Title
A Data-Driven Framework for Regional Assessment of Seismically Vulnerable Buildings

Author
Chen, Peng-Yu

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

A Data-Driven Framework for Regional Assessment of Seismically Vulnerable Buildings

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Civil Engineering

by

Peng-Yu Chen

2021

ABSTRACT OF THE DISSERTATION

A Data-Driven Framework for Regional Assessment of Seismically Vulnerable Buildings

by

Peng-Yu Chen
Doctor of Philosophy in Civil Engineering
University of California, Los Angeles, 2021
Professor Ertugrul Taciroglu, Chair

The urban region's seismic resilience is being actively studied in recent years as a measure for risk mitigation, where the identification of seismically vulnerable buildings and the assessment of their performance play indispensable roles. However, it is a labor-intensive and computationally expensive task to evaluate tens of thousands of buildings in a region because the identification requires professional judgment at a site and the seismic assessment demands comprehensive modeling depending on structure-specific data. Nevertheless, it is feasible with the aid of advanced development of the Internet of Things (IoT) and computer technology. In this study, a data-driven framework including two pipelines that focus on soft-story buildings and non-ductile reinforced concrete frames is proposed.

The first pipeline focuses on identifying soft-story buildings in the city of Santa Monica (California) through 3D point clouds and convolutional neural networks (CNNs). Although prior studies showed promising results in detecting soft-story buildings based on well-selected street-view images, false predictions are common when it is applied to real-world data. To address this issue, the pipeline implements point-cloud data where spatial information is available to segment building points and extract density features for training deep learning

models and identifying soft-story buildings. The transfer learning (TL) technique is adopted to avoid overfitting in deep neural networks, and the parameters within the pipeline are investigated for optimal performance. The results illustrate the potential applicability of the pipeline for developing pre- and post-event countermeasures.

The second pipeline focuses on another seismically vulnerable building, namely, the non-ductile reinforced concrete building (NDRCB). Prior studies indicated around 1,500 NDRCBs in Los Angeles that are urgently waiting for detailed assessment and mandatory retrofit or demolition if necessary. Because the fulfillment of these ordinances will last for decades, the potential risk of major losses will persist. To this end, an automatic method that harvests building information from IoT and imagery data, generates archetypal models, conducts probabilistic seismic assessment, and estimates the losses for NDRCB frames is hence developed. The accuracy of the data harvesting module using deep CNNs is validated with the existing inventory data. The archetypal frames are developed based on the era-specific representative code and are validated through nonlinear static and nonlinear dynamic analyses of previously investigated NDRCBs. State-of-the-practice loss estimation methodologies including HAZUS and FEMA P-58 are adopted in the pipeline for constructing damage fragility functions and corresponding losses. The regional application focuses on intensity-based assessment for thousands of individual buildings instead of a scenario-based assessment. The outcomes of expected losses and repair/reconstruction time emphasize the vulnerability of NDRCBs in Los Angeles, and the presented pipeline is believed to bridge the gaps between property owners, engineers, and decision-makers.

This research demonstrates how advanced data mining techniques and data-driven approaches can aid to solve civil engineering problems. While the framework currently focuses on soft-story and non-ductile frame buildings, it is expected to be extended in depth and breadth in the future. That is, more detailed models and other seismically vulnerable infrastructures can be included.

The dissertation of Peng-Yu Chen is approved.

Henry Burton

Jian Zhang

Ying Nian Wu

Zheng Yi Wu

Ertugrul Taciroglu, Committee Chair

University of California, Los Angeles

2021

To my parents . . .

who—among so many other things—

encouraged me discover science

since my childhood

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor, Professor Ertugrul Taciroglu, for his guidance, encouragement, and ingenuity. This dissertation would not have been possible without his steadfast support. I would like to thank Professors Henry Burton, Jian Zhang, Ying Nian Wu, and Dr. Zheng Yi Wu for serving on my dissertation committee and for their insightful comments. Special thanks to Dr. Zheng Yi Wu for his supervising my research internship at Bentley Systems.

Greatest appreciation to my loving family in Taiwan. Their selfless support allows me to fulfill my dreams. Of course, I have to express my gratitude to my lover Li-Hsuan for being considerate and understanding. Although we live in different places, our hearts are always connected.

Finally, I would like to thank my friends at UCLA. Dr.Abdoul Ghotbi, Dr.Wenyang Zhang, Dr.Mehrdad Shokrabadi, Dr.Han Sun, Dr.Yu Zhang, Dr.Dong Wang, Dr.Hua Kang, Dr.Xingquan Guan, Dr.Zhenxiang Yi, Tien-Shu Chang, Hao Lee, Chaowei Hu, Shuochuan Meng, Henan Mao and many others, who have offered help to me during the past five years.

Financial supports in the form of university fellowships and tuition waiver from the Department of Civil Engineering are also appreciated.

VITA

2009–2013 B.S. in Civil Engineering, National Central University, Taiwan

2013–2014 M.S. in Civil Engineering, National Central University, Taiwan

2018–2021 M.S. in Statistics, Universtiy of California, Los Angeles, USA

2016–2021 Ph.D. Candidate in Civil Engineering, Universtiy of California, Los Angeles, USA

SELECTED PUBLICATIONS

Chen, P.Y., Wu, Z. Y., and Taciroglu, E. (2021). "Classification of soft-story build-ings using deep learning with density features extracted from 3D point clouds" *Journal of Computing in Civil Engineering*, 35(3), 04021005. (published)

Chen, P.Y., Lesgidis, N. Y., Cetiner, B., Sextos, A., and Taciroglu, E. (2021). "A method for Automated Development of Model & Fragility Inventories of Non-ductile Reinforced Concrete Frame Buildings" *Earthquake Spectra* (under 2$^{\text{nd}}$-round review).

Chen, P.Y., Meng, S., and Taciroglu, E. (2021). "A data-driven framework for regional seismic loss estimation: a case study of non-ductile reinforced concrete buildings," *Engineering Structure* (ready to submit).

CHAPTER 1

Introduction

1.1 Background and Motivation

Seismic hazard is among one of the natural disasters that pose significantly destructive impacts to society, from damaging buildings, endangering lives to economic loss. From past earthquakes around the world, buildings having deficiencies in resisting seismic loads are identified to be the cause of loss of life and property. The vulnerability of these buildings is mainly because of era-specific design considering economic development and population growth, and the lacking of the concept of seismic requirements. The most ubiquitous cases are soft-story and non-ductile reinforced concrete buildings (NDRCBs) whose risks were evident in events including the 1971 San Fernando, 1989 Loma Prieta, 1994 Northridge, 1999 Chi-Chi, 2008 Sichuan, and 2009 L'Aquila earthquakes [42, 53, 62, 65, 92, 107, 120]. The existence of these seismically vulnerable buildings affects the losses (e.g., repair, replacement), post-earthquake occupancy rates, economic recovery, and even the number of fatalities during an earthquake, all of which can be used to explicitly quantify seismic resilience [18, 20] that is gradually valued in urban planning and risk mitigation.

Given that the built environment is made up of millions of buildings, a regional (e.g., city-scale) assessment is hence necessary because it provides valuable information that can be used for emergency preparedness in high seismic regions and populous modern cities. In a regional analysis, the following components are inevitable: (i) identifying target buildings and harvesting their fundamental information, (ii) constructing region-specific and structure-

specific representative analyses, and (iii) estimating potential damages and losses given a seismic hazard. While each ingredient has extensive background studies, many limitations and technical details render the practical application at a city scale. As a result, this study aims to propose a data-driven framework to integrate the aforementioned components for regional seismic assessment, which is expected to bridge knowledge gaps between property owners, engineers, and decision-makers.

Many inventories [11, 25, 26] have been done to identify the existence of seismically vulnerable buildings in California, which prompts ordinances for mandatory evaluation and retrofit of them if required. Specifically being discussed herein, the City of Los Angeles' Resilience by Design initiative (`https://www.lamayor.org/resilience-design-building-stronger-los-angeles`) released a task force to assess pre-1980 soft-story and NDRCBs. Many studies have also investigated and recommended retrofitting strategies [8, 19, 27, 35, 129]. Until this dissertation is being written, around 90% soft-story buildings have been assessed, 60% have been issued permits, and 40% have completed retrofit. However, less than 20% of non-ductile buildings have been assessed (`https://www.ladbs.org/services/core-services/plan-check-permit/plan-check-permit-special-assistance/mandatory-retrofit-programs`).

Given this circumstance, the motivations of choosing these buildings are quite different in the proposed framework where two pipelines are included as shown in Fig.1.1. For the soft-story buildings, the pipeline only focuses on improving the identification of these buildings at a city scale. Existing studies used 2D street-view images and recently developed deep learning (DL) techniques, which can only provide promising results when well-controlled images are given. To address this issue, the pipeline uses city-scale point cloud data, extracts density features, and identifies soft-story buildings through deep CNNs.

On the other hand, NDRCBs urgently need to be evaluated, so the specific pipeline aims to collect metadata for automated modeling and damage assessment of non-ductile buildings. Considering the scope of inventory, image-based DL techniques are used herein

to harvest building information. Analytical models are automatically developed following era-specific representative codes, and state-of-the-practice seismic assessment methodologies are implemented to estimate building responses, damage fragility functions, and losses. The automation of the process is trying to resolve practical restrictions and difficulties for the regional application.

The following sections will briefly introduce the backgrounds and objectives of the two pipelines, and the data mining module shown in Fig.1.1 will be discussed later when it is applied to the inventory of NDRCBs.

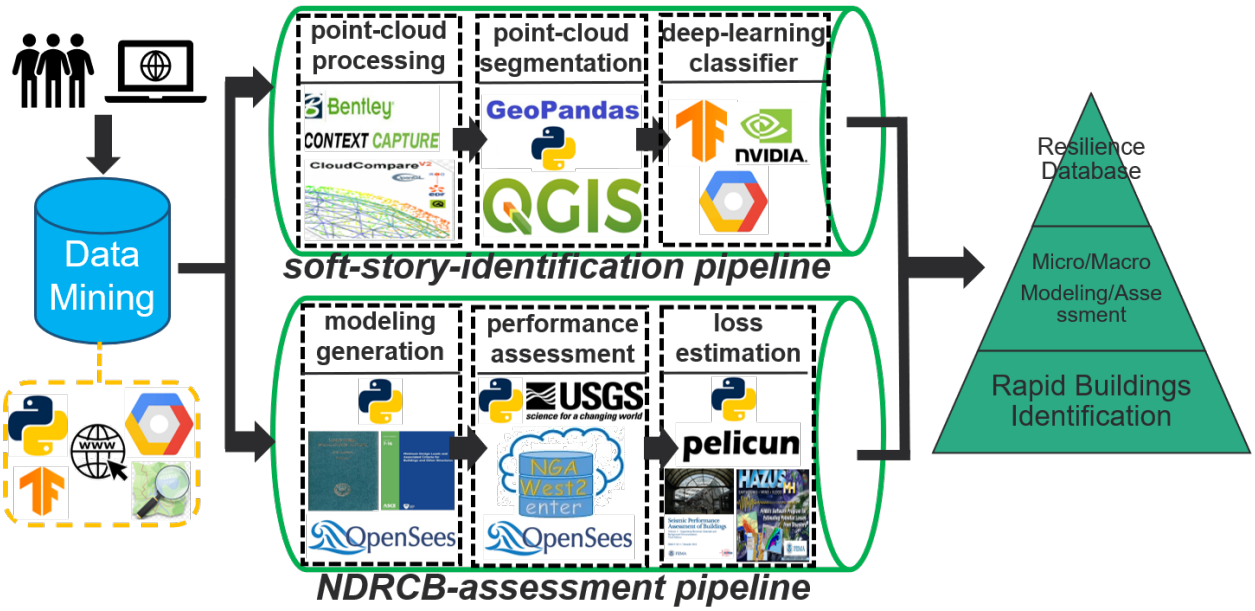


Figure 1.1. Overview of the proposed framework.

1.2 Identification of Soft-Story Buildings

A typical weakness of soft-story buildings is their open-space configuration on lower levels [11,42,129] as shown in Fig. 1.2. The open-space feature of soft-story buildings is a major hint for engineers to rapidly identify structural irregularities. Automating the detection of this characteristic is feasible and reliable given the recent developments of artificial intelligence

(AI) and DL in computer vision technology.

Convolutional neural networks (CNNs) are at the core of recent DL development. Distinguished from traditional computer vision and machine learning (ML) techniques, CNNs have the ability to capture millions of parameters learned from imagery data (pixel matrix). Many well-designed CNN architectures, such as VGGNet [112], GoogLeNet [117], and ResNet [58] have demonstrated their dominance in image recognition tasks. Several studies in civil engineering applications have used CNN-based models to detect and localize structural damage. For example, Zhang et al. [134, 135] developed a CNN-based pixel-level crack detection model, CrackNet, which outperforms traditional ML methods for detecting cracks on asphalt pavement. Alipour et al. [2] proposed a CNN-based model, CrackPix, to detect crack damage based on patch-level images, which allowed random-size input images. Cha et al. [22] designed a CNN architecture to detect crack damage in concrete structures, and their model was less sensitive to real-world situations (e.g., lighting, shadow changes). In addition to the application to imagery data, Kumar et al. [80] proposed a DL-based object detection model for automated defect classification and localization in a closed-circuit television video of sewers. Three object detection models, SSD [88], YOLOv3 [105], and Faster R-CNN [106] were evaluated for accuracy and speed in defect detection. The Faster R-CNN model demonstrated the highest accuracy in sewer inspection but required the most computational time. Although the above-mentioned studies improved the labor-intensive efforts for the inspection of civil infrastructure, the training of such CNN-based models from scratch is time-consuming and requires large amounts of data.

(a)

(b)

Figure 1.2. Typical soft-story buildings.

Transfer learning (TL) was developed to improve the training process and performance of CNN-based models [97]. In TL, knowledge learned from well-trained models with a base data set is transferred to a new model for retraining or fine-tuning on a smaller data set. Gao et al. [47] implemented TL for vision-based damage recognition at the pixel, component, and structural levels. In their study, the VGGNet model was retrained and fine-tuned for multilabel structural damage classification tasks. Kim et al. [75] established a region-based fully convolutional network to detect construction equipment for site management, where the ResNet model was implemented as the base model in TL. Kalfarisi et al. [72] developed two

DL-based approaches for crack detection and segmentation. The first approach integrated Faster R-CNN with structured random forest edge detection [38], whereas the second approach applied Mask R-CNN [57] with pre-trained CNN models. Their results showed that Mask R-CNN with Inception-ResNet-V2 [116] network architecture obtained higher crack detection and segmentation performance. In addition to component damage recognition and detection, some studies have implemented DL techniques for classification and detection at the regional scale. Kang et al. [73] trained VGGNet, GoogLeNet, and ResNet based on street-view images to classify building land use, and Srivastava et al. [115] used CNN models to classify land use based on aerial and street-view images.

For regional identification of soft-story buildings, Wu et al. [126, 127] and Yu et al. [132] developed TL-based models and achieved high accuracy of detection in street-view images. The detection confidence and accuracy decrease the more the buildings are blocked by objects, as shown in Fig. 1.3 (a) to (c). To overcome this limitation of street-view image-based methods, 3D point cloud data can be used to capture the soft-story features of buildings that are blocked or do not face the street. Fig.1.3 (d) shows an example of point cloud data (left panel) for a soft-story building blocked by objects (right panel); the point cloud data can be individually segmented from other objects. The objective of this pipeline is hence to develop a workflow that automatically segments point clouds and identifies soft-story buildings by applying DL models at the city scale. An overview of this pipeline is shown in Fig.1.4, and the detailed implementation will be discussed in Chapter 2

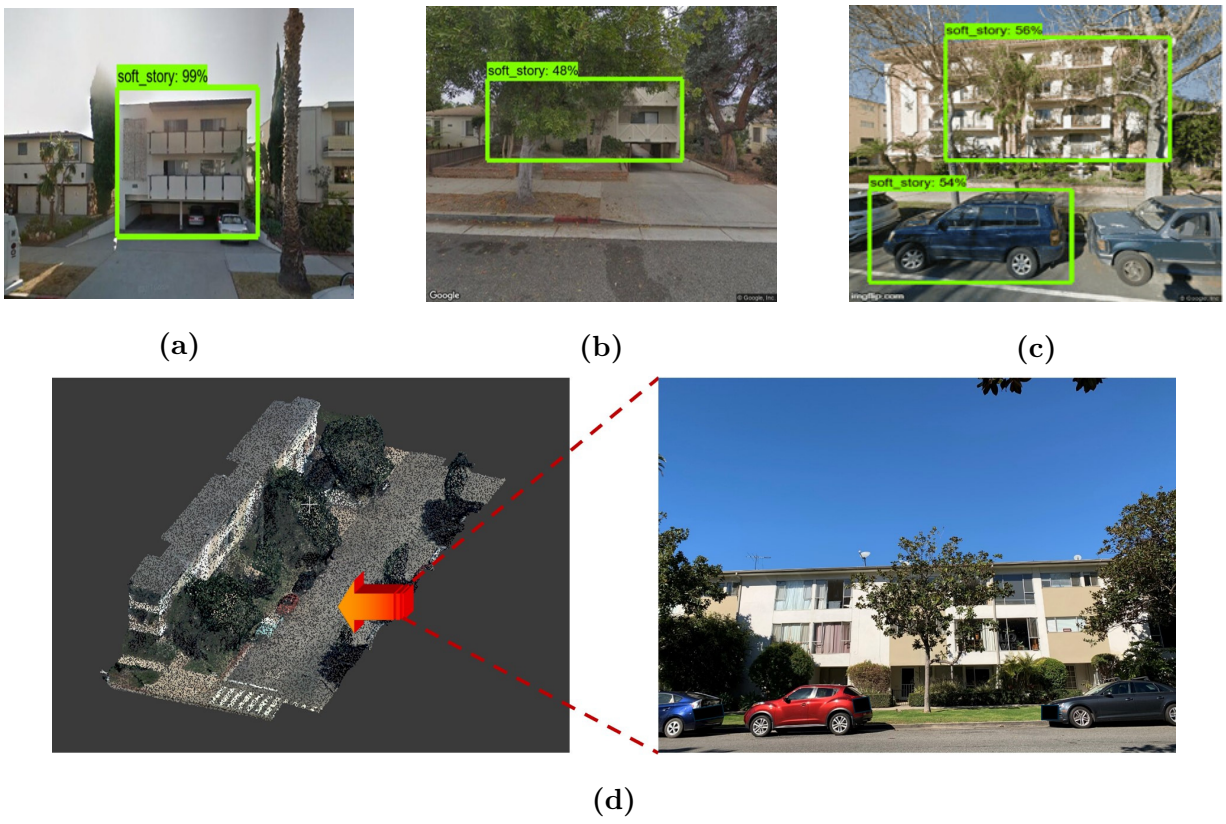


Figure 1.3. Soft-story detection with (a) non-blocked, (b) partially blocked, and (c) fully blocked images. (d) An example of soft-story point-cloud data.

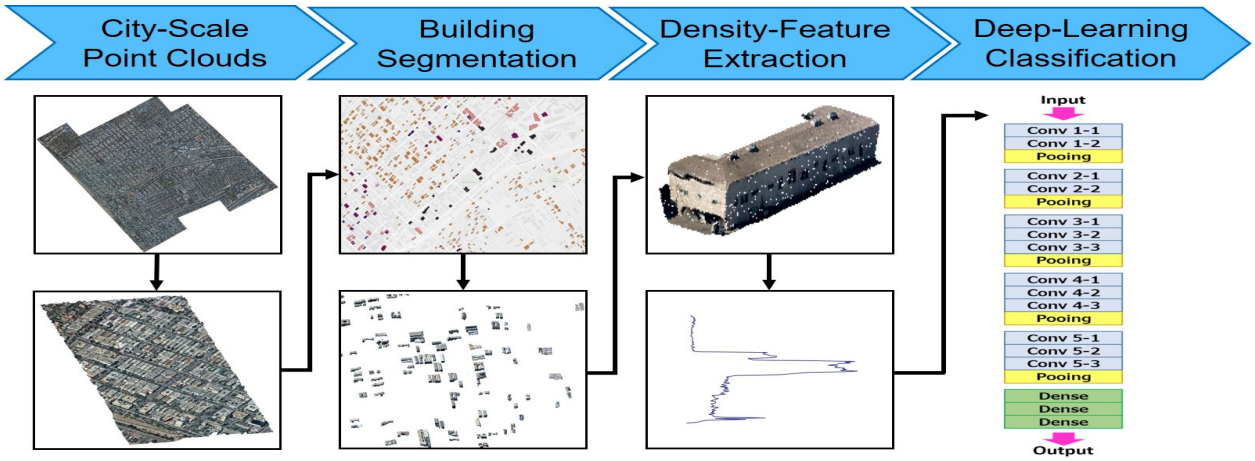


Figure 1.4. Schematic of identifying soft-story buildings.

1.3 Seismic Assessment of Non-Ductile Reinforced Concrete Buildings

Numerous efforts have been undertaken to identify and generate reliable databases of NDRCBs during the past decade to understand the risks posed by these collapse-prone buildings and to develop countermeasures. The Concrete Coalition—a network of engineer volunteers—identified 20,000 to 23,000 NDRCBs in California [25, 26], including residential, commercial, and public (e.g., school) buildings. The volunteer investigation took three years (2008 to 2011) to assemble a variety of data sources ranging from county assessor's files to satellite images. Site surveys were also conducted to corroborate data collected from multiple sources.

With the experience of Concrete Coalition, a case study of the City of Los Angeles was then conducted and approximately 1,500 NDRCBs were identified [5–7, 27]. The inventory has been compiled using more than 15 data sources and the attributes include: location, structural type, year built, number of stories, total size in square footage, and building usage. To address the seismic loss and the mitigation priority, Anagnos et al. [8] broke down the inventory into multiple groups that represent common construction typologies and estimated the loss for scenario events. The mandatory retrofit program of the City of Los Angeles took effect in 2005, yet the fulfillment of the ordinance will last for decades and more than one thousand NDRCBs are still pending assessment to determine their compliance with this ordinance.

Considering the expensive cost of on-site investigation, the complex modeling procedure, and the large demand for assessing thousands of buildings, this pipeline is inspired and developed. The goal of the pipeline is to: **(i) collect metadata through text and visual information available on the IoT to reduce labor-intensive site investigation; (ii) construct detailed and computationally efficient models to approximate as-built structures; and (iii) perform a large number of nonlinear dynamic analyses to construct fragility functions for seismic risk assessment.** An overview of this pipeline

is shown in Fig.1.5, and the detailed implementation will be discussed in Chapters 3 and 4.

This study is defined as pilot research, so it is expected to be extended and applied to any type of disaster and infrastructure in the future. However, we currently focus on moment-resisting frame buildings to narrow down the problem scale. According to prior studies [8,27], 64% of the total square footage and replacement values in Los Angeles' inventory concentrates on high-rise buildings (37%) where most of them have moment frames with shear-critical columns. These structures are thought to have a higher risk for regional loss estimation. While the pipeline is initially limited to frame structures, concrete construction with walls or masonry infill is expected to be included in the future.

In addition to assisting regional assessment, this pipeline is also developed to fill the knowledge gap of NDRCBs. For example, suitable intensity measures (IMs) for seismic assessment. IM plays a key role to connect seismic hazards and structural responses as it can provide an accurate prediction of building vulnerability with minimal uncertainty [34,78]. Although many researchers have studied optimal IM for a variety of structures, few of them focused on NDRCBs [4,66,67]. Besides, only limited types of frames (i.e., low-rise, mid-rise) and engineering demand parameters (i.e., maximal inter-story drift) were considered among them. Hence, the proposed pipeline is used to identify and recommend suitable IMs for the NDRCB frame based on efficiency, sufficiency, proficiency, and practicality [77, 89, 110].

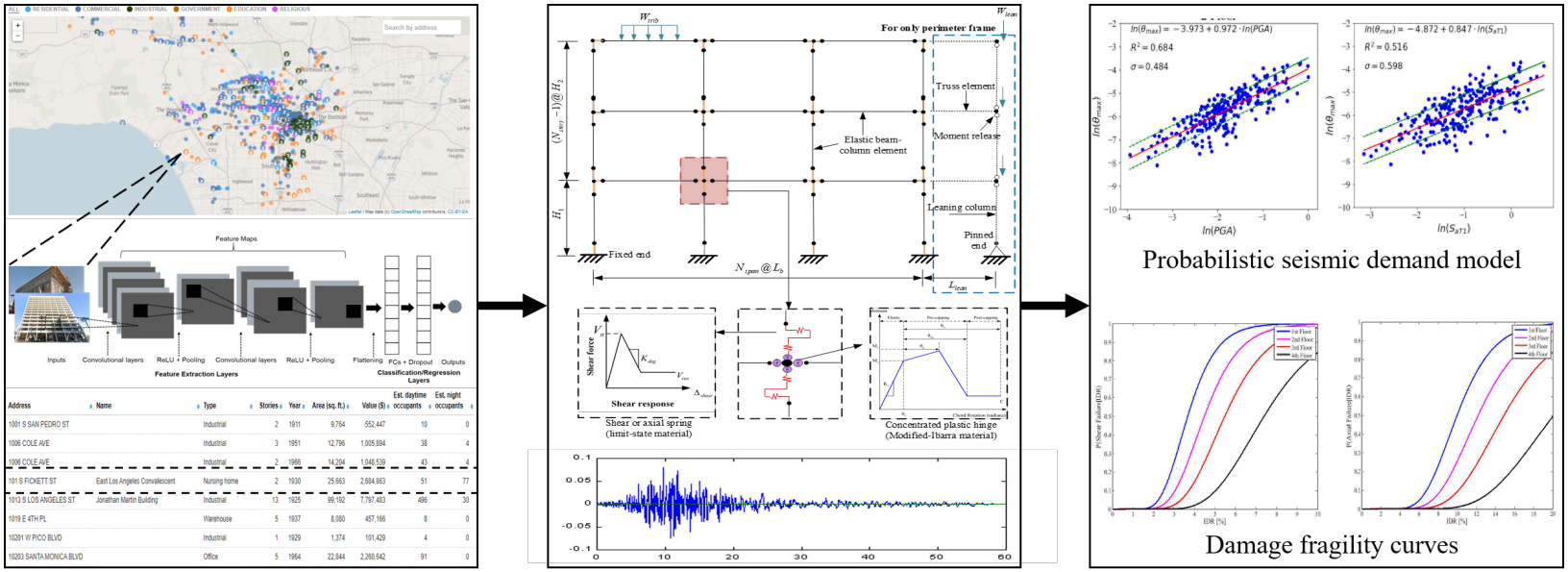


Figure 1.5. Schematic of assessing NDRCBs.

Another motivation of the study is to address limitations in applying state-of-the-practice risk assessment methodologies at a city scale. In particular, the multi-hazard loss estimation methodology (HAZUS [43]), and the FEMA P-58 method [44]. Buildings are classified into different groups based on their built year, material, lateral force resisting system, and the number of floors in HAZUS which was developed to provide a guideline for regional risk assessment and emergency response. The capacity spectrum method is used in HAZUS to obtain the demand and capacity of a single-degree-of-freedom (SDOF) system subject to a push-over load for different building groups, which is the advantage of HAZUS because it can rapidly assess a building-level performance and corresponding losses. However, the shortage of HAZUS is that the component-level performance is not available because the SDOF cannot differentiate the structural response. Besides, the capacity spectrum method cannot reflect the ground motion characteristic. On the other hand, FEMA P-58 method was developed by following the next-generation performance-based seismic design. In the P-58 method, the building performance is evaluated based on groups of components (e.g., columns at the first story, beam-column joints at the top floor), which requires comprehensive structural responses at different floors. P-58 method is believed to be able to provide much detailed seismic performance and loss estimation for a single building as long as refined modeling and dynamic analyses are conducted. While the P-58 method is dominant for a single building, it is challenging to apply it to a region because of the difficulty of obtaining detailed seismic responses. Given the shortage of HAZUS and the demand of the P-58 method, the contributions of this study are highlighted.

1.4 Organization and Outline

This dissertation consists of five chapters and their contents are elaborated below:

Chapter 1 discusses the background, motivation, and objectives of the research.

Chapter 2 begins with an overall description of the pipeline, followed by the segmentation of point-cloud data, and the calculation of density features for training the DL models to identify soft-story buildings. Multiple experiments are conducted to investigate the optimal parameters based on performance metrics. The TL techniques and state-of-the-art CNN models are then implemented to improve the performance. The final section summarizes the conclusions drawn for this pipeline.

Chapter 3 focuses on the automated development of NDRCB frames in the widely used Open System for Earthquake Engineering Simulation (OpenSees [91]). A series of reviews of prior and selected modeling techniques are discussed, followed by an elaboration of implementing era-specific code (i.e., Uniform Building Code 1976 [64]) to design archetypal frames. The procedure is then validated through nonlinear static and nonlinear dynamic analyses. The probabilistic seismic demand model (PSDM, [31]) is introduced to develop damage fragility functions and investigate suitable IMs for NDRCB frames. A sensitivity analysis is provided to identify the variation of structural responses computed through the proposed pipeline, which is followed by a summary of the chapter.

Chapter 4 applies the proposed pipeline of assessing NDRCB frames to Los Angeles' inventory. Python-based web scraping tools are conducted and combined with 2D image-based DL models to harvest the building information required in automated modeling. Hundreds of thousands of nonlinear time-history analysis (NTHAs) are carried out through high-performance cluster Hoffman2 at the University of California, Los Angeles. The chapter ends with an estimation of the potential seismic losses of NDRCBs in L.A., which is computed through HAZUS and P-58 methods.

Chapter 5 concludes the dissertation and provides a roadmap for future work.

CHAPTER 2

Pipeline for Identifying Soft-Story Buildings Using 3D Point-Cloud and Deep Learning

---

This chapter is partly based on the following publication:

Chen, P.Y., Wu, Z. Y., and Taciroglu, E. (2021). "Classification of soft-story buildings using deep learning with density features extracted from 3D point clouds," *Journal of Computing in Civil Engineering*, 35(3), 04021005. (published)

2.1 Overview

As introduced in the previous chapter (shown in Fig.1.4) the proposed automated pipeline for identifying soft-story buildings consists of four components: (1) preparation of city-scale point clouds; (2) segmentation of building point clouds using ground-truth soft-story building footprints; (3) computation of density-height variation plots; and (4) soft-story identification using CNNs.

The DL classifier starts with a naïve CNN architecture which is used to investigate the optimal hyper-parameters for the proposed pipeline. State-of-the-art CNNs are then introduced through TL techniques to improve the performance. Fundamental components of CNN (e.g., convolutional layer, maximum pooling, rectified linear unit, fully-connected layer) and selected models for TL will be introduced in this chapter, which is also used in

the next pipeline for the assessment of non-ductile concrete buildings.

2.2 Point-Cloud Data

Point clouds are a set of data points in space, where numerous points represent a 3D shape or object and each point has its set of X, Y, and Z coordinates. Point cloud data have been actively studied to reconstruct infrastructural models for facility management [23, 24], and post-event damage evaluation [74, 136] in recent developments due to its scalability of depth and dimension of objects. As shown in Fig.1.3 (d), the open-space feature of soft-story buildings can be reliably captured by 3D point clouds, but it may not always be visible in 2D images because it may be blocked from view by the orientation of the building and/or by objects (e.g., trees and cars). Furthermore, the point density of an opening space should be sparse compared with other regions, thus a point density feature is considered as a characteristic of soft-story buildings. As a result, point cloud data is selected in this study to improve the identification of soft-story buildings.

The city of Santa Monica conducted a virtual survey for seismically vulnerable buildings in 2015 and enacted an ordinance for a mandatory retrofit program in 2017 (`https://www.smgov.net/Departments/PCD/Programs/Seismic-Retrofit/`). Among the 2,000 seismically vulnerable buildings, 1,427 were identified as soft-story buildings and required to be retrofitted by July 2026. Santa Monica is selected as the target city in this study because of the extensive data inventory, including data on soft-story buildings. To generate point cloud data, 3D scanners or photogrammetry software are often used, which measure many points on the external surfaces of objects around them. The light detection and ranging (LiDAR) sensor is a well-known 3D scanner using ultraviolet, visible, or near-infrared light to measure the spatial relationships and shapes of objects and environments. While the application in civil engineering ranges from micro-damage detection of structural components [1] to macro-vulnerability estimation of an urban region [118], the expense to scan a city is too high.

Photogrammetry is implemented herein to create a 3D point cloud model for Santa Monica due to its affordability. In photogrammetry, evenly distributed points on images (e.g., street-view images) that depict an object are used to estimate the camera position with known and georeferenced 3D coordinates [123]. In this study, a well-developed photogrammetry software (ContextCapture [15]) is used to automatically register the shared points among images and calculates the distances in 3D space. The output is a set of point clouds with coordinates and appearance textures (e.g., R, G, B) as shown in Fig.2.1 For Santa Monica (as shown in Fig.2.2), 1.1 billion point cloud data points are generated, covering 18.16 km^2 at a density of 60 points/m^2.

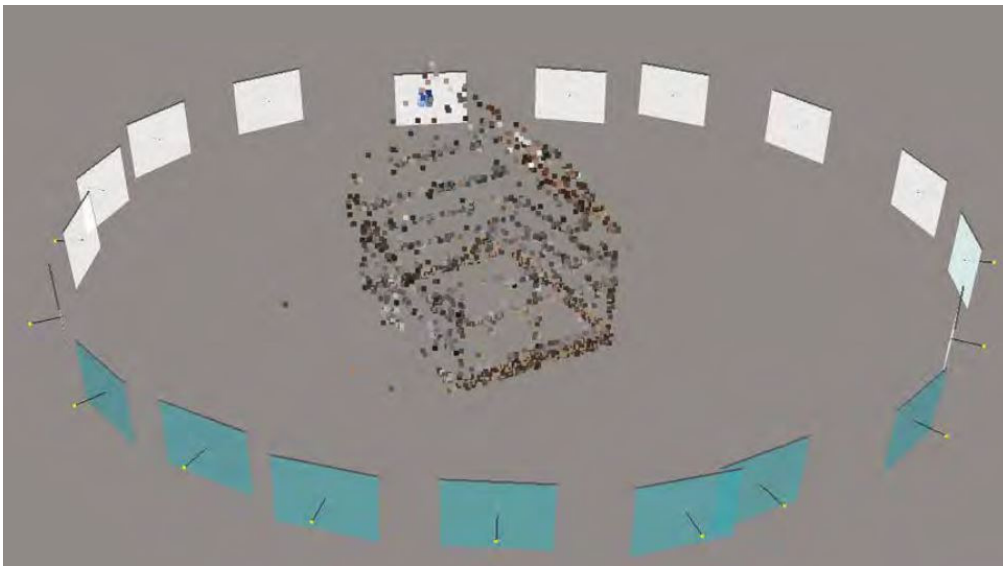


Figure 2.1. Example of 3D point clouds generated from 2D images [15].

Figure 2.2. Point cloud data of the city of Santa Monica.

2.3 Segmentation of Building Points

As shown in Fig.2.3, the original point clouds contained many redundant points (e.g., tree, road, car). To capture the open-space features of soft-story buildings, building points need to be segmented from the city-scale data. For this sake, building footprints are used to outline the boundaries of soft-story and non-soft-story buildings. The city government of Santa Monica provides an open-source geographic information system (GIS) database where footprint polygons are stored as shapefiles (.shp). The extraction process is performed using two Python packages, GeoPandas [69] and Shapely [48]. GeoPandas can access point clouds through *DataFrame* structure in Python and perform a spatial query based on the footprint

polygons read by Shapely. Based on the geocoordinates of point clouds, points within a polygon are segmented and stored for each building.

A parameter is introduced here, namely, the scaling ratio of footprint polygons. Using unscaled polygons sometimes causes flawed points on surfaces where only a few points are generated in photogrammetry. A scaling ratio is hence applied to ensure that the complete appearance of a building can be acquired. However, the larger polygon is used, the more points will be segmented. The effect of the scaling ratio on point density is hence necessary to be investigated for CNN models, which will be discussed later in this chapter.

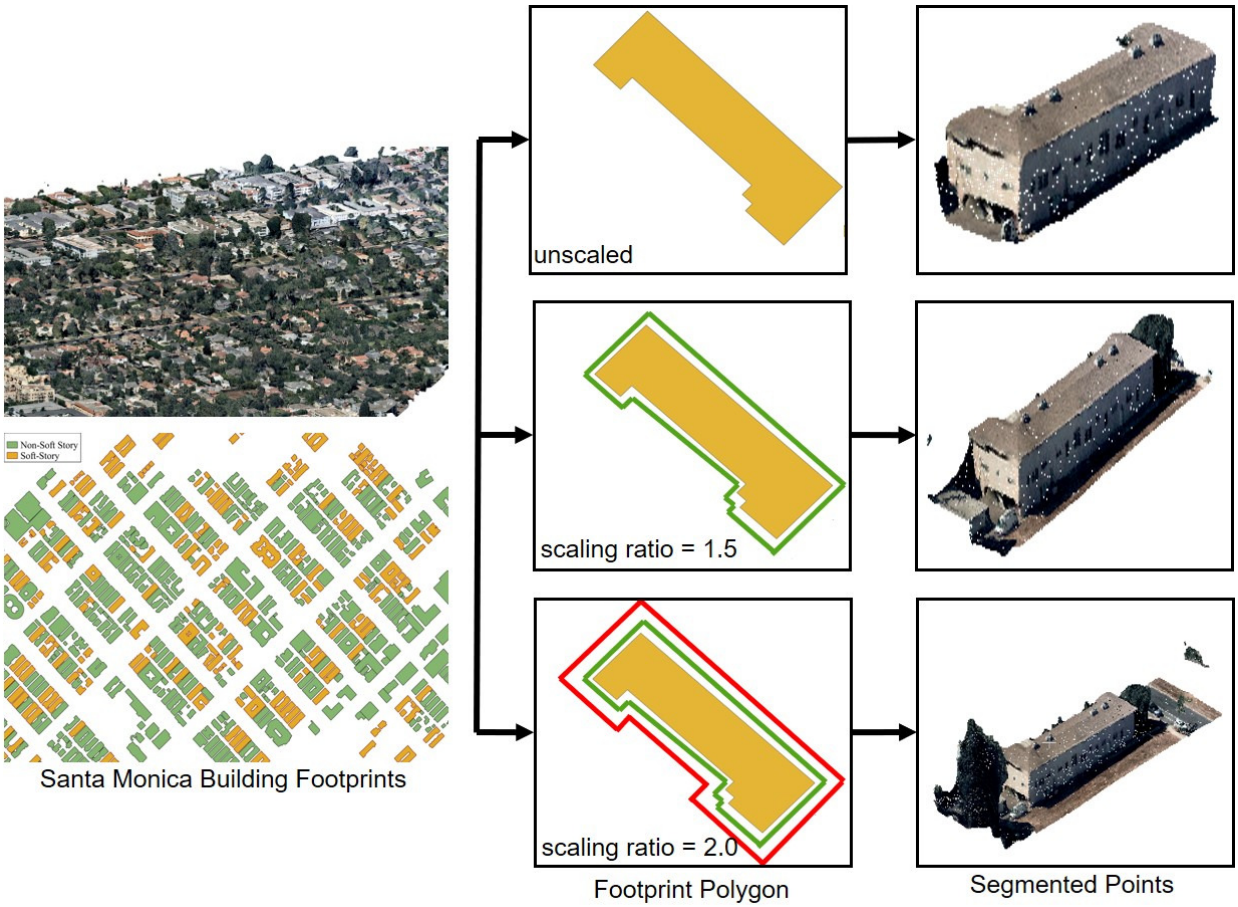


Figure 2.3. Point-cloud data of the city of Santa Monica.

2.4 Extraction of Density Features

As previously mentioned, a soft-story building usually has a ground level with a large opening space, which can be easily detected in 2D imagery data if there are no obstacles in front of it. When the opening is at the side of a building and cannot be captured by the street-view image, the identification becomes unreliable. In contrast, the opening space can always be described by point clouds, because a data point is only generated if there is an object. To use this characteristic, a moving box is introduced into the pipeline to depict the variation in the density of point clouds along with the building height.

As shown in Fig.2.4, a moving box with a fixed height is used to compute the point cloud density along the building height, where the moving step size is set to be the same as the box height. The relation between height and density can then be drawn and further implemented as imagery data for training the CNN-based classifiers. A gap/pit can be observed in the soft-story building's density-height figure, which is suitable for training a binary classifier. To ensure the robustness of the pipeline, the height and density data will be normalized before converted into a 2D image. The effect of box height on soft-story building identification is also evaluated and discussed later in this chapter.

