Earthq Eng & Eng Vib (2020) 19: 827-838 **DOI**: https://doi.org/10.1007/s11803-020-0598-2

Rapid visual screening of soft-story buildings from street view images using deep learning classification

Qian Yu[1†], Chaofeng Wang[2†], Frank McKenna[2‡], Stella X. Yu[1§], Ertugrul Taciroglu[3*], Barbaros Cetiner[3†] and Kincho H. Law[4*]

1. *International Computer Science Institute, University of California, Berkeley, CA, USA*
2. *Department of Civil and Environmental Engineering, University of California, Berkeley, CA, USA*
3. *Civil and Environmental Engineering, University of California, Los Angeles, CA, USA*
4. *Civil and Environmental Engineering, Stanford University, CA, USA*

Abstract: Rapid and accurate identification of potential structural deficiencies is a crucial task in evaluating seismic vulnerability of large building inventories in a region. In the case of multi-story structures, abrupt vertical variations of story stiffness are known to significantly increase the likelihood of collapse during moderate or severe earthquakes. Identifying and retrofitting buildings with such irregularities—generally termed as soft-story buildings—is, therefore, vital in earthquake preparedness and loss mitigation efforts. Soft-story building identification through conventional means is a labor-intensive and time-consuming process. In this study, an automated procedure was devised based on deep learning techniques for identifying soft-story buildings from street-view images at a regional scale. A database containing a large number of building images and a semi-automated image labeling approach that effectively annotates new database entries was developed for developing the deep learning model. Extensive computational experiments were carried out to examine the effectiveness of the proposed procedure, and to gain insights into automated soft-story building identification.

Keywords: soft-story building; deep learning; CNN; rapid visual screening; street view image

1 Introduction

Soft-story (SS) buildings are a common archetype that have distinct visual characteristics, such as having a large opening on the ground floor, e.g., a garage (One key criterion of defining a soft-story building is the stiffness of the ground floor relative to that of the floors above. In appearance, a soft-story building typically has an open space such as a garage on the ground floor). Such ample opening space can make a ground floor not as stiff as the higher floors, leading to the name ‘soft-story’. Consequently, an SS building is vulnerable to a moderate or severe earthquake (see Fig. 1). Take Los Angeles for example, in the 1994 Northridge earthquake, where two-thirds of the approximately 49,000 destroyed or damaged apartment units were SS buildings (https://la.curbed.com/2018/1/17/16871368/earthquake-apartments-safe-northridge). Since the west coast of the United States is situated along the belt of high seismicity, residents and properties are under the constant threat of earthquakes (see Fig. 2). As a result, a series of building reinforcement and mandatory retrofit projects have been launched since 2009, aiming to reduce structural deficiencies and to improve the performance of SS buildings during earthquakes.

Screening, which is labor intensive and is hard to organize, is often the first step of retrofit programs. In general, the screening process includes collecting buildings and evaluating structural integrity. Once the buildings of concern are identified, the government authority reports such information as building location, type, and status. The residents or property owners are then required to hire licensed professional engineers to conduct structural inspection and to file a report to the authorities. The reporting process can be costly and time-consuming. To reduce the economic burden on the residents and the time and labor cost imposed on government agencies, an automated building classification process is desirable.

The availability of street view images and recent advances in computer vision techniques make automated rapid screening of SS buildings viable. In recent years, street view images have attracted significant attention from researchers, and have emerged as a much sought resource because of their availability and the rich

Correspondence to: Chaofeng Wang, Department of Civil and Environmental Engineering, University of California, Berkeley, CA, USA
E-mail: c_w@berkeley.edu
†Postdoctoral Scholar; ‡Research Engineer; §Director (Vision); *Professor

Received July 3, 2020; **Accepted** September 4, 2020

visual information captured in the images. Studies have shown the potential of using the images for a variety of applications, ranging from predicting housing prices (Bency *et al*., 2017; Law *et al*., 2018) to evaluating the safety of neighborhoods (Naik *et al*., 2014; Liu *et al*., 2017). Given the distinctive visual appearance of SS buildings, the possibility of recognizing such buildings from street view images is promising.

The past several years have seen significant progress in machine learning technologies, which have been widely embraced by the computer vision community. Deep learning (DL) techniques such as convolutional neural networks (CNNs) have been applied and have achieved impressive performance in various applications, such as face verification (Sun *et al*., 2014) and object detection (Girshick, 2015; Ren *et al*., 2015).

In this work, a DL-based framework is proposed for automatically recognizing a soft-story building in an image. The SS building identification framework includes three tasks: collecting street images, building a classification model, and analyzing the results. To be specific, a large-scale building database was collected using the Google Street View Static API (Google Street View Static API: https://developers.google.com/maps/documentation/streetview/intro). Currently, the database contains 25K images captured from Google Street Map in five cities in California: Santa Monica, Oakland, San Francisco, San Jose, and Berkeley. The images are grouped by their source cities so that each set of images is geographically organized and representative of the city, and more importantly, the images can be used by municipal agencies or engineering companies.

