates records in Longport city are provided by local communities [56]. Three metrics, namely Root Mean Square Error (RMSE), Median Absolute Error (MAE), and recall of FFE whose absolute error is lower than 1 ft, are calculated between the estimated FFE and ground reference to evaluate the performance of the proposed method.

4. Results and discussions

4.1. Accuracy evaluation of the object detection

All models underwent fine-tuning for 300 epochs using pretrained weights from the COCO dataset. The fine-tuning process was performed on a machine equipped with 6 NVIDIA RTX5000 GPUs. The accuracy of the models was evaluated on the test set, and the evaluation results are presented in Table 3. Generally, larger models tend to exhibit better performance on the test set. Among the four model size candidates, the large (l) model demonstrates the highest accuracy. For instance, when considering an input image size of 640, all four metrics of the large model surpass those of the other models. Interestingly, the extra-large (xl) model with an input image size of 1280 does not outperform the large model, which may be attributed to the size of the training samples. Insufficient availability of high-quality training samples could lead to the underfitting of a more complex model. When comparing models of the same size, larger image sizes tend to yield improved performance. This improvement is particularly evident for the small (s) model, which exhibits a 0.064 increase in mAP@0.5:0.95. The accuracy of detection takes precedence over inference speed in this study, therefore, the large model with an image size of 1280 is selected for the subsequent detection of target objects and FFE estimation.

In order to improve the detection of small objects, slicing aided fine-tuning was applied to the selected yolov5 large model with an image size of 1280. The accuracy of the model with slicing aided fine-tuning, evaluated on the test set, is presented in Table 4. Among the three object categories, the door category poses the greatest challenge to the detector, with lower accuracy compared to the other two categories. Considering the mAP@0.5 metric, the window category (0.834) and the garage door category (0.925) exhibit significantly higher accuracies than the door category (0.756). The mAP@0.5:0.95 metric is more demanding and reflects the detector's localization ability. With slicing aided fine-tuning, the overall mAP@0.5:0.95 for all classes improves by 0.018. Analyzing the individual categories, both the door and window categories show increases in mAP@0.5:0.95, with the window category achieving a substantial improvement of 0.05. However, the garage door category experiences a slight decrease in mAP@0.5:0.95 (0.006). This could be attributed to the fact that garage doors are relatively larger than doors and windows, making them more prone to being fragmented during the slicing process.

Table 3
Accuracy of yolov5 models with different model size and input image size.

Image size	Model size	Precision	Recall	mAP@0.5	mAP@0.5:0.95
640	s	0.750	0.789	0.787	0.582
	m	0.694	0.792	0.781	0.603
	l	0.740	0.817	0.825	0.648
	xl	0.800	0.766	0.806	0.627
1280	s	0.806	0.789	0.819	0.646
	m	0.854	0.747	0.812	0.627
	l	0.789	0.845	0.842	0.671
	xl	0.761	0.798	0.810	0.658

Table 4
Accuracy of yolov5 large model with an image size of 1280 using slicing aided fine-tuning. The number in parentheses represents the difference achieved by using slicing aided fine-tuning compared to not using it.

	Precision	Recall	mAP@0.5	mAP@0.5:0.95
All	0.775	0.838	0.838	0.689 (+0.018)
Door	0.870	0.687	0.756	0.516 (+0.011)
Garage door	0.675	1.000	0.925	0.867 (−0.006)
Window	0.781	0.827	0.834	0.686 (+0.050)

The object detection samples obtained from the yolov5 large model, utilizing slicing aided fine-tuning with an image size of 1280, are showcased in Fig. 4. Residential buildings exhibit a wide range of visual appearance characteristics, including split-level structures, varying numbers of floors, and elevated components. The target objects, such as doors, windows, and garages, present challenges for the object detection algorithm due to their diverse sizes, colors, materials, and positions. In comparison to the optical Google Street View images, intensity-based front views projected from 3D point cloud data have certain disadvantages, namely sparsity and a lack of rich texture information such as color. Nevertheless, with effective fine-tuning, the model can accurately classify and localize windows, doors, and garages in the front view projections of most residential buildings. For certain mansions, particularly in coastal areas, such as the sixth example shown, the presence of doors on the second floor may pose challenges for accurately estimating the FFE. Another difficulty lies in distinguishing exterior glass doors, as they can closely resemble windows in the intensity-based front view. Additionally, the laser scanner's light signal cannot penetrate foreground objects like vegetation, leading to black holes in the point cloud projection, as observed in examples 7 and 8. Detecting objects concealed behind vegetation becomes extremely challenging under such circumstances. It can probably reduce the impact of vegetation occlusion through conducting mobile mapping and collecting point cloud data during autumn and winter seasons and facilitate more accurate estimation of FFE.

