Postevent Reconnaissance Image Documentation Using Automated Classification

Chul Min Yeum[1]; Shirley J. Dyke[2]; Bedrich Benes[3]; Thomas Hacker[4]; Julio Ramirez, M.ASCE[5]; Alana Lund[6]; and Santiago Pujol, M.ASCE[7]

Abstract: Reconnaissance teams are charged with collecting perishable data after a natural disaster. In the field, these engineers typically record their observations through images. Each team takes many views of both exterior and interior buildings and frequently collects associated metadata that reflect information represented in images, such as global positioning system (GPS) devices, structural drawings, timestamp, and measurements. Large quantities of images with a wide variety of contents are collected. The window of opportunity is short, and engineers need to provide accurate and rich descriptions of such images before the details are forgotten. In this paper, an automated approach is developed to organize and document such scientific information in an efficient and rapid manner. Deep convolutional neural network algorithms were successfully implemented to extract robust features of key visual contents in the images. A schema is designed based on the realistic needs of field teams examining buildings. A significant number of images collected from past earthquakes were used to train robust classifiers to automatically classify the images. The classifiers and associated schema were used to automatically generate individual reports for buildings. **DOI: 10.1061/(ASCE)CF.1943-5509.0001253.**

Introduction

Societal needs demand that attention is directed toward developing a built environment that is resilient and sustainable. To achieve this goal, actual performance must be compared systematically with design (expected) performance under extreme events such as earthquakes, windstorms, and associated hazards such as tsunami and storm surges in coastal areas. There is strong agreement in the earthquake engineering community that researchers need to learn more than they currently do from past earthquakes and windstorms. In the United States, the National Science Foundation recently funded a novel facility within the Natural Hazards Engineering Research Infrastructure (NHERI) network that is dedicated to supporting reconnaissance teams as they collect such field data (DesignSafe-CI 2016). The NHERI network has established a long-term science plan that highlights the crucial role of image information for use in disaster planning, mitigation, response, and recovery (DesignSafe-CI 2016).

The built environment is the most realistic laboratory that exists. Field observations provide tangible evidence in the form of the actual consequence of extreme conditions on the civil infrastructure and enable the documentation of its performance. These data are essential; they are meant to be used by researchers, practitioners, and students to identify gaps in the practical application of science and engineering knowledge together with economics and policy in providing evidence to motivate future research needed to improve the performance of infrastructure. Critical information related to structural performance is contained in those data centered around damaged structures and their components. Lessons learned from structures that were not damaged by an event are just as important.

In a typical mission to survey damage from earthquakes or windstorms, a group of engineers and scientists are dispatched to an area in which an event took place. To help empower a well-organized team, advance information about local construction, event severity, and regional maps are made available to facilitate planning. Further, it is recommended that those surveying the damage do so with a hypothesis to be tested with the evidence of the damage in order to focus observations. Often, the larger group is divided into small teams, each with experienced structural engineering evaluators.

In examining building-related damage, each team visits several buildings over the course of a day, collecting images and measurements from each building. The procedures employed by the team may be based on ATC-20 (for earthquakes) and ATC-45 (for flood and windstorms), which are designed for rapid structural evaluation after extreme events (ATC 1989, 2004). Recently, there has been an ongoing discussion in ACI-133 to establish the procedure for gathering information from major disasters and reporting their effects on concrete construction (ACI 2017). An enormous volume of images can be generated over just a few days. At the end of each day, evaluation teams regroup at a base station to share findings and to discuss next-day plans for expending time and effort to collect additional data.

Photographs collected by engineers are the most appropriate and widely used way to record findings and observations in the field. Meaningful scenes are captured from unique viewpoints and

[1]Professor, Dept. of Civil and Environmental Engineering, Univ. of Waterloo, Waterloo, ON, Canada N2L 3G1 (corresponding author). Email: cmyeum@uwaterloo.ca

[2]Professor, School of Mechanical Engineering, Purdue Univ., West Lafayette, IN 47907.

[3]Professor, Dept. of Computer Graphics Technology, Purdue Univ., West Lafayette, IN 47907.

[4]Professor, Computer and Information Technology, Purdue Univ., West Lafayette, IN 47907.

[5]Professor, Lyles School of Civil Engineering, Purdue Univ., West Lafayette, IN 47907.

[6]Graduate Student, Lyles School of Civil Engineering, Purdue Univ., West Lafayette, IN 47907.

[7]Professor, Lyles School of Civil Engineering, Purdue Univ., West Lafayette, IN 47907.

Note. This manuscript was submitted on February 14, 2018; approved on August 7, 2018; published online on December 5, 2018. Discussion period open until May 5, 2019; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Performance of Constructed Facilities***, © ASCE, ISSN 0887-3828.

locations that can be helpful to recall and confirm them in the future. Most images are focused on the visual appearances of the damaged/undamaged buildings and their components, although typical data sets also include other photographs taken by the teams, such as those of structural drawings or global positioning system (GPS) devices. These photographs are integral to understanding the data because they represent types of metadata that are needed for documentation of the data set. Because photographs are sequentially collected throughout a building, valuable temporal and spatial information is also contained in a data set. For example, an image of a structural component taken together with a measuring tape is often used to document the appearance of the target component as well as to record the measurement. Further, an observation associated with images taken on a higher floor of a building may trigger the need for additional images to subsequently be collected at the ground floor of the building.

In the field, a major challenge of dealing with such images is the limited time available to document and organize the data. Building reconnaissance teams may have anywhere from just a few minutes to a few hours to make decisions regarding where to next allocate teams and resources. Additionally, certain images require attached explanations of the purpose and intent behind them. Immediate documentation is critical before the important details are forgotten. With the large volume and wide variety of photographs included in these data sets, manual sorting these images would be tedious and time-consuming. Consider a case in which all images are automatically categorized and organized into useful image categories, and the metadata recorded in the images are extracted and accessible. Then, all extracted and processed information is displayed in the form of a report and can be converted to a shareable (e.g., web link or SharePoint in Microsoft) and storable format [e.g., portable document format (PDF), Microsoft Word, or Markdown]. This automated capability can save a great amount of time and effort for reconnaissance teams in the field.

