Automation in Construction 132 (2021) 103968

Contents lists available at ScienceDirect

Automation in Construction

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

Automatic detection of unreinforced masonry buildings from street view images using deep learning-based image segmentation

Chaofeng Wang [b,*,1], Sarah Elizabeth Antos [a,1], Luis Miguel Triveno [a,1]

[a] World Bank Group, United States
[b] M.E. Rinker, Sr. School of Construction Management, College of Design, Construction and Planning, University of Florida, FL, United States

ARTICLE INFO

Keywords:
Unreinforced masonry building
Instance segmentation
Deep learning
URM
Masonry building
Street view image

ABSTRACT

Mitigation of seismic risk is a challenge for 70+ countries in the world. Screening the building stock for potential structural defects is one way to locate structures that are vulnerable to strong ground motion. Often in developing countries, masonry buildings are not reinforced or confined to withstand earthquake loads. It has been observed that such buildings cannot withstand the lateral loads imposed by an earthquake. An estimated 77 percent of the fatalities in the earthquakes during the last 100 years were caused mainly by the collapse of masonry buildings. Given the probability of severe damage or collapse in the event of an earthquake, identification and retrofit of such masonry buildings are critical. Screening of masonry buildings by conventional methods is usually a time-consuming and labor-intensive process. This research presents an automated workflow for segmenting the presence of such buildings through the use of street-view images. The method uses deep learning techniques. An inventory composed of a set of street view images was collected from the streets of Oaxaca State, Mexico using a 360° camera mounted on a car. These images were then annotated by trained engineers. Using the annotated dataset, an instance segmentation model was trained to detect and classify masonry buildings from the street view images. Using the model, we performed city scale detections of masonry buildings. We found the spatial distribution pattern of masonry buildings is correlated with the urbanization paths of the cities. The model could be used to produce large-scale automated detection of buildings at a fraction of the cost and in the fraction of the time of the alternatives. With affordable, accurate and massive screenings, governments can target these buildings for retrofit efforts and companies can assess the business opportunity of prevention (or "Building Better Before").

1. Introduction

The consequences of earthquake can be devastating to the building system, causing immediate loss of lives and long-term damages to the economy. While precise prediction of earthquake occurrences is impossible [27], the seismic performance of certain construction types is well understood. Masonry, especially unreinforced masonry (URM), is a construction type known to be highly vulnerable to seismic activity. This is highlighted in the sequence of past major earthquakes [18,21,22,2,8,5]. Fig. 1 shows the greatest number of earthquake fatalities in the last 100 years, 77% owing to the collapse of masonry buildings. Masonry buildings are common all through Latin America, Eastern Europe, and many parts of Asia [4]. Active seismicity coalescing with rapid urbanization is increasingly threatening the safety of urban dwellers making pre-earthquake retrofitting of vulnerable masonry buildings an urgent need.

While retrofitting or upgrading of existing masonry buildings to improve their seismic performance is technically and economically feasible, the implementation at scale remains a challenge. One important obstacle is that the decision of whether to retrofit an existing building for better seismic performance is usually based on a professional risk screening, which is costly and time-consuming.

The availability of street view images and recent advances in computer vision techniques makes automated rapid screening of buildings viable. In recent years, street view images have attracted significant attention from researchers and have emerged as a much sought after resource because of their wide availability and rich visual information. Studies have shown the potential of using the images in a variety of

* Corresponding author.
E-mail addresses: chaofeng.wang@ufl.edu (C. Wang), santos1@worldbank.org (S.E. Antos), ltriveno@worldbank.org (L.M. Triveno).

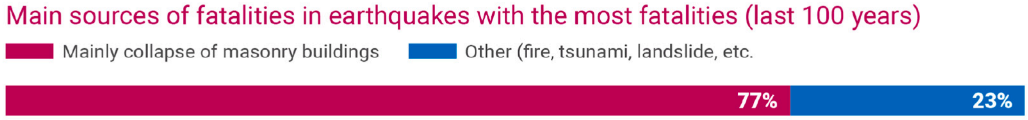
[1] These authors contributed equally

https://doi.org/10.1016/j.autcon.2021.103968
Received 12 February 2021; Received in revised form 12 September 2021; Accepted 16 September 2021
Available online 28 September 2021

Earthquakes causing the greatest number of fatalities in the last 100 years[1]

YEAR	MAIN CAUSE OF FATALITIES	LOCATION	FATALITIES	MAGNITUDE
1920	Masonry	Haiyuan, Ningxia (Ning-hsia), China	200,000	7.8
1923	Fire	Kanto (Kwanto), Japan	142,800	7.9
1927	Masonry	Gulang, Gansu (Kansu), China	40,900	7.6
1931	Mix	Near Fuyun (Koktokay), Xinjiang (Sinkiang), China	10,000	8
1933	Landslide/Mix	North of Maowen, Sichuan (Szechwan), China	9,300	7.5
1934	Masonry	Bihar, India-Nepal	10,700	8.1
1935	Masonry	Quetta, Pakistan (Baluchistan, India)	30,000	7.6
1939	Masonry	Erzincan, Turkey	32,700	7.8
1939	Masonry	Chillan, Chile	28,000	7.8
1944	Masonry	San Juan, Argentina	8,000	7.4
1948	Masonry	Ashgabat (Ashkhabad), Turkmenistan (Turkmeniya, USSR)	110,000	7.3
1949	Landslide	Khait, Tajikistan (Tadzhikistan, USSR)	12,000	7.5
1960	Masonry	Agadir, Morocco	15,000	5.7
1962	Masonry	Bu'in Zahra, Qazvin, Iran	12,225	7.1
1968	Earth	Dasht-e Bayaz, Iran	12,000	7.3
1970	Masonry	Chimbote, Peru	70,000	7.9
1970	Mixed	Tonghai, Yunnan Province, China	10,000	7.5
1972	Masonry	Iran: Qir, Karzin, Jahrom, Firuzabad	30,000	6.9
1972	Wood+Earth	Nicaragua: Managua	10,000	6.2
1974	Masonry	China	20,000	6.8
1976	Masonry	Tangshan, China	242,769	7.5
1976	Masonry	Guatemala	23,000	7.5
1976	Tsunami	Mindanao, Philippines	8,000	7.9
1978	Masonry	Iran: Tabas	15,000	7.8
1985	Mixed	Mexico: Michoacan: Mexico City	9,500	8
1988	Masonry	Spitak, Armenia	25,000	6.8
1990	Masonry	Western Iran	50,000	7.4
1993	Masonry	Latur-Killari, India	9,748	6.2
1999	Masonry	Turkey: Istanbul, Kocaeli, Sakarya	17,118	7.6
2001	Masonry	India: Gujarat: Bhuj, Ahmadabad, Rajokot; Pakistan	20,085	7.6
2003	Masonry	Iran: Southeastern: Bam, Baravat	31,000	6.6
2004	Tsunami	Indonesia: Sumatra: Aceh: Off West Coast	227,898	9.1
2005	Masonry	Pakistan: Muzaffarabad, Uri, Anantnag, Baramula	86,000	7.6
2008	Masonry	Eastern Sichuan, China	87,587	7.9
2010	Masonry	Haiti: Port-Au-Prince	316,000	7
2011	Tsunami	Japan: Honshu	20,896	9
2015	Masonry	Nepal: Kathmandu; India; China; Bangladesh	8,000	7.8

Main sources of fatalities in earthquakes with the most fatalities (last 100 years)



[1] Data from USGS (https://web.archive.org/web/20130411110008/http://earthquake.usgs.gov/earthquakes/world/world_deaths.php) and NOAA (National Geophysical Data Center / World Data Service (NGDC/WDS): Significant Earthquake Database. National Geophysical Data Center, NOAA. doi:10.7289/V5TD9V7K)

Fig. 1. Earthquake fatalities.

applications, ranging from predicting housing prices [3,23] to evaluating the safety of neighborhoods [28,24]. Given the distinct visual appearance of buildings, automatically recognizing such buildings from street view images is possible.

