Multi-output Image Classification to Support Post-Earthquake Reconnaissance

Ju An Park[1], Xiaoyu Liu[2], Chul Min Yeum[3], Shirley J. Dyke[4], Max Midwinter[5], Jongseong Choi[6], Zhiwei Chu[7], Thomas Hacker[8], Bedrich Benes[9]

[1]Graduate Student, Department of Civil Engineering, University of Waterloo, Waterloo, ON, Canada.

[2]Graduate Student, Lyles School of Civil Engineering, Purdue University, West Lafayette, IN, United States.

[3]Professor, Department of Civil Engineering, University of Waterloo, Waterloo, ON, Canada. Email: cmyeum@uwaterloo.ca

[4]Professor, School of Mechanical Engineering, Purdue University, West Lafayette, IN, United States.

[5]Graduate Student, Department of Civil Engineering, University of Waterloo, Waterloo, ON, Canada.

[6]Professor, Department of Mechanical Engineering, SUNY Korea – Stoney Brook University, Incheon, Republic of Korea.

[7]Graduate Student, Computer and Information Technology, Purdue University, West Lafayette, IN, United States.

[8]Professor, Computer and Information Technology, Purdue University, West Lafayette, IN, United States.

[9]Professor, Professor of Computer Science, Purdue University, West Lafayette, IN, United States.

ABSTRACT

After hazard events, large volumes of images are collected by reconnaissance teams to document the post-event state of structures and are used to assess their performance and improve design procedures and codes. The majority of the data are captured as images and manually labeled. This highly repetitive task requires considerable engineering expertise and time. Advances in deep learning have enabled researchers to rapidly classify reconnaissance images. Thus far, these classification methods remain limited to a simple classification schema in which the classes are either all mutually exclusive or independent. To date, an efficient classification system of a complex schema containing many classes arranged in a multi-level hierarchical structure is not available to support

earthquake reconnaissance. To address this gap, this paper introduces a comprehensive classification schema and a multi-output deep convolutional neural network (DCNN) model for rapid post-earthquake image classification. Compared to past works, a single multi-output DCNN classification model with a hierarchy-aware prediction is trained to enable the rapid organization of images. The performance of the proposed multi-output model is validated through comparisons with multi-label and multi-class models using a F1-score. As result, the multi-output model outperforms the other models. Then, the multi-output model developed in this study is deployed in a web-based platform called the Automated Reconnaissance Image Organizer, which can be used to easily organize earthquake reconnaissance images.

INTRODUCTION

The response of the built environment to extreme events such as earthquakes is an important source of information that can be used for improving the design and lifecycle analysis of infrastructure. Visual data are the most used method to document the performance of structures in field reconnaissance. Data collected in the field is used in subsequent investigations by researchers, practitioners, and students, seeking to identify gaps present today in how knowledge of engineering and science can be applied to the built environment. These investigations lead to actionable information when coupled with economics and policy considerations to promote and produce a safer built environment.

Significant progress has been made in earthquake engineering over the last several decades in learning lessons due to observations and data collected during such field missions, termed reconnaissance. The National Science Foundation (NSF) in the U.S. has funded numerous RAPID awards over the years since 2009, each focused on sending teams of engineers to the event site to collect specific data and photographs and prepare a report on their observations. The Earthquake Engineering Research Institute (EERI) also has a "Learning from Earthquakes" program, which documents reports and links to over 290 earthquakes that occurred since 1971 (EERI, 2016). Many images have been collected over the years during such field missions, and this practice has gradually become common in various types of natural and man-made hazard events. In 2016, a nationwide engineering research and

education infrastructure spanning many types of natural hazards was established, namely the Natural Hazards Engineering Research Infrastructure (NHERI) network. In 2017, NHERI established a long-term science plan which highlights the crucial role of field data in disaster planning, response, recovery, and mitigation. The NHERI RAPID facility, within the NHERI network, provides the expertise and support needed for reconnaissance teams to collect valuable field data, which can then be posted and shared to motivate new lines of inquiry (Wartman et al., 2017). In parallel, the NSF's structural extreme events reconnaissance network (StEER) was formed to coordinate reconnaissance missions in collaboration with other stakeholders (StEER Network, 2018) for academic researchers. The National Institute of Standards and Technology (NIST) has a similar program to study disasters and failure of infrastructure. The programs have conducted a technical investigation or preliminary reconnaissance of major events including the Joplin Tornado in 2011 or the Champlain Towers South collapse in 2021 (NIST, 2016). Globally, the World Bank's challenge funds have also supported disaster data-related research meant to enhance disaster risk information and assessment (GNDR, 2019).

Among the various types of field data that could be collected, readily available visual data stored in the form of images and video allow engineers to preserve evidence associated with changes in the appearance of a structure resulting from a natural disaster. These images may be used either to answer immediate questions about the event, or they may be stored and employed for future analysis perhaps in the form of longitudinal studies. By using inexpensive cameras or cell phones, the reconnaissance teams can quickly capture a large number of images to document damage or even the failure of a structural member. Images may contain evidence of spalling, shear cracks, large deformations, buckling, or even collapse.

In previous research, the authors' research group developed and implemented techniques for automatically classifying and documenting post-earthquake reconnaissance images of reinforced concrete buildings (Yeum et al., 2019). Automation was achieved by adapting and exploiting state-of-the-art deep convolutional neural network (DCNN) algorithms to analyze complex and unstructured real-world reconnaissance images. This prior work represents the next step in the practical use of image classification techniques to support scientific research while overcoming several challenges associated with this application and documenting the new knowledge generated.

This paper builds upon the previously developed approaches. In the prior work, an image classification methodology using several, single-level schemas is developed in collaboration with domain experts such as field engineers and researchers for earthquake reconnaissance (Yeum et al., 2018; Yeum et al., 2019). The previous work by the authors' group has two main limitations. First, the schema established in the prior research contains a small number of classes that are only useful for classifying images at a high level, such as describing the overall appearance of a building and its contents in the collected images. This paper expands the schema in collaboration with domain experts to include component-level observations and damage-level evaluations, since classifying images into more granular and detailed categories allow for more effective organization and meaningful analysis. Second, all categories (8 in total) in the schema previously developed are mutually exclusive. This assumption does not hold when expanding a schema into multiple levels of hierarchy. For example, if previously, a single class was used to represent buildings and building components, this class can be further described using the image depth and location the image was taken. The image depth would describe whether it is an image of a build component, overview, or a space, and the location would describe whether it is a scene taken inside or outside a building. Child classes can be recursively defined as many times as needed to provide a sufficient level of detail. Expanding the schema in this manner would allow a single image to be assigned one or more categories. This paper creates a single model that conforms to the various mutually exclusive and inclusive relationships between classes at several levels of detail in a class hierarchy.