Based on the database of collected images, a CNN model was developed to classify SS buildings. The classification task from street view images is not trivial. First of all, as illustrated in Fig. 3, street view images, which are captured on the roads by cameras instrumented on moving vehicles, are noisy. In a street view image, buildings may be heavily occluded by trees or cars. Furthermore, an image may not capture a building because of improper viewpoint. In addition to the noisy image problem, labeling a large number of images for training

(a) A typical soft-story building (b) Examples of damaged soft-story buildings

Fig. 1 Soft-story buildings in good and bad conditions. Soft-story buildings are common in many countries. Due to specific structure, such buildings are likely to collapse during a moderate or severe earthquake

Fig. 2 Urgently needed soft-story retrofit programs. Soft-story buildings are especially common on the west coast of the United States. As shown in the National Seismic Hazard Map released by U.S. Geological Survey (USGS) in 2014, the west coast is located in a seismic belt. Displayed here are five example images from the data sets collected from five west-coast cities

a classification model can be tedious, labor-intensive, and time-consuming. To address these challenges, the authors developed a semi-automatic labeling strategy to build a benchmark data set for training the classification model. An analysis module based on a class activation map was deployed to visualize the "cues" that the model used for the classification. The visualizations using the classification map can not only provide users with extra evidence for the predictions, but also help to diagnose the classification results and to improve the learning model.

The contributions of this work are three-fold: (1) this study represents the first effort towards developing a DL-based framework for automatic soft-story building classification; (2) a large-scale street view image database was compiled, accompanied with a semi-automatic data labeling strategy; (3) a benchmark data set was built for SS building to demonstrate the potential use of deep learning techniques in earthquake engineering and regional seismic vulnerability analysis. Based on the building inventory database collected from street view images, the authors conducted extensive experiments and produced a comprehensive analysis of the DL-based SS building classification framework. The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 describes the proposed framework. Experimental details and results are presented in Section 4, and an application example is provided in Section 5. Finally, the paper concludes with a brief summary and discussion.

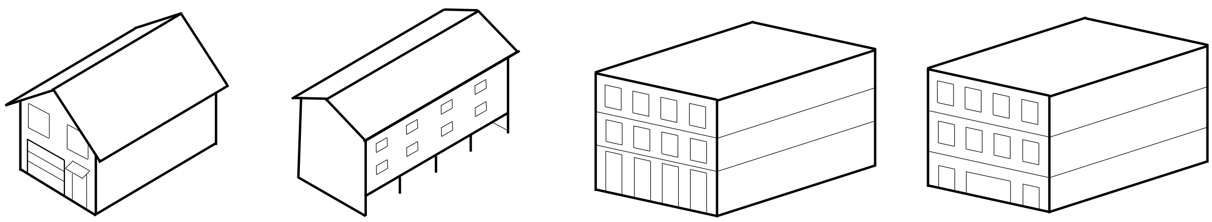
2 Related work

Screening of seismic vulnerable structures such as soft-story buildings requires comprehensive consideration of information from different areas of expertise, including the knowledge about seismic hazards, the performance of buildings during major earthquakes, seismic evaluation and performance assessment tools, etc. In 1988, ATC developed a handbook, FEMA 154 (ATC, 1988), which contains guidance and the technical basis for evaluating seismic performance of buildings using a scoring system. The handbook was later updated in ATC 2002 and ATC 2015. The objective of FEMA 154 was to provide for the earthquake community a methodology for evaluating the seismic safety of a large inventory of buildings quickly and inexpensively, with minimum access to the buildings (e.g., based on the visual clues manifested at building exteriors), and to determine whether a building requires a more detailed examination. The FEMA 154 method has been broadly used, or used for inspiration, in many rapid visual screening projects in different countries (Karbassi and Nollet, 2007; Wallace and Miller, 2008; Srikanth *et al*., 2010; Saatcioglu *et al*., 2013; Perrone *et al*., 2015; Ploeger *et al*., 2016; Ningthoujam and Nanda, 2018).

Visual-based screening methods work because the seismic performance of a building to a great extent, depends on its structure type, geometric irregularities, and foundation conditions, and these attributes can usually be identified based on visual clues. In a building structure, a story is called a soft story if its stiffness is dramatically weaker than other stories. However, it is impossible to quantitatively evaluate the strength/stiffness of each story in an as-built building if detailed information (e.g., the Building Information Model) of the building is not available. Fortunately, the existence of a soft story can often be identified using certain observable clues. For example, as shown in Fig. 4, if a building has an occupied space above a garage with limited / narrow wall widths on both sides of the garage opening, a large opening at the ground story, a story with less wall area

(a) Example images with improper viewpoints (b) Example images with heavy occlusions

Fig. 3 Challenges in classifying a soft-story building in a street view image



(a) Building with occupied space over a garage with limited / short wall lengths on both sides of the garage opening (b) Building with an open front at the ground story (c) Building with one story taller than the others (d) Building with one story having less wall or fewer columns than the others

Fig. 4 Soft-story buildings: buildings with severe vertical irregularity

or fewer columns than the other stories, or a story taller than the others, the building may potentially be identified as a soft-story building. These visual cues are obvious to a trained observer. In other words, visual screening of soft-story buildings can be performed by professionals following the rules and methods provided in FEMA 154. Although visual screening has been widely adopted, the process can be costly and error-prone, as it can be labor intensive in collecting a large amount of data (i.e., images) on the buildings, and human evaluations can be subjective, which may lead to differing interpretations and possibly erroneous results.