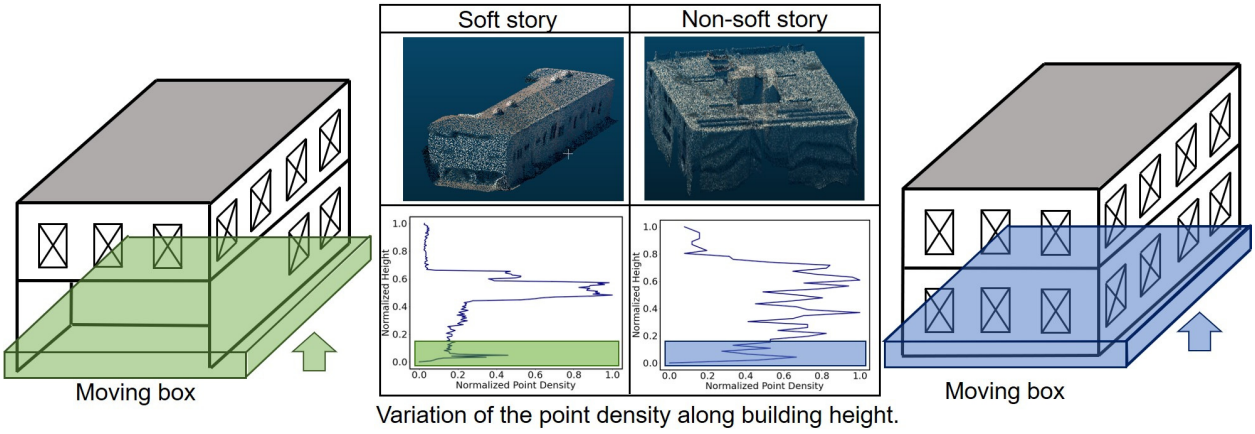


Figure 2.4. Extraction of density features along building height.

2.5 Naïve CNN Model

After converting the density feature into 2D imagery data, the identification of a soft-story building becomes a binary classification task that can be tackled using CNN models. Although many existing models have deep network architectures that can classify multiple objects in an image with a complicated background, this study started with a naïve CNN model because the density image is only composed of lines and is relatively simple compared to practical imagery data.

The schematic architecture of the naïve CNN model used in this study is shown in Fig.2.5. It generally contains three main types of layers: **convolutional layer**, **pooling layer**, and **fully-connected layer**. In more detail:

- Input layer: the raw pixel values of the image. For example, images with width 32, height 32, and with three color channels R,G, and B. The dimension is $[32 \times 32 \times 3]$.
- Convolutional layer: it adopts a set of trainable filters (e.g., 3×3, 5×5) and bias to slide over an image and compute their dot product. It captures the spatial information of each pixel and can be referred to as a feature map. This may result in a volume such as $[32 \times 32 \times 12]$ if 12 filters are used. An example is shown in Fig.2.6
- RELU layer: a nonlinear transformation function, namely, the rectified linear unit. It will apply an element-wise activation function, such as the $max(0, x)$ [94] thresholding at zero. This leaves the size of the volume unchanged ($[32 \times 32 \times 12]$).
- Pooling layer: it will perform a downsampling operation along the spatial dimensions (width, height) and maintain larger patterns in the feature map. For example, each pixel in each feature map is replaced by a maximum of the 2×2 patch around this pixel, which is called max-pooling [93], resulting in volume such as $[16 \times 16 \times 12]$.
- FC (i.e. fully connected) layer: it will compute the class scores, resulting in a volume of size $[1 \times 1 \times 2]$, where each of the 2 numbers corresponds to a class score, such as the 2 categories of buildings. As with ordinary neural networks and as the name implies,

each neuron in this layer will be connected to all numbers in the previous volume.

- Dropout: a regularization method that approximates training a large number of neural networks with different architectures in parallel to avoid over-fitting [114]. When applying dropout, some outputs are randomly ignored or "dropped out," which has the effect of making the layer look like and be treated like a layer with a different number of nodes and connectivity to the prior layer. This can make the training process noisy, so the model will not overfit a dataset with a few examples. A conventional dropout ratio of 0.5 is used for the naïve CNN model in this study.

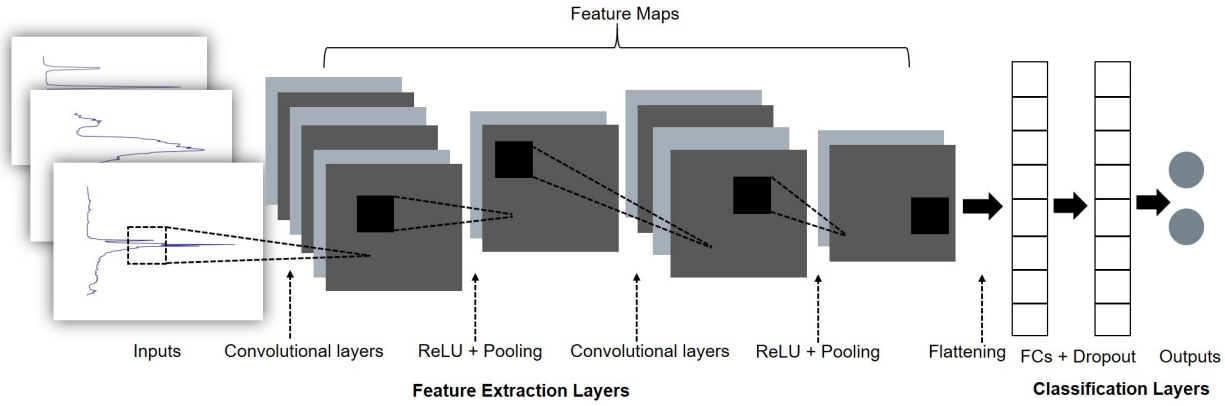


Figure 2.5. Schematic architecture of the naïve CNN model.

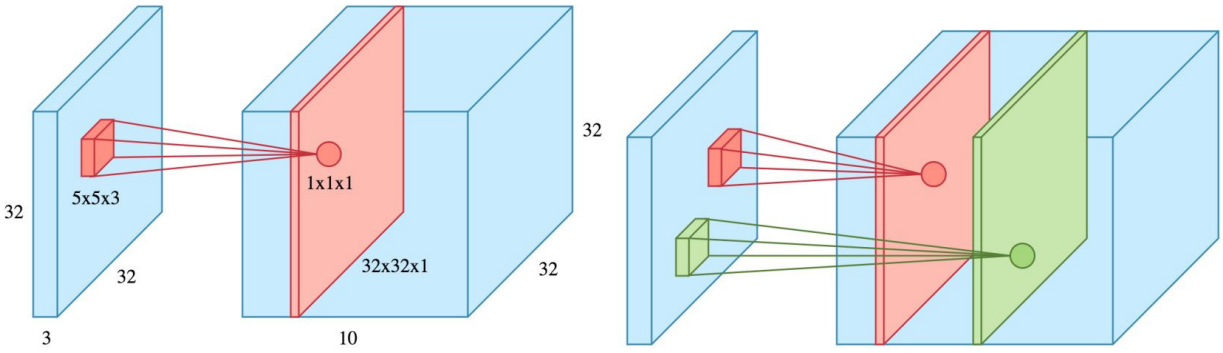


Figure 2.6. An example of convolutional filters.

Briefly speaking, the training process includes forward weight propagation and backward

error propagation. In the forward process, the weights are connected to features in the previous layer and the nonlinearity is introduced through the activation function. In the backpropagation process, the prediction made in the top layer is used to compute the error of the model, and the chain rule can be implemented to identify how much contribution made by each weight in each layer. The goal is to minimize the error or the loss function through updating weights. The network generally has a recursive structure as expressed in Eq.2.1, where $l = 1, \ldots, L$, denotes the layer; $h_0 = X$ is the input vector; W_l and b_l are weight matrix and bias vector respectively; and f_l is activation function at l layer (e.g., ReLU).

$$\begin{aligned} h_l &= f_l(s_l) \\ s_l &= W_l h_{l-1} + b_l \end{aligned} \tag{2.1}$$

To make the prediction of the input X_i into K categories, the output layer is a K dimensional vector $h_i = (h_{ik}, k = 1, \ldots, K)$, and the probability that the input X_i belongs to category k can be computed through $softmax$ probability expressed in Eq. 2.2. The input X_i is classified to category k whose h_{ik} is the maximum in h_i.

$$p_{ik} = \frac{e^{h_{ik}}}{\sum_{k'=1}^{K} e^{h_i k'}} \tag{2.2}$$

Suppose the training dataset is $(X_i, y_i), i = 1, \ldots, n$, where X_i is the input image, and y_i is the output category. Let $p(y|X, \theta)$ be the probability that the input image X belongs to the category y according to the above $softmax$ probability. θ can be estimated by maximizing the log-likelihood function expressed in Eq. 2.3 using gradient descent based on the back-propagation computation.

$$\sum_{i=1}^{n} logp(y_i|X_i, \theta) \tag{2.3}$$

The loss function in the back-propagation is the negative log-likelihood expressed in Eq. 2.4.

$$\mathcal{L} = -\sum_{i=1}^{n} logp(y_i|X_i, \theta) \tag{2.4}$$

If $y_i = k$, then

$$\begin{aligned}\frac{\partial log p(y_i|X_i,\theta)}{\partial h_{ik'}} &= 1(k'=k) - \frac{\partial}{\partial h_{ik'}} log \sum_{k'=1}^{K} e^{h_{ik'}} \\ &= 1(k'=k) - \frac{e^{h_{ik'}}}{\sum_{k'=1}^{K} e^{h_{ik'}}} \\ &= 1(k'=k) - p_{ik'}\end{aligned} \tag{2.5}$$

Let Y_i be the one-hot vector, i.e., if $y_i = k$, then the $k-th$ element of Y_i is 1, and 0 otherwise. Let $p_i = (p_{ik}, k = 1, \ldots, K)$ be the vector of predicted probabilities of the K category. Then Eq. 2.6 can be used to update the weights and biases using the chain rule in the back-propagation.

$$\frac{\partial log p(y_i|X_i,\theta)}{\partial h_{ik'}} = Y_i - p_i = e_i \tag{2.6}$$

To minimize the loss, the gradient descent for Eq. 2.4 is

$$\theta_{t+1} = \theta_t - \eta \mathcal{L}'(\theta_t) \tag{2.7}$$

where η is the step size or learning rate, and $\mathcal{L}'(\theta)$ is the gradient. This method may be time-consuming because the loss is summed over all the examples. A stochastic gradient descent algorithm can be used to address this problem by randomly selecting i from $1, 2, \ldots, n$. In this method, θ only need to be updated through the gradient for the $i-th$ example $\mathcal{L}'_{\rangle}(\theta_t)$.

The gradient descent algorithm goes downhills in the steepest direction in each step, but the steepest direction may not be the best direction. As shown in Fig.2.7, the black arrows are the gradient and the red arrows are the preferred direction, which is the direction of momentum. To accelerate the minimization, the momentum [103] should be considered in the stochastic gradient descent expressed in Eq. 2.8

$$\begin{aligned}v_t &= \gamma v_{t-1} + \eta_t g_t \\ \theta_t &= \theta_{t-1} - v_t\end{aligned} \tag{2.8}$$

where g_t is the average gradient computed from the batch of examples, and v_i is the momentum or velocity. γ is usually set at 0.9 for accumulating the momentum, and θ is updated based on the momentum.

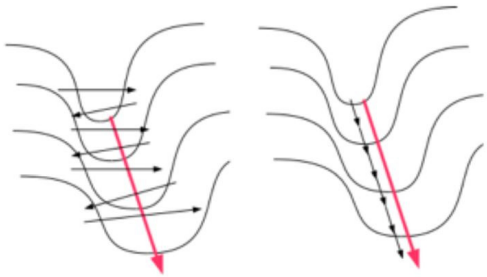


Figure 2.7. Direction of momentum in gradient descent algorithm.

Alternatively, the Adagrad algorithm [39] modifies the gradient descent in another direction by considering that the components of the gradient may be very uneven. The updating of the gradient is then expressed in Eq. 2.9

$$\begin{aligned} G_t &= G_{t-1} + g_t^2 \\ \theta_{t+1} &= \theta_t - \eta_t \frac{g_t}{\sqrt{G_t + \epsilon}} \end{aligned} \tag{2.9}$$

where ϵ is a small number to avoid dividing by 0. g_t^2 and $g_t/\sqrt{G_t + \epsilon}$ denote component-wise squares and division.

In the Adagrad algorithm, G_t is the sum over all time steps. RMSprop [61] improves it by only summing the recent time steps as expressed in Eq. 2.10 where β is often set to be 0.9.

$$\begin{aligned} G_t &= \beta G_{t-1} + (1-\beta) g_t^2 \\ &= (1-\beta)(\beta^{t-1} g_1^2 + \beta^{t-2} g_2^2 + \cdots + \beta g_{t-1}^2 + g_t^2) \end{aligned} \tag{2.10}$$

The state-of-the-art algorithm Adam optimizer [76] further combines RMSprop and momentum and has been widely used for training deep CNN models. The formulations are

expressed in Eq. 2.11.

$$
\begin{aligned}
v_t &= \gamma v_{t-1} + (1-\gamma) g_t \\
G_t &= \beta G_{t-1} + (1-\beta) g_t^2 \\
v_t &\leftarrow \frac{v_t}{1-\gamma} \\
G_t &\leftarrow \frac{G_t}{1-\beta} \\
\theta_{t+1} &= \theta - \eta \frac{v_t}{\sqrt{G_t + \epsilon}}
\end{aligned}
\tag{2.11}
$$

In this study, the above-mentioned layers and optimization algorithms are already developed and available through Python. For the naïve CNN model, a grid search technique is implemented to investigate the optimal learning rate and the algorithm for minimizing the loss. 0.01 learning rate and the stochastic gradient descent are concluded to be optimal and used in the proposed pipeline.

2.6 Investigation of Optimal Parameters

Three experiments are conducted to investigate the optimal parameters for the proposed pipeline: (1) moving-box height; (2) input-image size; and (3) scaling ratio of the building footprints. The implementation of these experiments is conducted on a TensorFlow platform using a self-assembled computer with a single GPU (CPU: Intel(R) Core i7-8700 @ 3.20 GHz, RAN:16.0 GB, and GPU: Nvidia RTX 2080).

It is worth noting that only one set of building points should be segmented from a footprint at a single address, which will only generate one density feature image. The moving box height will only affect the smoothness of the density lines. However, multiple polygons may be found in some addresses (i.e., footprints) where affiliated facilities (e.g., doghouses, swimming pools, storerooms) are also captured in the auto-segmenting process and classified as buildings. These false buildings can then be removed when the point set is shorter than the height of a typical building (i.e., 3 m).

2,842 density feature images are considered as real buildings and are used to conduct numerical experiments. The images are divided into a training set (2,274 $\approx$ 80%), a validation set (285 $\approx$ 10%), and a testing set (283 $\approx$ 10%). Each set has balanced images from two categories (i.e., soft-story and non-soft story). During the training process, the validation accuracy is monitored to track overfitting. The model with the highest validation accuracy is saved and used for testing. If no higher validation accuracy is observed for 10 consecutive epochs, the training process is interrupted (i.e., early stopping). For the input-image size experiment, four size categories are designed: small (64×64); median(128×128); VGG(224×224) implemented in VGGNet [112]; and large(256×256). The detailed configuration of the naïve CNN model is listed in Table 2.1. The number of parameters is related to the input, filter size, and the number of neurons in the fully connected layers.

Table 2.1. Configuration of naïve CNN model

Conv architecture			Input size and output shape			
Block	Layer	Filter size	64 x 64	128 x 128	224 x 224	256 x 256
Input	Input image	-	(N, 64, 64, 1)	(N, 128, 128, 1)	(N, 224, 224, 1)	(N, 256, 256, 1)
Conv block 1	Convolutional	3 x 3 (32)	(N, 62, 62, 32)	(N, 126, 126, 32)	(N, 222, 222, 32)	(N, 254, 254, 32)
	Convolutional	3 x 3 (32)	(N, 60, 60, 32)	(N, 124, 124, 32)	(N, 220, 220, 32)	(N, 252, 252, 32)
	Max pooling	-	(N, 30, 30, 32)	(N, 62, 62, 32)	(N, 110, 110, 32)	(N, 126, 126, 32)
Conv block 2	Convolutional	3 x 3 (64)	(N, 28, 28, 64)	(N, 60, 60, 64)	(N, 108, 108, 64)	(N, 124, 124, 64)
	Convolutional	3 x 3 (64)	(N, 26, 26, 64)	(N, 58, 58, 64)	(N, 106, 106, 64)	(N, 122, 122, 64)
	Max pooling	-	(N, 13, 13, 64)	(N, 29, 29, 64)	(N, 53, 53, 64)	(N, 61, 61, 64)
Conv block 3	Convolutional	3 x 3 (128)	(N, 11, 11, 128)	(N, 27, 27, 128)	(N, 51, 51, 128)	(N, 59, 59, 128)
	Convolutional	3 x 3 (128)	(N, 9, 9, 128)	(N, 25, 25, 128)	(N, 49, 49, 128)	(N, 57, 57, 128)
	Max pooling	-	(N, 4, 4, 128)	(N, 12, 12, 128)	(N, 24, 24, 128)	(N, 28, 28, 128)
Fully connected	Flatten	-	(N, 2,048)	(N, 18,432)	(N, 73,728)	(N, 100,352)
	Dense	-	(N, 1,024)	(N, 1,024)	(N, 1,024)	(N, 1,024)
	Dense	-	(N, 1,024)	(N, 1,024)	(N, 1,024)	(N, 1,024)
	Dropout	-	(N, 1,024)	(N, 1,024)	(N, 1,024)	(N, 1,024)
	Dense	-	(N, 2)	(N, 2)	(N, 2)	(N, 2)

2.7 Experimental Results

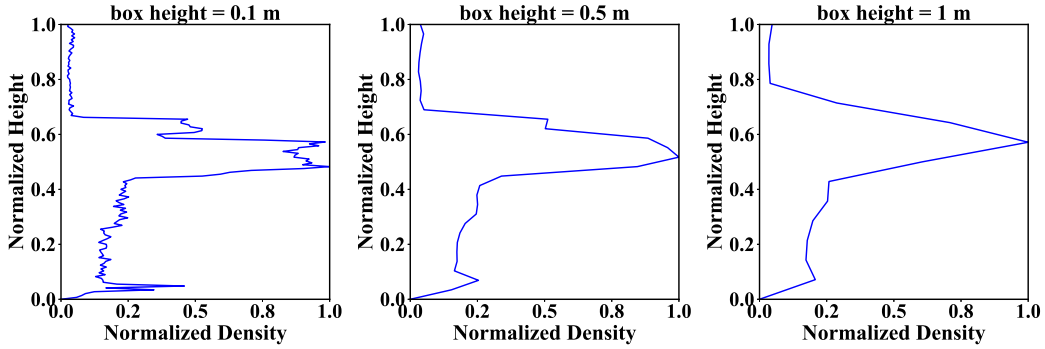
Parameter optimization is evaluated based on three performance metrics: *Recall*, *Precision*, and $F1$ *score* defined in Eqs. 2.12 to 2.14, respectively. Soft-story and non-soft-story buildings are referred to as positive and negative instances, respectively. TP, TN, FP, and FN are referred to as true positive, true negative, false positive, and false negative, respectively. *Recall* represents how many actual positives are captured by the model, whereas *Precision* refers to how precise or accurate the model is out of those predicted positives. $F1$ *score* is the harmonic mean of *Recall* and *Precision*.

$$Recall = \frac{TP}{TP + FN} \tag{2.12}$$

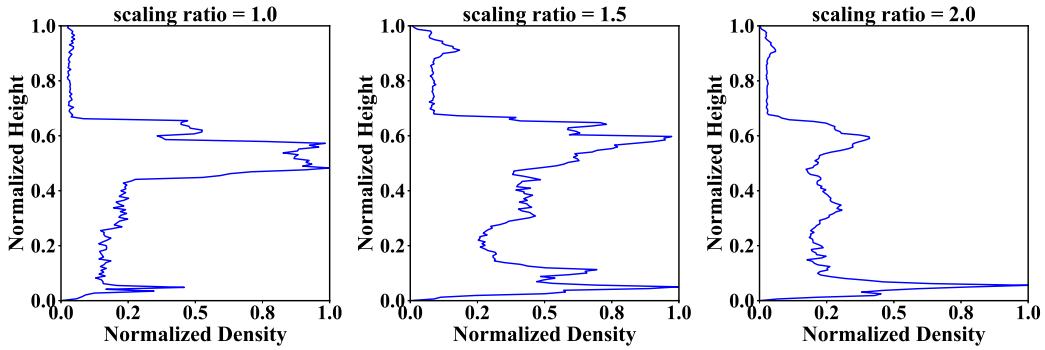
$$Precision = \frac{TP}{TP + FP} \tag{2.13}$$

$$F1\ score = \frac{2 * Recall * Precision}{Recall + Precision} \tag{2.14}$$

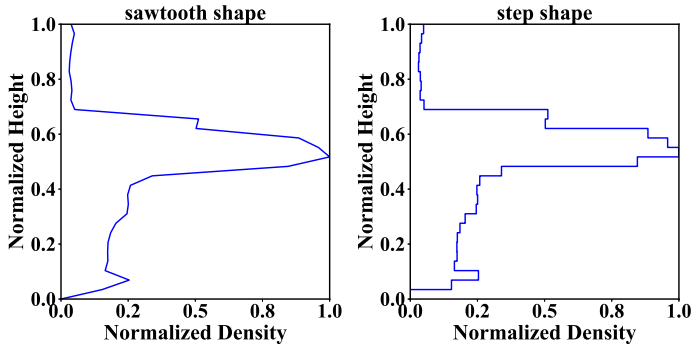
Fig.2.8 shows the density images computed using different parameters. The performance metrics of each experiment are shown in Fig.2.9. As shown in Fig.2.8 (a), the use of a higher moving box makes the depiction of the opening space in a lower level more difficult to identify. Hence, 0.1 m is the optimal moving-box height, dominating the performance metrics in Fig.2.9 (a). The median and large image sizes both lead to optimal results. Considering the computational cost, the median size (128×128) is recommended. In Fig.2.8 (b), more point clouds that do not belong to the building are included in the lower floors when 1.5 and 2.0 scaling ratios are used for the building footprints, which makes it difficult to identify the opening space of soft-story buildings. The unscaled building footprint is thus optimal, as shown in Fig.2.9 (b). Another attempt related to the density image (Fig.2.8 (c)) is also conducted, where density images depicted by either sawtooth or step shapes are used for training DL models. However, little improvement is obtained with this experiment.



(a) Density images computed using unscaled footprints and different moving-box heights.

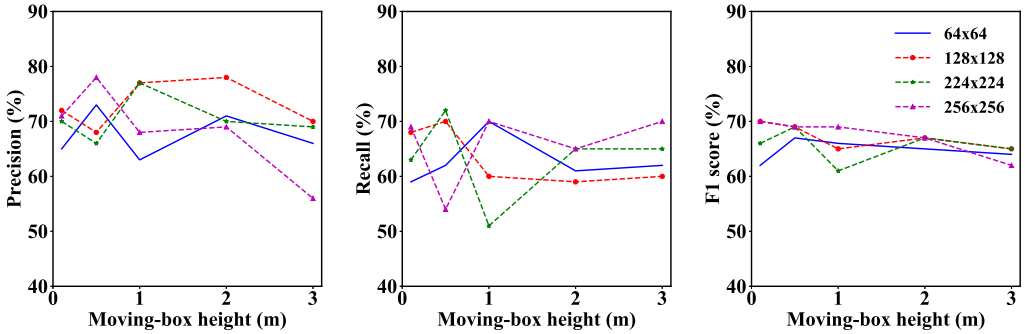


(b) Density images computed using scaled footprints and 0.1 m moving box.

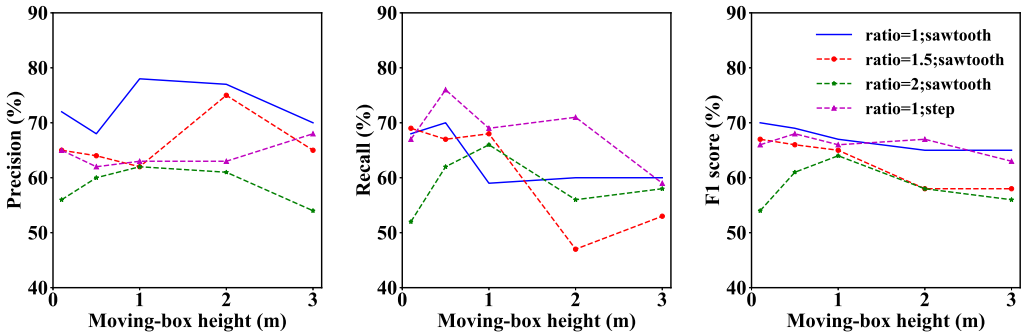


(c) Density images with different shapes.

Figure 2.8. Density images computed using different parameters.



(a) *Precsion*, *Recall*, and $F1score$ for different image sizes and box heights.



(b) *Precsion*, *Recall*, and $F1score$ for different scaling ratios and image shapes.

Figure 2.9. Performance comparison of experiments.

2.8 Transfer Learning

Although the optimal parameters are determined for the proposed pipeline, there remains room for improvement. State-of-the-art CNN architectures are hence implemented to assess the possibility of better performance. However, the dataset used in this study is too small to train a large number of parameters in these deep CNN models. As a result, TL is conducted. As mentioned previously, TL uses the experience and knowledge learned from the source domain and its learning task to perform well in the target domain. The transferability of TL has been verified; the initialization of transferred features can boost the performance relative to random features, even if there is a distance between the domain task and the

target task [131]. In this study, three deep CNN models are selected to implement TL on the target domain, namely, VGGNNet, Inception (a.k.a. GoogLeNet), and ResNet. The source domain of these models was trained on more than 1.2 million images and 1,000 labeled categories, as part of the ImageNet Large Scale Visual Recognition Challenge or ImageNet Challenge [37].

VGGNet was the winner of ImageNet in 2014. There are two versions, VGG16 and VGG19, which consist of 16 and 19 hidden layers, respectively. The numbers of convolutional blocks and fully connected layers are implemented within the hidden layers. The VGGNet filters are all 3×3, and approximately 138 million parameters are trained in VGGNet. Inception net was named after the movie "Inception", which includes the line "we need to go deeper." While Inception has 27 layers, it only contains 6.4 million trainable parameters due to its extensive use of 1×1 filters, which limits the number of input channels. Inception dominates the classification task in ImageNet after VGGNet. ResNet was developed to address the vanishing gradient problem in back-propagation, which is triggered when a deep CNN architecture is used. With the introduction of residual blocks with implemented identification functions, it is easier to find the derivative in the gradient descent method. However, this again increases the computational cost. For example, 60.3 million parameters are trainable in ResNet with 152 residual blocks. The architectures of these models are shown in Fig.2.10.

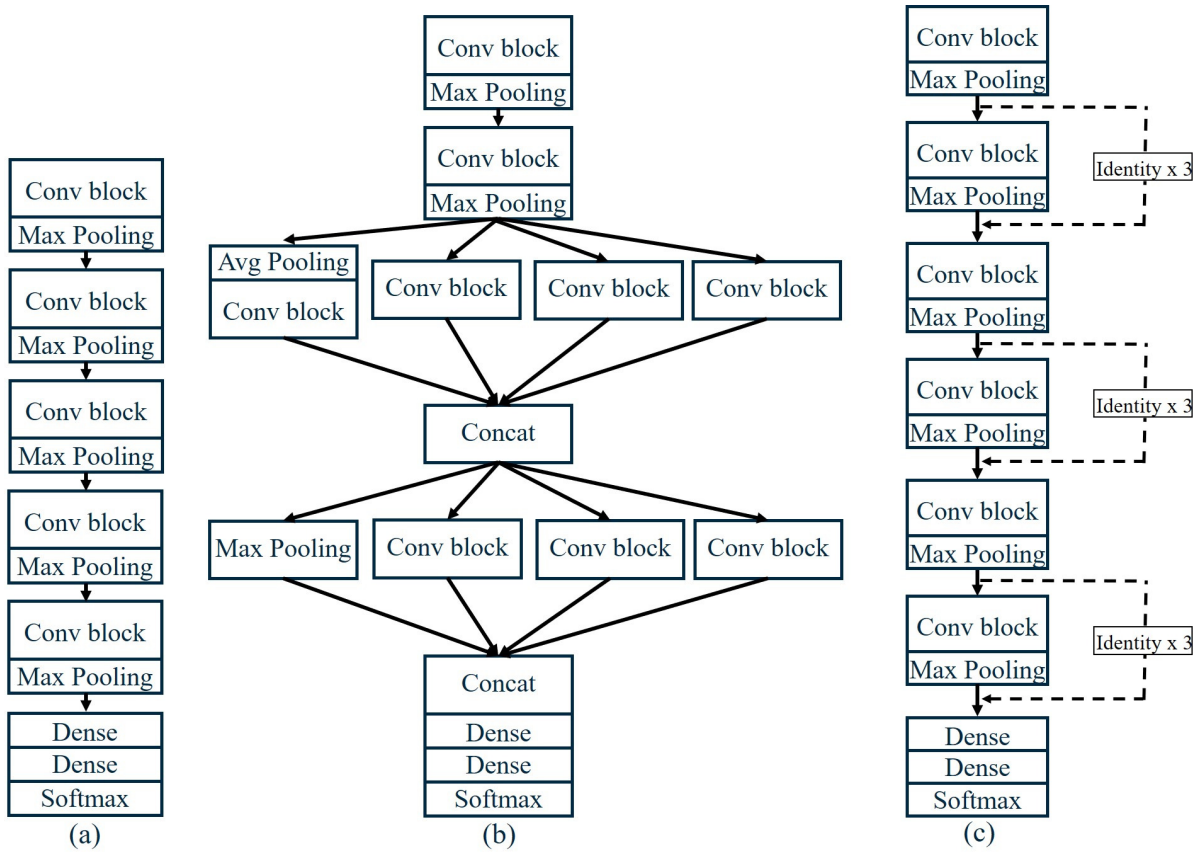


Figure 2.10. Deep CNN architectures: (a) VGG Net; (b) Inception Net; and (c) ResNet.

Feature extraction and fine-tuning are two widely used strategies when conducting TL. In feature extraction, parameters trained from the source domain are used in the forward procedure to capture features from the new domain and are not updated during back-propagation. Only the fully connected layers prior to the output are trained for the new task, which greatly reduces the training time and number of epochs. Fine-tuning also fixes some portions of the convolutional blocks where low-level features are captured and retrains the remaining convolutional blocks to adjust the mid-level and high-level features to the target domain. Compared with the general training process, fine-tuning is still less time-consuming. Both feature extraction and fine-tuning are conducted in this study for VGGNet, Inception net, and ResNet to investigate the optimal CNN model for the proposed pipeline.

Median size images computed through unscaled building footprints and moving boxes with a height of 0.1 m are identified to be optimal in the previous section and are used to evaluate the deep CNN models herein. The TL learning curves are shown in Fig.2.11, including the best naïve CNN model. Significant overfitting is observed in all learning curves except for the VGGNet result with implemented feature extraction. The performance on the test set for these models is listed in Table 2.2. VGGNet dominates in terms of *Precision*, *Recall*, and $F1$ *score*. Although the naïve CNN model shows similar performance to the VGGNet, the implementation of TL reduces the overfitting due to the pre-trained knowledge drawn from the ImageNet dataset. Table 2.3 summarizes the computational efficiency of each model. Roughly speaking, the training time increases when more parameters (i.e., deeper architecture) are included. However, VGGNet and Inception Net trained using fine-tuning techniques showed shorter training times due to triggering the early stopping mechanism. It is worth noting that the training and testing can be done in a few minutes due to the computational power of the GPU, which enhances the capability of the proposed pipeline for rapid identification of soft-story buildings at a city scale.

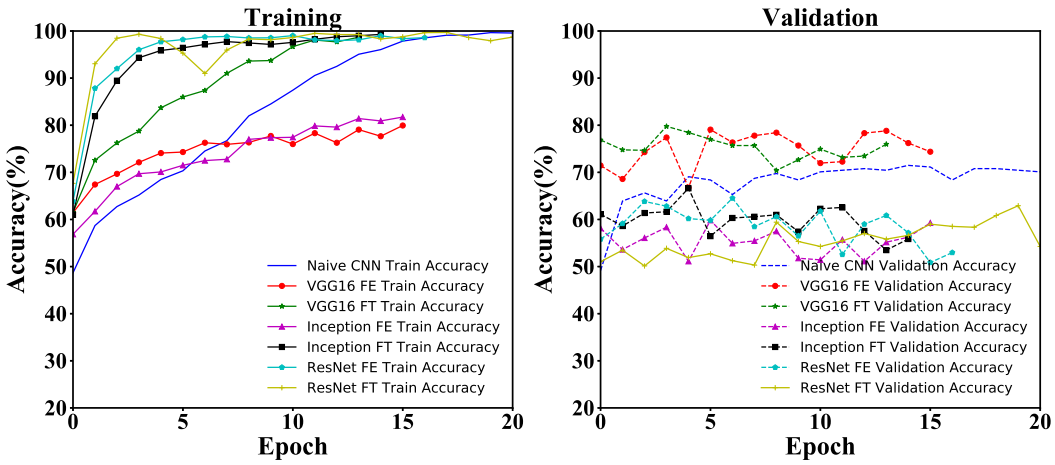


Figure 2.11. Performance comparison of CNN models.

Table 2.2. Performance of TL on the test set.

Performance Metrics	Naive CNN	VGG16 FE	VGG16 FT	Inception FE	Inception FT	ResNet FE	ResNet FT
Precision	0.67	0.76	0.74	0.5	0.61	0.58	0.64
Recall	0.73	0.66	0.66	0.4	0.36	0.4	0.38
F1 score	0.7	0.71	0.7	0.44	0.45	0.47	0.47

Table 2.3. Comparison of computational efficiency for different models.

Computational Efficiency (sec)	Naive CNN	VGG16 FE	VGG16 FT	Inception FE	Inception FT	ResNet FE	ResNet FT
Training	113.2	132.6	93.7	548.9	212.2	161.6	290.5
Testing	2.1	2.2	2.5	20.6	17	13.2	12.9

2.9 Summary

In this chapter, a pipeline for identifying soft-story buildings at a city scale was proposed. It includes (1) automated extraction of building point clouds from raw points generated by photogrammetry, (2) computation of the relation between point cloud density and building height through a moving box, and (3) training of DL models to identify soft-story buildings based on the density-height imagery data. Santa Monica was selected as the target city. Multiple experiments were conducted to investigate the optimal parameters for the naïve CNN model. The median size images (128 $\times$ 128) computed through unscaled building footprints and a moving box with a height of 0.1 m yield the highest *Recall*, *Precision*, and *F1 score* in the test set. Transfer learning was introduced into the proposed pipeline to improve the soft-story identification performance, and VGGNet with the implementation of feature extraction dominated the performance of the test set and reduced overfitting.

To apply the proposed pipeline to a different city, the building footprints are needed to segment the building point clouds. Building footprints are generally available in the GIS

database provided by city governments or can be downloaded via OpenStreetMap (`https://www.openstreetmap.org`), a collaborative project to create a free editable map of the world. In this study, ground-truth labeling can be used to evaluate the performance of the proposed pipeline, but a complete list of soft-story buildings may not be available for other cities. Hence, the prediction made by the proposed method for different cities may still require further validation.

As discussed earlier, the open space is a major hint for the identification of soft-story buildings due to its ease of observability. Hence, many studies, including this one, have focused on this characteristic of soft-story buildings. Point cloud data can accurately capture this open-space characteristic, thus overcoming the limitations of street-view image-based methods. However, soft-story buildings with structural irregularities may not always have visible open spaces. In addition, buildings with open spaces retrofitted with reinforcing structures (e.g., a garage of a single-story residential house) should not be identified as soft-story buildings. These issues are opportunities for further improvement of the proposed pipeline. Accordingly, CNN architectures trained on 3D point clouds can be investigated in the future as a means of obtaining detailed information on structural components. For example, PointNet [102] and PointCNN [82] have been proposed for the classification and segmentation of outdoor and indoor objects (e.g., buildings, cars, chairs, and desks) based on point-cloud data, and the features learned in those models can be transferred and used to assist soft-story detection on a large scale.

CHAPTER 3

Pipeline for Seismic Assessment of NDRCB Frames

This chapter is partly based on the following publication:

Chen, P.Y., Lesgidis, N. Y., Cetiner, B., Sextos, A., and Taciroglu, E. (2021). "A method for Automated Development of Model & Fragility Inventories of Non-ductile Reinforced Concrete Frame Buildings," *Earthquake Spectra* (under 2$^{\text{nd}}$-round review).

3.1 Overview

In this chapter, a pipeline for automated modeling and fragility computation for NDRCB archetypal frames is introduced. It starts with a design procedure that uses basic building information (i.e., number of floors, number of spans, floor area, floor height, span length) and era-specific codes to construct analytical models. Nonlinear elements used in the archetypal frame are discussed herein. Several well-studied NDRCB frames are utilized to validate the design process through nonlinear static and nonlinear dynamic analyses.

To compute fragility functions for discrete damage states, a set of ground motions including far-field and near-field is implemented in the pipeline to conduct the probabilistic seismic demand model (PSDM [31]). An inventory of fragility curves for archetypal frames is carried out and compared to HAZUS [43]. Suitable intensity measurements (IMs) are investigated for archetypal frames through multiple metrics including efficiency, sufficiency, proficiency,

and practicality. Given that the pipeline is approximating as-built buildings through minimum building information, the sensitivity of the pipeline to the building information is also identified herein through extensive nonlinear time-history analyses (NTHAs).

3.2 Modeling of NDRCB Frame

The vulnerabilities of NDRCB stocks were evident in past earthquakes. The poor performance of this building type is primarily due to the lack of confinement of the concrete cores of beams, columns, joints, and walls, which causes brittle—as opposed to the desirable ductile—behavior in structural components and even collapse. Specifically, shear failure or lap splice failure in columns, joint shear failure or bond failure due to discontinuous beam bottom reinforcement in joints, and premature column failure owing to the design concept of the weak column-strong beam. Fig.3.1 shows types of possible brittle failures in the non-ductile frames.

The shortage of seismic provisions in the early times (pre-1980s) of the United States is generally the main reason behind these deficient designs. Table 3.1 summarizes the evolution of reinforced concrete column design in the U.S. Every revision was mainly in response to experiences during major earthquakes in California such as the 1971 San Fernando, 1989 Loma Prieta, and 1994 Northridge earthquakes.

To reduce the (epistemic) uncertainties in seismic risk assessment, an era-representative code should be taken into account to generate analytical models and the models should be able to capture the above-mentioned critical behaviors. Furthermore, the models should be scalable and hence open-source softwares should be used, for example, OpenSees [91].

(a) **(b)** **(c)**

Figure 3.1. Failure of NDRCB including (a) axial failure at lap splice, (b) shear failure in a column, and (c) sideway collapse.

Table 3.1. Column dimension and detailing requirements in US building codes.

	1967 UBC [64]	ACI 318-71 [28]	ACI 318-02 [29]
Minimum width (b_c)	254 mm	NA	254 mm or $0.4d_c$
Minimum depth (d_c)	305 mm	NA	305 mm
ρ (longitudinal)	$0.01 \leq \rho \leq 0.08$	$0.01 \leq \rho \leq 0.06$	$0.01 \leq \rho \leq 0.06$
Spacing in the middle	smallest of $16\phi_b$,$48\phi_{b-tie}$, and $(d_b, d_c)_{min}$	smallest of $16\phi_b$,$48\phi_{b-tie}$, and $(d_b, d_c)_{min}$	smallest of $6\phi_b$, and 152 mm
Spacing in the end	NA	smallest of $16\phi_b$,$48\phi_{b-tie}$, and $(d_b, d_c)_{min}$	smallest of $b_c/2$, and 102 mm
Hooks	90°	135°	135°
Failure mode	brittle	brittle & ductile	ductile

Many laboratory experiments have been conducted to understand the failure mechanism of the NDRCB frame during earthquake events. From the component-level [16, 56, 60, 128] to full-scale tests [40, 70, 125]. These data have been used to qualitatively and quantitatively describe the behavior of non-ductile structural components/systems including shear failure and loss of lateral and axial load carrying capacity. As an initial development, these undesirable seismic behaviors are only considered in the columns of NDRCB frames in this study

for simplification.

3.2.1 Modeling shear failure in column

The most critical behavior of the NDRCB frame is that shear-dominated columns fail in shear and exhibit degradation of stiffness and strength until axial collapse occurs. It is important that the analytical models can capture this behavior and are computationally efficient and accurate. To this end, several existing shear failure models are discussed here to identify the most appropriate one for the pipeline.

Early studies [95, 113, 133] considered shear failure through modified flexural elements through postprocessing. While this method can detect shear failure, it cannot estimate shear deformation and strength degrading. To address the deficiency, some studies [96, 111, 130] developed continuum finite elements to capture the strength degrading after shear failure. Although it can address the shear deformations, the computational cost is expensive.

An alternative technique that is easy to implement is adding a shear spring in series with flexural elements [81, 100, 108]. These studies used the modified compression field theory (MCFT [122]) to define the backbone curve of the shear spring which is usually simulated through a zero-length element. While these studies can capture the strength degradation, the MCFT can only predict the backbone up to the point of maximum shear strength. It is not precise for the post-peak behavior.