In conjunction with using street view images for a variety of applications, the past several years have witnessed significant progress in deep learning (DL) technologies, which have been embraced by the computer vision community. Deep learning techniques (e.g., deep convolutional neural networks (CNNs)) have been applied and achieved impressive performance in various applications in civil and natural hazard engineering [13,10,16,11,37,33,15,35].

In this work, we propose an integrated workflow for automatic screening of masonry buildings at a city scale. The workflow consists of several components. First, street view images are collected using

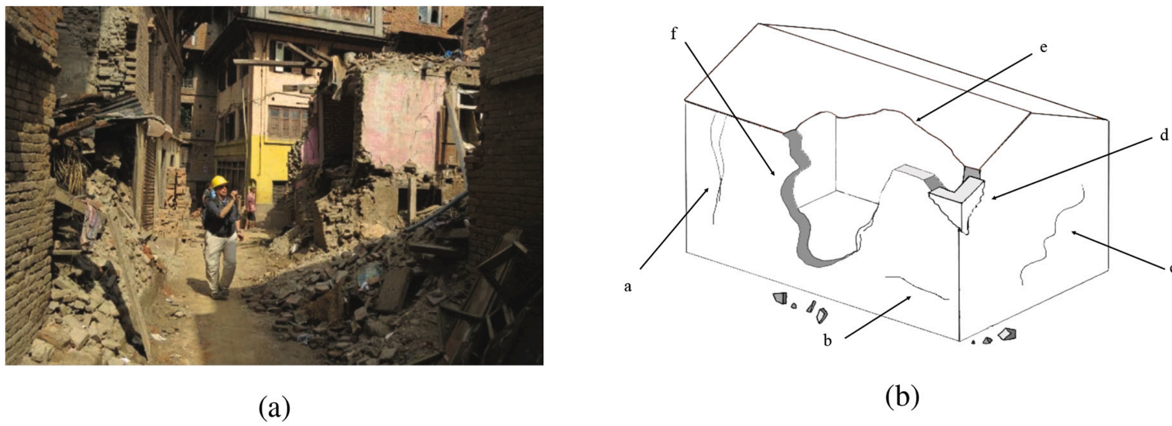


Fig. 2. (a) Observation of unreinforced masonry building collapse [12] (b) Common failure mechanisms for masonry structures: a. vertical cracking; b. horizontal shear cracking; c. diagonal shear cracking; d. corner damages and wall separation; e. roof collapse; f. wall out-of-plane collapse.

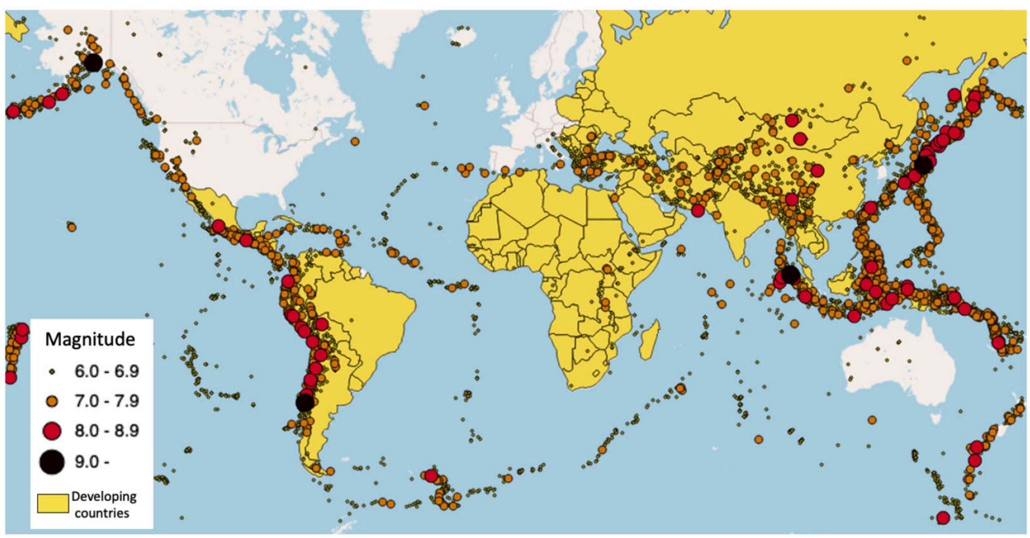


Fig. 3. Overlap of seismicity and developing countries: urgent needs for screening.

vehicles mounted with cameras. Second, a subset of images are selected and the masonry buildings are manually annotated. The annotated images are then used to train a instance segmentation model that can automatically detect masonry buildings in the rest of the images. The detected buildings are then registered in a database, which contains the footprints of all buildings, in a geocoding process. The key component of the workflow is the automatic detection of masonry buildings from street view images using the instance segmentation method. The automatic detection is challenging, because the street view images usually contain noises and complex scenes, posing difficulties (e.g., heavy occlusions caused by vegetation or vehicles) for the neural network to understand. In the following sections, we present the development of the workflow and two examples of its application.

2. Urgent need for screening of masonry buildings

Masonry buildings, especially unconfined or unreinforced masonry buildings (URMs), are well known for their seismic vulnerability. One reason is masonry materials, such as adobe, stone, brick, and concrete block, are typically strong in compression and able to bear gravity loads but are relatively weaker to carry lateral loads. Masonry buildings intrinsically exhibit a quasi-brittle nature and limited ductility, making them extremely susceptible to failure during earthquakes. Another reason is the walls of URM buildings are not secured to the roofs and floors. The break-away of walls can lead to partial or full collapse, hence leading to causalities and economic losses. Moreover, a considerable number of existing masonry buildings were built before the introduction of the first seismic design codes in their countries. Thus, their seismic performance has not been designed or detailed to withstand the demands of potential earthquakes. Common failure modes of URM buildings have been identified in historical earthquake events worldwide [6,7,20,26], including corner damage, wall separation, wall cracking and sliding, wall out-of-plane collapse, roof collapse, etc., as shown in Fig. 2.

Construction methods and technologies are dependent on local conditions and level of engineering expertise as well as demographic factors. For example, in North America, URM structures are common but building codes effectively regulate construction and determine how well protected structures are from seismic loading. In contrast, the rapid urbanization of developing countries has resulted in a huge rise in demand for quick and cheap housing without regulation or controls in place to ensure they are built to survive earthquakes. The technical studies behind the 2014 Disaster Risk Management Strategy from the Government of Oaxaca, Mexico, found two main sources of vulnerability in the housing stock: (i) poor quality of construction materials; and (ii) poor structural systems (e.g., URM). As a result, the 2017 Chiapas earthquake affected 70 thousand units in Istmo de Tehuantepec. SARC, a disaster risk management consultancy firm, estimated that retrofitting URM units could cut losses by half in future earthquakes Fig. 4.

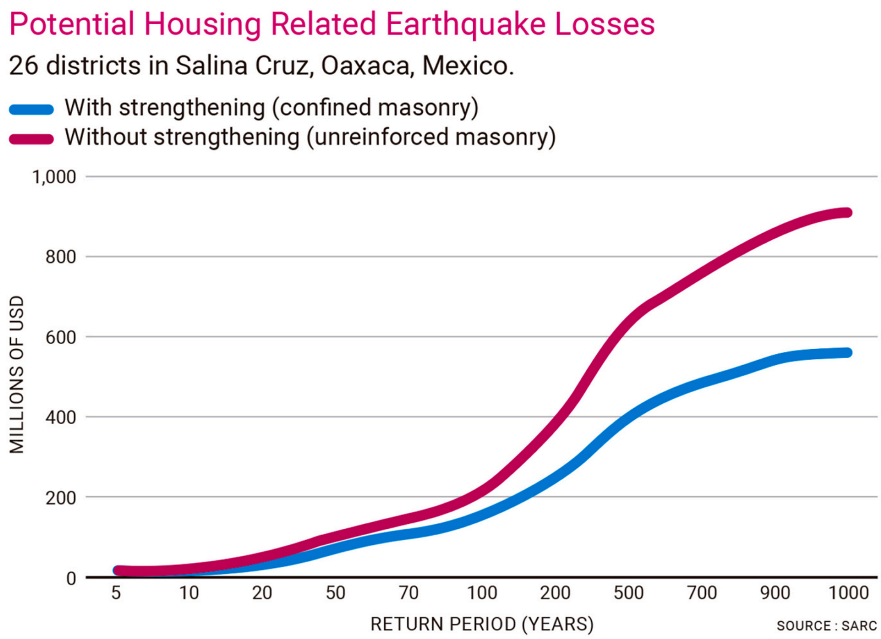


Fig. 4. Potential housing related earthquake losses in Salina Cruz, Oaxaca, Mexico.

