PACIFIC EARTHQUAKE ENGINEERING
RESEARCH CENTER

PEER Hub ImageNet (ϕ-Net):

A Large-Scale Multi-Attribute Benchmark Dataset of Structural Images

Yuqing Gao

Khalid M. Mosalam

Department of Civil and Environmental Engineering
University of California, Berkeley

PEER Report No. 2019/07
Pacific Earthquake Engineering Research Center
Headquarters at the University of California, Berkeley

November 2019

PEER 2019/07
November 2019

PEER Hub ImageNet (ϕ-Net):

A Large-Scale Multi-Attribute Benchmark Dataset of Structural Images

Yuqing Gao

Khalid M. Mosalam

Department of Civil and Environmental Engineering
University of California, Berkeley

PEER Report 2019/07
Pacific Earthquake Engineering Research Center
Headquarters at the University of California, Berkeley

November 2019

ABSTRACT

In this data explosion epoch, data-driven structural health monitoring (SHM) and rapid damage assessment after natural hazards have become of great interest in civil engineering research. This report introduces deep-learning (DL) approaches and their application to structural engineering, such as post-disaster structural reconnaissance and vision-based SHM. Using DL in vision-based SHM is a relatively new research direction in civil engineering. As researchers begin to apply these concepts to structural engineering concerns, two critical issues remain to be addressed: (1) the lack of a uniform automated detection principle or framework based on domain knowledge; and (2) the lack of benchmark datasets with well-labeled large amounts of data.

To address the first issue, an automated and hierarchical framework has been proposed: the PHI-Net or ϕ-Net for the PEER Hub Image-Net. This framework consists of eight basic benchmark detection tasks based on current domain knowledge and past reconnaissance experience. The second area of concern is based on the ϕ-Net framework; a large number of structural images was collected, preprocessed, and labeled to form an open-source online large-scale multi-attribute image dataset, namely, the ϕ-Net dataset. At the time of this writing, this dataset contains 36,413 images with multiple labels.

This report introduces herein three deep convolutional neuronal networks (CNN): VGG-16, VGG-19, and ResNet-50. The architecture design and network properties, etc., are described and discussed. For benchmarking purposes, a series of computer experiments are conducted. Multiple factors are considered in comparison studies under a fair setting of hyper-parameters and training approaches, i.e., using affine data augmentation (ADA) and transfer learning (TL). All experimental results are reported and discussed, which provide benchmark and reference values for future studies by other researchers developing new algorithms. These results reveal the great potential of using DL in vision-based SHM.

Finally, the first image-based challenge in structural engineering was held by the Pacific Earthquake Engineering Research (PEER) Center during the Fall of 2018. This challenge, designated as the ϕ-Net Challenge, served as a pre-event prior to the open sourcing of the ϕ-Net dataset and attracted worldwide attention and participation from researchers from around the globe.

ACKNOWLEDGMENTS

This project was sponsored by Tsinghua-Berkeley Shenzhen Institute (TBSI). The authors greatly appreciate the assistance from these student volunteers: Shuyuan Liang, Yujia Xu, Renjie Wu, Yidong Huang, Rongbing Zhou, Boyuan Kong, Jiachen Wang, Shuai Yao, Enze Cheng, Fan Hu, Peijie Li, Hagen Tam, Matthew Yeung, Lillian Lau, Victor Khong, Ke Xu, Jiawei Chen, Wen Tang, Ran Jiang, Jianfei Yang, Pengyuan Zhai, Chrystal Chern, Chenglong Li, and Sachit Shroff. The authors thank the Pacific Earthquake Engineering Research Center (PEER) staff members (Dr. Amaranth Kasalanati, Dr. Selim Günay, Mr. Gabriel Vargas, Ms. Grace Kang, and Ms. Erika Donald) and all the experts who participated in the ϕ-Net Challenge. The authors also appreciate the proofreading work by Ms. Fan Hu and Ms. Chrystal Chern, and the excellent editorial effort by Ms. Claire Johnson of this report. Any opinions, findings, and conclusions or recommendations expressed in this report are those of the authors and do not necessarily reflect those of PEER, TBSI, of the Regents of the University of California.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

1.1 MOTIVATION

During a structure's lifecycle, it may go through different types of loading stages, from the typical daily loads to severe destructive external loads caused by earthquakes, tsunami, hurricanes, etc. These external loads will introduce different degrees of structural damage that may lead to losses, injuries, and even fatalities. Thus, structural health monitoring (SHM) in the form of rapid and automated damage assessment after natural hazards has become an important focus in the field of structural engineering. A study of how SHM has evolved demonstrates that many of the approaches and damage criteria developed were based on human visual inspection, which is usually referred to as vision-based SHM. Until recently, vision-based SHM has been limited to visual inspection by engineers with experience on damage patterns to determine the level of damage that had occurred. The current focus aims to augment and eventually replace visual inspection by engineers by an automated recognition system based on artificial intelligence (AI) technology.

Both AI and machine-learning (ML) technologies have developed rapidly in recent decades, especially in the application of deep learning (DL) in computer vision (CV) [Goodfellow et al. 2016]. The objective of ML and DL implementation is that as computers perform labor-intensive repetitive tasks, they simultaneously "learn" from performing those tasks. Both ML and DL fall within the scope of empirical study, where data is the most essential component. In vision-based SHM, using images as data media is currently an active research direction. Structural images obtained from reconnaissance efforts or daily life are playing an increasing role as the success of ML and DL is contingent on the volume of data media available. The expectation is that eventually computers will be able to realize autonomous recognition of structural damage in daily life—under service conditions—or after an extreme event—say, a large earthquake or extreme wind. Many recent studies are addressing issues related to the relatively tedious manual efforts to catalog key vision features based on human knowledge with respect to damage patterns. This is referred to as "feature engineering," which facilitates making decisions based on such features (see Torok et al. [2013]; Yeum and Dyke [2015]; Yoon et al. [2016]; and Feng and Feng [2017]).

Features extracted based on human knowledge are known as handcrafted features, which are used for decision making in these studies and align well with the concept and procedures of

ML. Handcrafted features are concise, but they may be limited according to the perspective of image recognition. In some cases, useful damage features are abstract and exist in high-dimensional spaces. Humans are limited in having only a sense of low-dimensional features, e.g., locations, colors, or edges. Due to the complexity and diverse representations of image data, handcrafted features may not align well with the general cases and usually work for fewer limited scenarios. To address these issues, DL is being considered as a solution to address these limitations. Unlike traditional ML and feature engineering, the recent boosting of DL approaches uses the concept of representation learning with its abundant number of parameters being proposed to replace the manual feature engineering. In the past decade, great strides have been made by the DL research community, achieving state-of-the-art results in many vision-recognition tasks compared with traditional approaches [Goodfellow et al. 2016]. Until now, vision-based SHM applications have not fully benefited from the data-driven CV technologies, even as interest on this topic is ever increasing. Its application to structural engineering has been hamstrung mainly due to two factors: (1) the lack of a general automated detection principles or frameworks based on domain knowledge; and (2) the lack of benchmark datasets with well-labeled large amounts of data.

Generally, past studies were conducted within a very limited scope that treated damage detection tasks independently and waived the connections with the procedures of post-disaster reconnaissance or field inspection. Instead of borrowing technologies from computer science, insight unique to structural engineering should populate the domain knowledge to build a general and systematic framework. Detection tasks should be performed in a logical, step-by-step manner, analogous to how building design codes are developed.

The lack of uniform and quantified definition of structural attributes, e.g., structural damage patterns, increases the difficulties of labeling data for analysis. Benchmark detection tasks should be defined; most of the current DL applications in SHM fall into the scope of *supervised learning*[1] where labels and annotations of the data are crucial. In the CV domain, there exists several large-scale open-sourced datasets, e.g., ImageNet [Deng et al. 2009], MNIST [LeCun et al. 1998], and CIFAR-10 [Krizhevsky 2009], with about 15 million, 70,000, and 70,000 images, respectively.

Given that it is costly and time-consuming to obtain relevant labeled data applicable to structural engineering, establishing benchmark detection tasks has been another roadblock. Labeling structural damage data requires a significant amount of domain-specific professional knowledge and experience in structural engineering. Moreover, newly developed algorithms and novel applications in CV are usually tested first on benchmark datasets to compare the performance enhancement with the benchmark results. Since vision-based SHM is a relatively new direction in earthquake engineering research, researchers have been limited to using datasets that were collected or accessed individually. The lack of sharable and benchmark datasets for

[1] A machine-learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labeled training data consisting of a set of training examples.

both comparison's sake and validation of the datasets is yet another impediment to the training of DL models.

To address the above-mentioned two drawbacks, this report describes recent effort to build a large-scale open-sourced structural image database: the PEER[2] Hub ImageNet (PHI-Net or ϕ-Net). As of November 2019, this ϕ-Net dataset contains **36,413** images with multiple attributes for the following baseline recognition tasks: scene-level classification, structural component-type identification, crack existence check, and damage-level detection. It is now open sourced to the public via *https://apps.peer.berkeley.edu/phi-net/*. The ϕ-Net dataset uses a hierarchy-tree framework for automated structural detection tasks based on past experience from reconnaissance efforts for post-earthquakes and other hazards. Through a tree-branch mechanism, each structural image can be clustered into several sub-categories representing detection tasks. This acts as a sort of a filtering operation to decrease the complexity of the problem and improve the performance of the automated applications of the algorithms. To the best of the authors' knowledge, until now there is no open-sourced structural image dataset with multi-attribute labels and this volume of images in the vision-based SHM area. It is believed that this image dataset and its corresponding detection tasks and framework will provide the necessary benchmark for future studies of DL in vision-based SHM.

Inspired by the ImageNet Large Scale Visual Recognition Challenge (ILSVRC)[3], also known as the ImageNet Challenge [Deng et al. 2009], PEER organized the first image-based structural damage recognition competition as a pioneering study in the structural engineering area. This competition, referred to as the PEER Hub ImageNet (ϕ-Net) Challenge, was held from 23 August 23 2018 to 25 November 2018 as a pre-event prior to open- ourcing the ϕ-Net dataset. In the ϕ-Net Challenge, PEER provided the beta version of the ϕ-Net dataset and encouraged participants to develop their own algorithms and models for structural image classification to contribute to the establishment of automated vision-based SHM. The state-of-the-art algorithms tested in the ϕ-Net Challenge are expected to enhance the accuracy and generalization of vision-based SHM approaches. These approaches, including the effort documented in this report, aim towards the construction of a structural image dataset populated with enough images to construct robust algorithms and assess with confidence the built environment post-event.

1.2 RELATED WORK

Convolutional neural networks (CNN) have been at the heart of recent advances in DL. Compared with traditional CV and ML approaches, CNN no longer requires handcrafted low-level features or feature engineering; millions of parameters inside a typical network are capable of learning countless mid- to high-level image representations with input data obtained from a pixel matrix (tensor) data. Another unique characteristic of deep CNN is its depth of architecture.

[2] Pacific Earthquake Engineering Research Center
[3] *http://www.image-net.org/challenges/LSVRC/*

Many well-designed CNN architectures—such as VGGNet (Visual Geometry Group) [Simonyan and Zisserman 2014], GoogleNet [Szegedy et al. 2015], and Deep Residual Net [He et al. 2016]—have demonstrated great performance improvement with a substantial increase of the network depth. That said, issues like degradation and bias-variance tradeoff for very deep networks should be given more attention [He et al. 2016]. Although CNN has been used since the 1990s to solve handwritten digit recognition tasks [LeCun et al. 1989; 1998], its recent application on a much bigger scale wide can be attributed to advances in computing technology, such as the high-performance Graphic Processing Unit (GPU).

In addition, the ImageNet Challenge [Deng et al. 2009] has also played a big role in advertising its applicability. The ImageNet Challenge, which has been held for several years, is aimed at evaluating algorithms for object detection and image classification. Currently, more than 15 million images have been collected from a variety of domains and labeled, with over 1000+ classes. In the ImageNet Challenge of 2014, VGGNet, the model developed by VGG at the University of Oxford [Simonyan and Zisserman, 2014], achieved first place in localization tasks and second place in classification tasks; the VGG Model not only performed well for ImageNet dataset, but also had good generalization properties applicable to other datasets. In the ImageNet Challenge of 2015, Residual Net (ResNet), a new CNN architecture developed by a Microsoft research team [He et al. 2016], achieved first place in all main competition tracks, indicating great improvement of recognition accuracy compared to previous models and methods via its "shortcut connection" design. Boosted by the ImageNet Challenge, a continuously increasing number of deep CNN architectures are currently under development to achieve better performance. In the above-mentioned studies using images, several basic tasks are commonly performed, namely, image classification, object localization, object detection, and semantic segmentation[4]. If extended to vision/image-based SHM, classification for damage states and localization for damage patterns can be thought of as straightforward applications.

There has been an increasing trend of applying CNN technology in post-disaster reconnaissance and SHM. Soukup and Huber-Mörk [2014] applied CNN to detect railway defects; Yeum et al. [2016] applied VGG-F (F for "fast") CNN architecture to collapse classification based on large-scale images collected from reconnaissance efforts; Cha et al. [2017a; 2017b] used a deep CNN to detect concrete cracks without calculating the defect features and proposed a region-based algorithm for detecting multiple damage types; Zhang et al. [2017] designed a new CNN architecture, namely, CrackNet, for pavement crack detection at the pixel level; and Vetrivel et al. [2017] combined deep CNN and 3D point cloud[5] technologies to detect façade and roof damage of buildings; Xue and Li [2018] proposed a region-based full CNN to more efficiently classify and localize tunnel lining defects in geotechnical engineering.

Most of these studies have focused mainly on the binary problem of whether the structure is damaged or not and analyzed the images at the pixel level. In contrast, images collected by

[4] An image classification conducted at the pixel level.
[5] A set of data points in space produced by 3D scanners to measure many points on the external surfaces of objects.

engineers in more general reconnaissance efforts and SHM problems are usually related to the object level or even represent the whole structure; this makes utilizing current CNN technology more difficult to use without first conducting elaborate image pre-processing. A few researchers have tackled this problem, including Yeum et al. [2016], who started from the object or building level, instead of the pixel level, based on a variety of reconnaissance images. Subsequently, they studied the collapse classification problem and achieved promising results. But in reconnaissance decision making, conducting more complex tasks is essential and requires further study to determine, damage types and level of damage.

Motivated by this need for rapid information and dissemination in post-event scenarios, Gao et al. [2018] and Gao and Mosalam [2018] conducted a series of experiments on detecting structural component type and damage severity and type at the object level based on a small dataset, where transfer learning (TL) and online affine data augmentation (ADA) were applied. They demonstrated that using TL and ADA effectively alleviated the negative influence (namely, over-fitting) of labeled data deficiency and improved the performance of generalization on only 2000 images. To apply the advances made by Gao et al. [2018] to a broader range of scenarios, assembling a large volume of images for DL training is required.

As mentioned above, most studies have conducted experiments based on a small-scale or low-variety image set. Some cases conducted DL classification based on less than 10,000 images, and some studies obtained a large number of images by segmenting one high-resolution image into hundreds or thousands of small-size images. These images were homogenous and similar to each other in texture, lighting condition, etc., which may lead to a lack of generalization of the classifier to newly collected images with much less homogenous characteristics. Moreover, the above-mentioned studies reported their detection accuracies from their experiment, but some of them used different definitions for damage patterns, and each study used different datasets. This causes difficulties in comparing different algorithms and approaches, as it is well known that results of DL are sensitive to the source dataset.

Clearly, what is needed is an automated detection framework with basic/benchmark detection tasks. As reported herein, a publicly open-source, relatively large-scale structural image dataset with multi-attribute labels was proposed and populated. This dataset will serve as a benchmark dataset with well-defined detection tasks. Researchers are encouraged to test newly developed algorithms or models using this dataset with corresponding tasks and compare their performance with the benchmark results described herein.

1.3 REPORT ORGANIZATION

The report contains six chapters and one appendix. Chapter 2 describes the automated and hierarchical detection framework proposed, namely, ϕ-Net, and the corresponding large-scale multi-attribute ϕ-Net dataset, where eight basic/benchmark classification tasks are introduced with examples. Accordingly, the development procedure of such a dataset including data collection, pre-processing, labeling, and splitting is also introduced.

Chapter 3 describes three well-known DL models with their architecture designs and properties: VGG-16, VGG-19, and ResNet-50. These three models are treated herein as benchmark classifiers.

Chapter 4 describes the benchmark experiments with respect to the eight benchmark tasks. Reference values associated with comparisons and discussions of the ADA and TL are presented.

Chapter 5 describes the first image-based challenge in civil and structural engineering area, the ϕ-Net Challenge, held by PEER in the Fall of 2018.

Chapter 6 presents a summary of conclusions and future extensions.

Appendix A includes partial information on structural image collection with respect to earthquake events.

2 PEER Hub ImageNet (ϕ-Net)

Inspired by the establishment of ImageNet, a Structural ImageNet known as PEER Hub ImageNet (PHI-Net or ϕ-Net) was constructed. This image repository contains images relevant to civil engineering, such as buildings, bridges, substations, railways, etc., that catalogues structures in both the damaged and undamaged states. The establishment of ϕ-Net dataset can be used for recognition and vision-based problems in civil engineering for application in SHM concerns. The current stage of the study has kept the scope relatively narrow to address issues within structural engineering. The proposed ϕ-Net framework presented herein is intended to be used for detection and recognition of structural properties only.

2.1 FRAMEWORK OF ϕ-NET

Analogous to the classification and localization tasks in the ImageNet Challenge, the goal of the ϕ-Net framework was to construct similar recognition tasks designed for structural damage recognition and evaluation. Based on past experience from reconnaissance efforts (Sezen et al. [2003]; Li and Mosalam [2013]; Mosalam et al. [2014]; and Koch et al. [2015]), several issues affect the safety of structures post-event: the type of damaged component, the severity of damage in the component, and the type of damage. Because images collected from reconnaissance efforts vary broadly vary, including, different distances from objects, camera angles, and emphasized targets, it is useful to cluster these issues into different levels. That is, images taken from a very close distance or only containing part of the component belong to the pixel level; major targets in images such as single or multiple components belong to the object level; and images containing most of the structure belong to the structural level. Moreover, the corresponding evaluation criteria will be different for different levels: that is, images in the pixel level are more related to the material type and damage status; images on the structural level are more related to the structural type and failure status.

In a preliminary step in using visual data for damage classification, Yeum et al. [2016] proposed a detection method with four steps: metadata filtering, scene classification, object detection, and damage evaluation. Alternatively, proposed herein is a new processing framework with a hierarchy-tree structure shown in Figure 2.1, where images are classified as follows: (1) a raw image is clustered to different scene levels; (2) according to its level, corresponding recognition tasks are applied layer by layer following this hierarchy structure; and (3) each node is seen as one recognition task or a classifier, and the output of each node is seen as a characteristic or feature of the image to help with further analysis and decision making if required.

In the current ϕ-Net, we designed the following eight benchmark classification tasks:

1. three-class classification for scene level;
2. binary classification for damage state;
3. binary classification for spalling condition (material loss);
4. binary classification for material type;
5. three-class classification for collapse mode;
6. four-class classification for component type;
7. four-class classification for damage level and
8. four-class classification for damage type.

The proposed framework with a hierarchy-tree structure is depicted in Figure 2.1, where grey boxes represent the detection tasks for the corresponding attributes, white boxes within the dashed lines are possible labels of the attributes to be chosen in each task, and ellipsis in boxes represent other choices or conditions. In the detection procedure, one starts from a root leaf where recognition tasks are conducted layer by layer and node by node (grey box) until another leaf node. The output label of each node describes a structural attribute. Each structural image may have multiple attributes, i.e., one image can be categorized as being at the pixel level, concrete, damaged state, etc. In the terminology of CV, this is considered a multi-attribute (multi-label) classification problem. Given that this is a pilot study, these attributes are treated independently at this stage. More extensions such as the multi-label version of ϕ-Net will be updated in future studies.

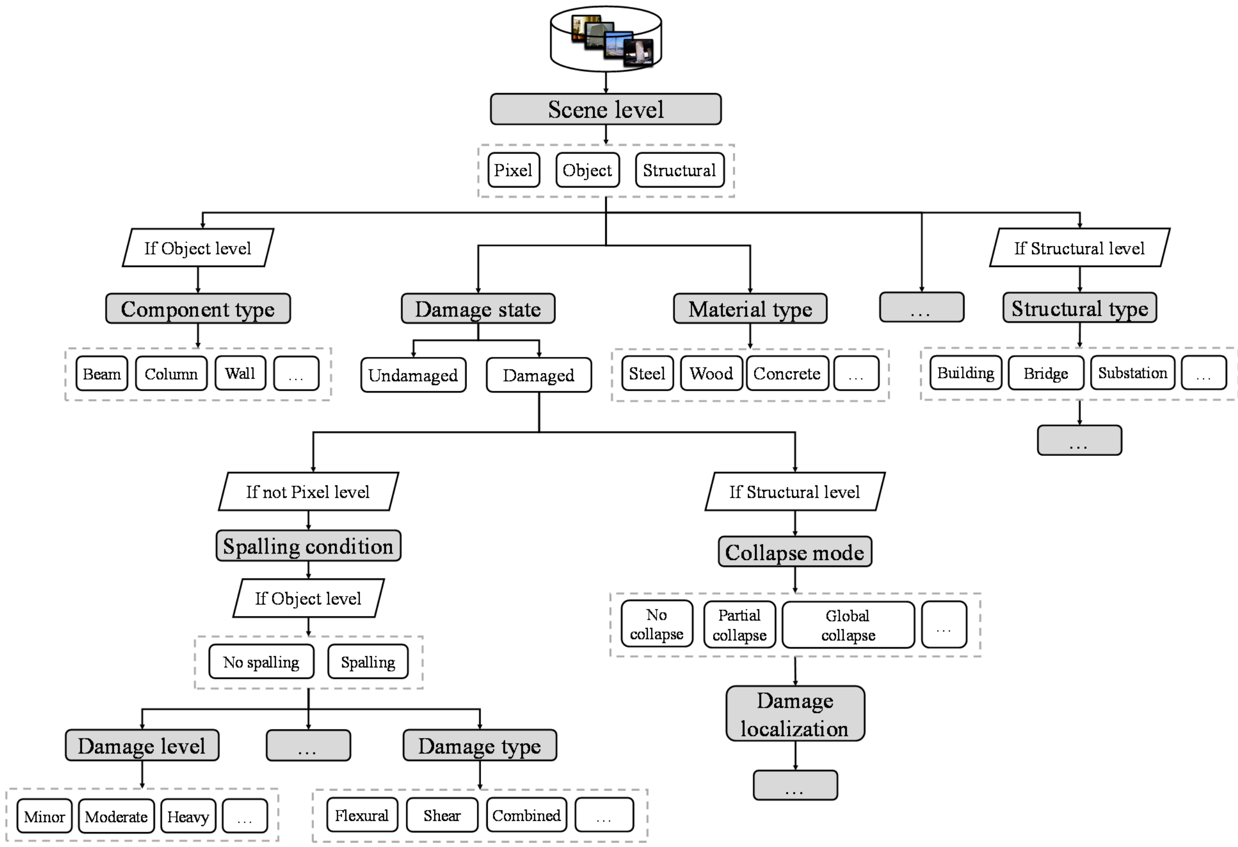
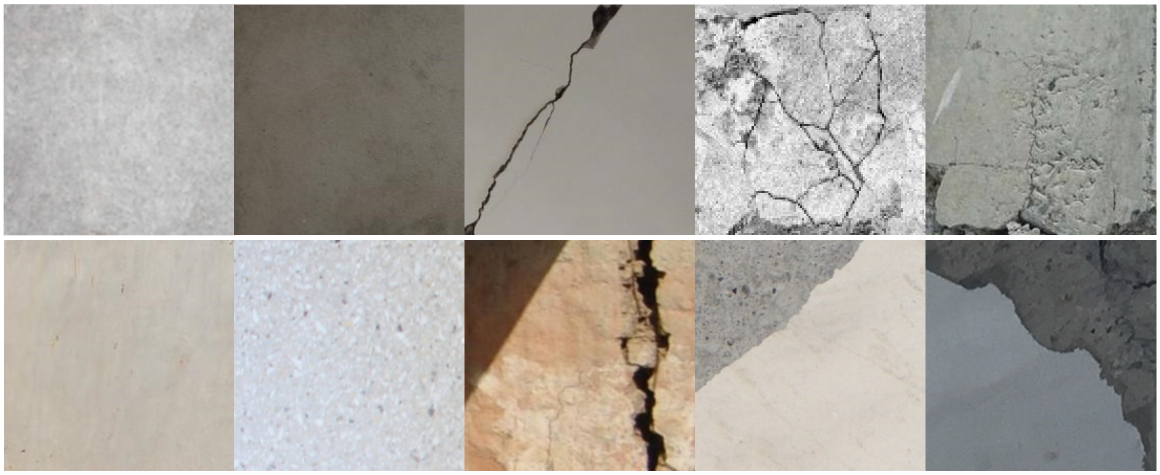


Figure 2.1 Framework of ϕ-Net.

2.2 BENCHMARK DETECTION TASKS

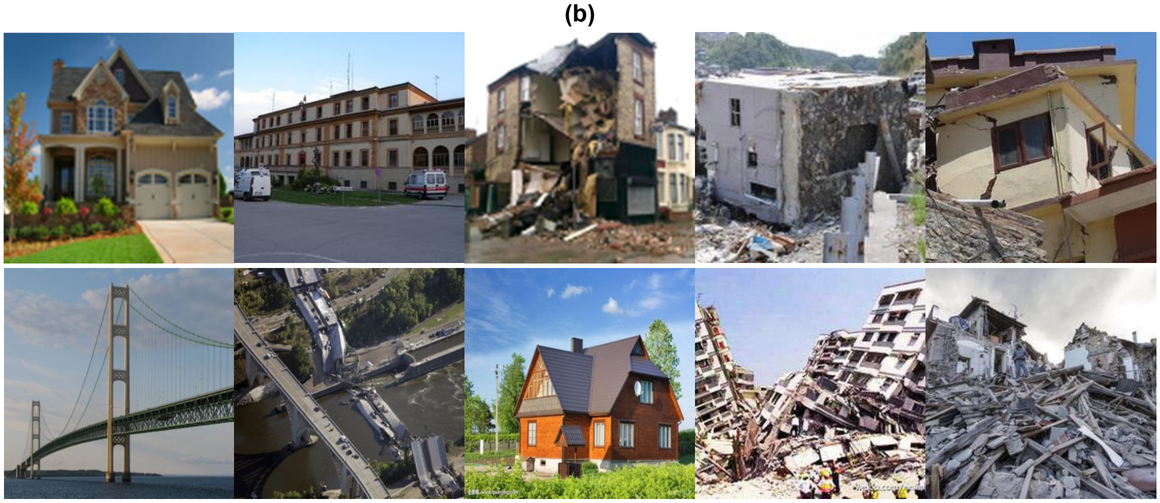
2.2.1 Task 1: Scene Level

From Figure 2.1, determining the scene-level classification is the initial task; subsequent tasks are pursued according to different levels. Because images collected from reconnaissance efforts broadly vary in terms of distance, camera angles, emphasized targets, etc., the initial establishment of the scene level can help decrease the doubt-of-scale issue in images and reduce the complexity of subsequent tasks. Through filtering out the irrelevant data and dividing images more granularly into different levels, the classifiers for subsequent tasks can achieve improved recognition performance. In the current ϕ-Net framework, the benchmark scene-level task is defined into three classes: pixel level (P), object level (O), and structural level (S). An image taken from a very close distance or only containing part of the component belongs to "P" ; see Figure 2.2(a). An image taken from mid-range and involving major targets such as a single structural component or multiple components belongs to "O"; see Figure 2.2(b). An image taken from a far distance and containing most parts of the structure or identifying its outline belongs to "S"; see Figure 2.2(c).



(a)

(b)



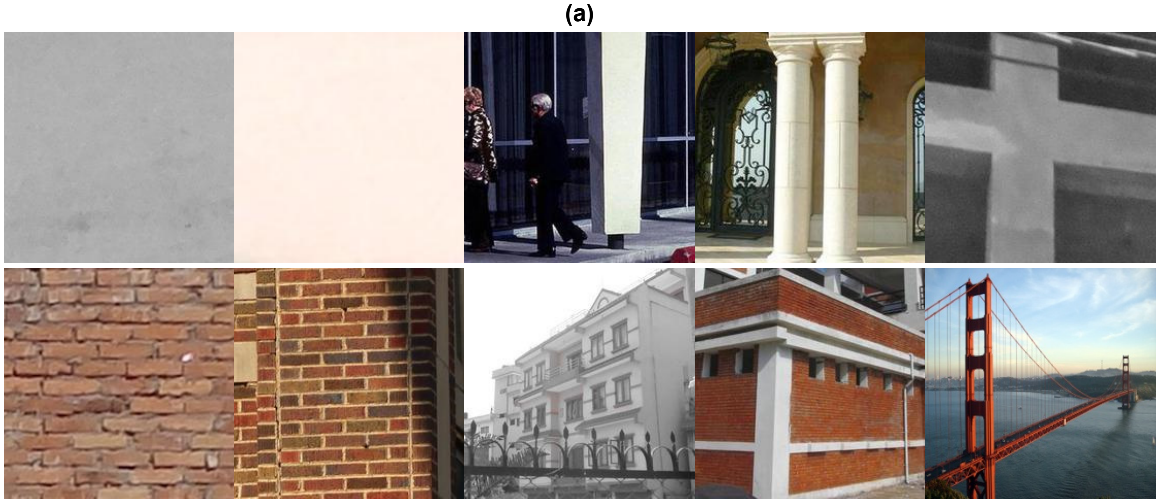
(c)

Figure 2.2 **Sample images used in scene-level detection: (a) pixel level; (b) object level; and (c) structural level.**

2.2.2 Task 2: Damage State

In the vision-based SHM, the damage state is one of the most important indices for structural health. In this task, the definition is straightforward: any observable damage pattern on the structural surface implies the damaged (D) case [shown in Figure 2.3(a)]; otherwise it is undamaged (UD) [shown in Figure 2.3(b)].