A solution for this shortage is the shear strength model using deformations to estimate the column shear strength [10, 101, 109]. However, these methods cannot predict a reliable drift capacity at shear failure which is an important contributor for displacement-based design for existing buildings. To overcome this, Elwood and Moehle [41] proposed an empirical model through a database of 50 flexure-shear-critical RC columns with configurations representative of NDRCB frames. This drift capacity model has been implemented into OpenSees [91] with the hysteretic uniaxial material, namely, the limit state material [40, 71] which can identify

a shear failure associated with column shear and column's total deformation. This model was further developed to consider the bar-slip, flexure-axial-shear interaction, and the loss of axial capacity [12, 14]. Elwood's model is selected in this study because of its accuracy, computational efficiency, and accessibility. The detailed computation is described below.

Fig.3.2 shows the mechanism of the flexure-shear-critical column using Elwood's limit state material. Once the column's total response exceeds a predetermined shear failure surface, namely, the limit shear curve, its shear response follows the constitutive law of the limit shear curve to include pinching and strength and stiffness degradation. The limit shear curve consists of the column shear capacity V_u, the degrading slope K_{deg}, and residual shear strength V_{res}. After a shear failure is detected, the degrading slope of the total response K^t_{deg} can be computed by Eq.3.1:

$$K^t_{deg} = \frac{V_u}{\Delta_a - \Delta_s} \tag{3.1}$$

where Δ_a is the displacement calculated under the axial load at the time of shear failure, and Δ_s is the displacement at the shear failure. If K^t_{deg} is determined, the degrading flexibility of the shear spring $1/K_{deg}$ can be calculated in Eq.3.2:

$$K_{deg} = (\frac{1}{K^t_{deg}} - \frac{1}{K_{unload}})^{-1} \tag{3.2}$$

where K_{unload} is the unloading stiffness of the beam-column element (i.e., $12EI_{eff}/L^3$).

Additionally, Δ_a and Δ_s can be obtained from Eq.3.2 and Eq.3.3:

$$\frac{\Delta_s}{L} = \frac{3}{100} + 4\rho_{sh} - \frac{v}{500\sqrt{f'_c}} - \frac{P}{A_g f'_c} \geq 0.01 \tag{3.3}$$

$$\frac{\Delta_a}{L} = \frac{4}{100} + \frac{1 + tan^2(\theta)}{tan(\theta) + P(s/A_v f_{yt} d_{cc} tan(\theta))} \tag{3.4}$$

where L is the length of the column, d_{cc} is the depth of column core from centerline to centerline of transverse reinforcement, s the spacing of transverse reinforcement, A_v is the area of transverse reinforcement, f_{yt} is the yield strength of transverse reinforcement, P is the axial load on the column, and θ the critical crack angle from horizontal (assumed to

be 65°). A very similar way to simulate the axial failure in Elwood [41], where the axial degrading stiffness is suggested in Eq.3.5:

$$K_{deg,a} = \frac{1}{100} K_{el,a} \tag{3.5}$$

where $K_{el,a}$ is the elastic axial stiffness of the column. In the proposed OpenSees model, shear and axial limit material are assigned in the shear and axial direction with the zero-length elements at both ends of the column coupled with the bilinear material as shown in Fig.3.3

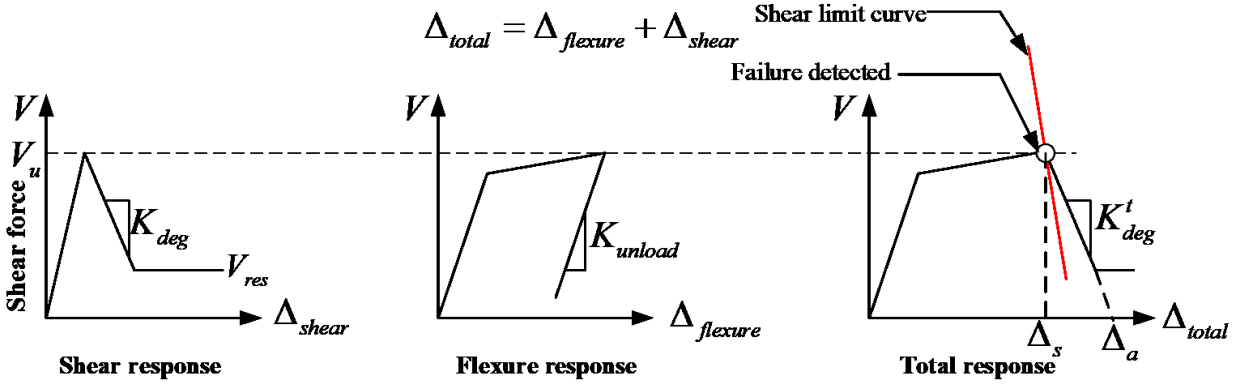


Figure 3.2. Schematic of shear response in column.

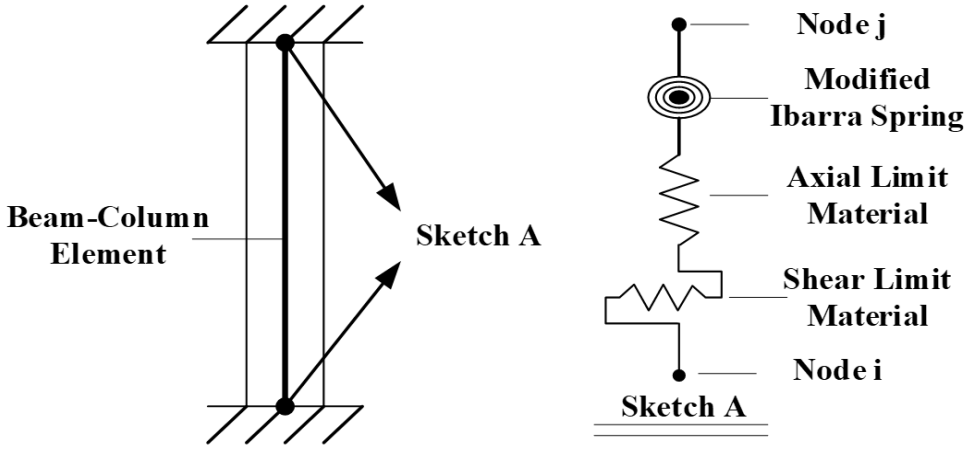
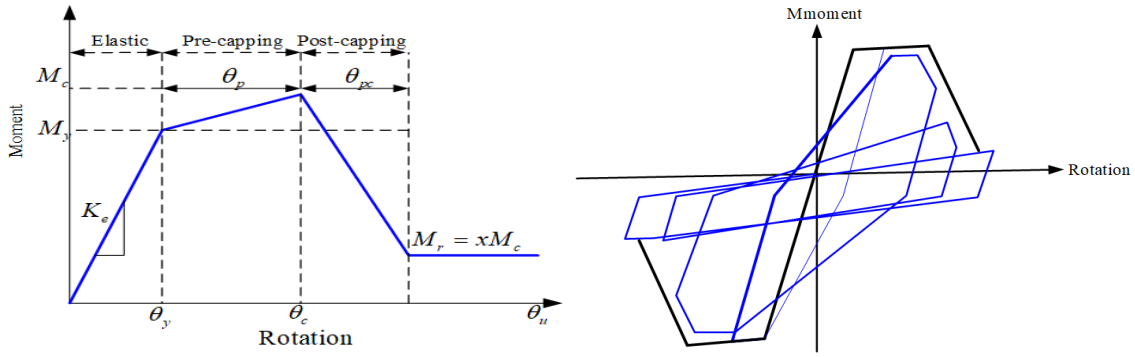


Figure 3.3. Modeling of flexure-shear-critical column.

3.2.2 The concentrated plastic hinge element

In addition to the shear element, the nonlinear plastic hinge element is also important for the collapse modeling of RC frame structures. In the archetypal model, the concentrated plasticity element is used to simulate plastic hinges in beam-column elements, which consist of a linear elastic portion with inelastic hinges at both ends that are typically represented as zero-length rotational springs. The inelastic response is often simulated through the modified Ibarra-Medina-Krawinkler model with a peak-oriented hysteretic response (IMK) [63, 87] which has been implemented in OpenSees.

Fig.3.4 shows the monotonic response of the IMK model. Six parameters are required to control the monotonic backbone curve, namely, the elastic stiffness K_e, yielding moment M_y, capping to yielding moment ratio M_c/M_y, plastic rotation in the post-yielding region θ_p, post-capping rotation θ_{pc}, and ultimate rotation θ_u. Three types of cyclic deterioration, including basic strength, post-capping strength, and unloading stiffness, are incorporated by defining λ and c describing the cyclic deterioration and the rate of deterioration, respectively. A residual strength, as a function of the ultimate strength, can also be assigned for the modified IMK. In the proposed model, 1% residual strength (i.e., M_r) is used for all beams and columns to avoid problems with numerical convergence.



(a) **(b)**

Figure 3.4. Modifed IMK material model: (a) monotonic backbone curve and (b) cyclic response.

While the modified IMK requires a total of 24 parameters in OpenSees, Haselton et al. [55] provided empirical equations and qualitative insights on how to determine each parameter through the calibration of 255 RC columns. The effective stiffness can be defined through Eq.3.6:

$$\frac{EI_y}{EI_g} = -0.07 + 0.59\frac{P}{Agf'_c} + 0.07\frac{L_s}{H} \tag{3.6}$$

where $0.2 \leq \frac{EI_y}{EI_g} \leq 0.6$, $v = P/Agf'_c$ is the axial load ration, and L_s/H is the shear-span ratio. The yielding moment M_y can be calculated through equations developed by Panagiotakos [98] or the conventional Whitney stress block approach. The plastic rotation θ_p can be predicted by Eq.3.7:

$$\theta_p = 0.13(1 + 0.55a_{sl})0.13^{v}(0.02 + 40\rho_{sh}0.57^{0.01f'_c}) \tag{3.7}$$

where a_{sl} is the bond-slip indicator variable, and ρ_{sh} is the lateral confinement ratio. The post-capping rotation θ_{pc} is computed through Eq.3.8:

$$\theta_{pc} = 0.76(0.031)^{v}(0.02 + 40\rho_{sh})^{1.02} \leq 0.1 \tag{3.8}$$

The capping to yield moment ratio M_c/M_y can be calculated using Eq.3.9:

$$\frac{M_c}{M_y} = 1.25(0.89)^{v}(0.91^{0.01f'_c}) \tag{3.9}$$

The cyclic energy dissipation capacity λ is defined through Eq.3.10:

$$\lambda = 170.7(0.27)^{v}(0.1)^{s/d} \tag{3.10}$$

where s/d is the ratio of lateral reinforcement spacing to the effective depth. The rate of deterioration c is generally set to be 1.0.

3.2.3 Archetypal model

Ideally, a large number of building cases should be developed to systematically capture variations in key structural characteristics (e.g., number of floors, number of spans, span lengths, and floor heights) for assessing the city-scale performance of NDRCB frames. Because the

objective of the pipeline is to approximate as-built buildings and address labor-intensive modeling processing, the concept of archetypal buildings is implemented to characterize NDRCB frames. Archetypal structures are a group of structures with similar geometric configuration and/or structural properties, which can represent the variety in design and performance possible among structures.

Archetypal structures have been actively used in recent years to investigate the seismic performance of building structures. For example, Applied Technology Council (ATC) used archetypal models to assess seismic design provisions for buildings' lateral force-resisting systems [36]. Similarly, FEMA P695 [32] also provides a methodology for quantifying the collapse vulnerability of different later force-resisting systems. For NDRCB frames, Liel et al. [83–86] utilized era-representative codes to design archetypal special moment frames and assess their collapse risk. Their investigation of the retrofitting policy as well as the corresponding costs and benefits greatly promoted the current ordinance in major cities of California.

Analytical archetypes proposed by Liel et al. can capture the global response of non-ductile columns through lumped plastic elements and IMK model, but cannot account for the shear behavior and the loss of axial force in columns. To address this, Jeon [66, 67] implemented the shear spring discussed in the previous section in series with flexural elements, which can simulate the stiffness and strength degradation of shear-dominated columns.

The archetypal model used in this pipeline is shown in Fig.3.5. Elastic beam-column elements are preferred over fiber-based beam-column elements in the proposed archetypal models to reduce the computational cost. For the same reason, the brittle failure mode considered here only captures the shear and axial failure in the columns but does not capture the joint shear failure as mentioned early. 5% Rayleigh damping in the first and the third mode is implemented to simulate the viscous damping of the structural system.

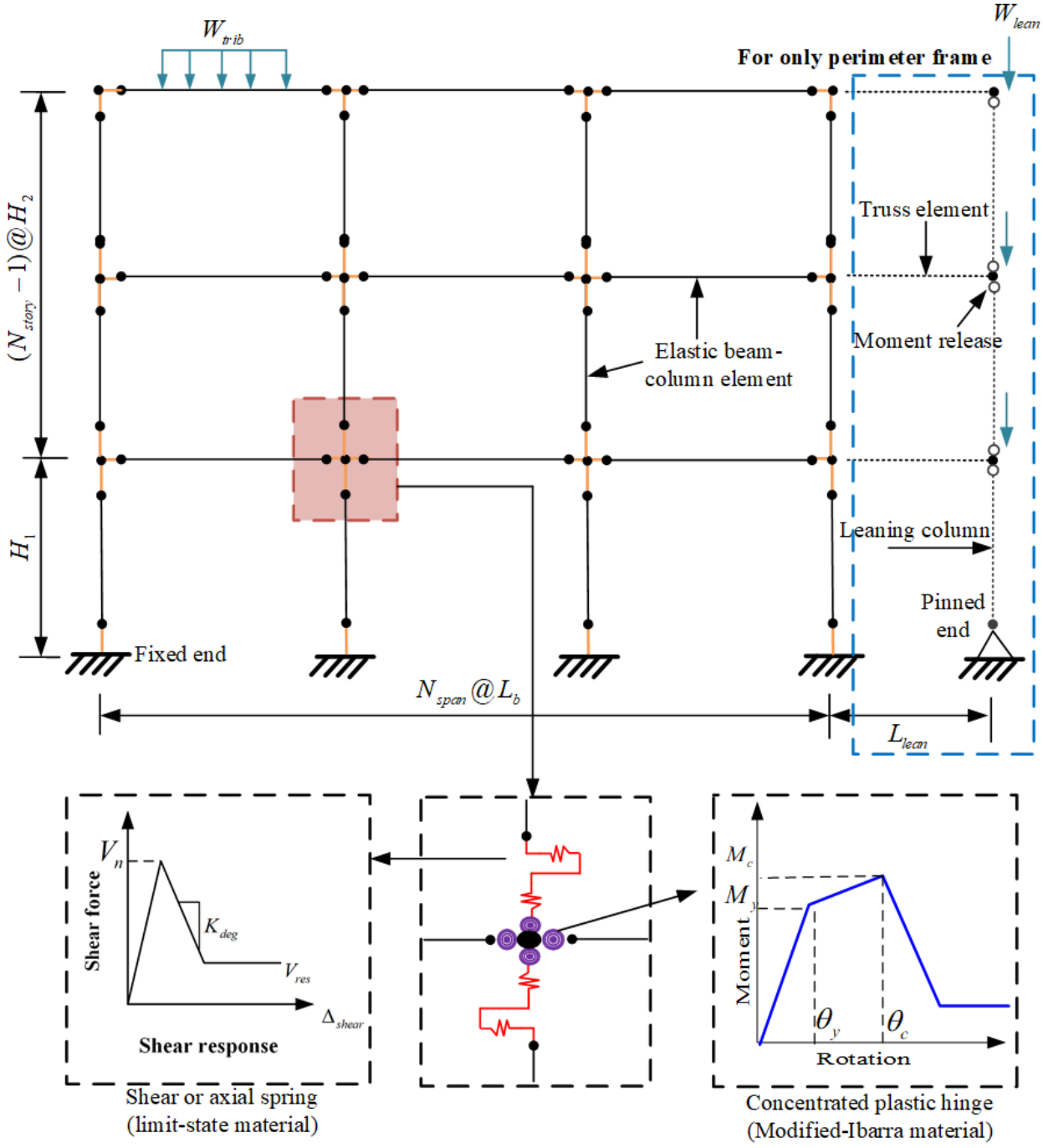


Figure 3.5. Proposed OpenSees model.

3.3 Era-representative design procedure

As discussed in previous sections, the existing modeling techniques can provide reliable predictions of the seismic response of NDRCB frames. However, many technical and implementation details render the practical application of these models difficult at a city scale. Specifically, the detailing (i.e., dimensions of components, reinforcement ratio, rebar spacing) of structural components is inevitable to decide nonlinear elements. This is a complex task requiring not only expertise in reinforced concrete behavior but also in nonlinear analysis. Scaling up the nominally detailed and labor-intensive level of modeling to a regional inventory is, thus, a challenging feat. This can be achieved by implementing a "design" procedure that requires only the basic configurations of a given NDRCB frame (e.g., the number of floors, square footage, etc.) to produce an analytical model that can approximate the as-built structure.

In the design procedure, the building code is a primary means of governing the earthquake response of the structure. NDRCB frames that exist in California were likely designed based on similar requirements in the 1967 Uniform Building Code (UBC) [64] because the California Building Code was formerly an amended version of UBC [83]. Although UBC introduced changes in the requirement for ties and stirrups, the ductile detailing was not mandatory until 1973. Besides, the earthquake design load was not significantly increased until 1976 [25]. The prior inventory [27] showed that more than 70% of NDRCBs were built before 1970, so the UBC is hence thought to be representative for the target region.

Two typical frame systems are provided in the proposed method—namely, space and perimeter frames, which are shown in Fig.3.6 [66, 83, 85]. As aforementioned, the archetypal frame is assumed to be plane symmetry and the geometric irregularities are not considered.

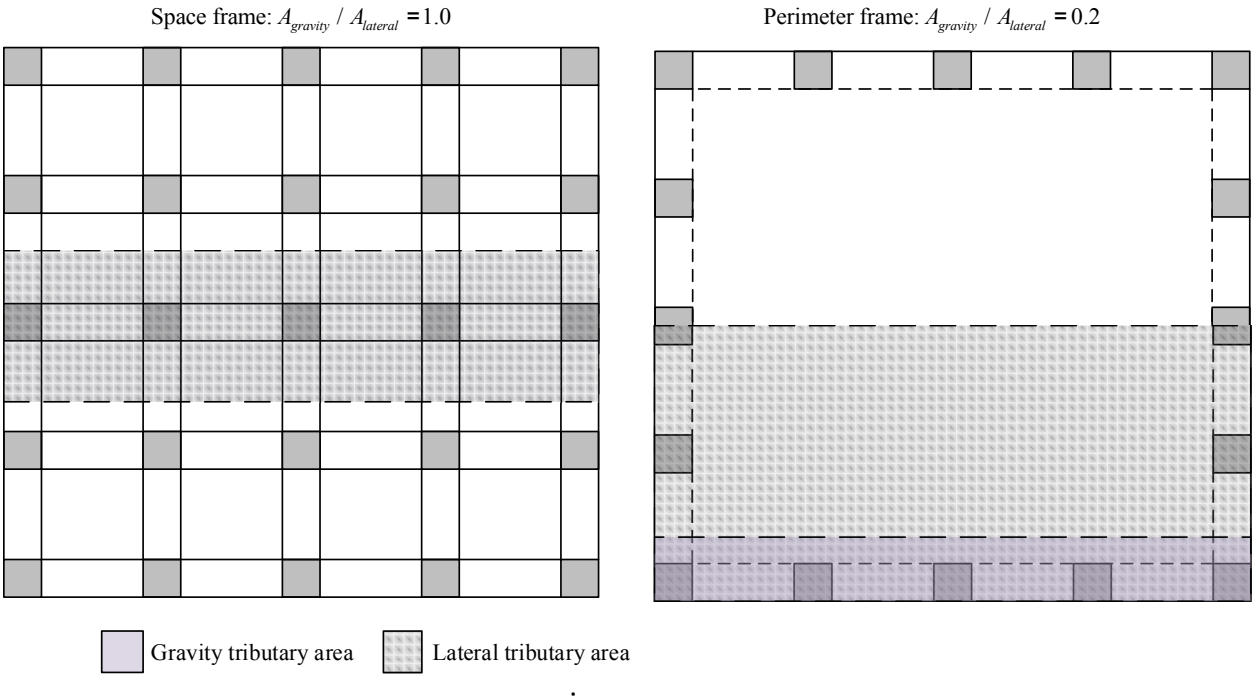


Figure 3.6. Plan view and ratio of gravity to lateral tributary areas of two archetypal structures.

In UBC, the required lateral force depends on the base shear (V) calculated as $V = CKW$. In this approximation formula, W is the weight of the building and C is the base shear coefficient determined by Eq.3.11, which, in turn, depends on the fundamental period of the structure and should not be larger than 0.1; K is the horizontal force factor varying from 0.67 (ductile) to 1.33 ($K = 1$ for non-ductile frames). In the proposed method, W is calculated based on the floor area, floor weight including dead and live loads, and the number of floors. The dead load is assumed to be 175 psf (8.4 KPa) and the live load is assigned based on the occupancy specified in UBC. The tributary areas of both space and perimeter frame configurations are illustrated in Fig.3.6, and are used to determine the weight contributing to the lateral forces. Additionally, C is calculated based on the structural period, which is assumed to be 0.1 times the total number of floors [9]; K is fixed to 1 for all NDRCB frames.

$$C = \frac{0.05}{\sqrt[3]{T}} \tag{3.11}$$

Once the base shear is obtained, the distributed lateral forces can be calculated using Eqs.3.12-3.14, wherein F_t and F_i represent the lateral forces applied to the roof and i-th floor; h_x, w_x, h_n, and w_n are the height and floor weight at the x-th floor and roof in the respective pairs; D_n is the plane dimension of the roof.

$$V = F_t + \sum_{i=1}^{n} F_i \tag{3.12}$$

$$F_t = 0.004V \left(\frac{h_n}{D_n} \right)^2 \tag{3.13}$$

$$F_x = \frac{(V - F_t)\, w_x h_x}{\sum_{i=1}^{n} w_i h_i} \tag{3.14}$$

The load combination required by UBC is used to analyze the maximum and second-largest internal forces and check the component capacity at the first floor and typical floors (2nd and above). The Ultimate Strength Design method and the detailing requirements specified in Section 2615, Section 2616, Section 2617, and Section 2619 of the UBC are applied to the archetypal structures.

Fig.3.7 delineates the design procedure for the analytical model. To obtain the component responses, a preliminary OpenSees model must be developed to perform static analysis. Six archetypal frames were hence selected from Leiel et al. [83] as the initial designs for structures with different configurations (i.e., number of floors, floor height, floor area, number of spans...etc). With the preliminary model, the observed forces can be used to check the component capacities. If the capacity is insufficient, then, the proposed algorithm will iteratively modify the reinforcement details (number and size of rebars) and member dimensions (depth and width). It is worth noting that the analytical model is not the as-built model but only its approximation. Moreover, although the algorithm is promising regarding convergence, the design outcome cannot exactly reproduce the decisions made by the original designers. The reason for implementing a design procedure is to generate unavailable information (e.g., dimensions of components, reinforcement ratio, rebar spacing) that is required to construct nonlinear elements for simulating the critical behavior of NDRCB frames. The generated

model can meet the minimum requirements of the code but is not an optimal design. Once the design procedure is completed, the proposed method automatically outputs tcl/tk files according to the element definitions summarized in Fig.3.5.

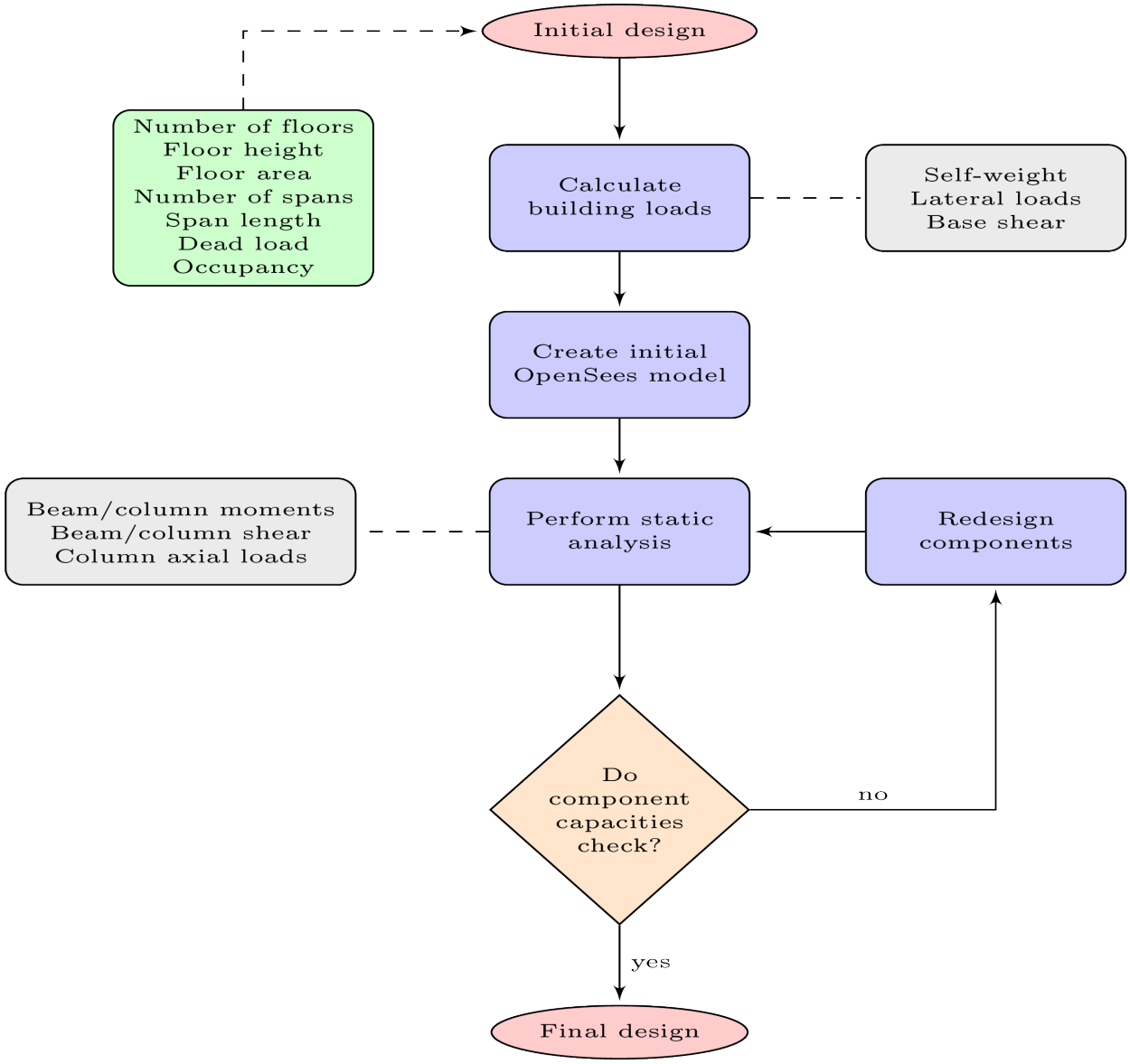


Figure 3.7. Design procedure in the proposed method.

3.4 Validation of the proposed procedure

To ensure that the auto-modeling procedure can reproduce the critical behaviors of the NDRCB frame, three cases were selected here for validating the analytical models generated through the proposed method. A summary of the reference models and the proposed model is presented in Table 3.2. The first one is the Van Nuys Hotel which was designed in 1965 according to the 1964 Los Angeles Building Code. Others are the archetypal frames investigated by prior studies. The design and modeling details of these buildings are described in [66, 79, 83].

Table 3.2. Comparison of modeling assumptions.

Property	Krawinkler [79]	Liel [83]	Jeon [66]	Proposed
Geometric configuration	7-story, 8-bay	8-story, 3-bay	4-story, 3-bay	as-referred
f_c' beam/column (ksi)	4/5	4/4	4/4	4/4
f_y beam/column (ksi)	50/75	40/60	40/60	40/60
Plastic hinge model	fiber section	concentrated plastic hinge	fiber section	concentrated plastic hinge
Shear failure considered	column	joint	column or joint	column
Shear failure model	force-distortion model	joint shear panel	joint spring or limit state material	limit state material
Loss of axial load capacity	not included	not included	included	included
P-Δ effect	not included	included	included	included
Gravity loads	not included	included	included	included

With the basic building configurations in reference, the proposed method is used to generate the numerical model for nonlinear static analysis (pushover), wherein the pushover loading pattern is calculated based on the equivalent lateral force specified in ASCE 7-10 [9]. The pushover results are presented in Fig. 3.8. All results reveal that these frames exhibit

shear failure after flexural yielding. The proposed model exhibits similar yielding strength and ultimate roof drift compared with the analyses in [83] and [66]. However, the ultimate roof drift is approximately 30% larger than the model reported by [79]. The reason for this discrepancy is because the difference in modeling the shear failure in the column. The shear force and shear distortion model used in [79] is independent of the loading history and the drift demand, while the limit-state material used in the proposed model takes the axial loading history into account to estimate the drift capacity. Although the proposed model implements elements that require less computational effort (e.g., the concentrated plastic hinge and limit state material model), the design and modeling results are still consistent with the results obtained by complex models.

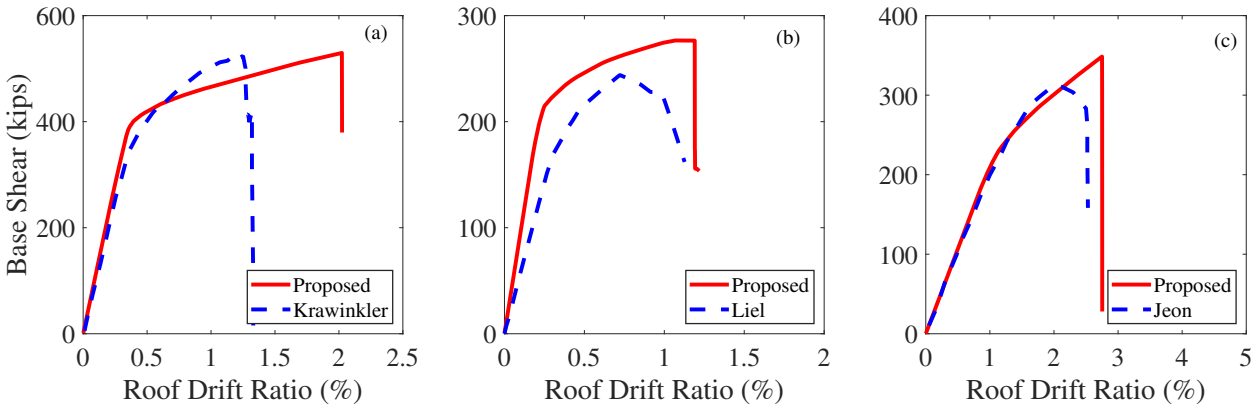


Figure 3.8. Pushover analysis of (a) Van Nuys Hotel, (b) eight-story perimeter frame, and (c) four-story perimeter frame.

In addition to static analysis, the actual data recorded at the Van Nuys Hotel are also used to validate the dynamic response of the proposed model. The Van Nuys Hotel was instrumented with the accelerometers in 1980 through the California Strong Motion Instrumentation Program (CESMIP). The 16-channel recording system considers both the north-south and east-west directions. The recorded acceleration data were obtained from the website of the Center for Engineering Strong Motion Data (`https://www.strongmotioncenter.org/`) and have been analyzed by researchers to obtain the fundamental period of buildings [49,119].

The natural period of the Van Nuys Hotel is approximately 1 s, owing to the light damage during the 1971 San Fernando earthquake (M6.6). The 1992 Big Bear earthquake (M6.5) and the 1994 Northridge earthquake (M6.7) were selected for calibrating the dynamic behavior in the linear and nonlinear regions.

Fig.3.9 shows the roof acceleration in the east-west direction of the Van Nuys Hotel subjected to the 1992 Big Bear and 1994 Northridge earthquake. Because most damage caused by the Northridge earthquake was observed in the east-west direction (north and south frames) [119], only the response in this direction is compared herein. The PGA measured during the 1992 Big Bear earthquake was 0.06 g; therefore, the building still remained in the linear region. The maximum response of the proposed model only has a difference of 3% compared with the actual data. While subjected to the Northridge earthquake, the building reached the nonlinear region because PGA was 0.59 g. The difference between the proposed model and the recorded data was approximately 5%. According to [119], several damages including diagonal/bi-diagonal cracks and flexural shear cracks along columns were observed from the ground floor to the 5th floor, which can be detected by the proposed model as shown in Fig.3.10. Although a mismatch is expected to exist between the proposed model and the actual data, owing to a certain degree of simplification, the proposed model can still capture the critical behavior of NDRCB frames, namely, shear failure.

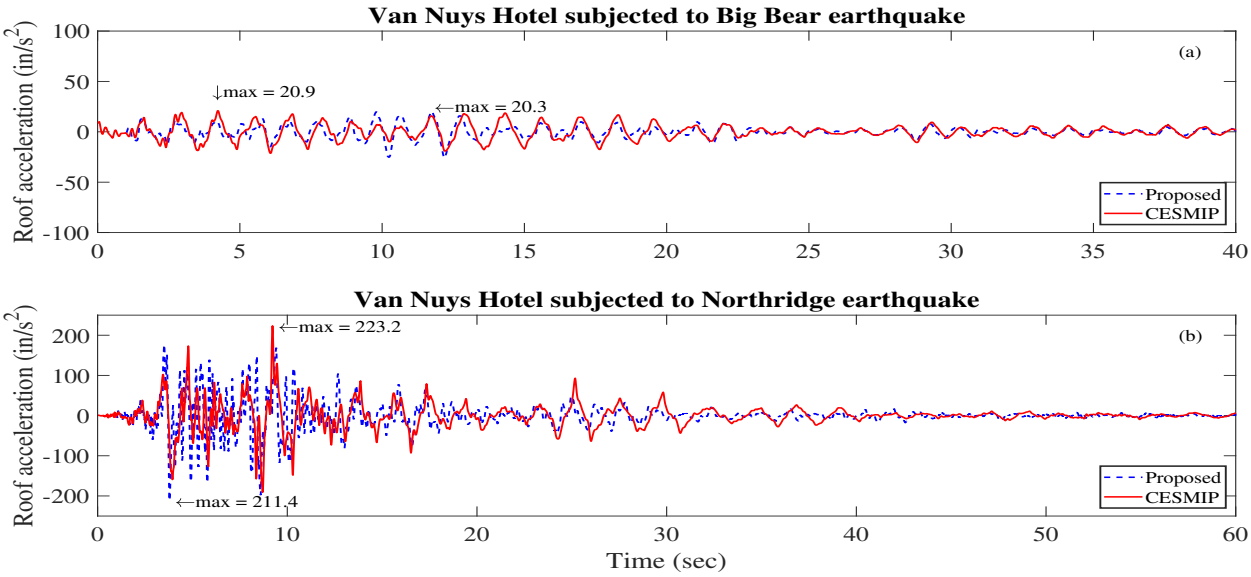


Figure 3.9. (a) Roof response to Big Bear and (b) Northridge earthquakes.

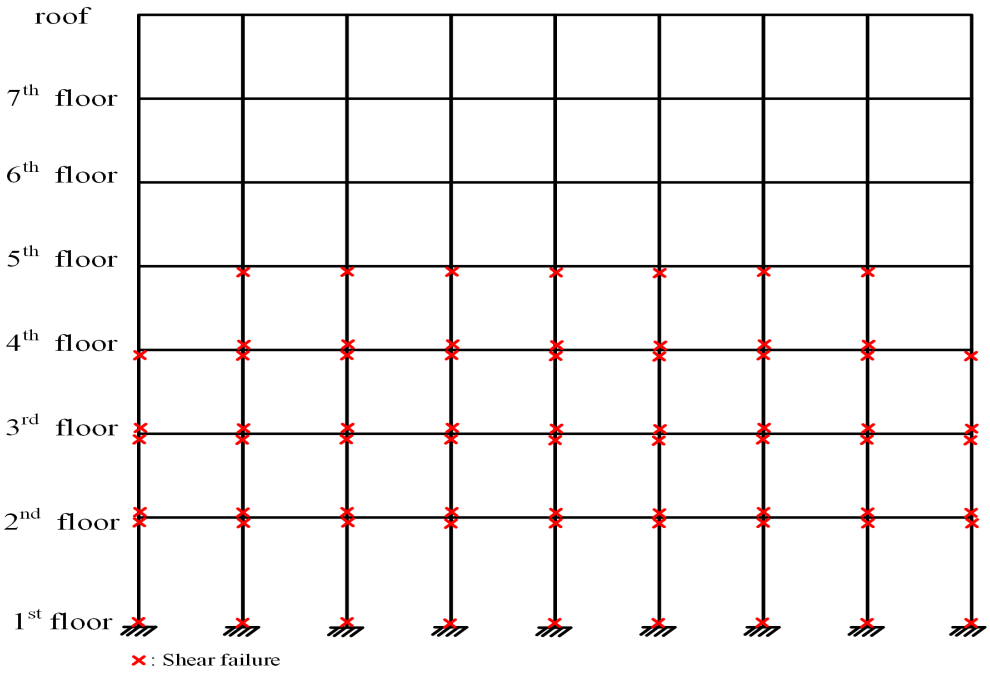


Figure 3.10. Detected shear failures in proposed model subjected to Northridge earthquake.

3.5 Construction of Fragility Functions

Once the archetypal model is prepared, a set of nonlinear time-history analyses (NTHAs) are required to compute the engineering demand parameters (EDPs) for damage assessment via fragility functions. To develop fragility curves, two probabilistic assessment methods are widely used. One is referred to as the probabilistic seismic demand modeling (PSDM) approach [31]. PSDM has a great advantage of being compatible with a closed form solution and does not require scaling of ground motions. Researchers have used PSDM to develop fragility curves for multiple performance levels of NDRCB [21, 68, 104]. The other method is the incremental dynamic analysis (IDA) [121] which requires intensive NTHAs with incrementally scaled ground motions. For instance, [14, 46, 85] developed fragility curves for system collapse using IDA. However, the scaling of ground motions without change in frequency content can produce unrealistic earthquake records that may not be representative of the seismic hazard in a region. Hence, PSDM is implemented in this section to construct fragility functions and quantitatively discuss the uncertainties/variability in the proposed method.

3.5.1 The probabilistic seismic demand model

As expressed by Eq.3.15, the PSDM describes the demand (D) as a function of an IM, wherein a and b represent the inverse logarithm of the vertical intercept and the slope that can be obtained after rearranging in the linear regression analysis, as expressed by Eq.3.16. The regression analysis data are obtained by subjecting a set of analytical models to a suite of N ground motions and recording the peak demand (e.g., the maximum inter-story drift and peak floor acceleration). The median demand is expressed by Eq.3.16, and the dispersion (logarithmic standard deviation) describing the record-to-record aleatoric uncertainty is calculated using Eq.3.17.

$$D = a(IM)^b \tag{3.15}$$

$$ln(D) = ln(a) + b \cdot ln(IM) \tag{3.16}$$

$$\beta_{D|IM} \cong \sqrt{\frac{\sum (ln(d) - ln(aIM^b))^2}{N-2}} \tag{3.17}$$

With the PSDM, a probabilistic curve describing the likelihood that the demand exceeds the capacity (C) given an intensity level can be constructed, namely, the fragility curve, as expressed by Eq.3.18. β_C and β_M denote the epistemic uncertainty in the capacity and modeling, which are assumed to be 0.3 [66] and 0.2 [21], respectively, for the NDRCB frames.

$$P(D > C|IM) = 1 - \Phi\left(\frac{ln(\hat{C}) - ln(a \cdot IM^b)}{\sqrt{\beta_{D|IM}^2 + \beta_C^2 + \beta_M^2}}\right) \tag{3.18}$$

3.5.2 Selected frames and ground motions

As a demonstration of the pipeline, archetypal frames developed by Liel et al. [83] are re-constructed through the proposed design procedure and the auto-modeling module to perform NTHAs and obtain EDPs. Analytical models (as shown in Fig.3.5) with the different numbers of stories (i.e., 2, 5, 12) and structural types (i.e., perimeter, space) are discussed herein. Following prior studies, all frames have three bays with a spacing of 25 ft (7.6 m). The first story and the typical story are set to be 15 ft (4.6 m) and 13 ft (3.96 m) in height, respectively.

To obtain structural responses with various intensities for regression analysis, the generic suite of ground motions proposed by Baker [13] are used in this study. The suite consists of four different sets including far- and near-field ground motions. In the first three sets, 120 pairs of broad-band ground motions with large magnitude and small distance are included: Set 1A(M_w = 7, R_{rup} = 10 km, soil site); Set 1B (M_w = 6, R_{rup} = 25 km, soil site); and Set 2 (M_w = 7, R_{rup} = 10 km, rock site). Set 4 contains 40 pairs of ground motions and strong velocity pulses are expected at near-fault directivity. The response spectrum for each set is shown in Fig.3.11. Each archetypal frame is subject to 320 ground motions (160 pairs) respectively, and all NTHAs are completed through the high-performance computing

platform Hoffman2 at the University of California, Los Angeles. It is worth noting that the vertical excitation is not considered in the NTHA although it may reduce the shear capacity of sehar-critical columns in NDRCB frame [52].