Previous research developed a fundamental computational method and associated schema to enable automatic extraction and analysis of the metadata and visual contents of such data sets (Yeum et al. 2017a, b, c). Such efforts are intended to support teams in the field by automatically classifying those data and displaying them in a favorable way to streamline the reconnaissance process. In the present study, deep learning using a convolutional neural network (CNN) is implemented and used to extract distinct features for categorizing the images according to a predetermined set of appropriate classes. Using large-scale real-world image data sets, robust image classifiers are trained for each class in the schema. Additional information including date, time, or GPS location is extracted from the exchangeable image file format (EXIF) for each image. The extracted information is systematically integrated and organized to enable rapid generation of reports according to a suitable digital format so that engineers can easily access and describe the images.

This automated report generation capability is intended to provide rapid decision support for postevent reconnaissance field investigations, as well as documentation for future data reuse. With this capability, engineers can browse through the report for a given building and rapidly find and examine particular images to identify an important image of interest. To demonstrate the technology and its capabilities, the authors first build a large-scale ground-truth image database to train the classifiers of the classes in the schema that are used for generating the report. This database contains real images collected during past earthquake reconnaissance missions from around the world. According to the schema developed to support earthquake report generation, multiclass and binary classifications are combined to support the class hierarchy by arranging distinctive image classes in each training scenario. Using these trained classifiers, sample reports are automatically generated from large sets of images collected from buildings damaged in the Ecuador and Taiwan earthquakes in 2016 (Sim et al. 2016; Purdue University and NCREE 2016).

The merit of this study is that the authors provide a practical and feasible solution to support actual field engineers in tasks that could not be automated without this capability. The authors do not collect new images for the purpose of validation in either training classifiers or evaluating them. Such data sets would be highly biased toward the quality of the data when they are used for the validation and would provide no assurance that the resulting classifiers would work with real-world images. Instead, the authors use real-world sets incorporating large numbers of complex and unstructured images. The images were collected from several previous earthquake events and contain hundreds of buildings from many regions and with varied appearances. Thus, the developed solution is unbiased, trustworthy, and practical, and can be immediately deployed during actual reconnaissance missions in the future.

Overview of the Approach

This automated approach was developed based on the challenges the authors faced during postdisaster reconnaissance data collection. Primary among these challenges was the lack of time. To survey the maximum number of structures within the time available for reconnaissance, data synthesis and planning meetings would often stretch far into the night. The toll of these long hours, usually in a new climate and time zone, impacted the quality of data compilation. Manually processing the content of a large number of images collected took hours of checking and rechecking among team members. With an automated compilation system, teams could immediately determine whether all the necessary data have been collected and rapidly plan for the next phase of the mission.

The concept driving the proposed approach to overcome this real challenge in the field is introduced in Fig. 1. To directly support efficient and comprehensive data collection in a region, the authors aim to establish the ability to rapidly gather and provide some organization and structure to the data as they are collected. Visual content and metadata are fully extracted for automatically classifying and cataloging the data collected. First, a large set of images is collected from several buildings by teams of field engineers. Although the main goal is to collect images from buildings that contain the visual appearance of damage, these image collections frequently will include metadata in the form of images such as photographs of drawings or measurements. In a short time, several hundred images may be collected from a single building.

Next, pretrained classifiers are applied to each image in the data collection to automatically analyze and extract the visual contents and metadata. A schema for organizing the images is designed in advance according to the specific needs of the domain and application. In this study, based on the general characteristics of prior reconnaissance data, the classes used to organize the data are designed according to two groups: building and building components (e.g., exterior or interior the building) and metadata records (e.g., GPS or drawing). Additionally, each of these groups may be further classified into subgroups. For instance, images of structural drawings may be separated from those of GPS images. Each image is then labeled with the results of the classification process. Finally, the extracted and classified information is organized into a report.

The authors have developed a format that is intended to directly support earthquake reconnaissance teams. The report begins with a brief description of the site and contents of the image collection,

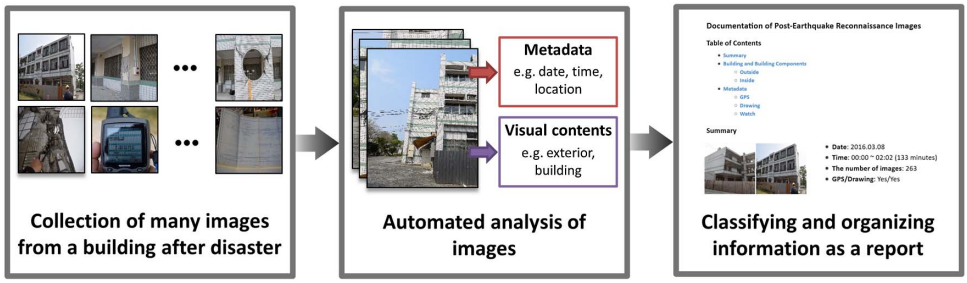


Fig. 1. Overview of the approach. Large volumes of images collected during postevent reconnaissance missions are automatically transformed into a well-organized report based on their visual contents and metadata. (Images from Shah et al. 2015; Sim et al. 2015, 2016; Purdue University and NCREE 2016.)

with a small set of representative images showing the overview of the appearance of the building. Beginning the report with such images helps the team to quickly recognize the subject building out of the many that are visited during the reconnaissance mission. Then, each of the images collected from that building is shown in the report within the classified categories. Images are shown in chronological order according to the time they are collected. The schema and associated format of this report could readily be modified for other types of extreme events or other purposes. This process typically requires less than a minute per building and rapidly turns the images into useful information for the field team.