This study proposes an alternative automated approach that collects street view images automatically from Google Maps and uses deep learning techniques to identify and classify soft-story buildings from the images. The proposed method has many advantages over traditional screening methods regarding cost efficiency, scalability, and consistency in the evaluation. Street view images are essentially photos taken on the road that capture such objects as buildings, trees, and cars. Such images have attracted much attention from researchers in recent years because: (1) street view images can be obtained easily; (2) they provide rich visual information; (3) computational software and hardware used to process images have been improved significantly. Researchers have demonstrated the potentials of street view images in many applications. Naik *et al*. (2014) predicted the perceived level of public safety by analyzing millions of street view images. This work shows that the visual appearance of urban environments can reflect the lives of residents in a neighborhood. Gebru *et al*. (2017) utilized the extracted information of cars to estimate the demographic makeup of neighborhoods. Law *et al*. (2018) used street view images to estimate housing prices in London, indicating a strong correlation between housing price and street appearance. Kang *et al*. (2018) used street view images together with satellite images to predict the function of a building. By leveraging the information contained in street view images, a variety of applications have been developed that can potentially benefit city planning and real estate marketing.

Recent years have seen significant progress in the field of computer vision, taking advantage of the advances in machine learning research, particularly on deep neural networks. A typical deep neural network consists of a stack of convolutional layers and fully connected layers. Each layer, except for the final layer, serves as a feature extractor. As opposed to traditional classification methods, which require the selection of an 'optimal' feature representation, a deep learning neural network can learn and extract features from the images themselves. When training a deep neural network for object classification, an image is used as an input and propagates forward through the network. Furthermore, a ground truth label is provided for supervision. The difference, i.e., the error, between the ground truth and the prediction from the training network is then propagated backward from the last layer to the first layer. The parameters of all the layers in the network are iteratively updated by repeating the forward and backward propagation process. Deep neural networks have been successfully applied in many computer vision tasks, such as object recognition, detection, and segmentation. Despite the significant progress made in computer vision and image processing, applying deep neural network methods to recognize an SS building from a raw image remains a challenging task since the visual features of a soft story are local and subtle.

3 Overall workflow for classifying soft stroy buildings

As shown in Fig. 5, there are three basic steps in the process of classifying soft-story buildings. The first step is to collect images of buildings and establish a building database for the cities of interest. Selected data

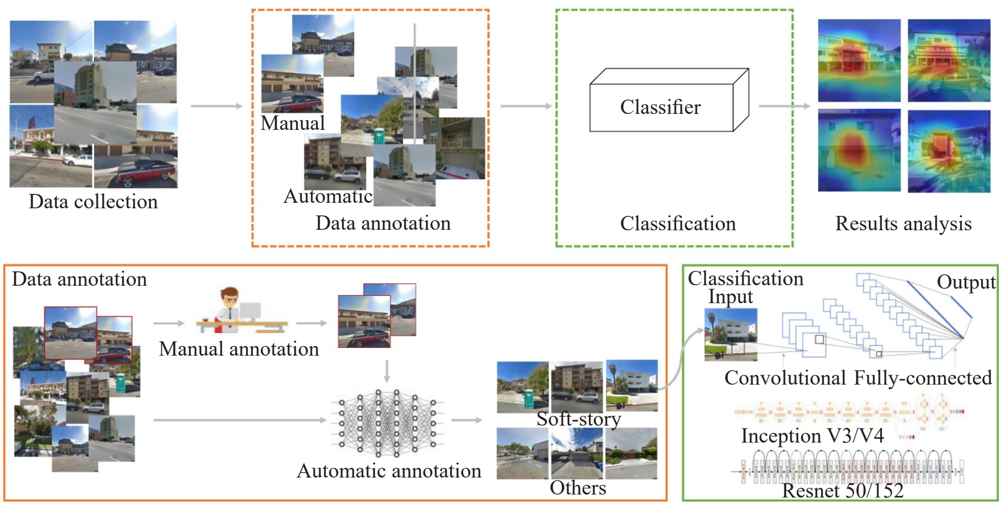


Fig. 5 Overall workflow of the proposed framework

sets collected are then used for training the classification models. The trained models are then applied to predict SS buildings in the cities of concern, and the predicted results are evaluated, for example, by using a visualization aid and scoring metrics

3.1 Create a building database

In this work, five cities in California, including Santa Monica, Oakland, San Francisco, San Jose, and Berkeley, were selected for the study of the SS building classification problem. The process of creating a building database for the five cities consists of two tasks, namely, image collection and image annotation.

Image collection Firstly, building addresses of each city are obtained from official websites. Based on the building addresses, street view images of individual buildings are downloaded using the Google Street View Static API (or "Google API" for short). Parameters of the API, such as field of view, pitch, and heading, are manually set and consistently applied to all cities. Altogether, as tabulated in Table 1, 25,340 images were collected. The data sets for Santa Monica, Oakland and San Francisco contain 16,665, 1,359 and 6,921 images, respectively. The images collected for Berkeley and San Jose have only 395 images, and are therefore grouped together as a single data set.