(a) A Confined URM building (b) Confinement highlighted

Fig. 5. Confined URM.

Unfortunately, as shown in Fig. 3, the global seismicity substantially overlaps with developing countries, which makes screening and retrofitting these buildings an urgent need, particularly given that building codes are rarely enforced.

3. Vision based screening of masonry buildings

Traditionally, seismic screening of buildings requires extensive examination of information from various fields of expertise, including the recognition of seismic risks, the performance of structures during significant earthquakes, seismic assessment and performance assessment tools, etc. For example, in the guidebook FEMA 154 [1], a system to analyze the seismic resistance of buildings, with minimal effort, without access to the buildings (e.g., based on the visual cues manifested at building exterior) was developed. The approach has been used broadly for evaluating structures in many countries and regions prone to damaging earthquakes [34,30,31,29]. The intuition behind vision-based screening methods is that, to a considerable degree, the seismic performance of a structure depends on its construction material and type, geometric irregularities, etc., which can be determined based on visual cues.

Therefore, it is possible to identify masonry buildings based on its visual cues captured in street view images. A building is considered to be a masonry building if its wall is constructed of masonry materials, such as adobe, brick, concrete block, etc. It is also important for the screening process to discriminate reinforced or confined masonry buildings from

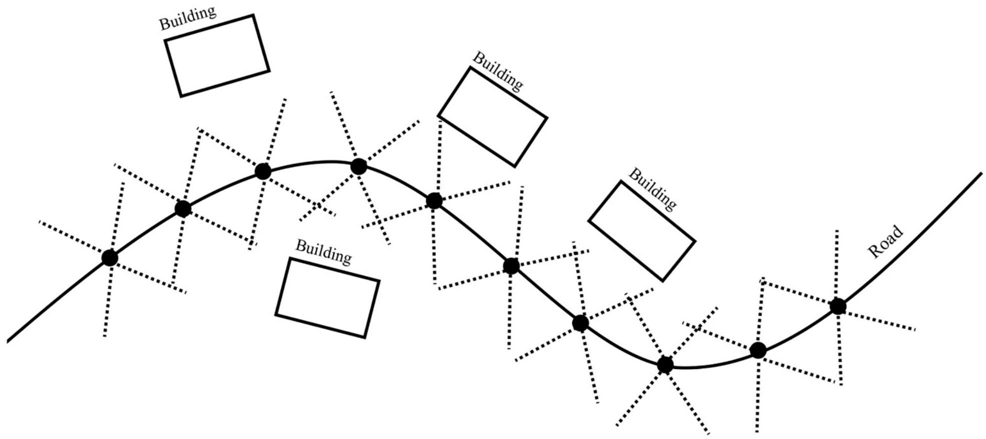


Fig. 6. Street view image harvesting.

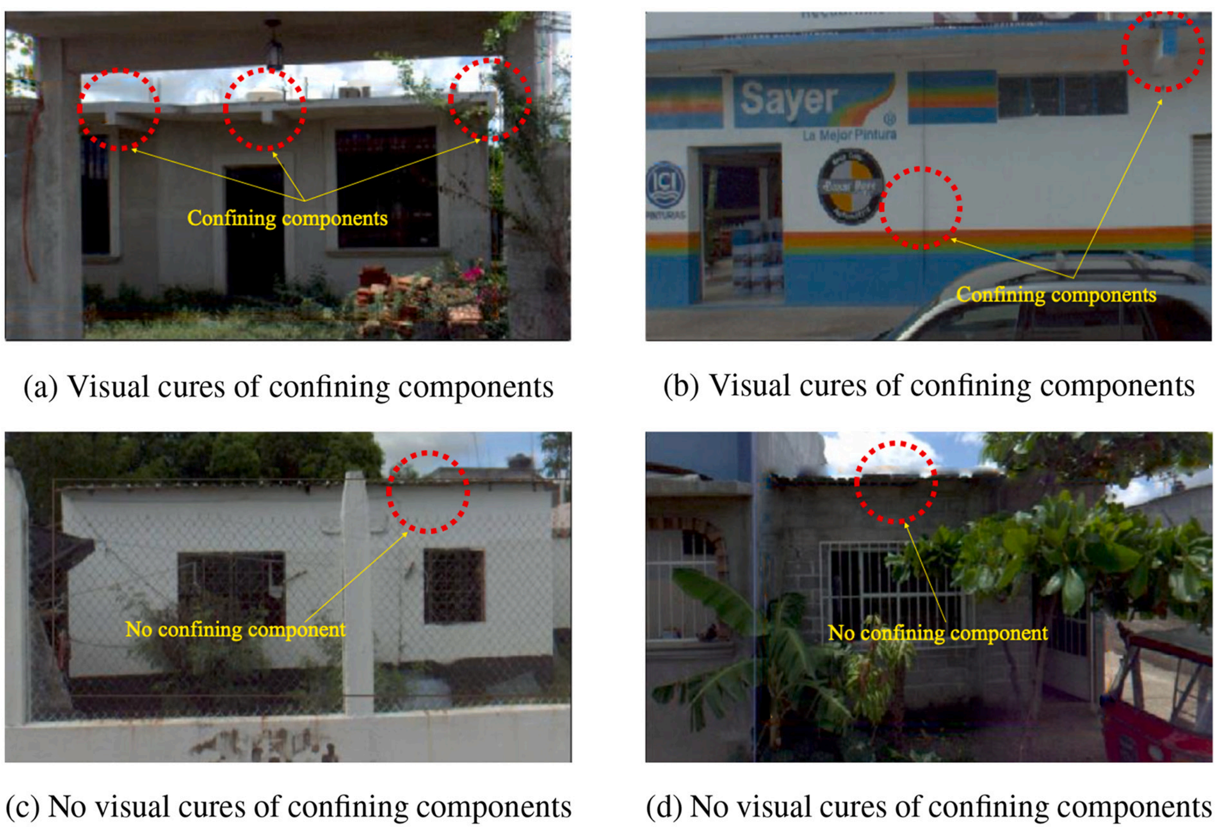


Fig. 7. Examples of images to be annotated.

URM buildings. Confined masonry can be defined as unreinforced masonry confined with reinforced concrete tie beams and columns around openings and masonry panels. The construction principle of confined masonry structures is to confine the unreinforced masonry components to strengthen their lateral resistance to seismic activity. The beams and columns tie the masonry walls together, preventing their separation at vertical connecting under the lateral inertia forces. It has proved that both the strength and ductility of masonry are improved by the confinement. Since its first use in the 1900s [9], confined masonry structures quickly became popular in many regions across the world [25] thanks to its satisfactory seismic performance, especially when the masonry densities in both horizontal and vertical directions are adequate. Confined masonry is now known as an economical and seismically superior construction type and has been adopted widely for low- and medium-rise buildings.

The reinforced concrete tie columns and tie beams are visually detectable from the exterior of a building. As the example shown in Fig. 5, the visual cues of the existence of tie columns and beams indicate the building is not a URM, but a confined masonry building or a concrete frame filled with masonry wall – both are more seismic resilient than a URM.

Although visual screening has been broadly adopted, such screening can be expensive, error-prone and labor-intensive. Decisions can be subjective, potentially leading to diverging interpretations and erroneous results, even if the screening process is implemented by professionals following guidelines, especially for large inventories.