(a)



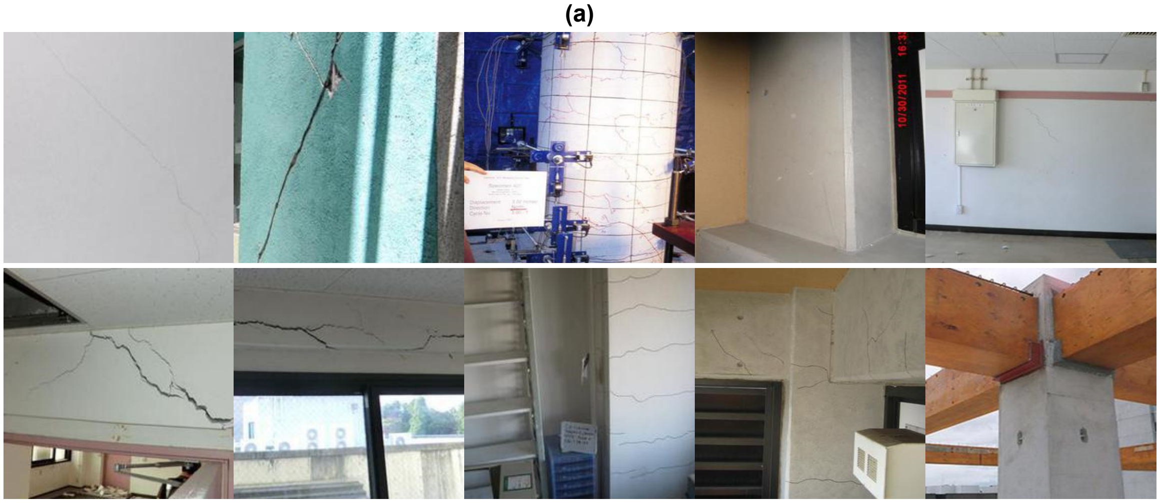
(b)

Figure 2.3 Sample images used in damage state detection: (a) damage state; and (b) undamaged state.

2.2.3 Task 3: Spalling Condition (Material Loss)

Spalling is usually defined as flakes of material that break off from a component or pertains to loss of cover of the component's surface. It often occurs in concrete and masonry structures, which may lead to severe consequences due to a reduction in the cross section and acceleration of corrosion in reinforcing steel bars. Spalling can be induced by expansion forces produced chemically or mechanically. Note: most images of spalling shown herein were collected from post-earthquake reconnaissance efforts. In other words, the spalling patterns in the ϕ-Net dataset are mainly due to earthquake loading. In this task, the spalling condition is a binary classification task with only two classes: spalling (SP) [shown in Figure 2.4(a)] and non-spalling (NSP) [shown in Figure 2.4(b)].

(a)



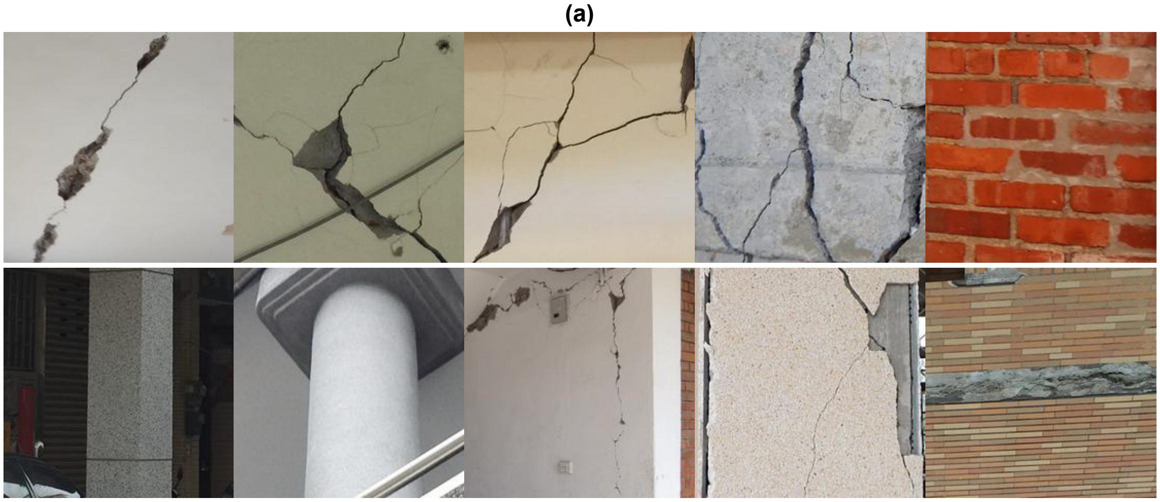
(b)

Figure 2.4 Sample images used in spalling condition (material loss) detection: (a) spalling; and (b) non-spalling.

2.2.4 Task 4: Material Type

Due to their mechanical properties, structural materials have a significant impact on structural response. Thus, identifying the type of material used for the structural component is important. The commonly used materials are steel, concrete, masonry, and wood. One of the difficulties in this task is that structural components are usually covered with non-structural components, e.g., plaster, making it difficult to accurately identify the material type from surface images. Having said that, some structures and components are made from exposed steel members such as braces and plates, which makes it simply to identify them based on their shape and surface texture. In this task, simplifications were made to limit the classification to two types: steel (ST) [shown in Figure 2.5(a)] and other materials (OM) [shown in Figure 2.5(b)].

(a)



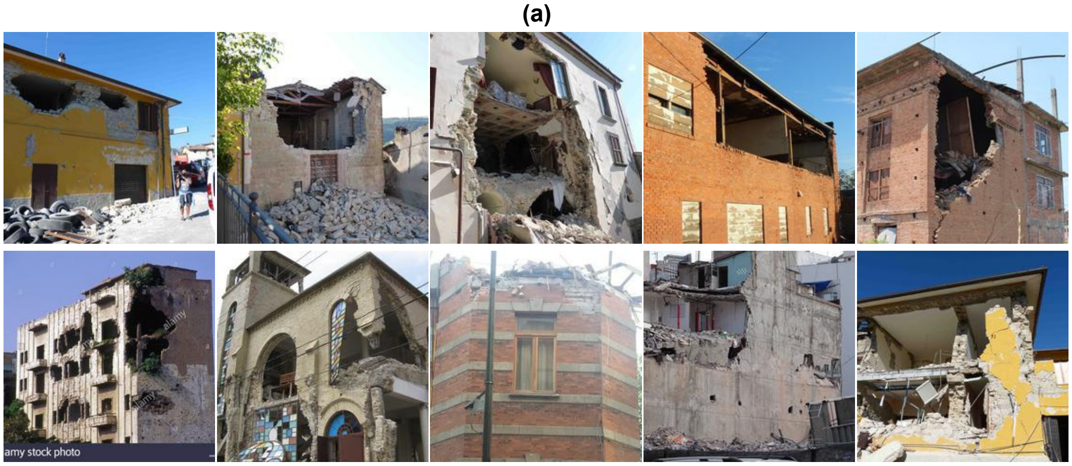
(b)

Figure 2.5 Sample images used in material type detection: (a) steel; and (b) other materials.

2.2.5 Task 5: Collapse Mode

Due to different photographic distances, camera angles, etc., it is reasonable to evaluate the severity of damage based on different scene levels. In the view of structural-level images, which contain the global information of the structure, the global performance of the structure, i.e., the collapse mode, is of the greatest concern. In terms of structural mechanics, collapse mode herein is defined as non-collapse (NC), partial collapse (PC), and global collapse (GC): although the NC [shown in Figure 2.6(a] includes undamaged or slightly damaged patterns, the structure remains intact; PC [shown in Figure 2.6(b)] corresponds to only part of the structure having collapsed while the remaining parts are still intact; and GC [shown in Figure 2.6(c)] represents catastrophic damage in the structure or evidence of permanent global excessive deformation.

(a)



(b)

(c)

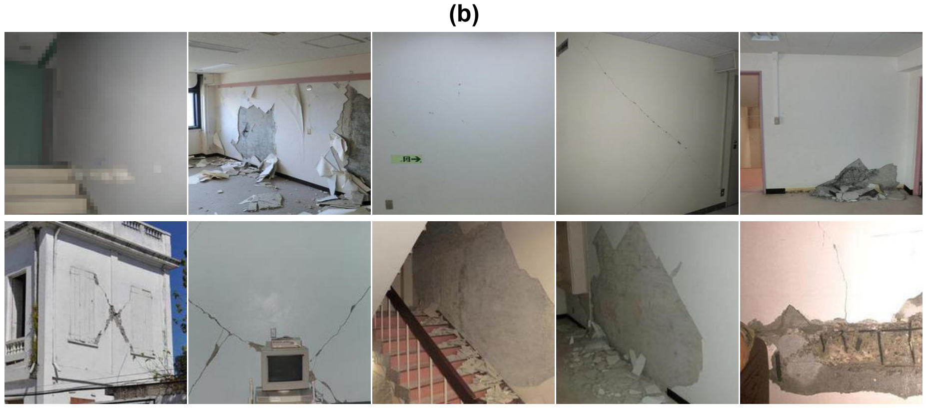
Figure 2.6 Sample images used in collapse mode detection: (a) non-collapse; (b) partial collapse; and (c) global collapse.

2.2.6 Task 6: Component Type

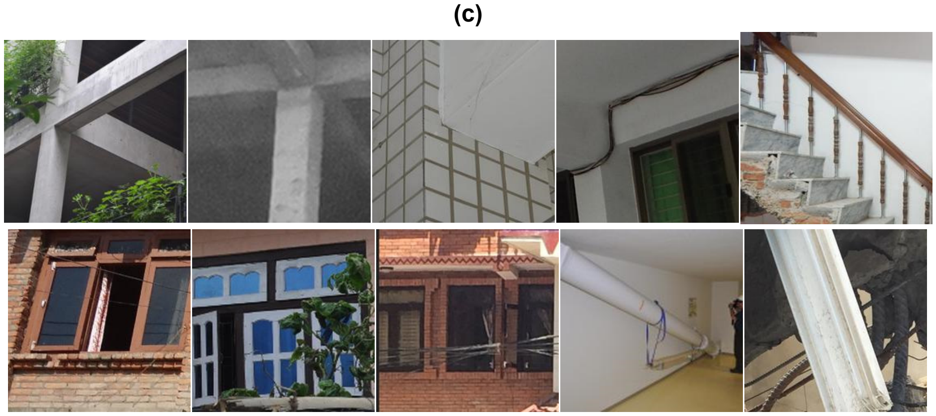
In a structural system, different structural components play different roles. Vertical components, i.e., columns and shear walls, provide lateral stiffness to resist lateral force induced by earthquakes; horizontal components, i.e., beams, diaphragms, and joints, transfer the horizontal load to the vertical components. In the seismic design of buildings, engineers have adopted concepts like the "strong-column–weak-beam" and "strong-joint–weak-component" to define the role they play in providing resistance to ground shaking [Moehle 2014]. Thus, identifying the type of structural component is informative for SHM. From the hierarchical relationship established in the ϕ-Net framework, structural component-type identification is considered as a subsequent task at the object level by four classes: Beam [shown in Figure 2.7(a)], Column [shown in Figure 2.7(b)], Wall [shown in Figure 2.7(c)], and Others [shown in Figure 2.7(d)]. Note: it is challenging to identify from images whether a wall is a shear wall or an infill wall even though there are major differences in how they provide lateral stiffness to the system. Thus, both types are grouped herein into one category. For components other than beams, columns, or walls, they are treated as other structural components, e.g., joints, staircases, braces, or even non-structural component such as windows, doors, etc.

(a)

(b)



(c)



(d)

Figure 2.7 **Sample images used in component type detection: (a) beam; (b) column; (c) wall; and (d) others.**

2.2.7 Task 7: Damage Level

This task refers specifically to damage-level evaluation of pixel and object-level images to complement the collapse mode task at the structural level, which is defined by four classes: Undamaged (UD), minor damage (MiD), moderate damage (MoD), and heavy damage (HvD); UD here is the same as that defined in the Task 2 damage state. As shown in Figure 2.8(a), MiD implies that there are only small and narrow cracks or a few spots where very minor spalling occurred on the cover of structural components. As the extent of the damage increases, i.e., cracks become wider and evidence of more extensive spalling occurs in areas but without failure, indicate MoD; see Figure 2.8(b). Shown in Figure 2.8(c), HvD means the damage area is large, and the structural component is approaching failure.

Figure 2.8 Sample images used in damage level detection; (a) minor damage; (b) moderate damage; and (c) heavy damage.

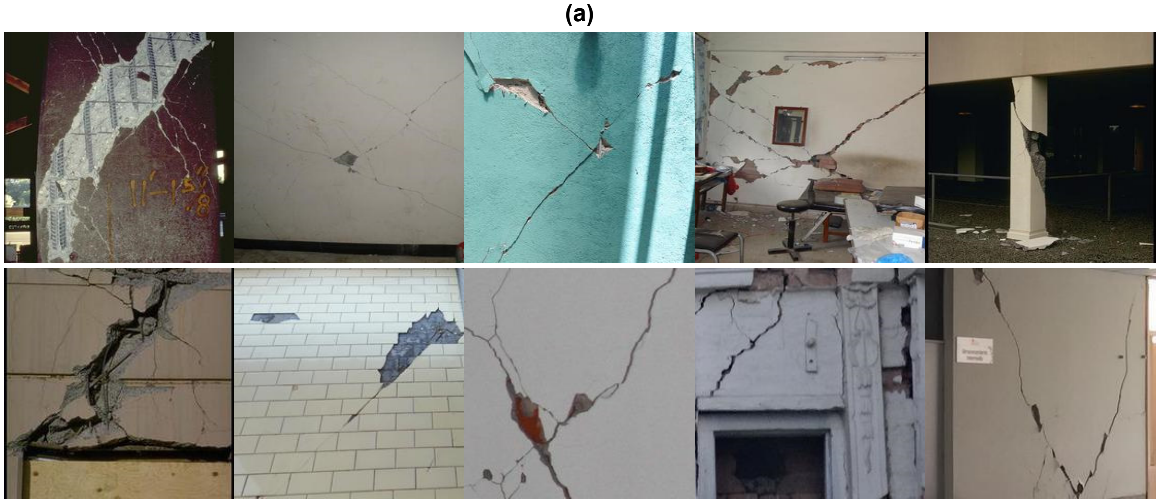
2.2.8 Task 8: Damage Type

Damage type is a more complicated and abstract classification task. Not restricted to a specific damage representation, e.g., reinforcement bar buckling, concrete spalling, etc., herein damage type refers to a complex and abstract semantic vision pattern, defined as flexural (FLEX), shear (SHEAR), and combined (COMB). These damage types have direct implications on mechanical properties and seismic design. Analogous to the failure types introduced in Moehle [2014], that is, flexural, shear, etc., and based on engineering judgment, label definitions were assigned as follows:

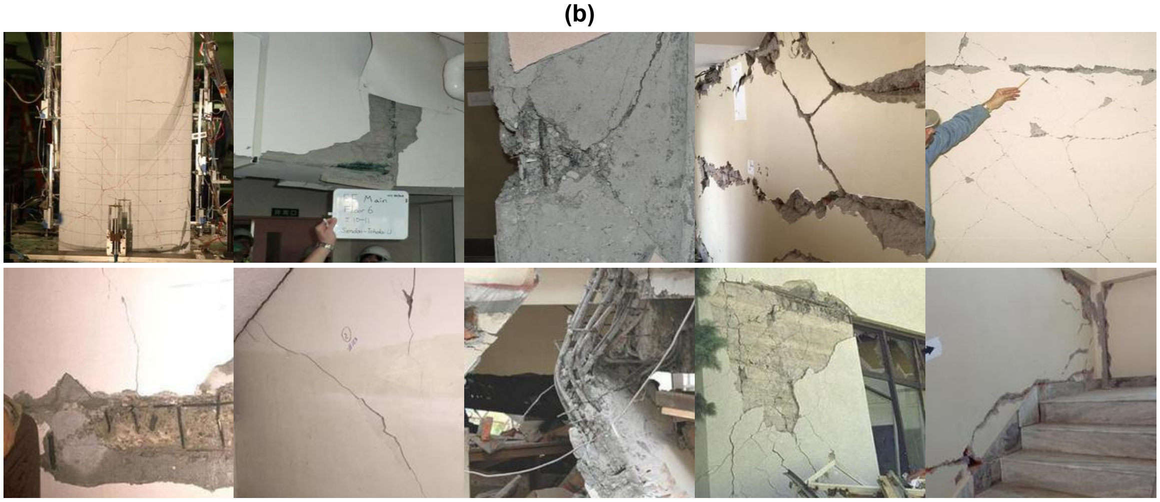
1. If most cracks occur in the horizontal or the vertical directions or at the end of a component with respective vertical or horizontal edges, it was considered flexural-type damage [shown Figure 2.9(a)];
2. If most cracks occur in a diagonal direction, or form an "X" or "V" patterns, it was considered shear-type damage [shown Figure 2.9(b)]; and
3. If the distribution of cracks was irregular or accompanied with heavy spalling, it was considered combined-type damage [shown Figure 2.9(c)].

According to Moehle [2014], FLEX represents ductile failure; it is a more desired pattern of failure compared to SHEAR, which leads to brittle failure. Thus, it is expected that damage types defined with concepts from seismic design will be very useful for the decision-making procedure.

(a)



(b)



(c)

Figure 2.9 Sample images used in damage-type detection; (a) flexural damage; (b) shear damage; and (c) combined damaged.

2.3 DATA COLLECTION, PRE-PROCESSING, LABELING, AND SPLITTING

2.3.1 Data Collection

Constructing a large-scale image dataset requires a large number of images; ImageNet contains over 15 million images. Researchers in structural engineering typically have large numbers of photographs taken from reconnaissance efforts or experimental tests, which are good resources for DL training. Moreover, there exist online resources in the form of databases and search engines that store raw structural images of pre- and post-disaster status. Unfortunately, these images are not labeled. In addition to self-taken and donated images, raw images for inclusion in the ϕ-Net dataset were collected from NISEE library archive[6], DESIGN SAFE[7], EERI Learning from Earthquake (LFE) Archive[8], Baidu Image[9], and Google Image[10]. As of June 2018, over 100,000 structural images with variant qualities were collected; a large number of them were photographed after recent earthquakes, e.g., 2017 Mexico City earthquake. More information related to the names and dates of earthquakes is listed in Table A.1 in Appendix A. As of this writing, multiple worldwide sources contributed to the variety of data in the current ϕ-Net dataset.

2.3.2 Data Pre-Processing

Since collected images vary broadly with differing degrees of quality, image pre-processing was conducted prior to labeling according to the following four steps.

1. Low-resolution images (lower than 448×448 pixels) and noisy images containing too many irrelevant objects, e.g., vehicles, people, furniture, etc., were manually eliminated;
2. To further increase the dataset size, especially for pixel and object-level images, subparts, e.g., beam, column, wall, etc., were cropped from some high-resolution structural images; see Figures 2.10 and 2.11;
3. To avoid significant distortions to image features due to stretching/rescaling used in training, cropping was applied to make the aspect ratio of the images roughly less than 2; and
4. To avoid images that were too low in quality and resolution, cropped images with resolutions lower than 224×224 pixels were eliminated.

[6] https://nisee.berkeley.edu/elibrary/
[7] https://www.designsafe-ci.org/
[8] https://www.eeri.org/projects/learning-from-earthquakes-lfe/lfe-reconnaissance-archive/
[9] https://image.baidu.com/
[10] https://images.google.com/

Figure 2.10 One structural level image includes multiple structural components. The solid red boxes indicate columns and dashed blue box indicates the beam.

Figure 2.11 Subparts cropped from high resolution structural level image.

2.3.3 Data Labeling

As discussed above, for benchmarking purposes and to reduce task complexity, eight benchmark tasks were independently treated. The labeling procedure followed the order from the framework shown in Figure 2.1, starting from labeling attribute in scene level and then progressing towards the corresponding subsequent tasks.

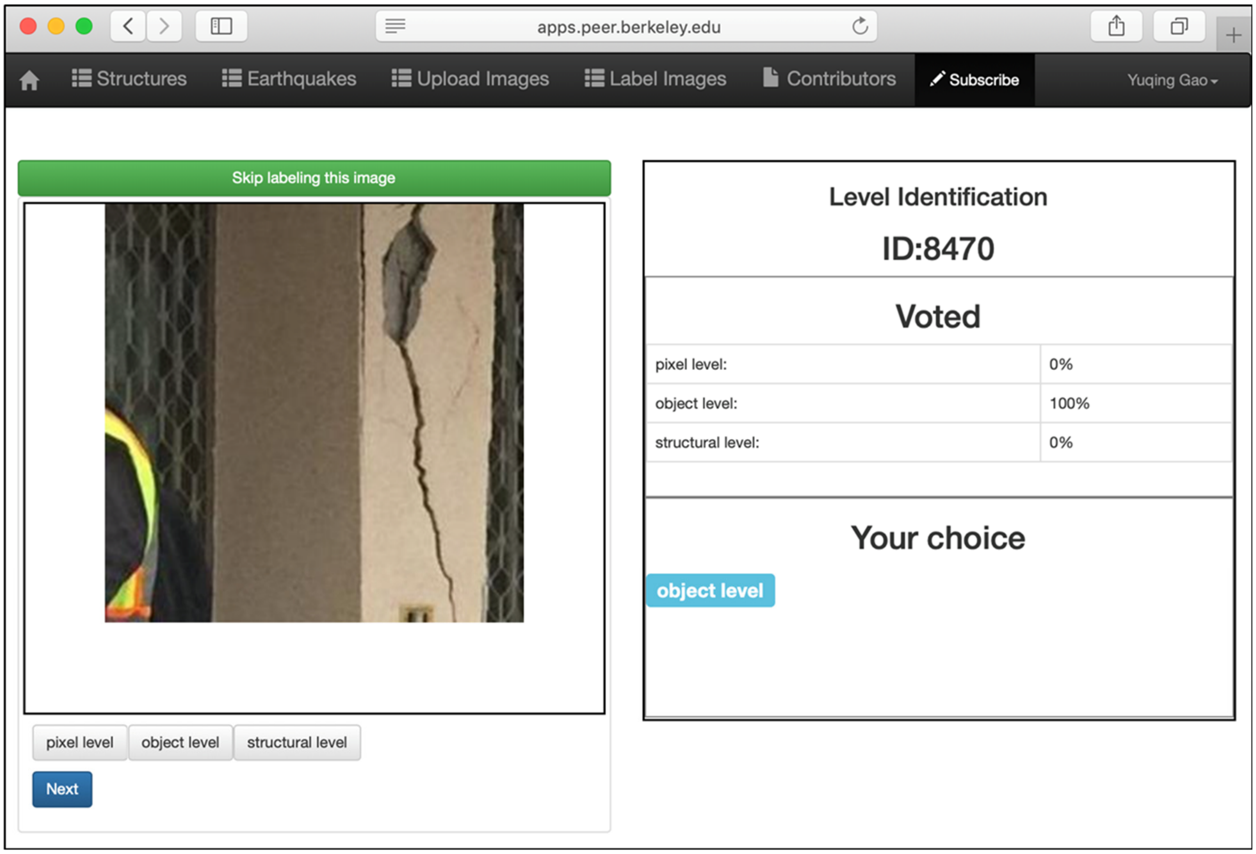


Figure 2.12 Graphic user interface of the developed online labeling tool.

The labeling procedure was crowdsourced via an online labeling tool[11], which provided a user-friendly graphic user interface for annotation where users just needed to click the button corresponding to the correct label; see Figure 2.12. To avoid bias, a majority voting mechanism was adopted where multiple people may label one image, but the final label for inclusion in the dataset was the one with the majority voting. In addition, a reference voting mechanism was introduced, as shown on the right column of Figure 2.12, which represented the percentage of votes for each label and provided opinions from others working on the same image. Based on the feedback from volunteers, the reference voting mechanism helped them in judging some ambiguous images. In some cases, where there were images unrelated to attributes or labels, or where the users were unable to make a firm decision regarding the appropriate label, a "skip" button was placed at the top of the image to bypass such images. Once the user completed labeling one attribute (task), the selected choices were shown on the bottom right corner, at which point the user could advance to the next task of the above-mentioned hierarchical relationship by clicking "next".

[11] *https://apps.peer.berkeley.edu/spo*

To assist in the labeling process, 20 volunteers with structural engineering backgrounds were recruited, in addition to worldwide contributions received from researchers and experts with the relevant background knowledge. In order to avoid bias and maintain good quality control of the labeling process in the current ϕ-Net dataset, each image was labeled with the majority voting among at least three people; it was discarded if it had less than three votes. As of November 2019, the ϕ-Net dataset has **36,413** labeled images.

2.3.4 Data Splitting

As a benchmark dataset for training and validating ML/DL models, splitting the raw dataset to deterministic training and testsets is required. Because the ϕ-Net is a multi-attribute dataset where each image contains multiple labels, the training/test split ratio for each task is different from one task to another. This not the case in common classification benchmark datasets, e.g., MNIST, or CIFAR-10. As an example, in the damage state task, 13,271 images have valid labels (UD and D), and the remaining 23,142 images are unlabeled for this task. Simply adopting a 9:1 training/test ratio across the entire image set may lead to the undesirable situation where all images with valid labels are in the training set, but no images with valid labels appear in the testset. To address this issue, a new split algorithm was developed where the training/test ratio was replaced by a range instead of a fixed number for each attribute. Usually, the range of the split ratio training/test is from 4:1 to 9:1. In this report, this split ratio is chosen in the range of 8:1 to 9:1.

Figure 2.13 shows the pseudo code of the multi-attribute split algorithm. First, according to the number of all images in the dataset (N_{total} = 36,413), a random sample of $\alpha \cdot N_{total}$ = 32,407 indices is selected (without replacement) from 1 to 36,413, which represents training indices where the split ratio factor α = 0.89 is assigned for better convergence performance after multiple trials; the remaining 4006 indices are for testing. Then, training/test ratios are checked for each attribute. If there exist any conflicts—where the ratio is out of range from 8:1 to 9:1—then the sampled indices are rejected, and the recursive loop is running again until it reaches a satisfactory split. As mentioned above, the process reported herein mainly focused on solving eight tasks individually. Thus, eight task-oriented sub-datasets for each attribute with training and test are further separated; see Tables 2.1 to 2.8.

Typically for ML and DL studies, it is suggested to keep a similar label distribution (percentage of each label with respect to all labels) among the training, test, and full datasets; therefore, distributions among these eight tasks were examined; Tables 2.9 to 2.16. All results indicated the ratios among these three datasets to be nearly consistent.

Required: function Generate_Index(x) for randomly generating x indices without replacement; function Generate_Remain_Index(N, y) to generate remaining y indices, where $N = x + y$; function Compute_Split(task_id, train_list, test_list) to compute the ratio for a specific task.

Define: trial split ratio α, number of total data N_{total}, number of training data $N_{training}$, number of test data N_{test}, and lower (Bound_L) and upper (Bound_U) bounds of the ratio.

```
Function Multi_Attribute_Split(α, N_total, Bound_L, Bound_U)
    Set done = False
    N_training ← α · N_total
    N_test ← (1 − α) · N_total
    while not done do
        training_index ← Generate_Index(N_training)
        test_index ← Generate_Remain_Index(N_total, N_test)
        # Loop over all attributes or tasks
        for task_id in [task1, task2, …, task8] do
            ratio ← Compute_Split(task_id, train_list, test_list)
            # Check if violating the boundary assumption
            if ratio < Bound_L or ratio > Bound_U do
                # Using recursion to speed up
                Multi_Attribute_Split(α, N_total, Bound_L, Bound_U)
            end if
        end for
        done ← True
    end while
    return training_index, test_index
```

Figure 2.13 Pseudo code of multi-attribute split algorithm.

Table 2.1 Number of images in scene level.

Set	Pixel level	Object level	Structural level	Total
Training	7690	8111	8508	24,309
Test	965	962	1070	2997
Total	9073	8655	9578	27,306

Table 2.2 Number of images in damage state.

Set	Damaged state	Undamaged state	Total
Training	6282	5529	11,811
Test	745	715	1460
Total	7027	6244	13,271

Table 2.3 Number of images in spalling condition.

Set	Non-spalling	Spalling	Total
Training	4294	2604	6898
Test	527	310	837
Total	4821	2914	7735

Table 2.4 Number of images in material type.

Set	Steel	Other materials	Total
Training	1806	6506	8312
Test	209	770	979
Total	2015	7276	9291

Table 2.5 Number of images in collapse mode.

Set	Non-collapse	Partial collapse	Global collapse	Total
Training	322	379	525	1226
Test	39	40	67	146
Total	361	419	592	1372

Table 2.6 Number of images in component type.

Set	Beam	Column	Wall	Others	Total
Training	511	1618	2268	358	4755
Test	60	205	265	49	579
Total	571	1823	2533	407	5334

Table 2.7 Number of images in damage level.

Set	No	Minor	Moderate	Heavy	Total
Training	1551	869	799	919	4138
Test	207	93	104	94	498
Total	1758	962	903	1013	4636

Table 2.8 Number of images in damage type.

Set	No	Flexural	Shear	Combined	Total
Training	1598	476	826	1193	4093
Test	215	46	99	132	492
Total	1813	522	925	1325	4585

Table 2.9 Label distribution in training, test, and full dataset in scene level.

Set	Pixel level	Object level	Structural level
Training	31.6%	33.4%	35.0%
Test	32.2%	32.1%	35.7%
Full	31.7%	33.2%	35.1%

Table 2.10 Label distribution in training, test, and full dataset in damage state.

Set	Damaged state	Undamaged state
Training	53.2%	46.8%
Test	51.0%	49.0%
Full	53.0%	47.0%

Table 2.11 Label distribution in training, test, and full dataset in spalling condition.

Set	Non-spalling	Spalling
Training	62.2%	37.8%
Test	63.0%	37.0%
Full	62.3%	37.7%

Table 2.12 Label distribution in training, test, and full dataset in material type.

Set	Steel	Other materials
Training	21.7%	78.3%
Test	21.3%	78.7%
Full	21.7%	78.3%

Table 2.13 Label distribution in training, test, and full dataset in collapse mode.

Set	Non-collapse	Partial collapse	Global collapse
Training	26.3%	30.9%	42.8%
Test	26.7%	27.4%	45.9%
Full	26.3%	30.5%	43.1%

Table 2.14 Label distribution in training, test, and full dataset in component type.