Image annotation For model training purposes, the building images need to be annotated or labelled. Manually annotating the entire database with a large number of images is labor intensive and time-consuming. A semi-automatic labeling strategy was developed to facilitate the annotation of the images. The basic idea is to first train a classifier on a small manually annotated data set and then apply the classifier to categorize the rest of the unlabeled images. This process can be repeated depending on the accuracy and the amount of data available and needed for annotation. In this work, an expert was recruited to annotate 1,302 images selected from the Santa Monica data set, with about half of the images being SS buildings. A CNN model was then trained and used to classify the other images in Santa Monica and Oakland. This labelling strategy was employed for annotating 18,419 images for the cities of Santa Monica and Oakland. Additionally, given the small number of images, the data set for Berkeley and San Jose was manually annotated and was used as a (ground truth) data set for testing the generalization of the trained model on other cities. Finally, the images for the city of San Francisco were left as raw data to simulate the application of the SS building classification model on an unseen location.

3.2 Training a building classifier

The goal of this work was to develop a classifier which could automatically recognize an SS building from a street view image. A convolutional neural network (CNN) was used as the classifier. Given an input image i, the network outputs a label with possibility p_i, indicating how confident the model is about the prediction. Since there were only two classes, SS and non-SS buildings, a cross-entropy loss was used for training the model:

$$L(c,\, p) \;=\; -\frac{1}{N}\sum_{i=0}^{N}(c_i \log p_i + (1-c_i)\log(1-p_i)) \quad (1)$$

where p_i and c_i are, respectively, the predicted probability and the ground truth label. The CNN model was first pre-trained on ImageNet (Deng *et al*., 2009) and then fine-tuned using the collected building data sets on the cities selected for the training purpose. Details of the fine tuning process are discussed in the experiment studies.

3.3 Visualization for interpretation of predictive results

In addition to the labelling of images and the confidence score for model training and prediction, a visualization module was developed to enhance users' understanding of how the CNN model makes classification decisions. As indicated in Zhou *et al*. (2016), a class activation map obtained from a convolutional layer (normally the last convolutional layer in the network) can be used to interpret the prediction results.

Given an input image i, the class activation map can be constructed as:

$$\mathrm{Map}_c = w_k^c f_k(x, y) \quad (2)$$

where c refers to a class, i.e., soft-story or non-soft-story building; w_k^c are the learned weights, indicating the contribution of the channel k to class c; and $f_k(x, y)$ represents the activation of k-th filter at location (x, y).

Map_c, which in essence is a linear combination of the activation maps obtained by each filter, is capable of showing the regions triggering the prediction of class c.

Table 1 Statistics of the building database

City	Images	Annotation method
Santa Monica	16,665	Automatic
Oakland	1,359	Automatic
Berkeley & San Jose	395	Manual
San Francisco	6,921	Raw

4 Experimental studies and results

This section describes the experiments conducted on the data collected from the five cities in California. First, the data sets from Santa Monica and Oakland are used to illustrate the machine learning approach for developing the soft story building classification model. The implementation details and the different CNN architectures tested are discussed. Generalization of the classifier is then tested using the manually labelled data set for the cities of San Jose and Berkeley. To illustrate the potential application of the machine learning model to a "new" city with raw and unlabelled data, the classifier is then applied to the data set for San Francisco. Lastly, to illustrate the application of the approach, the machine learning model is incorporated in a regional seismic vulnerability assessment framework to predict the regional distribution of soft story buildings, using the city of Oakland as an illustrative example.

4.1 Model development of soft story building classifier

4.1.1 Data selection

As discussed in the previous section, the first task for building a predictive model is to evaluate and select of the data for training the model. As shown in Table 1, the data set collected from Santa Monica, had far more non-SS than SS building images, which poses a data imbalance problem that may lead to bias. To address this problem, random images were drawn from the non-SS class to create a training data set with 3,203 SS and 3,921 non-SS buildings. On the other hand, the data collected from Oakland was well balanced with, respectively, 717 SS and 642 non-SS buildings. For each data set, as shown in Table 2, the data used for training and testing was split at a ratio of 9:1.

4.1.2 Implementation

Four well established convolutional neural networks, namely, InceptionV3 (Szegedy *et al*., 2016), InceptionV4 (Szegedy *et al*., 2017), ResNet50 (He *et al*., 2016), and ResNet152 (He *et al*., 2016), were employed in this experimental study. The four CNNs have different network architectures, each with a different number of layers (depth), and they have been applied and achieved excellent results on the publicly available ImageNet dataset. As discussed in Section 3, for each CNN, the initial model pretrained using the ImageNet data was adopted. The model was then fine-tuned using the SS building training data sets. During the fine-tuning, the number of output classifiers was changed from 1000 to 2 (for either SS or non-SS building type). Furthermore, a step training strategy was adopted to first fine tune the last fully connected layer while keeping all prior layers frozen and then fine tune all layers. This training strategy helped speed up the convergence.

When training a model using InceptionV3/V4, the initial learning rate and momentum are set to be 0.01 and 0.0004, and the RMSProp optimizer is used with weight decay 0.00004. When training ResNet50/152, the initial learning rate was set to be 0.001, and the Adam optimizer was employed with a weight decay of 0.0001. For training a model, each batch contained 64 images, and all input images were first resized to 256×256 and then randomly cropped to the size of 224×224. During training, the authors first fine-tuned the final fully connected layer for 5000 iterations and then fine-tuned all layers for another 40,000 iterations. All the experiments were implemented with Tensorflow and conducted on NVIDIA Titan Xp GPUs.