4.2. Accuracy evaluation of FFE

As shows in Table 5, the ablation study conducted in this research examines the effectiveness of three different techniques: Slicing Aided Fine-Tuning (SF), Post-Processing (PP), and Slicing Aided Hyper Inference (SAHI). The study focuses on evaluating their accuracy results evaluated on the manually annotated ground references of Ventnor City. The base model used for the study is the yolov5 large model, which utilizes an input image size of 1280 and slicing aided fine-tuning. The results obtained from this base model show an RMSE of 2.21 ft and a MAE of 0.27 ft. Additionally, the recall of FFE, where true positive is defined as an absolute error lower than 1 ft, is measured at 0.63. To further refine the estimations, a rule-based post-processing technique is applied. This post-processing approach significantly reduces the RMSE to 1.81 ft and it does not decrease the rate of accurately estimated FFE values. While the slicing aided hyper inference technique does not demonstrate consistent improvement in the accuracy of estimated FFE. Considering these factors, the practical estimation of FFE can benefit from the combination of slicing aided fine-tuning and rule-based post-processing techniques.

The scatter plots depicting the estimated FFE and ground references for the three study areas are presented in Fig. 5. Specifically, when examining the residential buildings in the inland community, Manville, a wider range of FFE values is observed, primarily ranging from 40 to 80 ft. The estimated FFE values for 1819 residential buildings in Manville exhibit an absolute error lower than 1 ft, accounting for 61% of the total

Fig. 4. Visual inspection of the object detection results. The first and third rows display Google Street View images, while the second and fourth rows depict intensity-based front views with detected bounding boxes.

Table 5
Ablation study of yolov5 large model with an image size of 1280 evaluated on the test set of Ventnor city. SF, PP, SAHI and FI denote slicing aided fine-tuning, post-processing, slicing aided hyper inference, and full inference, respectively. The measurement unit of RMSE and MAE is feet.

Model	RMSE	MAE	Recall
SF	2.21	0.27	0.63
SF + PP	1.81	0.27	0.63
SF + PP + SAHI	2.14	0.35	0.63
SF + PP+(SAHI+FI)	2.18	0.36	0.63

2960 manually annotated buildings in this area. It shows a relatively higher level of accuracy compared to the two coastal cities. The corresponding RMSE and MAE metrics for Manville are measured at 0.75 ft and 0.2 ft, respectively. Ventnor and Longport exhibit similar patterns, with the elevation of residential buildings in these two cities being significantly lower than that of Manville. In Ventnor, the FFE of residential buildings is primarily below 20 ft, and in Longport, it is even lower, below 15 ft. Among the manually annotated FFE values in these areas, 63% of the 4124 in Ventnor and 70% of the 872 in Longport can be accurately estimated using the proposed automated method. As the largest study area among the three selected sites, Ventnor demonstrates the highest error in terms of both RMSE, which is 1.81 ft, and MAE, which is 0.27 ft. On the other hand, Longport exhibits a notably higher recall of highly accurately estimated FFE values, and its RMSE and MAE are measured at 1.55 ft and 0.24 ft, respectively. Taking the FFE from 145 validated elevation certificates record as reference, it demonstrates lower accuracy and the RMSE, MAE and recall metrics are measured at 1.93 ft, 0.44 ft and 0.56 in Longport.

The relationship between the estimated FFE and manually annotated for different object detection confidence intervals is illustrated in Fig. 6. The confidence score associated with a detected bounding box reflects the accuracy of the object's classification and localization. The majority of detected doors exhibit high confidence scores. For instance, the yolov5 model detects doors with a confidence score higher than 0.8 for 1417 residential buildings in Manville, which accounts for over 70% of the total properties. The figure also demonstrates a positive correlation between the confidence level and the accuracy of the estimated FFE. The group with a confidence interval of (0.9,1.0] exhibits the highest accuracy across all three study areas. In Manville, this group achieves an RMSE and MAE of 0.44 ft and 0.17 ft, respectively. In Ventnor, the corresponding metrics are 1.28 ft and 0.20 ft, while in Longport, they are 0.68 ft and 0.19 ft. However, it is important to note that there are instances where groups with lower confidence scores demonstrate higher accuracy compared to certain groups with higher confidence scores. This phenomenon is likely influenced by the sample size of the groups. For example, the group with a confidence interval of (0.2,0.3] in Longport exhibits lower error than the group with a confidence interval of (0.3,0.4]. Nevertheless, it is crucial to consider that the former group consists of only 18 residential building samples. Based on the findings, it is concluded that the accuracy of the estimated FFE can be improved by utilizing more powerful detectors, such as RT-DETR, which can enhance the object detection performance [57].