Two sample reports will be provided subsequently in Fig. 7. With such reports, the team can efficiently review the image sets rather than looking through a massive collection of mixed and unstructured individual images. This report is automatically generated from a set of images with no manual processing required, although users can freely add comments to the images or report.

Image Classification Using Convolutional Neural Networks

CNN algorithms provide the foundation for developing the ability to extract and classify the visual contents of the disaster images. CNNs have recently mobilized researchers in many communities to pursue computer vision applications because they enable the learning of features (both deep and high-level) for image recognition using large-scale databases (LeCun et al. 1990; Krizhevsky et al. 2012). CNNs often use one or multiple convolutional layers linked with weights and pooling layers that work to extract translation, scale, and rotation-tolerant features. These CNNs with fully connected layers associated with these features can be used to classify images or identify categorical membership of objects. Conceptually, CNNs work by using many convolutional filters to identify the features that best describe the given images. CNNs are most successful when training classifiers using a larger set of images as a source, along with a large set of parameters. Graphical processing units (GPUs) can also be exploited to implement CNNs, and dropout regularization can also be used. Several CNN architectures have been described in the literature, and their accuracy has been greatly improved. Nonetheless, an active research area is in devising optimized network architectures along with the configuration of categories and input images for various application domains. However, when natural images are to be analyzed, it is clear that CNNs can provide exceptional performance (Russakovsky et al. 2015).

CNNs may be implemented in many ways and for various purposes. This study will focus on scene (image) classification to organize the visual data relevant to the specified application domain (Zhou et al. 2014; Karpathy et al. 2014; Johnson et al. 2015). In general, scenes can be described via a subject (e.g., fishing or auto racing) or a place (e.g., street or room) that generally indicates what is represented by the image. Scenes from images may be interpreted by the presence of one or more objects along with their spatial arrangement. Consequently, scenes are regarded as an indication of a class of an image. For example, a scene of a building can be interpreted as containing objects with a particular spatial arrangement. The objects may be the columns and walls, or a set of borders of windows or the floor. The process of classifying a scene can be accomplished through the use of coarse-resolution images because scenes can typically be recognized by low-dimensional features (such as colors, composition, or general shape). Thus, interpretation of these features does not require the detailed appearance of objects. For instance, although details are not visible in a thumbnail of an image, there is still sufficient information available to understand the scene of that image. In practice, scene classification is typically performed to extract the features that represent the class of the images. Thus, the concept of the scene can be interpreted by image-level classification.

An overview of the approach to image classification based on CNNs is provided in Fig. 2. Training the classifiers requires that ground-truth images are labeled according to the designed category ("Design of the Schema" section). For training classifiers, it is essential to establish a clear rule for defining each class of images to maximize visual discriminative boundaries among the images.

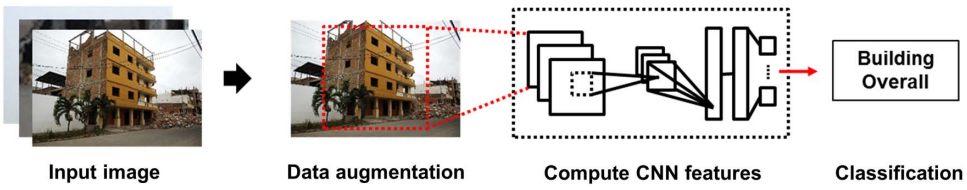


Fig. 2. Overview of image classification using a convolutional neural network. (Images from Shah et al. 2015; Sim et al. 2015, 2016; Purdue University and NCREE 2016.)

The authors implement both multiclass and two-class (binary) classification and use the same network structures except in the final layer for classification. The authors use a logistic loss layer for binary classification and a softmax layer for multiclass classification.

Typically, CNNs require low-resolution images (200–300 pixels per side) as input (Krizhevsky et al. 2012). Thus, the images may be resized or stretched, which depends on the original resolution of the image. Then, data augmentation is applied to the resized images to avoid overfitting with respect to translation, color, or light variations. Augmentation expands the set of training images without the need to add new images or expending additional effort in ground-truth labeling. For example, for Alexnet, which is used for this study, all the images were resized to 256×256 pixels, and random square regions of 224×224 pixels were used for exaction.

Once the inputs to be used for training the CNN were prepared, the algorithm sought to minimize the error in determining the correct (true positive) labels of the training images by automatically tuning a large set of parameters at the fully connected, convolutional, and pooling layers with the goal of extracting robust features. The stochastic gradient descent algorithm along with a defined batch size of a set of images were used to optimize the CNN parameters. Over each training epoch, an assigned batch at each iteration used randomly ordered images after augmentation. The training process typically required a time span of 2 h to several weeks. The training time depended on learning rates, number of input images, CNN network configuration (number of layers and neurons), and specification of the computer hardware used for training. A more detailed discussion of the implementation used in the experimental study is given in the "Configuration of Convolutional Neural Network" section.

Design of the Schema

Discussions were held with field engineers to develop appropriate categories and an associated hierarchical structure that support the needs of the application. The classes selected must be useful for the specific application, although classification can only proceed if images in the classes can be visually distinguished. A hierarchy can also be determined to enable efficiency and accuracy in the classification. Thus, visual contents in images are classified sequentially, reducing the amount of processing required and improving overall accuracy. For instance, only after an image is already classified as being related to buildings and their components would a classifier be applied to determine if that image should be classified as an exterior building or interior building image.