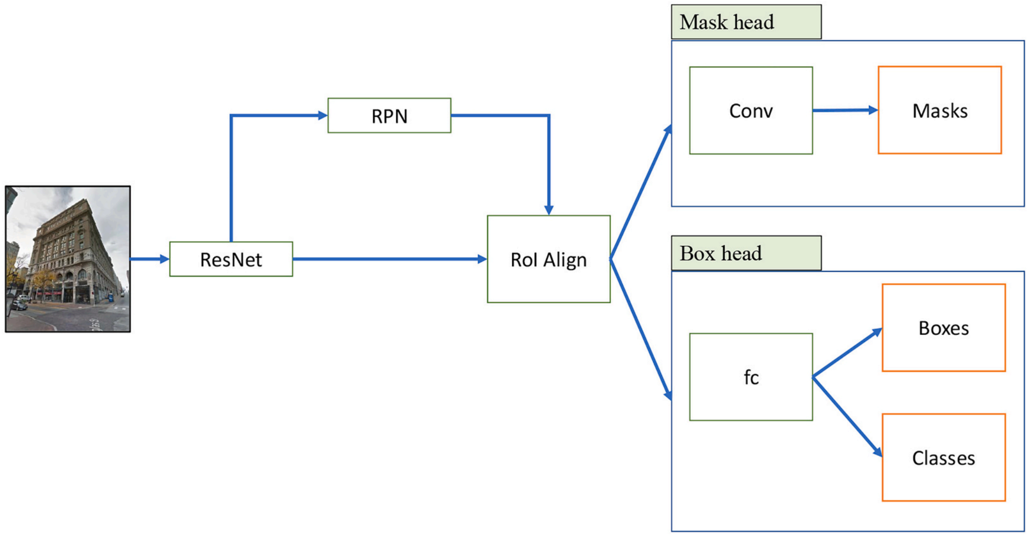


Fig. 8. Mask R-CNN [17].

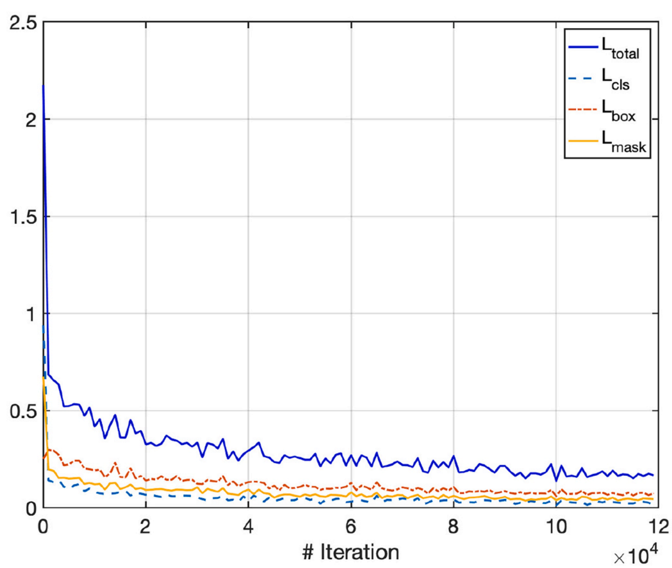


Fig. 9. Training loss.

4. Automated screening by instance segmentation of street view images

Street view images are photos taken from roadways that capture objects such as buildings, vegetation, vehicles. Recently, street view images have drawn considerable attention because of the rich visual information and the relatively low cost for collecting them. Furthermore, the computational hardware and software used for processing these images have improved substantially. All these advantages lead to a wide spectrum of applications of street view images in a variety of disciplines. For example, [28] calculated the perceived degree of public security by analyzing numerous street view images. It showed that the visual presence of urban surroundings can indicate the lives of inhabitants in a neighborhood. In another study, [14] utilized the information of cars extracted from street view images to gauge the demographic makeup of cities. [23] used street view images to measure housing prices in London and found a strong correlation between street presence and housing prices. [19] combined street view images together with satellite images to infer the function of a building. In a recent study [36,37], street view image classification was successfully applied in building seismic risk analyses.

Further, recent advancements in the field of computer vision technology, such as deep learning and convolutional neural networks (CNN) have boosted street view analysis. A class of image analysis approaches, instance segmentation, that can categorize and group pixels in an image into predefined classes, has been developed and applied in various disciplines. Building upon an image segmentation approach, this study introduces an alternative screening procedure that collects images systematically from the street and employs a trained instance segmentation model for detecting masonry buildings from these images. The proposed approach possesses many advantages when compared with traditional screening approaches, in terms of the consistency in the evaluation, cost efficiency, and scalability.

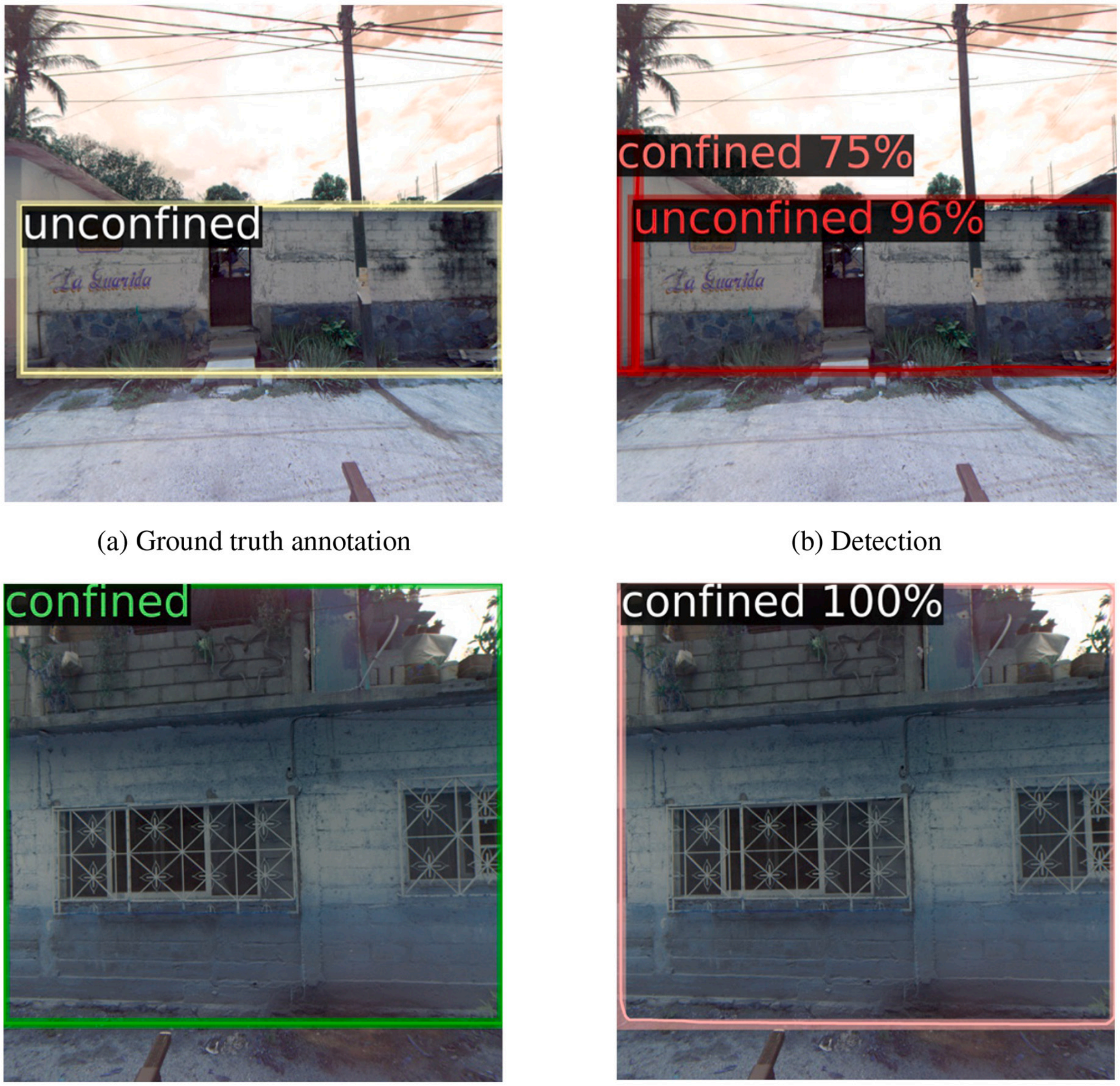
4.1. Objectives of the screening

The primary objective of the screening process is to detect masonry buildings from street view images. Moreover, detected masonry buildings should be classified based on their construction type which determines the seismic vulnerability level. To this end, the screening process will further classify the detected masonry buildings into two class depending on whether the masonry walls are confined with ties. Furthermore, the developed workflow is capable of performing citywide screening, which produces evaluations of most buildings in a city or a region. Such screening results are extremely important for guiding the governmental and emergency agencies in disaster preparedness and retrofitting plans.