4.3. Spatial analysis of first floor elevation

Fig. 7 depicts the height of the first floor above bare ground along with its corresponding distribution. However, it is important to note that the proposed method used in this study was unable to estimate FFE for occluded buildings, and those properties within the communities were not displayed on the maps. The height of residential buildings above the

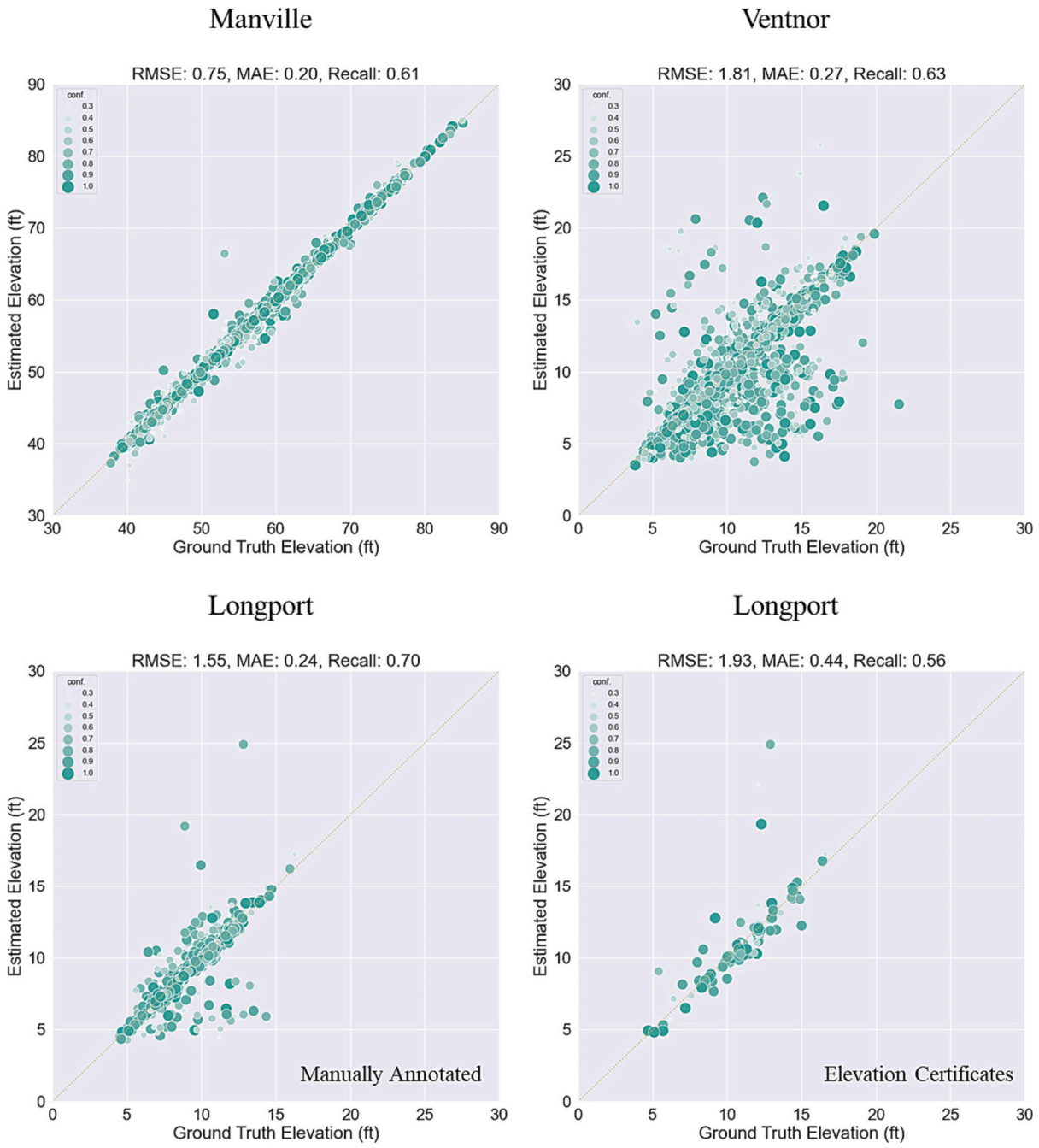


Fig. 5. Estimated FFE using proposed method and ground reference (i.e., manually annotated FFE and elevation certificates) in feet and the larger and darker points indicate detections with higher confidences. The FFE of the three study sites is estimated by the yolov5 large model with an image size of 1280 and slicing aided fine-tuning.

bare ground in the inland city is measured lower than the coastal cities. In Manville, the median difference between the first floor and ground is 2.62 ft, while in Ventnor and Longport, these values are slightly higher at 3.14 ft and 3.45 ft, respectively. It is worth mentioning that in Manville, the majority of residential buildings have heights ranging from 2 to 4 ft above bare ground. It is common to observe houses elevated more than 6 ft as a retrofitting measure to mitigate the risks associated with flooding and rising sea levels in coastal areas such as Ventnor and Longport. In the inland city of Manville, the primary type of flood hazard is pluvial flooding, and the flood zone AE is located near the river with low elevation (as depicted in Fig. 1). Houses exposed to higher flood risks should be constructed at higher elevations to comply with flood protection requirements. In coastal communities, particularly in Ventnor, houses located near the ocean exhibit higher first floor. The availability of the First Floor Elevation (FFE) map proves instrumental in conducting flood vulnerability analyses and facilitating informed decision-making processes aimed at constructing flood-resilient communities. It is important to emphasize that the accurate FFE information derived from this study holds significant potential in assisting FEMA in developing more precise risk-based flood insurance premiums. The map highlights the substantial variations in flood exposure risks among houses even within a relatively small area. By shifting from setting premiums based on national averages to a pricing methodology that better reflects the actual flood risk can yield significant societal benefits [9].