To successfully classify the images into the selected categories, a clear definition is needed for each category. Definitions are essential because they guide human annotators in establishing consistent and meaningful ground-truth data that are suitable for training. Additionally, the classification results will be more accurate when there are clear boundaries to distinguish the visual features of the images in different classes. Here, the authors describe the definition for each category used in the report. These definitions were used by human annotators in order to label the images used for the demonstration in the "Experimental Validation" section. A few sample images within each of the following classes are presented in Fig. 3:

- Building and building components (BBC): This category is applied to images that contain the visual appearance of a physical building and includes interior or exterior building images. Typical examples are images that contain either undamaged or damaged structural components, close-ups of damaged areas, or building components. This category includes images that contain measurements, typically of structural components or damage. This category does not include images taken to record metadata, such as drawings or GPS devices. Images in this category are also classified further (discussed subsequently).
- Overview (OV): Overview images are defined as those that represent the complete external appearance of the building, usually from a distance. These images contain basic information, for instance, allowing the users e.g., to estimate how many floors are present, observe the architectural type or structural type, approximate the age. Images in this category should include one entire side view and/or a front view of the building. Partial obstruction may exist, but about more than 70% of the building should be visible in the image. Images that contain the entire building façade, a diagonal view of the building, or a complete side of the building are included in this category.
- Building interior (BIN): This category is defined as images that have a general sense of inside space. For example, these images might show a space that is surrounded by walls or windows, or they might contain interior building components or a clearly indoor region (e.g., room, basement, or corridor).
- Building exterior (BEX): This category has the opposite meaning as the BIN category. These images show a space having no ceiling or surrounding walls.
- Measurement (MEAS): Images in this category are to record the absolute or relative size of the building components (e.g., column or rebar spacing) or damage (e.g., cracks or spalling). The size may be measured using a measuring tape, crack gauge, or fixed size objects (e.g., pen, hand, or note). The images are typically captured at a close distance to the target object so that the digits of the measurement tool are visible.
- WATCH: Images that focus on an analog or digital timepiece are commonly collected during reconnaissance missions to document the time associated with a particular stage in the data collection process. Images in this category are particularly useful when images are reformatted/resized and the EXIF information is lost. They are also useful when the date and time of the digital camera have not been corrected according to the local time zone.
- GPS: Images are collected frequently to document the GPS location values shown on a GPS device if they are not geotagged. As in the previous case, these are helpful in case the geotagging images are reformatted which often results in the elimination of the EXIF information.
- Document (DOC): This category is defined as an image primarily containing text and/or drawings. For example, structural drawings, or photos of signs or text documents are often collected, and these are included in this category. In many cases, digital versions of the structural drawings for the subject buildings do not exist. Photographs taken of the entire set of structural drawings provide valuable reference information that will enable a better understanding of the subject building.

These categories are designed to directly support reconnaissance missions. Expansion or modification of the categories defined here is possible, although it should be done according to the needs of a particular application and expected quality of the images.

With a large volume of images labeled within each class according to these definitions, the ground-truth trained the robust images classifiers. The classifiers can then be applied to past and future image collections having a visual appearance similar to those included in the training data. As in the implementation, the classifiers for each class are trained in a hierarchical order. Fig. 4 shows the class tree that was designed to classify the raw images into appropriate categories according to the hierarchy. Initially, a multiclass classifier is applied to all images in the collection to distinguish among BBC, GPS, WATCH, and DOC. Subsequently, those images classified as BBC are further classified into the subcategories included in the proposed schema. BBC images are categorized as

Fig. 3. Sample images in each of the designed classes: (a) building overview; (b) building interior; (c) building exterior; (d) measurement; (e) watch; (f) GPS; and (g) document. These labeled images were used for training classifiers to analyze the visual contents of the images. All images were rescaled as square thumbnails for convenient arrangement. (Images from Shah et al. 2015; Sim et al. 2015, 2016; Purdue University and NCREE 2016.)

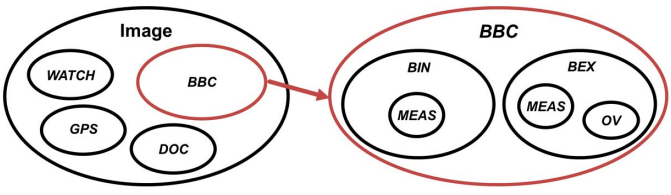


Fig. 4. Hierarchy of the classes in the schema.

either BIN and BEX. The BEX image class may include OV and MEAS images, but images in the BIN category include MEAS. Independent binary classifiers are used for this step to classify all of the BBC images into each of these subcategories. This approach suits this particular application because images not related to building appearance are largely eliminated from the BBC class at a prior level. For example, for classification of interior building images, the algorithm does not need to revisit the images that were not previously classified as BBC, such as GPS or DOC images. Thus, using such a hierarchical procedure results in high classification accuracy as well as efficient processing.

Report Generation

The report developed here is aimed to support rapidly searching for and finding important images of interest. Cluttered listings of images and their information may impose additional effort and time to analyze and review. Information extracted from the images should be

neatly compiled with the images in the report so that the users can readily recall important memories and experiences during the reconnaissance mission. To enable such capability in the generated report, the authors document the images with their information in the following ways (sample reports are shown in Fig. 7).

First, general information extracted from the images is placed at the beginning of the report. Such information plays a key role to make a quick judgment regarding whether the image set includes important images of interest. Text information (e.g., building name, address, or street number) correlated to each building that the engineers visited are not as intuitive and memorable as overview images, which represent a general view of the building. Thus, a couple of the images classified as OV with the highest classification scores are placed at the beginning of the report (summary section in the report in Fig. 7). In addition, some metadata information is also included for referencing the building, such as date and time or a number of images.