4.2. Overall workflow

To achieve the aforementioned goals, we developed and validated an integrated workflow for automatic large-scale screening of masonry buildings. Components of the workflow include:

- Data Collection: The first step is to collect street level images, together with the GPS coordinates and parameters of the camera.
- Data Annotation: The objective of this component is to obtain the training data, i.e., the annotations of and building objects in images. The annotation should mark the locations of buildings in a given image and also the attributes of each building.
- Model Training: The annotated dataset is used for training the instance segmentation models.



(a) Ground truth annotation (b) Detection

(c) Ground truth annotation (d) Detection

Fig. 10. Examples of ground truth annotations compared with inferences.

- Inference: The trained models are used for detecting buildings and their attributes from all collected images.
- Geocoding: The detected building information is integrated into building footprints.

The core objective of the workflow is to successfully train an instance segmentation model that is capable of detecting and classifying masonry building instances from street view images. The complex contexts in street view images present a unique challenge. To succeed, the model must be trained on a large dataset to facilitate the identification of building objects and the further classification of them into different categories.

4.3. Data collection

The street view images are collected from cameras mounted on a ground vehicle. Another option is to use unmanned aerial vehicle (UAV), which has the advantages such as potentials for reducing heavy occlusions in the collected images. We chose to use a ground vehicle instead of UAV based on the following considerations: 1. Obtaining a permit for a UAV to be hover in the middle of the street, just above the traffic, would have not been possible or the administration process could be prolonged. 2. Using a UAV would be prohibitively time consuming because the battery would die frequently (every 30–60 min). As shown in Fig. 6, while the vehicle drives through the roads, the cameras take images at a constant frequency from angles that are perpendicular to the travel direction. The cameras also record the GPS coordinate of each captured image and the heading direction of the travel at the moment the image is taken.

4.4. Data annotation

The task requires that we not only annotate the shape of the buildings in the training dataset, but also annotate the class of the buildings, which demands professional knowledge of the annotators to discriminate between different construction types. To classify the building as 'unconfined', the annotators must observe visual cues of confining components on the facade, which is challenging. Examples of the images

(a) (b)

Fig. 11. Multiple adjacent buildings.

(a) Closeup shot (b) Long shot

Fig. 12. Different distances from camera.

are shown in Fig. 7. As can be seen, the identification of the visual cues depends on the annotators' judgement of the existing of structural elements that might hold concrete and rebars.

A subset of street view images are randomly selected from the dataset collected in the city of Salina Cruz, Oaxaca, Mexico. Bounding boxes are drawn on images to annotate masonry buildings. Each bounding box is associated with a class name either 'confined' or 'unconfined'. The 'confined' class indicates a masonry building with confinement or reinforcement components, such as tie columns and beams that confine the masonry walls. The 'unconfined' class represents masonry buildings without reinforcement or confinement components. For each masonry building in an image, only its front facade is annotated with a rectangular bounding box. A total of 10,541 buildings are annotated, among which there are 8839 confined masonry buildings and 1702 unconfined masonry buildings. These annotations together with the annotated images are used as the input for training a instance segmentation model.

4.5. *Segmentation model*

We employ the Mask R-CNN approach, which is originally developed by [17], as our segmentation model. This model is an extension of the Faster R-CNN algorithm [32] by adding a branch for predicting segmentation masks on each Region of Interest (RoI) in parallel with the existing branch for classification and bounding box regression. Fig. 8 shows the architecture of the Mask R-CNN.

The street view image is first input into a pre-trained backbone CNN (ResNet), which is connected with a Region Proposal Network (RPN) for proposing regions of interest (RoI) from the feature maps output by the backbone. The RoIs are then fed into two parallel branches: the mask branch and the bounding box branch. The mask branch is a small fully convolutional network (FCN) applied to each RoI, predicting a segmentation mask in the pixel level. The box branch is a set of fully-connected (fc) layers that yield the clasoftmax and the bounding box regression. During the backpropagation, the lois calculated for each RoI: $L = L_{cls} + L_{box} + L_{mask}$, where L is the total training loss; L_{cls} is the loss the

(a) Vehicle (b) Vegetation

Fig. 13. Occlusions.

(a) (b)

Fig. 14. Examples of sloped street.

classification and L_{box} is the loof bounding box; and L_{mask} is the loof the mask. The detailed definitions of L_{cls}, L_{box} and L_{mask} are referred to [32] and [17].

4.6. Training

This section describes the training of the instance segmentation model. The input of the model is a raw image. The output of the model is polygons and bounding boxes, indicating the locations of the detected masonry buildings and the corresponding classes (confined/ unconfined).

90% of the annotated data is used for training the model. The initial base learning rate of the training is set as 0.0025. The training metrics of this model are presented in Fig. 9, which shows that the training converged after 10,000 iterations. The performance of a segmentation model can be evaluated by the average precision (AP) of all classes at different Intersection over Union (IoU) levels, which is defined as the intersection of the prediction area and ground truth area divided by the union of prediction and ground truth areas. Another popular metric for measuring the performance of segmentation models is the mean average precision (mAP) over a series of IoU levels. This model is tested on the left 10% of the ground truth annotations: the mAP is 71.41% at IoU $\geq$ 0.75, 78.97% at IoU $\geq$ 0.5, which is satisfactory. It should be noted that the presented model is trained and tested on data collected from the region of interested, where the construction types, development and economy levels are similar.

4.7. Inference

This section presents a few examples of the inference, followed by a discussion on the performance of the model. The first row of Fig. 10 shows the ground truth annotation and the inference of an unconfined masonry building, while the second row shows the example of a confined masonry building. Compared with the ground truth annotations, the masonry buildings are detected by the model and their classes are predicted correctly, with high confidence.

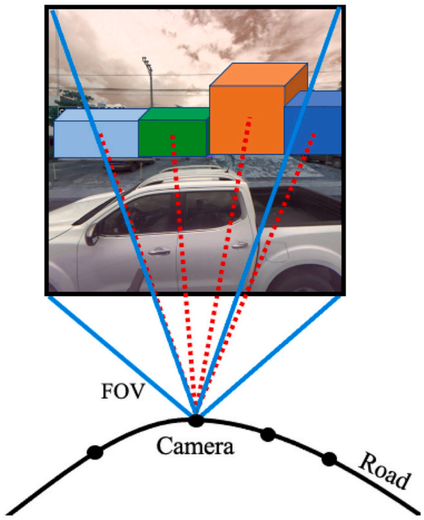


Fig. 15. Find the direction of each detected facade.

The generalization ability of the trained model is essential. In the following, we show a couple of examples for examining the model's performance under complicated street scenes. In dense urban area, buildings are often constructed adjacent to or totally connected with each other. It is important for the model to be able to identify each of them, instead of misclassifying them as one building. As the examples shown in Fig. 11, the model is capable of detecting multiple masonry buildings that are closely constructed.

Another common issue during the data collection is the varying distance between the camera and the building. For images collected from narrow streets, buildings appear in a closeup shot style. In this case, the buildings are partially captured in the image. On the opposite side, distant buildings usually appear in a long shot style, occupying only a very small fraction of the street view image. Both cases will cause difficulties for the model, especially if the model is trained on a relatively small dataset. Therefore, we intentionally included all these cases in the annotation effort to increase the diversity in the training dataset. Successful detection examples of such cases are shown in Fig. 12.