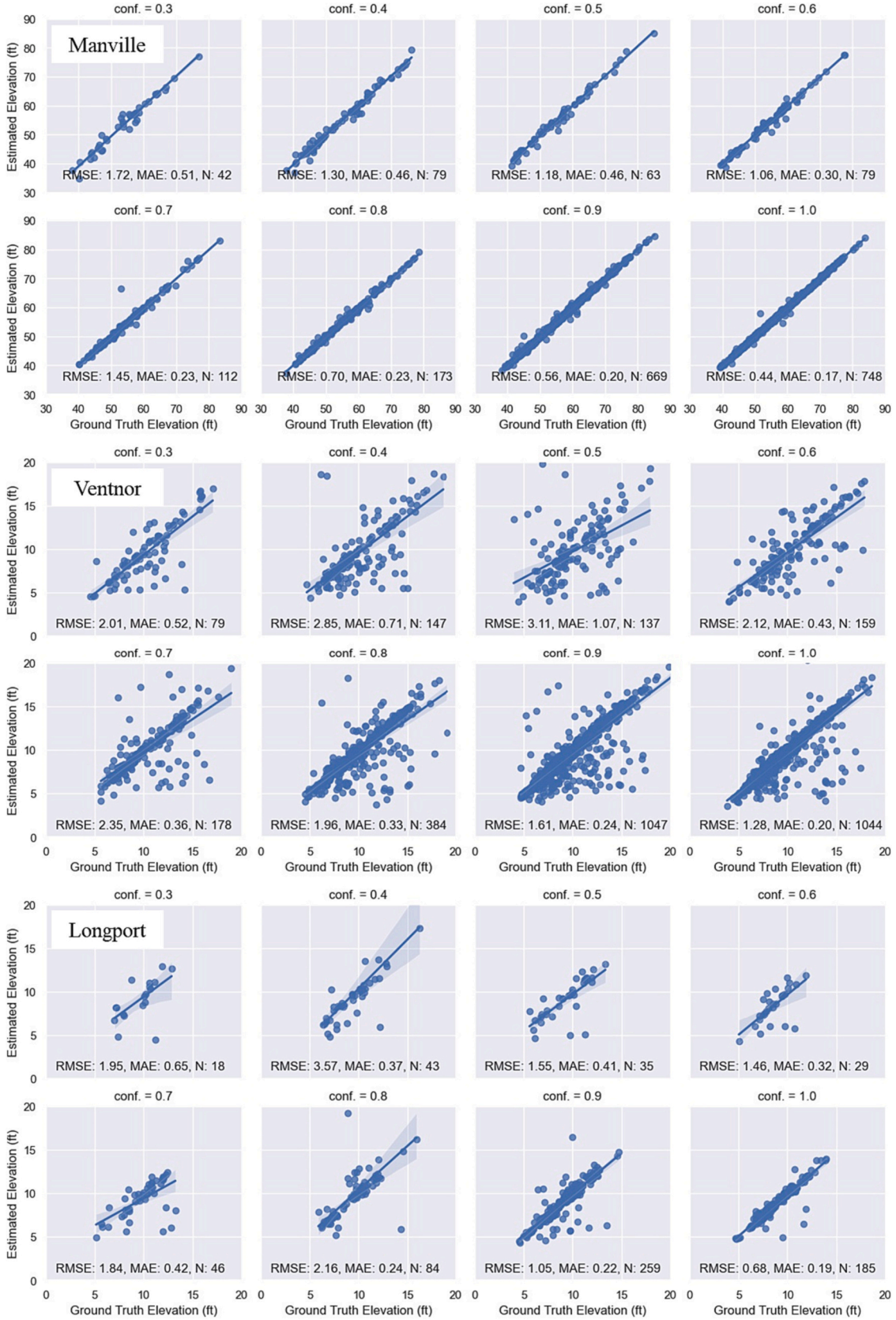


Fig. 6. Estimated FFE and manually annotated FFE under different object detection confidence. The confidence value for each subplot is the large value of the interval, for example, conf. = 1.0 represents confidence range (0.9,1.0].