Next, the images are arranged according to their classes. Images classified as BBC (including BIN and BEX) and metadata (GPS, DOC, and WATCH) are arranged in the order shown in Fig. 7. Because photographs are sequentially collected throughout each building, the closeness in the capture time of the images means they may contain a similar spatial context. Thus, the images in each class are sorted in chronological order according to the time each is collected. MEAS images are typically captured at a close distance to target components so that the digits of the measurement tool are legible. These images require spatial context to identify the location of the measured component on the building. Although these images are intended to record the measurement values as metadata, they are simply marked in the classified BBC images rather than arranging them in a separate category (BBC images marked with a dark gray box in Fig. 7).

All images are arranged as a grid after being resized to an identical small square so that the engineers can rapidly look over the information in the report. However, each image is also linked to its raw (higher resolution) version for ready access. For instance, MEAS images, or images having small damage such as a crack, may require high resolution. The report is written using GitHub Flavored Markdown language (Github 2007) and can be easily converted into different types of formatted documents and directly upload to GitHub for sharing and versioning.

Experimental Validation

In this section, detailed procedures are described to demonstrate the approach on several complex real-world unstructured data sets from previous earthquake reconnaissance teams. Similar capabilities for other applications could readily be implemented with an application-appropriate well-defined schema and similar procedures.

Ground-Truth Labeling Using Postdisaster Image Database

Yeum (2016) and Yeum et al. (2016, 2017a, b, c) developed a large postevent reconnaissance image database for use in training classifiers for each category. An extensive collection of approximately 100,000 color images has been acquired by practitioners and researchers after past natural disasters. The database includes images from hurricane, tornado, and seismic events (e.g., from disaster responders, Purdue University datacenterhub.org, Canterbury Earthquake Digital Archive, and Earthquake Engineering Research Institute image collection) (UC CEISMIC 2011; EERI 2009; Datacenterhub 2014). However, to discuss this demonstration, the authors focus on earthquake images. Some sample collections included in the database are shown in Fig. 5 (Datacenterhub 2014). These images were collected from events such as earthquakes (e.g., Haiti in 2010, L'Aquila in 2009, Christchurch in 2011, Nepal in 2015, and Taiwan in 2016), hurricanes (e.g., Katrina in 2011, Florida in 2004, and Texas in 2008), and tornadoes (Florida in 2007 and Greensburg in 2007). To date, the images have been distributed according to the type of disaster: earthquake (90%), hurricane (5%), tornado (4%), and others (1%). The authors plan to continue collecting reconnaissance images to integrate into the database from natural disasters.

For this research, images collected from earthquake reconnaissance missions in Düzce, Turkey in 1999; Bingöl, Turkey in 2003; Peru in 2007; Haiti in 2010; Nepal in 2015; Taiwan in 2016; and Ecuador in 2016 (Shah et al. 2015; Sim et al. 2015, 2016; Purdue University and NCREE 2016) were used. Data sets from approximately 750 reinforced concrete or masonry buildings from these events are available, and much of the reconnaissance data for each building are documented, including, for instance, measurements of structural components and assigned building damage levels, in addition to the images of that building. Images from earthquake events in Turkey in 1999 and 2003 were digitized from film slides. In most cases, images were captured using digital cameras (from Peru in 2007), and thus a large number of images were readily collected from each building. The average number of images from each building is 44, although the number of images collected in each event is growing steadily as the cost of collecting and storing images becomes more affordable. All of these images are available to the public (Datacenterhub 2014). The authors labeled these images into the developed classes in the schema.

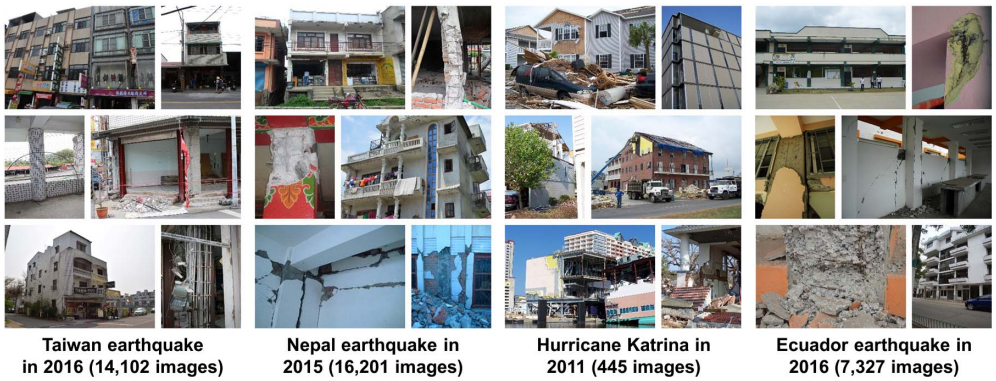


Fig. 5. Sample image collections from the authors' postdisaster image database. (Images from Shah et al. 2015; Sim et al. 2015, 2016; Purdue University and NCREE 2016.)

J. Perform. Constr. Facil., 2019, 33(1): 04018103

Table 1. Classification results for various image classes

	Multiclass				Binary		Binary	
Category	BBC	GPS	WATCH	DOC	BIN	BEX	OV	MEAS
Number of labeled images	16,747	835	320	3,283	6,407	9,650	1,531	690
Number of testing images	4,126	187	80	827	1,609	2,337	360	172
Precision (%)	99.7	93.8	86.4	97.6	82.2	90.8	50.9	37.4
Recall (%)	99.1	97.0	95.0	98.7	87.2	87.0	90.0	79.8

To generate the ground-truth training data, the authors designed a web-based annotation tool that enables multiple annotators to work together as a team to label a large number of images within a short period of time. The tool is designed to enable users to rapidly annotate images using only a keystroke. Annotators can also add comments, as needed, to indicate where there may be some uncertainty about the classification. Three annotators with training in civil engineering took part in the annotation of these images. Each labeled image was reviewed separately by at least two times by three annotators. When beginning, the annotators were given some time to learn each class category by being provided with samples of labeled images. If an annotator was uncertain whether an image met the definition of a certain class, those images could be skipped without being categorized.