In this study, automatic post-disaster reconnaissance image organization is explored using a hierarchical schema and a single multi-output model. A comprehensive hierarchical schema is designed for the categorization of these images into one or more associated classes for each image. A single, multi-output model is developed and trained to predict the schema classes while following the hierarchical rules enforced by the schema (where hierarchical rules denote interclass relationships like mutual independence or exclusivity between classes). The multi-output model efficiently replaces the need for several single-output models. The capability of the technique is thoroughly evaluated using real-world post-earthquake images collected from past earthquake events. The classifier trained with the hierarchical schema is deployed into an online tool, the Automated Reconnaissance Image

Organizer (ARIO). This allows for cross-collaboration between teams of field engineers, where they can easily combine datasets, even if they have been previously labeled using different schemas. This is possible as the model can intelligently relabel the images under a single comprehensive schema within minutes instead of engineers spending days to months manually labeling each image. This enables immediate and fast inter-team collaboration. ARIO allows multiple users to readily access the web-based image organizer and the data contained therein without any installation process or computing hardware.

BACKGROUND

During post-disaster reconnaissance, it is common for the inspector to take photographs for the record, reference, or report. However, no standards or guidelines are applied for gathering or classifying these inspection images. For example, the Applied Technology Council's (ATC) guidelines, ATC-20-1 and ATC-45 are commonly accepted as standard practice for rapid evaluations, however, this standard does not provide instructions for the task of documenting inspection images (ATC, 2004; ATC, 2005). FEMA P-58 attempts to build a probabilistic model of the structure by assigning structural components to fragility curves after the events (ATC, 2018). Lastly, to conduct detailed and engineering evaluations, FEMA 306 and FEMA 352 provide a detailed guide to classifying earthquake damage. However, the guidance provided is idealized and the damage definitions are tied to how the damage affects the performance of a given structural element, which may be difficult to discern from inspection images alone (ATC, 1998; SAC Joint Venture, 2000).

A consequence of not defining standards for post-disaster image collection or classification leads to discrepancies between reconnaissance databases. In published post-earthquake reconnaissance image databases, images should be manually tagged by the data collectors, the authors of these important digital products, to allow for easier navigation and data reuse. Based on an extensive assessment of existing repositories, no image tagging standard has been widely adopted, which has resulted in popular publishers (e.g., American Concrete Institute (ACI), EERI Clearinghouse, and Purdue University) having differing labeling schemes (EERI, 2009; Datacenterhub, 2014; Jafarzadeh et al., 2015; Laughery et al., 2020). For example, ACI classifies damage as light, moderate, and severe;

while the EERI Clearinghouse has none, moderate, severe, and total collapse classes as well as an option to specify the structural element that is being tagged.

The most useful post-disaster databases allow researchers to easily review important features and draw conclusions from aggregate data (EERI, 2009; Datacenterhub, 2014; DesignSafe-CI, 2016). Currently, the authors or data curators must tag every image manually, which is a tedious and costly endeavor. To address this issue, several researchers have leveraged advances in computer vision and artificial intelligence to develop methods to automate the categorization of images for infrastructure assessment after a disaster. The scope of ongoing research spans a vast domain from bridge damage, post-earthquake building damage, infrastructure damage, or geotechnical failures (Bray et al., 2019; Azimi et al., 2020). A popular benchmark for image-based structural assessment is the PEER Hub ImageNet (PHI) Challenge. The PHI Challenge is divided into eight separate classification tasks: material type, component type, damage type, damage level, damage state, scene level, spalling condition, and collapse mode where each classification task contains two to four distinct classes.

A large collection of open-source labeled images known as PEER Hub ImageNet was recently developed to provide a general automated framework for the development and evaluation of deep learning models in structural health monitoring (Gao and Mosalam, 2020). As a multi-attribute structural image dataset, PEER Hub ImageNet envisions a hierarchical tree, where at each node would be a bespoke multi-class classifier or detector (e.g., scene level, damage state, spalling, or material type), designed to progress the classification down the tree. Gao and Mosalam demonstrated their work by developing a classification system for one branch of the proposed hierarchy tree, which required two binary, three-class, and four-class classifiers. For the proposed Structural ImageNet hierarchy tree, 13 classifiers are required, and classifying one image can require the application of up to six classifiers. While the proposed hierarchy tree is effective, the classification system involves numerous deep learning models, challenges to tune hyperparameters for each model and highly expensive computational costs to test.

METHODOLOGY

In this study, an automated classifier is trained to categorize images into pre-defined, task-oriented, and hierarchical classes that are designed to be useful for organizing and documenting post-disaster reconnaissance data (Yeum et al., 2019). Since these image collections are large in number, diverse in contents, and very complex, manually sifting through these images to identify scenes of interest is extremely cumbersome and time-consuming. Implementation of the classifier is deployed on an interactive and web-accessible platform (ARIO, http://ario.tech.purdue.edu) to directly support field engineers in the rapid gathering and organization of the data as they are collected. With this capability and platform, engineers in the field can quickly organize their data and directly utilize the classified image sets to get actionable information and make informed judgments about the present state of structures, and then plan for the next phase of the mission.

Design of a Hierarchical Schema

To establish appropriate categories for the application and a hierarchical structure for the categories, field engineers were consulted to understand the types of categories needed. These categories were designed for the visual classification application, which relies on the images being visually distinguishable for classification to proceed. The class hierarchy in this study is designed to enable the informative classification of reconnaissance images. Herein, the relationship between the classes does not need to be mutually exclusive, which means that each image can be categorized into multiple classes. The multi-output classification algorithm discussed in the section titled "CONVOLUTIONAL NEURAL NETWORK-BASED HIERARCHICAL CLASSIFICATION" allows for the output of multiple probable classes without compromising the classification accuracy.

Each category in Fig. 1, was developed to guide human annotators in the annotation process needed to establish consistent ground truth training data. The boxes with a solid blue fill are the classes used in training the classifier in this study. The remainder of the classes are defined as part of the schema but are not used to train the classifier due to an insufficient number of image samples. Note that the schema is designed to provide rapid categorization into the shallow set of classes that engineers would explore or need for documentation and further analysis. Thus, this structure provides basic classes to reduce the effort required to explore the images and is not

designed for exhaustive classification and/or analysis. As such, engineers can augment the model to add a custom set of classes to the schema and the associated classifiers as needed.

Figure 1. Hierarchy of the classes in the schema: The boxes with a solid blue are the classes used for image classification in this study. The black arrow edges represent the relationship between a superior (parent) and subordinate(s) (child). The solid red edges indicate that the two classes at the ends of each edge are mutually exclusive. A group of classes chain-linked by red lines form a single set of mutually exclusive classes (e.g., CD0, CD1, CDR, and CDM form a set where each class is mutually exclusive from one another).

After data collection, the reconnaissance images, shown as RIMG in Fig. 1, are first classified into one of eight classes, seven of which are classes containing visual contents relating to metadata (shown as drawing (DWG), GPS (GPS), irrelevant (IRR), nonstructural element (NON), sign (SGN), watch (WAT), and measurement (MEAS) images) and the other containing visual contents of buildings and building components (shown as BBC). These metadata classes and their definitions were defined in the authors' prior work (Yeum et al., 2019). A summary definition of all the categories in Fig. 1 is available in Table 1 and their sample images are shown in Fig. 2.