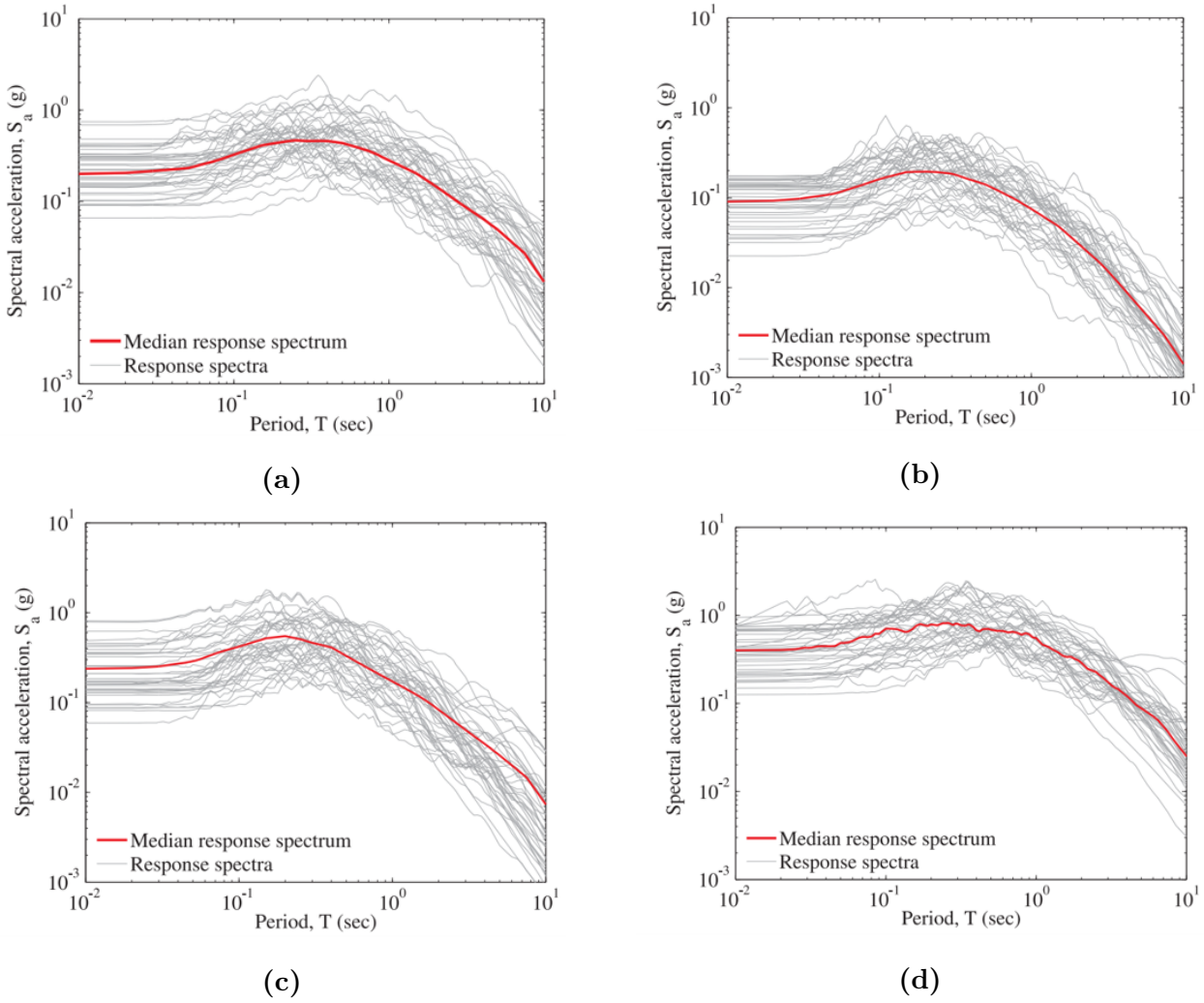


Figure 3.11. Response spectra for (a) set 1A, (b) set 1B, (c) set 2, (d) set 4 in Baker [13].

3.5.3 Damage fragility curves

With EDPs from NTHAs, linear regression analyses are then used to obtain the intercept and slope expressed in Eq.3.16. The peak inter-story drift ratio (θ_{max}) and spectral displacement

(S_d) are selected as the demand and IM to compare with HAZUS where the fragility is a function of peak displacement response. Thresholds of multiple damage states defined in HAZUS are used to generated fragility curves expressed in Eq.3.17.

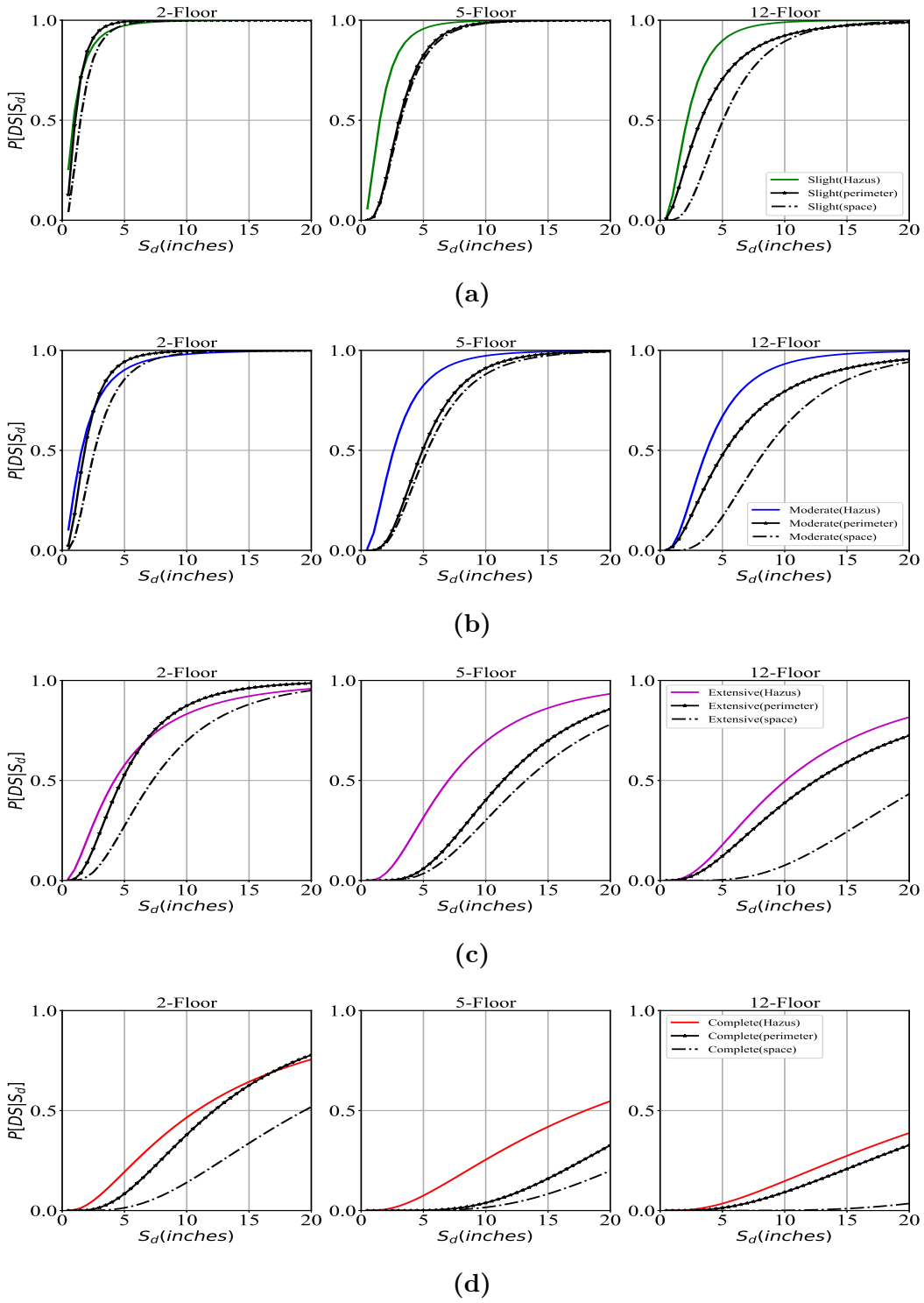


Figure 3.12. Fragility functions for (a) slight, (b) moderate, (c) extensive, and (d) complete damage.

Fig.3.12 shows four levels of damage fragility curves for each archetypal frame. The perimeter frame is vulnerable than the space frame because the $P-\Delta$ effect is considered in the numerical modeling, which cannot be found in HAZUS' fragility curves constructed through the capacity-spectrum method and SDOF systems. While the difference in modeling (i.e., SDOF v.s. MDOF) and analytical methods (i.e., static push-over v.s. nonlinear time-history) could be the reason behind the discrepancy between fragility curves presented here, the proposed design procedure is conservative by only providing two sections (i.e., first floor, typical floor) based on first and second-largest internal forces. It can be further improved when detailed drawings are provided or a better design algorithm can be introduced in the future. This section illustrates the fulfillment of the damage assessment module in the proposed pipeline. The selected archetypal frames can be further modified according to structure-specific data when applying the proposed method to the Los Angeles inventory in the next chapter.

3.6 Investigation of Intensity Measurements

As early mentioned, limited NDRCB frames (i.e., low-rise, mid-rise) and EDP(i.e., maximal inter-story drift) were selected to investigate the suitable IMs in prior studies. It is hence discussed herein by including additional IMs, structure type (i.e., high-rise), and EDP (i.e., peak floor acceleration) to provide a more comprehensive IM evaluation.

To identify the optimal IM, regression analyses of PSDM described in the previous sections are used to evaluate metrics including efficiency, sufficiency, proficiency, and practicality. Several IMs are investigated in this study, including the conventional amplitude-based measures (i.e., S_{aT1}, S_d, PGA, PGV), and cumulative and duration-based measures (i.e., CAV, AI, $Ds575$, $Ds595$). The maximum inter-story drift (θ_{max}) and peak floor acceleration (PFA) are selected as EDPs in PSDM because they are often used to evaluate the damage states for structural and non-structural components.

A lower $\beta_{D|IM}$ in the PSDM represents a smaller variation of the estimated demand for a given IM, and thus indicates an efficient IM. The practicality is used to evaluate the dependence of the demand and IM, which can be measured by the slope b in the regression analysis. If the slope is close to zero, the dependence of the demand and IM can be ignored; therefore, the IM is impractical. The proficiency is a composite measurement of efficiency and practicality, which can be calculated using Eq.3.19. A lower value of ζ indicates a more proficient IM and lower uncertainty in the demand model by the selection of the IM. Additionally, the coefficient of determination R^2 can validate the PSDM assumption of the linear relationship in lognormal space. A higher R^2 value indicates the goodness of the regression fit and reduces the dispersion among the dataset, and can be used to quantify the efficiency and proficiency.

$$\zeta = \frac{\beta_{EDP|IM}}{b} \tag{3.19}$$

Sufficiency is another property that can be used to measure IM variability. A sufficient IM is independent of the ground motion characteristics such as the magnitude (M_w) and epicentral distance (R), which can be evaluated by the statistical p-value. The p-value is the probability of obtaining results as extreme as the results obtained by a statistical hypothesis test, wherein the null hypothesis is the independence of the IM and ground motion characteristics. If the p-value is greater than the significance level (e.g., 5% [59], M_w and R are not statistically significant, and thereby the null hypothesis holds and the IM is sufficient. The p-value can be obtained by the linear regression of the residuals from PSDM against M_w or R.

As shown in Table 3.3, PGV dominates the efficiency, proficiency, R^2, and sufficiency for θ_{max}. On the other hand, PGA dominates the practicality, proficiency, and sufficiency for PFA as summarized in Table 3.4. Fig.3.13 shows the regression analyses of θ_{max} and PFA that are recommended by this investigation. As has been mentioned, the goodness of the linear fit in log space demonstrates that the IM is efficient, practical, and proficient. Moreover, if the regression line of the residual (e.g., $\epsilon_{\theta_{max}|PGV}$ or $\epsilon_{PFA|PGA}$) against the

ground motion characteristics (e.g., M_w or R) is flat, the IM is sufficient. It must be noted that the damage fragility functions constructed in the previous section are using the spectral displacement as IMs to follow the conventional risk assessment methodologies (i.e., HAZUS). It can be easily changed because PSDMs for multiple IMs have been conducted. Complete results of regression analyses are provided in Appendix A.

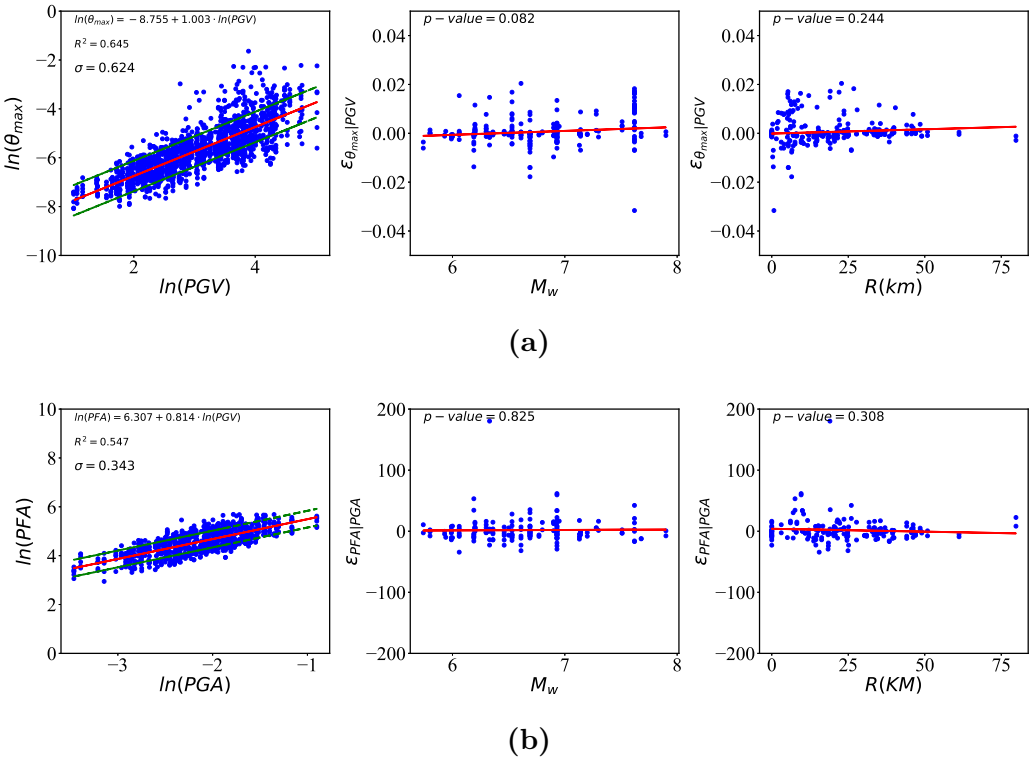


Figure 3.13. Regression analyses for (a) θ_{max} and (b) PFA.

Table 3.3. Evaluation metrics of IMs for θ_{max}.

IM	Efficiency	Practicality	Proficiency		Sufficiency	
	$\beta_{D\|IM}$	b	ζ	R^2	M_w	R
S_{aT1} (g)	0.73	0.71	1.03	0.33	0.0	**0.33**
S_d (inches)	0.68	0.96	0.71	0.62	0.02	0.0
PGA (g)	0.59	**1.05**	0.57	0.46	0.0	0.0
PGV (m/s)	**0.39**	1.0	**0.39**	**0.65**	**0.08**	0.24
I_a (m/s)	0.5	0.68	0.73	0.55	0.0	0.12
CAV (m/s)	1.08	0.34	3.17	0.02	0.0	0.0
$Ds575$ (sec)	1.03	0.19	5.4	0.02	0.0	0.0
$Ds595$ (sec)	1.02	0.35	2.9	0.04	0.0	0.0

Table 3.4. Evaluation metrics of IMs for PFA.

IM	Efficiency	Practicality	Proficiency		Sufficiency	
	$\beta_{D\|IM}$	b	ζ	R^2	M_w	R
S_{aT1} (g)	**0.1**	0.2	0.53	**0.61**	0.39	0.0
S_d (inches)	0.51	0.69	0.74	0.47	0.0	0.0
PGA (g)	0.12	**0.81**	**0.15**	0.55	**0.83**	**0.31**
PGV (m/s)	0.18	0.36	0.51	0.29	0.21	0.0
I_a (m/s)	0.15	0.34	0.43	0.43	0.06	0.13
CAV (m/s)	0.26	2.1	0.12	0.0	0.02	0.0
$Ds575$ (sec)	0.49	0.31	1.6	0.1	0.02	0.0
$Ds595$ (sec)	0.48	0.41	1.2	0.13	0.05	0.0

3.7 Sensitivity Analysis

As mentioned before, the goal of this pipeline is to provide a single method that automates the modeling and damage assessment for NDRCB frames. When it is available to the public, EDPs computed through the proposed method are expected to be varying when using structure-specific data provided by users, and the damage state fragility curves may also be varying. Given this circumstance, the sensitivity of the proposed method is necessary to be identified. Specifically, which variable of building configurations needs more attention during the future application?

Considering that potential users may not precisely provide the building configurations, a deterministic sensitivity analysis is conducted in this study to quantify the variation introduced by the user-provided variables (i.e., first and typical floor height, floor area, span length) in EDPs. Archetypal frames applied in the previous sections are first used as baseline models, and each variable is then changed to an extreme value at a time for conducting NTHAs. Each variable has three values: first floor height (i.e., 10 ft, 15 ft, 20 ft); typical floor height (i.e., 8 ft, 13 ft, 18 ft); floor area (i.e., 2812 ft^2, 5625 ft^2, 8437 ft^2); span length (i.e., 12.5 ft, 25 ft, 37.5 ft).

The variability of the maximum inter-story drift (θ_{max}) and peak floor acceleration (PFA) are quantified through the quartile coefficient of dispersion (QCD) [17] which is less prone to the outlier influence. The QCD is calculated by $(Q3 - Q1)/(Q3 + Q1)$, where $Q3$ and $Q1$ denote the third and first quartile of the dataset. The QCD of θ_{max} and the PFA for archetypal frames are summarized through the box plots shown in Fig.3.14 and 3.15. These results indicate that PFA is relatively less sensitive to the user-defined variables than θ_{max}. The effect of a single variable is not significant compared to the type of structure (i.e., high-rise perimeter frame). It must be noted that the sensitivity is investigated given the condition that users only change basic building configurations instead of modeling assumptions in the proposed method. The detailing decisions are not investigated because the design procedure

only generates two types of sections (i.e., first floor, typical floor) that satisfy the minimum requirement of the UBC of the time. Given this simplification, the archetypal models are believed to be conservative, and the detailing will not vary too much.

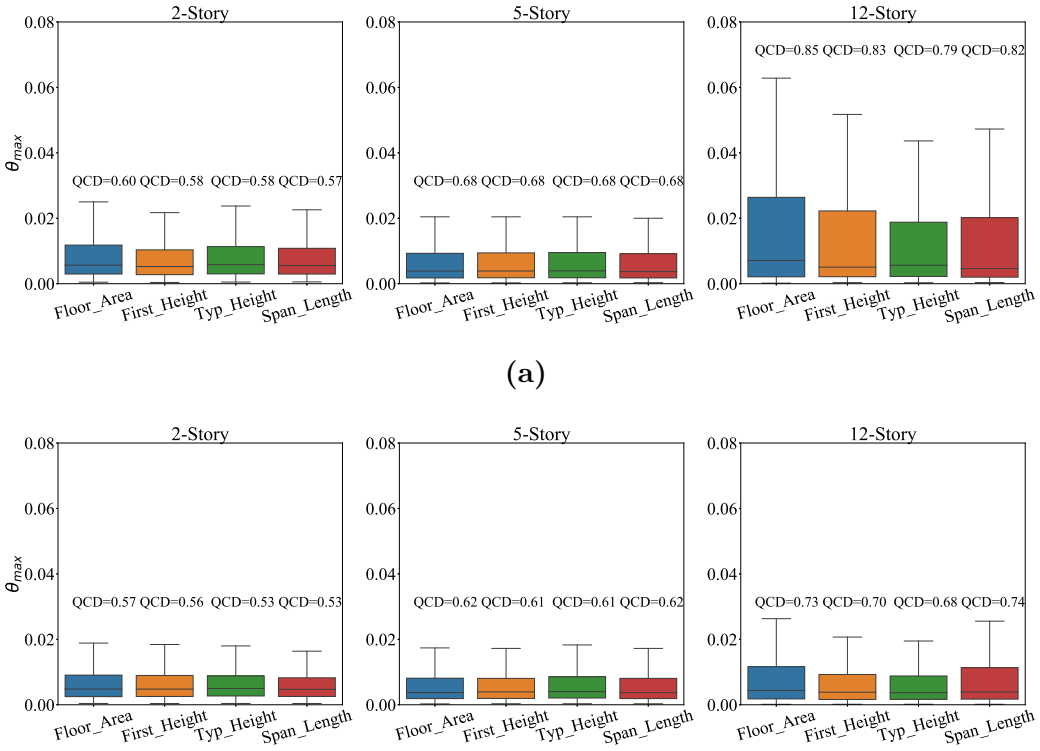
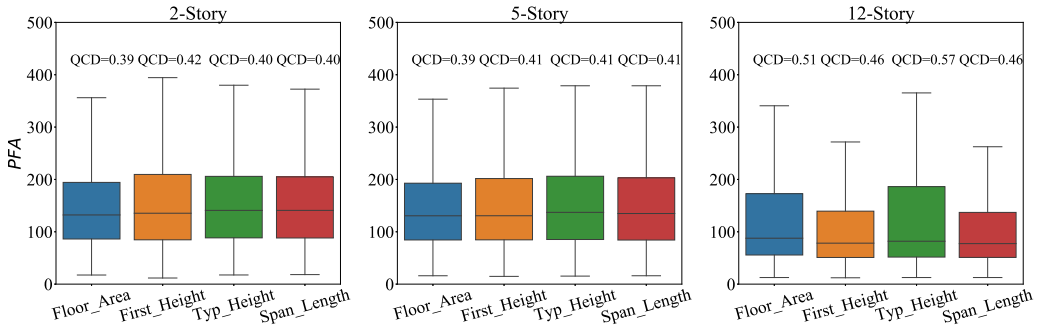
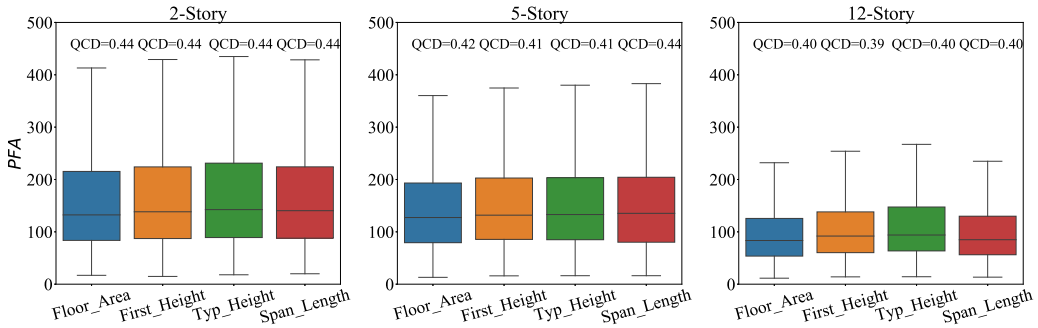


Figure 3.14. Box plots of θ_{max} for (a) perimeter and (b) space frames.



(a)



(b)

Figure 3.15. Box plots of PFA for (a) perimeter and (b) space frames.

3.8 Summary

In this chapter, a pipeline for automating the development of models and damage fragility functions was proposed for NDRCB frames. Archetypal frames were constructed through basic building configurations and a design procedure following the 1967 UBC. Nonlinear elements that are capable of simulating the stiffness degradation and brittle failure in columns were integrated into the proposed method, and the OpenSees Tcl/Tk files were automatically prepared. The method was validated through nonlinear static and nonlinear dynamic analyses with comparison to previously investigated NDRCB frames.

A generic suite of ground motions was used to conduct nonlinear time-history analyses

of archetypal frames for developing the probabilistic seismic demand model, which was further used to construct the damage state fragility curves. The damage state fragility curves were constructed using PSDM and HAZUS, which illustrated that the perimeter frame is vulnerable than the space frame due to the consideration of the $P-\Delta$ effect. While the proposed method can improve the shortage of using SDOF in HAZUS and can satisfy the demand of detailed structural response in the P-58 method, the design outcome is conservative with only providing two sections (i.e., first floor, typical floor), which can be further improved in the future with detailed drawings or a better design algorithm. The efficiency, practicality, proficiency, and sufficiency metrics were investigated for suitable IMs, and the PGV and PGA are suggested for the maximum inter-story drift and peak floor acceleration, respectively. A deterministic sensitivity analysis was conducted to quantify the variability of the proposed method with different building information. The sensitivity analysis indicated that PFA is relatively less sensitive to the user-defined variables than θ_{max}, and the effect of a single variable is not significant compared to the type of structure (i.e., high-rise perimeter frame).

CHAPTER 4

Application to Los Angeles' Inventory

This chapter is partly based on the following publication:

Chen, P.Y., Meng, S., and Taciroglu, E. (2021). "A data-driven framework for regional seismic loss estimation: a case study of non-ductile reinforced concrete buildings." *Engineering Structure* (ready to submit).

4.1 Overview

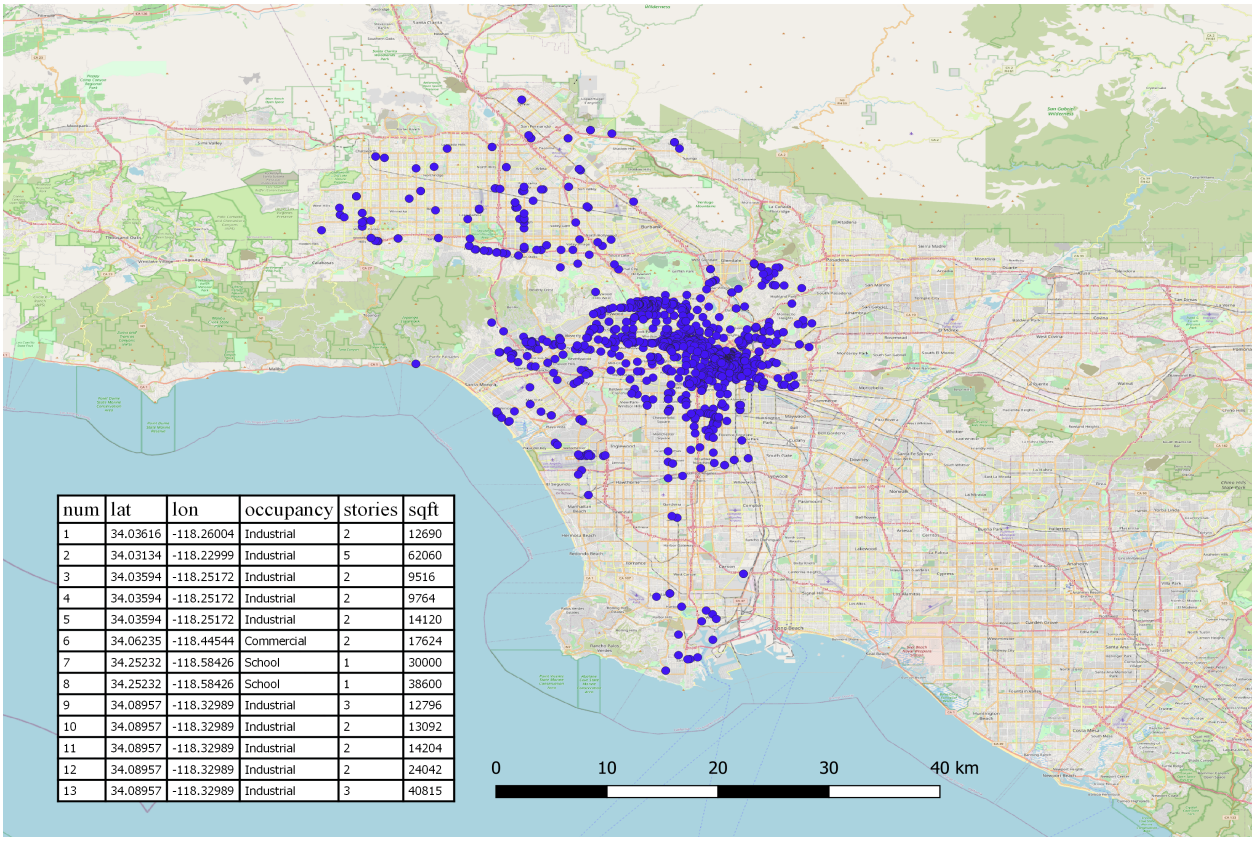
In this chapter, an application of the proposed pipeline to NDRCBs in the city of Los Angeles is presented. Prior studies provide preliminary information on these vulnerable buildings, yet more data needs to be excavated for constructing archetypal models and calculating structural responses for loss estimation. To address it, a data mining module combining web-scraping techniques and deep learning models to enhance the missing data is proposed. Given the previously investigated data as ground-truth information, CNNs are used to obtain essential building configurations such as number of floors, floor height, number of spans, span length, and floor area. With this building information, city-scale modeling and analyses are carried out. Engineering demand parameters obtained from the nonlinear time-history analyses are then used to estimate the seismic loss. An open-source Python application programming interface (API), namely, *pelicun* is integrated into the pipeline to implement

HAZUS [43] and FEMA P-58 [44] methodologies.

4.2 Data Collection and Validation

4.2.1 Inventory of Los Angeles

A series of studies done by Anagnos et al. [5–8] developed an inventory including 1,452 non-ductile reinforced concrete buildings in Los Angeles. This inventory (`https://graphics.latimes.com/la-concrete-buildings/`) compiled data set includes: structural type, use, year built, number of stories, total square footage, and occupancy. A spatial distribution of these NDRCBs and example data are shown in Fig.4.1. The majority ($\approx 80\%$) of them located in the City of Los Angeles within Los Angeles County.



num	lat	lon	occupancy	stories	sqft
1	34.03616	-118.26004	Industrial	2	12690
2	34.03134	-118.22999	Industrial	5	62060
3	34.03594	-118.25172	Industrial	2	9516
4	34.03594	-118.25172	Industrial	2	9764
5	34.03594	-118.25172	Industrial	2	14120
6	34.06235	-118.44544	Commercial	2	17624
7	34.25232	-118.58426	School	1	30000
8	34.25232	-118.58426	School	1	38600
9	34.08957	-118.32989	Industrial	3	12796
10	34.08957	-118.32989	Industrial	2	13092
11	34.08957	-118.32989	Industrial	2	14204
12	34.08957	-118.32989	Industrial	2	24042
13	34.08957	-118.32989	Industrial	3	40815

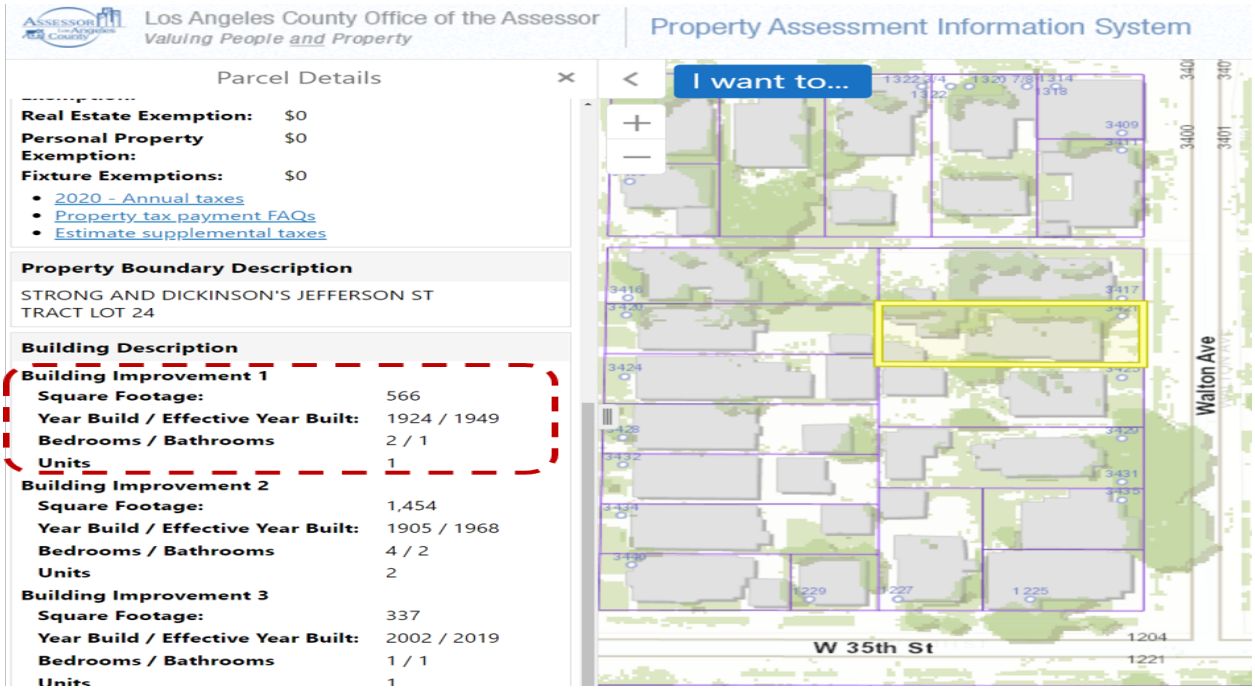
Figure 4.1. Los Angeles NDRCB inventory map.

Many data sources are the backbone of the inventory, which includes:

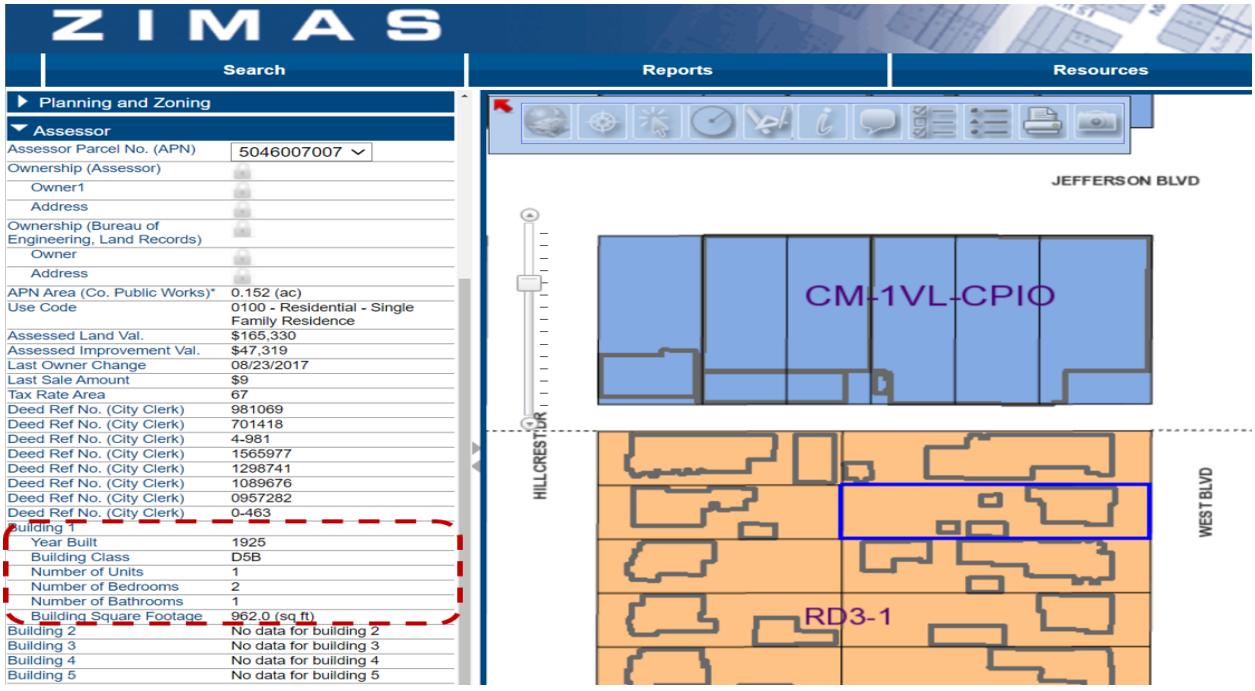
- LA County Assessor Map: `https://maps.assessor.lacounty.gov/m/`
- LA Zooming Information and Map Access System (ZIMAS): `http://zimas.lacity.org/`
- Los Angeles City Department of Building and Safety (LADBS): `https://www.ladbs.org/services/check-status/online-building-records`
- Google Maps: street-view and aerial images
- Sidewalk visits and building surveys
- Input from volunteer local engineers through cooperation with the Concrete Coalition [25, 26]

Assessor records are available in the assessor map where the size, number of units, year of built, tax records, square footage, property type, and land value may be found through an interactively searching interface. During the development of the inventory, researchers found that many buildings have multiple entries in the assessors' data when there are multiple owners. ZIMAS was hence used extensively to validate and verify those data. For a particular address or parcel number, permit information may even be available and which leads to the data of LADBS. Examples from the assessor map and ZIMAS are shown in Fig4.2.

Building permit data supposed to provide the most direct help for modeling. For example, plane drawings could offer interior dimensions in addition to the perimeter length and the floor area as shown in Fig.4.3. However, obtaining all permits for these vulnerable buildings is a challenge because plans for older buildings are sporadically available and can only be achieved if paid for by owners.



(a)



(b)

Figure 4.2. Available data from (a) L.A. assessor map, and (b) ZIMAS.

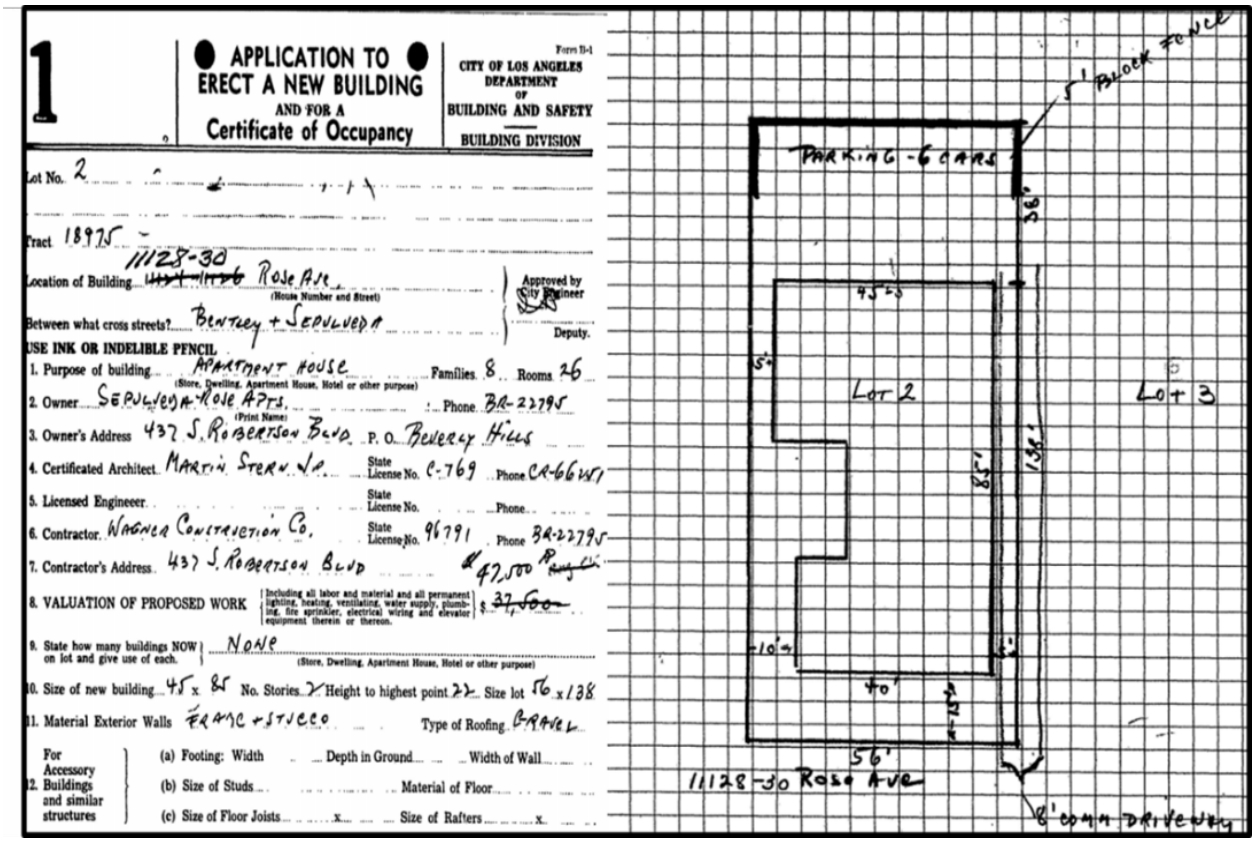


Figure 4.3. Example of building permit from LADBS.