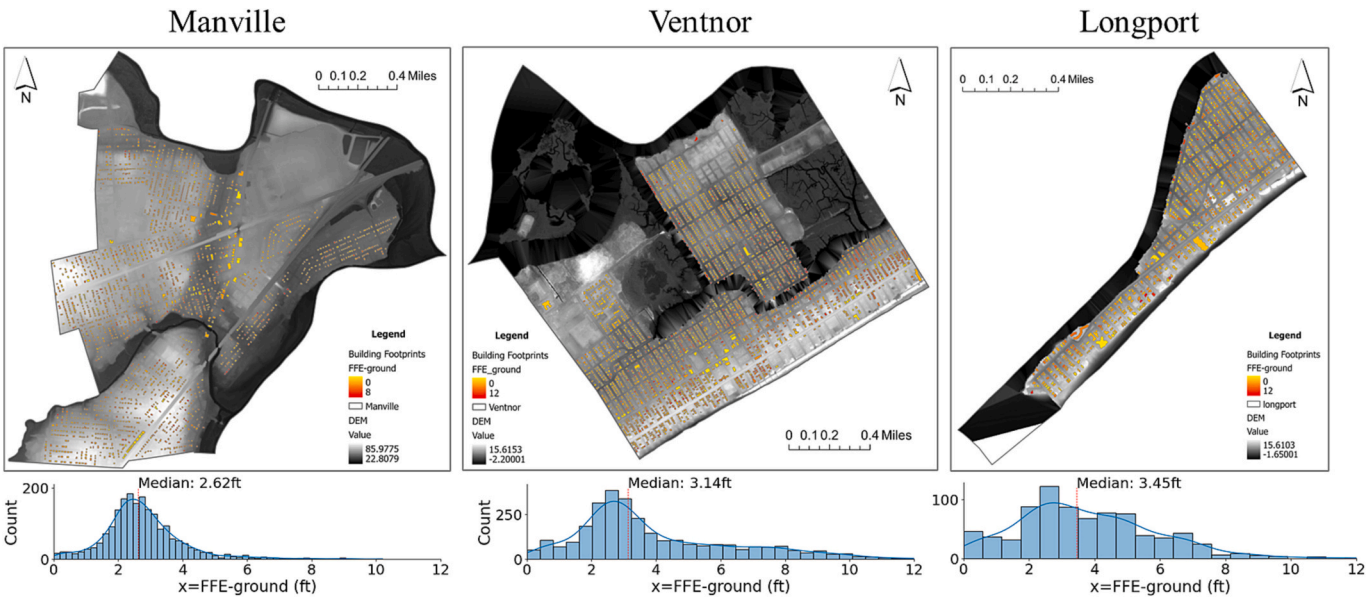


Fig. 7. Map of building footprints colorized with height of first floor above ground. The bare earth Digital Elevation Model (DEM) with a resolution of 10 ft is from NJOGIS and the building footprints were generated by Microsoft (https://github.com/microsoft/USBuildingFootprints).

5. Conclusions

Due to the escalating flood risk and the extensive damage caused by this hazard, researchers have become increasingly interested in implementing flood management measures and enhancing the resilience of flood-prone communities. Flood insurance is one of the most crucial preparedness actions of defense against flood damage and the newly implemented Risk Rating 2.0 of FEMA relies on FFE data to assess a property's flood risk more accurately in the United States. Our proposed methodology, which is automated and scalable, aims to address the data gap concerning missing FFE information for residential buildings across large geographic areas. Leveraging the projected intensity view of the point cloud, we employed a foundational computer vision model, the yolov5 large model, with an image size of 1280, incorporating slicing-aided fine-tuning. This significantly reduced the need of annotated mobile LiDAR point cloud data. This approach achieved an mAP@0.5:0.95 of 0.689 for all classes. Furthermore, by incorporating rule-based post-processing, we evaluated the MAE metric for the estimated FFE in Manville, Ventnor, and Longport at 0.2 ft, 0.27 ft, and 0.24 ft, respectively, based on manually annotated ground reference data. In Longport, validation against elevation certificates yielded an MAE of 0.44 ft for the estimated FFE. When considering the height of the first floor above bare ground, residential buildings in inland cities were found to have lower elevations compared to coastal cities. Specifically, the median difference between the first floor and ground measured 2.62 ft, 3.14 ft, and 3.45 ft for Manville, Ventnor, and Longport, respectively. The FFE data holds the potential to provide valuable guidance for setting flood insurance premiums and facilitating benefit-cost analyses of buyout programs targeting residential buildings with a high flood risk.

While our current work presents significant advancements, it is important to acknowledge the existing limitations. One limitation lies in the detection-based paradigm we employed, which does not provide accurate estimates of First Floor Elevation (FFE) for occluded buildings. To overcome this challenge, potential solutions could involve employing regression techniques utilizing alternative building attributes or exploring the fusion of visual data collected from both ground-based platforms and airborne platforms such as UAVs. By integrating multiple data sources, a more comprehensive estimation of FFE can be achieved. Another challenge we face is the issue of updating FFE information. In our future research endeavors, we aim to address this by developing efficient methods to locate the reconstruct properties affected by natural or human factors and minimize the required efforts while enabling the estimation of FFE for these properties. Last, but not the least, it is imperative to establish a direct linkage between the accuracy of FFE extraction with flood risk reduction in future studies. This linkage will make the method proposed here widely applicable to floodplain management practices.