Set	Beam	Column	Wall	Others
Training	10.7%	34.0%	47.7%	7.5%
Test	10.4%	35.4%	45.8%	8.5%
Full	10.7%	34.2%	47.5%	7.6%

Table 2.15 Label distribution in training, test, and full dataset in damage level.

Set	No	Minor	Moderate	Heavy
Training	37.5%	21.0%	19.3%	22.2%
Test	41.6%	18.7%	20.9%	18.9%
Full	37.9%	20.8%	19.5%	21.9%

Table 2.16 Label distribution in training, test, and full dataset in damage type.

Set	No	Flexural	Shear	Combined
Training	39.0%	11.6%	20.2%	29.1%
Test	43.7%	9.3%	20.1%	26.8%
Full	39.5%	11.4%	20.2%	28.9%

2.4 DATA SUMMARY

The ϕ-Net dataset is composed of 36,413 effectively labeled images that can be treated either as a one multi-attribute (multi-label) dataset or as consisting of eight independent sub-datasets corresponding to its attribute/task; label statistics are illustrated in Figure 2.14.

The above-mentioned eight well-labeled datasets with both training and testsets online[12] are open source, and researchers can download the data according to instructions for non-commercial usage. All eight dataset files are compressed in the form of zip files and share the

[12] https://apps.peer.berkeley.edu/phi-net/

same data storage architecture containing **X_train.npy** (i.e., training data), **y_train.npy** (i.e., training label), **X_test.npy** (i.e., test data), and **y_test.npy** (i.e., test label). All image data are converted to NumPy array and can be easily loaded by multiple computer languages, e.g., Python, MatLab, etc.

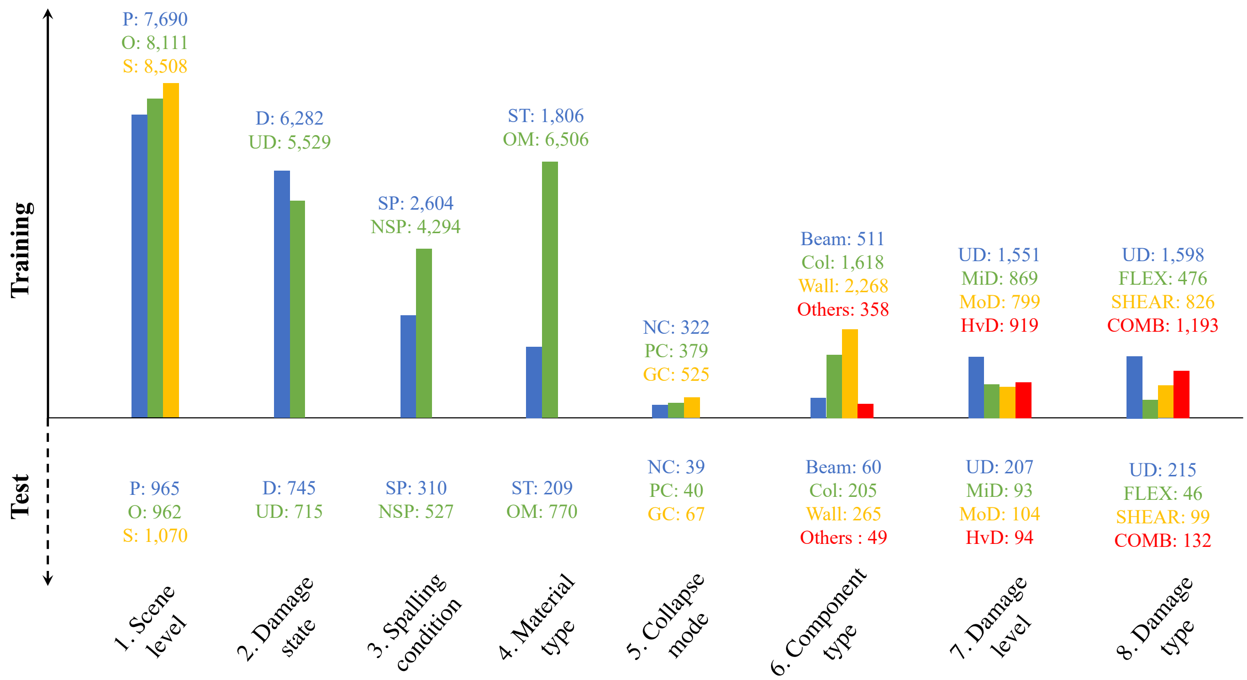


Figure 2.14 Statistics of the labels of the ϕ-Net dataset.

2.5 EXTENSIONS ON DATASET HIERARCHY

The flowchart of the ϕ-Net framework shown in Figure 2.1 indicates the hierarchital relationships between tasks, and its dependence on the engineer's domain knowledge and past experience. In a recent study by Gao et al. [2019], a new version of the framework was proposed, namely, the hierarchy ϕ-Net, which combines several branches for efficiency in computation and introduces conditional nodes (diamonds) for tree-branch selections. The grey nodes in Figure 2.15 represent detection tasks. All eight benchmark tasks are included; therefore, the ϕ-Net dataset is also suitable for this extended framework. Hierarchy ϕ-Net has a clearer order and stronger correlation between the root and leaf tasks than the original framework by additional emphasis on the two types of diamond nodes (i.e., scene-level and damage-state tasks). The remaining nodes are now connected and play the most important roles in the whole framework. Thus, it is expected that hierarchy ϕ-Net will be beneficial for future studies; however, the order and selection of node tasks are not optimal, and more studies to explore and to better utilize the hierarchical relationships are encouraged.

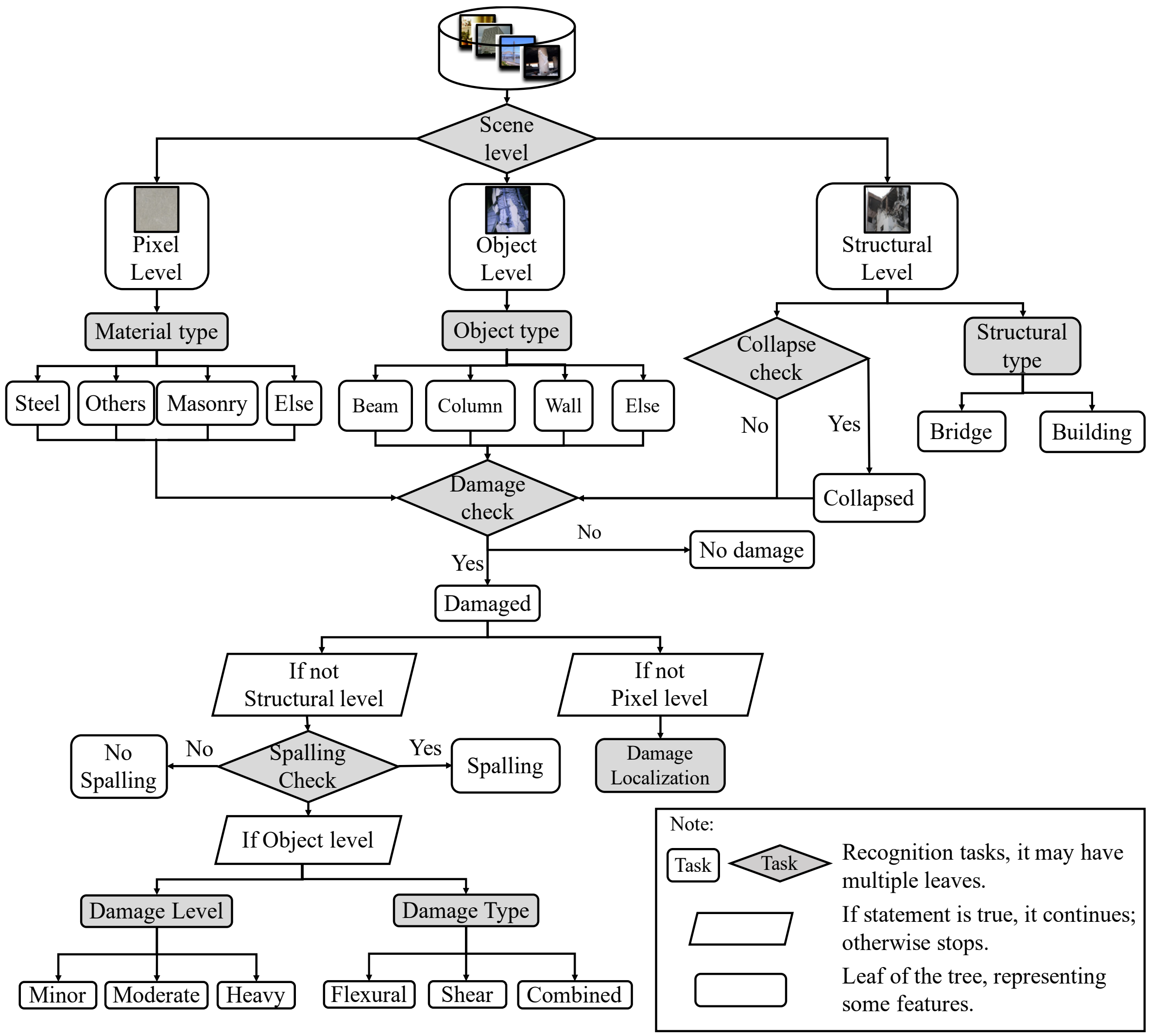


Figure 2.15 Hierarchy of the extended ϕ-Net framework.

3 Deep Convolutional Neural Networks

As mentioned in previous chapters, several famous Deep CNN architectures have been proposed and validated effectively, e.g., VGGNet, and ResNet. This chapter introduces both VGGNet and ResNet as benchmark DL classifiers.

3.1 VGGNET

Inherited from the two well-known architectures of LeNet [LeCun et al. 1998] and AlexNet [Krizhevsky et al. 2012], VGGNet uses similar configurations such as multiple convolutional (conv) blocks, two fully connected layers (*fc*-layers) representing image features, and one *fc*-layer for classification usage. Compared to previous network architectures, VGG goes much deeper with most CNN layers—exemplifying the concept of "very deep network" [Simonyan and Zisserman 2014] —that provides a direction for design and efficient usage of CNN. The most prominent contribution of the VGG architecture is to prove that by applying a small filter size (or kernel size) such as 1×1, or 3×3, and increasing the depth of the network, one can effectively improve model performance. The pre-trained VGG Model on ImageNet has very good generalization on other datasets and provides an environment for a transfer learning (TL) study, where the DL classifier can learn from a source domain to transfer the learning to another target domain with less data.

The input image size used in Simonyan and Zisserman (2014) is fixed at 224×224; pixel intensity in an image is reduced by the mean RGB channel as the only preprocessing procedure. For the conv layer [Goodfellow et al. 2016], only a small filter size (typically 3×3 or even 1×1) is used for the convolution operation. In order to maintain the size of the feature maps (generated by convolution), the stride is taken as one with zero padding while performing the convolution. Pooling layers can reduce the dimensionality of the representation and create an invariance to small shifts and distortions. In order to save computational time and pursue better performance of translation invariance, after several rounds of convolution, max pooling layers are added to the conv layers. The stride for the max pooling is 2 with a 2×2 window size. Max pooling has some positive effects on extracting superior invariant features [Nagi et al. 2011], which might help with identifying features like cracks. Usually there are multiple conv layers before the pooling layer. This report denotes a conv block as several conv layers ended by a max pooling layer.

In the original design [Simonyan and Zisserman 2014] after a series of conv blocks, three *fc*-layers follow, where each of the first two layers has 4096 neurons and the third one has 1000 neurons matching 1000 classes of the detection targets (also called a Softmax layer). According to different sizes of datasets and different qualities of images between the original ImageNet and the current ϕ-Net dataset, some modifications were made and tested, including the following: (1) the shape of the Softmax layer was changed according to the number of classes to be identified in the classification tasks, e.g., for binary damage detection, the shape of the Softmax layer was (N, 2) where N represents the number of input data; and (2) different neurons and different numbers of *fc*-layers were applied after the conv blocks, which is also known as the adaptation layer.

According to different numbers of conv and *fc*-layers, VGGNet has several variations, e.g., VGG-16, and VGG-19. The detailed configurations including filter size and output shape for each layer of the two kinds of VGGNet adapted in the ϕ-Net dataset are listed in Tables 3.1 and 3.2. Here, the only difference between the adapted VGG-16 and VGG-19 is that VGG-19 added one more conv layer in conv blocks 3, 4, and 5, i.e., it has 3 more layers than VGG-16 (which explains its name).

In order to avoid overfitting, dropout [Srivastava et al. 2014] is applied in the adaptation layers (*fc*-layers), with probability of p = 0.5. For simplicity, all the hidden layers were applied with the activation function ReLU [Nair and Hinton 2010], which favorably increases the nonlinearity of the network. Given that VGGNet is not only a powerful classifier but also performs well in localization and object detection tasks [Simonyan and Zisserman 2014], it is proposed to learn more about these capabilities of the detection tasks using future versions of ϕ-Net.

Table 3.1 Adapted VGG-16 configuration.

Block	Layer (type)	Filter size (#)	Output shape
Input	Input Image	-	(N, 3, 224, 224)
Conv Block 1	Convolutional	3×3 (64)	(N, 64, 224, 224)
	Convolutional	3×3 (64)	(N, 64, 224, 224)
	Max Pooling	-	(N, 64, 112, 112)
Conv Block 2	Convolutional	3×3 (128)	(N, 128, 112, 112)
	Convolutional	3×3 (128)	(N, 128, 112, 112)
	Max Pooling	-	(N, 128, 56, 56)
Conv Block 3	Convolutional	3×3 (256)	(N, 256, 56, 56)
	Convolutional	3×3 (256)	(N, 256, 56, 56)
	Convolutional	3×3 (256)	(N, 256, 56, 56)
	Max Pooling	-	(N, 256, 28, 28)
Conv Block 4	Convolutional	3×3 (512)	(N, 512, 28, 28)
	Convolutional	3×3 (512)	(N, 512, 28, 28)
	Convolutional	3×3 (512)	(N, 512, 28, 28)
	Max Pooling	-	(N, 512, 14, 14)
Conv Block 5	Convolutional	3×3 (512)	(N, 512, 14, 14)
	Convolutional	3×3 (512)	(N, 512, 14, 14)
	Convolutional	3×3 (512)	(N, 512, 14, 14)
	Max Pooling	-	(N, 512, 7, 7)
Fully-connected Layer	Flatten	-	(N, 25,088*)
	Dense	-	(N, 256)
	Dense	-	(N, 128)
	Dense	-	(N, 2/3/4)**

*Note 1: $512 \times 7 \times 7 = 25{,}088$.

**Note 2: N denotes the number of data, and 2/3/4 corresponds to different classes to be classified.

Table 3.2 Adapted VGG-19 configuration.

Block	Layer (type)	Filter size (#)	Output shape
Input	Input Image	-	(N, 3, 224, 224)
Conv Block 1	Convolutional	3×3 (64)	(N, 64, 224, 224)
	Convolutional	3×3 (64)	(N, 64, 224, 224)
	Max Pooling	-	(N, 64, 112, 112)
Conv Block 2	Convolutional	3×3 (128)	(N, 128, 112, 112)
	Convolutional	3×3 (128)	(N, 128, 112, 112)
	Max Pooling	-	(N, 128, 56, 56)
Conv Block 3	Convolutional	3×3 (256)	(N, 256, 56, 56)
	Convolutional	3×3 (256)	(N, 256, 56, 56)
	Convolutional	3×3 (256)	(N, 256, 56, 56)
	Convolutional	3×3 (256)	(N, 256, 56, 56)
	Max Pooling	-	(N, 256, 28, 28)
Conv Block 4	Convolutional	3×3 (512)	(N, 512, 28, 28)
	Convolutional	3×3 (512)	(N, 512, 28, 28)
	Convolutional	3×3 (512)	(N, 512, 28, 28)
	Convolutional	3×3 (512)	(N, 512, 28, 28)
	Max Pooling	-	(N, 512, 14, 14)
Conv Block 5	Convolutional	3×3 (512)	(N, 512, 14, 14)
	Convolutional	3×3 (512)	(N, 512, 14, 14)
	Convolutional	3×3 (512)	(N, 512, 14, 14)
	Convolutional	3×3 (512)	(N, 512, 14, 14)
	Max Pooling	-	(N, 512, 7, 7)
Fully-connected Layer	Flatten	-	(N, 25,088*)
	Dense	-	(N, 256)
	Dense	-	(N, 128)
	Dense	-	(N, 2/3/4)**

*Note 1: 512×7×7 = 25,088.

**Note 2: *N* denotes the number of data, and 2/3/4 corresponds to different classes to be classified.

3.2 RESNET

A deep network has more trainable parameters compared to a shallow network, and the increasing complexity might help solve more complicated problems and improve performance. Unfortunately, as demonstrated by He et al. [2016], vanishing/exploding gradients and degradation issues will occur with the increasing network depth, which make it difficult to effectively train. To address this concern, He et al. [2016] developed a new CNN architecture with a shortcut connection named the Residual Network (ResNet). As a reflection of its name, the layers are reformatted with reference to the inputs through the shortcut connection, and the network learns residual functions instead of unreferenced functions as is the case with traditional CNN. These shortcuts act like highways, and the gradients can easily flow back, resulting in faster training and support for stacking more layers. Mathematically, assuming $H(x)$ is any desired mapping function, the traditional way assumes that convolutional operations fit $H(x)$. In contrast, in ResNet these convolutional operations are used to fit the residual function (x), and then mapping can be represented as the sum of functions of input and residual terms, i.e., $H(x) = F(x) + h(x)$. If the additive term $h(x)$ is the original input, this shortcut connection is considered as identity mapping; see Figure 3.1(a). In addition, more complicated mapping was also designed for the shortcut connection using a convolutional shortcut; see Figure 3.1(b). He et al. [2016] also studied and validated more detailed derivations. The CNN structures shown in Figures 3.1(a) and (b) are denoted as an identity residual unit and a convolutional residual unit, respectively.

The benchmark experiments implemented ResNet-50, which consists of stacking multiple identity and convolutional residual units; Figure 3.2 shows the detailed CNN configuration. The first residual unit in one conv block is set as convolutional unit, and the latter units are all identity units. Similar to VGGNet, there are multiple variations of ResNet based on depth, e.g., ResNet-101 and ResNet-152, or the more advanced design, i.e., ResNeXt. Interested readers can find more information in He et al. [2016] and Xie et al. [2017].

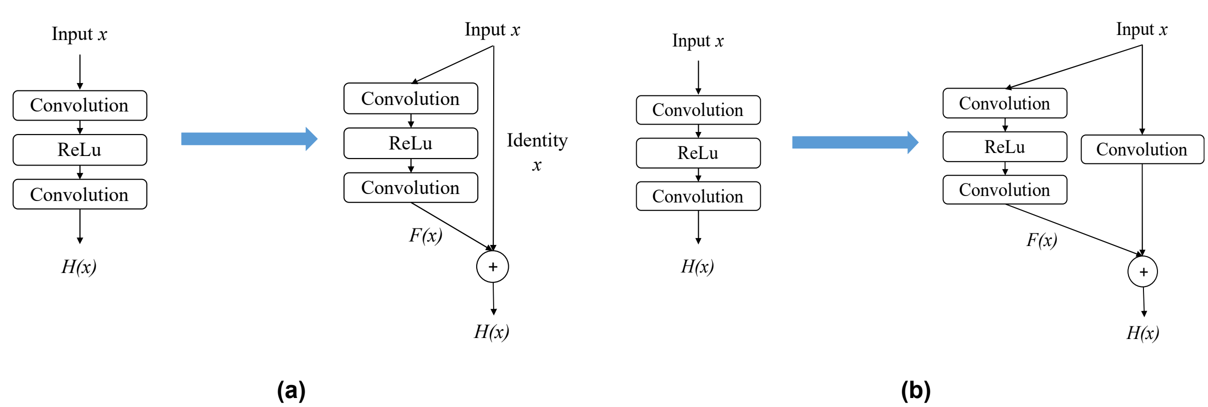


Figure 3.1 **Two kinds of shortcut connection in ResNet: (a) identity mapping; and (b) convolutional shortcut.**

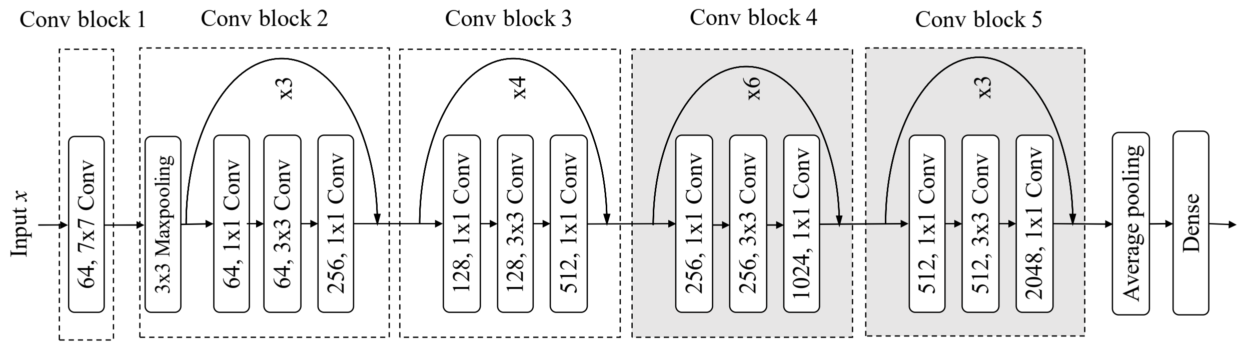


Figure 3.2 Illustration of ResNet-50 with shortcut connection.

4 Benchmark Experiments and Discussions

With the well-defined detection tasks and corresponding well-established datasets, benchmark experiments were conducted; the results described below. The purpose of these experiments was to provide a reference performance for some well-known deep CNN models and to establish a benchmark for future studies. In addition, the detection abilities of these DL models on the current scale of the ϕ-Net dataset with different training strategies was explored.

4.1 BENCHMARK EXPERIMENTS

One of the primary goals of the benchmark experiments was to compute the accuracy of each task. In addition, through conducting multiple comparison experiments, the influence of the following was explored: (1) different DL models, i.e., VGGNet and ResNet; (2) affine data augmentation (ADA); and (3) transfer learning (TL). Some detection tasks only contained very few images and should be considered as small datasets: e.g., the Task 5 collapse mode only contains 1372 images total. Thus, part of this research was to apply DL approaches in cases with a data deficit where a regular DL training approach may not be effective.

Through applying affine transformation towards a raw image, e.g., linear translation, rotation, zoom in and out, etc., the raw image can be transformed into new one with different possible angles, scales, etc. This operation, which aims to increase the amount of data, is shown to be efficient in many DL applications [Gao et al. 2018; Gao and Mosalam 2018]. Similarly, the advantage of TL is to help improve the prediction function in learning the target task by using the knowledge from the source domain [Pan and Yang 2010]. It also helps to mitigate the great dependency of ML/DL on the available dataset size and maximizes the use of existing data. Gao et al. [2018] and Gao and Mosalam [2018] conducted a series of experiments on detecting structural component type, damage severity, and type at the object level based on a small dataset with only 2000 images and where ADA and TL were applied. These two methods effectively alleviated the negative influence (namely, overfitting) of labeled data deficiency and achieved promising detection accuracy.

To establish a reference performance for some well-known deep CNN models to achieve Objective (1), we selected VGG-16, VGG-19, and ResNet-50 models, which were introduced in

Chapter 3. For Objectives 2 and 3 regarding ADA and TL, four different cases were compared for each model and each task; these cases are listed as follows:

1. Train model from scratch without ADA, denoted as baseline (BL)
2. Train model from scratch with ADA (BL-ADA)
3. Fine-tune the pre-trained model from ImageNet without ADA (TL)
4. Fine-tune the pre-trained model from ImageNet with ADA (TL-ADA)

For the comparison, several experimental settings for each task were fixed. The stochastic gradient descent (SGD) was applied along with a piecewise decayed learning rate for all cases, where the initial learning rate was 0.01; the entire learning schedule is shown in Figure 4.1. The learning schedule was adopted from He et al. [2016], which stipulates that the learning rate decreases by a scale (usually 10) when test accuracy does not increase,. In constrast to He et al. [2016], who made adjustments based on the observation of accuracy history, this study set empirically several timing epochs and automatically decreased the learning rate. The intuition behind this method is that in the early stage of training, a larger learning rate helps to avoid the model becoming trapped in local minima; the model is expected to “explore” somewhat. After training for the early epochs, the learning rate decreases such that the smaller learning rate stabilizes the model and avoids large oscillation.

In Cases (3) and (4) using TL, model parameters in conv layers were loaded from ImageNet pre-trained models, and parameters in the following adaptation layers (the output was consistent with the number of classes for different tasks) were initialized randomly. In cases (2) and (4) using ADA, affine transformations were performed from random combinations of six cases: (a) horizontal translation within 10% of total width; (b) vertical translation within 10% of total height; (c) rotation within 5^{o}; (d) zoom in less than 120% of the original size; (e) zoom out less than 80% of the original size; and (f) horizontal flip.

The batch size was taken as 64, and the total number of training epochs was selected as 50 where test accuracy was computed and reported at the end of each epoch. All cases were performed only once, except for some cases where the training loss did not change in the first 10 epochs. Future experiments are in the planning stage. The implementation of the benchmark experiments was conducted on TensorFlow and Keras platforms, and performed on CyberpowerPC with single GPU (CPU: Intel Core i7-8700K@3.7GHz 6 Core, RAM: 32 GB and GPU: Nvidia Geforce RTX 2080Ti). For convenience of notation, VGG-16 and VGG-19 are denoted as simply VGGNets.

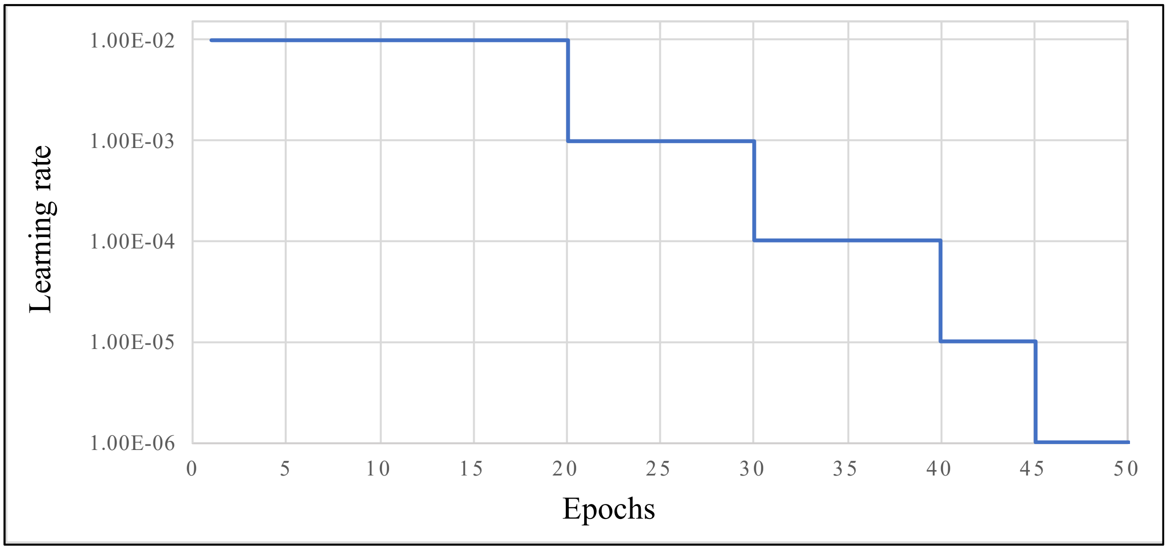


Figure 4.1 Learning schedule.

4.1.1 Task 1: Scene Level

Results for test accuracy and training history of Task 1 (scene level) are shown in Table 4.1 and Figure 4.2, respectively. Task 1 is a three-class classification, where the random guess is 33.33%. From Table 4.1, BL achieved promising results (~87%), which is significantly better than a random guess. There is not much difference between the different DL models, but ResNet-50 performed slightly better than the other two, achieving the best test accuracy (93.39%). Because of the sharing a similar architecture, the performances of VGG-16 and VGG-19 are also similar. The BL-ADA performance was worse than the solely BL's performance. Similarly the TL-ADA performance was worse than the solely TL's performance. In general, ADA did not contribute significantly to improving the results, and the results were even worse when training from scratch. In contrast, the TL worked well and clearly improved the performance by 6% for the best model. Figures 4.2(b), (d), and (f) present a clear view for the whole training history whereby it is obvious that models using TL outperformed those that did not use TL. Moreover, training history curves of models using TL are smoother and more stable than those that did not use TL, especially for the ResNet-50 model. The results clearly indicate that the TL provides a better initialization and more stable training environment than training from scratch.

Table 4.1 Test accuracy (%) for each case in scene level.