The lesson from this inventory shows that an address-by-address search is labor-intensive and cumbersome. Not to mention that some data will still be missing even after going through each source. While the existing inventory data can partially satisfy the need for this study, it is a good opportunity to develop automatic data-harvesting tools through the power of the internet of the thing (IoT). The goal is to automatically scrape the available data sources by only giving the location of the target building. The algorithm currently focuses on NDRCBs in Los Angeles but can be further applied to different properties in different urban regions.

4.2.2 Data harvesting via web-scraping

Given the previously mentioned demand for large number data mining, the automatic data-harvesting module is proposed as shown in Fig.4.4. Two roadmaps are included and both of them are developed through open-source APIs. The first roadmap focuses on web scraping and is constructed through Python and relevant packages.

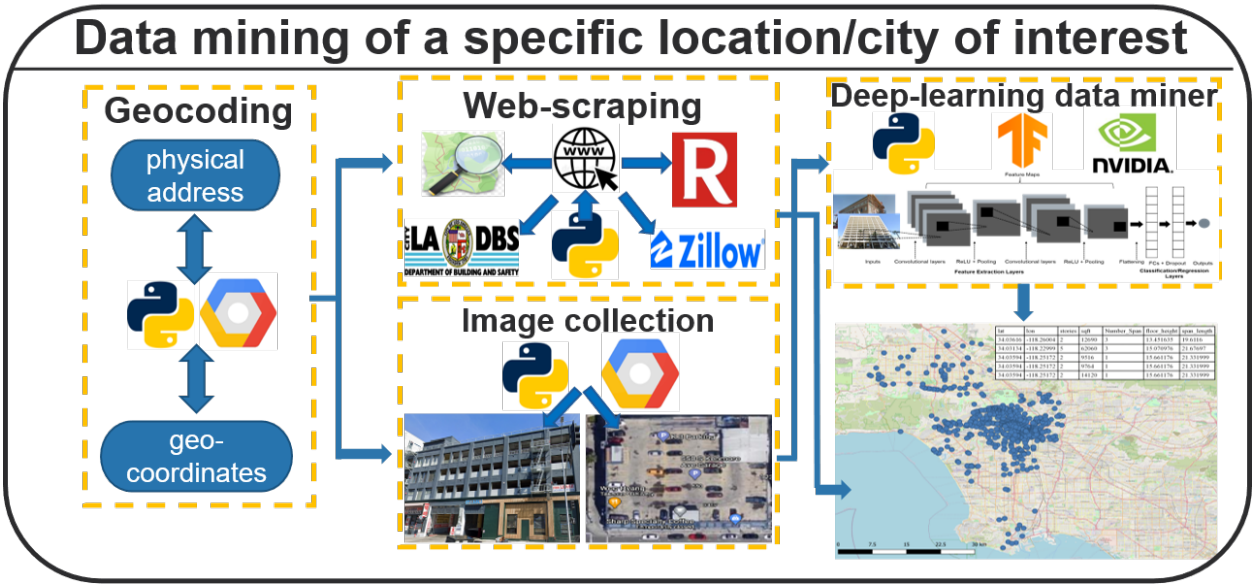


Figure 4.4. Schematic of data-harvesting module.

As shown in Fig.4.5, the procedure starts with a given location which can be either a physical address or geolocation. Most of the time, a physical address will be used in the automated web scraping, so the conversion needs to be done through Google Map API where a user-specific activation key is required. Once the address is prepared, a Python-based package called *selenium* (`https://pypi.org/project/selenium/`) will use a web browser (e.g., Google Chrome) and get access to a website (e.g., LADBS, ZIMAS, Zillow). After entering the web page, XPath on the HTML will be identified to fill in the provided address and the search button will be executed by *selenium*. An example is shown in Fig.4.6 where the XPath can lead to the total parcel area of the building.

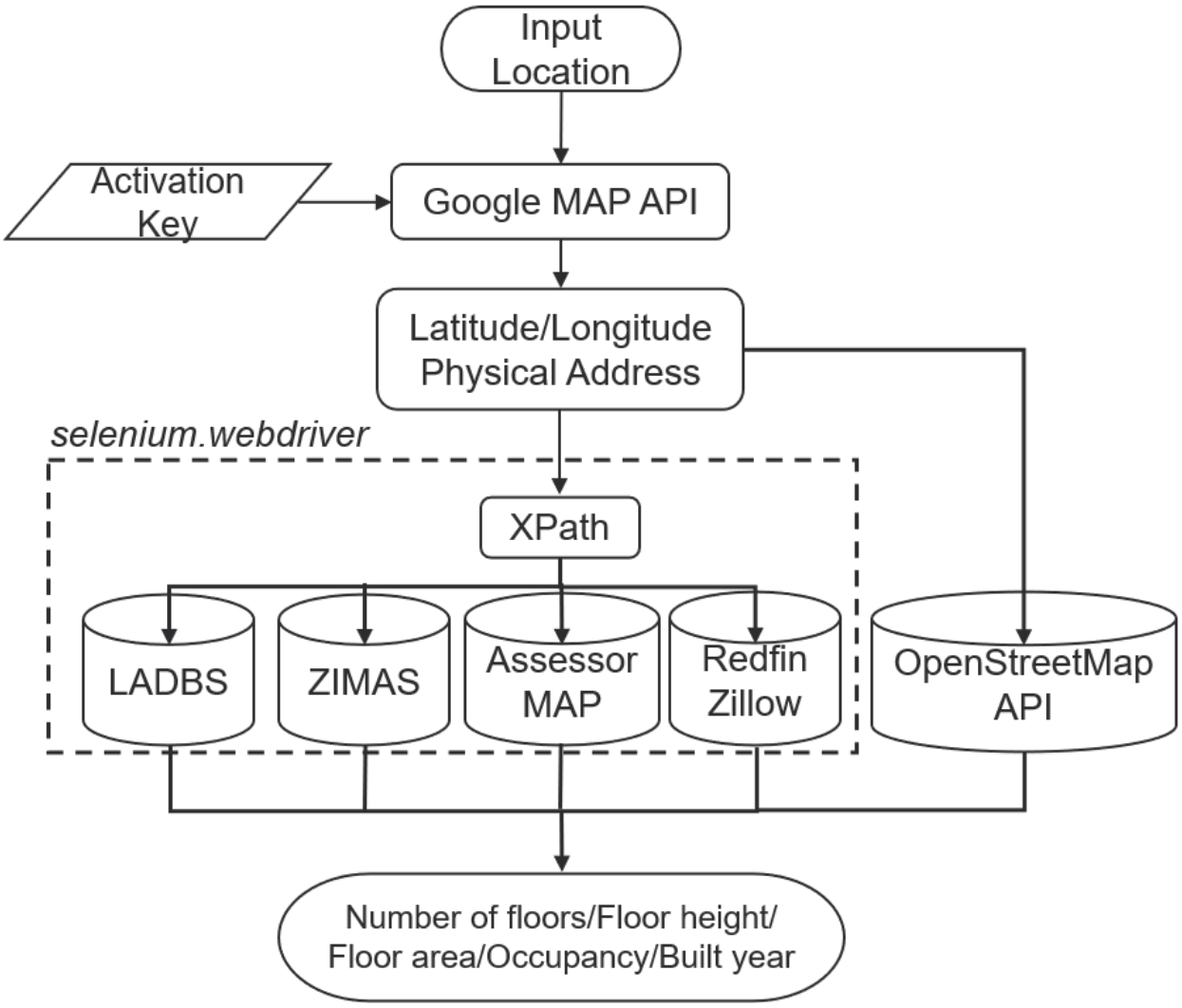


Figure 4.5. Flowchart of web-scraping.

Similar procedures can be done for different sites. In addition to the govern-owned website, some commercial websites such as Zillow and Redfin are also useful. For example, the rental price and can be found and can be used to calculate non-direct loss during an earthquake. Table 4.1 summarizes the data that may be obtained from these websites.

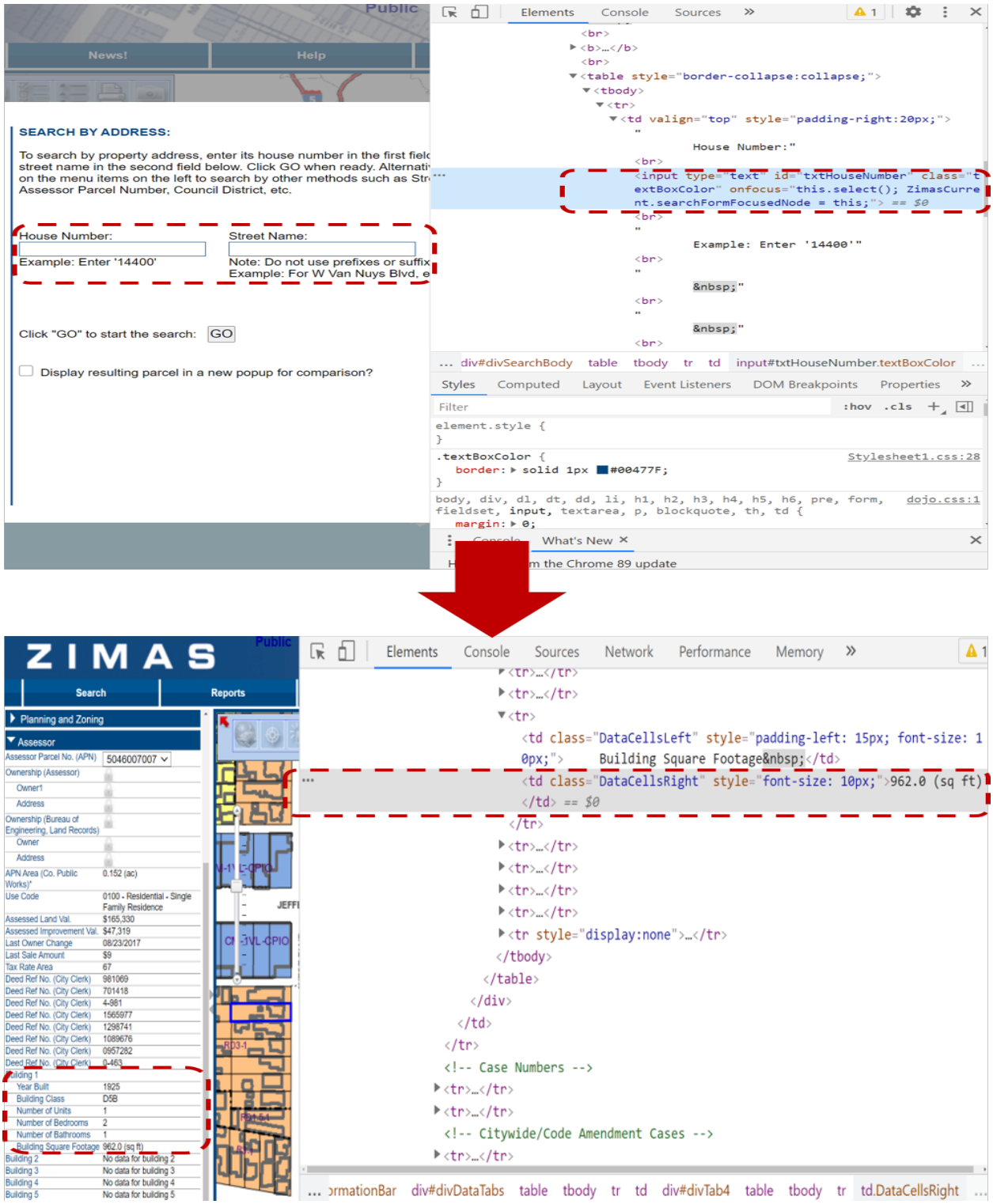


Figure 4.6. Xpath of ZIMAS.

Table 4.1. Summary of available data.

Data	LADBS	Assessor Map	ZIMAS	Zillow
Building Permit	✓			
Address	✓	✓	✓	✓
Usage	✓	✓	✓	✓
Lot/Parcel Area	✓	✓	✓	✓
Story	✓			
Height	✓			
Units	✓	✓	✓	✓
Soil Type	✓			
Year Build	✓	✓	✓	✓
Bathroom/Bedroom		✓	✓	✓
Fault Information			✓	
Land Value		✓	✓	
Rental Estimation				✓

Data sources discussed before seldom provide the building height, which is an important variable to conduct structural modeling and performance assessment. In most of the urban regions of the US, Open Street Map (OSM) (https://www.openstreetmap.org), a collaborative project to create a free editable map of the world, contains millions of building footprints including polygons with geocoordinates, building attributes, elevation, and most importantly, height. The height information shown in Fig.4.7 has a unit in meters and it is the total height of the building. That is to say, if the number of floors is available, the story height can be computed accordingly. Unlike other data sources using XPath, OSM has an API (`https://osmnx.readthedocs.io/en/stable/`) available through Python and hence can directly be accessed to get the height.

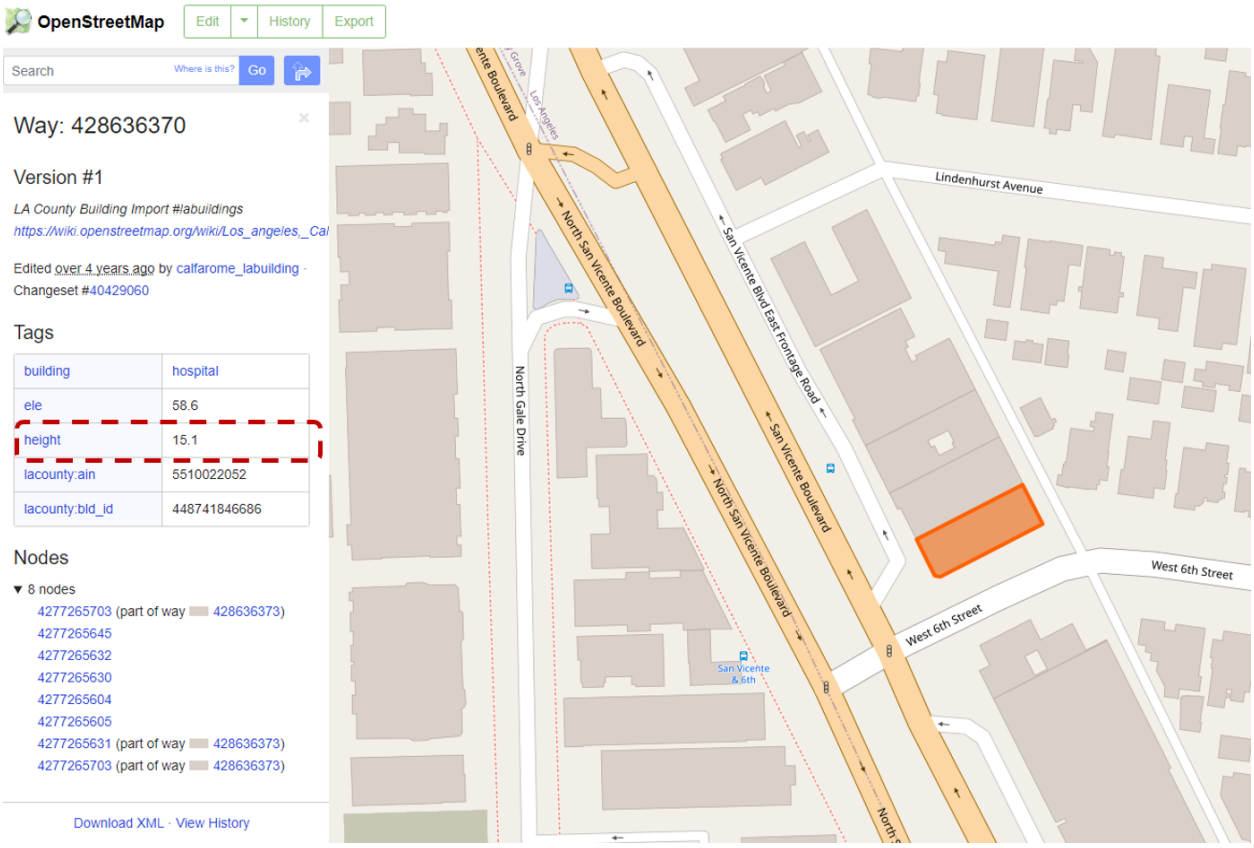


Figure 4.7. Example of Open Street Map.

The web-scraping roadmap currently focuses on Los Angeles and data sources used in the proposed roadmap need to be changed for application to other regions. For example, the building permit information should be investigated to see if available in local departments. There is no guarantee that a single methodology can collect all required information. It highly depends on whether those data have been digitized. Besides, building footprints may not be always available on OSM, which could be different region by region. Alternatively, another scheme is using images that almost always available from Google Map and Google Earth. The existing building information, as well as height obtained through OSM, will be the ground truth for training CNN models in the image-based roadmap shown in Fig.4.4.

4.2.3 Data harvesting via deep learning

For modeling archetypal buildings using the proposed pipeline, some building information is already available from the existing inventory but some are not. For example, the number of spans and the span length. Given the recently well-developed deep learning techniques in computer vision, data-harvesting via street-view and aerial images is potentially promising and hence implemented. Tasks including detection of objects (i.e., floor, span) and regression of variables (i.e., height, length, area) are carried out through CNN models. Although some data are available in the inventory, specific agents for detection and regression are still trained for future application to different regions.

In Chapter 2, CNN models were developed for a classification task. The last layer used the *sigmoid* function to compute the probability of each category. For the regression case developed in the image-based roadmap, a linear function is used to compute a continuous variable (i.e., building height, floor, span length). For example, the top layer of an input image can be 1×256 after several convolutional layers and fully connected layers. A linear function $a\mathbf{X} + b$ where $\mathbf{X} = [x_1, x_2, \ldots, x_{256}]^T$ then converts it into 1×1 output. Another difference for the regression task is the loss function. Compared to Eq.2.4, the mean squared error shown in Eq.4.1 is generally used as the loss function to update the weights for the regression task. The back-propagation is similar to the classification task described earlier.

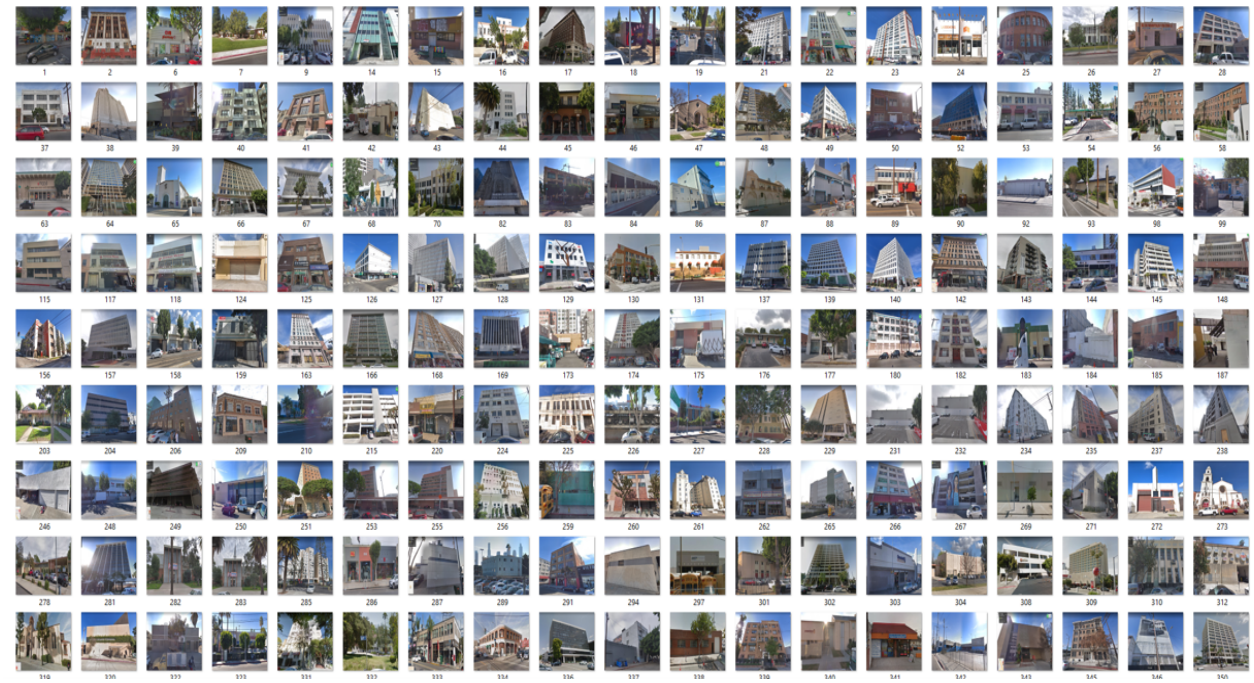
$$MSE(y, \hat{y}) = \frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y_i})^2 \tag{4.1}$$

For the regression tasks, three geometric variables are carried out, including building height, span length, and floor area. While these tasks seem to be promising for state-of-the-art CNN models, it highly depends on the training data set and the ground-truth labeling which is unavoidably time-consuming. Because obtaining and labeling a large number of images is not feasible, the TL technique introduced in Chapter 2 is implemented again to train models on a small group of images to avoid overfitting.

With the known addresses of NDRCBs, Google Map API is used to collect street-view

and aerial images as shown in Fig.4.8. These images are used to train CNN models for predicting floor height and floor area which are given in the inventory. It is worth noting that fully automated collection is not able to get a clear observation of target buildings. 1,452 street-view and aerial images are selected manually, and around 80% of them are the training set and the rest ($\approx 20\%$) are the validation set.

Not like the building height and floor area, the ground truth of span length is not available in the inventory. Fortunately, a database of instrumented buildings is available for California. The California Strong Motion Instrumentation Program (CSMIP, `https://www.strongmotioncenter.org/`) installed accelerometers on tens of thousands of infrastructures including bridges and buildings to record strong-motion time histories. For buildings, different types of stations can be selected. For example, 703 records are available from 77 reinforced concrete buildings whose configurations are also provided. As shown in Fig.4.9, the plan dimension can be obtained from CSMIP and can be used as ground truth to train CNN models.



(a)

(b)

Figure 4.8. Collected (a) street-view, and (b) aerial images of NDRCBs.

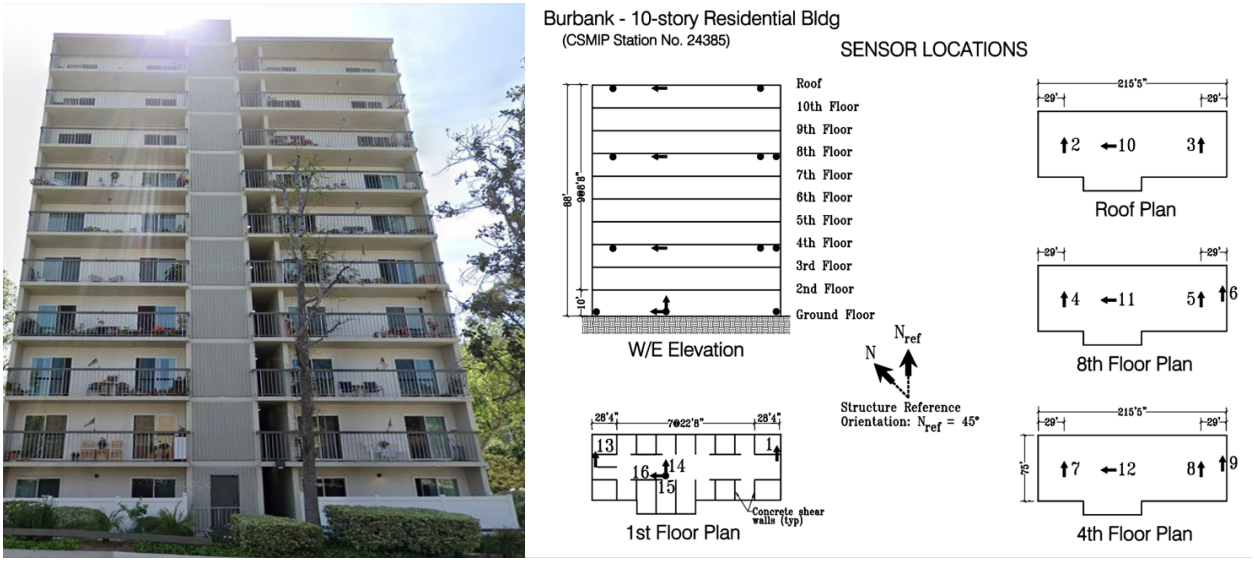


Figure 4.9. Side and plan view of CSMIP station.

However, using only 77 images (i.e., stations) as the training data set is not enough and will easily cause overfitting. As a result, some image augmentation techniques are implemented herein to enhance the training data. For instance, multiple images for the same site can be created by mirroring, rotating, and brightening the original image as shown in Fig.4.10 This can be easily fulfilled through the Python Imaging Library (PIL, `https://pillow.readthedocs.io/en/stable/`), and eventually, 770 images were used for training regression models of span length.

The training was done through the same machine with Nvidia GPU mentioned in Chapter 2. The performances are summarized in Table 4.2. The evaluation was done through 10-fold cross-validation and only the best model for each task is presented. The pre-trained CNNs implemented in TL have been introduced in Chapter 2 including VGG Net, Inception Net, and Res Net. The MSE may not be an intuitive performance metric, so the coefficient of determination (R^2) is also included here to illustrate how well the observed outcomes are replicated by the model. The R^2 can be computed in Eq.4.2.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{4.2}$$

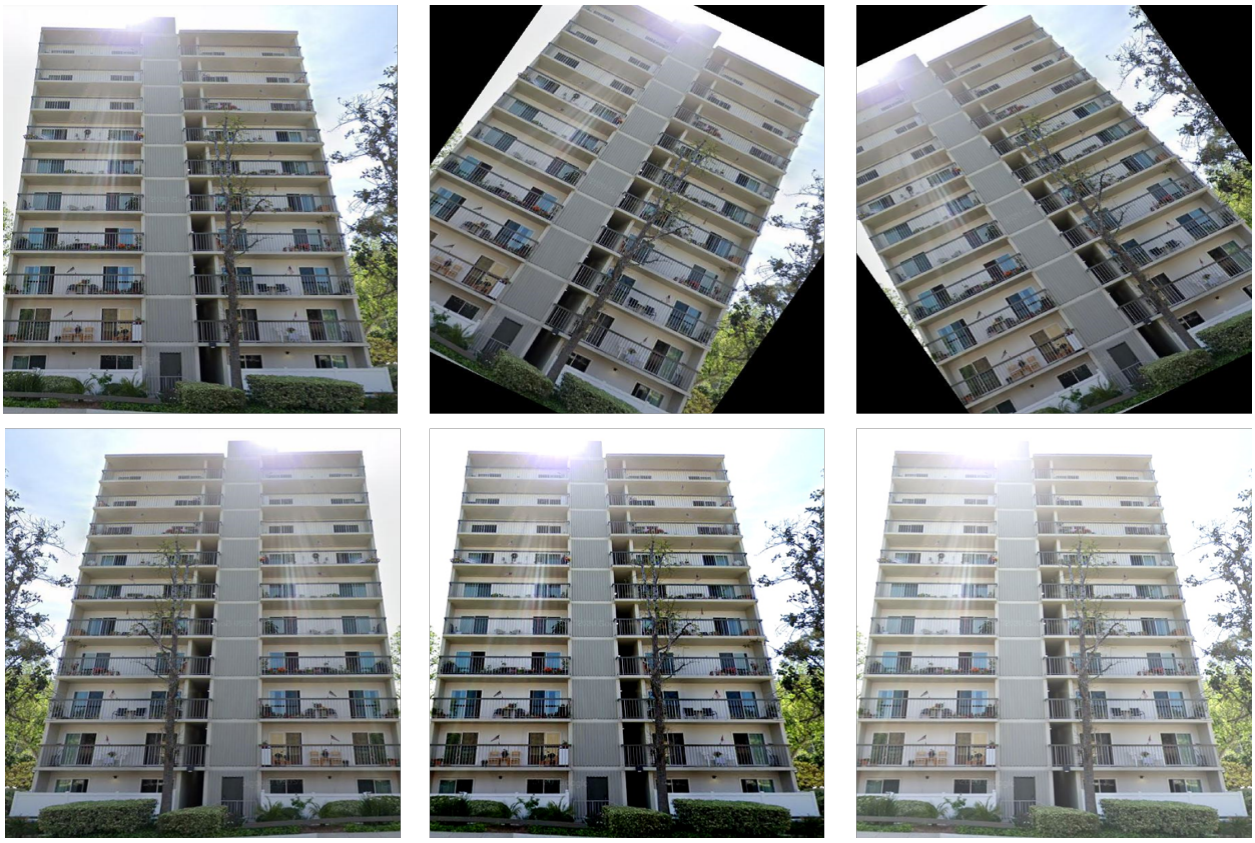


Figure 4.10. Image augmentation including rotation, mirror, and brightness.

Table 4.2. Performance of CNN models for regression tasks.

Task	Building height (ft)	Span length (ft)	Floor area (ft^2)
Image	street-view	street-view	aerial
Ground-truth	L.A. inventory	CSMIP	L.A. inventory
Training	1307	693	1307
Validation	72	38	72
Test	73	39	73
TL baseline	VGG	VGG	ResNet
μ_{MSE} | CV_{MSE}	1.38 | 0.62	6.79 | 0.59	21.5 | 0.6
μ_{R^2} | CV_{R^2}	0.94 | 0.1	0.94 | 0.06	0.89 | 0.06

The rest task for harvesting building information is to detect the number of floors and spans through CNN models, which is a mature application in the field of object detection. The main procedure of the object detection task is to first locate the potential object through a bounding box and then classify the type of it. The region-based convolutional neural network (R-CNN) [51] family has dominated these image detection and segmentation tasks for several years. The first version of R-CNN is straightforward yet computationally expensive. It starts with proposing potential regions of objects and uses CNN to extract features. A linear classifier is then applied to these features and identifies the target class. The disadvantage is the computational cost of features from each proposed region. To address this, a fast version of RCNN (Fast R-CNN [50]) was hence proposed. In Fast R-CNN, all proposed regions are treated as input and their features are extracted by a single CNN where a pre-trained model can be used (e.g., VGG) to speed up the computation. The output of feature extraction is a custom layer called a region of interest pooling layer (i.e., RoI pooling). RoI pooling layer is followed by fully connected layers, an object classifier, and a bounding box regressor. While the model is significantly faster to train and to make predictions, it still requires a set of candidate regions to be proposed along with each input image.

A faster version is further proposed to improve the burden of region proposals, namely, the Faster R-CNN [106]. The Faster R-CNN introduces a region proposal network (RPN) into the Fast R-CNN as part of the training process. In this case, only one feature map extracted through CNN is established for each input image, and both RPN and the classifier or regressor operate on the same feature map, which greatly improves the efficiency. A summary of the R-CNN family is shown in Fig.4.11, and the Faster R-CNN was used in this research for floor and span detection.

1,452 street-view images were manually labeled through a Python-based labeling tool (`https://github.com/tzutalin/labelImg`). Floor-detection and span-detection models were trained separately, as well as their labeling. An image was labeled with multiple bounding boxes identified as floor or span. Examples of prediction are shown in Fig.4.12.

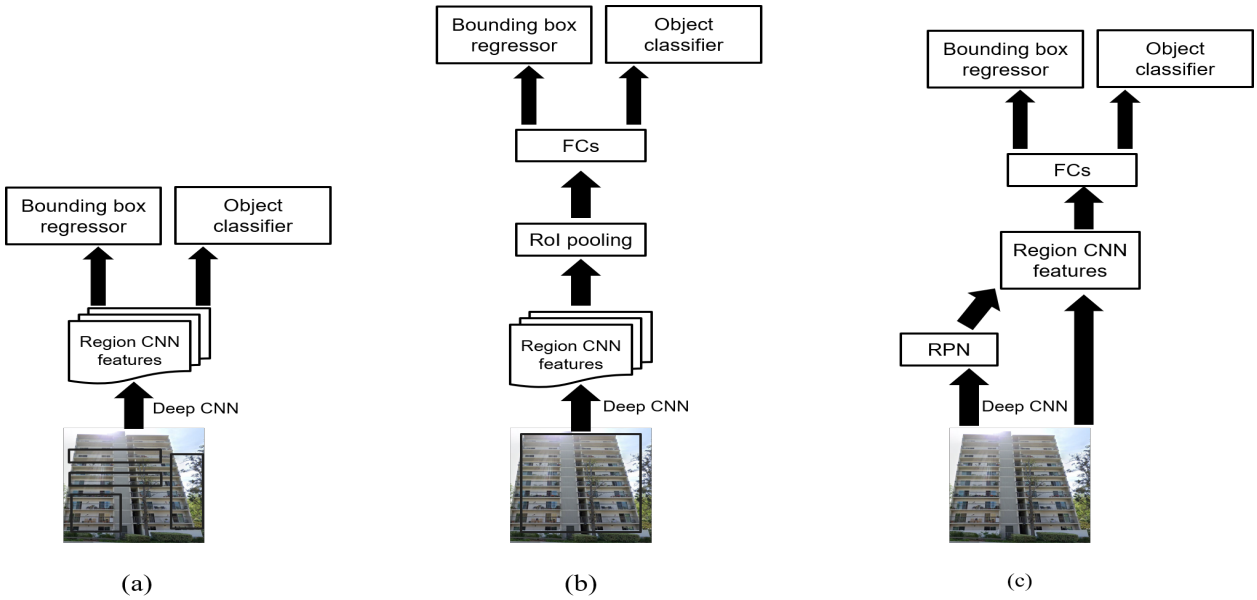


Figure 4.11. Architectures of (a)R-CNN, (b)Fast R-CNN, and (c)Faster R-CNN.

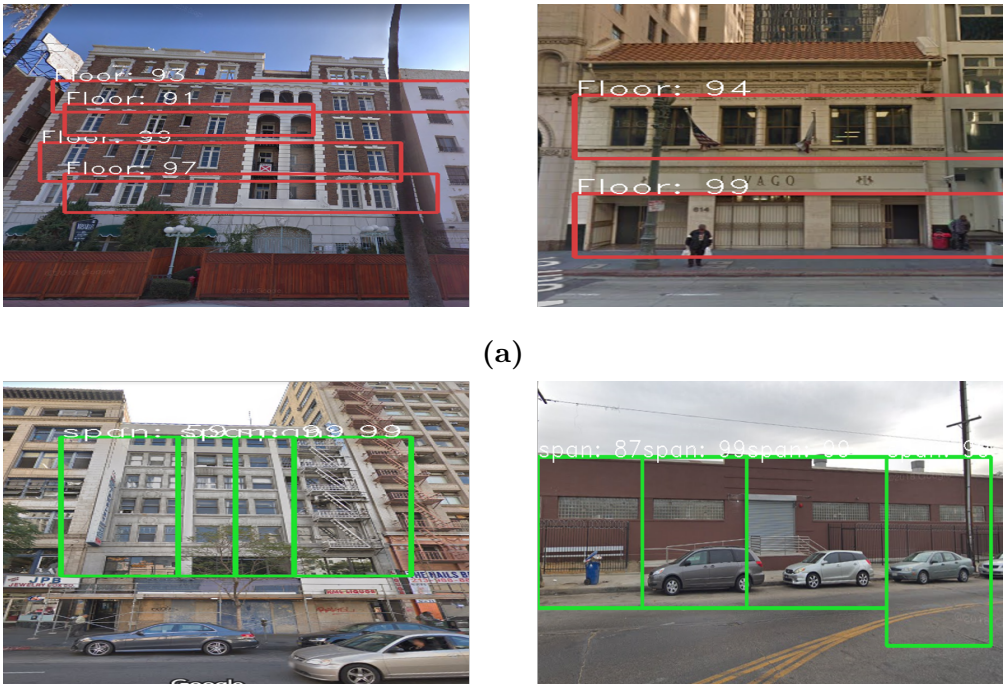


Figure 4.12. Example of predictions: (a)floor, and (b)span.

The performances of the detection tasks are summarized in Table 4.3. The same metrics for the classification task described in Eqs. 2.12 to 2.14 are implemented, and the ground truth used in these equations is the number of floors or spans. The mean value and the coefficient of variation of the performance metrics for each model are computed through the 10-folder cross-validation technique. It must be noted that the ground-truth number of spans was defined by the author, so there is a certain degree of bias that cannot be eliminated. The judgement of a span depends on the location of the columns, which should be apparent for moment-resisting systems (e.g., parking lot). However, if the lateral force-resisting system is a shear wall or contains dual systems, the current ground truth may not be applicable. While the proposed roadmap still has room to be improved, the whole pipeline allows users to provide exact geometric information for automated modeling, which can reduce the uncertainty made by the false prediction from image-based data harvesting.

Table 4.3. Performance of Faster R-CNN for detection tasks.

Task	Floor	Span
Image	street-view	street-view
Ground-truth	L.A. inventory	self judgement
Training	1307	1307
Validation	72	72
Test	73	73
TL baseline	ResNet	VGG
μ_{recall} | CV_{recall}	0.86 | 0.09	0.68 | 0.12
$\mu_{precision}$ | $CV_{precision}$	0.81 | 0.15	0.81 | 0.1
$\mu_{f1-score}$ | $CV_{f1-score}$	0.83 | 0.12	0.74 | 0.11

4.2.4 Summary of the data set

With the prior inventory and the proposed data-harvesting roadmaps, the building information required by the proposed pipeline for assessing NDRCBs is ready. As described in Chapter 3, the input for the automated modeling includes the number of floors, floor height, number of spans, span length, floor area, and occupancy. Histograms of these data are shown in Fig.4.13. It can be noted that the ranges of these variables (i.e., number of stories, floor height, floor area, and span length) are contained in the sensitivity analysis conducted in Chapter 3 for constructing potential archetypal models. In the discovery of the sensitivity analysis, much variance in EDP can be observed for perimeter frames. Hence, the following sections will use the collected data and the proposed perimeter frames to evaluate the seismic losses of intensity associated with the Design Basis Earthquake (DBE) and the Maximum Considered Earthquake (MCE).

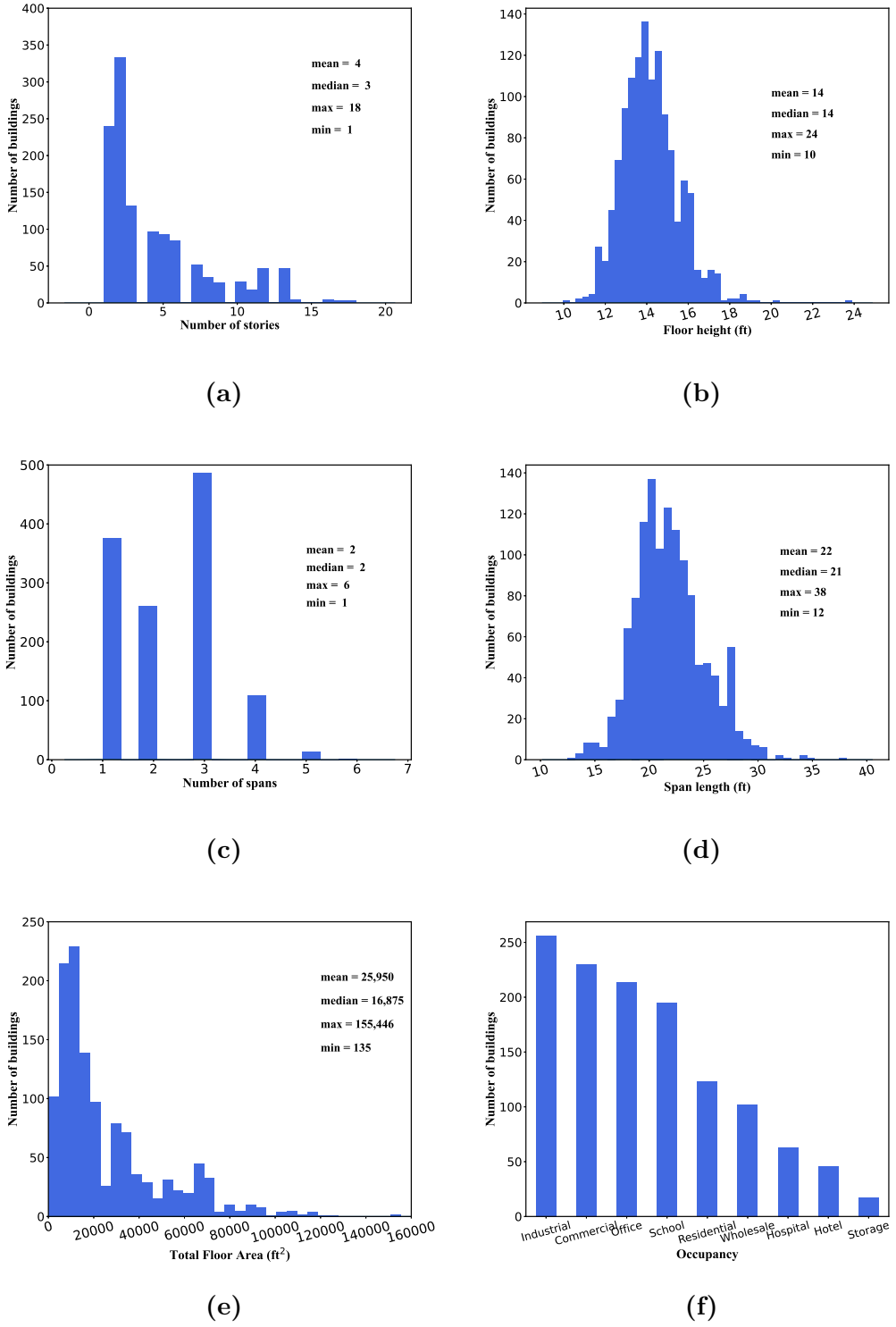


Figure 4.13. Histograms of (a) number of stories, (b) floor height, (c) number of spans (d) span length, (e) total floor area, and (f) occupancy.