A total of 21,185 images were labeled as belonging to the classes designed in this study. Table 1 provides the number of labeled images in each class. Recall that BBC images, as a parent category, include BIN, BEX, OV, and MEAS images. Also, BEX images include detailed images of the building exterior as well as OV images. Only 103 WATCH images were found in the developed database. Not all reconnaissance teams recorded time and date using images, so just a small number of these were available. However, because timepieces are everyday objects, the authors were able to supplement the training data for this class of images with images obtained through a Google image search. A combination of the keywords wrist and watch was used to search for these images, and 207 images were added to the ground-truth data sets for WATCH. Samples of the supplemental images are marked in gray in Fig. 3(e). For demonstrating the report generation capabilities in the "Sample Reports" section, the authors selected two sets of building images from the labeled data set, thus showing how the system would work for actual users who will collect new data in future reconnaissance missions. Because these image sets are being tested for report generation, they cannot be used for training and testing classifiers and were removed from consideration.

Configuration of Convolutional Neural Network

The authors implemented an ImageNet CNN model called Alexnet (TorontoNet in Caffe), a popular model that was framed in the MatConvNet library (Vedali and Lenc 2015). In the 2012 ImageNet image classification competition, Alexnet exhibited the implementation of CNNs for computer vision applications. Alexnet has been widely used for assessing the performance of new CNN models in benchmark tests (Krizhevsky et al. 2012). The Alexnet network architecture is discussed in detail in the literature (Krizhevsky et al. 2012). Alexnet was chosen here because it is an established method and a simple general CNN model to use for this research and demonstration. Since 2012, some improved architectures have also been proposed and may be tested in the future, although the results are not expected to be very sensitive to such variations.

To train a classifier in each category, the labeled images are first transformed into inputs for the CNNs. All images, regardless of the category, are isotopically (preserving their aspect ratio) resized so that the shorter side of each image is set to 256 pixels, and this setup is followed by cropping the center square region so that each image is reduced to a 256×256 (pixels) square. Data augmentation is applied by randomly cropping 227×227 patches from the 256×256 images in each epoch. The training set is further augmented by random horizontal image flipping. Additionally, random color shifting is applied to vary (jitter) the intensities of red, green, and blue (RBG) images.

The last 1,000-way softmax layer is modified from the original implementation used in the ImageNet competition to use a four-way softmax for multiclass classification and logistic for binary classification. The layers in the neural network are initialized using a Gaussian distribution with a zero mean and a variance equal to 0.1. The same hyperparameters used in Alexnet are applied in this implementation. The authors trained the models using stochastic gradient descent using a batch size of 512 images, momentum value of 0.9, and weight decay of 0.0005. The authors trained the network for 120 epochs, and the learning rate was logarithmically decreased from 10×10^{-2} to 10×10^{-5} (10×10^{-3} to 10×10^{-5} for the MEAS classifier) during training. A computer PC workstation containing a Xeon E5-2620 CPU and NVidia Titan X with a 12-GB video memory GPU was used for training and testing the algorithm.

The number of labeled images in the training data set was highly unbalanced because the number of images gathered to record metadata was much lower than the main focus of the data collection effort, which is the BBC images (MEAS images are categorized as BBC but their purposes are metadata recording). For instance, the number of BBC images in the training data set is around 50 times higher than the number of WATCH images. If training samples were assigned randomly, the classifier would tend to overfit to the majority of classes (herein, these are BBC and DOC). To sidestep this issue, the authors assigned the same number of images from each category to a batch. For example, for four-way multiclass classification, 128 images from each class are randomly selected when the batch size is 512. As a result, several repetitions of images in the minority classes (e.g., GPS or WATCH) are permitted to complete a batch of training data. Thus, various data augmentation methods were used. Two binary classifiers for identifying OV and MEAS were also trained after balancing the number of images (samples) in each batch.

To train the classifiers and then test their capabilities, all labeled image sets are divided into groups as 50%, 25%, and 25% for training, validating, and testing, respectively. However, for demonstrating report generation, the authors assign 75% of these for training and 25% for validation to train each classifier. This ratio is selected so that many more images are involved in the actual training of classifiers while the testing set was removed for evaluating the performance of each classifier. As mentioned previously, the images from the two sets of images to be used for demonstrating the report generation ("Sample Reports" section) are not included when training the classifiers.

Classification Results

The results from the one multiclass classification and the three binary classifications are summarized in Table 1. Based on the

Table 2. Data sets used for generating the sample reports

Identifier	Event	Date	Latitude	Longitude	Number of images	Structural damage	Masonry wall damage
Set 1	Ecuador earthquake	July 16, 2016	−0.594472222	−80.42313889	93	Moderate	Severe
Set 2	Taiwan earthquake	March 8, 2016	23.12341667	120.4699444	263	Severe	Moderate

(a)

(b)

Fig. 6. Two sets of the original images collected from the actual earthquake reconnaissance missions: (a) 93 images from Ecuador, 2016; and (b) 263 images from Taiwan, 2016. Images are ordered based on the time when each image was taken and are rescaled as square thumbnails for convenience. (Images from Sim et al. 2016; Purdue University and NCREE 2016.)