CRediT authorship contribution statement

Jiahao Xia: Writing – review & editing, Writing – original draft, Validation, Formal analysis, Data curation, Conceptualization. **Jie Gong:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization.

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.

Data availability

Data will be made available on request.

Acknowledgments

This material is based upon work supported by the U.S. National Science Foundation under award 2103754, by FEMA under HMGP DR4488, and the U.S. Department of Homeland Security under Grant Award 22STESE00001-03-02. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. National Science Foundation and U.S. Department of Homeland Security.

References

[1] J. Rentschler, M. Salhab, B.A. Jafino, Flood exposure and poverty in 188 countries, Nat. Commun. 13 (1) (2022) 3527, https://doi.org/10.1038/s41467-022-30727-4.

[2] R.J. Nicholls, D. Lincke, J. Hinkel, S. Brown, A.T. Vafeidis, B. Meyssignac, S. E. Hanson, J.-L. Merkens, J. Fang, A global analysis of subsidence, relative sea-level change and coastal flood exposure, Nat. Clim. Chang. 11 (4) (2021) 338–342, https://doi.org/10.1038/s41558-021-00993-z.
[3] B. Jongman, P.J. Ward, J.C.J.H. Aerts, Global exposure to river and coastal flooding: long term trends and changes, Glob. Environ. Chang. 22 (4) (2012) 823–835, https://doi.org/10.1016/j.gloenvcha.2012.07.004.
[4] B. Tellman, J.A. Sullivan, C. Kuhn, A.J. Kettner, C.S. Doyle, G.R. Brakenridge, T. A. Erickson, D.A. Slayback, Satellite imaging reveals increased proportion of population exposed to floods, Nature 596 (7870) (2021) 80–86, https://doi.org/10.1038/s41586-021-03695-w.
[5] H.C. Winsemius, Jeroen C.J.H. Aerts, Ludovicus P.H. van Beek, Marc F.P. Bierkens, A. Bouwman, B. Jongman, Jaap C.J. Kwadijk, W. Ligtvoet, Paul L. Lucas, Detlef P. van Vuuren, Philip J. Ward, Global drivers of future river flood risk, Nat. Clim. Chang. 6 (4) (2016) 381–385, https://doi.org/10.1038/nclimate2893.
[6] T. Tingsanchali, Urban flood disaster management, Proc. Eng. 32 (2012) 25–37, https://doi.org/10.1016/j.proeng.2012.01.1233.
[7] E.J. Plate, Flood risk and flood management, J. Hydrol. 267 (1) (2002) 2–11, https://doi.org/10.1016/S0022-1694(02)00135-X.
[8] FEMA, Home Builder's Guide to Coastal Construction, Federal Emergency Management Agency, Washington, DC, 2010. URL: https://www.fema.gov/sites/default/files/2020-08/fema499_2010_edition.pdf (Accessed: 12 December 2023).
[9] L.T. de Ruig, T. Haer, H. de Moel, S.D. Brody, W.J.W. Botzen, J. Czajkowski, J.C.J. H. Aerts, How the USA can benefit from risk-based premiums combined with flood protection, Nat. Clim. Chang. 12 (11) (2022) 995–998, https://doi.org/10.1038/s41558-022-01501-7.
[10] D.P. Horn, National Flood Insurance Program: The Current Rating Structure and Risk Rating 2.0, Congressional Research Service, 2021. URL: https://sgp.fas.org/crs/homesec/R45999.pdf (Accessed: 12 December 2023).
[11] D.F. Maune, GPS elevation surveys-a key to proactive flood plain management, in: IAHS Publications-Series of Proceedings and Reports-Intern Assoc Hydrological Sciences 239, 1997, pp. 331–336. URL: https://digitalcommons.usf.edu/nhcc/84 (Accessed 12 December 2023).
[12] H. Ning, Z. Li, X. Ye, S. Wang, W. Wang, X. Huang, Exploring the vertical dimension of street view image based on deep learning: a case study on lowest floor elevation estimation, Int. J. Geogr. Inf. Sci. 36 (7) (2022) 1317–1342, https://doi.org/10.1080/13658816.2021.1981334.
[13] H. Needham, Nick McIntyre, Analyzing the Vulnerability of Buildings to Coastal Flooding in Galveston, Texas, URL: https://repository.library.noaa.gov/view/noaa/38177, 2018.
[14] N. Bruno, R. Roncella, Accuracy assessment of 3d models generated from google street view imagery, Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. XLII-2 (W9) (2019) 181–188, https://doi.org/10.5194/isprs-archives-XLII-2-W9-181-2019.
[15] Dewberry, National Enhanced Elevation Assessment Final Report, URL: https://www.dewberry.com/services/geospatial-mapping-and-survey/national-enhanced-elevation-assessment-final-report, 2012 (Accessed 12 December 2023).