Model	BL	BL-ADA	TL	TL-ADA
VGG-16	87.29	85.39	91.93	92.49
VGG-19	87.35	83.98	92.09	92.09
ResNet-50	87.69	88.02	93.39	92.99

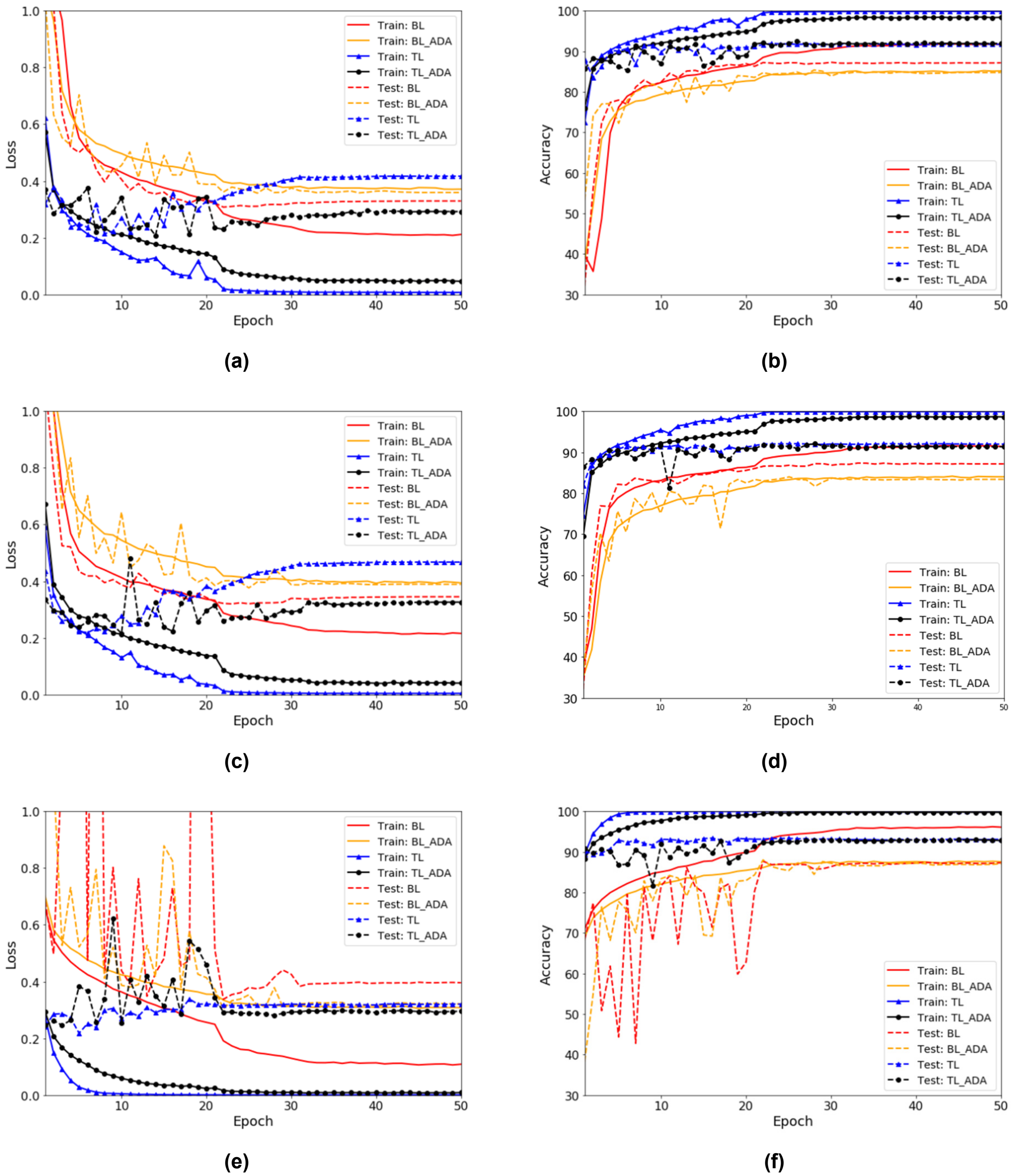


Figure 4.2 **Training history of DL models for 50 epochs in scene level comparing the four cases: (a) loss history of VGG-16; (b) accuracy history of VGG-16; (c) loss history of VGG-19; (d) accuracy history of VGG-19; (e) loss history of ResNet-50; and (f) accuracy history of ResNet-50.**

4.1.2 Task 2: Damage State

Results for test accuracy and training history of Task 2 (damage state) are shown in Table 4.2 and Figure 4.3, respectively. Task 2 is a binary classification where the random guess is 50%. As another basic and common detection task, the best model achieved 88.94%, which is approximately close to human performance. The best result was achieved by ResNet-50 using both TL and ADA. Note that in this experiment under the TL setting, ADA improved the performance slightly, in constrast to that observed in Task 1. A more interesting observation here is that compared to ResNet-50, VGGNets training from scratch did not produce stable and acceptable results (~63% and 61%). Test accuracy was just barely higher than a random guess: such results are not acceptable for sensitive damage detection. Thus, in Task 2, training from scratch only is far from acceptable. In contrast, the best performances of each model that adopted ADA and TL are close, with ResNet-50 using TL-ADA slightly better. Such observations indicate that all models have good potential if trained using appropriate strategies.

Figure 4.3 provides more information for the effectiveness of TL, which cannot be reflected in Table 4.2. All three models indicate similar trends in training: that the accuracy of training the models from scratch remains trapped at an early stage even with a large learning rate; however, the training accuracy increases slightly along with the decrease of the learning rate, and the whole history curve becomes flat. In contrast, the accuracy of models using TL has a good initial value and keeps increasing, and converges when the learning rate decreases, which helps stabilize the performance.

Table 4.2 Test accuracy (%) for each case in damage state.

Model	BL	BL-ADA	TL	TL-ADA
VGG-16	63.97	60.72	87.29	87.60
VGG-19	62.67	60.72	87.29	87.36
ResNet-50	76.71	75.27	87.26	88.94

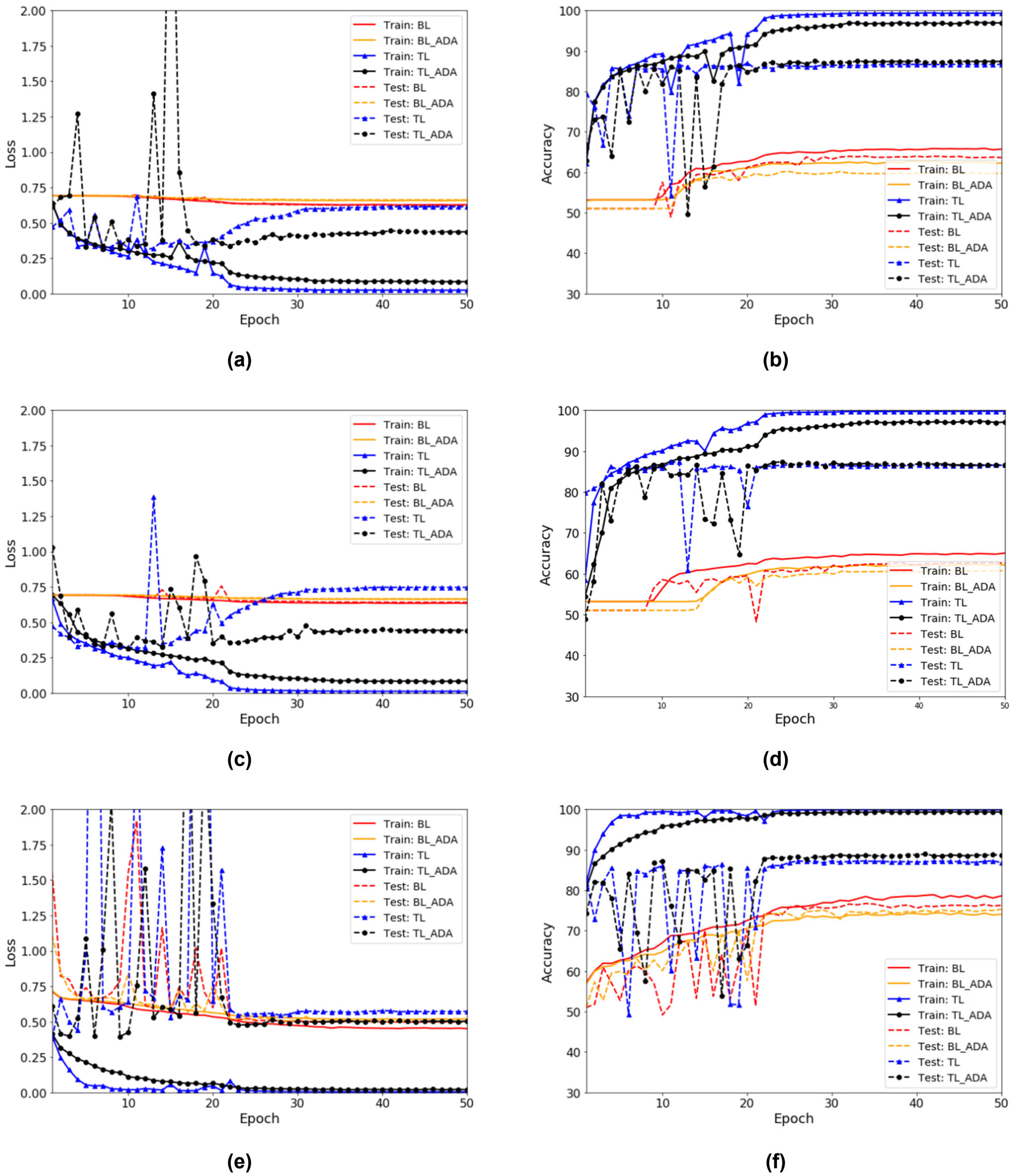


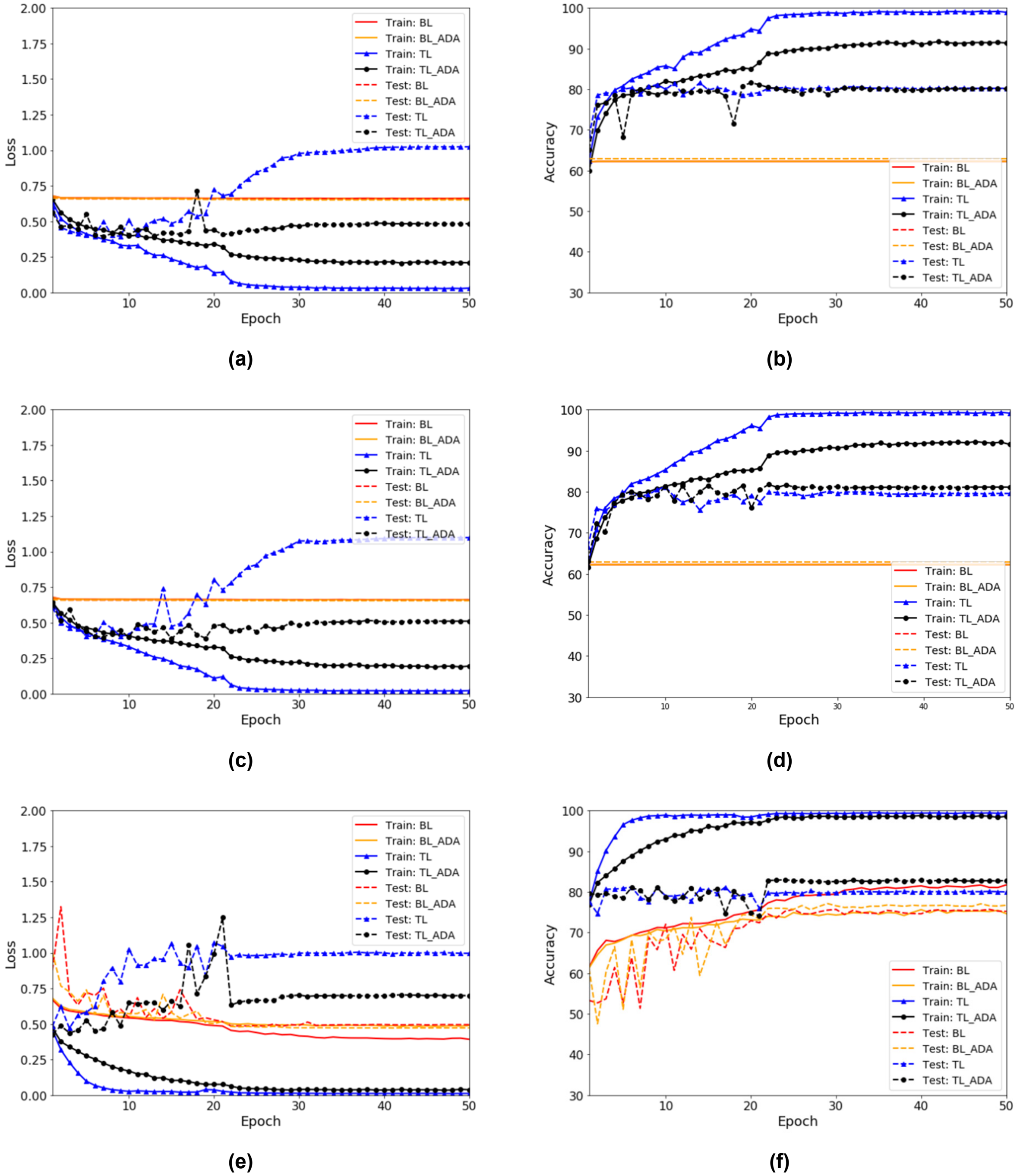
Figure 4.3 **Training history of DL models for 50 epochs in damage state comparing the four cases: (a) loss history of VGG-16; (b) accuracy history of VGG-16; (c) loss history of VGG-19; (d) accuracy history of VGG-19; (e) loss history of ResNet-50; and (f) accuracy history of ResNet-50.**

4.1.3 Task 3: Spalling Condition

Results for test accuracy and training history of Task 3 (spalling condition) are shown in Table 4.3 and Figure 4.4, respectively. Even though this task is also a binary classification like Task 2, its smaller amount of data (for a total 7735, which is only 75% of that in Task 2) increases the difficulty somewhat. This explains the performance deterioration in the accuracy, where the best accuracy is 82.97% as obtained by ResNet-50 using both TL and ADA. In the view of CV, the vision patterns of spalling are complex and irregular, which makes it difficult to identify, especially within the damaged component surface. Similar conclusions were obtained for Task 2, but the issues were more severe during the training of the model without using TL. As shown in Figures 4.4(a) to (d), loss and accuracy for training and testing of VGGNets became trapped. The models did not learn and, therefore, could not update anything. Unlike VGGNets, ResNet-50 without TL is still trainable; however, even after minor improvement, the model quickly converged to bad local minima. Similar to observations seen in Task 2, using TL improved the model performance significantly; for ADA the improvement was not obvious.

Table 4.3 Test accuracy (%) for each case in spalling condition.

Model	BL	BL-ADA	TL	TL-ADA
VGG-16	62.96	62.96	81.60	81.48
VGG-19	62.96	62.96	81.18	81.78
ResNet-50	75.69	77.12	81.72	82.97



(a) **(b)**

(c) **(d)**

(e) **(f)**

Figure 4.4 **Training history of DL models for 50 epochs in spalling condition comparing the four cases: (a) loss history of VGG-16; (b) accuracy history of VGG-16; (c) loss history of VGG-19; (d) accuracy history of VGG-19; (e) loss history of ResNet-50; and (f) accuracy history of ResNet-50.**

4.1.4 Task 4: Material Type

Results for test accuracy and training history of Task 4 (material type) are shown in Table 4.4 and Figure 4.5, respectively. Unlike Tasks 2 and 3, the models have an easier time identifying material type. The scope for material type observed from the surface is narrow, being defined into only two categories: steel and others. Because there are sharp differences between these two vision patterns, i.e., texture, color, shape, etc., such distinct categories explain the high accuracy obtained for all cases, where the best result is achieved by ResNet-50 using both TL and ADA. In this case, ADA still does not contribute with any degree of significance toward improving the performance. The results are even worse when training from scratch, which is similar to those results observed in Task 1. Another observation similar to Task 1 is that BL achieved high accuracy, whereas the TL only achieved limited improvement: ~3%. In general, the best performance at 98.52% is close to or even better than human judgment under some conditions. In the early training stages of ResNet, the models trained from scratch oscillated significantly no matter whether ADA was adopted or not. This is also similar to that observed in Task 1. Overall, the DL approach was considered effective in identifying between steel and non-steel materials. It is expected to extend this benchmark by including finer classification categories, i.e., masonry, wood, concrete, etc.

Table 4.4 Test accuracy (%) for each case in material type.

Model	BL	BL-ADA	TL	TL-ADA
VGG-16	93.92	90.76	97.04	96.99
VGG-19	94.43	90.30	96.53	96.88
ResNet-50	95.05	93.11	98.31	98.52

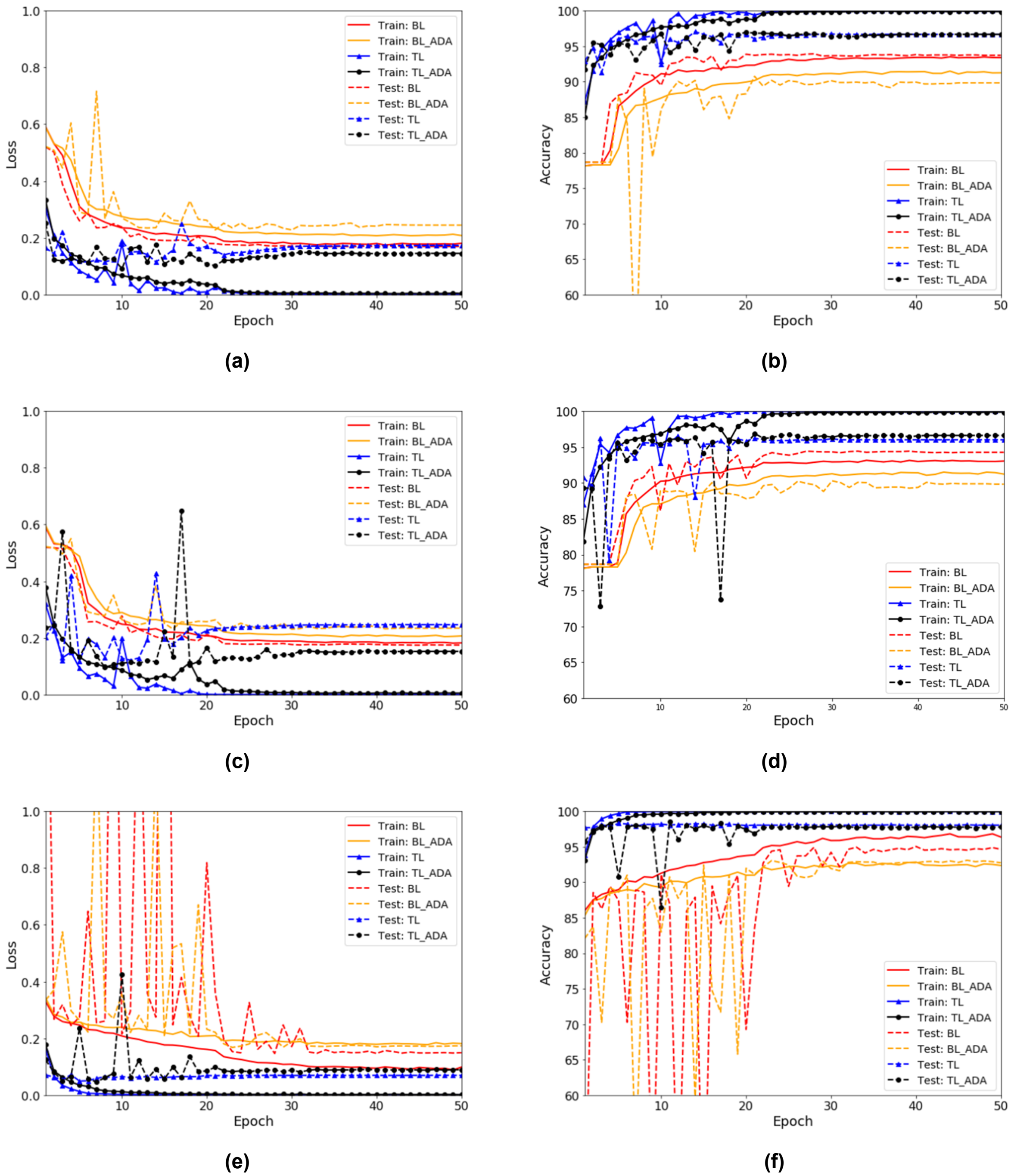


Figure 4.5 Training history of DL models for 50 epochs in material type comparing the four cases: (a) loss history of VGG-16; (b) accuracy history of VGG-16; (c) loss history of VGG-19; (d) accuracy history of VGG-19; (e) loss history of ResNet-50; and (f) accuracy history of ResNet-50.

4.1.5 Task 5: Collapse Mode

Results for test accuracy and training history of Task 5 (collapse mode) are shown in Table 4.5 and Figure 4.6, respectively. Task 5 is a three-class classification, where the random guess is 33.33%. As mentioned above, this task is considered difficult due to very limited data (a total of 1372), which is defined as an application of low-data regime. Thus, it was expected to obtain fewer promising results in terms of its baseline performance. Similar to issues noted in Task 2 (spalling condition), VGGNets were untrainable when training from scratch, and parameters did not change during the entire training history. The results for ResNet-50 were slightly better, but the final accuracy was still not acceptable. For models training from scratch, ADA did not did not perform satisfactorily. In contrast, using TL improved the model's performance significantly. As shown in Table 4.5, for the best model of VGG-16, VGG-19, and ResNet-50, accuracies increased by ~30%, 25%, and 13%, respectively, where ADA contributed to the performance for VGG-19 but did not improve the performance of the other two models. Similar to previous tasks, ResNet-50 using TL obtained the best test accuracy: 78.08%.

Table 4.5 Test accuracy (%) for each case in collapse mode.

Model	BL	BL-ADA	TL	TL-ADA
VGG-16	45.89	45.89	75.34	71.92
VGG-19	45.89	45.89	70.55	72.60
ResNet-50	65.07	61.64	78.08	74.66

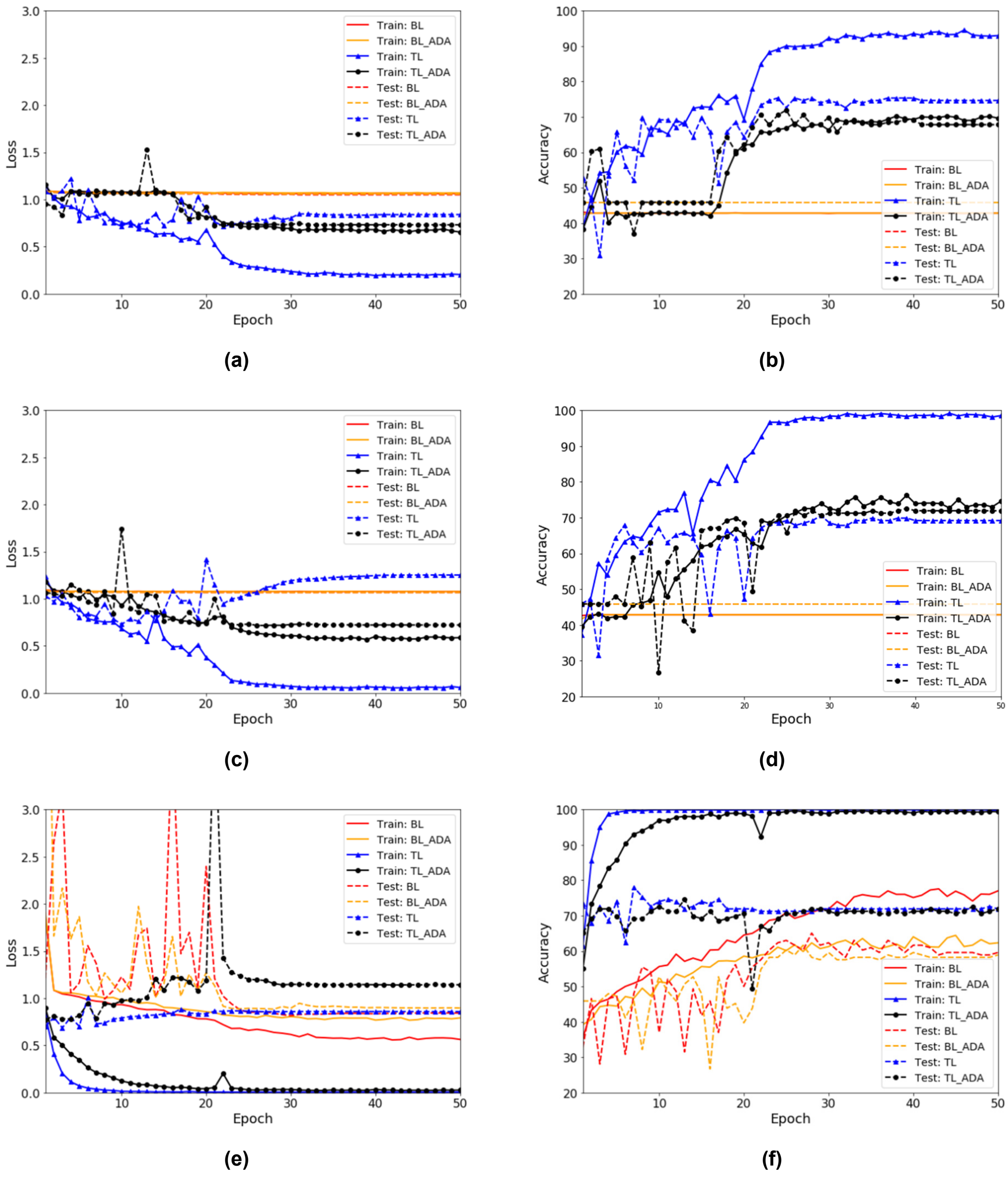


Figure 4.6 **Training history of DL models for 50 epochs in collapse mode comparing the four cases: (a) loss history of VGG-16; (b) accuracy history of VGG-16; (c) loss history of VGG-19; (d) accuracy history of VGG-19; (e) loss history of ResNet-50; and (f) accuracy history of ResNet-50.**

4.1.6 Task 6: Component Type

Results for test accuracy and training history of Task 6 (component type) are shown in Table 4.6 and Figure 4.7, respectively. Task 6 is a four-class classification, where the random guess is only 25%. The number of categories compared to previous tasks was expanded to four, which increased the difficulty for classification. Moreover, according to the hierarchy of the ϕ-Net dataset shown in Figure 2.15, the component type is a leaf task of Task 1 scene level; therefore, only images belonging to object level will flow through this task. This leads to a smaller image collection (a total of 5334) for Task 6 compared to Task 1 scene level (a total of 27,306) or Task 2 damage state (a total of 13,271). These two issues underscore the major difficulty facing component-type identification.

In general, the BL models obtained mediocre results but were better than the results obtained in Task 5 (collapse mode); once again, ADA did not perform satisfactorily. Similar to that observed in previous tasks, also TL performed, it did not provide significant improvement. One interesting and valuable observation is that VGG-19 achieved the best accuracy (77.03%) as opposed to ResNet-50. Such observations indicate the necessity of taking DL models as hyper-parameters where different models may be suitable for different application scenarios.

Table 4.6 Test accuracy (%) for each case in component type.

Model	BL	BL-ADA	TL	TL-ADA
VGG-16	64.77	64.77	75.13	76.86
VGG-19	65.11	65.11	77.03	76.86
ResNet-50	69.78	68.22	75.82	76.34

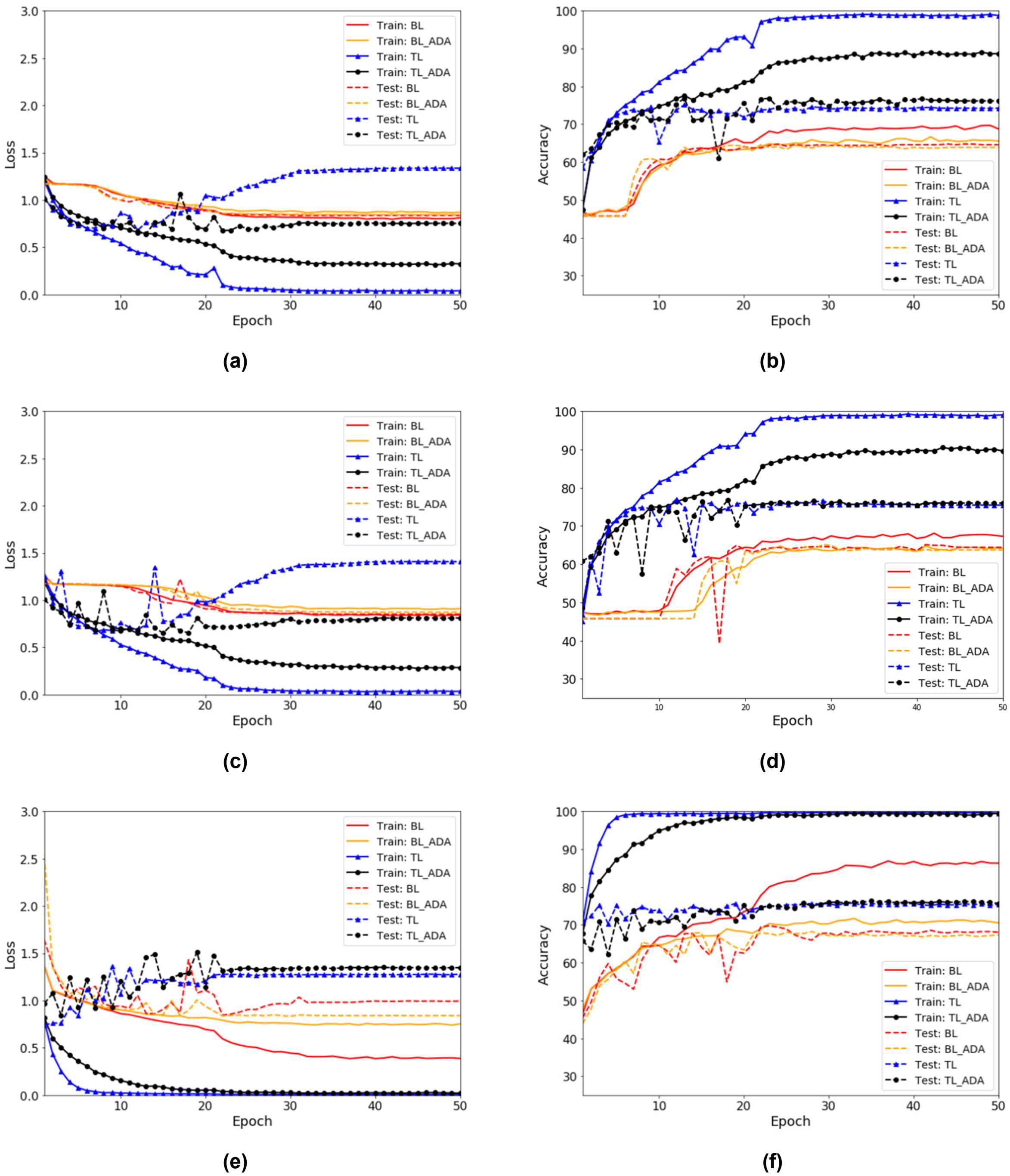


Figure 4.7 **Training history of DL models for 50 epochs in component type comparing the four cases: (a) loss history of VGG-16; (b) accuracy history of VGG-16; (c) loss history of VGG-19; (d) accuracy history of VGG-19; (e) loss history of ResNet-50; and (f) accuracy history of ResNet-50.**

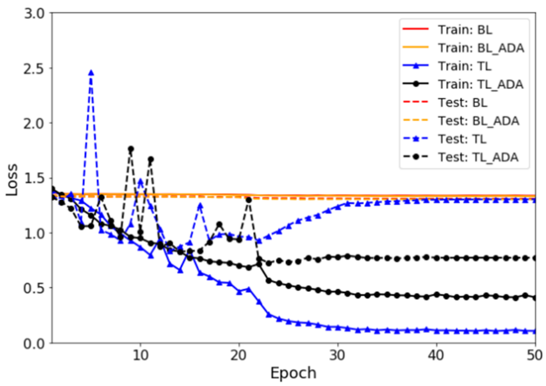
4.1.7 Task 7: Damage Level

The results for test accuracy and training history of Task 7 (damage level) are shown in Table 4.7 and Figure 4.8, respectively. Task 7 is a four-class classification. Compared to Task 6 (the component-type task), it is more abstract and has semantic meaning, which significantly increases the difficulty. Here, the so-called abstract and semantic meaning mean that the categories cannot be easily identified by low-level vision features such as edge, color, etc., as is the case in component-type identification. Classification needs to be elevated to high-dimensional spaces to identify some key features linked to various damage patterns semantically. For example, the orientation and boundary of the component can lead to the specific component type. As pointed out in Section 2.2.7, definition of the damage level relates more to the severity and development of damage recognized by humans, which is often subjective and not easily quantifiable. Thus, it requires the DL model to find features and patterns in abstract and high-dimensional spaces related to semantic meaning of "minor," "moderate," or "heavy," which is more difficult compared to previous tasks. As is the case in Task 6, the total amount of images available is only 4636, which is not considered adequate.