4.3 Regional Application

With the data described before, the proposed pipeline carries out the automated modeling and prepares all Tcl/Tk files for running nonlinear time-history analyses. A flowchart shown in Fig.4.14 summaries the data exchange between different files during the regional application.

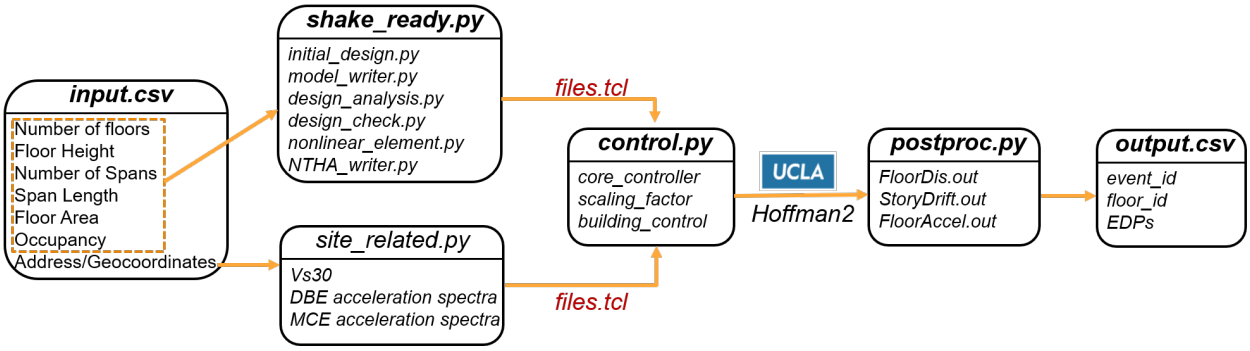


Figure 4.14. Data flowchart for regional application.

The data first goes to the main Python script *shakeready.py* to generate an initial model which will then be constantly updated to satisfy the minimum requirements of the design code (i.e., UBC 1967). With the detailing information from the finalized building, Tcl/Tk files containing elastic elements, nonlinear elements, and recorders will be generated. At the same time, the address or geocoordinates will be used to obtain the site-specific information. For example, the shear-wave velocity at the depth of 30 m V_{s30}, DBE, and MCE spectrums.

V_{s30} is available from the USGS database (`https://earthquake.usgs.gov/data/vs30/us/`) where the spatial distribution of hundreds of data can be used to compute the value for a specific site. For these 1,452 sites, most of their V_{s30} fall in the range of 180-360 m/s and are be defined as class D soil based on NEHRP [33]. In order to estimate site-specific and hazard-consistent loss, uniform-hazard spectrums at the specific site are necessary for scaling ground motions. Multiple websites provide a user interface that can calculate spectral parameters for DBE and MCE levels. Through *selenium*, the automation

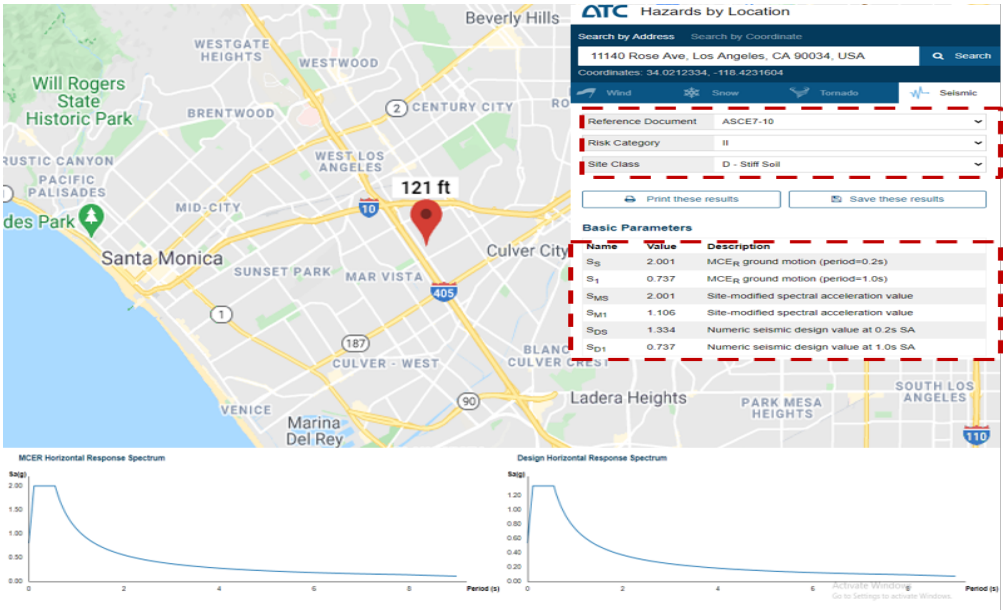
of obtaining site-specific data can then be achieved. For example, on the ATC hazards by location tool (`https://hazards.atcouncil.org/#/`), the address, site class, risk category can be automatically filled in, and the spectral parameters including the MCE spectral acceleration at short period S_s, MCE spectral acceleration at a period of 1 S_1, MCE spectral acceleration at short period adjusted for site class effects S_{MS}, MCE spectral acceleration at a period of 1 adjusted for site class effects S_{M1}, DBE spectral acceleration at short period S_{DS}, and DBE spectral acceleration at a period of 1 S_{D1} can be obtained. Examples of harvesting V_{s30} and spectra parameters are shown in Fig.4.15.

With the spectral parameters, a set of ground motions can be selected and scaled for a specific site to obtain unbiased estimates of median structural response at a given intensity. The selection of ground motion is not the main focus of this study and hence is not introduced into the proposed pipeline. Besides, because most of the NDRCB stocks in the inventory locate close to each other and concentrate in the downtown Los Angeles region, selected ground motions for different sites are not expected to be very different. As a result, the set of 44 (22 pairs) far-field ground motion records specified in FEMA P695 [32] are used. In Chapter 3, four sets provided by Baker et al. [13] were adopted to construct PSDM with scaling. The focus of the previous chapter was to investigate the capability of the proposed pipeline and hence needs to include diverse ground motions. However, this chapter focuses on application. Consequently, those sets are not reused herein considering the practicality and following the conventional assessment method.

In FEMA's set, the magnitude varies from M6.5 to M7.6 with an average of M7.0. Thirty-two (16 pairs) of the ground motions were recorded at sites classified as site class D and the remaining records are from site class C locations, which covers the Los Angeles inventory. The peak ground acceleration for the recordset varies from 0.21 g to 0.82 g with an average of 0.43 g. To estimate the corresponding loss for a given intensity level, scaling of ground motions is performed. For each site, two scaling factors are computed and adopted so that the median spectral acceleration of the record set matches the specified intensity levels (i.e.,

DBE, MCE).

(a)



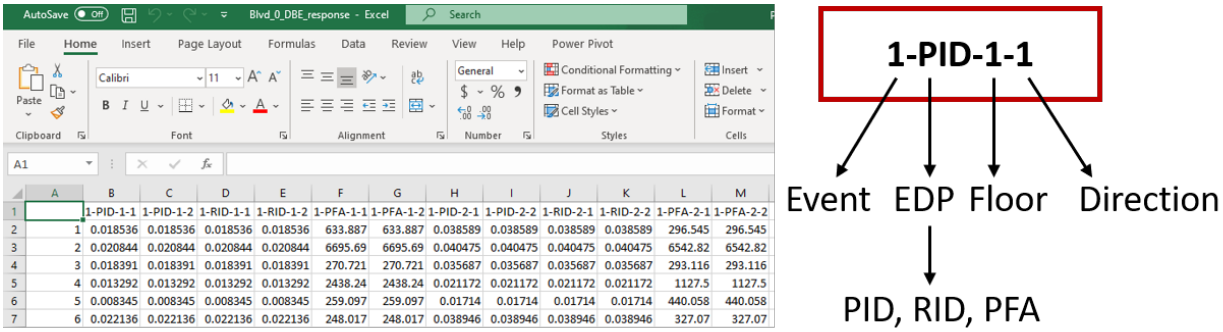
(b)

Figure 4.15. Examples of (a) USGS V_{s30} map, and (b) ATC hazard tool.

To efficiently conduct NTHAs, parallel computation is realized through the high-performance

cluster platform at UCLA, namely, Hoffman2. As shown in Fig.4.14, three indices are introduced to communicate with Hoffman2. The *core controller* connects ground motion records to computing cores in the cluster. In our case, two scaling factors are applied to each record, so 88 records need to be performed for each building, which means 88 computing cores. Considering the limitation of available memory for each user, the original recorded set is accessed by each core and the scaling factor will be assigned based on the index of the core. For example, core 1 corresponds to a building's DBE scaling factor for the first ground motion, and core 2 corresponds to the same building's MCE scaling factor for the first ground motion. For each core, at most 30 buildings' NTHAs are carried out due to the limitation of memory. In summary, 30 buildings' 88 NTHAs (i.e., 44 records, 2 scaling factors) can be conducted during each running, which approximately costs 1 hour.

Once all NTHAs are finished, postprocessing is performed to extract EDPs for loss estimation. The peak inter-story drift at each floor is used to evaluate the damage of structural components and drift-sensitive nonstructural components. The peak floor acceleration is used for acceleration-sensitive nonstructural components, and the residual inter-story drift is used to decide whether a building can be repaired or need to be demolished. An example output of the post-processing is shown in Fig.4.16, which will be the input for loss estimation.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1		1-PID-1-1	1-PID-1-2	1-RID-1-1	1-RID-1-2	1-PFA-1-1	1-PFA-1-2	1-PID-2-1	1-PID-2-2	1-RID-2-1	1-RID-2-2	1-PFA-2-1	1-PFA-2-2
2	1	0.018536	0.018536	0.018536	0.018536	633.887	633.887	0.038589	0.038589	0.038589	0.038589	296.545	296.545
3	2	0.020844	0.020844	0.020844	0.020844	6695.69	6695.69	0.040475	0.040475	0.040475	0.040475	6542.82	6542.82
4	3	0.018391	0.018391	0.018391	0.018391	270.721	270.721	0.035687	0.035687	0.035687	0.035687	293.116	293.116
5	4	0.013292	0.013292	0.013292	0.013292	2438.24	2438.24	0.021172	0.021172	0.021172	0.021172	1127.5	1127.5
6	5	0.008345	0.008345	0.008345	0.008345	259.097	259.097	0.01714	0.01714	0.01714	0.01714	440.058	440.058
7	6	0.022136	0.022136	0.022136	0.022136	248.017	248.017	0.038946	0.038946	0.038946	0.038946	327.07	327.07

Figure 4.16. An example output of the post-processing.

4.4 Seismic Loss Estimation

The seismic loss of infrastructure can be explicitly used as a measurement for seismic risk. It is also a meaningful decision value to stakeholders compared to the engineering demand parameter (EDP) which is meaningful to engineers only. Hence, the last part of the study is to conduct the loss estimation for NDRCBs based on their EDPs associated with DBE and MCE intensities.

Two state-of-the-practice loss assessment methodologies, namely, HAZUS [43] and FEMA P-58 [44] are both implemented herein. For both methods, the loss is accumulated from discrete damage states associated with structural components and nonstructural components. To decide the triggering of damages, fragility curves are used to describe the probability of exceeding damage states given EDPs, which is the main difference between these methods. While HAZUS provides building-level fragility curves, P-58 focuses on component-level fragility functions.

In HAZUS, fragility curves are developed for different damage states of a building's components including structural, non-structural drift-sensitive, and nonstructural acceleration-sensitive components. Four types of damage states including slight, moderate, extensive, and complete are defined based on building types associated with the height (i.e., low-rise, mid-rise, high-rise), design category (i.e., pre-code, low-code, moderate-code, high-code), material types (i.e., concrete, steel, wood), and model types (e.g., pure moment frame, moment frame with shear walls, moment frame with unreinforced masonry infilled walls). For example, fragility curves shown in Fig.3.12 describe the probability of four different damage states for a pre-code concrete moment frame (C1).

The building loss in HAZUS is directly transformed from building damage. HAZUS provides default repair/replacement costs per square foot for buildings based on their occupancy (e.g., single-family dwelling, hospital, parking, education-school), which can be used to estimate the total cost with total area. With the total cost, the relationship between damage

states and repair/replacement cost for different component types can be defined. Examples are shown in Tables 4.4 and 4.5, and complete tables can be found in Tables 11.1 to 11.4 in HAZUS [43]. In view of the rapid evaluation of regional seismic, HAZUS' methodology is feasible because the damage is associated with building-level responses and the regional loss is only the accumulation of all buildings. However, the calculation is too conservative because it does not consider the difference of fragility curves for different components (i.e., beam, column, joint, wall) even their performance is governed by story drift. The computation of cost based on only four damage states may underestimate or overestimate the total cost. For instance, [30] shows that HAZUS predicts a lower loss at lower intensities and a higher loss at higher intensities.

Table 4.4. Example of replacement cost per square foot.

Occupancy code	Descriptions	Means cost 2002
RES1	Single Family Dwelling	$62
COM1	Retail Trade	$59
IND1	Heavy	$68
EDU1	Schools/Libraries	$86

Table 4.5. Example loss rates for single-family residences of light-frame wood construction.

Damage state	Structural components	Nonstructural drift-sensitive	Nonstructural accel.-sensitive	Total building	Contents	Buildings with contents
Slight	0.38	0.8	0.43	1.6	0.4	2.0
Moderate	1.88	2.0	2.13	8	2.0	10
Extensive	9.38	20	10.63	40	10	50
Complete	18.75	40	21.25	80	20	100

The other loss estimation method that has been widely used in recent years is the FEMA P-58, which was developed by the Applied Technology Council (ATC) for the next-generation

performance-based seismic design methodology [3]. The essential difference between the FEMA P-58 and HAZUS is its probabilistic nature. Instead of binning a structure into a loss group, FEMA P-58 simulates ground motions and compiles building losses using Monte-Carlo simulations. FEMA P-58 uses probabilistic seismic hazard analysis at the site of interest to obtain the mean annual frequency of exceedance of specific ground motion intensity levels (or hazard curves). Nonlinear structural responses (i.e., EDPs) to hazard-specific excitations are then estimated through NTHAs. Component-level fragility curves are provided in P-58 to connect the EDPs to physical damage, and statistical loss functions are used to describe the relationship between physical damage and repair or replacement costs. An intensity-based loss estimation procedure is shown in Fig.4.17.

For a given intensity level (e.g., DBE, MCE), NTHAs are conducted to obtain building responses and generate the multivariate log-normal distribution of EDPs (e.g., peak inter-story drift, residual inter-story drift, peak floor acceleration). Through Monte-Carlo simulations, random realizations of EDPs are used to first check if the drift limit of collapse is reached. If the building collapses or the residual drift is too large and the building cannot be repaired, the replacement cost of this realization is counted. Otherwise, EDPs are used to sample damage states of components based on the provided fragility curves. The costs of repairing or replacing components are then conditioned on the damage states. The cost of non-collapse building is taken as the sum of costs for repairing individual components in all directions and stories, and the loss conditioned on a single intensity is taken as the mean of all Monte Carlo realizations.

FEMA P-58 provides detailed damage and loss models calibrated using data from laboratory tests of components, historical data on repairs, and estimates from contractors. Fragility parameters (i.e., median, dispersion) of damage states, repair cost, and repair time parameters (i.e., min, average, max) of each damage state are available in the supporting materials of P-58 [44]. An example of component damageability and cost data is shown in Table 4.6.

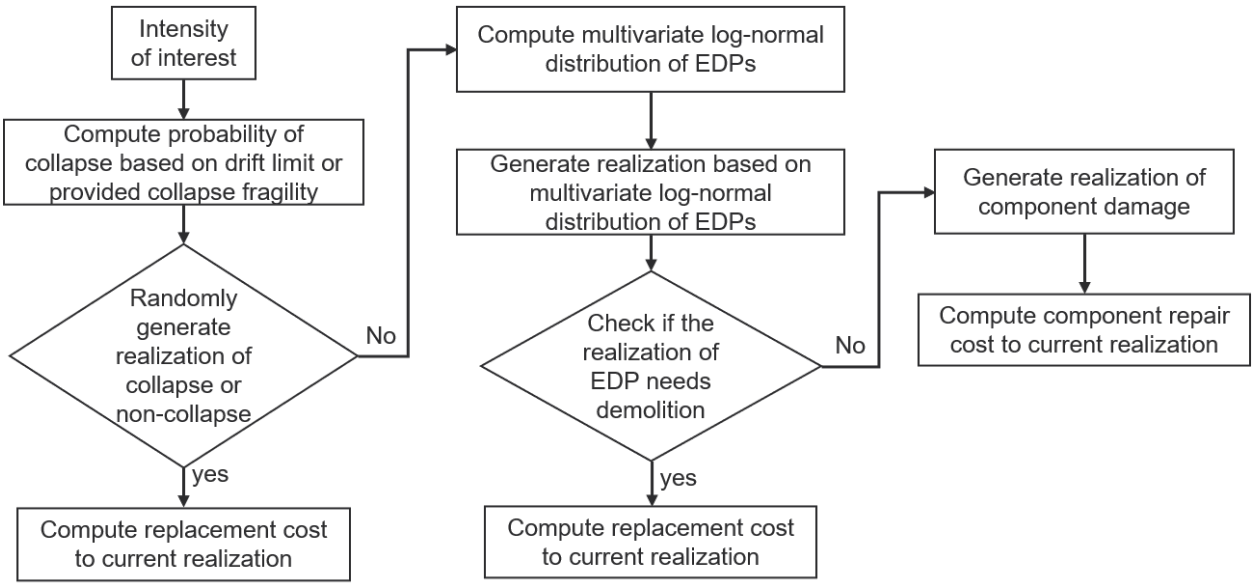


Figure 4.17. FEMA P-58 intensity-based loss estimation procedure.

Table 4.6. An example of component damageability and cost data.

Component	Description	DS: Median	Unit Repair Cost Range
B1041.041a	ACI 318 OMF with weak joints and column flexural response, Conc Col & Bm =24" x 24", Bean one side	DS1 (IDR): 0.015 DS2 (IDR): 0.0175 DS3 (IDR): 0.02	$\$23,131 - \$34,697$ $\$35,077 - \$52,615$ $\$43,176 - \$64,764$
B1041.041b	ACI 318 OMF with weak joints and column flexural response, Conc Col & Bm =24" x 24", Beam both sides	DS1 (IDR): 0.015 DS2 (IDR): 0.0175 DS3 (IDR): 0.02	$\$23,131 - \$34,679$ $\$35,077 - \$52,615$ $\$43,176 - \$64,764$
B1049.001a	Reinforced concrete flat slabs-columns without shear reinforcing 0<Vg/Vo<.2, no continuity reinf	DS1 (IDR): 0.025 DS2 (IDR): 0.04	$\$41,570 - \$61,132$ $\$56,841 - \$83,590$

4.4.1 NHERI-SimCenter *Pelicun*

Both HAZUS and P-58 has been implemented in software tools, such as PACT (Performance Assessment Calculation Tool [45]) and SP3 (Seismic Performance Prediction Program [54]), which provide engineers the ability to evaluate the seismic performance of design solutions for new buildings and retrofit of existing structures. Although these software have comprehensive functions, they are not open-source tools, which makes it difficult to be integrated into a single framework or platform for a specific research purpose.

To address this problem, the Computational Modeling and Simulation Center (SimCenter, `https://simcenter.designsafe-ci.org/`) under the Natural Hazards Engineering Research Infrastructure (NHERI) program funded by the National Science Foundation (NSF) recently developed an open-source Python-based tool for probabilistic estimation of losses, injury and community resilience under natural disasters, namely, *pelicun* [137].

Pelicun provides multiple modules for users and allows them to modify them for customized data. As shown in Fig.4.18, users can define specific event files for the interest of sites. BIM files including building information (i.e., number of floors, floor area, built year, occupancy) are also provided by users, which can be used to estimate building responses and define the performance group (e.g., RC columns on the first floor in the first direction). For the response estimation, users can choose their preference and desired accuracy to prepare the EDP files. Examples shown in Fig.4.16 are the EDPs of the Los Angeles' inventory and they already followed the format requirements of *pelicun*. Based on the raw EDP files, *pelicun* can compute the parameters for multivariate log-normal distribution and the collapse probability, which will then be used for Monte-Carlo realizations. Currently, the damage and loss data shown before for both methods are provided with *pelicun*, and users only need to define a configuration file that includes component lists to associate with the database for loss estimation. An example configuration following Table 4.6 is shown in Table 4.7.

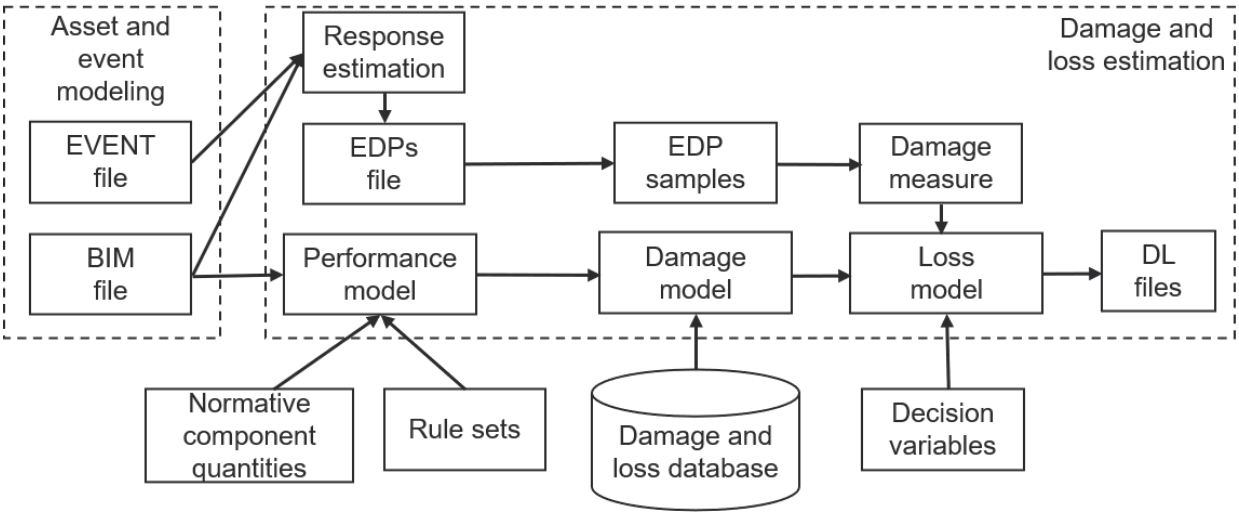


Figure 4.18. The workflow of *pelicun*.

Table 4.7. Configuration of components for *pelicun*.

Component	Direction	Levels	Total Units	Qty. Per Unit	Total Qty. in Building
B1041.041a	1	1-2	8	1 Each	8 Each
	2	1-2	8		8 Each
B1041.041b	1	1-2	12	1 Each	12 Each
	2	1-2	12		12 Each
B1049.001a	ND	1-2	18	1 Each	18 Each

With the configuration file, *pelicun* disaggregates a building into three hierarchies to estimate the loss of each realization. The highest level of the hierarchy is the *Fragility Groups* (FGs), where each FG collects components with similar fragility and are controlled by a specific type of EDP. This FG can then be broken down into the second hierarchy, namely, *Performance Groups* (PGs). In each PG, all components are governed by the same EDP in the same direction and same location. The last hierarchy is the *Component Groups* (CGs), where each CG has components with the same damage states during a realization. The number of components in each CG is then used to compute the corresponding consequences such as repair time and cost.

4.4.2 Regional estimation

With EDPs computed from NTHAs and *pelicun*, the rest work of this study is to investigate the potential losses of these NDRCBs under the DBE-level and MCE-level intensities, their spatial distribution, and the compositions of losses. In *pelicun*, the loss is computed based on the expected replacement value of the building. To estimate the losses that can reflect current construction prices for the target region, SP3 [54] software is utilized because it considers the latest RSMeans cost data. Examples of expected replacement values are shown in Table 4.8, which is estimated with consideration of original building cost and an additional 25% cost of debris removal.

For the configuration file required by *pelicun*, the level of detail depends on the loss methodology used. For example, if HAZUS is implemented, the configuration only needs to provide the location (i.e., floor), the direction (i.e., 1, 2), the unit (i.e., each, ft, ft^2), the component type (i.e., structural component, non-structural component sensitive to acceleration, non-structural component sensitive to drift). and the number of components. On the other hand, the detailed ID of components is required when using the P-58 method, as shown in Table 4.7. SP3 is also used to assist in listing components for all buildings including different heights and occupancy. The Complete configuration lists used for *pelicun* can be found in Appendix B.

Through *pelicun*, losses of thousands of buildings can be computed within an acceptable time (e.g., 1 hour). The spatial distribution of losses is shown in FIg.4.19. The loss ratio is the expected loss of 2,000 random realizations divided by the expected replacement cost. Detailed summaries of the losses are shown in Table.4.9 and Table.4.10, where HAZUS estimates the losses more conservatively. This can be expected because HAZUS only considers building-level responses and does not take component-level fragility curves into account. However, P-58 does consider it story by story, and the components in each story may have different damage states.

Similarly, the distribution of expected time for repair or reconstruction is shown in Fig.4.20, where HAZUS also has a conservative estimation. The repair or reconstruction for each building is assumed to be conducted simultaneously, which is not realistic when a huge event occurs and multiple buildings are damaged in a region. The repair time can be used to assess the building closure time and corresponding indirect losses, yet it is not the focus of this study and hence is not discussed herein.

Table 4.8. Expected replacement value of different types.

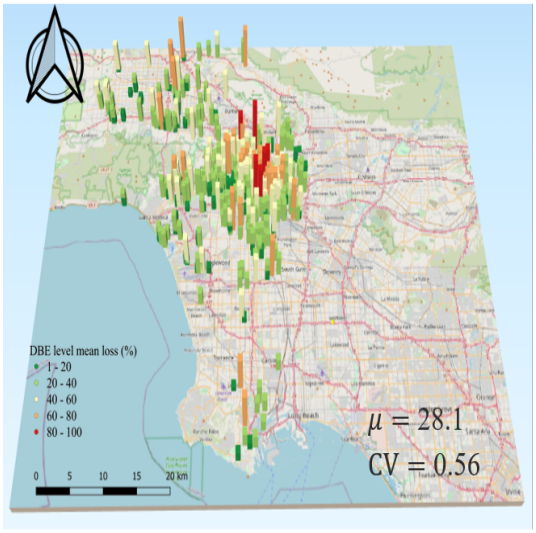
million	Low-rise (2-story)	Mid-rise (5-story)	High-rise (12-story)
Industrial	\$3	\$7	\$15
Commercial-office	\$5	\$13	\$41
School	\$6	\$14	\$36

Table 4.9. Mean values of losses obtained through P-58.

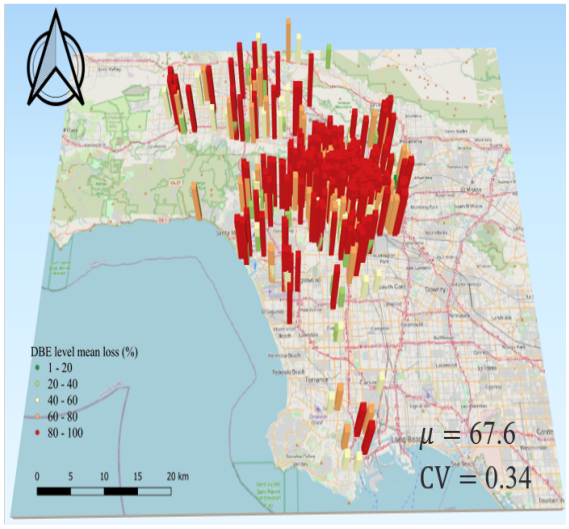
	Low-rise		Mid-rise		High-rise	
	DBE	MCE	DBE	MCE	DBE	MCE
cost ratio (%)	27	40	28	39	29	43
repair time (days)	186	214	199	217	203	223

Table 4.10. Mean values of losses obtained through HAZUS.

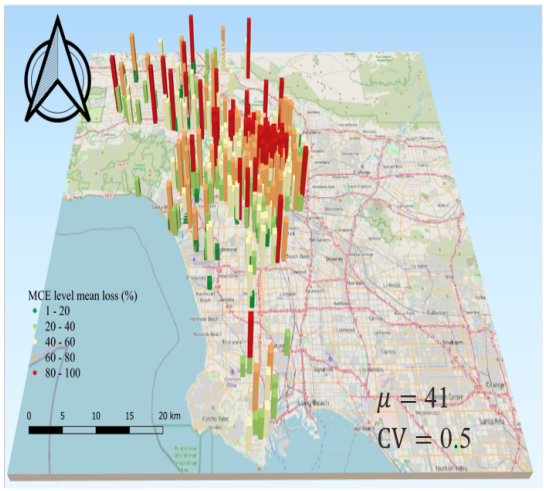
	Low-rise		Mid-rise		High-rise	
	DBE	MCE	DBE	MCE	DBE	MCE
cost ratio(%)	70	74	70	78	76	83
repair time (days)	317	337	318	340	322	344



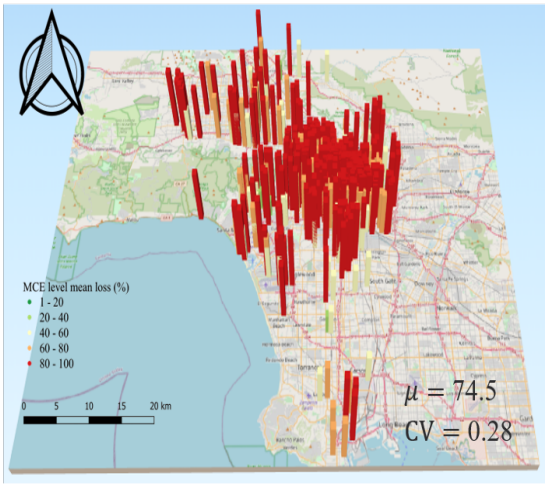
(a) DBE-level via P-58



(b) DBE-level via HAZUS

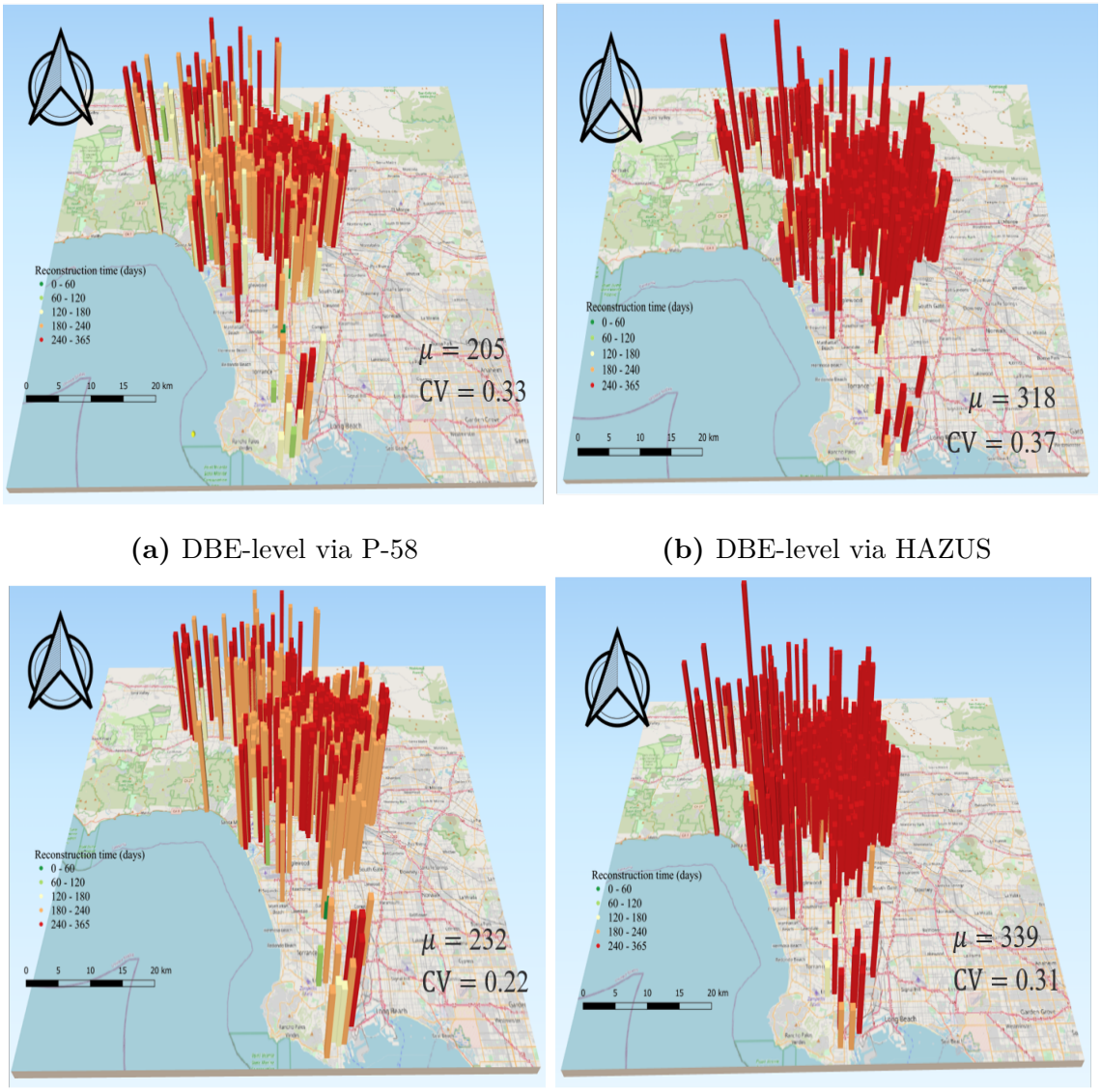


(c) MCE-level via P-58



(d) MCE-level via HAZUS

Figure 4.19. Spatial distribution of losses estimated through different methods at different intensities.



(a) DBE-level via P-58

(b) DBE-level via HAZUS

(c) MCE-level via P-58

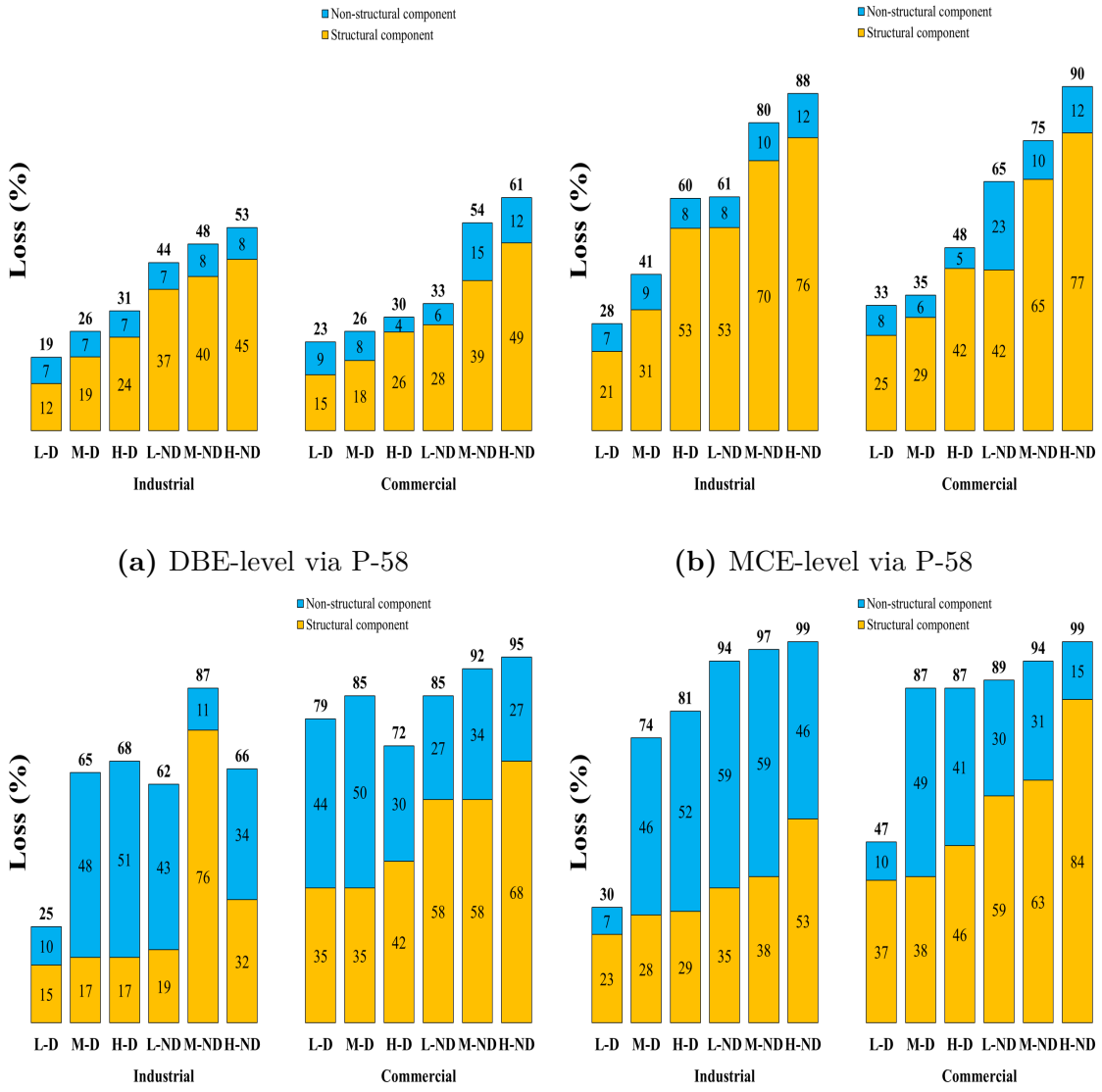
(d) MCE-level via HAZUS

Figure 4.20. Spatial distribution of repair/reconstruction time estimated through different methods at different intensities.

To better understand the compositions of losses, Fig.4.21 shows the disassemble of losses for buildings with a different height (i.e., L: low-rise, M: mid-rise, H: high-rise), different types (i.e., D: ductile, ND: non-ductile), and different occupancy (i.e., industrial, commercial).

Only industrial and commercial buildings are discussed here because they account for more than 50% in the original inventory. The compositions show that HAZUS estimates the losses of non-structural components on average 4 times than the P-58 method. In terms of the total loss, HAZUS is about 1.5 times of P-58. For comparison purposes, the losses of several ductile reinforced concrete moment frames are also listed here. The models are similar to non-ductile buildings but without consideration of shear-limit elements. The losses of structural components of ductile buildings are around 1.5 times less than non-ductile buildings, which results in less total loss ratio.

The regional losses estimated in this study is a summation of all expected losses of individual buildings given DBE and MCE intensities. However, it is not realistic because not all the buildings will reach DBE or MCE intensity at a given event. That is to say, is a conservative estimation for a region. In fact, the intensity differs by the site during a scenario event. Many studies have shown that uncertainties in ground motion associated with different sites and different earthquake events could effect the distribution of aggregate regional losses [90, 99, 124]. The spatial correlation exhibited in ground motions is hence necessary to be considered as a potential future study.



(a) DBE-level via P-58

(b) MCE-level via P-58

(c) DBE-level via HAZUS

(d) MCE-level via HAZUS

Figure 4.21. Disassemble of losses.

4.5 Summary

In this chapter, a data-harvesting module was proposed based on existing inventory data and collected street-view/aerial images of NDRCBs, which was introduced to assist the

numerical modeling. The auto-modeling module discussed in Chapter 3 was applied to 1,452 NDRCB stocks in the City of Los Angeles to conduct NTHAs and compute responses of these buildings subjected to earthquake ground motions. EDPs at DBE-level and MCE-level intensities were obtained and used to estimate the seismic losses through *pelicun*, a recently developed open-source tool implementing HAZUS and FEMA P-58.