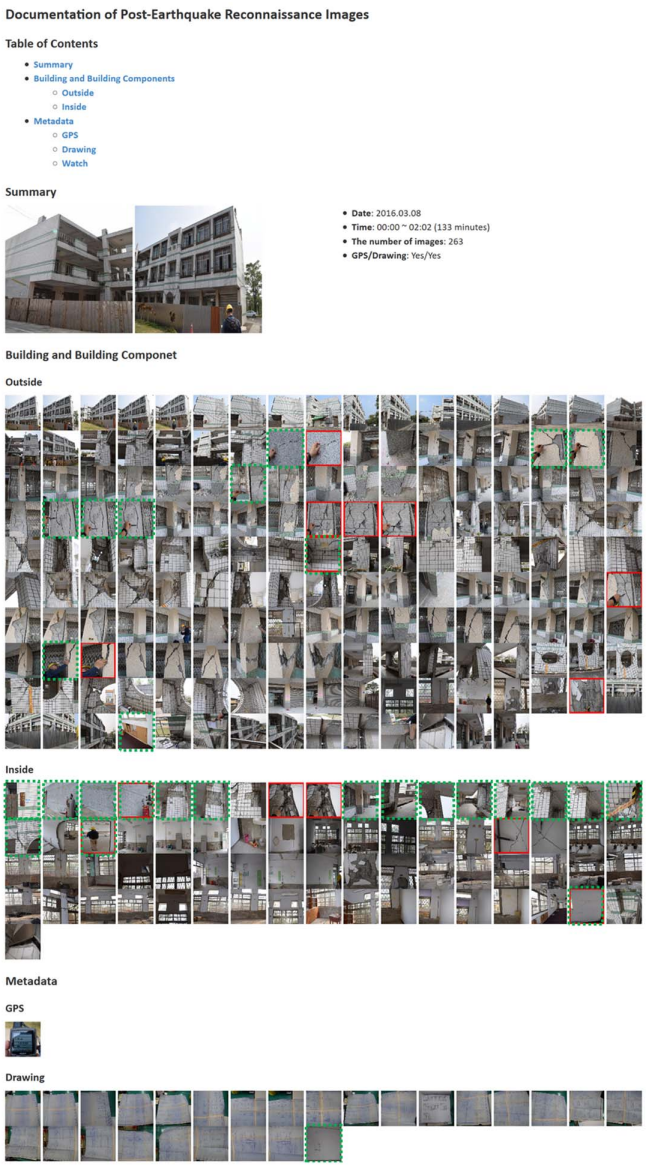
testing results, the classification is successful in achieving high recall and precision. Overall, the recall obtained in each class is greater than 80.0%. Herein, it is helpful to point out two findings related to these results. First, in the binary classification results for OV or MEAS, precision values are relatively low compared with recall values. This outcome is due to the fact that the number of testing samples is highly unbalanced, indicating that the number of negative samples is much greater than the number in the target class (OV or MEAS). In the authors' real-world image collection, images in these classes are relatively small portions of the number in BIN or BEX. Thus, when these classifiers are applied to OV and MEAS, a large portion of false-positive samples is generated. However, in actual implementation, high recall far outweighs high precision because it is more critical to identify all of the images in such minority classes without overlooking them.

Second, the precision of the MEAS classifier is relatively low compared with multiclass precision values. In multiclass classification, the images in each class must have very distinct boundaries between the visual features in the classes, and the images within a class share similar appearances. This characteristic yields high classification accuracy. However, with binary classification, the images tend to look more similar, with less clear boundaries between classes. For example, in the MEAS class, the only visual difference may be the inclusion of a measurement tool such as a measuring tape or crack gauge. This limitation could be possibly overcome by adding more training images in the MEAS class to generate more features in the network that are based on these differences.

Sample Reports

All of the steps discussed to this point are brought together in the generation of a single report for each building using trained classifiers. Here, the classifiers are applied to extract information from the data set of a given building and document the data and metadata in a ready to use format. Two sets of images (the full set of images from two different buildings) collected after different earthquakes are used for demonstrating report generation. Images used for sample report generation were not used for training as described in the

(a)



(b)

Fig. 7. (a and b) Sample reports automatically generated from original images in Figs. 6(a and b). MEAS images are marked with boxes. Dotted boxes indicate false classification. (Images from Purdue University and NCREE 2016.)

J. Perform. Constr. Facil., 2019, 33(1): 04018103

previous section. Detailed information about these two data sets is presented in Table 2. The external appearances and location of each building can be viewed in Google Street View using the provided GPS information. The definitions used here for the various damage levels assigned to structural and masonry wall damage in Table 2 are available from Datacenterhub (2014).

The images collected from each building are shown in Fig. 6. All images in Fig. 6 are presented as square thumbnails to enable convenient representation of the data set. Each image set does not necessarily include images from all classes in the schema. For example, if a digital copy of the structural drawings is available, the team would not need to collect drawings as images, and they would not need to be classified.

The two reports automatically generated from each data set using the proposed approach are shown in Fig. 7. All images are linked in the file to the original (high) resolution ones. The processing time for generating each report is 9 and 14 s, respectively. The images classified as MEAS are marked with a dark gray box. The images that are incorrectly classified, i.e., false classification, are marked here with a dashed gray border. Overall, the performance of the classifiers was quite successful. Among a total of 93 and 263 images in each data set, 85 and 235 images were correctly classified, corresponding to recalls of 91.3% and 89.4%, respectively, which are comparable with the results in the previous section. An engineer could readily use such reports to review the general building appearance and information collected in the field. Moreover, they would easily be able to identify images of special interest that require extra attention, such as those providing specific forensic information, without the distraction created by unrelated images.

Conclusion

In this study, the authors developed a novel approach for rapidly and autonomously classifying and organizing postevent reconnaissance building images. Automation was achieved by exploiting and adapting recent developments in convolutional neural networks to analyze this type of complex and unstructured real-world images. An appropriate schema was designed that is useful for organizing and browsing through the many images. The report generation capability will directly support field data collection and documentation efforts, providing proper descriptive information, and enabling engineers to readily find images of special interest. The authors demonstrated the approach through the organization of images from damaged buildings during past earthquake reconnaissance missions in which some of the authors took part. Using a large volume of real-world images from past missions, the authors showed the capabilities of the classifiers for image classification. With classifiers successfully trained, two sample reports were generated from two sets of building images, each containing many images. Using the pretrained classifiers, images from a single building could automatically be used to generate a report in under 1 min. This report generation capability will directly support field data collection and documentation efforts, help in providing proper descriptive information, and enable engineers to readily find images of special interest. As the use of drones and other data collection system increases and more and more images are collected in future missions, such automation will be essential to organize and understand those data.