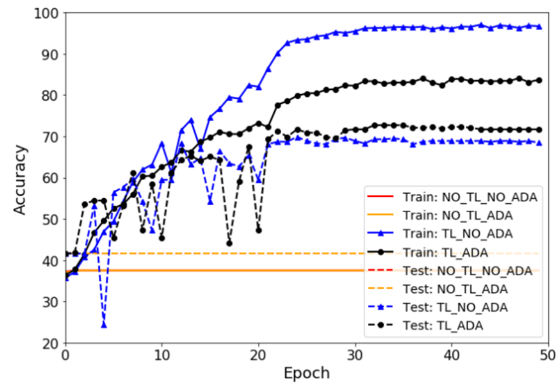
From Table 4.7, the performances of BL are consistent with the concern that VGGNets are untrainable no matter whether ADA is adopted or not. Test accuracies were unacceptable—only ~41%. The observation in baselines was similar to that found in Task 6, but there were some differences; the ResNet-50 trained from scratch was capable of obtaining a much better BL performance compared to VGGNets, and using ADA further improved the performance by ~4%. Due to the relatively poor performance in BL, models with TL were considered significantly improved from those trained from scratch, with improvements of ~30%, 23%, and 18% for VGG-16, VGG-19, and ResNet-50, respectively. ResNet-50 still performed the best with 74.5% test accuracy, which is acceptable for this task using the current dataset.

Table 4.7 Test accuracy (%) for each case in damage level.

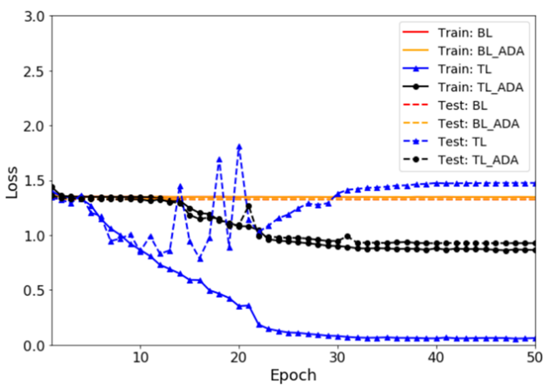
Model	BL	BL-ADA	TL	TL-ADA
VGG-16	41.57	41.57	69.88	72.69
VGG-19	41.57	41.57	70.68	64.66
ResNet-50	54.62	58.43	74.50	72.69



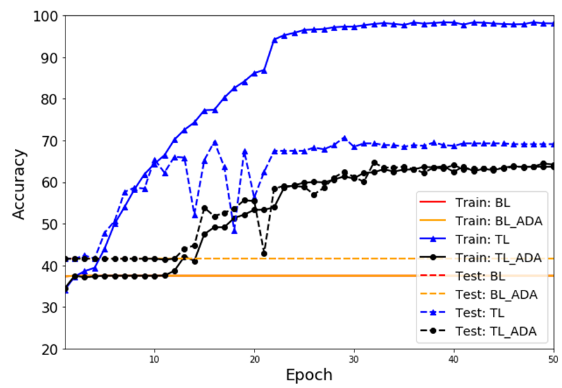
(a)



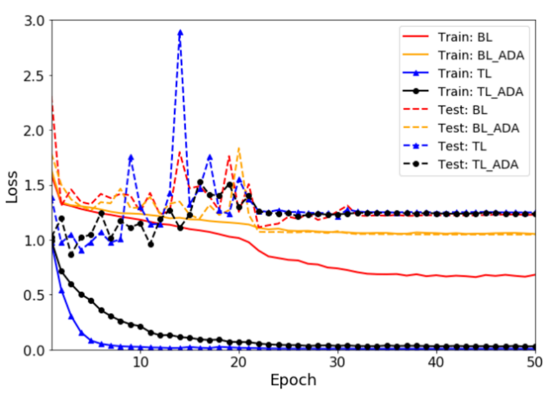
(b)



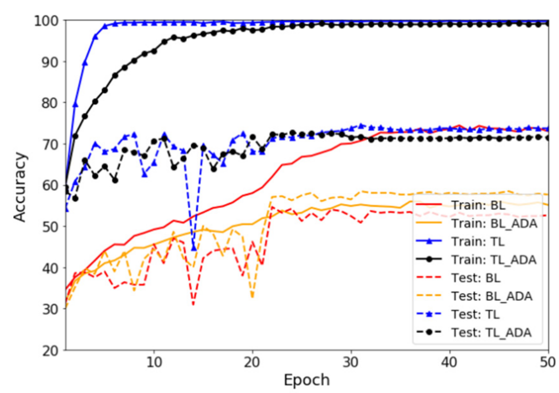
(c)



(d)



(e)



(f)

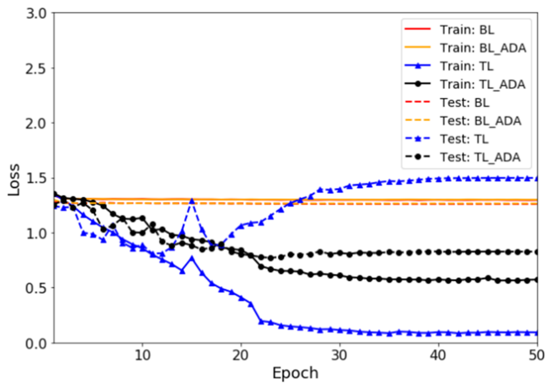
Figure 4.8 Training history of DL models for 50 epochs in damage level comparing the four cases: (a) loss history of VGG-16; (b) accuracy history of VGG-16; (c) loss history of VGG-19; (d) accuracy history of VGG-19; (e) loss history of ResNet-50; and (f) accuracy history of ResNet-50.

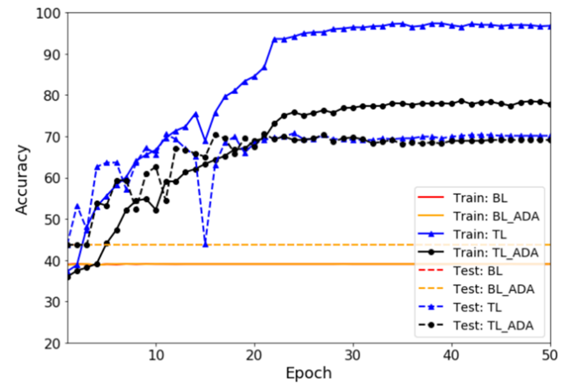
4.1.8 Task 8: Damage Type

Results for test accuracy and training history of Task 8 (damage type) are shown in Table 4.8 and Figure 4.9, respectively. As with Task 7, Task 8 is also thought of as abstract with various vision patterns corresponding to specific types mentioned in Section 2.2.8. From the perspective of the relative lack of data, a total of 4585 images is not considered representative enough for training purposes. Similar issues occurred in BL models for Task 7, i.e., VGGNets proved to be untrainable, and ResNet-50 achieved relatively low accuracy. In contrast, using TL achieved significant improvement in this task, and the best accuracies of VGG-16, VGG-19, and ResNet-50 increased by ~27%, 29%, and 15%, respectively. Among the above three models, VGG-19 performed the best (as was also found for Task 6). Note: the performance of the best ResNet-50 model was even worse than the performance for the VGG-16 model. This underscores the importance of selecting the appropriate model for future applications instead of relying on only one single model.

Table 4.8 Test accuracy (%) for each case in damage type.

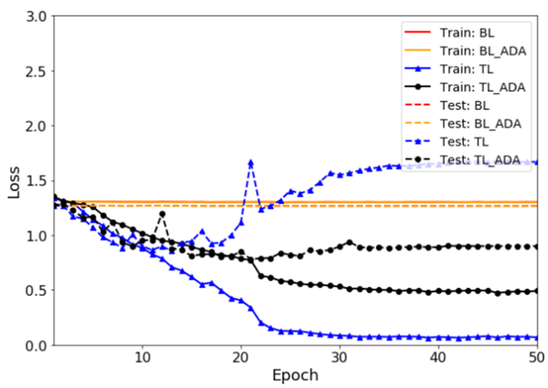
Model	BL	BL-ADA	TL	TL-ADA
VGG-16	43.70	43.70	70.73	70.53
VGG-19	43.70	43.70	71.14	72.36
ResNet-50	54.27	56.10	68.90	68.09

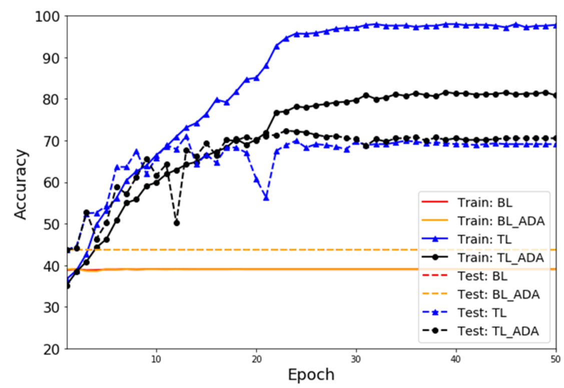




(a

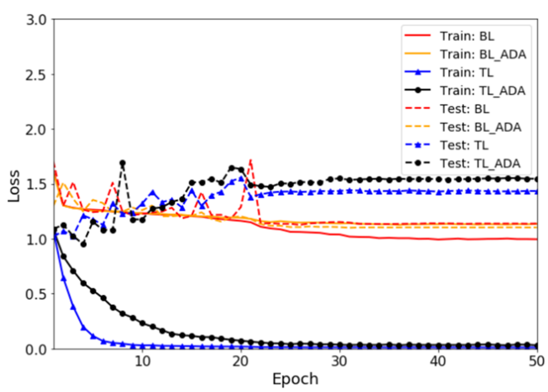
(b)

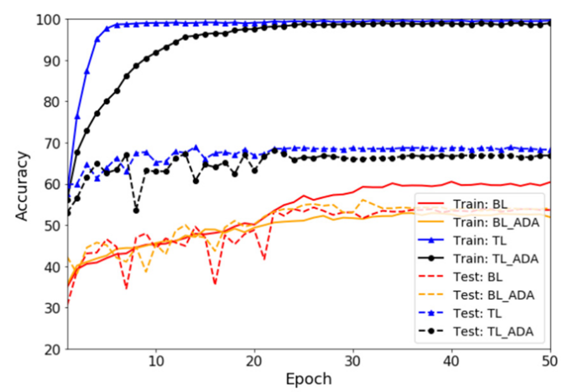




(c)

(d)





(e)

(f)

Figure 4.9 Training history of DL models for 50 epochs in damage type comparing the four cases: (a) loss history of VGG-16; (b) accuracy history of VGG-16; (c) loss history of VGG-19; (d) accuracy history of VGG-19; (e) loss history of ResNet-50; and (f) accuracy history of ResNet-50.

4.2 RESULTS SUMMARY

Sections 4.1.1 to 4.1.8 presented the reference results including history plots for all cases in the identified eight tasks; Table 4.9 lists the best reference values. Since the ultimate goal of this study is to adopt DL for automated detection, accuracy should be within a reasonable range. For the easier tasks (Tasks 1 to 4), the test accuracies appear very promising, especially for classification between steel and non-steel materials. For the harder tasks (Tasks 5 to 8), although the best results are still reasonable and acceptable, they are expected to be improved with more sophisticated DL models, training approaches, and tuned hyper-parameters.

In a comparison of the DL models, ResNet-50 appeared to be more generalized because it achieved the highest test accuracy in most cases (except in Tasks 6 and 8). Thus, to achieve the state-of-the-art performance, it is suggested to perform model selection among multiple DL models. From the perspective of the training approach, TL worked well in all eight tasks, with significant improvement in many cases, and could be seen as the necessary training strategy for the current scale of the dataset. ImageNet pre-trained models already have some sense of features of basic and natural objects from the source domain. Thus, transferring source-domain knowledge to the target domain of structural images may provide a better understanding for the deep CNN models towards some general structural vision patterns, e.g., object boundary, material texture, etc. Moreover, fine-tuning from ImageNet pre-trained models will obtain better parameter initialization than directly training models from scratch where parameters are initialized randomly. These findings were more evident in the harder tasks with limited amount of data or abstract semantic meanings: e.g., TL achieved 10–30% enhancements over the BL case. Compared to TL, ADA did not significantly improve the accuracy of the test and in some cases performed poorly compared to other approaches. This can be attributed to two major issues: (1) randomness and uncertainty due to only having one run for each case; and (2) different ADA settings, e.g., range of rotation angle, translation, etc., which may lead to different performance.

More information about a model's performance can be extracted from the history plots. Thus, the best results for VGG-16, VGG-19, and ResNet-50 were plotted in one figure for each task; see Figures 4.10(a) to (h). In all benchmark tasks for all models, training accuracies were high and close to 100%; however, gaps between training and test accuracies are evident where the latter varied from one task to another. This is an indicator of overfitting and fully training, i.e., the models contained enough parameters and complexity to learn the task well. In ML/DL, usually applying regularizations, i.e., dropout (applied herein but the dropout rate can be adjusted), batch normalization [Ioffe and Szegedy 2015], etc., may help to alleviate overfitting. This issue is worthy for future studies. From the perspective of training stability, DL models are less stable in the early training stages when the learning rate is large. During this period, the parameters are explored and updated more with gradient descent, and the history curves are more oscillating. With a scheduled decayed learning rate—especially after 20 epochs—the training curves become flat and smooth, indicating that the model performance is now stable.

Table 4.9 Summary of best reference performance in each task.

Task	Best test accuracy (%)	Model	Training approach	Over BL (%)	Over BL-ADA (%)	Over TL (%)	Over TL-ADA (%)
Scene level	93.4	ResNet-50	TL	+5.7	+5.4	-	+0.4
Damage state	88.9	ResNet-50	TL-ADA	+12.2	+13.7	+1.7	-
Spalling cnditions	83.0	ResNet-50	TL-ADA	+7.3	+5.9	+1.8	-
Material type	98.5	ResNet-50	TL-ADA	+3.5	+5.4	+0.2	-
Collapse mode	78.1	ResNet-50	TL	+13.0	+16.3	-	+3.4
Component type	77.0	VGG-19	TL	+11.9	+11.9	-	+0.2
Damage level	74.5	ResNet-50	TL	+19.9	+16.1	-	+1.8
Damage type	72.4	VGG-19	TL-ADA	+28.7	+28.7	+1.2	-

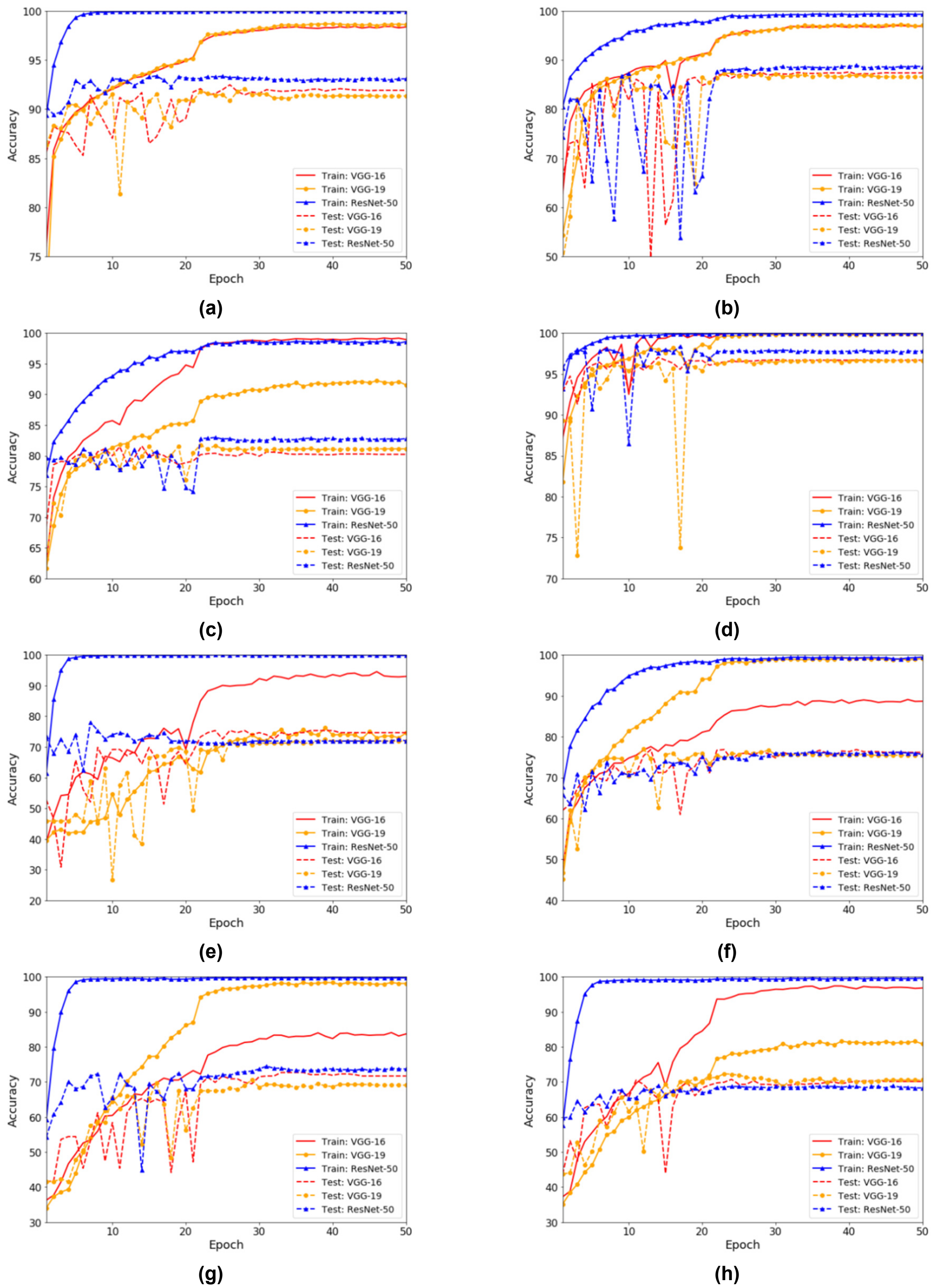


Figure 4.10 **Accuracy history of best models for 50 epochs in all tasks: (a) scene level; (b) damage state; (c) spalling condition; (d) material type; (f) component types; (g) damage level; and (h) damage type.**

Computational efficiency is another important measurement in DL applications. Tables 4.10 to 4.17 list average computational times per training epoch for each model under different training approaches, thus providing a reference for future studies. Without ADA, ResNet-50 costs the least computational time, VGG-16 costs the second least, and VGG-19 is the most expensive as it contains more parameters than the other two; there are roughly 21 million, 26 million, and 25 million parameters in VGG-16, VGG-19, and ResNet-50, respectively. Even though ResNet-50 is deeper than VGGNets, it uses a smaller size of filter (1×1 rather than the 3×3 applied in VGG(Nets), and shortcut connections that make gradient flow faster accelerate the training speed. With ADA, all models have increasing time costs and similar training time per epoch due to the extra cost of performing affine transformations on raw images. As for TL, it does not affect the average training time, because it neither introduces new parameters nor performs image-processing operations (per ADA). Thus, using TL proved to be a more efficient and economical method in the experiments reported herein.

In summary, the reference results and history plots for all cases and tasks were determined, and then compared with the performances of the best DL models, the influences of ADA and TL were analyzed, and the computational time cost for each task was assessed. For benchmarking purposes, a comparison with the same settings was conducted for all tasks. Note: although the results shown herein are not the state-of-the-art, they provide a reference for future DL applications. In order to obtain better results, tuning hyper-parameters (e.g., learning rate, dropout rate, and the number of neurons in the *fc*-layers) and developing new models and training approaches are proposed for future studies. It is recommended to make full use of the ϕ-Net dataset, using conclusions from this report to enlarge the dataset, quantify finer categories, using images for new vision tasks, etc.

Table 4.9 Average computational time per training epoch in scene level (seconds).

Model	BL	BL-ADA	TL	TL-ADA
VGG-16	145	178	146	187
VGG-19	169	185	169	188
ResNet-50	131	183	123	190

Table 4.10 Average computational time per training epoch in damage state (seconds).

Model	BL	BL-ADA	TL	TL-ADA
VGG-16	71	79	71	81
VGG-19	82	83	82	83
ResNet-50	64	83	63	83

Table 4.11 Average computational time per training epoch in spalling condition (seconds).

Model	BL	BL-ADA	TL	TL-ADA
VGG-16	42	46	42	46
VGG-19	48	49	48	49
ResNet-50	37	47	37	47

Table 4.12 Average computational time per training epoch in material type (seconds).

Model	BL	BL-ADA	TL	TL-ADA
VGG-16	52	67	52	67
VGG-19	59	69	59	69
ResNet-50	46	68	46	68

Table 4.13 Average computational time per training epoch in collapse mode (seconds).

Model	BL	BL-ADA	TL	TL-ADA
VGG-16	8	8	8	8
VGG-19	9	9	9	9
ResNet-50	7	8	7	8

Table 4.14 Average computational time per training epoch in component type (seconds).

Model	BL	BL-ADA	TL	TL-ADA
VGG-16	28	30	28	30
VGG-19	33	33	33	33
ResNet-50	26	30	26	31

Table 4.15 Average computational time per training epoch in damage level (seconds).

Model	BL	BL-ADA	TL	TL-ADA
VGG-16	25	26	25	26
VGG-19	29	29	29	29
ResNet-50	22	27	22	27

Table 4.16 Average computational time per training epoch in damage type (seconds).

Model	BL	BL-ADA	TL	TL-ADA
VGG-16	25	25	25	25
VGG-19	28	29	28	29
ResNet-50	22	27	22	27

5 ϕ-Net Challenge 2018

Chapters 1 to 4 described the background, definitions, development procedure, deep-learning (DL) methods, and benchmark results of the ϕ-Net Challenge. Before finalizing and open sourcing the benchmark datasets and results, during the Fall 2018, PEER released a partial dataset of the ϕ-Net dataset and held a worldwide challenge to call for interested researchers to participate. The results of the challenge provided a preliminary reference for the establishment of the benchmark dataset. The challenge also served to alert the engineering community and encourage researchers in civil engineering to interact with other disciplines towards transdisciplinary research[13] activities, and to learn and apply the state-of-the-art approaches in computer vision (CV) for application to civil engineering. More details of the ϕ-Net Challenge are introduced in the following sections.

5.1 MOTIVATION OF THE CHALLENGE

Inspired by several famous CV competitions in the computer science area, such as the ILSVRC[14], PASCAL VOC[15], and COCO[16] challenges, during the Fall of 2018 PEER organized the first image-based structural damage identification competition: the PEER Hub ImageNet (ϕ-Net) Challenge. Figure 5.1 shows the advertising flyer for the challenge.

In this competition, PEER provided eight image datasets that corresponded to the tasks introduced in Section 2.2. Each team received training and test sets for these tasks. Note: since the ϕ-Net Challenge acted as a pre-event for the open sourcing of the ϕ-Net dataset, the datasets released in the challenge were different from the finalized benchmark datasets summarized in Section 2.4. Based on these well-labeled images, each team was expected to develop and use algorithms to train their recognition models and then predict labels for the unlabeled test datasets. The labels predicted for test datasets would then be compared against ground truth. In machine

---

[13] Transdisciplinary research are research efforts conducted by investigators from different disciplines working jointly to create new conceptual, theoretical, methodological, and translational innovations that integrate and move beyond discipline-specific approaches to address a common problem.

[14] http://image-net.org/challenges/LSVRC/

[15] http://host.robots.ox.ac.uk/pascal/VOC/

[16] http://cocodataset.org/#home

learning (ML), ground truth is factual data that have been observed, labeled, or measured, and can be analyzed objectively. Teams with the highest accuracy would be declared the winners of the competition. The challenge started on 23 August 2019, at 00:00 am UTC and was closed on 25 November 2019 at 11:59 pm UTC. For more details about the ϕ-Net Challenge refer to the website: *https://apps.peer.berkeley.edu/phichallenge/*.

The ϕ-Net dataset and challenge effort is part of PEER's strategic plan to equip the earthquake engineering community with the tools of the digital revolution era, including ML, DL, AI, high-performance computing (HPC), and emerging technology. The main objective of the challenge was to engage fully the earthquake engineering community (or more broadly the natural hazards communities) in all stages of the competition, including preparation for the computation, execution of the computation, and processing of results. In the competition preparation stage, members of the community contributed significantly by uploading images that could be used in the competition and by labeling these images through the developed tool mentioned in Section 2.3.3. PEER welcomes contributions of newly labeled images to further expand the current ϕ-Net dataset.

In summary, the goal of the ϕ-Net Challenge was to evaluate algorithms for structural image classification using large-scale datasets based on service conditions, past reconnaissance efforts, and laboratory experiments for condition assessments following extreme events. The state-of-the-art algorithms to be tested in the ϕ-Net Challenge are expected to enhance the accuracy and the generalization of vision-based approaches for SHM, and contribute to the construction of a large structural image dataset used to solve societal-scale problems of SHM and perform assessments of the built environment.

PEER

Φ Challenge 2018

PEER Hub ImageNet (PHI) Challenge

PEER has developed the first structural engineering dataset that incorporates machine-learning models of detecting and categorizing damage in images. The PEER Hub ImageNet (PHI) dataset tool will enhance the field and application of vision-based structural health monitoring for researchers and practitioners in natural hazards engineering.

CALL FOR PARTICIPATION

- Enter the PHI Challenge contest!
  Registration is now open at
  http://apps.peer.berkeley.edu/phichallenge/registration/

PHI CHALLENGE CONTEST

PEER is organizing the first image-based structural damage recognition competition to enhance the field of vision-based structural health monitoring. The contest will **open July 15 and close November 25, 2018.**

- **Train** your algorithm and model.

A dataset of labeled and annotated images will be provided so that contestants can train their machine-learning algorithms and models.

- **Test** and submit results.

A dataset of unlabeled images will be provided so that predications of labels and annotations can be generated. A few classifications of identification are the following: Component type, Damage check, Damage level, Damage type, and Material type. The contest submittal should include the labeled dataset as well as a brief report that summarizes assumptions, methodology, tools, and detailed computations.

- **Apply and register** to be eligible.

Researchers, practitioners, and students in the fields of structural and civil engineering, computer science, data science, and other related fields are encouraged to participate.

The registration period for individuals and teams will open July 15, 2018. Two contestant pools are as follows: (1) Computer Science / Data Science (CS/DATA) and (2) non-CS/DATA. Notification of eligibility in either of these two contestant pools will be issued upon receipt of application.

- **Winners**

Winners in each category for single detection tasks as well as for overall performance will be announced in December 2018.

WHO BENEFITS?

Algorithms developed in the PHI Challenge can be used to automatically detect damage from images taken after earthquakes or other natural hazards. Accurate and automated labeling encourages crowd-sourced data and allows engineers to focus more on the interpretation of the image data, which can increase the efficiency of a tagging process after a major natural hazard.

Researchers will have a large, accurately labeled structural image dataset that can be used as a basis for future studies. Practitioners can test their machine-learning and deep-learning algorithms and models on the PHI dataset.

Automated damage detection is an advanced tool for determining the extent of damage, and it will enhance the ability of the engineering community to respond in a timely manner to the aftermath of natural hazards, thereby serving the population at large.

MORE information

http://apps.peer.berkeley.edu/phichallenge

Headquarters at: University of California, Berkeley, 325 Davis Hall, Berkeley, CA 94720-1792
Phone: (510) 642-3437 ▪ Fax: (510) 642-1655 ▪ http://peer.berkeley.edu ▪ peer_center@berkeley.edu

Figure 5.1 Flyer for the ϕ-Net Challenge of Fall 2018.

5.2 PARTICIPATION OF THE ϕ-NET CHALLENGE

5.2.1 Team Setup

A team consisted of either an individual or a maximum of four team members. The team setting rules encouraged communication and collaboration between different disciplines. In the ϕ-Net Challenge of Fall 2018, PEER received a total of 68 team applications from different countries and regions from around the world. The geographic distribution of the registered teams is shown in Figure 5.2, where over half of the teams were from the United States. As expected, many other disciplines besides structural engineering were actively engaged in the challenge, including mechanical engineering, transportation engineering, environmental engineering, computer science, statistics, etc. The challange considered the background advantage of those participants from computer science and data science in ML/DL, and two different pools were created to address this issue: computer science/data science (CS/DATA) and engineering (Non-CS/DATA). Each team was assigned to the pool corresponding to the team members' educational fields as declared in the team application. If at least one team member was majoring or had a major in computer science/data science, that team was assigned to the pool of CS/DATA and vice versa.