4.2 Performance results and analysis

4.2.1 Evaluation metrics

In this work, the CNN classification models were evaluated using two metrics, namely average accuracy and F1 score. Average accuracy can be obtained straight forwardly by averaging the classification accuracies of each class. However, this measure is likely biased towards the class with more training data and does not reflect the real performance of the model, particularly when the distribution of the data is imbalanced among the output classes. The F1 score, also known as F-measure, was calculated based on precision P and recall R as follows:

$$\mathrm{F1} = 2 \times P \times R/(P+R) \tag{3}$$

The F-measure avoids class bias and thus provides a better measurement for comparing the different models. The performance of all the CNN learning models are reported using these two metrics.

Tables 3 and 4 show the performance of the four CNN architectures obtained on the data sets from Santa Monica and Oakland, respectively. The average accuracy results were obtained for doing a single-crop of the image from a size of 256×256 to 224×224 during testing, where multi-crops may further boost the performance by 1%–2%. From the results, the following can be observed:

1. Good performances were achieved for all four CNN models. However, the performances of the models were different for the two data sets. As shown in the tables, ResNet50 performed the best on the Santa Monica

Table 2 The training/test splits of the data sets

City	# SS	# non-SS	# train	# test
Santa Monica	3,203	3,921	6,421	712
Oakland	717	642	1,224	135
Berkeley & San Jose	198	197	–	395

Table 3 Performance of four networks on Santa Monica data set

Model	Average acc.	*P*	*R*	F1
ResNet50	85.94%	84.16%	82.80%	0.8347
ResNet152	85.03%	82.32%	83.12%	0.8271
InceptionV3	84.38%	81.39%	83.77%	0.8256
InceptionV4	83.20%	80.52%	80.52%	0.8052

Table 4 Performance of four networks on Oakland data set

Model	Average acc.	*P*	*R*	F1
ResNet50	82.29%	81.54 %	82.81%	0.8217
ResNet152	79.69%	77.94%	82.81%	0.8030
InceptionV3	80.21%	80.65%	78.13%	0.7937
InceptionV4	84.38%	82.81%	82.81%	0.8281

Fig. 6 Examples of correctly classified building images. Top: SS buildings; Bottom: non-SS buildings

data set, while InceptionV4 showed the best performance on the Oakland data set;

2. A deeper network does not always outperform a shallower counterpart, as shown in the results for ResNet50 vs. ResNet152.

3. ResNet50, ResNet152, and InceptionV3 performed better on the Santa Monica data set than on the Oakland data set, while InceptionV4 had opposite results.

4.2.2 Effect of semi-automatic labeling strategy

As discussed in Section 3, a semi-automatic labelling strategy was used to annotate the 18,024 images from Santa Monica and Oakland. Initially, the labelled images included only 1,302 images from Santa Monica. With these initial labelled images, a preliminary classifier was trained and applied to annotate the rest of the images (of about 6× the labeled data). More training data, although being "noisy" as produced from the preliminary classifier, are expected to lead to a better-performing classifier. To assess the enhancement from the semi-automatic labelling strategy, the performance of the preliminary classifier built based on the manually labelled 1,302 images was compared with the final classifier, as presented, with the 7,124 images labelled for the Santa Monica data set. The two classifiers, both using ResNet50 as the backbone model, were tested on the Berkeley and San Jose data sets. As shown in Table 6, the final classifier, taking advantage of the labelled images from the semi-automatic strategy, achieved higher accuracy and F1 score than the preliminary classifier built using the manually labelled 1,302 images. This result shows the effectiveness of the labelling strategy as well as the benefits of using a larger set of training data (even possibly with noise).

4.2.3 Visualizations of predictive results

As discussed in Section 3, an analysis module was developed to visualize the class activation maps and to enhance interpretation of the predictive results with visual evidence. Figure 7 shows several examples for the visualization of class activation maps.

On the top row of Fig. 7, the highlighted parts indicate the regions where the CNN model was used to predict an SS building. In the middle row of the figure, the highlighted regions show the contributions that the model used for the prediction of non-SS class. It can be seen that, when conducting the classification task, the model learned to look for a building on an image, regardless of the size and location of the building. Interestingly, when deciding if an image contained a non-SS building, the model also attended to trees and roads (as illustrated in the leftmost image in the middle row of the figure). This result may be due to the fact that in the data set, many of non-SS building images have heavy occlusions such as trees and cars, such that the model may have taken these visual cues into account when classifying a building. The occlusions contributed to the noise of the data, which is a common issue in machine learning. One of the most efficient ways to remove noise is to manually clean the data, which is recommended when funding is sufficient.