Table 1. Definitions of classes in a schema

The Building and Building Components (BBC) category contains visual contents broadly relating to buildings and their structural components. Images in the BBC category are divided into image depth (DEP) and location (LOC). DEP categorizes the distance between the building (or building component), and the image capture location. The reason for categorizing the images by depth is that the appearance and type of visual content on the images change dramatically depending on their depth. Images in this category are classified into three different depths and each depth contains different visual contents: building components (BCP) which are often captured near the object, building overview (BOV) which are captured from a distance to record the complete external appearance of buildings, and building space (BSP) containing multiple building components to understand their spatial context. LOC categorizes the image as taken of the building's exterior or interior. LOCEX is defined as images that have a

general sense of outside space. Images labeled as LOCEX might show the external face of a wall, be very bright due to sunlight, or contain the overview scenery of a building. LOCIN has the opposite meaning as the LOCEX category. For example, these images could show an interior space surrounded by walls or windows, or contain interior building components or regions (e.g., hallway, basement, or room).

BCP has two subclasses: (1) building component damage (CPDMG): this category describes the level of damage on a building component; and (2) building component type (CPTYPE): This category specifies the building component. The subclasses in CPTYPE include the scenes of specific building components, such as beams (CPBEAM), columns (CPCOL), and walls (CPWAL). Note that occasionally, scenes of walls, columns, and beams may overlap with each other, but a given class is distinguished from the other two by being the center of focus of the image (e.g., the camera is pointed at the beam or takes up more space in the image than other components). The definition of the classes indicating component-level damage in CPDMG are present here:

- CD0: This label indicates that the component has little (e.g., crack, minor spalling) to no visual damage on the building component of interest such as a wall or column. The component may contain a small portion of a fragment or chip like spalling;
- CD1: This category is defined by the presence of severe cracks or flaking/peeling of a large portion of a thin layer of plaster from the surface due to severe cracks (here, the “large” is estimated based on its size relative to the underlying building components). In the case of a wall, it can be visually identified as a structural or masonry wall. Also, it often contains wide (and deep) cracks on the concrete surface where paint and plaster can be seen peeling off. Fragments generated from crack intersections may be produced;
- CDR: The characteristic of this category inherited from CD1 in that it also describes concrete damage on structural components but is a classification of higher severity. Images with this category include damage such as exposed rebar due to severe spalling, severe fractures, chipped off concrete cover around rebar, and/or large deformation of the rebar (buckling) due to a lack of confinement;
- CDM: Images in this category are described by the presence of severe masonry damage. Images in this class include flaking or peeling of a large portion of a thin layer of plaster from the surface due to severe cracking.

The flaking of the plaster enables engineers to visually identify such scenes as a masonry wall. The category can also contain scenes of wide (stair step) cracking following the masonry joints, damage to masonry blocks, and/or holes in a depth direction.

Note that the subclass CPDMG has an associated CPTYPE, and vice versa, which together describe the damage and type of a building component. For example, if a given image is annotated as CD0 and CPCOL using the trained classifier, it indicates that there is little to no damage to the building column captured in the image.

BOV has two subclasses: (1) building overview damage (OVDMG): this category describes the level of damage visible on a building exterior. The extent of damage that is visually present on the building overview can be categorized using OVDMG. There are two different damage levels: light/moderate (ODM) and severe (ODS) damage. (2) building overview angle (OVANG): This category describes the angle relative to the building the image was captured. OVANG has two categories: OVCAN indicating the building view is "canonical" (angled, showing more than one face of the building), and OVFRT, showing a single-sided view of the building. The definition of the classes in OVDMG are present here:

- ODM: This category indicates that the building in the image is from slightly to moderately damaged, with visual evidence of cracks, broken windows, and/or some minor spalling. The building in this category does not contain significant structural damage;
- ODS: This category indicates that the building in the image is significantly damaged, where the structure looks unstable (e.g., leaning), large chunks of the structure missing (e.g., missing columns/walls), or entire sections of the building are demolished. Images in this category often show overview scenes of component-level damages such as CDR or CDM.

For BSP, building spaces, the classes within this category are the building/room entrance (SPENT), first floor in a building (SPFFL), room space (SPROOM), or building basement (SPBSM). Note that these subclasses are presented in the schema but are not used to train the model because there were an insufficient number of images for these classes to train the model in this study. For the same reason, CPBEAM is not considered in this study.

Figure 2. Sample images for each of the designed classes introduced in the schema in Fig. 1 [Shah et al., 2015; Sim et al., 2015, 2017; Purdue University and NCREE, 2016].

The proposed schema for this study presents sets of both mutually exclusive and overlapping classes in the nodes of the hierarchy tree. In Fig. 1, the black arrow edges indicate the relationship between a superior (parent) and subordinates (children). All images with subordinate classes have all their parent class(es) (immediate and all ancestors). For example, an image of CDR is also labeled as RIMG, BBC, DEP, BCP, and CPDMG. In this study, only the classes with a solid blue color were considered (e.g., BCP). Also, classes horizontally connected by solid red edges represent a set of classes that are mutually exclusive from one another. For example, CD0, CD1, CDR, and CDM form a set, where each class in this set is mutually exclusive with all the other classes in the set. In other words, the immediate child nodes of the CPDMG node form a distinct set that is all mutually exclusive from each other. Thus, immediate child node classes of RIMG, DEP, LOC, CPDMG, CPTYPE, OVDMG, and OVANG form eight distinct sets for the schema as shown in Fig. 1.

The rule of a class assignment for an image can be explained using a top-down approach. First, a class from the immediate child classes of RIMG is assigned. If the class is non-BBC, the class assignment is complete. If the class is BBC, then both DEP and LOC are assigned to the image, since these two classes are not mutually exclusive. Then, one child class from DEP and LOC each is assigned to the image. For example, BCP and LOCIN, or BOV and LOCOUT. If the image is assigned a BCP label, then both CPDMG and CPTYPE classes are assigned. Then one class from CPDMG and CPTYPE each is assigned to the image. On the other hand, if the image is assigned BOV instead of BCP, then the image is assigned classes OVANG and OVDMG, and then also assigned one child class from each of the two classes. Lastly, if the image is assigned BSP instead of BCP or BOV, then the image is assigned SPTYPE, and one of its child classes. As a result, a single image may fall into one or more classes that follow this strict hierarchical rule. This rule for the class assignment will be integrated into the multi-output classifier to avoid classifications that would be inconsistent with the hierarchical schema.

CONVOLUTIONAL NEURAL NETWORK-BASED HIERARCHICAL CLASSIFICATION

As illustrated in the subsection titled "Design of a Hierarchical Schema", the hierarchy proposed for this study presents a complex set of mutually exclusive classes in the nodes of the hierarchy tree. For example, the child classes of CPTYPE are mutually exclusive among each other but are independent of the child classes of CPDMG. This structure allows labeling of both the component type and corresponding damage level. In the design of the network, the output must conform to the hierarchical relationships of mutual exclusivity and/or independence. Unfortunately, using a single classification scheme, such as multi-class or multi-label, would not be sufficient to ensure model predictions conform to the schema. Multi-class, which is typically used to predict a single class from a set of a mutually exclusive set of classes, does not allow for the prediction of multiple classes and assumes all the classes are independent of each other. Multi-label, while allowing the prediction of multiple classes for a single image, assumes mutual independence between all classes, potentially leading to an illogical set of predictions (e.g., an image being predicted to be both IRR and WAT). As a result, a multi-output DCNN design and configuration are adopted here to ensure model predictions conform to the hierarchical schema.