[16] S.G. Kakade, Q.M. Nalawala, R.S. Srinivasan, Preliminary research in UAV-based estimation of lowest floor elevation for flood hazard pre-disaster management, in: 20th International Conference on Sustainable Environment & Architecture, 2020. URL: http://drf.shimberg.ufl.edu/Kakade-et-al-2020-SENVAR.pdf (Accessed 12 December 2023).
[17] N.D. Diaz, W.E. Highfield, S.D. Brody, B.R. Fortenberry, Deriving first floor elevations within residential communities located in Galveston using UAS based data, Drones 6 (4) (2022) 81, https://doi.org/10.3390/drones6040081.
[18] A. Gordon, Benjamin McFarlane, Developing First Floor Elevation Data for Coastal Resilience Planning in Hampton Roads, URL: https://www.hrpdcva.gov/uploads/docs/Developing_First_Floor_Elevation_Data_for_Coastal_Resilience_Planning_in_Hampton_Roads_HRPDC.pdf, 2019 (Accessed 12 December 2023).
[19] M. Guo, J. Gong, J.L. Whytlaw, Large-scale cloud-based building elevation data extraction and flood insurance estimation to support floodplain management, Int. J. Disast. Risk Reduct. 69 (2022) 102741, https://doi.org/10.1016/j.ijdrr.2021.102741.
[20] R. Klette, Concise Computer Vision, Springer, 2014, https://doi.org/10.1007/978-1-4471-6320-6.
[21] H.S. Munawar, A.W.A. Hammad, S.T. Waller, A review on flood management technologies related to image processing and machine learning, Autom. Constr. 132 (2021) 103916, https://doi.org/10.1016/j.autcon.2021.103916.
[22] C. Wang, Q. Yu, K.H. Law, F. McKenna, S.X. Yu, E. Taciroglu, A. Zsarnóczay, W. Elhaddad, B. Cetiner, Machine learning-based regional scale intelligent modeling of building information for natural hazard risk management, Autom. Constr. 122 (2021) 103474, https://doi.org/10.1016/j.autcon.2020.103474.
[23] Z. Zhou, J. Gong, X. Hu, Community-scale multi-level post-hurricane damage assessment of residential buildings using multi-temporal airborne LiDAR data, Autom. Constr. 98 (2019) 30–45, https://doi.org/10.1016/j.autcon.2018.10.018.
[24] S. Shirowzhan, S.M.E. Sepasgozar, H. Li, J. Trinder, P. Tang, Comparative analysis of machine learning and point-based algorithms for detecting 3D changes in buildings over time using bi-temporal lidar data, Autom. Constr. 105 (2019) 102841, https://doi.org/10.1016/j.autcon.2019.102841.
[25] M. Kamari, Y. Ham, AI-based risk assessment for construction site disaster preparedness through deep learning-based digital twinning, Autom. Constr. 134 (2022) 104091, https://doi.org/10.1016/j.autcon.2021.104091.
[26] N. Haghighatgou, S. Daniel, T. Badard, A method for automatic identification of openings in buildings facades based on mobile LiDAR point clouds for assessing impacts of floodings, Int. J. Appl. Earth Obs. Geoinf. 108 (2022) 102757, https://doi.org/10.1016/j.jag.2022.102757.
[27] Z. Zhengxia, C. Keyan, S. Zhenwei, G. Yuhong, Y. Jieping, Object detection in 20 years: a survey, Proc. IEEE 111 (3) (2023) 257–276, https://doi.org/10.1109/JPROC.2023.3238524.
[28] J. Ding, N. Xue, G.S. Xia, X. Bai, W. Yang, M.Y. Yang, S. Belongie, J. Luo, M. Datcu, M. Pelillo, L. Zhang, Object detection in aerial images: a large-scale benchmark and challenges, IEEE Trans. Pattern Anal. Mach. Intell. 44 (11) (2022) 7778–7796, https://doi.org/10.1109/TPAMI.2021.3117983.
[29] Z.Q. Zhao, P. Zheng, S.T. Xu, X. Wu, Object detection with deep learning: a review, IEEE Trans. Neural Netw. Learn. Syst. 30 (11) (2019) 3212–3232, https://doi.org/10.1109/TNNLS.2018.2876865.
[30] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110, https://doi.org/10.1023/B:VISI.0000029664.99615.94.
[31] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) Vol. 1, 2005, pp. 886–893, vol. 881, https://doi.org/10.1109/CVPR.2005.177.
[32] M.A. Hearst, S.T. Dumais, E. Osuna, J. Platt, B. Scholkopf, Support vector machines, IEEE Intell. Syst. Appl. 13 (4) (1998) 18–28, https://doi.org/10.1109/5254.708428.
[33] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, Microsoft Coco: Common Objects in Context, Springer International Publishing, Cham, 2014, pp. 740–755, https://doi.org/10.1007/978-3-319-10602-1_48.
[34] R. Girshick, Fast r-cnn, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1440–1448, https://doi.org/10.1109/ICCV.2015.169.
[35] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 580–587, https://doi.org/10.1109/CVPR.2014.81.