Apart from helping to interpret the decision made by a CNN model, the analysis module can also be used to diagnose and improve the model. The bottom row of Fig. 7 shows the case examples for misclassifications from the CNN model where the buildings in the images are non-SS buildings but are mistakenly classified as SS buildings. The visualizations of the activation maps indicate that the prediction of SS buildings is triggered

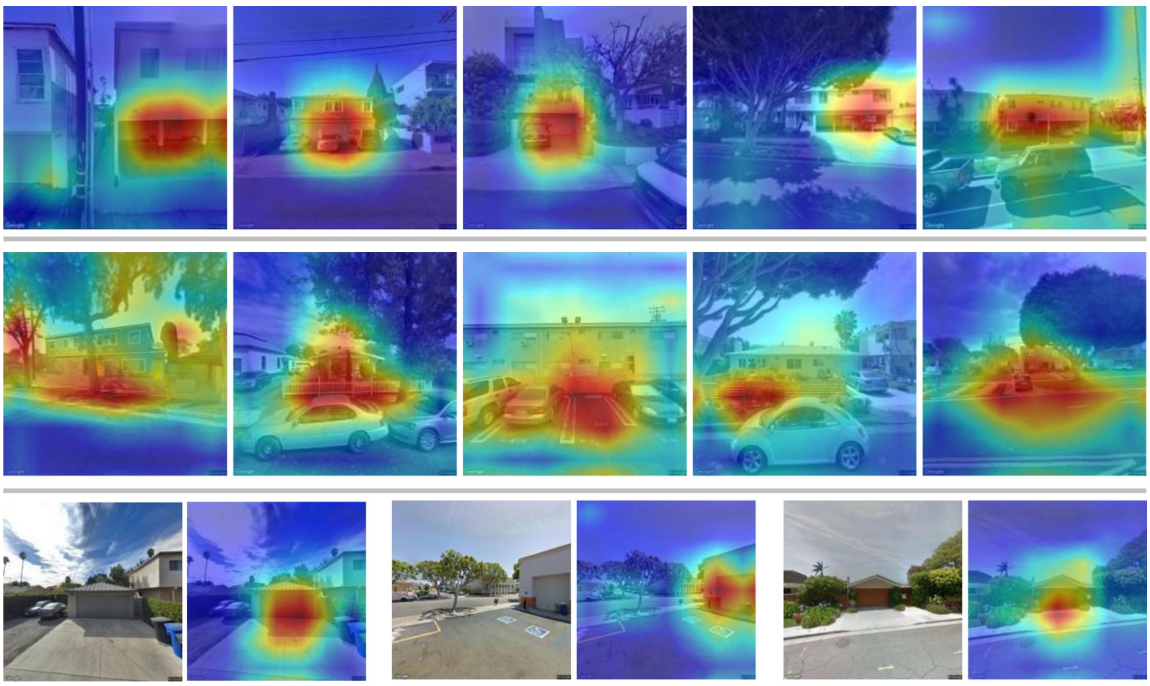


Fig. 7 Visualizations generated by the analysis module. First and Second row: correctly classified SS and non-SS buildings; Bottom row: example images which are misclassified as soft-story buildings and their corresponding class activation maps. All visualizations are generated by the ResNet50 model trained on Santa Monica subset

by the garage regions, suggesting that the model may have mistaken the existing of a garage as a sufficient condition. To solve this problem, it may be necessary to broaden the training data set to include a more diverse set of images that enable the model to take into consideration both the holistic structure and the local details of a building.

4.3 Generalization and applications

4.3.1 Generalization of SS building classification models

To test the generalization ability of the classification models, the models trained on Santa Monica were directly applied to classify the manually labelled (ground truth) images for Berkeley and San Jose data sets *without* any fine-tuning or modification. As shown in Table 5, the models performed very well even with the unseen image data from other cities.

To further illustrate generalization and application of the SS building classification model, the best model with the Santa Monica data set, ResNet50, was tested on the images collected in San Francisco. Since the images are not labelled, the evaluation metrics are not applicable for analyzing the results. Here, some classification results of SS buildings are shown in Fig. 8. Due to its unique topography, the SS buildings in San Francisco look quite different from those of other cities. For example, SS buildings in SF often have conjoined terraces, such as the three images shown on the first row of Fig. 8.

The proposed framework for SS building classification can find other potential applications for rapid regional screening and seismic vulnerability evaluation of SS buildings. As discussed in the introduction, one of the objectives of this study is to automatically recognize an SS building from street view images, thereby potentially facilitating the screening process of building reinforcement projects. When city-wide images are available, the framework can efficiently process the large number of images, and the prediction results can then be used to generate a SS distribution map for the city. Knowing the distribution of seismic-sensitive buildings in a city provides valuable information for local government agencies to plan seismic reinforcement work.

Take the city of Oakland as an example. Oakland is a major port city in California, located in an area prone to seismic hazards. Using the 1,359 street view images captured for the city and the prediction results on SS buildings from the CNN model, a distribution map of SS buildings can be generated using the SimCenter Uncertainty Research Framework, SURF (Wang, 2019), which is designed and developed to analyze the spatial patterns of geo-tagged data sets based on random field

Correctly classified SS buildings

Correctly classified non-SS buildings (bad angle images included)

Fig. 8 Classification results on San Francisco data set

theory and machine learning (Wang *et al*., 2017; Wang and Chen, 2018). Figure 9 shows a heat map produced to display which areas of the city are likely to be occupied by SS buildings.

5 Summary and discussion

In summary, a framework to automatically identify soft-story buildings is proposed. Traditional approaches for seismic screening of soft story buildings is labor intensive and time consuming. One novelty of this study is that it takes advantage of DL techniques to automate the seismic screening process, which is revolutionary because it doesn't require humans in the loop compared with the traditional approach. Though it does not intend to replace detailed evaluation results from numerical analysis, such as incremental dynamic analysis (IDA) and fragility analysis, the methodology can hopefully help reduce the burden of government agencies and residents during the screening step of the seismic vulnerability assessment process. The approach is to take advantage of publicly available street view images and advances in deep learning. The framework presented in this study can potentially be used for large scale rapid screening of soft story buildings using publicly available street view images. Furthermore, an effective semi-automatic data labeling strategy has been presented to alleviate the laborious and time consuming image annotation effort. Last but not least, the framework is demonstrated with raw street images and the results can potentially find many applications to enhance the assessment and planning for regional seismic vulnerability evaluation.