Web scraping was adopted to harvest site- and structural- specific information through IoT, and deep CNNs were trained to identify geometric information that is missing in the current inventory but is vital for numerical modeling. The web scraping only focused on known data sources (e.g., LADBS, ZIMAS, Zillow) and specific information (i.e., built year, floor area, building height), yet it can be further applied to any web sources. Deep CNNs for detecting the number of floors/spans and estimating span-length/floor-height were trained through the TL technique mentioned in Chapter 2. The cross-validation technique was implemented to identify the goodness of hyper-parameters and the uncertainty of prediction. Although the harvested images focus on the region of interest and the CNNs are only for specific tasks (i.e., detection, regression), the idea of using artificial intelligence to reduce labor-intensive work for collecting data and helping risk mitigation is feasible.

The estimation of regional losses for NDRCBs was practiced as a demonstration of the outcomes of this study. With a high-performance computational cluster, hundreds of thousands NTHAs were carried out and the time reduced from weeks to days. State-of-the-practice loss assessment methodologies help to depict the spatial distribution of losses at a city scale and identify the composition of them for different types of buildings, which highlights the vulnerability of NDRCB stocks.

CHAPTER 5

Conclusions and Future Work

5.1 Conclusions of Chapters

This dissertation proposed a data-driven framework for regional assessment of seismically vulnerable buildings. Two pipelines using data-driven artificial intelligence methods are included in the framework for soft-story buildings and non-ductile reinforced concrete frames, respectively. While detailed summaries can be found in each chapter, brief conclusions are described below.

The first pipeline focuses on identifying soft-story buildings at a city scale and is introduced in Chapter 2. The goal is to improve the accuracy of existing studies using street-view image-based deep-learning models which highly rely on well-controlled images with clear target objects. The pipeline used photogrammetry to construct city-scale point clouds and used building footprints to segment building points. A heuristic algorithm was developed to collect density features of point clouds along with the building height, which can conquer the issue that open-space characteristics of soft-story buildings are often being blocked in 2D images. The density feature can depict the opening first floor or parking garage and hence improve the accuracy of DL models. The transfer learning technique was introduced to overcome the over-fitting and sets of hyper-parameters were investigated for optimizing convolutional neural networks. The work did not carry out the soft-story modeling and seismic analysis because many studies have already done that and many of these buildings have been retrofitted or are being retrofitted.

The second pipeline discussed in Chapter 3 aims to fulfill an automated assessment for NDRCB frames, which includes automated modeling and damage assessment. Because generating a large number of models is time-consuming and requires many implementation details, the pipeline developed an automated design procedure using minimal building information (i.e., number of floors, floor height, number of spans, span length, floor area) to construct archetypal models satisfying era-specific building codes as an approximation of as-built buildings. The archetypal frames were validated with existing NDRCBs through nonlinear static and nonlinear dynamic analyses, which showed that the proposed OpenSees model can capture the stiffness degradation and shear failure in columns that are often observed in NDRCBs.

The probabilistic seismic demand model (PSDM) was adopted in the pipeline to develop damage fragility curves because it has a closed-form formulation and does not require any scaling of earthquake ground motions. Based on PSDM, regression analyses were carried out for log-normal coefficients used to describe the probability of exceedance of a damage state given an intensity measurement (IM). Suitable IMs were also investigated for the maximum inter-story drift and peak floor acceleration based on efficiency, sufficiency, proficiency, and practicality. PGV and PGA are the most suitable IMs for maximum inter-story drift and peak floor acceleration respectively. A sensitivity analysis was conducted to identify the variation of engineering demand parameters (EDPs) given different structural information, and the result revealed that the structural type had more effect on the EDPs than user-defined building information. For example, the high-rise building with a perimeter frame as the lateral force-resistant system had a larger variance in the maximum inter-story drift.

Chapter 4 introduced a data-harvesting module to assist a real-world application of the proposed pipeline to regional assessment and loss estimation for NDRCBs. Web-scraping techniques and deep CNNs were adopted to collect site- and structural-specific information for the auto-modeling. For existing inventory data, the module automated the data-mining procedure in each data source and can be further adapted to other web sources for studies of

interest. On the other hand, deep CNNs were trained for investigating missing information that is vital for modeling such as number of floors, number of spans, and span length. TL technique and image augmentation were used to enhance the performance of CNN models on detection and regression tasks. The detection task was trained for identifying the number of objects (i.e., floor, span), while the regression task was trained for predicting values (i.e., height, length, area). Optimal CNNs and the uncertainties in their predictions for both tasks were obtained through cross-validation.

With the harvested metadata, 1,452 NDRCBs in the city of Los Angeles were assessed through the proposed pipeline. Hundreds of thousands NTHAs were conducted to obtain the responses of archetypal frames subjected to earthquake ground motions. The EDPs were used to compute log-normal distributions for random realizations in the Monte-Carlo simulation, which were then used to evaluate the damage state and the corresponding repair time and cost. Regional losses and their composition highlighted the vulnerability of NDRCBs.

5.2 Research Impacts and Limitations

This research has some potential impacts and also some limitations for using it.

Potential impacts include:

- The data-driven framework is applicable to reduce the burden of labor-intensive large-scale inventories for natural hazard mitigation in urban regions.
- The framework integrates existing tools from multidisciplinary fields, which allows users with different backgrounds to adopt it based on their level of interest.
- The spatial distribution of seismically vulnerable buildings is informative for developing pre-event planning and post-event countermeasures. For example, the arrangement of emergency rescue routes.
- While image-based methods have been widely used for a city-scale risk assessment, 3D data such as point clouds have not. The geocoordinates, texture, and depth information

of points can be used to enhance the level of detailing in modeling digital cities.

- The auto-modeling and design procedures are currently applicable to pre-1980 reinforced concrete buildings but can be extended to any era and any type of structure.

Limitations of the proposed framework are:

- The pipeline for identifying soft-story buildings relies on the quality of point cloud data which are not as common as street-view images. That is, to get promising results, an advanced tool may be required to obtain refined point clouds, which may increase the computational cost.
- The data-harvesting module focuses on non-ductile reinforced concrete buildings in Los Angeles. For other regions, the available data sources will be different and it may still need some labor work to find suitable sources. Similarly, images used to train deep CNNs may also need to be re-collected and labeled for other regions because the appearance of buildings may be different even if they are all considered as non-ductile buildings.
- Current archetypal models only have moment-resistant frames, but shear-wall buildings and moment frames with masonry infilled walls also have a high percentage in the inventory. These types of buildings must be introduced into the proposed pipeline to accurately estimate the regional risk.
- Similarly, only 2D archetypal frames are considered in this study. 3D models with consideration of irregularity are also necessary to be included because some undesired behavior of NDRCBs may be worse when the irregularity exists.
- The selection of site-specific ground motions is not the focus of this study, so the current pipeline isn't a fully end-to-end (i.e., site-specific hazard to loss) method yet. This module is inevitable for a more precise assessment.

5.3 Recommendation of Future Work

With the aforementioned limitations, the potential future work of this study includes:

- The point cloud data has been well labeled and can be applied to recently developed 3D CNN models such as PointNet [102] and PointCNN [82]. These models can directly train neural networks on raw point clouds without any transformation, which can further improve the efficiency of the proposed pipeline.
- Current models are simplified. More types of NDRCBs and extra complexity need to be considered. In addition to the aforementioned 3D analysis and irregularity, detailed elements such as fiber elements can be introduced to simulate the bidirectional interaction of axial loads and bending moments.
- The effect of vertical seismic excitation on the shear capacity of shear-critical columns should be considered. Soil nonlinearity and interaction with superstructures can also be included in the pipeline.
- The spatial cross-correlation in ground motions should be considered and scenario-based loss assessment should be discussed in order to truly respond the distribution of aggregate regional losses.
- The image-harvesting still needs some manual filtering to obtain clear side-view images for training deep CNNs, which should be improved with the help of automated positioning and selection of imaging parameters.
- A user-friendly interface should be developed to incorporate all the scripts and modules developed in this study so that the benefits of the framework can be maximized.

Appendix A

Regression Analyses of PSDM

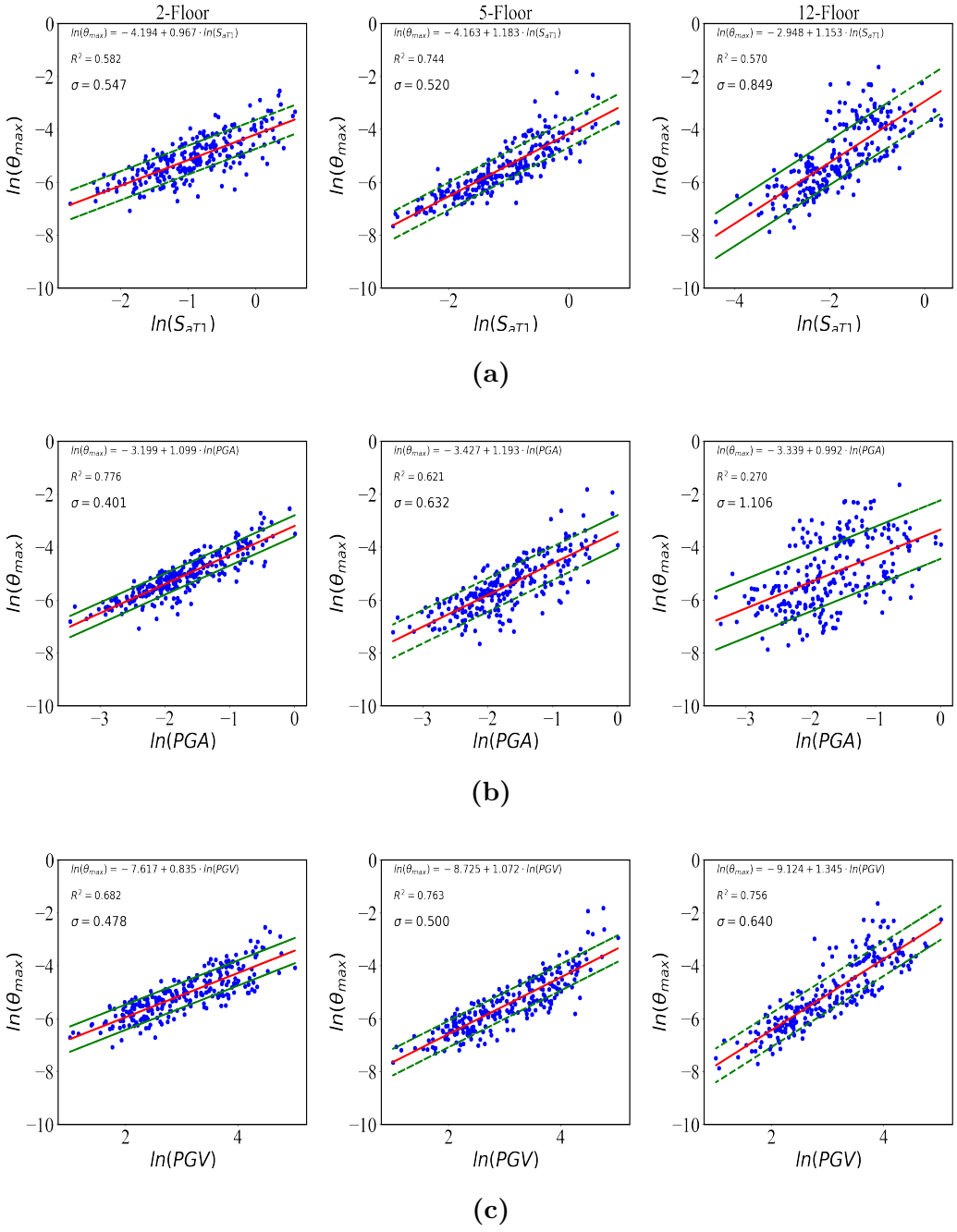


Figure A.1. Regression analyses of perimeter frames for θ_{max} v.s. (a) S_{aT1}, (b) PGA, and (c) PGV.

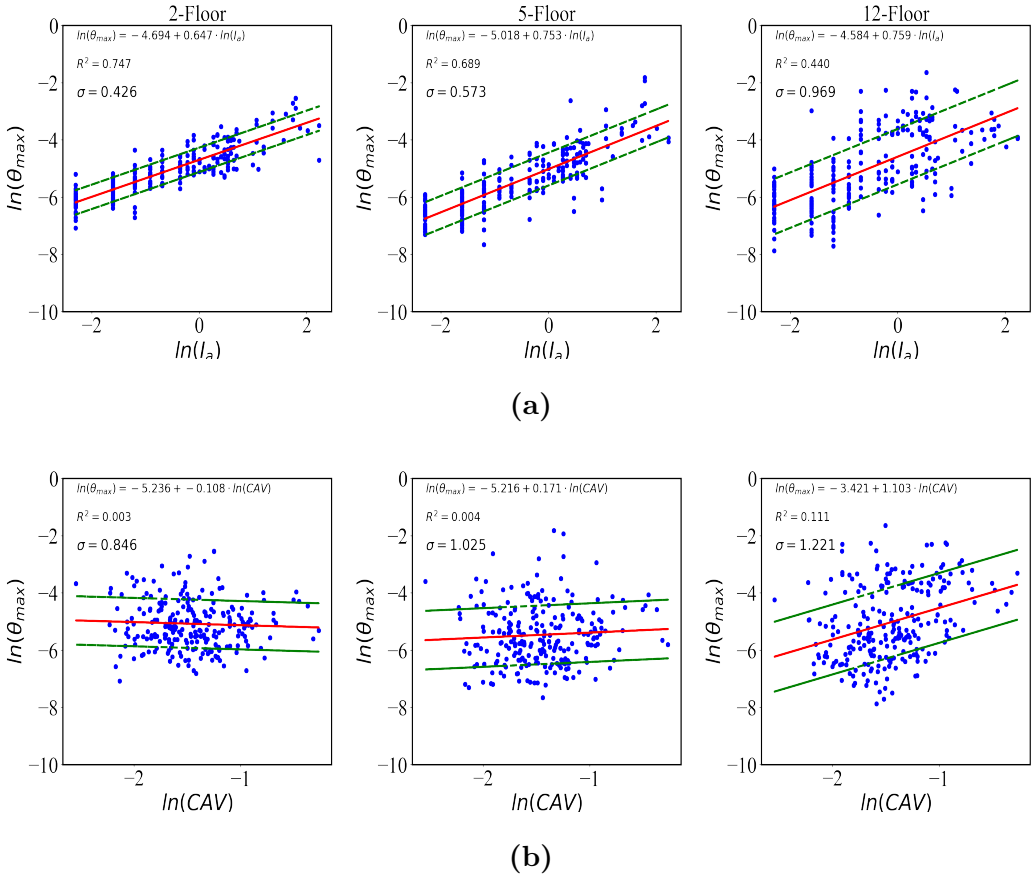
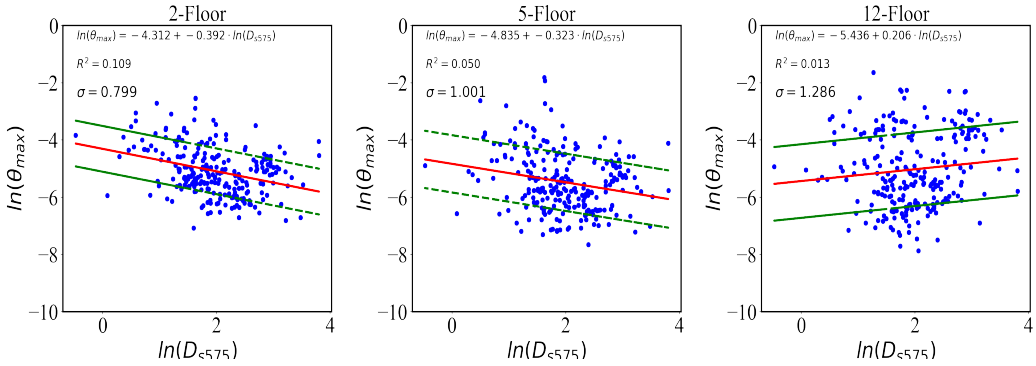
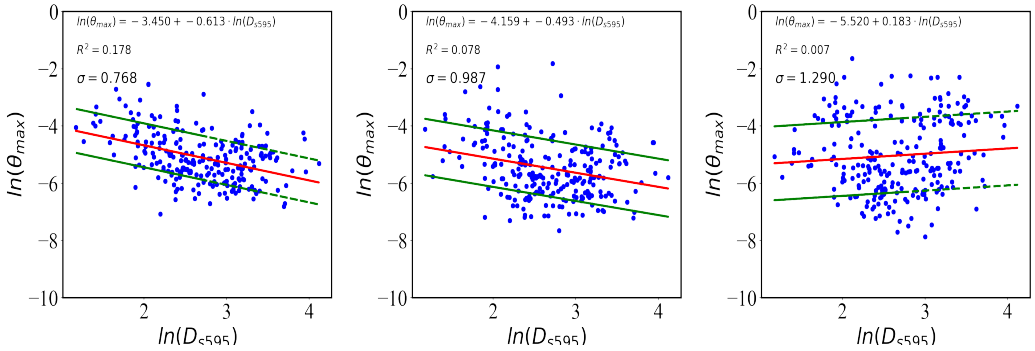


Figure A.2. Regression analyses of perimeter frames for θ_{max} v.s. (a) I_a, and (b) CAV.

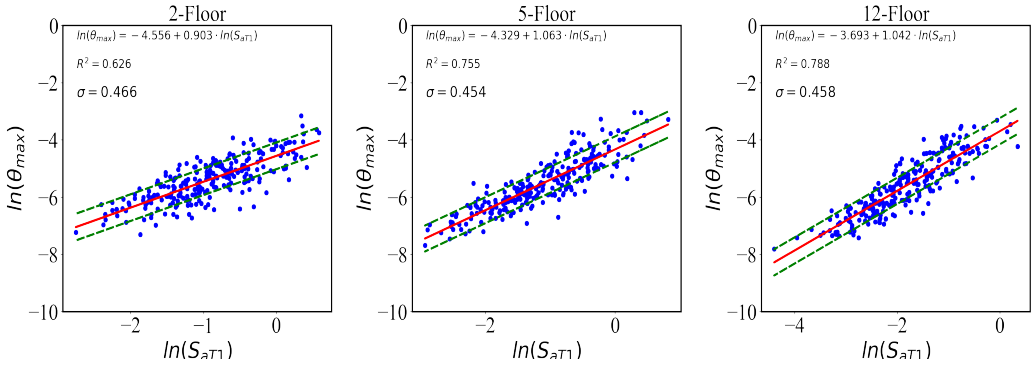


(a)

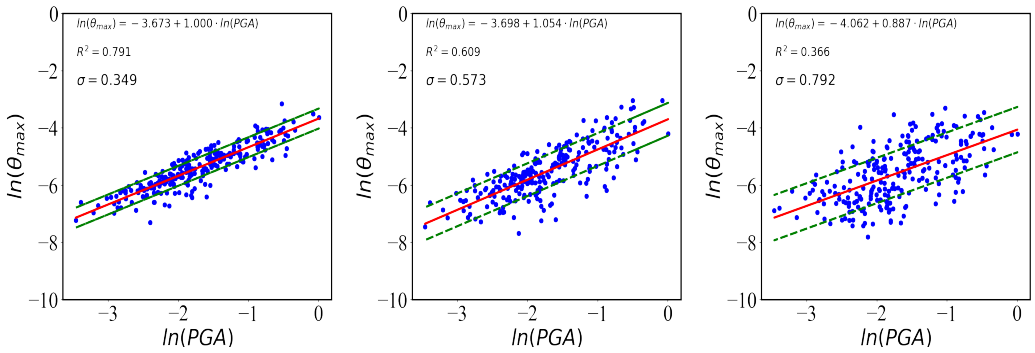


(b)

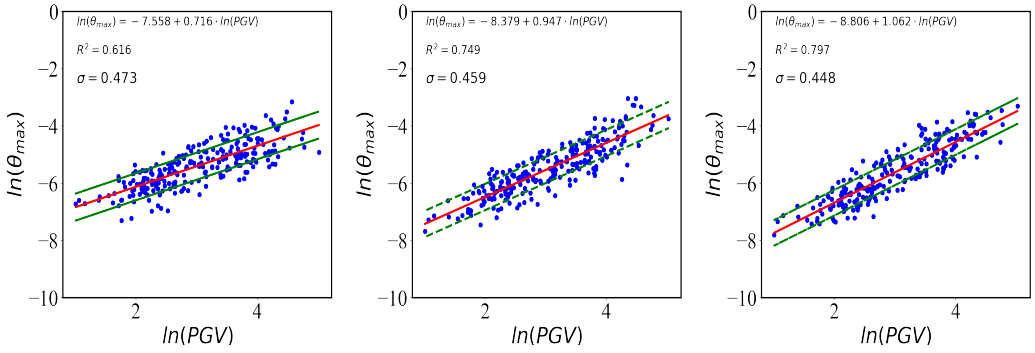
Figure A.3. Regression analyses of perimeter frames for θ_{max} v.s. (a) D_{s575}, and (b) D_{s595}.



(a)



(b)



(c)

Figure A.4. Regression analyses of space frames for θ_{max} v.s. (a) S_{aT1}, (b) PGA, and (c) PGV.

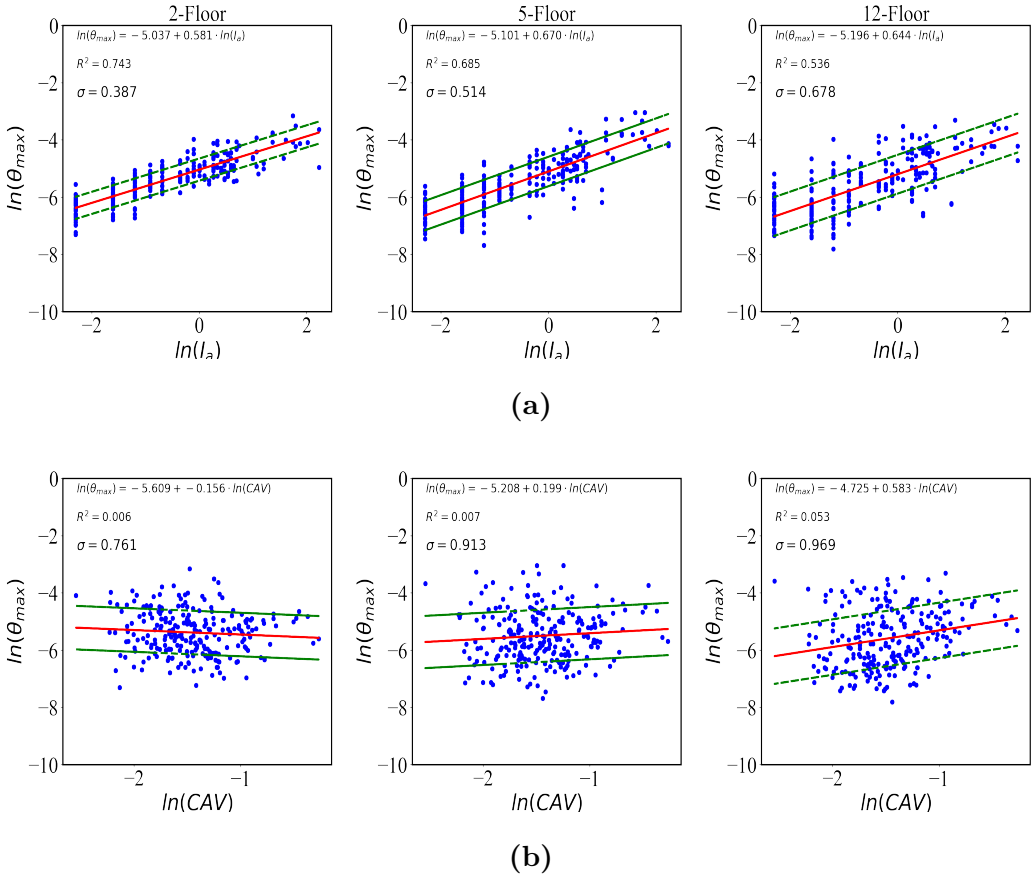


Figure A.5. Regression analyses of space frames for θ_{max} v.s. (a) I_a, and (b) CAV.

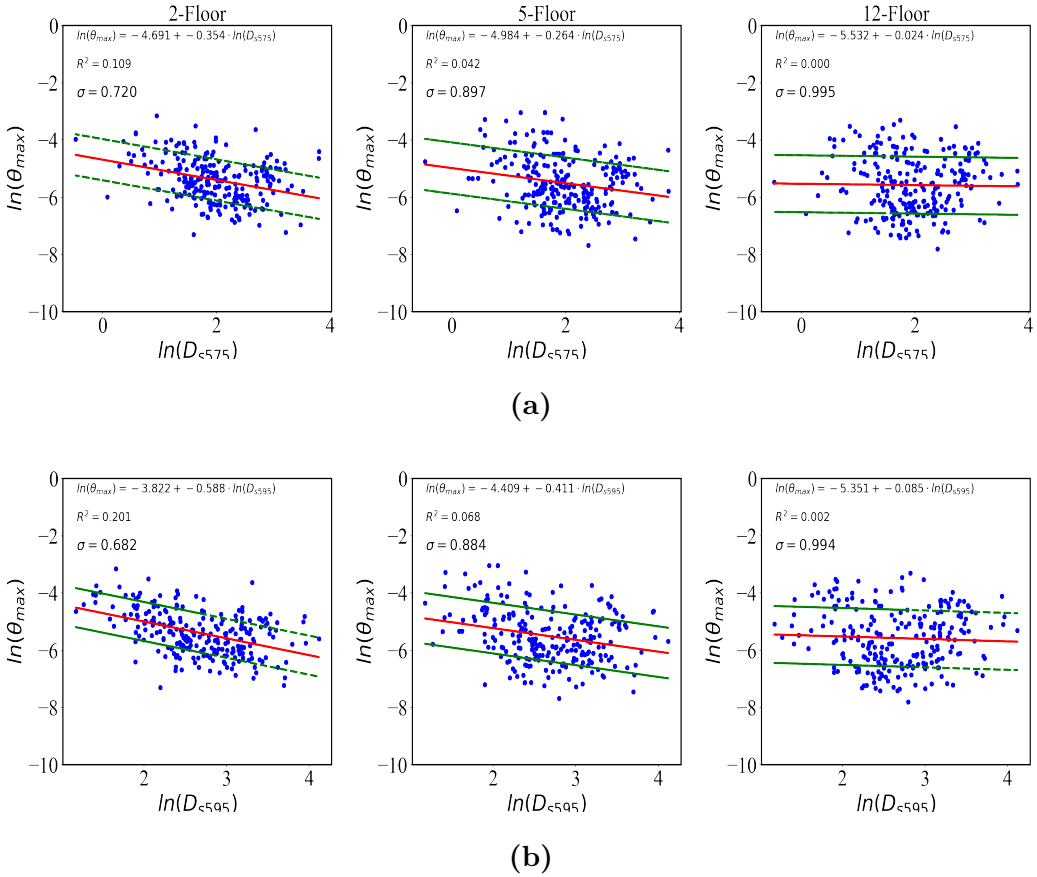
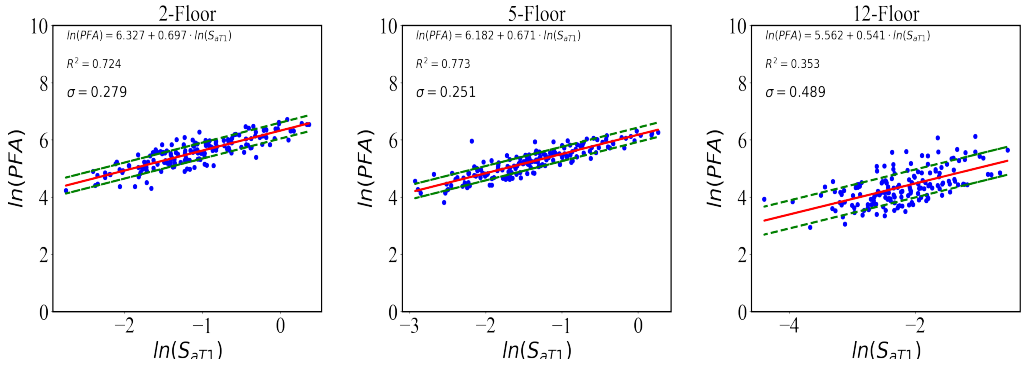
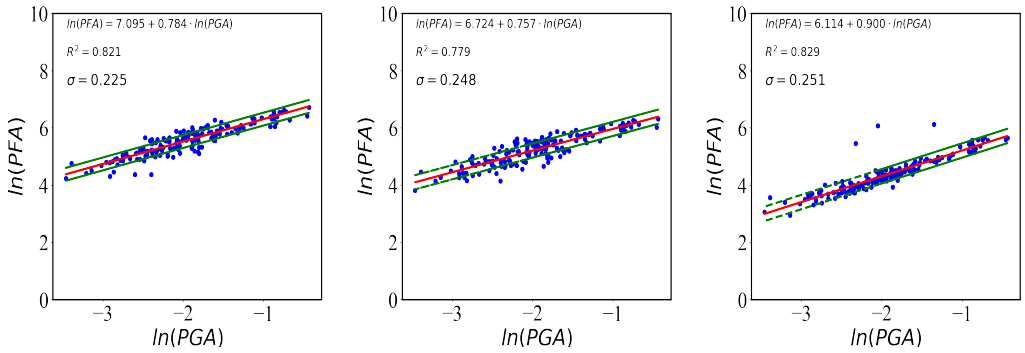


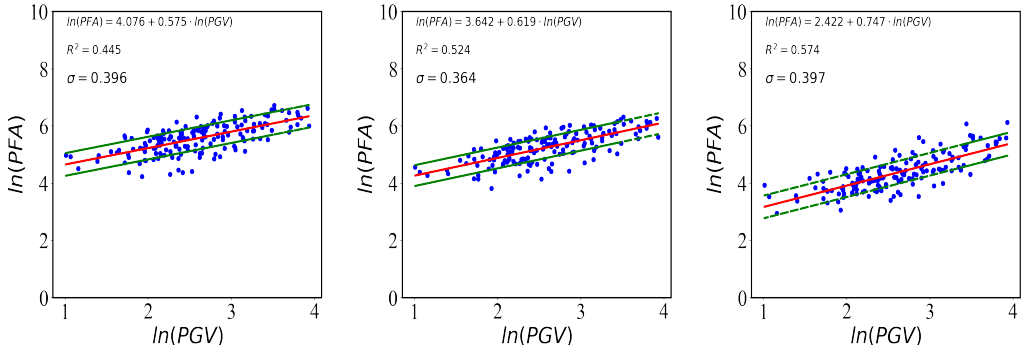
Figure A.6. Regression analyses of space frames for θ_{max} v.s. (a) D_{s575}, and (b) D_{s595}.



(a)



(b)



(c)

Figure A.7. Regression analyses of perimeter frames for PFA v.s. (a) S_{aT1}, (b) PGA, and (c) PGV.

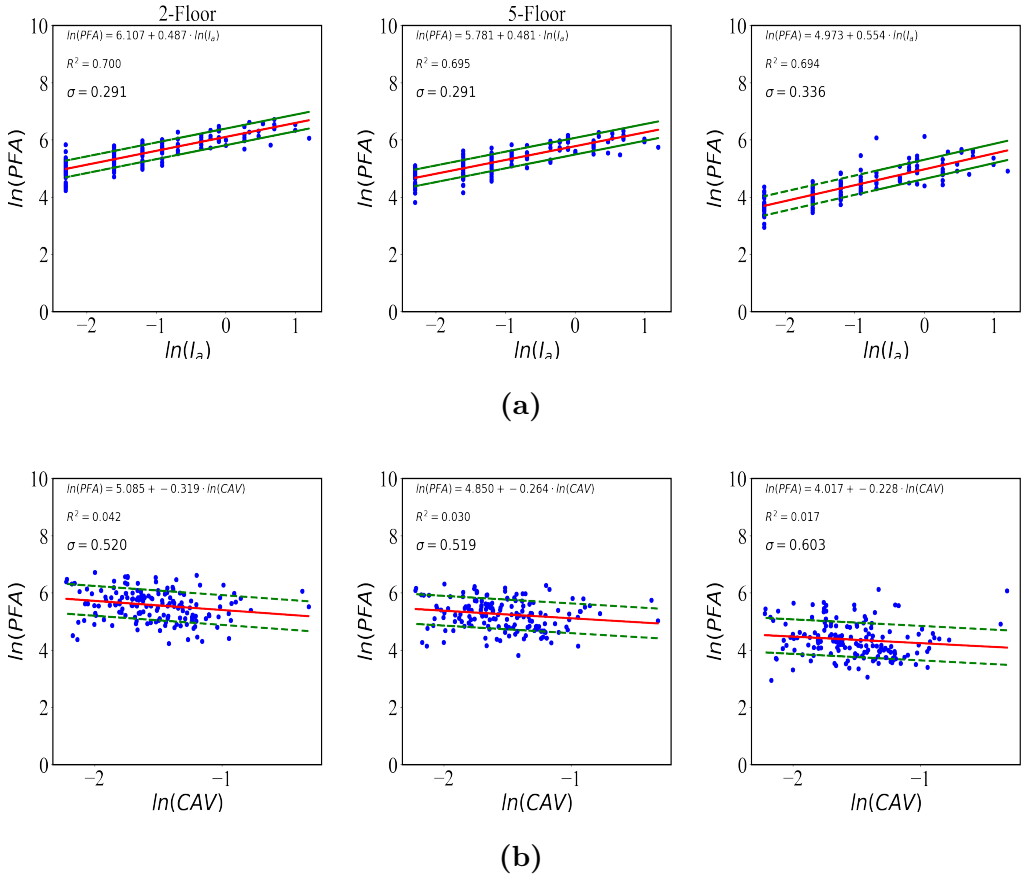


Figure A.8. Regression analyses of perimeter frames for PFA v.s. (a) I_a, and (b) CAV.

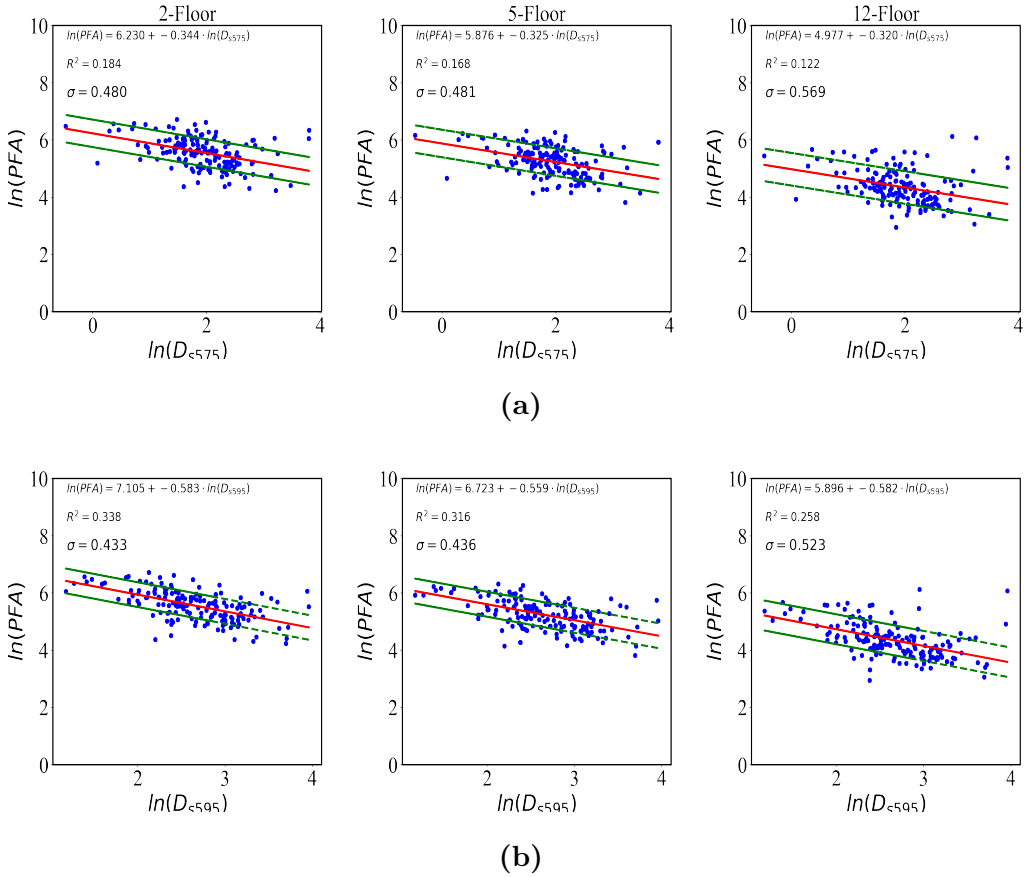
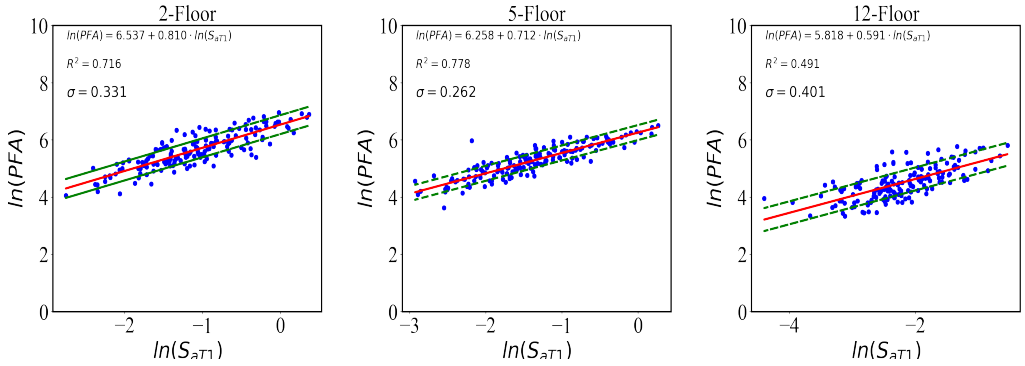
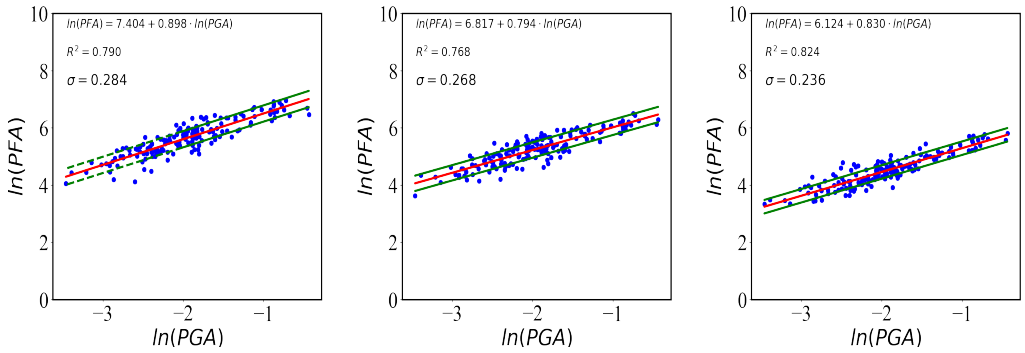


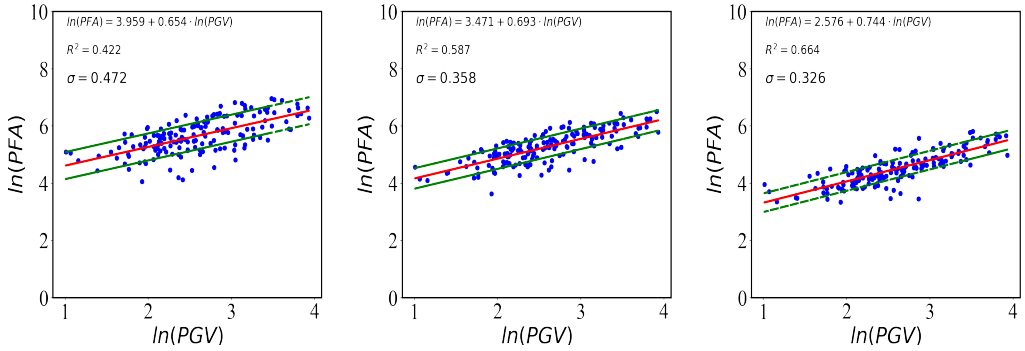
Figure A.9. Regression analyses of perimeter frames for PFA v.s. (a) D_{s575}, and (b) D_{s595}.



(a)



(b)



(c)

Figure A.10. Regression analyses of space frames for PFA v.s. (a) S_{aT1}, (b) PGA, and (c) PGV.