Figure 5.2 Registered team distribution of the ϕ-Net Challenge of Fall 2018.

5.2.2 Access to the Challenge

The ϕ-Net Challenge followed the current trend of deploying data science competitions on Kaggle[17], which is an online community of data scientists and machine learners. Kaggle provides

[17] https://www.kaggle.com/

a flexible platform to host online challenges where the host can release the dataset and launch the challenge in one of multiple ways, including: public challenges, invited challenges, or private challenges. In the ϕ-Net Challenge of Fall 2018, eight challenge webpages were established via Kaggle as invited challenges, and only the teams who applied could receive the links to access the data and later submit their results; see Figure 5.3. The Kaggle webpages for the eight benchmark tasks mentioned above are as follows:

- Task 1 (Scene level): *https://www.kaggle.com/c/phi2018task1*
- Task 2 (Damage state): *https://www.kaggle.com/c/phi2018task2*
- Task 3 (Spalling condition): *https://www.kaggle.com/c/phi2018task3*
- Task 4 (Material type): *https://www.kaggle.com/c/phi2018task4*
- Task 5 (Collapse check): *https://www.kaggle.com/c/phi2018task5*
- Task 6 (Component type): *https://www.kaggle.com/c/phi2018task6*
- Task 7 (Damage level): *https://www.kaggle.com/c/phi2018task7*
- Task 8 (Damage type): *https://www.kaggle.com/c/phi2018task8*

To ensure good participation and submission of results, Task 1 (Scene level) and Task 2 (Damage state) were mandatory in the *Call for Participation*. Only after the participants submitted the results for Tasks 1 and 2 at least once did they have access to the remaining tasks. These first two tasks were judged as the most basic detection tasks in vision-based SHM according to the ϕ-Net framework; see Figure 2.1. After obtaining full access to all the tasks, the participating teams could choose and perform some or all of the tasks that were of interest to them. The winners from each of the two pools of participants were determined via individual performance in a single task and in overall performance of all tasks.

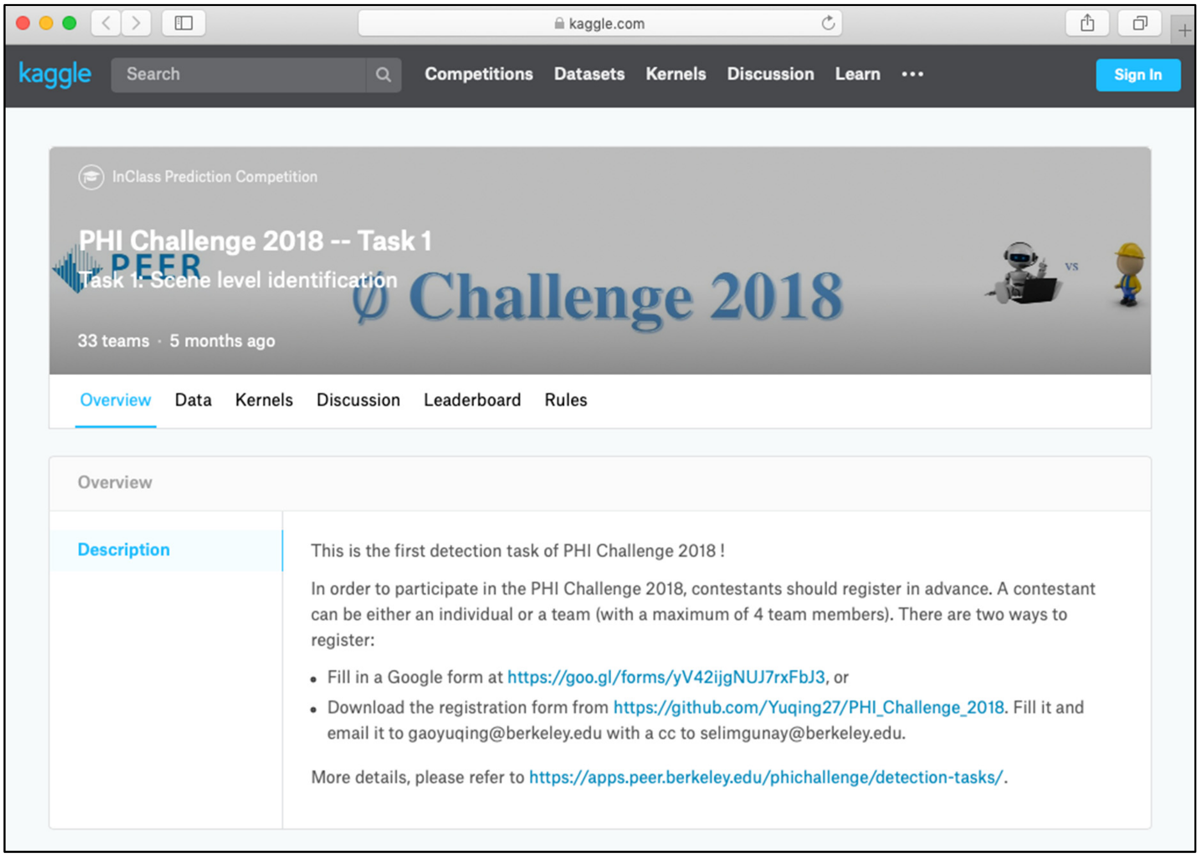


Figure 5.3 Homepage of Task 1 scene-level classification of the ϕ-Net Challenge of Fall 2018.

5.2.3 Data Download

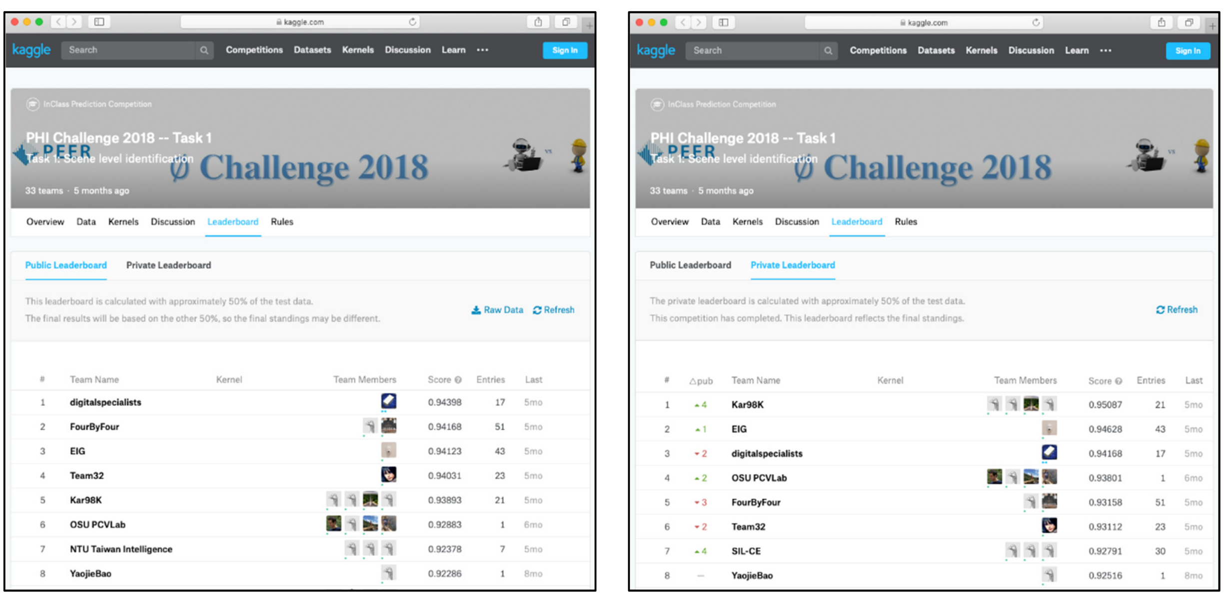
Participants were able to download the data in the **Data** tab from each task's webpage. There were four types of files in each task: X_train.npy, y_train.npy, X_test.npy, and sample_submission.csv.

- **X_train.npy** was the training data, whose shape was as follows: (Number of training data, 224, 224, 3).
- **y_train.npy** was the training label, whose value field was integer: {0, 1} for binary classification, {0, 1, 2} for three-class classification, and {0, 1, 2, 3} for four-class classification. Refer to the Kaggle page for the meaning of each value for each task.
- **X_test.npy** was the test data, whose shape was as follows: (Number of test data, 224, 224, 3).
- **sample_submission.csv** was the template for submission. The first column contained the data index, and the second column contained the prediction results. For each detection task, contestants needed to replace the second column with their own prediction results.

5.2.4 Results Submission and Leaderboard

As mentioned above, Kaggle provides a great platform for hosting online challenges; it automatically collect the results submitted by each team and shows the ranking with scores. Teams submitted their predictions through the **Submit Prediction** tab and viewed their rankings on the **Leaderboard** page. In order to avoid excessive submissions, the submission of an individual was considered a "team" submission, and the number of submissions per team was limited to a maximum of four submissions per day.

There were two leaderboards: a public leaderboard and a private leaderboard. The public leaderboard was calculated with approximately 50% of the test data; the private leaderboard was calculated with all test data and reflected the final score. Participants only had access to the public leaderboard. Screenshots of both leaderboards for Task 1 are shown in Figure 5.4. A comparison of the left image to the right one shows that the true scores and rankings slightly changed, e.g., No. 1 in the public board became No. 3 in the final ranking, and No. 5 in the public board became No. 1 in the final ranking. Giving participants access to only the public leaderboard increased the interest and motivated the teams to compete with one another. Note: the top-ranked team in the public leaderboard may not have been the "real" winner of the task; blindly pursuing the best performance on the public leaderboard may have led to overfitting issues.



(a) (b)

Figure 5.4 Rank of score in leaderboard: (a) public leaderboard; and (b) private leaderboard.

5.3 STATISTICS OF THE CHALLENGE RESULTS

Sixty-eight teams applied, and thirty-four teams submitted at least one entry, for a 50% final submission rate. Considering the resource-consuming process of conducting big-data analysis and given the short time period between the announcement of the challenge and the submission deadline, this is an amazing level of participation. Each team was required to submit the results via the Kaggle pages (e.g., *https://www.kaggle.com/c/phi2018task1*); the statistics of team submissions for each detection task are shown in Figure 5.5. Completing Tasks 1 and 2 were mandatory for access to the remaining six tasks and received the most submissions, where almost half of the teams submitted their first several predictions. Optional Tasks 3 to 8 still received nearly 20 team submissions each.

The final results for the individual tasks—evaluated by percentage accuracy (the number of correct predictions in test dataset/number of all data in the test dataset) and obtained from the private leaderboard—are listed in Table 5.1. The distributions of all results are presented in Figures 5.6(a) to (h). As mentioned above, note that the data used in the ϕ-Net Challenge were culled from the beta version. Thus, the final results of the challenge are only for reference purposes and cannot be used for calibration or considered as benchmark results.

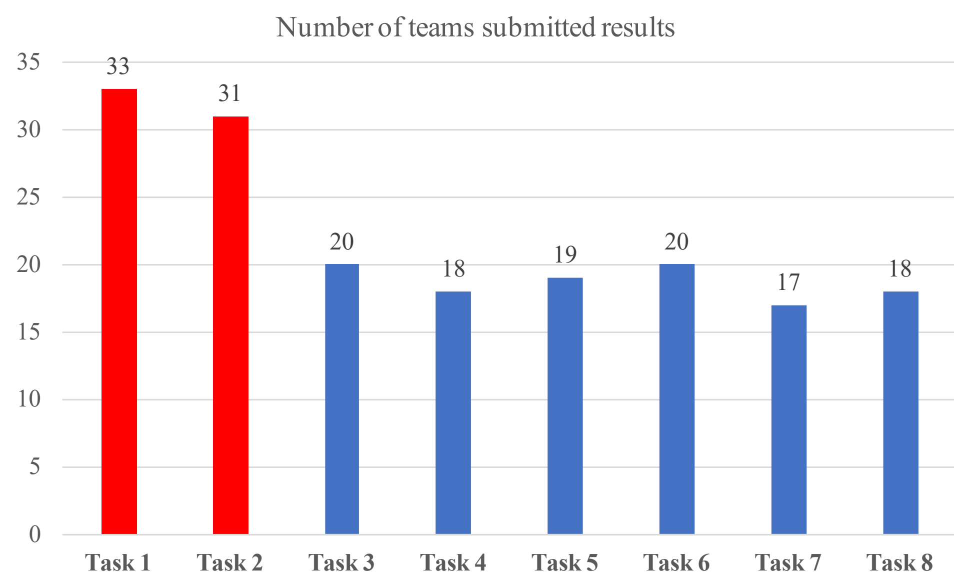
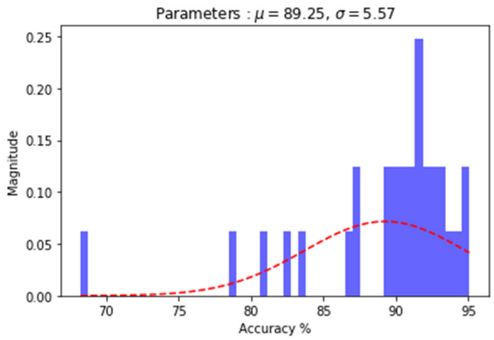


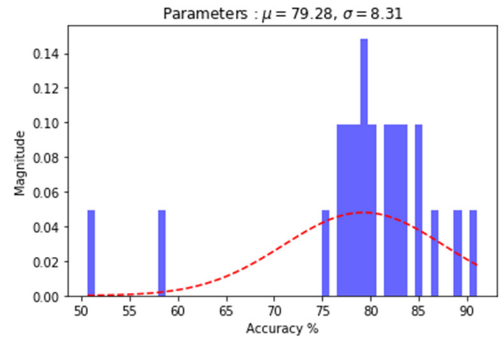
Figure 5.5 Team submissions for each task of the ϕ-Net Challenge of Fall 2018.

Table 5.1 Final results of all tasks from the ϕ-Net Challenge of Fall 2018.

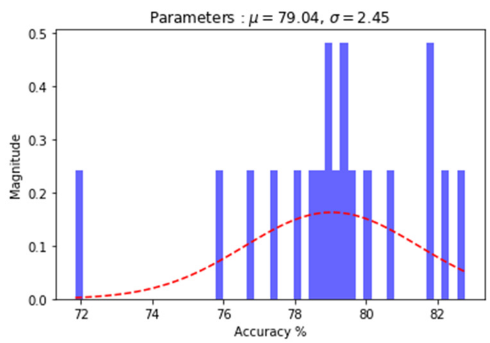
Task	Highest (%)	Lowest (%)	Mean μ	Standard deviation σ
1	95.09	31.82	89.25	5.57
2	91.08	47.16	79.28	8.31
3	82.77	71.84	79.04	2.45
4	99.82	94.84	98.16	1.34
5	70.77	46.15	62.91	5.28
6	76.29	64.44	71.61	3.30
7	79.94	55.01	66.23	5.28
8	76.90	36.78	64.89	8.56



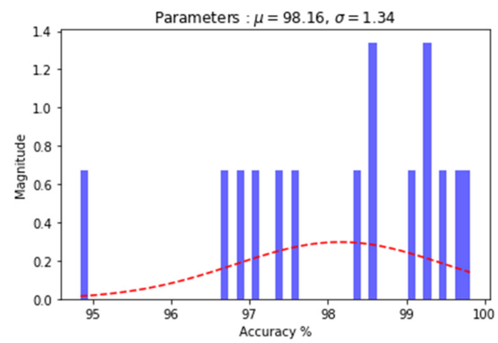
(a)



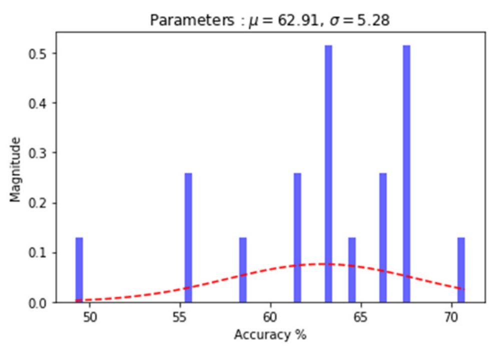
(b)



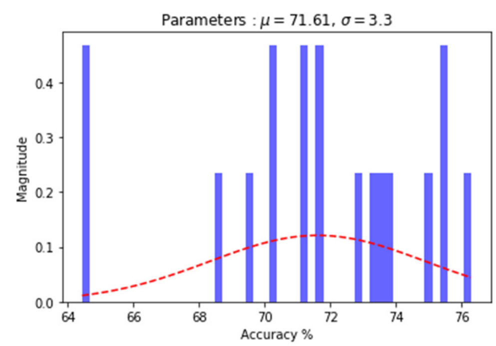
(c)



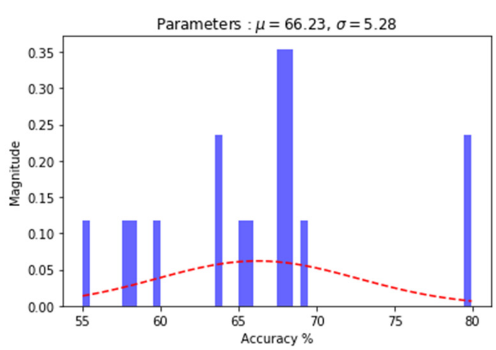
(d)



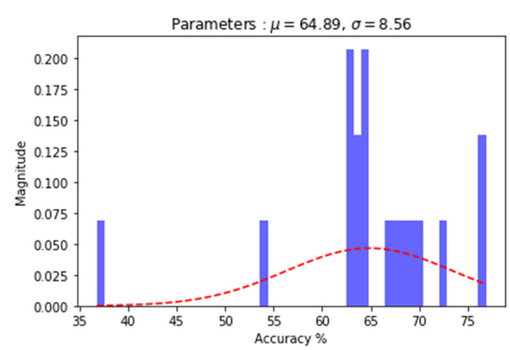
(e)



(f)



(g)



(h)

Figure 5.6 **Team submissions for each task of the ϕ-Net Challenge of Fall 2018: (a) Task 1: scene level; (b) Task 2: damage state; (c) Task 3: spalling condition; (d) Task 4: material type; (e) Task 5: collapse mode; (f) Task 6: component type; (g) Task 7: damage level; and (h) Task 8: damage type.**

5.4 SCORING OF OVERALL PERFORMANCE OF THE CHALLENGE

For overall performance evaluation, a special metric, Equation (5.1), considered the natural difficulty of tasks with both prior knowledge and the team's performance.

$$\text{Score} = \sum_i Weight_i \times Test\ Accuracy_1 \tag{5.1}$$

$$Weight_i = e^{(100-\mu_i)/\sum_i(100-\mu_i)} \times e^{\sigma_i/\sum_i \sigma_i} \times \phi_i \Big/ \sum_i \left[e^{(100-\mu_i)/\sum_i(100-\mu_i)} \times e^{\sigma_i/\sum_i \sigma_i} \times \phi_i \right] \tag{5.2}$$

This metric computed the weighted average of the score (accuracy) from each task where the weights, Equation (5.2), were based on the distribution of each team performance reflected by the distribution parameters (μ and σ), and the prior expected task difficulty factor ϕ. The μ and σ for each task were fitted through the final challenge results data—see Figure 5.6—and the ϕ was determined by the technical committee members with domain knowledge. Note: in the final submissions of Tasks 1 and 2, there were several invalid results, e.g., results close to 31.82% (Task 1) and 47.16% (Task 2), which were due to one team simply predicting all test data as one category; this resulted in a close-to-random guess for a relatively balanced dataset for three-class and binary classification problems, respectively. Thus, while using distribution parameters to compute the weights, some invalid submissions in Tasks 1 and 2 were ruled out.

Intuitively, with a higher average accuracy and a smaller standard deviation, the task became easier. Thus, $100 - \mu_i$ represented the difficulty of the task (the larger the value, the harder the task), and the magnitude of σ was also positively related to the difficulty level. Normalizing these two terms and taking their exponential, i.e., $e^{(100-\mu_i)/\sum_i(100-\mu_i)}$ and $e^{\sigma_i/\sum_i \sigma_i}$, may result in values slightly greater than 1, which are thought of as amplification factors. Multiplying them together along with the difficulty factor ϕ and further normalizing leads to the final weights for each task, which sum up to 1. Therefore, in some sense the final score computed by Equation (5.1) represented an average accuracy for all tasks. In summary, according to Equations (5.1) and (5.2), weights for overall performance were computed and reported; see Table 5.2.

Table 5.2 Combination weight considering task difficulty and distribution parameters of the ϕ-Net Challenge of Fall 2018.

Task	μ	σ	ϕ_i	Weights
1	89.25	5.57	50	0.130103498
2	79.28	8.31	30	0.088120542
3	79.04	2.45	30	0.076233949
4	98.16	1.34	30	0.067004295
5[18]	62.91	5.28	5	0.014852228
6	71.61	3.30	50	0.134993156
7	66.23	5.28	80	0.233489716
8	64.89	8.56	80	0.255202616

5.5 WINNERS OF THE CHALLENGE

Based on a comparison of the submitted predictions and ground truth, winners were identified by the contest judges in two categories: Computer Science/Data Science (CS/DATA) and Engineering (Non-CS/DATA). The top-three overall performance in all tasks, top-one performance in a single task, and best technical report were awarded for both categories. According to the weights computed in Section 5.4, the top-three winners in overall performance were evaluated and are listed in Table 5.3. For more details about winners, refer to the challenge website.

Table 5.3 Top-3 overall performance of the ϕ-Net Challenge of Fall 2018.

Place	CS/DATA		Non-CS/DATA	
	Team name	Score	Team name	Score
1	Kar98K	82.57	Stanford EIG	77.93
2	digitalspecialists	82.56	SIL-CE	77.29
3	OSU PCVLab	75.82	FourByFour-UCLA	76.97

[18] Due to two issues in Task 5 where (1) there were several duplicated and mislabeled images, i.e., two identical images with different labels; and (2) data leakage, i.e., several identical images occur in both training and test sets, the ϕ in Task 5 was taken as a low value of 5.

6 Summary and Conclusions

This report investigated applying deep-learning (DL) approaches to address on-going issues in civil engineering, such as post-disaster structural reconnaissance and structural health monitoring (SHM). Many factors impede the development of such DL applications in vision-based SHM, but the two most serious concerns are as follows: (1) there are no general automated detection principles or frameworks based on domain knowledge; and (2) the lack of a benchmark dataset with well-labeled large amounts of data. In order to address these concerns, an automated and hierarchical framework was proposed named the PEER Hub ImageNet (ϕ-Net) ϕ-Net consists of eight basic benchmark detection tasks based on the current domain knowledge and past reconnaissance experiences. These tasks are as follows: (1) scene level; (2) damage state; (3) spalling condition (material loss); (4) material type; (5) collapse mode; (6) component type; (7) damage level; and (8) damage type. To populate this framework, a large number of structural images were collected, preprocessed, labeled, and formed into an online open-source, large-scale, and multi-attribute image dataset, named ϕ-Net. As of November 2019, ϕ-Net contains 36,413 images with labels corresponding to the above eight attributes. Introduced herein were three deep convolutional neural network (CNN) models that provided a basic introduction to the concept of DL: VGG-16, VGG-19, and ResNet-50. With the well-defined tasks and well-built datasets, benchmark experiments were conducted whereby multiple influences and affine data augmentation (ADA) and transfer learning (TL) were considered in comparison studies under equivalent settings of hyper-parameters. All experimental results are reported and discussed herein, thus providing reference values for future studies. As a pre-event prior to open-sourcing the ϕ-Net dataset, in the Fall of 2018 PEER held the ϕ-Net Challenge, which is the first image-based challenge in civil engineering with emphasis on structural engineering. The rules, statistics, and winners of the challenge were also presented.

In summary, this report highlights the potential of using DL in vision-based SHM and automated damage recognition. Future plans include extending theϕ-Net dataset as follows: (1) the addition more fine-grained categories; (2) the investigation of additional DL models with hyper-parameter tuning; and (3) application of more complex training approaches. In addition, it is planned to make full use of these images in the ϕ-Net dataset by relabeling and exploring more computer-vision problems other than just classification, e.g., object detection, localization, semantic segmentation, etc.

REFERENCES

Cha Y.J., Choi W., Buyukozturk O. (2017a). Deep learning-based crack damage detection using convolutional neural networks, *Comput-Aided Civ. Inf.*, 32(5): 361–378.

Cha Y.J., Mahmoudkhani S., Buyukozturk O. (2017b). Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types, *Comput-Aided Civ. Inf.*, 33(9): 731–747.

Deng J., Wei D., Richard S., Li-Jia L., Kai L., Fei-Fei L. (2009). Imagenet: A large-scale hierarchical image database, *Proceedings, IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, Miami, FL.

Feng D., Feng M.Q. (2017). Experimental validation of cost-effective vision-based structural health monitoring, *Mech. Syst. Signal Pr.*, 88: 199–211.

Gao Y., Li K., Mosalam K.M., Günay S. (2018). Deep residual network with transfer learning for image-based structural damage recognition, *Proceedings, Eleventh U.S. National Conference on Earthquake Engineering*, Los Angeles, CA

Gao Y., Mosalam K.M. (2018). Deep transfer learning for image-based structural damage recognition, *Comput-Aided Civ. Inf.*, 33(9): 748-768.

Gao Y., Kong B., Mosalam K.M. (2019). Deep leaf-bootstrapping generative adversarial network for structural image data augmentation, *Comput-Aided Civ. Inf.*, 34(9): 755–773.

Goodfellow I., Bengio Y., Courville A. (2016). *Deep Learning*, MIT Press, Cambridge, MA.

He K., Zhang X., Ren S., Sun J. (2016). Deep residual learning for image recognition, *Proceedings, IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV.

Ioffe S., Szegedy C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift, *arXiv preprint arXiv:1502.03167*.

Koch C., Georgieva K., Kasireddy V., Akinci B., Fieguth P. (2015). A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure, *Adv. Eng. Inform.*, 29(2): 196–210.

Krizhevsky A. (2009). Learning multiple layers of features from tiny images, Technical Report TR-2009, University of Toronto, Toronto, Canada.

Krizhevsky A., Sutskever I., Hinton G.E. (2012), Imagenet classification with deep convolutional neural net- works, *Proceedings, Advances in Neural Information Processing Systems*, pp. 1097–1105, Lake Tahoe, NV.

LeCun Y., Bottou L., Bengio Y., Haffner P. (1998). Gradient-based learning applied to document recognition, *Proceedings, IEEE*, 86(11):2278–2324.

LeCun Y., Haffner P., Bottou L., Bengio Y. (1999). Object recognition with gradient-based learning, *Shape, Contour and Grouping in Computer Vision*; Forsyth, D.A., Mundy, J.L., di Gesú, V., Cipolla, R., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, pp. 319–345.

Li B., Mosalam K.M. (2012). Seismic performance of reinforced-concrete stairways during the 2008 Wenchuan earthquake, *J. Perform. Constr. Fac.*, 27(6): 721–730.

Moehle J.P. (2014). *Seismic Design of Reinforced Concrete Buildings*, McGraw Hill Professional, New York, NY.

Mosalam K.M., Takhirov S.M., Park S. (2014). Applications of laser scanning to structures in laboratory tests and field surveys, *Struct Control Hlth*, 21(1): 115–34.

Nagi J., Ducatelle F., Di Caro G.A., Cireşan, D., Meier U., Giusti A., Nagi F., Schmidhuber J., Gambardella L.M. (2011). Max-pooling convolutional neural networks for vision-based hand gesture recognition, *Proceedings, IEEE International Conference on Signal and Image Processing Applications*, pp. 342–347, Kuala Lumpur, Malaysia.

Nair V., Hinton G.E. (2010), Rectified linear units improve restricted Boltzmann machines, *Proceedings, 27th International Conference Machine Learning*, pp. 802–814, Haifa, Israel.

Pan S J., Yang Q. (2010). A survey on transfer learning, *IEEE Knowl Data Eng*., 22(10): 1345–1359.

Sezen H., Whittaker A.S., Elwood K.J., Mosalam K.M. (2003). Performance of reinforced concrete buildings during the August 17, 1999 Kocaeli, Turkey earthquake, and seismic design and construction practice in Turkey, *Eng. Struct*., 25(1) 103–14.

Simonyan K., Zisserman A. (2014). Very deep convolutional networks for large-scale image recognition, *arXiv:1409.1556*.

Soukup D., Huber-Mörk R. (2014), Convolutional neural networks for steel surface defect detection from photometric stereo images, *Proceedings, International Symposium Visual Computing*, pp. 668–677, Las Vegas, NV.

Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R. (2014), Dropout: a simple way to prevent neural networks from overfitting, *J. Mach. Learn. Res*., 15: 1929–1958.

Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. (2015). Going deeper with convolutions, *Proceedings, IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1–9, Boston, MA.

Torok M.M., Golparvar-Fard M., Kochersberger K. (2013). Image-based automated 3D crack detection for post-disaster building assessment, *J. Comp. Civil Eng*., 28(5): A4014004.

Vetrivel A., Gerke M., Kerle N., Nex F., Vosselman G. (2017). Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning, *ISPRS J. Photogramm.*, 140: 45–59, doi.org/10.1016/j.isprsjprs.2017.03.001.

Xie S., Girshick R., Dollár P., Tu Z., He K. (2017). Aggregated residual transformations for deep neural networks, *Proceedings, IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1492–1500.

Xue Y., Li Y. (2018). A fast detection method via region‐based fully convolutional neural networks for shield tunnel lining defects. *Comput-Aided Civ. Inf.*, 33(8): 638–654.

Yeum C.M., Dyke S.J. (2015). Vision-based automated crack detection for bridge inspection, *Comput-Aided Civ. Inf.*, 30(10): 759–70.

Yeum C.M., Dyke S.J., Ramirez L., Benes B. (2016). Big visual data analysis for damage evaluation in civil engineering, *Proceedings, International Conference Smart Infrastructure and Construction*, Cambridge, U.K.

Yoon H., Elanwar H., Choi H., Golparvar-Fard M., Spencer B.F. (2016). Target-free approach for vision-based structural system identification using consumer-grade cameras, *Struct. Control Hlth*, 23(12): 1405–1416.