Both quantitative and qualitative results have been presented to demonstrate the effectiveness of the data collection and machine learning framework. To further enhance the methodology, several issues need to be addressed in future work:

1. Data Bias. Currently, the trained model takes garage as a sufficient condition to classify a building as a soft-story structure. For example, Fig. 7 shows an image with a single garage and the building is thus predicted as having a soft story. The prediction is probably caused by a data bias problem that a large number of SS buildings in the training data set have garages. To address this problem, in addition to increasing the diversity of buildings within different characteristics in the training data, other approaches, such as hard mining, that dynamically select hard data to train the model can be introduced to help improve the performance.

Table 5 Performance of four networks on Berkeley/San Jose data set

Model	Average acc.	*P*	*R*	F1
ResNet50	86.61%	84.26 %	89.34%	0.8670
ResNet152	83.26%	80.37%	87.31%	0.8370
InceptionV3	87.72%	84.26%	92.39%	0.8814
InceptionV4	83.93%	81.25%	91.37%	0.8601

Table 6 Effect of the semi-automatic labeling strategy

Model	Average acc.	*P*	*R*	F1
Preliminary	85.06%	85.57%	84.26%	0.8491
Final	86.61%	84.26 %	89.34%	0.8670

The performance was obtained on Berkeley and San Jose data sets; both classifiers used the ResNet-50 as the backbone.

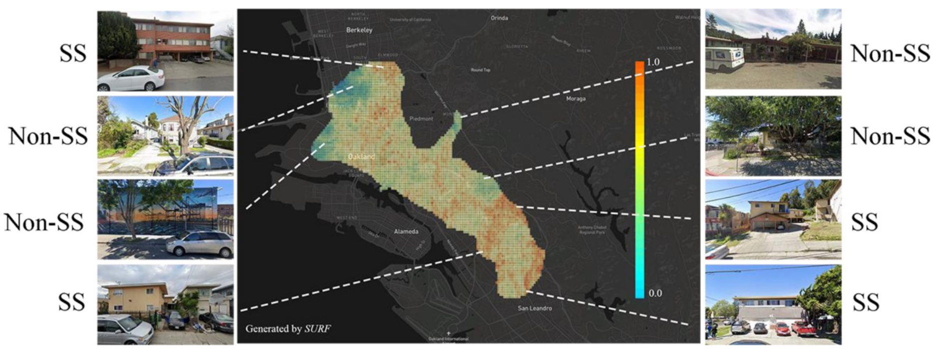


Fig. 9 Application example: predicted soft-story building distribution in Oakland. This distribution map is generated based on the outputs of the SS building classification framework. The value ranging from 0.0 to 1.0 indicate the probability of being identified as a soft-story building

Fig. 10 Problems to be addressed in the future. Top: Example of problem images: occlusions (1st) and multi-buildings (2nd and 3rd). In the 3rd image, the left building is actually a soft-story; however, its opening space is located on the other side of the building. Bottom: Example images of a retrofitted soft-story. As shown in the 3rd image, some reinforced structure is added inside the building, which may not be captured by street view images. The 1st and 2nd images are from the official website of the LA Department of Building and Safety, and the 3rd is from a California Foundation Repair Contractor

2. Data Noise. As shown in the top row of Fig. 10, many of the images collected show only part of a building, have no buildings, or contain several buildings. The images are "noisy" due to occlusions or being captured from an improper viewpoint. To address this problem, an object detection module can be introduced to first locate the existence of a building before feeding the image into the classifier.

3. Intrinsic limitations of street view images. Street view images used in this work only capture the front side of a building. However, many key visual cues, such as large open space, may be present on the other sides of the building that are not captured on the images. Furthermore, as shown in the bottom row of Fig. 10, it is difficult to detect any structural reinforcements located inside the building that cannot be viewed from the outdoor street scenes. This limitation is inherent to the use of publicly available street images. One solution is to allow images taken from multiple sides or inside the building, uploaded, for example, by the building occupants. Accordingly, the classifier can be trained using a multi-branch architecture, where each branch is designed to classify images captured from a specific view and then the results of all branches are fused to get the final classification result.

Acknowledgement

This study is based upon work supported by the US National Science Foundation under Grant No. 1612843. NHERI DesignSafe (Rathje *et al*., 2017) and Texas Advanced Computing Center (TACC) are acknowledged for the generous allotment of compute resources.

References

ATC (1988), "Rapid Visual Screening of Buildings for Potential Seismic Hazards: A Handbook, FEMA 154," *Federal Emergency Management Agency, Washington, DC*, *USA*.

ATC (2002), "Rapid Visual Screening of Buildings for Potential Seismic Hazards: A Handbook, FEMA 154, second edition," *Applied Technology Council, National Earthquakes Hazards Reduction Program*, *USA*.