Three network configuration types are designed and examined in this study, namely multi-class (MC), multi-label (ML), and multi-output (MO), to assess and compare the model performance of the three different approaches. Their top layer configurations are shown in Fig. 3. All three configurations take as input the feature vector of dimension **10 × 10 × 1792** from the base model (in Fig. 4) and have two sets of layers including a global 2D averaging pooling layer and a dropout layer. A **1 × 1792** vector is fed into the last layer. The last layer has different structures for each configuration. For MC and ML in Figs. 3(a) and (b), a 21-way (for 21 classes) densely connected neural networks are designed but Softmax and Logistic activation functions are used, respectively. For MO in Fig. 3(c), six groups of densely connected layers with a Softmax activation function are used, and each group is responsible for producing a single prediction for a given set of classes. They contain the number of classes corresponding to their ancestor (e.g., for CPTYPE, a 2-way dense layer is used to label either CPCOL or CPWAL).

In this study, the performance of three different configurations, MC, ML, and MO are assessed to demonstrate that MO is the best for supporting the designed schema. The MC configuration is not feasible for

training from a dataset with a complex hierarchy since it is designed for single class prediction. Alternatively, the ML configuration, while it can be used to predict multiple classes for a single image, does not enforce the hierarchical relationship of mutual exclusivity/independence discussed previously. As a result, the MO configuration with a hierarchy-aware prediction algorithm was adopted, which uses several dense layers with Softmax functions to enable a multi-output framework to better satisfy the mutual exclusivity of relationships in the hierarchy.

The MO is beneficial in the sense that only a single DCNN classifier is developed to predict all required classes by utilizing a multi-output structure. The multi-output structure means that classes belonging to different categories (e.g., LOCEX and LOCIN belonging to category LOC, and CD0, CD1, CDM, and CDR belonging to category CPDMG) can be grouped by categories to utilize different activation and loss functions relevant to each category using features trained at the base network. For this study, the Softmax activation function is used for the MO configuration, and the maximum probabilities of each category are selected to be the positive predictions. In the actual implementation of the proposed model, the leaf nodes, BCP, and BOV have been included in the set of possible model predictions at the top classes in the schema instead of BBC to distinguish whether the image has BCP or BOV-relevant labels. However, simply selecting the classes with the maximum probability of each category in the MO architecture still violates the hierarchical rules of the schema. For example, if an image has a WAT label, the same image cannot be associated with any other class. On the other hand, if an image has a BOV label, then it will also have classes related to LOC, OVANG, and OVDMG. A custom prediction algorithm must be further implemented for the MO configuration.

(a) (b)

(c)

Figure 3. Top layer configurations for (a) multi-class, (b) multi-label, and (c) multi-output: The base network design is shown in Fig. 4, which is shared by all three configurations. The output vector dimensions are marked beside each layer

To enable a model prediction conforming to the hierarchy for the MO configuration, a custom prediction algorithm is proposed in Algorithm 1. The proposed hierarchy-aware prediction algorithm follows a top-down approach, where prediction results of the top-level nodes (e.g., RIMG) guide the results of downstream nodes (e.g., LOC, CPTYPE, OVDMG). First, the class with the highest probability among child classes of RIMG and nodes BCP and BOV is selected. If the class is one of the immediate child classes of RIMG (DWG, GPS, IRR, NON, SGN, WAT, or MEAS), then the prediction algorithm terminates with the image label being assigned the said class. If BCP is selected, then one class with the highest probability is selected from the child classes of CPDMG, CPTYPE, and LOC each (e.g., LOCIN, CD0, CPWAL). If the selected class is BOV, then one class with the highest probability is selected from the child classes of LOC, OVANG, and OVDMG each (LOCEX, OVFRT, and ODM). After these child classes are selected, the algorithm terminates class assignments.

Algorithm 1: Hierarchy-Aware Model Prediction

Input: Raw model probabilities of 21 classes

Result: Model predictions conforming to the hierarchy

1. Create an empty prediction list, P
2. RIMG class = argmax(Raw probabilities of RIMG child classes)
3. Add RIMG class to P
4. **If** RIMG class = BCP **then**
   a. LOC class = argmax(raw probabilities of LOC child classes)
   b. CPDMG class = argmax(Raw probabilities of CPDMG child classes)
   c. CPTYPE class = argmax(Raw probabilities of CPTYPE child classes)
   d. Add LOC class, CPDMG class, and CPTYPE class to P
5. **Else If** RIMG class = BOV **then**
   a. LOC class = argmax(Raw probabilities of LOC child classes)
   b. OVANG class = argmax(Raw probabilities of OVANG child classes)
   c. OVDMG class = argmax(Raw probabilities of OVDMG child classes)
   d. Add LOC class, OVANG class, and OVDMG class to P
6. **Return** P

In general, the algorithm developed here enables a single multi-output model to make a hierarchy conforming set of predictions by first relying on the initial guess of whether the image is one of the immediate child

classes of RIMG or one of BCP or BOV. Only after this initial class prediction can the algorithm decide to include related classes of BCP (LOC, CPDMG, and CPTYPE) or BOV (LOC, OVANG, and OVDMG), or include neither. The proposed algorithm heavily relies on the initial set of predictions. If the initial set of predictions is not accurate, model performance deteriorates significantly. Thus, it is important for the model to accurately predict the initial set of classes. Typically, it is easier for models to predict visually distinct classes than classes with an overlap in their visual features. Child classes of the RIMG and BCP and BOV which are the initial set of classes for this study are visually distinct and as such, the risk of misprediction for these classes is relatively low compared to the other classes. For example, the metadata classes, BCP and BOV at the top level are more visually distinct than the other mutually exclusive pairs in subordinate classes (e.g., CPDMG, CPTYPE). And, since images labeled at BCP and BOV level in the hierarchy include images labeled at all child classes, more labeled images are used for training the classifier.

EXPERIMENTAL VALIDATION

In this section, the dataset used to train the DCNN with each of the top layer variations, the training configuration, and the training results are discussed in detail.

Ground truth labeling

The research team uses the post-earthquake reconnaissance image database (Yeum et al., 2019). The images in this database have been collected during dozens of past earthquake reconnaissance missions, including Düzce and Bolu, Turkey in 1999; Peru in 2007; Bingöl, Turkey in 2003; Taiwan in 2016; Haiti in 2010; Nepal in 2015; and Ecuador in 2016 (Shah et al., 2015; Sim et al., 2015, 2017; Purdue University and NCREE, 2016). The reconnaissance data sets that house the images used in this study are publicly available at datacenterhub.org (Purdue University). A total of 9173 images were labeled using the designed schema. In total, labeled images were captured from 992 reinforced concrete buildings from these events (for an average of 11.9 images per building).