[36] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2961–2969, https://doi.org/10.1109/ICCV.2017.322.
[37] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: towards real-time object detection with region proposal networks, Adv. Neural Inf. Proces. Syst. 28 (2015). https://doi.ieeecomputersociety.org/10.1109/TPAMI.2016.2577031.
[38] J. Redmon, S. Divvala, R. Girshick, 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, https://doi.org/10.1109/CVPR.2016.91.
[39] A. Bochkovskiy, C.-Y. Wang, H.-Y.M. Liao, Yolov4: Optimal speed and accuracy of object detection, arXiv Prepr. (2020), https://doi.org/10.48550/arXiv.2004.10934 arXiv:2004.10934.
[40] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, An image is worth 16x16 words: transformers for image recognition at scale, arXiv Prepr. (2020), https://doi.org/10.48550/arXiv.2010.11929 arXiv:2010.11929.
[41] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, S. Zagoruyko, End-to-end object detection with transformers, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16, Springer, 2020, pp. 213–229, https://doi.org/10.1007/978-3-030-58452-8_13.
[42] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, M. Bennamoun, Deep learning for 3D point clouds: a survey, IEEE Trans. Pattern Anal. Mach. Intell. 43 (12) (2021) 4338–4364, https://doi.org/10.1109/TPAMI.2020.3005434.
[43] Y. Wu, Y. Wang, S. Zhang, H. Ogai, Deep 3D object detection networks using LiDAR data: a review, IEEE Sensors J. 21 (2) (2021) 1152–1171, https://doi.org/10.1109/JSEN.2020.3020626.
[44] B. Li, T. Zhang, T. Xia, Vehicle detection from 3d lidar using fully convolutional network, arXiv Prepr. (2016), https://doi.org/10.48550/arXiv.1608.07916 arXiv:1608.07916.
[45] K. Saleh, A. Abobakr, M. Attia, J. Iskander, D. Nahavandi, M. Hossny, S. Nahvandi, Domain adaptation for vehicle detection from Bird's eye view LiDAR point cloud data, in: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019, pp. 3235–3242, https://doi.org/10.1109/ICCVW.2019.00404.
[46] X. Chen, H. Ma, J. Wan, B. Li, T. Xia, Multi-view 3d object detection network for autonomous driving, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1907–1915, https://doi.org/10.1109/CVPR.2017.691.
[47] Y. Zhou, O. Tuzel, Voxelnet: end-to-end learning for point cloud based 3d object detection, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (2018) 4490–4499, https://doi.org/10.1109/CVPR.2018.00472.
[48] H. Zhao, L. Jiang, J. Jia, P.H. Torr, V. Koltun, Point transformer, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 16259–16268, https://doi.org/10.1109/ICCV48922.2021.01595.
[49] G. Li, M. Muller, A. Thabet, B. Ghanem, Deepgcns: can gcns go as deep as cnns?, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 9267–9276, https://doi.org/10.1109/TPAMI.2021.3074057.
[50] Y. Wang, Y. Sun, Z. Liu, S.E. Sarma, M.M. Bronstein, J.M. Solomon, Dynamic graph cnn for learning on point clouds, ACM Trans. Graph. (tog) 38 (5) (2019), https://doi.org/10.1145/3326362. Article 146.
[51] NOAA National Centers for Environmental Information (NCEI), US Billion-Dollar Weather and Climate Disasters, 2023, https://doi.org/10.25921/stkw-7w73.
[52] Ultralytics, ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation, 2022, https://doi.org/10.5281/zenodo.7347926.

[53] F.C. Akyon, S.O. Altinuc, A. Temizel, Slicing aided hyper inference and fine-tuning for small object detection, in: 2022 IEEE International Conference on Image Processing (ICIP), IEEE, 2022, pp. 966–970, https://doi.org/10.1109/ICIP46576.2022.9897990.
[54] T. Lin, LabelImg, URL: https://github.com/tzutalin/labelImg, 2015 (Accessed 12 December 2023).
[55] M. Schütz, Potree: Rendering Large Point Clouds in Web Browsers, Technische Universität Wien, Wiedeń, 2016. URL: https://publik.tuwien.ac.at/files/publik_252607.pdf (Accessed 12 December 2023).
[56] Flood Information, Longport, NJ, URL: https://longportnj.withforerunner.com/properties, 2023 (Accessed 12 December 2023).
[57] W. Lv, S. Xu, Y. Zhao, G. Wang, J. Wei, C. Cui, Y. Du, Q. Dang, Y. Liu, Detrs beat yolos on real-time object detection, arXiv Prepr. (2023), https://doi.org/10.48550/arXiv.2304.08069 arXiv:2304.08069.