ATC (2015), "Rapid Visual Screening of Buildings for Potential Seismic Hazards: A Handbook, FEMA 154, third edition," *Federal Emergency Management Agency, Washington DC*, *USA*.

Bency AJ, Rallapalli S, Ganti RK, Srivatsa M and Manjunath B (2017), "Beyond Spatial Auto-Regressive Models: Predicting Housing Prices with Satellite Imagery," *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*.

Deng J, Dong W, Socher R, Li LJ, Li K and FeiFei L (2009), "ImageNet: A Large-Scale Hierarchical Image Database," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Gebru T, Krause J, Wang Y, Chen D, Deng J, Aiden EL and FeiFei L (2017), "Using Deep Learning and Google Street View to Estimate the Demographic Makeup of Neighborhoods Across the United States," *Proceedings of the National Academy of Sciences*, **114**(50): 13108–13113.

Girshick R (2015), "Fast r-cnn," *Proceedings of the IEEE International Conference on Computer Vision*.

He K, Zhang X, Ren S and Sun J (2016), "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Kang J, Körner M, Wang Y, Taubenböck H and Zhu XX (2018), "Building Instance Classification Using Street View Images," *ISPRS Journal of Photogrammetry and Remote Sensing*, **145**: 44–59.

Karbassi A and Nollet M (2007), "The Adaptation of the FEMA 154 Methodology for the Rapid Visual Screening of Existing Buildings in Accordance with Nbcc-2005," *Proceedings of the 9th Canadian Conference on Earthquake Engineering*, Ottawa, Ont, 27–29.

Law S, Paige B and Russell C (2018), "Take a Look Around: Using Street View and Satellite Images to Estimate House Prices," *arXiv Preprint arXiv:1807.07155*.

Liu X, Chen Q, Zhu L, Xu Y and Lin L (2017). "Place-Centric Visual Urban Perception with Deep Multi-Instance Regression," *Proceedings of the 25th ACM International Conference on Multimedia*.

Naik N, Philipoom J, Raskar R and Hidalgo C (2014), "Streetscore-Predicting the Perceived Safety of One Million Streetscapes," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*.

Ningthoujam M and Nanda RP (2018), "Rapid Visual Screening Procedure of Existing Building Based on Statistical Analysis," *International Journal of Disaster Risk Reduction*, **28**: 720–730.

Perrone D, Aiello MA, Pecce M and Rossi F (2015), "Rapid Visual Screening for Seismic evaluation of RC Hospital Buildings," *Structures*, Vol. 3, Elsevier, 57–70.

Ploeger S, Sawada M, Elsabbagh A, Saatcioglu M, Nastev M and Rosetti E (2016), "Urban Rat: New Tool for Virtual and Site-Specific Mobile Rapid Data Collection for Seismic Risk Assessment," *Journal of Computing in Civil Engineering*, **30**(2): 04015006.

Rathje EM, Dawson C, Padgett JE, Pinelli JP, Stanzione D, Adair A, Arduino P, Brandenberg S J, Cockerill T, Dey C, *et al*. (2017), "Designsafe: New Cyberinfrastructure for Natural Hazards Engineering," *Natural Hazards Review*, **18**(3): 06017001.

Ren S, He K, Girshick R and Sun J (2015), "Faster R-Cnn: Towards Real-Time Object Detection with Region Proposal Networks," *Advances in Neural Information Processing Systems*.

Saatcioglu M, Shooshtari M and Foo S (2013), "Seismic Screening of Buildings Based on the 2010 National Building Code of Canada," *Canadian Journal of Civil Engineering*, **40**(5): 483–498.

Srikanth T, Kumar RP, Singh AP, Rastogi BK and Kumar S (2010), "Earthquake Vulnerability Assessment of Existing Buildings in Gandhidham and Adipur Cities Kachchh, Gujarat (India)," *European Journal of Scientific Research*, **41**(3): 336–353.

Sun Y, Chen Y, Wang X and Tang X (2014), "Deep Learning Face Representation by Joint Identification-Verification," *Advances in Neural Information Processing Systems*.

Szegedy C, Ioffe S, Vanhoucke V and Alemi AA (2017), "Inception-V4, Inception-Resnet and the Impact of Residual Connections on Learning," *Proceedings of the 31st AAAI Conference on Artificial Intelligence*.

Szegedy C, Vanhoucke V, Ioffe S, Shlens J and Wojna Z (2016), "Rethinking the Inception Architecture for Computer Vision," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Wallace NM and Miller TH (2008), "Seismic Screening of Public Facilities in Oregon's Western Counties," *Practice Periodical on Structural Design and Construction*, **13**(4): 189–197.

Wang C (2019). "NHERI-SimCenter/SURF: v0.2.0," <https://doi.org/10.5281/zenodo.3463676> (September).

Wang C and Chen Q (2018), "A Hybrid Geotechnical and Geological Data-Based Framework for Multiscale Regional Liquefaction Hazard Mapping," *Géotechnique*, **68**(7): 614–625.

Wang C, Chen Q, Shen M and Juang CH (2017), "On the Spatial Variability of Cpt-Based Geotechnical Parameters for Regional Liquefaction Evaluation," *Soil Dynamics and Earthquake Engineering*, **95**: 153–166.

Zhou B, Khosla A, Lapedriza A, Oliva A and Torralba A (2016), "Learning Deep Features for Discriminative Localization," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.