Zhang A., Wang K.C., Li B., Yang E., Dai X., Peng Y., Fei Y., Liu Y., Li J.Q., Chen C. (2017). Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network, *Comput-Aided Civ. Inf.*, 32(10): 805–819.

Appendix A Data Collection Information

In this appendix, we present part of the information in the structural images collection of the ϕ-Net dataset with respect to earthquake events as listed in Table A.1.

Table A.1 Data collection information.

Date (Y-M-D)	Earthquake	# collected raw images
2017-09-19	Mexico city, Mexico	2,689
2016-11-13	Kaikoura, New Zealand	453
2016-09-09	Pawnee and Cushing, Oklahoma	1,609
2016-08-24	Central Italy, Italy	46
2016-02-05	Southern Taiwan, Taiwan	4,788
2016-04-15	Kumamoto, Japan	13
2015-04-25	Gorkha, Nepal	6,858
2010-01-12	Haiti	35
2014-08-24	South Napa, USA	3,218
2016-04-16	Musine, Ecuador	5,299
2004-09-28	Parkfield, USA	6
2016-09-03	Oklahoma, USA	1,502
2010-02-07	Chile	118
2011-03-11	Tohoku, Japan	201
2011-08-23	Virginia	67
2010-09-04	Canterbury, New Zealand	28
2015-05-20	Emilia, Italy	60

PEER REPORTS

PEER reports are available as a free PDF download from https://peer.berkeley.edu/peer-reports. In addition, printed hard copies of PEER reports can be ordered directly from our printer by following the instructions at https://peer.berkeley.edu/peer-reports. For other related questions about the PEER Report Series, contact the Pacific Earthquake Engineering Research Center, 325 Davis Hall, Mail Code 1792, Berkeley, CA 94720. Tel.: (510) 642-3437; and Email: peer_center@berkeley.edu.

PEER 2019/06 *Fluid-Structure Interaction and Python-Scripting Capabilities in OpenSees*. Minjie Zhu and Michael H. Scott. August 2019.

PEER 2019/05 *Expected Earthquake Performance of Buildings Designed to the California Building Code (California Alfred E. Alquist Seismic Safety Publication 19-01).* Grace S. Kang, Sifat Muin, Jorge Archbold, Bitanoosh Woods, and Khalid Mosalam. July 2019.

PEER 2019/04 *Aftershock Seismic Vulnerability and Time-Dependent Risk Assessment of Bridges*. Sujith Mangalathu, Mehrdad Shokrabadi, and Henry V. Burton. May 2019.

PEER 2019/03 *Ground-Motion Directivity Modeling for Seismic Hazard Applications.* Jennifer L. Donahue, Jonathan P. Stewart, Nicolas Gregor, and Yousef Bozorgnia. Review Panel: Jonathan D. Bray, Stephen A. Mahin, I. M. Idriss, Robert W. Graves, and Tom Shantz. May 2019.

PEER 2019/02 *Direct-Finite-Element Method for Nonlinear Earthquake Analysis of Concrete Dams Including Dam–Water–Foundation Rock Interaction.* Arnkjell Løkke and Anil K. Chopra. March 2019.

PEER 2019/01 *Flow-Failure Case History of the Las Palmas, Chile, Tailings Dam.* R. E. S. Moss, T. R. Gebhart, D. J. Frost, and C. Ledezma. January 2019.

PEER 2018/08 *Central and Eastern North America Ground-Motion Characterization: NGA-East Final Report*. Christine Goulet, Yousef Bozorgnia, Norman Abrahamson, Nicolas Kuehn, Linda Al Atik, Robert Youngs, Robert Graves, and Gail Atkinson. December 2018.

PEER 2018/07 *An Empirical Model for Fourier Amplitude Spectra using the NGA-West2 Database*. Jeff Bayless, and Norman A. Abrahamson. December 2018.

PEER 2018/06 *Estimation of Shear Demands on Rock-Socketed Drilled Shafts subjected to Lateral Loading*. Pedro Arduino, Long Chen, and Christopher R. McGann. December 2018.

PEER 2018/05 *Selection of Random Vibration Procedures for the NGA-East Project*. Albert Kottke, Norman A. Abrahamson, David M. Boore, Yousef Bozorgnia, Christine Goulet, Justin Hollenback, Tadahiro Kishida, Armen Der Kiureghian, Olga-Joan Ktenidou, Nicolas Kuehn, Ellen M. Rathje, Walter J. Silva, Eric Thompson, and Xiaoyue Wang. December 2018.

PEER 2018/04 *Capturing Directivity Effects in the Mean and Aleatory Variability of the NGA-West 2 Ground Motion Prediction Equations.* Jennie A. Watson-Lamprey. November 2018.

PEER 2018/03 *Probabilistic Seismic Hazard Analysis Code Verification*. Christie Hale, Norman Abrahamson, and Yousef Bozorgnia. July 2018.

PEER 2018/02 *Update of the BCHydro Subduction Ground-Motion Model using the NGA-Subduction Dataset.* Norman Abrahamson, Nicolas Kuehn, Zeynep Gulerce, Nicholas Gregor, Yousef Bozorgnia, Grace Parker, Jonathan Stewart, Brian Chiou, I. M. Idriss, Kenneth Campbell, and Robert Youngs. June 2018.

PEER 2018/01 *PEER Annual Report 2017–2018*. Khalid Mosalam, Amarnath Kasalanati, and Selim Günay. June 2018.

PEER 2017/12 *Experimental Investigation of the Behavior of Vintage and Retrofit Concentrically Braced Steel Frames under Cyclic Loading*. Barbara G. Simpson, Stephen A. Mahin, and Jiun-Wei Lai, December 2017.

PEER 2017/11 *Preliminary Studies on the Dynamic Response of a Seismically Isolated Prototype Gen-IV Sodium-Cooled Fast Reactor* (*PGSFR*). Benshun Shao, Andreas H. Schellenberg, Matthew J. Schoettler, and Stephen A. Mahin. December 2017.

PEER 2017/10 *Development of Time Histories for IEEE693 Testing and Analysis (including Seismically Isolated Equipment).* Shakhzod M. Takhirov, Eric Fujisaki, Leon Kempner, Michael Riley, and Brian Low. December 2017.

PEER 2017/09 *“”R” Package for Computation of Earthquake Ground-Motion Response Spectra.* Pengfei Wang, Jonathan P. Stewart, Yousef Bozorgnia, David M. Boore, and Tadahiro Kishida. December 2017.

PEER 2017/08 *Influence of Kinematic SSI on Foundation Input Motions for Bridges on Deep Foundations*. Benjamin J. Turner, Scott J. Brandenberg, and Jonathan P. Stewart. November 2017.

PEER 2017/07 *A Nonlinear Kinetic Model for Multi-Stage Friction Pendulum Systems*. Paul L. Drazin and Sanjay Govindjee. September 2017.

PEER 2017/06 *Guidelines for Performance-Based Seismic Design of Tall Buildings, Version 2.02*. TBI Working Group led by co-chairs Ron Hamburger and Jack Moehle: Jack Baker, Jonathan Bray, C.B. Crouse, Greg Deierlein, John Hooper, Marshall Lew, Joe Maffei, Stephen Mahin, James Malley, Farzad Naeim, Jonathan Stewart, and John Wallace. May 2017.

PEER 2017/05 *Recommendations for Ergodic Nonlinear Site Amplification in Central and Eastern North America*. Youssef M.A. Hashash, Joseph A. Harmon, Okan Ilhan, Grace A. Parker, and Jonathan P. Stewart. March 2017.

PEER 2017/04 *Expert Panel Recommendations for Ergodic Site Amplification in Central and Eastern North America*. Jonathan P. Stewart, Grace A. Parker, Joseph P. Harmon, Gail M. Atkinson, David M. Boore, Robert B. Darragh, Walter J. Silva, and Youssef M.A. Hashash. March 2017.

PEER 2017/03 *NGA-East Ground-Motion Models for the U.S. Geological Survey National Seismic Hazard Maps*. Christine A. Goulet, Yousef Bozorgnia, Nicolas Kuehn, Linda Al Atik, Robert R. Youngs, Robert W. Graves, and Gail M. Atkinson. March 2017.

PEER 2017/02 *U.S.–New Zealand–Japan Workshop: Liquefaction-Induced Ground Movements Effects, University of California, Berkeley, California, 2–4 November 2016.* Jonathan D. Bray, Ross W. Boulanger, Misko Cubrinovski, Kohji Tokimatsu, Steven L. Kramer, Thomas O’Rourke, Ellen Rathje, Russell A. Green, Peter K. Robinson, and Christine Z. Beyzaei. March 2017.

PEER 2017/01 *2016 PEER Annual Report*. Khalid M. Mosalam, Amarnath Kasalanati, and Grace Kang. March 2017.

PEER 2016/10 *Performance-Based Robust Nonlinear Seismic Analysis with Application to Reinforced Concrete Bridge Systems*. Xiao Ling and Khalid M. Mosalam. December 2016.

PEER 2017/09 *Detailing Requirements for Column Plastic Hinges subjected to Combined Flexural, Axial, and Torsional Seismic Loading.* Gabriel Hurtado and Jack P. Moehle. December 2016.

PEER 2016/08 *Resilience of Critical Structures, Infrastructure, and Communities*. Gian Paolo Cimellaro, Ali Zamani-Noori, Omar Kamouh, Vesna Terzic, and Stephen A. Mahin. December 2016.

PEER 2016/07 *Hybrid Simulation Theory for a Classical Nonlinear Dynamical System*. Paul L. Drazin and Sanjay Govindjee. September 2016.

PEER 2016/06 *California Earthquake Early Warning System Benefit Study*. Laurie A. Johnson, Sharyl Rabinovici, Grace S. Kang, and Stephen A. Mahin. July 2006.

PEER 2016/05 *Ground-Motion Prediction Equations for Arias Intensity Consistent with the NGA-West2 Ground-Motion Models.* Charlotte Abrahamson, Hao-Jun Michael Shi, and Brian Yang. July 2016.

PEER 2016/04 *The M_W 6.0 South Napa Earthquake of August 24, 2014: A Wake-Up Call for Renewed Investment in Seismic Resilience Across California.* Prepared for the California Seismic Safety Commission, Laurie A. Johnson and Stephen A. Mahin. May 2016.

PEER 2016/03 *Simulation Confidence in Tsunami-Driven Overland Flow.* Patrick Lynett. May 2016.

PEER 2016/02 *Semi-Automated Procedure for Windowing time Series and Computing Fourier Amplitude Spectra for the NGA-West2 Database.* Tadahiro Kishida, Olga-Joan Ktenidou, Robert B. Darragh, and Walter J. Silva. May 2016.

PEER 2016/01 *A Methodology for the Estimation of Kappa (κ) from Large Datasets: Example Application to Rock Sites in the NGA-East Database and Implications on Design Motions.* Olga-Joan Ktenidou, Norman A. Abrahamson, Robert B. Darragh, and Walter J. Silva. April 2016.

PEER 2015/13 *Self-Centering Precast Concrete Dual-Steel-Shell Columns for Accelerated Bridge Construction: Seismic Performance, Analysis, and Design*. Gabriele Guerrini, José I. Restrepo, Athanassios Vervelidis, and Milena Massari. December 2015.

PEER 2015/12 *Shear-Flexure Interaction Modeling for Reinforced Concrete Structural Walls and Columns under Reversed Cyclic Loading*. Kristijan Kolozvari, Kutay Orakcal, and John Wallace. December 2015.

PEER 2015/11 *Selection and Scaling of Ground Motions for Nonlinear Response History Analysis of Buildings in Performance-Based Earthquake Engineering*. N. Simon Kwong and Anil K. Chopra. December 2015.

PEER 2015/10 *Structural Behavior of Column-Bent Cap Beam-Box Girder Systems in Reinforced Concrete Bridges Subjected to Gravity and Seismic Loads. Part II: Hybrid Simulation and Post-Test Analysis*. Mohamed A. Moustafa and Khalid M. Mosalam. November 2015.

PEER 2015/09 *Structural Behavior of Column-Bent Cap Beam-Box Girder Systems in Reinforced Concrete Bridges Subjected to Gravity and Seismic Loads. Part I: Pre-Test Analysis and Quasi-Static Experiments*. Mohamed A. Moustafa and Khalid M. Mosalam. September 2015.

PEER 2015/08 *NGA-East: Adjustments to Median Ground-Motion Models for Center and Eastern North America*. August 2015.

PEER 2015/07 *NGA-East: Ground-Motion Standard-Deviation Models for Central and Eastern North America*. Linda Al Atik. June 2015.

PEER 2015/06 *Adjusting Ground-Motion Intensity Measures to a Reference Site for which V_{S30} = 3000 m/sec*. David M. Boore. May 2015.

PEER 2015/05 *Hybrid Simulation of Seismic Isolation Systems Applied to an APR-1400 Nuclear Power Plant.* Andreas H. Schellenberg, Alireza Sarebanha, Matthew J. Schoettler, Gilberto Mosqueda, Gianmario Benzoni, and Stephen A. Mahin. April 2015.

PEER 2015/04 *NGA-East: Median Ground-Motion Models for the Central and Eastern North America Region*. April 2015.

PEER 2015/03 *Single Series Solution for the Rectangular Fiber-Reinforced Elastomeric Isolator Compression Modulus*. James M. Kelly and Niel C. Van Engelen. March 2015.

PEER 2015/02 *A Full-Scale, Single-Column Bridge Bent Tested by Shake-Table Excitation.* Matthew J. Schoettler, José I. Restrepo, Gabriele Guerrini, David E. Duck, and Francesco Carrea. March 2015.

PEER 2015/01 *Concrete Column Blind Prediction Contest 2010: Outcomes and Observations.* Vesna Terzic, Matthew J. Schoettler, José I. Restrepo, and Stephen A Mahin. March 2015.

PEER 2014/20 *Stochastic Modeling and Simulation of Near-Fault Ground Motions for Performance-Based Earthquake Engineering*. Mayssa Dabaghi and Armen Der Kiureghian. December 2014.

PEER 2014/19 *Seismic Response of a Hybrid Fiber-Reinforced Concrete Bridge Column Detailed for Accelerated Bridge Construction*. Wilson Nguyen, William Trono, Marios Panagiotou, and Claudia P. Ostertag. December 2014.

PEER 2014/18 *Three-Dimensional Beam-Truss Model for Reinforced Concrete Walls and Slabs Subjected to Cyclic Static or Dynamic Loading*. Yuan Lu, Marios Panagiotou, and Ioannis Koutromanos. December 2014.

PEER 2014/17 *PEER NGA-East Database.* Christine A. Goulet, Tadahiro Kishida, Timothy D. Ancheta, Chris H. Cramer, Robert B. Darragh, Walter J. Silva, Youssef M.A. Hashash, Joseph Harmon, Jonathan P. Stewart, Katie E. Wooddell, and Robert R. Youngs. October 2014.

PEER 2014/16 *Guidelines for Performing Hazard-Consistent One-Dimensional Ground Response Analysis for Ground Motion Prediction*. Jonathan P. Stewart, Kioumars Afshari, and Youssef M.A. Hashash. October 2014.

PEER 2014/15 *NGA-East Regionalization Report: Comparison of Four Crustal Regions within Central and Eastern North America using Waveform Modeling and 5%-Damped Pseudo-Spectral Acceleration Response*. Jennifer Dreiling, Marius P. Isken, Walter D. Mooney, Martin C. Chapman, and Richard W. Godbee. October 2014.

PEER 2014/14 *Scaling Relations between Seismic Moment and Rupture Area of Earthquakes in Stable Continental Regions*. Paul Somerville. August 2014.

PEER 2014/13 *PEER Preliminary Notes and Observations on the August 24, 2014, South Napa Earthquake*. Grace S. Kang and Stephen A. Mahin, Editors. September 2014.

PEER 2014/12 *Reference-Rock Site Conditions for Central and Eastern North America: Part II – Attenuation (Kappa) Definition*. Kenneth W. Campbell, Youssef M.A. Hashash, Byungmin Kim, Albert R. Kottke, Ellen M. Rathje, Walter J. Silva, and Jonathan P. Stewart. August 2014.

PEER 2014/11 *Reference-Rock Site Conditions for Central and Eastern North America: Part I - Velocity Definition*. Youssef M.A. Hashash, Albert R. Kottke, Jonathan P. Stewart, Kenneth W. Campbell, Byungmin Kim, Ellen M. Rathje, Walter J. Silva, Sissy Nikolaou, and Cheryl Moss. August 2014.

PEER 2014/10 *Evaluation of Collapse and Non-Collapse of Parallel Bridges Affected by Liquefaction and Lateral Spreading*. Benjamin Turner, Scott J. Brandenberg, and Jonathan P. Stewart. August 2014.

PEER 2014/09 *PEER Arizona Strong-Motion Database and GMPEs Evaluation*. Tadahiro Kishida, Robert E. Kayen, Olga-Joan Ktenidou, Walter J. Silva, Robert B. Darragh, and Jennie Watson-Lamprey. June 2014.

PEER 2014/08 *Unbonded Pretensioned Bridge Columns with Rocking Detail*. Jeffrey A. Schaefer, Bryan Kennedy, Marc O. Eberhard, and John F. Stanton. June 2014.

PEER 2014/07 *Northridge 20 Symposium Summary Report: Impacts, Outcomes, and Next Steps*. May 2014.

PEER 2014/06 *Report of the Tenth Planning Meeting of NEES/E-Defense Collaborative Research on Earthquake Engineering*. December 2013.

PEER 2014/05 *Seismic Velocity Site Characterization of Thirty-One Chilean Seismometer Stations by Spectral Analysis of Surface Wave Dispersion.* Robert Kayen, Brad D. Carkin, Skye Corbet, Camilo Pinilla, Allan Ng, Edward Gorbis, and Christine Truong. April 2014.

PEER 2014/04 Effect of Vertical Acceleration on Shear Strength of Reinforced Concrete Columns. Hyerin Lee and Khalid M. Mosalam. April 2014.

PEER 2014/03 *Retest of Thirty-Year-Old Neoprene Isolation Bearings.* James M. Kelly and Niel C. Van Engelen. March 2014.

PEER 2014/02 *Theoretical Development of Hybrid Simulation Applied to Plate Structures.* Ahmed A. Bakhaty, Khalid M. Mosalam, and Sanjay Govindjee. January 2014.

PEER 2014/01 *Performance-Based Seismic Assessment of Skewed Bridges.* Peyman Kaviani, Farzin Zareian, and Ertugrul Taciroglu. January 2014.

PEER 2013/26 *Urban Earthquake Engineerin*g. Proceedings of the U.S.-Iran Seismic Workshop. December 2013.

PEER 2013/25 *Earthquake Engineering for Resilient Communities: 2013 PEER Internship Program Research Report Collection.* Heidi Tremayne (Editor), Stephen A. Mahin (Editor), Jorge Archbold Monterossa, Matt Brosman, Shelly Dean, Katherine deLaveaga, Curtis Fong, Donovan Holder, Rakeeb Khan, Elizabeth Jachens, David Lam, Daniela Martinez Lopez, Mara Minner, Geffen Oren, Julia Pavicic, Melissa Quinonez, Lorena Rodriguez, Sean Salazar, Kelli Slaven, Vivian Steyert, Jenny Taing, and Salvador Tena. December 2013.

PEER 2013/24 *NGA-West2 Ground Motion Prediction Equations for Vertical Ground Motions*. September 2013.

PEER 2013/23 *Coordinated Planning and Preparedness for Fire Following Major Earthquakes*. Charles Scawthorn. November 2013.

PEER 2013/22 *GEM-PEER Task 3 Project: Selection of a Global Set of Ground Motion Prediction Equations.* Jonathan P. Stewart, John Douglas, Mohammad B. Javanbarg, Carola Di Alessandro, Yousef Bozorgnia, Norman A. Abrahamson, David M. Boore, Kenneth W. Campbell, Elise Delavaud, Mustafa Erdik, and Peter J. Stafford. December 2013.

PEER 2013/21 *Seismic Design and Performance of Bridges with Columns on Rocking Foundations*. Grigorios Antonellis and Marios Panagiotou. September 2013.

PEER 2013/20 *Experimental and Analytical Studies on the Seismic Behavior of Conventional and Hybrid Braced Frames*. Jiun-Wei Lai and Stephen A. Mahin. September 2013.

PEER 2013/19 *Toward Resilient Communities: A Performance-Based Engineering Framework for Design and Evaluation of the Built Environment.* Michael William Mieler, Bozidar Stojadinovic, Robert J. Budnitz, Stephen A. Mahin, and Mary C. Comerio. September 2013.

PEER 2013/18 *Identification of Site Parameters that Improve Predictions of Site Amplification*. Ellen M. Rathje and Sara Navidi. July 2013.

PEER 2013/17 *Response Spectrum Analysis of Concrete Gravity Dams Including Dam-Water-Foundation Interaction*. Arnkjell Løkke and Anil K. Chopra. July 2013.

PEER 2013/16 *Effect of Hoop Reinforcement Spacing on the Cyclic Response of Large Reinforced Concrete Special Moment Frame Beams.* Marios Panagiotou, Tea Visnjic, Grigorios Antonellis, Panagiotis Galanis, and Jack P. Moehle. June 2013.

PEER 2013/15 *A Probabilistic Framework to Include the Effects of Near-Fault Directivity in Seismic Hazard Assessment*. Shrey Kumar Shahi, Jack W. Baker. October 2013.

PEER 2013/14 *Hanging-Wall Scaling using Finite-Fault Simulations.* Jennifer L. Donahue and Norman A. Abrahamson. September 2013.

PEER 2013/13 *Semi-Empirical Nonlinear Site Amplification and its Application in NEHRP Site Factors*. Jonathan P. Stewart and Emel Seyhan. November 2013.

PEER 2013/12 *Nonlinear Horizontal Site Response for the NGA-West2 Project.* Ronnie Kamai, Norman A. Abramson, Walter J. Silva. May 2013.

PEER 2013/11 *Epistemic Uncertainty for NGA-West2 Models.* Linda Al Atik and Robert R. Youngs. May 2013.

PEER 2013/10 *NGA-West 2 Models for Ground-Motion Directionality.* Shrey K. Shahi and Jack W. Baker. May 2013.

PEER 2013/09 *Final Report of the NGA-West2 Directivity Working Group.* Paul Spudich, Jeffrey R. Bayless, Jack W. Baker, Brian S.J. Chiou, Badie Rowshandel, Shrey Shahi, and Paul Somerville. May 2013.

PEER 2013/08 *NGA-West2 Model for Estimating Average Horizontal Values of Pseudo-Absolute Spectral Accelerations Generated by Crustal Earthquakes*. I. M. Idriss. May 2013.

PEER 2013/07 *Update of the Chiou and Youngs NGA Ground Motion Model for Average Horizontal Component of Peak Ground Motion and Response Spectra*. Brian Chiou and Robert Youngs. May 2013.

PEER 2013/06 *NGA-West2 Campbell-Bozorgnia Ground Motion Model for the Horizontal Components of PGA, PGV, and 5%-Damped Elastic Pseudo-Acceleration Response Spectra for Periods Ranging from 0.01 to 10 sec.* Kenneth W. Campbell and Yousef Bozorgnia. May 2013.

PEER 2013/05 *NGA-West 2 Equations for Predicting Response Spectral Accelerations for Shallow Crustal Earthquakes*. David M. Boore, Jonathan P. Stewart, Emel Seyhan, and Gail M. Atkinson. May 2013.

PEER 2013/04 *Update of the AS08 Ground-Motion Prediction Equations Based on the NGA-West2 Data Set*. Norman Abrahamson, Walter Silva, and Ronnie Kamai. May 2013.

PEER 2013/03 *PEER NGA-West2 Database.* Timothy D. Ancheta, Robert B. Darragh, Jonathan P. Stewart, Emel Seyhan, Walter J. Silva, Brian S.J. Chiou, Katie E. Wooddell, Robert W. Graves, Albert R. Kottke, David M. Boore, Tadahiro Kishida, and Jennifer L. Donahue. May 2013.

PEER 2013/02 *Hybrid Simulation of the Seismic Response of Squat Reinforced Concrete Shear Walls*. Catherine A. Whyte and Bozidar Stojadinovic. May 2013.

PEER 2013/01 *Housing Recovery in Chile: A Qualitative Mid-program Review*. Mary C. Comerio. February 2013.

PEER 2012/08 *Guidelines for Estimation of Shear Wave Velocity.* Bernard R. Wair, Jason T. DeJong, and Thomas Shantz. December 2012.

PEER 2012/07 *Earthquake Engineering for Resilient Communities: 2012 PEER Internship Program Research Report Collection.* Heidi Tremayne (Editor), Stephen A. Mahin (Editor), Collin Anderson, Dustin Cook, Michael Erceg, Carlos Esparza, Jose Jimenez, Dorian Krausz, Andrew Lo, Stephanie Lopez, Nicole McCurdy, Paul Shipman, Alexander Strum, Eduardo Vega. December 2012.

PEER 2012/06 *Fragilities for Precarious Rocks at Yucca Mountain.* Matthew D. Purvance, Rasool Anooshehpoor, and James N. Brune. December 2012.

PEER 2012/05 *Development of Simplified Analysis Procedure for Piles in Laterally Spreading Layered Soils.* Christopher R. McGann, Pedro Arduino, and Peter Mackenzie–Helnwein. December 2012.

PEER 2012/04 *Unbonded Pre-Tensioned Columns for Bridges in Seismic Regions.* Phillip M. Davis, Todd M. Janes, Marc O. Eberhard, and John F. Stanton. December 2012.

PEER 2012/03 *Experimental and Analytical Studies on Reinforced Concrete Buildings with Seismically Vulnerable Beam-Column Joints*. Sangjoon Park and Khalid M. Mosalam. October 2012.

PEER 2012/02 *Seismic Performance of Reinforced Concrete Bridges Allowed to Uplift during Multi-Directional Excitation*. Andres Oscar Espinoza and Stephen A. Mahin. July 2012.

PEER 2012/01 *Spectral Damping Scaling Factors for Shallow Crustal Earthquakes in Active Tectonic Regions.* Sanaz Rezaeian, Yousef Bozorgnia, I. M. Idriss, Kenneth Campbell, Norman Abrahamson, and Walter Silva. July 2012.

PEER 2011/10 *Earthquake Engineering for Resilient Communities: 2011 PEER Internship Program Research Report Collection*. Heidi Faison and Stephen A. Mahin, Editors. December 2011.

PEER 2011/09 *Calibration of Semi-Stochastic Procedure for Simulating High-Frequency Ground Motions*. Jonathan P. Stewart, Emel Seyhan, and Robert W. Graves. December 2011.

PEER 2011/08 *Water Supply in regard to Fire Following Earthquake*. Charles Scawthorn. November 2011.

PEER 2011/07 *Seismic Risk Management in Urban Areas*. Proceedings of a U.S.-Iran-Turkey Seismic Workshop. September 2011.

PEER 2011/06 *The Use of Base Isolation Systems to Achieve Complex Seismic Performance Objectives*. Troy A. Morgan and Stephen A. Mahin. July 2011.

PEER 2011/05 *Case Studies of the Seismic Performance of Tall Buildings Designed by Alternative Means.* Task 12 Report for the Tall Buildings Initiative. Jack Moehle, Yousef Bozorgnia, Nirmal Jayaram, Pierson Jones, Mohsen Rahnama, Nilesh Shome, Zeynep Tuna, John Wallace, Tony Yang, and Farzin Zareian. July 2011.

PEER 2011/04 *Recommended Design Practice for Pile Foundations in Laterally Spreading Ground.* Scott A. Ashford, Ross W. Boulanger, and Scott J. Brandenberg. June 2011.

PEER 2011/03 *New Ground Motion Selection Procedures and Selected Motions for the PEER Transportation Research Program*. Jack W. Baker, Ting Lin, Shrey K. Shahi, and Nirmal Jayaram. March 2011.

PEER 2011/02 *A Bayesian Network Methodology for Infrastructure Seismic Risk Assessment and Decision Support*. Michelle T. Bensi, Armen Der Kiureghian, and Daniel Straub. March 2011.

PEER 2011/01 *Demand Fragility Surfaces for Bridges in Liquefied and Laterally Spreading Ground.* Scott J. Brandenberg, Jian Zhang, Pirooz Kashighandi, Yili Huo, and Minxing Zhao. March 2011.

PEER 2010/05 *Guidelines for Performance-Based Seismic Design of Tall Buildings.* Developed by the Tall Buildings Initiative. November 2010.

PEER 2010/04 *Application Guide for the Design of Flexible and Rigid Bus Connections between Substation Equipment Subjected to Earthquakes.* Jean-Bernard Dastous and Armen Der Kiureghian. September 2010.

PEER 2010/03 *Shear Wave Velocity as a Statistical Function of Standard Penetration Test Resistance and Vertical Effective Stress at Caltrans Bridge Sites.* Scott J. Brandenberg, Naresh Bellana, and Thomas Shantz. June 2010.

PEER 2010/02 *Stochastic Modeling and Simulation of Ground Motions for Performance-Based Earthquake Engineering.* Sanaz Rezaeian and Armen Der Kiureghian. June 2010.

PEER 2010/01 *Structural Response and Cost Characterization of Bridge Construction Using Seismic Performance Enhancement Strategies.* Ady Aviram, Božidar Stojadinović, Gustavo J. Parra-Montesinos, and Kevin R. Mackie. March 2010.

PEER 2009/03 *The Integration of Experimental and Simulation Data in the Study of Reinforced Concrete Bridge Systems Including Soil-Foundation-Structure Interaction.* Matthew Dryden and Gregory L. Fenves. November 2009.

PEER 2009/02 *Improving Earthquake Mitigation through Innovations and Applications in Seismic Science, Engineering, Communication, and Response*. Proceedings of a U.S.-Iran Seismic Workshop. October 2009.

PEER 2009/01 *Evaluation of Ground Motion Selection and Modification Methods: Predicting Median Interstory Drift Response of Buildings*. Curt B. Haselton, Editor. June 2009.

PEER 2008/10 *Technical Manual for Strata.* Albert R. Kottke and Ellen M. Rathje. February 2009.

PEER 2008/09 *NGA Model for Average Horizontal Component of Peak Ground Motion and Response Spectra.* Brian S.-J. Chiou and Robert R. Youngs. November 2008.