Training and Hyperparameter Configuration

For this study, the EfficientNet-B4 architecture is adopted as the base model, followed by one of three possible configurations of FCNNs consisting of dropout and dense layers with one or more activation functions at the last layer, depending on the configuration (Tan and Le, 2019). The EfficientNet architecture starts with a single convolutional layer and is followed by a series of MBConv blocks which allow the network to achieve state-of-the-art accuracies on the ImageNet leaderboards with a substantially lower number of parameters (by an order of magnitude) as compared to preceding state-of-the-art models, such as ResNet, Inception, DenseNet. The network architecture and top layer configurations are shown in Figs. 3 and 4, respectively. In Fig. 4, the model takes as input color image size of **299 × 299 × 3**, which is pushed through a sequence of convolutional and MBConv(*t*) layers, and output as a feature vector of size **10 × 10 × 1792**. Here, *t* is the expansion factor parameter. In Fig. 4, the "#L = n" on the top of each layer denotes the number of times the corresponding layer is replicated. This feature vector is then input into one of the top layer configurations shown in Fig. 4, and the output is the predictions for the 21 classes. One CCE loss function is applied to the MC configuration and one BCE loss function to the ML configuration. For the MO configuration, six CCE loss functions are applied, one to each grouping of classes (RIMG, LOC, CPDMG, CPTYPE, OVANG, OVDMG). Three different top layer configurations are tested to identify an optimal top layer configuration for the given hierarchical schema and dataset of the classification task.

Figure 4. EfficientNet-B4 base network architecture

For model training, the dataset consisting of 9173 images is randomly split into 80% (7338 images) training and 20% (1835 images) validation sets. Some training parameters have been empirically optimized for each top layer configuration, and thus do not share the same training hyperparameters. An epoch of 30, batch size of 16, and a stochastic gradient descent optimizer with learning rates between 1×10^{-3} and 1×10^{-2}, a momentum of 0.9, and a decay of 1×10^{-3} are used to fine-tune the model. Before each image is input into the model it is automatically resized to **299 × 299** pixels, respectively, and randomly augmented to artificially increase the number of unique samples the model sees during the training phase. Augmentations implemented for this study include horizontal

flips, minor brightness range augmentations ±10%, minor horizontal and vertical shifts ±5% of the image width and height, respectively, minor zoom augmentation ±3%, and minor rotations up to ±15 degrees. A Linux workstation with an Intel Core i9-7940 CPU, NVIDIA GTX 1080 Ti GPU with 11 GB video memory, and 64GB of RAM was used to train the three different top layer configurations of the DCNN.

Classification Results

A summary of the performance of the three different top layer configurations is shown in Table 2. Precision (Eq. 1), recall (Eq. 2), and F1-score (Eq. 3) metrics are used to assess the performance of each configuration, defined as

$$\text{precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{1}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{2}$$

$$\text{F1-score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision+recall}} \tag{3}$$

where TP, FP, and FN are the total respective number of true-positives, false-positives, and false-negatives for a given class. Precision assesses the accuracy of a model's positive predictions for a given class. Recall assesses how well the model accurately predicts the true labels. F1-score is a metric that incorporates both precision and recall, giving a well-rounded performance estimate. Specifically, recall is a measure of how well a model retrieves labels of a specific class, while precision is a measure of how often a model is correct when it predicts a positive prediction. For example, in the dataset, there are 494 DWG images. The model correctly predicts 474 of the 494, yielding a recall of 474/494, which is 0.96. However, it is crucial to note that the recall value for DWG is not affected when the model wrongly predicts images of other classes as DWG (although the recall values of other classes will decrease). Whether the model predicts 10 or 100 GPS images as DWG, the recall value for DWG will not change, as it is a metric of information retrieval, aimed at answering the question "how well does the model identify the class correctly?". However, the recall value for GPS will suffer, as images classified as GPS are wrongly classified as DWG. On the other hand, if the model predicted that 499 images are DWG, but only 494 of the images are actually DWG, then the corresponding precision would be 494/499, which is 0.99. As in the previous example, if

388 the number of incorrect predictions increases (FP), the precision would go down. Thus, both recall and precision
389 are used to assess the overall performance of the model, or the F1-score, which is a combination of the two metrics.
390 Furthermore, macro-average and weighted averages are calculated to estimate the model's overall
391 performance across all classes. Macro-averages are computed as the simple mean of a metric (e.g., precision, recall,
392 or F1-score) over all the classes, and can be used to indicate whether the model generally predicts well for all classes.
393 On the other hand, weighted averages are a mean of a metric over all the classes, weighted by the number of samples
394 in each class, and can be used to assess model performance in terms of the number of images used for the assessment.
395 In addition, the combination of the two metrics is a good initial indicator for general model performance. For
396 example, a high weighted average and a low macro-average are likely an indicator of class imbalance, where the
397 model performs well for classes with many samples, but not so well against classes with a small number of samples.

Table 2. Performance of the three different top layer configurations

398 From Table 2, the MO configuration achieves a balanced average precision and recall ratio, which results
399 in the highest F1-score macro average of 89% and a weighted average of 92% among the three configurations. As
400 expected, the performances of the MC and ML configurations are highly biased to either precision or recall. For
401 MC, the model produces the most probable single class. Although the estimated class is very likely one of the true
402 classes labeled on the test image, the rest of the classes go undetected. Thus, the precision is much higher with MC
403 than with the other two configurations, while the recall is significantly lower. On the other hand, the ML
404 configuration is trained to predict any relevant classes without any knowledge of the schema hierarchy. This
405 approach often causes the model to generate more predictions than the set of true labels for each image, leading to
406 high numbers of false-positive detections. Contrary to the MC case, the ML configuration thus obtains high recall
407 values but very low precision, reducing the reliability and usability of the classifier. In addition, the effects of
408 training data class imbalance impact the ML configuration considerably more than the MO configuration, where

classes with a low number of training samples often perform significantly worse than classes with a higher number of training samples, especially for classes that are more visually varied (e.g., NON, SGN, CDM, and CDR)

By comparing the three configurations, it is clear that the MO model outperforms ML and MC by achieving both high precision and recall ratios while conforming to the hierarchy relationships. The MO model in combination with the hierarchy-aware prediction algorithm better preserves precision than the ML model and ensures that the output labels do not overlap with each other, preserving the hierarchical rules of the schema. Some representative sample classification results are shown in Fig. 5. Overall, the MO model is clearly both more precise and more accurate in multiple class predictions than the other two configurations.

Figure 5. Sample predictions using MC, ML, and MO configurations [Sim et al., 2015, 2017].

The MO model does have one limitation not evident in the other two models. Because the model makes an initial prediction among RIMG child classes to determine whether to generate more predictions (pertaining to BCP or BOV child classes) in Algorithm 1, some recall is lost in the case of wrong initial predictions. For example, if an image is incorrectly predicted as IRR when the true label is BCP, this results in all child classes of CPDMG and CPTYPE, or LOC becoming FN. As a result, the accuracy of BCP and BOV classification affects the performance of each of their subordinate classes. However, as shown in Table 2, the performance of BCP and BOV is a class with a quite high accuracy, and the risk of mis- or un-detection here is relatively low compared to the other classes.