Another challenge encountered was the heavy occlusions in the street view, which are usually caused by trees and vehicles. For example, the facade of the building in Fig. 13(a) is occluded by a truck and (b) is occluded by trees, leaving only a part of the buildings visible in the image. As shown in the figures, the model successfully detected these buildings.

For sloped roads shown in Fig. 14, though the buildings in the images are leaning, they are also successfully detected.

4.8. Geocoding

The next step is to link the detected buildings to a city-scale building footprint database. The camera recorded the GPS coordinate and the heading direction of the travel when an image is captured. For each captured image, the field of view (FOV) is constant and known. Since the camera direction is perpendicular to the travel direction, it is straightforward to calculate the compass direction of each detected building based on its relative position in the image, as shown in Fig. 15.

In the horizontal plane, a line that originates from the camera and parallels the calculated compass direction can be drawn. The class detected from the image at the origin of the line will be assigned to the first footprint polygon the line encounters. Fig. 16 visualizes the lines that link the camera locations to the building footprints. A building could appear in multiple images captured at different camera location. Only the detection with the highest confidence is selected as the final prediction of the building.

4.9. Application example

The aforementioned model is trained on images collected from the city of Salina Cruz, Mexico. In this section, we use the model for detecting masonry buildings not only in Salina Cruz but also another city, Juchitán, in the same state.

Salina Cruz is a city located on the Pacific coast. It is a major seaport and the fourth-largest municipality of the state of Oaxaca, Mexico in terms of population. About 2,052,840 images are collected along the majority of the roads in the city, as shown in Fig. 17(a). 15,452 masonry buildings are detected from these images, among which 13,128 are classified as confined/reinforced masonry buildings, the left are classified as unconfined masonry buildings. Comparing Fig. 17(a) and (b), it shows that most confined masonry buildings (blue) are found in the areas where there the roads and streets are straight indicating engineered urbanization. Meanwhile, the density of unconfined masonry buildings (red) is much higher in the periphery of those urban areas.

Juchitán de Zaragoz is a city of similar size and is located 30 km to the northeast of Salina Cruz. We collected 695,524 images in the city along the roads shown in Fig. 18(a). 7029 buildings are detected from these images, among which 6315 are confined masonry buildings, the left are classified as unconfined masonry buildings. The screening results visualized in Fig. 18(b) show distribution patterns similar to Salina Cruz, i.e., unconfined masonry buildings are more frequently spotted on the periphery of the city.

5. Discussion

Although machine learning and deep learning have been employed for image-based information acquisition in multiple disciplines, using this technique for hazard risk mitigation has been rare. The objective of

Fig. 16. Register detections with footprints.

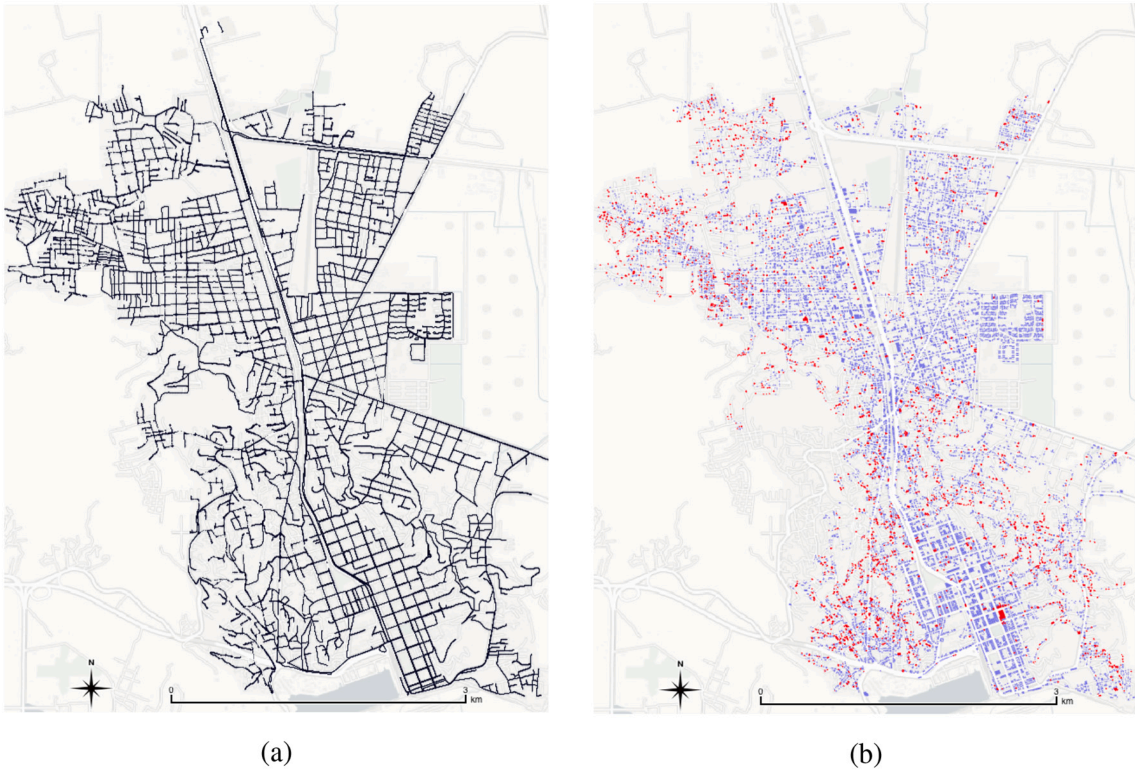


Fig. 17. Screening at Salina Cruz, Mexico: (a) Data collection routes (b) Screened building: blue is confined masonry buildings and red is unconfined masonry buildings. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

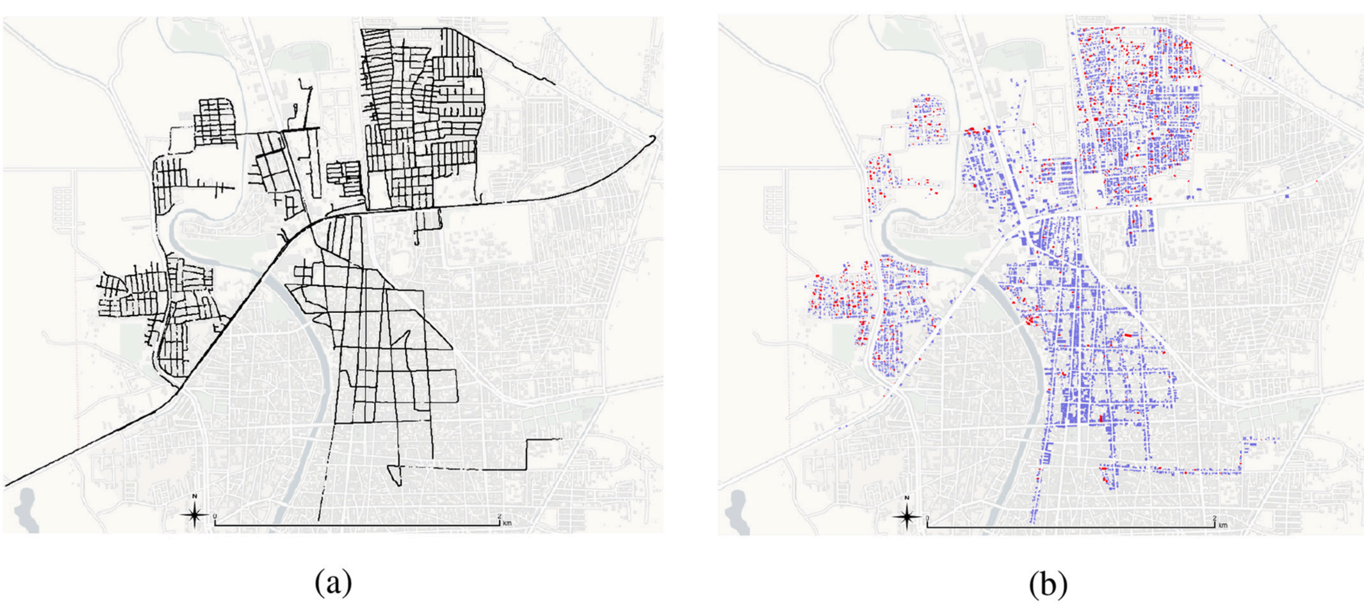
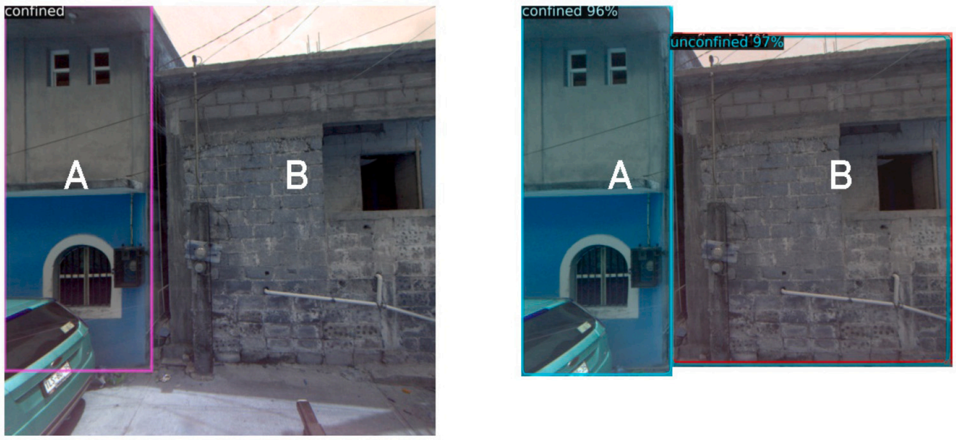