Acknowledgments

The authors wish to acknowledge partial support from National Science Foundation under Grant Nos. NSF-1608762 and DGE-1333468, and the valuable image contributions from the Center for Earthquake Engineering and Disaster Data (CrEEDD) at Purdue University (datacenterhub.org), the EUCentre (Pavia, Italy), the Instituto de Ingenieria of UNAM (Mexico), FEMA, and the EERI collections. The authors also acknowledge the NVIDIA Corporation for the donation of a high-end GPU board.

References

ACI (American Concrete Institute). 2017. *Disaster reconnaissance*. ACI 133. Farmington Hills, MI: ACI.

ATC (Applied Technology Council). 1989. *Building safety evaluation forms and placards*. ATC 20. Redwood City, CA: ATC.

ATC (Applied Technology Council). 2004. *Field manual: Safety evaluation of buildings after windstorms and floods*. ATC 45. Redwood City, CA: ATC.

Datacenterhub. 2014. Accessed October 9, 2017. http://datacenterhub.org.

DesignSafe-CI. 2016. "Rapid experimental facility." Accessed October 9, 2017. https://rapid.designsafe-ci.org.

EERI (Earthquake Engineering Research Institute). 2009. "Earthquake clearinghouse." Accessed October 9, 2017. www.eqclearinghouse.org.

Github. 2007. Accessed October 9, 2017. http://github.com.

Johnson, J., R. Krishna, M. Stark, L.-J. Li, D. A. Shamma, M. S. Bernstein, and L. Fei-Fei. 2015. "Image retrieval using scene graphs." In *Proc., 2015 IEEE Conf. on Computer Vision and Pattern Recognition*, 3668–3678. Washington, DC: IEEE.

Karpathy, A., G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei. 2014. "Large-scale video classification with convolutional neural networks." In *Proc., 2014 IEEE Conf. on Computer Vision and Pattern Recognition*, 1725–1732. Washington, DC: IEEE.

Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2012. "Imagenet classification with deep convolutional neural networks." In *Proc., 25th Int. Conf. on Advances in Neural Information Processing Systems*, 1097–1105. Lake Tahoe, Nevada: Curran Associates, Inc.

LeCun, Y., B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel. 1990. "Handwritten digit recognition with a back-propagation network." In *Proc., Advances in Neural Information Processing Systems*, 396–404. Lake Tahoe, NV: Curran Associates, Inc.

Purdue University and NCREE (National Center for Research on Earthquake Engineering). 2016. "Performance of reinforced concrete buildings in the 2016 Taiwan (Meinong) earthquake." Accessed October 28, 2017. https://datacenterhub.org/resources/14098.

Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, and S. Ma. 2015. "ImageNet large scale visual recognition challenge." *Int. J. Comput. Vision* 115 (3), 211–252. https://doi.org/10.1007/s11263-015-0816-y.

Shah, P., S. Pujol, A. Puranam, and L. Laughery. 2015. "Database on performance of low-rise reinforced concrete buildings in the 2015 Nepal earthquake." Accessed October 9, 2017. https://datacenterhub.org/resources/238.

Sim, C., C. Song, N. Skok, A. Irfanoglu, S. Pujol, and M. Sozen. 2015. "Database of low-rise reinforced concrete buildings with earthquake damage." Accessed October 9, 2017. https://datacenterhub.org/resources/123.

Sim, C., E. Villalobos, J. P. Smith, P. Rojas, S. Pujol, A. Y. Puranam, and L. Laughery. 2016. "Performance of low-rise reinforced concrete buildings in the 2016 Ecuador earthquake." Accessed October 9, 2017. https://datacenterhub.org/resources/14160.

UC CEISMIC. 2011. "Canterbury earthquake digital archive." Accessed October 9, 2017. www.ceismic.org.nz.

Vedali, A., and K. Lenc. 2015. "MatConvNet: Convolutional neural networks for MATLAB." In *Proc., 23rd ACM Int. Conf. on Multimedia*, 689–692. New York: ACM.

Yeum, C. M. 2016. "Computer vision-based structural assessment exploiting large volumes of images." Ph.D. dissertation, Lyles School of Civil Engineering, Purdue Univ.

Yeum, C. M., S. J. Dyke, B. Benes, T. Hacker, J. Ramirez, A. Lund, and S. Pujol. 2017a. "Rapid, automated image classification for

documentation." In *Proc., 7th Conf. on Advances in Experimental Structural Engineering*. Pavia, Italy: EUCENTRE Foundation.

Yeum, C. M., S. J. Dyke, and J. Ramirez. 2017b. "Visual data classification in post-event building reconnaissance." *Eng. Struct.* 155: 16–24. https://doi.org/10.1016/j.engstruct.2017.10.057.

Yeum, C. M., S. J. Dyke, J. Ramirez, and B. Benes. 2016. "Big visual data analysis for damage evaluation in civil engineering." In *Proc., Int. Conf. on Smart Infrastructure and Construction*, 569–574. Westminster, London: ICE Publishing.

Yeum, C. M., S. J. Dyke, J. Ramirez, T. Hacker, S. Pujol, and C. Sim. 2017c. "Annotation of image data from disaster reconnaissance." In *Proc., 16th World Conf. on Earthquake Engineering*. Santiago, Chile: Chilean Association of Seismology and Earthquake Engineering.

Zhou, B., A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. 2014. "Learning deep features for scene recognition using places database." In *Proc., 27th Conf. on Advances in Neural Information Processing Systems*, 487–495. Lake Tahoe, NV: Curran Associates, Inc.