Training and validation curves for the MO model configuration are shown in Fig. 6, where (a) is the loss and (b) is the F1-score. Each color in the graph denotes the loss or F1-score for a single category group (e.g., RIMG, CPTYPE). The training loss and F1-score are plotted as solid lines, while validation ones are plotted as dashed lines. The curve denoted as “All” and “Val. All” in Fig. 6(a) represents the summation of the loss of all the categories, while in Fig. 6(b), it represents the average of the F1-score of all the categories. From Fig. 6(b), the RIMG category has the highest F1-score, followed by CPTYPE, LOC, OVDMG, and OVANG categories, and the CPDMG class performs the worst out of all the categories. The worse performance of CPDMG compared to the rest of the

categories is likely due to it containing more classes than the others while containing fewer training samples for each class.

(a)

(b)

Figure 6. Training and validation loss in (a) and F1-score curve in (b) of the model with MO configuration. The acronym Val. denotes validation. Each color represents a single category group, and the training curves are shown in solid lines, while validation curves are shown in dashed lines. "All" and "Val. All" curves represent the summation of the training and validation loss for (a), and the average of the training and validation F1-score for (b).

A confusion matrix of the trained model with the MO configuration is shown in Fig. 7. The values of each cell are calculated as the number of occurrences divided by the total number of actual labels. For example, for the cell corresponding to actual class CD0 and predicted class CD1: if there are 13 occurrences of images and a total of 150 actual CD0 labels, the value of the cell is 13/150, yielding 0.09. The diagonal values of the matrix are the recall values of each class and are the same as the recall values in Table 2. The cells with no color and number in this confusion matrix are irrelevant pairs and are not included. For example, images with IRR can be classified as BOV or BCP but cannot be classified as BOV' or BCP's child classes because the MO configuration utilizes different Softmax groups. Overall, the model predicts true classes for most test samples with high accuracy.

Figure 7. Confusion matrix of the validation image set using the model with MO configuration: confusion matrices from the six groups in the MO configurations are combined into one for simplicity.

Fig. 8 shows some representative sample predictions in (a) using the trained model with the MO configuration and visualization of activation maps using Grad-CAM++ in (b) to understand the trained model [Chattopadhay et al., 2018]. Fig. 8(a) shows cases of successful classification in the first two rows and cases of erroneous classification in the last row. Fig. 8(b) shows class activation maps for CDR, OVFRT, and MEAS classes for the first, second, and third images, respectively. The class activation maps show that the model uses logically

447 coherent portions of the scene when deciding. For example, to determine the label CDR, the model uses the rebar
448 and the column visual information. For OVFRT, the model uses a majority of the face of the building. For MEAS,
449 the model uses the presence of fingers and the ruler.

(a)

(b)

Figure 8. (a) Sample predictions using the MO configuration and (b) their analysis using Grad-CAM++ [Sim et al., 2015, 2017; Purdue University and NCREE, 2016].

450 Regardless of high overall accuracy, there is some degree of error, as seen in the confusion matrix. However,
451 these errors are reasonable and explainable. First, images with either MEAS, SGN, or IRR often contain scenes of
452 building in their background, resulting in an overlap of visual features with other classes (e.g., BCP or BOV).
453 Despite this, the model can accurately classify images falling into these three classes a large majority of the time.
454 This capability is because, despite overlapping visual features, the presence of unique visual features allows the
455 model to differentiate between these classes and other classes. For example, the second image in Fig. 5 shows that
456 the measuring tape is a key visual feature that can differentiate between the classes in BCP and MEAS. Likewise,
457 SGN and IRR can be classified by unique foreground features (e.g., sign, human) even though most regions on
458 these images contain visual features pertaining to BCP or BOV. However, the key visual features can often vary
459 wildly, for example, in terms of color, shape, or orientation. As a result, if the model is not adequately trained to
460 recognize key visual features, it could easily become confused and mispredict a given image. Second, the visual
461 indications that distinguish one class from another are often ambiguous, such as damage levels in CPDMG or
462 viewing angles in OVANG. While the damage level is categorized into discrete classes, the damage is, in fact,
463 continuous. In the subsection titled “Design of a Hierarchical Schema”, although clear guidance to label four
464 different classes in CPDMG was established in this paper, labeling damage level is somewhat subjective and often
465 difficult to assign to one specific class. For example, some images may lie at the boundary between two damage
466 levels (e.g., CD0-CD1 or CD1-CDR) and it may be difficult to assign one or the other. As a result, there is some

overlap of model predictions between CD0 and CD1, as shown in the third image in the third row in Fig. 8(a). Similarly, the OVANG has two discrete child classes, the front and canonical view. However, some images look ambiguous such as the first and second images in the third row of Fig. 8(a), which are labeled as OVFRT, but contain minor visual scenes of the other side of the building. As a result, the model struggles with predicting the correct building angle for images with such ambiguous scenes. Third, some images have insufficient resolution to determine whether the label should be ODM or ODS. BOV images are usually taken at a distance. Although the raw images likely have sufficient resolution to ensure the damaged structural components are visible, as the image is resized to the model input image size of **299 × 299** pixels, these detailed visual features may be too small for the model to detect. In the second image in the third row in Fig. 8(a), the damage class of the building is mispredicted as ODM. From afar, the building looks undamaged. On a closer look, however, there is significant damage on the first-floor columns and walls. Lastly, the co-visibility of key features of the set of mutually exclusive classes negatively affects the classification accuracy. As mentioned in the definition of BCP, structural components may be co-visible in a single image (e.g., visual features of CPWAL and CPCOL in one image). Although the class of such images is determined by the focus of the image (e.g., the image is labeled as CPCOL if the component is placed at the center), the presence of other classes' visual features can negatively impact model training and inference. Similarly, some images with OVFRT include minor portions of building sides, as shown in the first image in the third row in Fig. 8(a).

In general, some of these issues can be mitigated in part by collecting and consistently labeling more training data. However, handling real-world images for applications such as this one does involve many grey areas such as overlapping class boundaries or co-visible mutually exclusive features, leaving them as a topic for future study.

AUTOMATED RECONNAISSANCE IMAGE ORGANIZER (ARIO)

ARIO is web-based image classification, visualization, and documentation system made possible by the techniques developed by the authors (accessible at http://ario.tech.purdue.edu). Users can upload a set of images collected

from each building to ARIO. Then, ARIO will automatically classify each of the images into the designed hierarchical schema (in Fig. 1) using the classifier with the MO configuration developed in the section titled "CONVOLUTIONAL NEURAL NETWORK-BASED HIERARCHICAL CLASSIFICATION". As of the time this paper is written, the classifier is deployed on a server equipped with Intel i7-5930K CPU, NVIDIA GeForce RTX 2070 Super with 8GB of memory, and 16GB of system memory in the HPC Lab in Knoy Hall at Purdue University. The classified images are organized in the form of a web report. A sample report is shown in Fig. 9. The report includes building and event information, image content distribution, and categorized images. In addition, the report supports interactive image filtering and clustering to allow users to search for useful and relevant images across a single report or across all reports. The time needed for one user to upload images to a single report was measured, and found that the classifier server was able to return the classified image categories for each image with an average time of less than a half-second per image. The web report including the data and categorized information can be downloaded to a local machine as a pdf or HTML for offline documentation. ARIO is a web-based platform so it can be easily accessible in various operating systems (e.g., Windows, Android, Mac).