Fig. 18. Screening at Juchitán de Zaragoza, Mexico: (a) Data collection routes (b) Screened building: blue is confined masonry buildings and red is unconfined masonry buildings. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

this study was to explore the possibility of detecting buildings from images for assisting decision-making for earthquake preparedness. We demonstrated that deep convolutional neural networks (Mask R-CNN) can be used to help identify and classify masonry buildings using street view images.

A few prior studies investigated the application of using deep learning for street view image analysis, most if not all of them focused on the whole image classification approach, which is problematic because the image classification model makes decisions based on the whole image, without paying attention to the objects of interest in the image. Thus, not only the objects of the interest but also the noise are taken into consideration when deciding which class the building belongs to. In contrast to the existing literature that uses the full image classification method, this study explored using the segmentation technique in building information acquisition from street view images and has demonstrated that this technique can overcome the noise and occlusion



(a) No label, false prediction

(b) False label, correct prediction

(c) Model failed to detect the building

Fig. 19. Examples of false labels and false predictions.

problems.

To better understand the difficulties the model is facing and what confuses the model, it is important to discuss on examples of false predictions. A few annotated images (on the left) versus segmented building instances (on the right) with predictions are shown in Fig. 19(a–c). In (a), we focus on building B, which is not annotated. There are two predictions for this building: unconfined (97%) and confined (74%), which means the model found both confined and unconfined features in this building and the unconfined features outweighed the confined features. A closer look into building B reveals visual cues of confining components, but the cues are not strong, which confuses the model. In (b), the building is labeled as unconfined building, which is a false label because there are visual cues of confining components. The model successfully classified this building as 'confined'. In (c), the model failed to detect the buildings in the image. The most possible reason is: there were not enough annotated buildings similar to those in figure (d) – only the front facades are annotated in our training dataset, but figure (c) did not capture the front facades of those buildings. These examples show that (1) there are slight noises in our training dataset, like any other dataset. (2) The model encounters difficulties when the visual cues of confining components are weak. (3) The model performs best for front facades. Despite the aforementioned false prediction examples, the overall performance of the model is satisfying, with a 78.97% mean average precision on a randomly selected testing dataset.

6. Conclusion

The present study aimed at developing a method for the rapid detection of masonry buildings from street view images. The method used is instance segmentation, which is based on a deep convolutional neural network. A segmentation model was successfully trained that performs the following tasks:

- Detect masonry building instances from street view images
- Classify the identified instances of a masonry building in said image into two categories: confined and unconfined.

Using the developed framework, we can perform city scale detections of masonry buildings. In the region being studied, we found the spatial distribution pattern of masonry buildings is correlated with the urbanization paths of the cities.

This work appears to be the first study to use deep learning for the seismic screening of masonry buildings, and the first study that employs segmentation approach for building instance identification using street view images. When compared with the image classification method, one of the advantages of the present segmentation method is its ability to locate instances of buildings in an image. This leads to accurate classification of the building since the decision can be made on the building itself without interference from noise in the image itself. This has significant implications for using machine learning to understand building risks from images.

Although this method can be used for the rapid screening of masonry buildings, it is not intended to replace more rigorous assessments of building performance using numerical simulations or statistical methods. The present approach aims to provide an automated and inexpensive method for large-scale regional examinations that can benefit policymakers in determining seismic risk. With affordable, accurate and massive screenings, governments can target these buildings for retrofit efforts and companies can assess the business opportunity of prevention (or "Building Better Before"). This method, however, is intended to allow interested parties to improve the target of scarce resources and does not claim to be a substitute of much needed physical inspections and detailed assessments prior to the design and implementation of retrofits.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgments

This methodology, analysis and data was produced by the World Bank's Global Program for Resilient Housing (GPRH) as part of the technical support provided to Mexico's National Housing Commission (CONAVI).