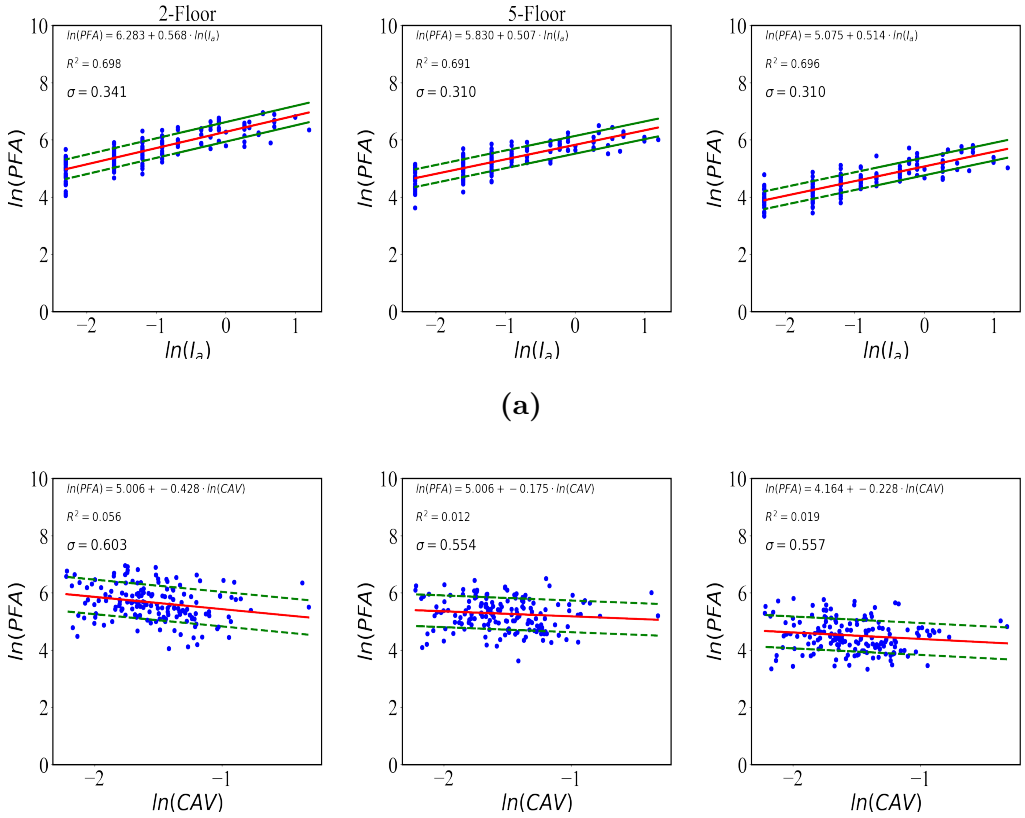


Figure A.11. Regression analyses of space frames for PFA v.s. (a) I_a, and (b) CAV.

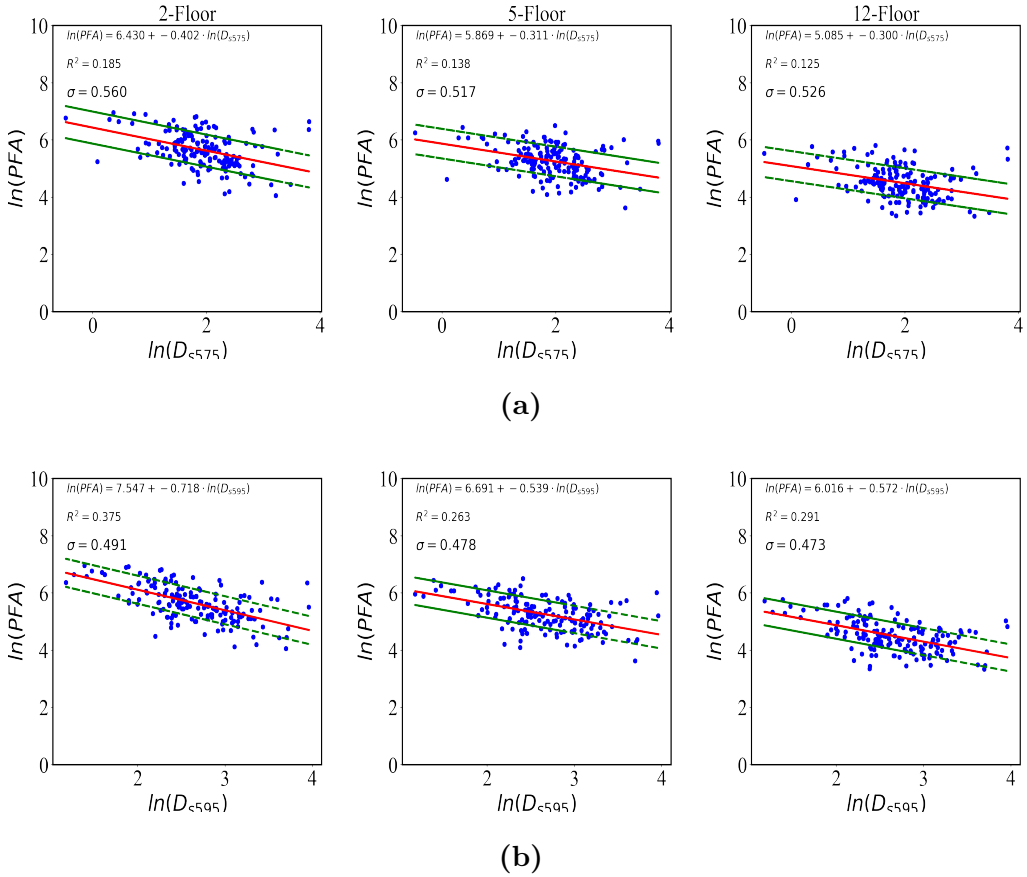


Figure A.12. Regression analyses of space frames for PFA v.s. (a) D_{s575}, and (b) D_{s595}.

Appendix B

Component Details for Loss Estimation

Table B.1. Structural component list.

Component	Description	DS: Median	Unit Repair Cost Range
B1041.041a	ACI 318 OMF with weak joints and column flexural response, Conc Col & Bm =24" x 24", Bean one side	DS1 (IDR): 0.015 DS2 (IDR): 0.0175 DS3 (IDR): 0.02	$\$23,131 - \$34,697$ $\$35,077 - \$52,615$ $\$43,176 - \$64,764$
B1041.041b	ACI 318 OMF with weak joints and column flexural response, Conc Col & Bm =24" x 24", Beam both sides	DS1 (IDR): 0.015 DS2 (IDR): 0.0175 DS3 (IDR): 0.02	$\$23,131 - \$34,679$ $\$35,077 - \$52,615$ $\$43,176 - \$64,764$
B1049.001a	Reinforced concrete flat slabs-columns without shear reinforcing 0<Vg/Vo<.2, no continuity reinf	DS1 (IDR): 0.025 DS2 (IDR): 0.04	$\$41,570 - \$61,132$ $\$56,841 - \$83,590$

Table B.2. Exterior finishes component list.

Component	Description	DS: Median	Unit Repair Cost Range
B2011.201a	Precast concrete panels 4.5 inches thick - in plane deformation	DS1a (IDR): 0.0182 DS1b (IDR): 0.0182	$\$9,122 - \$14,824$ $\$45,611 - \$74,118$
B2011.201b	Precast concrete panels 4.5 inches thick - out of plane deformation Curtain Walls - Generic Midrise Stick-Built Curtain wall, Confg: Insulating	DS1 (PFA): 6.01	$\$68,416 - \$111,176$
B2022.002	Glass Units (dual pane), Lamination: Unknown, Glass Type: Unknown, Details: Aspect ratio = 6:5, Other details Unknown	DS1 (IDR): 0.021 DS2 (IDR): 0.024	$\$2,127 - \$3,989$ $\$2,127 - \$3,989$

Table B.3. Partition walls component list.

Component	Description	DS: Median	Unit Repair Cost Range
C1011.001a	Wall Partition, Type: Gypsum with meta studs, Full Height, Fixed Below, Fixed Above	DS1 (IDR): 0.005 DS2 (IDR): 0.01 DS3 (IDR): 0.021	$\$2,457 - \$4,606$ $\$4,913 - \$9,213$ $\$10,040 - \$14,174$

Table B.4. Lighting component list.

Component	Description	DS: Median	Unit Repair Cost Range
C3034.001	Independent Pendant Lighting - non seismic	DS1 (PFA): 0.6	$\$1 - \$3$

Table B.5. Other nonstructural component list.

Component	Description	DS: Median	Unit Repair Cost Range
C2011.011b	Non-monolithic precast concrete stair assembly with concrete stringers and treads with no seismic joint	DS1 (IDR): 0.005 DS2 (IDR): 0.017 DS3 (IDR): 0.028	$\$567 - \$1,890$ $\$2,997 - \$9,989$ $\$18,588 - \$61,9,589$
C3011.011a	Wall Partition, Type: Gypsum + Wallpaper, Full Height, Fixed Below, Fixed Above	DS1 (IDR): 0.0021	$\$2,916 - \$4,374$
C3027.002	Raised Access Floor, seismically rated	DS1 (PFA): 1.5	$\$124 - \$186$
D4011.022a	Fire Sprinkler Water Piping - Horizontal Mains and Branches - Old Style Victaulic - Thin Wall Steel - No bracing, SDC C, PIPING FRAGILITY	DS1 (PFA): 1.1 DS2 (PFA): 2.4	$\$425 - \$520$ $\$3,219 - \$3,935$
D4011.032a	Fire Sprinkler Drop Standard Threaded Steel - Dropping into unbraced lay-in til SOFT ceiling - 6 ft. long drop maximum, SDC C	DS1 (PFA): 0.75 DS2 (PFA): 0.95	$\$607 - \$742$ $\$607 - \$742$
D5012.013a	Motor Control Center - Capacity: all - Unanchored equipment that is not vibration isolated - Equipment fragility only	DS1 (PFA): 0.73	$\$5,042 - \$6,162$
D5012.021d	Low Voltage Switchgear - Capacity: 1200 to 2000 Amp - Unanchored equipment that is not vibration isolated - Equipment fragility only	DS1 (PFA): 1.28	$\$62,444 - \$76,321$

Table B.6. Ceiling component list.

Component	Description	DS: Median	Unit Repair Cost Range
C3032.001a	Suspended Ceiling, SDC A,B,C, Area (A): A $<$ 250, Vert support only	DS1 (PFA): 1.17	$\$391 - \$587$
		DS2 (PFA): 1.58	$\$3,064 - \$4,596$
		DS3 (PFA): 1.82	$\$6,304 - \$9,456$
C3032.001b	Suspended Ceiling, SDC A,B,C, Area (A): 250 $<$ A $<$ 1000, Vert support only	DS1 (PFA): 1.01	$\$940 - \$1,409$
		DS2 (PFA): 1.45	$\$7,354 - \$11,031$
		DS3 (PFA): 1.69	$\$15,129 - \$22,694$
C3032.001c	Suspended Ceiling, SDC A,B,C, Area (A): 1000 $<$ A $<$ 2500, Vert support only	DS1 (PFA): 0.7	$\$2,819 - \$4,228$
		DS2 (PFA): 1.2	$\$22,062 - \$33,093$
		DS3 (PFA): 1.43	$\$45,388 - \$68,081$
C3032.001d	Suspended Ceiling, SDC A,B,C, Area (A): A $>$ 2500, Vert support only	DS1 (PFA): 0.56	$\$3,915 - \$5,872$
		DS2 (PFA): 1.08	$\$30,642 - \$45,963$
		DS3 (PFA): 1.31	$\$63,038 - \$94,558$

Table B.7. Elevator component list.

Component	Description	DS: Median	Unit Repair Cost Range
D1014.012	Traction Elevator - Applies to most California Installations prior to 1976, most western states installations prior to 1982 and most other U.S installations prior to 1998.	DS1a (PFA): 0.31	$\$3,564 - \$11,879$
		DS1b (PFA): 0.31	$\$15,145 - \$50,485$
		DS1c (PFA): 0.31	$\$12,959 - \$43,195$
		DS1d (PFA): 0.31	$\$2,025 - \$6,749$

Table B.8. Piping component list.

Component	Description	DS: Median	Unit Repair Cost Range
D2021.012a	Cold or Hot Potable - Small Diameter Threaded Steel - (2.5 inches in diameter or less), SDC C, PIPING FRAGILITY	DS1 (PFA): 1.5 DS2 (PFA): 2.6	$\$352 - \$431$ $\$3,219 - \$3,935$
D2021.022a	Cold or Hot Potable Water Piping (dia > 2.5 inches), SDC C, PIPING FRAGILITY	DS1 (PFA): 1.5 DS2 (PFA): 1.1	$\$283 - \$945$ $\$2,713 - \$9,044$
D2022.012a	Heating hot Water Piping - Small Diameter Threaded Steel - (2.5 inches in diameter or less), SDC C, PIPING FRAGILITY	DS1 (PFA): 0.55 DS2 (PFA): 2.6	$\$352 - \$431$ $\$3,219 - \$3,935$
D2022.012b	Heating hot Water Piping - Small Diameter Threaded Steel - (2.5 inches in diameter or less), SDC C, BRACING FRAGILITY	DS1 (PFA): 1.2 DS2 (PFA): 2.4	$\$462 - \$564$ $\$4,617 - \$5,642$
D2022.022a	Heating hot Water Piping - Large Diameter Welded Steel - (greater than 2.5 inches in diameter), SDC C, PIPING FRAGILITY	DS1 (PFA): 1.5 DS2 (PFA): 2.6	$\$283 - \$945$ $\$2,713 - \$9,044$
D2031.022a	Sanitary Waste Piping - Cast Iron w/bell and spigot couplings, SDC C, PIPING FRAGILITY	DS1 (PFA): 2.25	$\$2,713 - \$9,044$
D2031.022b	Sanitary Waste Piping - Cast Iron w/bell and spigot couplings, SDC C, BRACING FRAGILITY	DS1 (PFA): 1.2 DS2 (PFA): 2.4	$\$324 - \$1,080$ $\$3,523 - \$11,744$

Table B.9. Hvac component list.

Component	Description	DS: Median	Unit Repair Cost Range
D3031.011b	Chiller - Capacity: 100 to <350 Ton - Unanchored equipment that is not vibration isolated - Equipment fragility only	DS1 (PFA): 0.2	$\$185,876 - \$227,181$
D3031.021b	Chiller - Capacity: 100 to <350 Ton - Unanchored equipment that is not vibration isolated - Equipment fragility only	DS1 (PFA): 0.5	$\$94,760 - \$115,818$
D3041.011c	HVAC Galvanized Sheet Metal Ducting less than 6 sq. ft in cross sectional area, SDC D, E, or F	DS1 (PFA): 1.5 DS2 (PFA): 2.25	$\$790 - \$965$ $\$7,714 - \$9,429$
D3041.012c	HVAC Galvanized Sheet Metal Ducting - 6 sq. ft cross sectional area or greater, SDC D, E, or F	DS1 (PFA): 3.75 DS2 (PFA): 4.5	$1,154 - \$1,411$ $\$9,658 - \$11,805$
D3052.011d	Sanitary Waste Piping - Cast Iron w/bell and spigot couplings, SDC C, BRACING FRAGILITY	DS1 (PFA): 1.2 DS2 (PFA): 2.4	$\$2,430 - \$2,970$ $\$225,966 - \$276,181$

REFERENCES

[1] Mehdi M Akhlaghi, Supratik Bose, M Ebrahim Mohammadi, Babak Moaveni, Andreas Stavridis, and Richard L Wood. Post-earthquake damage identification of an rc school building in nepal using ambient vibration and point cloud data. *Engineering Structures*, 227:111413, 2021.

[2] Mohamad Alipour, Devin K Harris, and Gregory R Miller. Robust pixel-level crack detection using deep fully convolutional neural networks. *Journal of Computing in Civil Engineering*, 33(6):04019040, 2019.

[3] C Allin. Progress and challenges in seismic performance assessment. *PEER newsletter*, 2000.

[4] Anita Amirsardari, Pathmanathan Rajeev, Elisa Lumantarna, and Helen M Goldsworthy. Suitable intensity measure for probabilistic seismic risk assessment of nonductile australian reinforced concrete buildings. *Bulletin of earthquake engineering*, 17(7):3753–3775, 2019.

[5] T Anagnos, MC Comerio, C Goulet, H Na, J Steele, and JP Stewart. Los angeles inventory of nonductile concrete buildings for analysis of seismic collapse risk hazards. In *Proceedings of the 14th World Conference on Earthquake Engineering*, 2008.

[6] T Anagnos, MC Comerio, C Goulet, J Steele, and JP Stewart. Development of a concrete building inventory: Los angeles case study for the analysis of collapse risk. In *Proc. 9th National & 10th Canadian Conf. on Eq. Eng*, 2010.

[7] Thalia Anagnos, Mary C Comerio, Christine Goulet, Peter J May, Marjorie Greene, David L McCormick, and David Bonowitz. Developing regional building inventories: Lessons from the field. *Earthquake Spectra*, 28(4):1305–1329, 2012.

[8] Thalia Anagnos, Mary C Comerio, and Jonathan P Stewart. Earthquake loss estimates and policy implications for nonductile concrete buildings in los angeles. *Earthquake Spectra*, 32(4):1951–1973, 2016.

[9] ASCE. *Minimum Design Loads for Buildings and Other Structures (ASCE/SEI 7-10)*. American Society of Civil Engineers, 2013.

[10] Mark Aschheim and Jack P Moehle. *Shear strength and deformability of RC bridge columns subjected to inelastic cyclic displacements*, volume 92. Earthquake Engineering Research Center, University of California, 1992.

[11] ATC. Here today—here tomorrow: The road to earthquake resilience in san francisco, earthquake safety for soft-story buildings, 2009.

[12] Mohammad Reza Azadi Kakavand and Mohammad Khanmohammadi. Seismic fragility assessment of local and global failures in low-rise non-ductile existing rc buildings: Empirical shear-axial modelling vs. asce/sei 41 approach. *Computational Engineering and Physical Modeling*, 1(1):38–57, 2018.

[13] Jack W. Baker, Ting Lin, Shrey K. Shahi, and Nirmal Jayaram. *New Ground Motion Selection Procedures and Selected Motions for the PEER Transportation Research Program*. Number PEER 2011/03. Pacific Earthquake Engineering Research Center, Berkeley, CA, 2011.

[14] Shoraka M Baradaran and Kenneth J Elwood. Mechanical model for non ductile reinforced concrete columns. *Journal of Earthquake Engineering*, 17(7):937–957, 2013.

[15] Bentley System. *ContextCapture: Quick Start Guide*. Watertown, CT: Bentley Systems, 2017.

[16] Michael Berry, Myles Parrish, and Marc Eberhard. Peer structural performance database user's manual (version 1.0). *University of California, Berkeley*, 2004.

[17] Douglas G Bonett. Confidence interval for a coefficient of quartile variation. *Computational statistics & data analysis*, 50(11):2953–2957, 2006.

[18] Michel Bruneau, Stephanie E Chang, Ronald T Eguchi, George C Lee, Thomas D O'Rourke, Andrei M Reinhorn, Masanobu Shinozuka, Kathleen Tierney, William A Wallace, and Detlof Von Winterfeldt. A framework to quantitatively assess and enhance the seismic resilience of communities. *Earthquake spectra*, 19(4):733–752, 2003.

[19] Henry Burton, Aryan Rezaei Rad, Zhengxiang Yi, Damian Gutierrez, and Koyejo Ojuri. Seismic collapse performance of los angeles soft, weak, and open-front wall line woodframe structures retrofitted using different procedures. *Bulletin of earthquake engineering*, 17(4):2059–2091, 2019.

[20] Henry V Burton, Gregory Deierlein, David Lallemant, and Ting Lin. Framework for incorporating probabilistic building performance in the assessment of community seismic resilience. *Journal of Structural Engineering*, 142(8):C4015007, 2016.

[21] Ozan Cem Celik and Bruce R Ellingwood. Seismic fragilities for non-ductile reinforced concrete frames–role of aleatoric and epistemic uncertainties. *Structural Safety*, 32(1):1–12, 2010.

[22] Young-Jin Cha, Wooram Choi, and Oral Büyüköztürk. Deep learning-based crack damage detection using convolutional neural networks. *Computer-Aided Civil and Infrastructure Engineering*, 32(5):361–378, 2017.

[23] Jingdao Chen, Zsolt Kira, and Yong K Cho. Deep learning approach to point cloud scene understanding for automated scan to 3d reconstruction. *Journal of Computing in Civil Engineering*, 33(4):04019027, 2019.

[24] Yun-Jian Cheng, Wenge Qiu, and Jin Lei. Automatic extraction of tunnel lining cross-sections from terrestrial laser scanning point clouds. *Sensors*, 16(10):1648, 2016.

[25] C Comartin, D Bonowitz, M Greene, D McCormick, P May, and E Seymour. The concrete coalition and the california inventory project: An estimate of the number of pre-1980 concrete buildings in the state. *September, Oakland, CA: The Earthquake Engineering Research Institute*, 2011.

[26] Craig D Comartin, David McCormick, Marjorie Greene, and David Bonowitz. Building a volunteer engineering network: the experience of the concrete coalition. In *Proceedings of the 9th US National and 10th Canadian Conference on Earthquake Engineering, Toronto, CA*, 2010.

[27] MC Comerio and T Anagnos. Los angeles inventory: Implications for retrofit policies for nonductile concrete buildings. In *Proc. 15th World Conf. Earthquake Eng*, pages 24–28, 2012.

[28] ACI Committee. Commentary on building code requirements for reinforced concrete (aci 318-71). American Concrete Institute, 1971.

[29] ACI Committee. Building code requirements for structural concrete:(aci 318-02) and commentary (aci 318r-02). American Concrete Institute, 2002.

[30] D Cook, K Fitzgerald, T Chrupalo, and CB Haselton. Comparison of fema p-58 with other building seismic risk assessment methods. *Chico, California, available at: http://static-assets. hbrisk. com/reports/RiskAssessmentComparisonReport_FEMA-P-58-to-OtherMethods_2017_03_08. pdf*, 2017.

[31] Allin C. Cornell, Fatemeh Jalayer, Ronald O. Hamburger, and Douglas A. Foutch. Probabilistic basis for 2000 sac federal emergency management agency steel moment frame guidelines. *Journal of Structural Engineering*, 128(4):526–533, 2002.

[32] Applied Technology Council and United States. Federal Emergency Management Agency. *Quantification of building seismic performance factors*. US Department of Homeland Security, FEMA, 2009.

[33] Building Seismic Safety Council. Nehrp recommended provisions for seismic regulations for new buildings and other structures (fema 450). *Washington, DC*, 2003.

[34] Marco De Biasio, Stephane Grange, Frederic Dufour, Frederic Allain, and Ilie Petre-Lazar. A simple and efficient intensity measure to account for nonlinear structural behavior. *Earthquake Spectra*, 30(4):1403–1426, 2014.

[35] Gregory Deierlein and Abbie Liel. Benefit–cost evaluation of seismic risk mitigation in existing non-ductile concrete buildings. In *Advances in Performance-Based Earthquake Engineering*, pages 341–348. Springer, 2010.

[36] Gregory G Deierlein, Abbie B Liel, Curt B Haselton, and Charles A Kircher. Atc 63 methodology for evaluating seismic collapse safety of archetype buildings. In *Structures Congress 2008: Crossing Borders*, pages 1–10, 2008.

[37] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[38] Piotr Dollár and C Lawrence Zitnick. Structured forests for fast edge detection. In *Proceedings of the IEEE international conference on computer vision*, pages 1841–1848, 2013.

[39] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7), 2011.

[40] Kenneth J. Elwood and Jack P. Moehle. *Shake Table Tests and Analytical Studies on the Gravity Load Collapse of Reinforced Concrete Frames*. Number PEER 2003/01. Pacific Earthquake Engineering Research Center, Berkeley, CA, 2003.

[41] Kenneth J. Elwood and Jack P. Moehle. Drift capacity of reinforced concrete columns with light transverse reinforcement. *Earthquake Spectra*, 21(1):71–89, 2005.

[42] FEMA. Seismic evaluation and retrofit of multi-unit wood-frame buildings with weak first stories, 2012.

[43] Hazus-MH FEMA. Mr3 technical manual. *Multi-hazard loss estimation methodology earthquake model*, 2003.

[44] P FEMA. Seismic performance assessment of buildings. *Methodology". CA, USA, Applied Technology Council for the Federal Emergency Management Agency*, 1, 2012.

[45] P FEMA. Seismic performance assessment of buildings – performance assessment calculation tool (pact). *Methodology". CA, USA, Applied Technology Council for the Federal Emergency Management Agency*, 3, 2012.

[46] Panagiotis H Galanis and Jack P Moehle. Development of collapse indicators for risk assessment of older-type reinforced concrete buildings. *Earthquake Spectra*, 31(4):1991–2006, 2015.

[47] Yuqing Gao and Khalid M Mosalam. Deep transfer learning for image-based structural damage recognition. *Computer-Aided Civil and Infrastructure Engineering*, 33(9):748–768, 2018.

[48] Sean Gillies. The shapely user manual. *URL https://pypi. org/project/Shapely*, 2013.

[49] UM Gilmartin, SA Freeman, and SS Rihal. Using earthquake strong motion records to assess the structural and nonstructural response of the 7-story van nuys hotel to the northridge earthquake of january 17, 1994. In *Proceedings of 6th US National Conference on Earthquake Engineering*, number 268, Seattle, Washington, 1998.

[50] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.

[51] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[52] Cody C Harrington and Abbie B Liel. Collapse assessment of moment frame buildings, considering vertical ground shaking. *Earthquake Engineering & Structural Dynamics*, 45(15):2475–2493, 2016.

[53] Stephen K Harris and John A Egan. Effects of ground conditions on the damage to four-story corner apartment buildings. *The Loma Prieta, California, Earthquake of October 17, 1989–Marina District*, pages F181–F194, 1992.

[54] CB Haselton and J Baker. Seismic performance prediction program (sp3). *Haselton Baker Risk Group, Chico, CA*, 2018.

[55] Curt B. Haselton, Christine A. Goulet, Judith Mitrani-Reiser, James L Beck, Gregory G. Deierlein, Keith A. Porter, Jonathan P. Stewart, and Ertugrul Taciroglu. *An assessment to benchmark the seismic performance of a code-conforming reinforced-concrete moment-frame building.* Number PEER 2007/1. Pacific Earthquake Engineering Research Center, Berkeley, CA, 2008.

[56] Wael M Hassan and Jack P Moehle. Experimental assessment of seismic vulnerability of corner beam-column joints in older concrete buildings. In *Proc. 15th World Conf. Earthquake Eng*, pages 24–28, 2012.

[57] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[58] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[59] Joseph F. Healey. *The essentials of statistics: A tool for social research.* Nelson Education, 2012.

[60] Kurt Henkhaus, Santiago Pujol, and Julio Ramirez. Axial failure of reinforced concrete columns damaged by shear reversals. *Journal of structural engineering*, 139(7):1172–1180, 2013.

[61] Geoffrey Hinton, Nitish Srivastava, and Kevin Swersky. Neural networks for machine learning lecture 6a overview of mini-batch gradient descent. *Cited on*, 14(8), 2012.

[62] WT Holmes and P Somers. Northridge earthquake reconnaissance report, vol. 2. *Earthquake Spectra*, 11:125–176, 1996.

[63] Luis Francisco Ibarra and Helmut Krawinkler. *Global collapse of frame structures under seismic excitations*. Number 152. The John A. Blume Earthquake Engineering Center, Stanford,CA, 2005.

[64] ICBO. *Uniform Building Code*. International Conference of Building Officials, Pasadena, CA, 1967.

[65] M Saiful Islam. Analysis of the northridge earthquake response of a damaged non-ductile concrete frame building. *The structural design of tall buildings*, 5(3):151–182, 1996.

[66] Jong-Su Jeon. *Aftershock vulnerability assessment of damaged reinforced concrete buildings in California*. PhD thesis, Georgia Institute of Technology, 2013.

[67] Jong-Su Jeon, Reginald DesRoches, Laura N Lowes, and Ioannis Brilakis. Framework of aftershock fragility assessment–case studies: Older california reinforced concrete building frames. *Earthquake Engineering & Structural Dynamics*, 44(15):2617–2636, 2015.

[68] Jong-Su Jeon, Laura N Lowes, Reginald DesRoches, and Ioannis Brilakis. Fragility curves for non-ductile reinforced concrete frames that exhibit different component response mechanisms. *Engineering Structures*, 85:127–143, 2015.

[69] K Jordahl. Geopandas: Python tools for geographic data. *URL: https://github.com/geopandas/geopandas*, 2014.

[70] T Kabeyasawa, T Matsumori, H Katsumata, and K Shirai. Design of the full-scale six-story reinforced concrete wall-frame building for testing at e-defense. In *US-Japan Workshop*, 2005.

[71] MR Azadi Kakavand. Limit state material manual. *Available on www. opensees. berkeley. edu, University of California, Berkeley, California*, 2012.

[72] Rony Kalfarisi, Zheng Yi Wu, and Ken Soh. Crack detection and segmentation using deep learning with 3d reality mesh model for quantitative assessment and integrated visualization. *Journal of Computing in Civil Engineering*, 34(3):04020010, 2020.

[73] Jian Kang, Marco Körner, Yuanyuan Wang, Hannes Taubenböck, and Xiao Xiang Zhu. Building instance classification using street view images. *ISPRS journal of photogrammetry and remote sensing*, 145:44–59, 2018.

[74] Alireza G Kashani, Patrick S Crawford, Sufal K Biswas, Andrew J Graettinger, and David Grau. Automated tornado damage assessment and wind speed estimation based on terrestrial laser scanning. *Journal of Computing in Civil Engineering*, 29(3):04014051, 2015.

[75] Hongjo Kim, Hyoungkwan Kim, Yong Won Hong, and Hyeran Byun. Detecting construction equipment using a region-based fully convolutional network and transfer learning. *Journal of computing in Civil Engineering*, 32(2):04017082, 2018.

[76] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[77] Mohsen Kohrangi, Paolo Bazzurro, and Dimitrios Vamvatsikos. Vector and scalar ims in structural response estimation: Part ii – building demand assessment. *Earthquake Spectra*, 32:1525–1543, 2016.

[78] Mohsen Kohrangi, Dimitrios Vamvatsikos, and Paolo Bazzurro. Implications of intensity measure selection for seismic loss assessment of 3-d buildings. *Earthquake Spectra*, 32(4):2167–2189, 2016.

[79] Helmut Krawinkler. *Van Nuys Hotel Building Testbed Report: Exercising Seismic Performance assessment*. Number PEER 2005/11. Pacific Earthquake Engineering Research Center, Berkeley, CA, 2005.

[80] Srinath Shiv Kumar, Mingzhu Wang, Dulcy M Abraham, Mohammad R Jahanshahi, Tom Iseley, and Jack CP Cheng. Deep learning–based automated detection of sewer defects in cctv videos. *Journal of Computing in Civil Engineering*, 34(1):04019047, 2020.

[81] Do Hyung Lee and Amr S Elnashai. Seismic analysis of rc bridge columns with flexure-shear interaction. *Journal of Structural Engineering*, 127(5):546–553, 2001.

[82] Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. Pointcnn: Convolution on x-transformed points. In *Advances in neural information processing systems*, pages 820–830, 2018.

[83] Abbie B. Liel. *Assessing the Collapse Risk of California's Existing Reinforced Concrete Frame Structures: metrics for Seismic Safety Decisions*. PhD thesis, Stanford University, 2008.

[84] Abbie B Liel and Gregory G Deierlein. Cost-benefit evaluation of seismic risk mitigation alternatives for older concrete frame buildings. *Earthquake Spectra*, 29(4):1391–1411, 2013.

[85] Abbie B. Liel, Curt B. Haselton, and Gregory G. Deierlein. Seismic collapse safety of reinforced concrete buildings. ii: Comparative assessment of nonductile and ductile moment frames. *Journal of Structural Engineering*, 137(4):492–502, 2011.

[86] Abby B. Liel and Gregory G. Deierlein. Using collapse risk assessments to inform seismic safety policy for older concrete buildings. *Earthquake Spectra*, 28:1495–1521, 2012.

[87] Dimitrios G. Lignos and Helmut Krawinkler. Development and utilization of structural component databases for performance-based earthquake engineering. *Journal of Structural Engineering*, 139(8):1382–1394, 2013.

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

[89] Nicolas Luco and Allin C. Cornell. Structure-specific scalar intensity measures for near-source and ordinary earthquake ground motions. *Earthquake Spectra*, 23(2):357–392, 2007.

[90] M Markhvida, L Ceferino, and J Baker. Effect of ground motion correlation on regional seismic loss estimation: application to lima, peru using a cross-correlated principal component analysis model. In *Proc., 12th Int. Conf. on Structural Safety and Reliability*, volume 1853, 2017.

[91] Frank McKenna. Opensees: a framework for earthquake engineering simulation. *Computing in Science & Engineering*, 13(4):58–66, 2011.

[92] HK Miyamoto, AS Gilani, and Akira Wada. Reconnaissance report of the 2008 sichuan earthquake, damage survey of buildings and retrofit options. In *Proceedings of the 14th World Conference on Earthquake Engineering*, pages 12–17, 2008.

[93] Jawad Nagi, Frederick Ducatelle, Gianni A Di Caro, Dan Cireşan, Ueli Meier, Alessandro Giusti, Farrukh Nagi, Jürgen Schmidhuber, and Luca Maria Gambardella. Max-pooling convolutional neural networks for vision-based hand gesture recognition. In *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pages 342–347. IEEE, 2011.

[94] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *ICML*, 2010.

[95] Shunsuke Otani and Mete Avni Sozen. Behavior of multistory reinforced concrete frames during earthquakes. Technical report, University of Illinois Engineering Experiment Station. College of . . . , 1972.

[96] Joško Ožbolt, Utz Mayer, and Henrik Vocke. Smeared fracture fe-analysis of reinforced concrete structures—theory and examples. In *Modeling of inelastic behavior of RC structures under seismic loads*, pages 234–256. ASCE, 1999.

[97] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

[98] Telemachos B Panagiotakos and Michael N Fardis. Deformations of reinforced concrete members at yielding and ultimate. *Structural Journal*, 98(2):135–148, 2001.

[99] Joonam Park, Paolo Bazzurro, Jack W Baker, et al. Modeling spatial correlation of ground motion intensity measures for regional seismic hazard and portfolio loss estimation. *Applications of statistics and probability in civil engineering*, pages 1–8, 2007.

[100] JoséA Pincheira, Faridun S Dotiwala, and Jonathan T D'Souza. Seismic analysis of older reinforced concrete columns. *Earthquake Spectra*, 15(2):245–272, 1999.

[101] Santiago Pujol, JA Ramfrez, and Mete A Sozen. Drift capacity of reinforced concrete columns subjected to cyclic shear reversals. *Special Publication*, 187:255–274, 1999.

[102] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017.

[103] Ning Qian. On the momentum term in gradient descent learning algorithms. *Neural networks*, 12(1):145–151, 1999.

[104] Pathmanathan Rajeev and S Tesfamariam. Seismic fragilities of non-ductile reinforced concrete frames with consideration of soil structure interaction. *Soil Dynamics and Earthquake Engineering*, 40:78–86, 2012.

[105] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[106] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.

[107] Paolo Ricci, Flavia De Luca, and Gerardo Mario Verderame. 6th april 2009 l'aquila earthquake, italy: reinforced concrete building performance. *Bulletin of earthquake engineering*, 9(1):285–305, 2011.

[108] H Sezen and T Chowdhury. Hysteretic model for the lateral behavior of reinforced concrete columns including shear deformation. *ASCE Journal of Structural Engineering*, 135(2):139–146, 2009.

[109] Halil Sezen. Seismic behavior and modeling of reinforced concrete building columns. 2004.

[110] Abdollah Shafieezadeh, Karthik Ramanathan, Jamie E. Padgett, and Reginald DesRoches. Fractional order intensity measures for probabilistic seismic demand modeling applied to highway bridges. *Earthquake Engineering & Structural Dynamics*, 41(3):391–409, 2012.

[111] PB Shing and B Spencer. Modeling of shear behavior of rc bridge structures, modeling of inelastic behavior of rc structures under seismic load. 2001.

[112] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[113] Enrico Spacone, Filip C Filippou, and Fabio F Taucer. Fibre beam–column model for non-linear analysis of r/c frames: Part i. formulation. *Earthquake Engineering & Structural Dynamics*, 25(7):711–725, 1996.

[114] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[115] Shivangi Srivastava, John E Vargas-Muñoz, and Devis Tuia. Understanding urban landuse from the above and ground perspectives: A deep learning, multimodal solution. *Remote sensing of environment*, 228:129–143, 2019.

[116] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-first AAAI conference on artificial intelligence*, 2017.

[117] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[118] Yolanda Torres, José Juan Arranz, Jorge M Gaspar-Escribano, Azadeh Haghi, Sandra Martínez-Cuevas, Belén Benito, and Juan Carlos Ojeda. Integration of lidar and multispectral images for rapid exposure and earthquake vulnerability estimation. application in lorca, spain. *International Journal of Applied Earth Observation and Geoinformation*, 81:161–175, 2019.

[119] Mihailo Dimitrije Trifunac and TY Hao. 7-Storey Reinforced Concrete Building in Van Nuys, California: Photographs of the Damage from the 1994 Northridge Earthquake. Technical Report USC CE 01-05, University of Southern California, Los Angeles, California, 2001.

[120] KC Tsai, Chiang Pi Hsiao, and Michel Bruneau. Overview of building damages in 921 chi-chi earthquake. *Earthquake Engineering and Engineering Seismology*, 2(1):93–108, 2000.

[121] Dimitrios Vamvatsikos and C Allin Cornell. Applied incremental dynamic analysis. *Earthquake spectra*, 20(2):523–553, 2004.

[122] Frank J Vecchio and Michael P Collins. The modified compression-field theory for reinforced concrete elements subjected to shear. *ACI J.*, 83(2):219–231, 1986.

[123] Ruisheng Wang. 3d building modeling using images and lidar: A review. *International Journal of Image and Data Fusion*, 4(4):273–292, 2013.

[124] GA Weatherill, V Silva, H Crowley, and P Bazzurro. Exploring the impact of spatial correlations and uncertainties for portfolio analysis in probabilistic seismic loss estimation. *Bulletin of Earthquake Engineering*, 13(4):957–981, 2015.

[125] Chiun-lin Wu, Wu-Wei Kuo, Yuan-Sen Yang, Shyh-Jiann Hwang, Kenneth J Elwood, Chin-Hsiung Loh, and Jack P Moehle. Collapse of a nonductile concrete frame: Shaking table tests. *Earthquake Engineering & Structural Dynamics*, 38(2):205–224, 2009.

[126] Z. Y. Wu and M. Hmosze. Soft-story detection by deep learning. Technical report, Bentley Systems, Incorporated, Watertown, CT: Bentley Systems, 2018.

[127] Z. Y. Wu, M. Hmosze, and R. Kalfarisi. Soft-story building detection using deep convolutional neural network for complying building resilience ordinance. In *ASCE Engineering Mechanics Institute Conference*, Caltech, Pasadena, CA., 2019.

[128] Soheil Yavari, Kenneth J Elwood, and Chiun-lin Wu. Collapse of a nonductile concrete frame: Evaluation of analytical models. *Earthquake engineering & structural dynamics*, 38(2):225–241, 2009.

[129] Zhengxiang Yi, Henry V Burton, Mehrdad Shokrabadi, and Omar Issa. Multi-scale cost-benefit analysis of the los angeles soft-story ordinance. *Engineering Structures*, 214:110652, 2020.

[130] Hirozo Mihashi Yoshio Kaneko and Seiichiro Ishihara. Entire load-displacement characteristics for direct shear failure of concrete. *Modeling of inelastic behavior of RC structures under seismic loads*, page 175, 2001.

[131] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? In *Advances in neural information processing systems*, pages 3320–3328, 2014.

[132] Qian Yu, Chaofeng Wang, Frank McKenna, X Yu Stella, Ertugrul Taciroglu, Barbaros Cetiner, and Kincho H Law. Rapid visual screening of soft-story buildings from street view images using deep learning classification. *Earthquake Engineering and Engineering Vibration*, 19(4):827–838, 2020.

[133] Christos A Zeris and Stephen A Mahin. Behavior of reinforced concrete structures subjected to biaxial excitation. *Journal of structural engineering*, 117(9):2657–2673, 1991.

[134] Allen Zhang, Kelvin CP Wang, Yue Fei, Yang Liu, Siyu Tao, Cheng Chen, Joshua Q Li, and Baoxian Li. Deep learning–based fully automated pavement crack detection on 3d asphalt surfaces with an improved cracknet. *Journal of Computing in Civil Engineering*, 32(5):04018041, 2018.

[135] Allen Zhang, Kelvin CP Wang, Baoxian Li, Enhui Yang, Xianxing Dai, Yi Peng, Yue Fei, Yang Liu, Joshua Q Li, and Cheng Chen. Automated pixel-level pavement crack detection on 3d asphalt surfaces using a deep-learning network. *Computer-Aided Civil and Infrastructure Engineering*, 32(10):805–819, 2017.

[136] Zixiang Zhou, Jie Gong, and Mengyang Guo. Image-based 3d reconstruction for posthurricane residential building damage assessment. *Journal of Computing in Civil Engineering*, 30(2):04015015, 2016.

[137] A Zsarnóczay and GG Deierlein. Pelicun–a computational framework for estimating damage, loss and community resilience. In *Proceedings, 17th World Conference on Earthquake Engineering,(Sendai: WCEE)*, 2020.