The web report page is divided into four major regions: Region A lists the general information of the report that is collected from both the image EXIF and the user inputs. Information in the image EXIF (if available, e.g., GPS, date, and time) or image composition (e.g., the number of images, presence of DWG, or GPS images) will be automatically extracted after all the image classification results are returned. If GPS coordinates in the image EXIF are available, the building of the report is also pinned to Google Map. The users are initially asked to fill in basic information (e.g., year of the event, data source, earthquake magnitude) pertaining to the "Report information" in Region A and the "Event information" in Region B. The two images in Region A are overview images representing the building. They are automatically selected as the images having the highest-class probabilities among OVFRT and OVCAN images. In Region B, the statistics of the contents (classes) of the input image set are displayed. The statistics are also automatically generated and displayed after the classification results of all the images are returned. The number of images in each category is shown in bar and pie charts. The information in Region B helps users to have a quick understanding of the contents of the image set in this report.

Figure 9. Sample report in ARIO: Note that the screenshots of the web report have been modified to display all the information in a single image.

Regions C and D display classification results and interactively visualize their information. Initially, a set of all input images are listed, with the option to filter out irrelevant classes as required. Based on the proposed schema, each image can have more than one label. For example, the third image in Region C is categorized into three different classes. The user can click the thumbnail of an image to view it in full resolution (raw). Filters in Region D enable users to quickly navigate through many classified images to easily view images with relevant labels. Set operations, *Union* and *Intersection*, can further refine the image search by leveraging images with multiple labels. The images that are filtered out using these operations are grayed out to allow users to view the relevant images without distraction.

Lastly, users can browse across multiple reports by selecting specific image categories. In the Categories tab in Fig. 10(a), the definition of each image class used in ARIO is clearly explained. All images of each class that were categorized and stored in ARIO are linked to the corresponding text (blue texts in Fig. 10(a)). Fig. 10(b) shows all images that are labeled as concrete damage (CDR). While looking through these images, users may find it necessary to examine more information about a specific image. To do so, they can simply click the image on this page, and they will be directed to the report page which contains the target image. This functionality dramatically increases the speed of image browsing and searching in the reports.

(a) (b)

Figure 10. Categories page in ARIO: (a) definition of each class is displayed and corresponding images categorized and stored in ARIO are linked, and (b) all images labeled as "Concrete damage" (CDR) in ARIO.

CONCLUSION

In this study, a novel automated image classification algorithm to support the sorting and filtering of post-earthquake reconnaissance images was developed and validated. The main contributions of this study are fourfold. First, a comprehensive hierarchical schema is designed to support the rapid categorization of images for a real-world

application that needs to leverage artificial intelligence, post-earthquake reconnaissance. The schema allows engineers and researchers to readily access useful information when the images are categorized according to the schema. Second, a multi-output DCNN model with a hierarchy-aware prediction algorithm is developed and trained to classify each image into multiple relevant classes defined in the schema. This multi-output model addresses the limitations of existing multi-class and multi-label DCNN configurations, and specifically addresses those limitations in cases when the classes are hierarchical and not mutually exclusive. Third, the multi-output DCNN model is trained and validated on a large volume of images collected from past earthquake reconnaissance missions, such as the Taiwan or Ecuador earthquake in 2016 which mostly focus on reinforced concrete buildings. The multi-output model developed herein outperforms existing models and achieves 89% and 92% of macro and weighted average F1-scores, respectively. The macro and weighted F1-scores are quite similar, indicating that the model is less likely to be significantly biased towards a specific class. Finally, the trained multi-output DCNN model is deployed on ARIO to enable collaboration between field engineers and researchers by rapidly combining several earthquake reconnaissance image sets under a single schema. It is expected that the schema and the multi-output DCNN model, now integrated into ARIO, will enable rapid image classification and documentation of the post-earthquake state of buildings. This study aims to support the use and reuse of these images for scientific purposes and building code provisions. As a result, a hierarchical schema coupled with a multi-output DCNN model will be useful for many other fields that make use of images for scientific purposes.

DATA AVAILABILITY STATEMENT

Reconnaissance images used for developing ARIO are publicly available at dataceterhub.org. The source code for the multi-output image classification model will be shared through GitHub upon acceptance of the paper. The ground-truth labels of the images will be available from the corresponding author by request.

ACKNOWLEDGEMENT

Our research team acknowledges the support from the Natural Sciences and Engineering Research Council of Canada under RGPIN-2020-03979, the National Science Foundation under Grant No. NSF 1835473, and the valuable image contributions from the Center for Earthquake Engineering and Disaster Data at Purdue University.

REFERENCES

ATC (Applied Technology Council). (1998). *Evaluation of Earthquake Damaged Concrete and Masonry Wall Buildings*.

ATC (Applied Technology Council). (2004). *Field Manual: Safety Evaluation of Buildings after Windstorms and Floods* (1st ed., Vol. 45).

ATC (Applied Technology Council). (2005). *Field Manual: Postearthquake Safety Evaluation of Buildings* (2nd ed.).

ATC (Applied Technology Council). (2018). *Seismic Performance Assessment of Buildings* (2nd ed.).

Azimi, M., Eslamlou, A. D., & Pekcan, G. (2020). Data-Driven Structural Health Monitoring and Damage Detection through Deep Leaning: State-of-the-Art Review. Sensors, 20.

Bray Jonathan D., Frost James D., Rathje Ellen M., & Garcia Fernando E. (2019). Recent Advances in Geotechnical Post-earthquake Reconnaissance. Frontiers in Built Environment, 5, 5.

Chattopadhay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2018). Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. *2018 IEEE winter conference on applications of computer vision (WACV)*, 839-847.

Datacenterhub. (2014). https://datacenterhub.org

DesignSafe-CI. (2016). Rapid experimental facility. https://rapid.designsafe-ci.org

EERI (Earthquake Engineering Research Institute). (2016). Learning from Earthquake program.

FEMA. (2019). *Post-disaster Building Safety Evaluation Guidance, Report on the Current State of Practice including Recommendations Related to Structural and Nonstructural Safety and Habitability* (FEMA P-2055). Federal Emergency Management Agency.

Gao, Yuqing, & Mosalam, K. M. (2020). PEER Hub ImageNet: A Large-Scale Multiattribute Benchmark Data Set of Structural Images. *Journal of Structural Engineering*, *146*(10), 04020198.

GNDR (Global Network of Civil Society Organizations for Disaster Reduction). (2019). *Challenge Fund: Program Overview (English)*. World Bank Group.