References

[1] ATC, Rapid Visual Screening of Buildings for Potential Seismic Hazards, A Handbook, FEMA 154 Third Edition, Federal Emergency Management Agency, Washington DC, USA, 2015. ISBN 9780160926754.
[2] A. Bayraktar, N. CoŞkun, A. Yalçin, Damages of masonry buildings during the July 2, 2004 doğubayazıt (ağrı) Earthquake in Turkey, Eng. Fail. Anal. 14 (1) (2007) 147–157, https://doi.org/10.1016/j.engfailanal.2005.11.011.
[3] A.J. Bency, S. Rallapalli, R.K. Ganti, M. Srivatsa, B. Manjunath, Beyond spatial auto-regressive models: predicting housing prices with satellite imagery, in: Proceedings of the IEEE Winter Conference on Applications of Computer Vision, 2017, pp. 320–329, https://doi.org/10.1109/WACV.2017.42.
[4] S. Bhattacharya, S. Nayak, S.C. Dutta, A critical review of retrofitting methods for unreinforced masonry structures, Int. J. Disaster Risk Reduct. 7 (2014) 51–67, https://doi.org/10.1016/j.ijdrr.2013.12.004.
[5] G. Brando, D. Rapone, E. Spacone, M.S. O'Banion, M.J. Olsen, A.R. Barbosa, M. Faggella, R. Gigliotti, D. Liberatore, S. Russo, et al., Damage reconnaissance of unreinforced masonry bearing wall buildings after the 2015 Gorkha, Nepal Earthquake, Earthq. Spectra 33 (2017) 243–273, https://doi.org/10.1193/010817eqs009m.
[6] M. Bruneau, Performance of masonry structures during the 1994 northridge (Los Angeles) Earthquake, Can. J. Civil Eng. 22 (2) (1995) 378–402, https://doi.org/10.1139/l95-048.
[7] M. Bruneau, K. Yoshimura, Damage to masonry buildings caused by the 1995 Hyogo-Ken Nanbu (Kobe, Japan) Earthquake, Can. J. Civil Eng. 23 (3) (1996) 797–807. https://www.humanitarianlibrary.org/ (accessed 09.12.21).
[8] S. Brzev, M. Astroza, O. Moroni, Performance of Confined Masonry Buildings in the February 27, 2010 Chile Earthquake, Earthquake Engineering Research Institute (EERI), Oakland, California, 2010. https://www.confinedmasonry.org (accessed 09.12.21).
[9] S. Brzev, K. Mitra, Earthquake-Resistant Confined Masonry Construction, NICEE, National Information Centre of Earthquake Engineering, 2007. ISBN: 978-93-80903-06-4.
[10] Y.-J. Cha, W. Choi, G. Suh, S. Mahmoudkhani, O. Büyüköztürk, Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types, Comput.-Aided Civil Infrastruct. Eng. 33 (9) (2018) 731–747, https://doi.org/10.1111/mice.12334.
[11] T. Czerniawski, F. Leite, Automated segmentation of rgb-d images into a comprehensive set of building components using deep learning, Adv. Eng. Inform. 45 (2020) 101131, https://doi.org/10.1016/j.aei.2020.101131.
[12] D. Dizhur, J. Ingham, M. Griffith, D. Biggs, A. Schultz, Performance of unreinforced masonry and infilled RC buildings during the 2015 Gorkha, Nepal Earthquake sequence, in: 16th International Brick and Block Masonry Conference, IBMAC, 2016, pp. 2399–2408, https://doi.org/10.1201/b21889-314.
[13] Y. Gao, K.M. Mosalam, Deep transfer learning for image-based structural damage recognition, Comput.-Aided Civil Infrastruct. Eng. 33 (9) (2018) 748–768, https://doi.org/10.1111/mice.12363.
[14] T. Gebru, J. Krause, Y. Wang, D. Chen, J. Deng, E.L. Aiden, L. Fei-Fei, Using deep learning and google street view to estimate the demographic makeup of neighborhoods across the United States, Proc. Natl. Acad. Sci. USA 114 (50) (2017) 13108–13113, https://doi.org/10.1073/pnas.1700035114.
[15] D. Gonzalez, D. Rueda-Plata, A.B. Acevedo, J.C. Duque, R. Ramos-Pollan, A. Betancourt, S. Garcia, Automatic detection of building typology using deep learning methods on street level images, Build. Environ. 177 (2020) 106805, https://doi.org/10.1016/j.buildenv.2020.106805.
[16] J. Guo, Q. Wang, Y. Li, P. Liu, Façade defects classification from imbalanced dataset using meta learning-based convolutional neural network, Comput.-Aided Civil Infrastruct. Eng. (2020) 1403–1418, https://doi.org/10.1111/mice.12578.
[17] 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.1995.466933.
[18] K. Jagadish, S. Raghunath, K.N. Rao, Behaviour of masonry structures during the Bhuj Earthquake of January 2001, J. Earth Syst. Sci. 112 (3) (2003) 431–440, https://doi.org/10.1007/BF02709270.
[19] J. Kang, M. Körner, Y. Wang, H. Taubenböck, X.X. Zhu, Building instance classification using street view images, ISPRS J. Photogramm. Remote Sens. 145 (2018) 44–59, https://doi.org/10.1016/j.isprsjprs.2018.02.006.
[20] H. Kato, S. Tajiri, T. Mukai, Preliminary Reconnaissance Report of the Chile Earthquake 2010, Building Research Institute Japan, 2010. http://www.kenken.go.jp (accessed 09.12.21).
[21] J. Kiyono, A. Kalantari, Collapse Mechanism of Adobe and Masonry Structures During the 2003 Iran Bam Earthquake, Bulletin of the Earthquake Research Institute, University of Tokyo, 2004, pp. 157–161. http://www.researchgate.net (accessed 09.12.21) 79.
[22] R.E. Klingner, Behavior of masonry in the northridge (US) and Tecomán-colima (Mexico) Earthquakes: lessons learned, and changes in US design provisions, Construct. Build. Mater. 20 (4) (2006) 209–219, https://doi.org/10.1016/j.conbuildmat.2005.08.024.
[23] S. Law, B. Paige, C. Russell, Take a Look Around: Using Street View and Satellite Images to Estimate House Prices, 2018. arXiv:1807.07155 (arXiv preprint).
[24] X. Liu, Q. Chen, L. Zhu, Y. Xu, L. Lin, Place-centric visual urban perception with deep multi-instance regression, in: Proceedings of the 25th ACM International Conference on Multimedia, 2017, pp. 19–27, https://doi.org/10.1145/3123266.3123271.
[25] R. Meli, S. Brzev, M. Astroza, T. Boen, F. Crisafulli, J. Dai, M. Farsi, T. Hart, A. Mebarki, A. Moghadam, et al., Seismic Design Guide for Low-Rise Confined

Masonry Buildings, Earthquake Engineering Research Institute (EERI), 2011. https://www.humanitarianlibrary.org/ (accessed 09.12.21).
[26] L.M. Moon, D. Dizhur, J.M. Ingham, M.C. Griffith, Seismic performance of masonry buildings in the Christchurch Earthquakes 2010–2011: a progress report, in: Australian Earthquake Engineering Society 2012 Conference, 2012. https://www.aees.org.au (accessed 09.21.21).
[27] F. Mulargia, P.B. Stark, R.J. Geller, Why is probabilistic seismic hazard analysis (psha) still used? Phys. Earth Planet. Inter. 264 (2017) 63–75, https://doi.org/10.1016/j.pepi.2016.12.002.
[28] N. Naik, J. Philipoom, R. Raskar, C. Hidalgo, Streetscore-predicting the perceived safety of one million streetscapes, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2014, pp. 793–799, https://doi.org/10.1109/CVPRW.2014.121.
[29] M. Ningthoujam, R.P. Nanda, Rapid visual screening procedure of existing building based on statistical analysis, Int. J. Disaster Risk Reduct. 28 (2018) 720–730, https://doi.org/10.1016/j.ijdrr.2018.01.033.
[30] D. Perrone, M.A. Aiello, M. Pecce, F. Rossi, Rapid Visual Screening for Seismic Evaluation of RC Hospital Buildings. Structures, vol. 3, Elsevier, 2015, pp. 57–70, https://doi.org/10.1016/j.istruc.2015.03.002.
[31] S. Ploeger, M. Sawada, A. Elsabbagh, M. Saatcioglu, M. Nastev, E. Rosetti, Urban rat: new tool for virtual and site-specific mobile rapid data collection for seismic risk assessment, J. Comput. Civil Eng. 30 (2) (2016), https://doi.org/10.1061/(ASCE)CP.1943-5487.0000472, 04015006.
[32] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, Adv. Neural Inf. Process. Syst. 28 (2015) 91–99, https://doi.org/10.1109/TPAMI.2016.2577031.
[33] D. Rueda-Plata, D. González, A. Acevedo, J. Duque, R. Ramos-Pollán, Use of deep learning models in street-level images to classify one-story unreinforced masonry buildings based on roof diaphragms, Build. Environ. 189 (2021) 107517, https://doi.org/10.1016/j.buildenv.2020.107517.
[34] N.M. Wallace, T.H. Miller, Seismic screening of public facilities in Oregon's western countries, Pract. Period. Struct. Des. Constructi. 13 (4) (2008) 189–197, https://doi.org/10.1061/(ASCE)1084-0680(2008)13:4(189.
[35] C. Wang, Q. Yu, K.H. Law, F. McKenna, X.Y. Stella, E. Taciroglu, A. Zsarnóczay, W. Elhaddad, B. Cetiner, Machine learning-based regional scale intelligent modeling of building information for natural hazard risk management, Autom. Construct. 122 (2021) 103474, https://doi.org/10.1016/j.autcon.2020.103474.
[36] Q. Yu, C. Wang, B. Cetiner, S.X. Yu, F. Mckenna, E. Taciroglu, K.H. Law, Building Information Modeling and Classification by Visual Learning at a City Scale, 2019. arXiv:1910.06391 (arXiv preprint).
[37] Q. Yu, C. Wang, F. McKenna, X.Y. Stella, E. Taciroglu, B. Cetiner, K.H. Law, Rapid visual screening of soft-story buildings from street view images using deep learning classification, Earthq. Eng. Eng. Vib. 19 (4) (2020) 827–838, https://doi.org/10.1007/s11803-020-0598-2.