Jafarzadeh, R., Ingham, J. M., & Wilkinson, S. (2015). *NEES: A Database for Seismic Retrofit Construction Cost of Concrete and Steel Framed Schools in Iran*. https://datacenterhub.org/resources/252

Laughery, L. A., Puranam, A. Y., Segura, C. L. Jr., & Behrouzi, A. A. (2020). The Institute's Team for Damage Investigations. *Concrete International*, *42*(12).

National Institute of Standards and Technology (NIST). (2016). The Disaster and Failure Studies Program.

Purdue University, & NCREE (National Center for Research on Earthquake Engineering). (2016). Performance of Reinforced Concrete Buildings in the 2016 Taiwan (Meinong) Earthquake. https://datacenterhub.org/resources/14098

SAC Joint Venture. (2000). *Recommended Postearthquake Evaluation and Repair Criteria for Welded Steel Moment-Frame Buildings*. Applied Technology Council.

Shah, P., Pujol, S., Puranam, A., & Laughery, L. (2015). Database on Performance of Low-Rise Reinforced Concrete Buildings in the 2015 Nepal Earthquake, https://datacenterhub.org/resources/238

Sim, C., Song, C., Skok, N., Irfanoglu, A., Pujol, S., & Sozen, M. (2015). Database of low-rise reinforced concrete buildings with earthquake damage. https://datacenterhub.org/resources/123

Sim, C., Villalobos, E., Smith, J. P., Rojas, P., Pujol, S., Puranam, A. Y., & Laughery, L. A. (2017). *Performance of Low-rise Reinforced Concrete Buildings in the 2016 Ecuador Earthquake*.

StEER Network. (2018). Structural Extreme Events Reconnaissance Network. https://www.steer.network

Tan, M., & Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *ArXiv Preprint ArXiv:1905.11946*.

Wartman, J., Berman, J., Olsen, M. J., Irish, J. L., Miles, S., Gurley, K., Lowes, L., & Bostrom, A. (2017). The NHERI RAPID Facility: Enabling the Next-Generation of Natural Hazards Reconnaissance. *AGU Fall Meeting Abstracts*, *44*.

612 Yeum, C. M., Dyke, S. J., Benes, B., Hacker, T., Ramirez, J., Lund, A., & Pujol, S. (2019). Postevent
613 Reconnaissance Image Documentation Using Automated Classification. *Journal of Performance of Constructed*
614 *Facilities*, *33*(1), 04018103.

615 Yeum, C. M., Dyke, S. J., & Ramirez, J. (2018). Visual data classification in post-event building reconnaissance.
616 *Engineering Structures*, *155*, 16–24.

617

618 **Table 1.** Definitions of classes in a schema

Acronym	Definition
RIMG	Reconnaissance image set
DWG	Drawings relating to a building (e.g., structural, schematics, or floor plans)
GPS	GPS device containing positional information of a building
IRR	Irrelevant images (e.g., people, scenery, or animals)
NON	Non-structural assets of a building (e.g., ceilings, generators, or cables)
SGN	Signage, writings, or notices on or near buildings
WAT	Analog or digital timepiece denoting the time of inspection
MEAS	Images that record the relative size of building components and damage (e.g., ruler placed on a cracked surface)
BBC	Buildings and building components
DEP	Image depth, the distance the image is taken from a scene (e.g., building component, room, or building)
LOC	Location the image was taken (interior or exterior)
BCP	Building components such as columns, beams, or wall
BOV	Building overview of entire buildings
BSP	Building spaces, such as rooms or corridors
LOCEX	Outside of a building
LOCIN	Inside of a building
CPDMG	Building component damage, from minor damage to major damage exposing masonry or rebar.
CPTYPE	Building component types, such as beam, column, or wall
OVANG	Viewing angle where the building image was taken
OVDMG	Level of damage to the entire structure
SPTYPE	Type of space, such as room or corridor
CD0	Zero to minor damage on a building component
CD1	Moderate damage on a building component
CDR	Severe concrete damage exposing rebar
CDM	Severe damage exposing masonry
CPCOL	Column component
CPWAL	Wall component
CPBEAM	Beam component
OVCAN	A canonical view of a building
OVFRT	A front view of a building
ODM	Slight to moderate damage on a building overall
ODS	Severe damage on a building overall
SPENT	Building/room entrance
SPFFL	Building's first floor
SPROOM	Room space in a building
SPBSM	Basement of a building

619

620 **Table 2.** Performance of the three different top layer configurations

	Precision	Recall	F1-score	Number of images

Class	MO	ML	MC	MO	ML	MC	MO	ML	MC	Testing	Training
DWG	0.99	0.99	1	0.96	1	0.99	0.98	0.99	**1**	494	1970
GPS	1	0.99	1	1	1	1	**1**	**1**	**1**	244	1037
IRR	0.97	0.7	0.95	0.88	0.98	0.93	0.92	0.81	**0.94**	98	352
NON	0.83	0.38	0.77	0.89	0.97	0.87	**0.86**	0.54	0.81	38	184
SGN	0.63	0.53	0.86	0.89	0.97	0.97	0.74	0.69	**0.91**	37	134
WAT	0.95	1	1	0.9	1	1	0.93	**1**	**1**	21	75
MEAS	0.96	0.98	0.99	0.93	1	0.96	0.94	**0.99**	0.98	213	793
BCP	0.94	0.93	1	0.97	0.99	0	0.95	**0.96**	0.01	401	1624
BOV	0.98	0.98	1	0.98	0.99	0.01	**0.98**	**0.98**	0.03	289	1169
LOCEX	0.93	0.86	0.9	0.92	0.95	0.02	**0.92**	0.9	0.04	430	1715
LOCIN	0.85	0.78	1	0.9	0.95	0.1	**0.88**	0.86	0.18	260	1078
CD0	0.8	0.56	0.86	0.88	0.95	0.61	**0.84**	0.71	0.72	150	636
CD1	0.77	0.38	0.8	0.71	0.93	0.46	**0.74**	0.54	0.58	107	411
CDR	0.78	0.57	0.9	0.9	0.96	0.83	0.83	0.71	**0.86**	98	367
CDM	0.94	0.4	1	0.74	0.89	0.48	**0.83**	0.55	0.65	46	210
CPCOL	0.9	0.85	0.95	0.96	0.98	0.17	**0.93**	0.91	0.28	251	941
CPWAL	0.89	0.67	0.94	0.86	0.94	0.31	**0.87**	0.78	0.46	150	683
OVCAN	0.88	0.65	0.92	0.85	0.99	0.46	**0.86**	0.78	0.62	123	489
OVFRT	0.87	0.69	0.97	0.88	0.95	0.5	**0.88**	0.8	0.66	166	680
ODM	0.91	0.74	0.94	0.9	0.98	0.36	**0.9**	0.84	0.52	184	727
ODS	0.86	0.68	0.95	0.85	0.9	0.54	**0.85**	0.77	0.69	105	442
Macro avg.	0.89	0.73	**0.94**	0.89	**0.96**	0.55	**0.89**	0.82	0.62		
Weighted avg.	0.92	0.82	**0.96**	0.92	**0.97**	0.45	**0.92**	0.88	0.51		

621