PEER 2008/08 *Toward Earthquake-Resistant Design of Concentrically Braced Steel Structures*. Patxi Uriz and Stephen A. Mahin. November 2008.

PEER 2008/07 *Using OpenSees for Performance-Based Evaluation of Bridges on Liquefiable Soils*. Stephen L. Kramer, Pedro Arduino, and HyungSuk Shin. November 2008.

PEER 2008/06 *Shaking Table Tests and Numerical Investigation of Self-Centering Reinforced Concrete Bridge Columns*. Hyung IL Jeong, Junichi Sakai, and Stephen A. Mahin. September 2008.

PEER 2008/05 *Performance-Based Earthquake Engineering Design Evaluation Procedure for Bridge Foundations Undergoing Liquefaction-Induced Lateral Ground Displacement*. Christian A. Ledezma and Jonathan D. Bray. August 2008.

PEER 2008/04 *Benchmarking of Nonlinear Geotechnical Ground Response Analysis Procedures.* Jonathan P. Stewart, Annie On-Lei Kwok, Youssef M. A. Hashash, Neven Matasovic, Robert Pyke, Zhiliang Wang, and Zhaohui Yang. August 2008.

PEER 2008/03 *Guidelines for Nonlinear Analysis of Bridge Structures in California.* Ady Aviram, Kevin R. Mackie, and Božidar Stojadinović. August 2008.

PEER 2008/02 *Treatment of Uncertainties in Seismic-Risk Analysis of Transportation Systems*. Evangelos Stergiou and Anne S. Kiremidjian. July 2008.

PEER 2008/01 *Seismic Performance Objectives for Tall Buildings.* William T. Holmes, Charles Kircher, William Petak, and Nabih Youssef. August 2008.

PEER 2007/12 *An Assessment to Benchmark the Seismic Performance of a Code-Conforming Reinforced Concrete Moment-Frame Building.* Curt Haselton, Christine A. Goulet, Judith Mitrani-Reiser, James L. Beck, Gregory G. Deierlein, Keith A. Porter, Jonathan P. Stewart, and Ertugrul Taciroglu. August 2008.

PEER 2007/11 *Bar Buckling in Reinforced Concrete Bridge Columns.* Wayne A. Brown, Dawn E. Lehman, and John F. Stanton. February 2008.

PEER 2007/10 *Computational Modeling of Progressive Collapse in Reinforced Concrete Frame Structures*. Mohamed M. Talaat and Khalid M. Mosalam. May 2008.

PEER 2007/09 *Integrated Probabilistic Performance-Based Evaluation of Benchmark Reinforced Concrete Bridges*. Kevin R. Mackie, John-Michael Wong, and Božidar Stojadinović. January 2008.

PEER 2007/08 *Assessing Seismic Collapse Safety of Modern Reinforced Concrete Moment-Frame Buildings*. Curt B. Haselton and Gregory G. Deierlein. February 2008.

PEER 2007/07 *Performance Modeling Strategies for Modern Reinforced Concrete Bridge Columns*. Michael P. Berry and Marc O. Eberhard. April 2008.

PEER 2007/06 *Development of Improved Procedures for Seismic Design of Buried and Partially Buried Structures*. Linda Al Atik and Nicholas Sitar. June 2007.

PEER 2007/05 *Uncertainty and Correlation in Seismic Risk Assessment of Transportation Systems*. Renee G. Lee and Anne S. Kiremidjian. July 2007.

PEER 2007/04 *Numerical Models for Analysis and Performance-Based Design of Shallow Foundations Subjected to Seismic Loading.* Sivapalan Gajan, Tara C. Hutchinson, Bruce L. Kutter, Prishati Raychowdhury, José A. Ugalde, and Jonathan P. Stewart. May 2008.

PEER 2007/03 *Beam-Column Element Model Calibrated for Predicting Flexural Response Leading to Global Collapse of RC Frame Buildings*. Curt B. Haselton, Abbie B. Liel, Sarah Taylor Lange, and Gregory G. Deierlein. May 2008.

PEER 2007/02 *Campbell-Bozorgnia NGA Ground Motion Relations for the Geometric Mean Horizontal Component of Peak and Spectral Ground Motion Parameters*. Kenneth W. Campbell and Yousef Bozorgnia. May 2007.

PEER 2007/01 *Boore-Atkinson NGA Ground Motion Relations for the Geometric Mean Horizontal Component of Peak and Spectral Ground Motion Parameters*. David M. Boore and Gail M. Atkinson. May 2007.

PEER 2006/12 Societal Implications of Performance-Based Earthquake Engineering. Peter J. May. May 2007.

PEER 2006/11 *Probabilistic Seismic Demand Analysis Using Advanced Ground Motion Intensity Measures, Attenuation Relationships, and Near-Fault Effects*. Polsak Tothong and C. Allin Cornell. March 2007.

PEER 2006/10 *Application of the PEER PBEE Methodology to the I-880 Viaduct*. Sashi Kunnath. February 2007.

PEER 2006/09 *Quantifying Economic Losses from Travel Forgone Following a Large Metropolitan Earthquake*. James Moore, Sungbin Cho, Yue Yue Fan, and Stuart Werner. November 2006.

PEER 2006/08 *Vector-Valued Ground Motion Intensity Measures for Probabilistic Seismic Demand Analysis*. Jack W. Baker and C. Allin Cornell. October 2006.

PEER 2006/07 *Analytical Modeling of Reinforced Concrete Walls for Predicting Flexural and Coupled–Shear-Flexural Responses*. Kutay Orakcal, Leonardo M. Massone, and John W. Wallace. October 2006.

PEER 2006/06 *Nonlinear Analysis of a Soil-Drilled Pier System under Static and Dynamic Axial Loading*. Gang Wang and Nicholas Sitar. November 2006.

PEER 2006/05 *Advanced Seismic Assessment Guidelines.* Paolo Bazzurro, C. Allin Cornell, Charles Menun, Maziar Motahari, and Nicolas Luco. September 2006.

PEER 2006/04 *Probabilistic Seismic Evaluation of Reinforced Concrete Structural Components and Systems*. Tae Hyung Lee and Khalid M. Mosalam. August 2006.

PEER 2006/03 *Performance of Lifelines Subjected to Lateral Spreading.* Scott A. Ashford and Teerawut Juirnarongrit. July 2006.

PEER 2006/02 *Pacific Earthquake Engineering Research Center Highway Demonstration Project.* Anne Kiremidjian, James Moore, Yue Yue Fan, Nesrin Basoz, Ozgur Yazali, and Meredith Williams. April 2006.

PEER 2006/01 *Bracing Berkeley. A Guide to Seismic Safety on the UC Berkeley Campus.* Mary C. Comerio, Stephen Tobriner, and Ariane Fehrenkamp. January 2006.

PEER 2005/17 *Earthquake Simulation Tests on Reducing Residual Displacements of Reinforced Concrete Bridges.* Junichi Sakai, Stephen A Mahin, and Andres Espinoza. December 2005.

PEER 2005/16 *Seismic Response and Reliability of Electrical Substation Equipment and Systems.* Junho Song, Armen Der Kiureghian, and Jerome L. Sackman. April 2006.

PEER 2005/15 *CPT-Based Probabilistic Assessment of Seismic Soil Liquefaction Initiation.* R. E. S. Moss, R. B. Seed, R. E. Kayen, J. P. Stewart, and A. Der Kiureghian. April 2006.

PEER 2005/14 *Workshop on Modeling of Nonlinear Cyclic Load-Deformation Behavior of Shallow Foundations.* Bruce L. Kutter, Geoffrey Martin, Tara Hutchinson, Chad Harden, Sivapalan Gajan, and Justin Phalen. March 2006.

PEER 2005/13 *Stochastic Characterization and Decision Bases under Time-Dependent Aftershock Risk in Performance-Based Earthquake Engineering*. Gee Liek Yeo and C. Allin Cornell. July 2005.

PEER 2005/12 *PEER Testbed Study on a Laboratory Building: Exercising Seismic Performance Assessment.* Mary C. Comerio, Editor. November 2005.

PEER 2005/11 *Van Nuys Hotel Building Testbed Report: Exercising Seismic Performance Assessment.* Helmut Krawinkler, Editor. October 2005.

PEER 2005/10 *First NEES/E-Defense Workshop on Collapse Simulation of Reinforced Concrete Building Structures.* September 2005.

PEER 2005/09 *Test Applications of Advanced Seismic Assessment Guidelines.* Joe Maffei, Karl Telleen, Danya Mohr, William Holmes, and Yuki Nakayama. August 2006.

PEER 2005/08 *Damage Accumulation in Lightly Confined Reinforced Concrete Bridge Columns.* R. Tyler Ranf, Jared M. Nelson, Zach Price, Marc O. Eberhard, and John F. Stanton. April 2006.

PEER 2005/07 *Experimental and Analytical Studies on the Seismic Response of Freestanding and Anchored Laboratory Equipment.* Dimitrios Konstantinidis and Nicos Makris. January 2005.

PEER 2005/06 *Global Collapse of Frame Structures under Seismic Excitations*. Luis F. Ibarra and Helmut Krawinkler. September 2005.

PEER 2005//05 *Performance Characterization of Bench- and Shelf-Mounted Equipment.* Samit Ray Chaudhuri and Tara C. Hutchinson. May 2006.

PEER 2005/04 *Numerical Modeling of the Nonlinear Cyclic Response of Shallow Foundations.* Chad Harden, Tara Hutchinson, Geoffrey R. Martin, and Bruce L. Kutter. August 2005.

PEER 2005/03 *A Taxonomy of Building Components for Performance-Based Earthquake Engineering.* Keith A. Porter. September 2005.

PEER 2005/02 *Fragility Basis for California Highway Overpass Bridge Seismic Decision Making*. Kevin R. Mackie and Božidar Stojadinović. June 2005.

PEER 2005/01 *Empirical Characterization of Site Conditions on Strong Ground Motion.* Jonathan P. Stewart, Yoojoong Choi, and Robert W. Graves. June 2005.

PEER 2004/09 *Electrical Substation Equipment Interaction: Experimental Rigid Conductor Studies.* Christopher Stearns and André Filiatrault. February 2005.

PEER 2004/08 *Seismic Qualification and Fragility Testing of Line Break 550-kV Disconnect Switches.* Shakhzod M. Takhirov, Gregory L. Fenves, and Eric Fujisaki. January 2005.

PEER 2004/07 *Ground Motions for Earthquake Simulator Qualification of Electrical Substation Equipment.* Shakhzod M. Takhirov, Gregory L. Fenves, Eric Fujisaki, and Don Clyde. January 2005.

PEER 2004/06 *Performance-Based Regulation and Regulatory Regimes.* Peter J. May and Chris Koski. September 2004.

PEER 2004/05 *Performance-Based Seismic Design Concepts and Implementation: Proceedings of an International Workshop.* Peter Fajfar and Helmut Krawinkler, Editors. September 2004.

PEER 2004/04 *Seismic Performance of an Instrumented Tilt-up Wall Building.* James C. Anderson and Vitelmo V. Bertero. July 2004.

PEER 2004/03 *Evaluation and Application of Concrete Tilt-up Assessment Methodologies.* Timothy Graf and James O. Malley. October 2004.

PEER 2004/02 *Analytical Investigations of New Methods for Reducing Residual Displacements of Reinforced Concrete Bridge Columns.* Junichi Sakai and Stephen A. Mahin. August 2004.

PEER 2004/01 *Seismic Performance of Masonry Buildings and Design Implications.* Kerri Anne Taeko Tokoro, James C. Anderson, and Vitelmo V. Bertero. February 2004.

PEER 2003/18 *Performance Models for Flexural Damage in Reinforced Concrete Columns.* Michael Berry and Marc Eberhard. August 2003.

PEER 2003/17 *Predicting Earthquake Damage in Older Reinforced Concrete Beam-Column Joints.* Catherine Pagni and Laura Lowes. October 2004.

PEER 2003/16 *Seismic Demands for Performance-Based Design of Bridges.* Kevin Mackie and Božidar Stojadinović. August 2003.

PEER 2003/15 *Seismic Demands for Nondeteriorating Frame Structures and Their Dependence on Ground Motions.* Ricardo Antonio Medina and Helmut Krawinkler. May 2004.

PEER 2003/14 *Finite Element Reliability and Sensitivity Methods for Performance-Based Earthquake Engineering*. Terje Haukaas and Armen Der Kiureghian. April 2004.

PEER 2003/13 *Effects of Connection Hysteretic Degradation on the Seismic Behavior of Steel Moment-Resisting Frames.* Janise E. Rodgers and Stephen A. Mahin. March 2004.

PEER 2003/12 *Implementation Manual for the Seismic Protection of Laboratory Contents: Format and Case Studies*. William T. Holmes and Mary C. Comerio. October 2003.

PEER 2003/11 *Fifth U.S.-Japan Workshop on Performance-Based Earthquake Engineering Methodology for Reinforced Concrete Building Structures*. February 2004.

PEER 2003/10 *A Beam-Column Joint Model for Simulating the Earthquake Response of Reinforced Concrete Frames.* Laura N. Lowes, Nilanjan Mitra, and Arash Altoontash. February 2004.

PEER 2003/09 *Sequencing Repairs after an Earthquake: An Economic Approach.* Marco Casari and Simon J. Wilkie. April 2004.

PEER 2003/08 *A Technical Framework for Probability-Based Demand and Capacity Factor Design (DCFD) Seismic Formats*. Fatemeh Jalayer and C. Allin Cornell. November 2003.

PEER 2003/07 *Uncertainty Specification and Propagation for Loss Estimation Using FOSM Methods.* Jack W. Baker and C. Allin Cornell. September 2003.

PEER 2003/06 *Performance of Circular Reinforced Concrete Bridge Columns under Bidirectional Earthquake Loading*. Mahmoud M. Hachem, Stephen A. Mahin, and Jack P. Moehle. February 2003.

PEER 2003/05 *Response Assessment for Building-Specific Loss Estimation.* Eduardo Miranda and Shahram Taghavi. September 2003.

PEER 2003/04 *Experimental Assessment of Columns with Short Lap Splices Subjected to Cyclic Loads*. Murat Melek, John W. Wallace, and Joel Conte. April 2003.

PEER 2003/03 *Probabilistic Response Assessment for Building-Specific Loss Estimation*. Eduardo Miranda and Hesameddin Aslani. September 2003.

PEER 2003/02 *Software Framework for Collaborative Development of Nonlinear Dynamic Analysis Program*. Jun Peng and Kincho H. Law. September 2003.

PEER 2003/01 *Shake Table Tests and Analytical Studies on the Gravity Load Collapse of Reinforced Concrete Frames.* Kenneth John Elwood and Jack P. Moehle. November 2003.

PEER 2002/24 *Performance of Beam to Column Bridge Joints Subjected to a Large Velocity Pulse.* Natalie Gibson, André Filiatrault, and Scott A. Ashford. April 2002.

PEER 2002/23 *Effects of Large Velocity Pulses on Reinforced Concrete Bridge Columns.* Greg L. Orozco and Scott A. Ashford. April 2002.

PEER 2002/22 *Characterization of Large Velocity Pulses for Laboratory Testing.* Kenneth E. Cox and Scott A. Ashford. April 2002.

PEER 2002/21 *Fourth U.S.-Japan Workshop on Performance-Based Earthquake Engineering Methodology for Reinforced Concrete Building Structures*. December 2002.

PEER 2002/20 *Barriers to Adoption and Implementation of PBEE Innovations.* Peter J. May. August 2002.

PEER 2002/19 *Economic-Engineered Integrated Models for Earthquakes: Socioeconomic Impacts.* Peter Gordon, James E. Moore II, and Harry W. Richardson. July 2002.

PEER 2002/18 *Assessment of Reinforced Concrete Building Exterior Joints with Substandard Details.* Chris P. Pantelides, Jon Hansen, Justin Nadauld, and Lawrence D. Reaveley. May 2002.

PEER 2002/17 *Structural Characterization and Seismic Response Analysis of a Highway Overcrossing Equipped with Elastomeric Bearings and Fluid Dampers: A Case Study*. Nicos Makris and Jian Zhang. November 2002.

PEER 2002/16 *Estimation of Uncertainty in Geotechnical Properties for Performance-Based Earthquake Engineering*. Allen L. Jones, Steven L. Kramer, and Pedro Arduino. December 2002.

PEER 2002/15 *Seismic Behavior of Bridge Columns Subjected to Various Loading Patterns*. Asadollah Esmaeily-Gh. and Yan Xiao. December 2002.

PEER 2002/14 *Inelastic Seismic Response of Extended Pile Shaft Supported Bridge Structures.* T.C. Hutchinson, R.W. Boulanger, Y.H. Chai, and I.M. Idriss. December 2002.

PEER 2002/13 *Probabilistic Models and Fragility Estimates for Bridge Components and Systems.* Paolo Gardoni, Armen Der Kiureghian, and Khalid M. Mosalam. June 2002.

PEER 2002/12 *Effects of Fault Dip and Slip Rake on Near-Source Ground Motions: Why Chi-Chi Was a Relatively Mild M7.6 Earthquake*. Brad T. Aagaard, John F. Hall, and Thomas H. Heaton. December 2002.

PEER 2002/11 *Analytical and Experimental Study of Fiber-Reinforced Strip Isolators*. James M. Kelly and Shakhzod M. Takhirov. September 2002.

PEER 2002/10 *Centrifuge Modeling of Settlement and Lateral Spreading with Comparisons to Numerical Analyses***.** Sivapalan Gajan and Bruce L. Kutter. January 2003.

PEER 2002/09 *Documentation and Analysis of Field Case Histories of Seismic Compression during the 1994 Northridge, California, Earthquake*. Jonathan P. Stewart, Patrick M. Smith, Daniel H. Whang, and Jonathan D. Bray. October 2002.

PEER 2002/08 *Component Testing, Stability Analysis and Characterization of Buckling-Restrained Unbonded Braces™*. Cameron Black, Nicos Makris, and Ian Aiken. September 2002.

PEER 2002/07 *Seismic Performance of Pile-Wharf Connections.* Charles W. Roeder, Robert Graff, Jennifer Soderstrom, and Jun Han Yoo. December 2001.

PEER 2002/06 *The Use of Benefit-Cost Analysis for Evaluation of Performance-Based Earthquake Engineering Decisions*. Richard O. Zerbe and Anthony Falit-Baiamonte. September 2001.

PEER 2002/05 *Guidelines, Specifications, and Seismic Performance Characterization of Nonstructural Building Components and Equipment*. André Filiatrault, Constantin Christopoulos, and Christopher Stearns. September 2001.

PEER 2002/04 *Consortium of Organizations for Strong-Motion Observation Systems and the Pacific Earthquake Engineering Research Center Lifelines Program: Invited Workshop on Archiving and Web Dissemination of Geotechnical Data, 4–5 October 2001*. September 2002.

PEER 2002/03 *Investigation of Sensitivity of Building Loss Estimates to Major Uncertain Variables for the Van Nuys Testbed*. Keith A. Porter, James L. Beck, and Rustem V. Shaikhutdinov. August 2002.

PEER 2002/02 *The Third U.S.-Japan Workshop on Performance-Based Earthquake Engineering Methodology for Reinforced Concrete Building Structures*. July 2002.

PEER 2002/01 *Nonstructural Loss Estimation: The UC Berkeley Case Study.* Mary C. Comerio and John C. Stallmeyer. December 2001.

PEER 2001/16 *Statistics of SDF-System Estimate of Roof Displacement for Pushover Analysis of Buildings.* Anil K. Chopra, Rakesh K. Goel, and Chatpan Chintanapakdee. December 2001.

PEER 2001/15 *Damage to Bridges during the 2001 Nisqually Earthquake.* R. Tyler Ranf, Marc O. Eberhard, and Michael P. Berry. November 2001.

PEER 2001/14 *Rocking Response of Equipment Anchored to a Base Foundation.* Nicos Makris and Cameron J. Black. September 2001.

PEER 2001/13 *Modeling Soil Liquefaction Hazards for Performance-Based Earthquake Engineering*. Steven L. Kramer and Ahmed-W. Elgamal. February 2001.

PEER 2001/12 *Development of Geotechnical Capabilities in OpenSees*. Boris Jeremić. September 2001.

PEER 2001/11 *Analytical and Experimental Study of Fiber-Reinforced Elastomeric Isolators*. James M. Kelly and Shakhzod M. Takhirov. September 2001.

PEER 2001/10 *Amplification Factors for Spectral Acceleration in Active Regions.* Jonathan P. Stewart, Andrew H. Liu, Yoojoong Choi, and Mehmet B. Baturay. December 2001.

PEER 2001/09 *Ground Motion Evaluation Procedures for Performance-Based Design.* Jonathan P. Stewart, Shyh-Jeng Chiou, Jonathan D. Bray, Robert W. Graves, Paul G. Somerville, and Norman A. Abrahamson. September 2001.

PEER 2001/08 *Experimental and Computational Evaluation of Reinforced Concrete Bridge Beam-Column Connections for Seismic Performance*. Clay J. Naito, Jack P. Moehle, and Khalid M. Mosalam. November 2001.

PEER 2001/07 *The Rocking Spectrum and the Shortcomings of Design Guidelines.* Nicos Makris and Dimitrios Konstantinidis. August 2001.

PEER 2001/06 *Development of an Electrical Substation Equipment Performance Database for Evaluation of Equipment Fragilities*. Thalia Agnanos. April 1999.

PEER 2001/05 *Stiffness Analysis of Fiber-Reinforced Elastomeric Isolators.* Hsiang-Chuan Tsai and James M. Kelly. May 2001.

PEER 2001/04 *Organizational and Societal Considerations for Performance-Based Earthquake Engineering*. Peter J. May. April 2001.

PEER 2001/03 *A Modal Pushover Analysis Procedure to Estimate Seismic Demands for Buildings: Theory and Preliminary Evaluation*. Anil K. Chopra and Rakesh K. Goel. January 2001.

PEER 2001/02 *Seismic Response Analysis of Highway Overcrossings Including Soil-Structure Interaction*. Jian Zhang and Nicos Makris. March 2001.

PEER 2001/01 *Experimental Study of Large Seismic Steel Beam-to-Column Connections.* Egor P. Popov and Shakhzod M. Takhirov. November 2000.

PEER 2000/10 *The Second U.S.-Japan Workshop on Performance-Based Earthquake Engineering Methodology for Reinforced Concrete Building Structures*. March 2000.

PEER 2000/09 *Structural Engineering Reconnaissance of the August 17, 1999 Earthquake: Kocaeli (Izmit), Turkey.* Halil Sezen, Kenneth J. Elwood, Andrew S. Whittaker, Khalid Mosalam, John J. Wallace, and John F. Stanton. December 2000.

PEER 2000/08 *Behavior of Reinforced Concrete Bridge Columns Having Varying Aspect Ratios and Varying Lengths of Confinement*. Anthony J. Calderone, Dawn E. Lehman, and Jack P. Moehle. January 2001.

PEER 2000/07 *Cover-Plate and Flange-Plate Reinforced Steel Moment-Resisting Connections.* Taejin Kim, Andrew S. Whittaker, Amir S. Gilani, Vitelmo V. Bertero, and Shakhzod M. Takhirov. September 2000.

PEER 2000/06 *Seismic Evaluation and Analysis of 230-kV Disconnect Switches.* Amir S. J. Gilani, Andrew S. Whittaker, Gregory L. Fenves, Chun-Hao Chen, Henry Ho, and Eric Fujisaki. July 2000.

PEER 2000/05 *Performance-Based Evaluation of Exterior Reinforced Concrete Building Joints for Seismic Excitation*. Chandra Clyde, Chris P. Pantelides, and Lawrence D. Reaveley. July 2000.

PEER 2000/04 *An Evaluation of Seismic Energy Demand: An Attenuation Approach*. Chung-Che Chou and Chia-Ming Uang. July 1999.

PEER 2000/03 *Framing Earthquake Retrofitting Decisions: The Case of Hillside Homes in Los Angeles.* Detlof von Winterfeldt, Nels Roselund, and Alicia Kitsuse. March 2000.

PEER 2000/02 *U.S.-Japan Workshop on the Effects of Near-Field Earthquake Shaking*. Andrew Whittaker, Editor. July 2000.

PEER 2000/01 *Further Studies on Seismic Interaction in Interconnected Electrical Substation Equipment.* Armen Der Kiureghian, Kee-Jeung Hong, and Jerome L. Sackman. November 1999.

PEER 1999/14 *Seismic Evaluation and Retrofit of 230-kV Porcelain Transformer Bushings*. Amir S. Gilani, Andrew S. Whittaker, Gregory L. Fenves, and Eric Fujisaki. December 1999.

PEER 1999/13 *Building Vulnerability Studies: Modeling and Evaluation of Tilt-up and Steel Reinforced Concrete Buildings*. John W. Wallace, Jonathan P. Stewart, and Andrew S. Whittaker, Editors. December 1999.

PEER 1999/12 *Rehabilitation of Nonductile RC Frame Building Using Encasement Plates and Energy-Dissipating Devices*. Mehrdad Sasani, Vitelmo V. Bertero, James C. Anderson. December 1999.

PEER 1999/11 *Performance Evaluation Database for Concrete Bridge Components and Systems under Simulated Seismic Loads*. Yael D. Hose and Frieder Seible. November 1999.

PEER 1999/10 *U.S.-Japan Workshop on Performance-Based Earthquake Engineering Methodology for Reinforced Concrete Building Structures*. December 1999.

PEER 1999/09 *Performance Improvement of Long Period Building Structures Subjected to Severe Pulse-Type Ground Motions*. James C. Anderson, Vitelmo V. Bertero, and Raul Bertero. October 1999.

PEER 1999/08 *Envelopes for Seismic Response Vectors.* Charles Menun and Armen Der Kiureghian. July 1999.

PEER 1999/07 *Documentation of Strengths and Weaknesses of Current Computer Analysis Methods for Seismic Performance of Reinforced Concrete Members*. William F. Cofer. November 1999.

PEER 1999/06 *Rocking Response and Overturning of Anchored Equipment under Seismic Excitations*. Nicos Makris and Jian Zhang. November 1999.

PEER 1999/05 *Seismic Evaluation of 550 kV Porcelain Transformer Bushings.* Amir S. Gilani, Andrew S. Whittaker, Gregory L. Fenves, and Eric Fujisaki. October 1999.

PEER 1999/04 *Adoption and Enforcement of Earthquake Risk-Reduction Measures.* Peter J. May, Raymond J. Burby, T. Jens Feeley, and Robert Wood. August 1999.

PEER 1999/03 *Task 3 Characterization of Site Response General Site Categories*. Adrian Rodriguez-Marek, Jonathan D. Bray and Norman Abrahamson. February 1999.

PEER 1999/02 *Capacity-Demand-Diagram Methods for Estimating Seismic Deformation of Inelastic Structures: SDF Systems*. Anil K. Chopra and Rakesh Goel. April 1999.

PEER 1999/01 *Interaction in Interconnected Electrical Substation Equipment Subjected to Earthquake Ground Motions*. Armen Der Kiureghian, Jerome L. Sackman, and Kee-Jeung Hong. February 1999.

PEER 1998/08 *Behavior and Failure Analysis of a Multiple-Frame Highway Bridge in the 1994 Northridge Earthquake*. Gregory L. Fenves and Michael Ellery. December 1998.

PEER 1998/07 *Empirical Evaluation of Inertial Soil-Structure Interaction Effects.* Jonathan P. Stewart, Raymond B. Seed, and Gregory L. Fenves. November 1998.

PEER 1998/06 *Effect of Damping Mechanisms on the Response of Seismic Isolated Structures*. Nicos Makris and Shih-Po Chang. November 1998.

PEER 1998/05 *Rocking Response and Overturning of Equipment under Horizontal Pulse-Type Motions*. Nicos Makris and Yiannis Roussos. October 1998.

PEER 1998/04 *Pacific Earthquake Engineering Research Invitational Workshop Proceedings, May 14–15, 1998: Defining the Links between Planning, Policy Analysis, Economics and Earthquake Engineering*. Mary Comerio and Peter Gordon. September 1998.

PEER 1998/03 *Repair/Upgrade Procedures for Welded Beam to Column Connections*. James C. Anderson and Xiaojing Duan. May 1998.

PEER 1998/02 *Seismic Evaluation of 196 kV Porcelain Transformer Bushings.* Amir S. Gilani, Juan W. Chavez, Gregory L. Fenves, and Andrew S. Whittaker. May 1998.

PEER 1998/01 *Seismic Performance of Well-Confined Concrete Bridge Columns.* Dawn E. Lehman and Jack P. Moehle. December 2000.

PEER REPORTS: ONE HUNDRED SERIES

The following PEER reports are available by Internet only at http://peer.berkeley.edu/publications/peer_reports_complete.html.

PEER 2012/103 *Performance-Based Seismic Demand Assessment of Concentrically Braced Steel Frame Buildings.* Chui-Hsin Chen and Stephen A. Mahin. December 2012.

PEER 2012/102 *Procedure to Restart an Interrupted Hybrid Simulation: Addendum to PEER Report 2010/103.* Vesna Terzic and Bozidar Stojadinovic. October 2012.

PEER 2012/101 *Mechanics of Fiber Reinforced Bearings.* James M. Kelly and Andrea Calabrese. February 2012.

PEER 2011/107 *Nonlinear Site Response and Seismic Compression at Vertical Array Strongly Shaken by 2007 Niigata-ken Chuetsu-oki Earthquake.* Eric Yee, Jonathan P. Stewart, and Kohji Tokimatsu. December 2011.

PEER 2011/106 *Self Compacting Hybrid Fiber Reinforced Concrete Composites for Bridge Columns.* Pardeep Kumar, Gabriel Jen, William Trono, Marios Panagiotou, and Claudia Ostertag. September 2011.

PEER 2011/105 *Stochastic Dynamic Analysis of Bridges Subjected to Spacially Varying Ground Motions.* Katerina Konakli and Armen Der Kiureghian. August 2011.

PEER 2011/104 *Design and Instrumentation of the 2010 E-Defense Four-Story Reinforced Concrete and Post-Tensioned Concrete Buildings.* Takuya Nagae, Kenichi Tahara, Taizo Matsumori, Hitoshi Shiohara, Toshimi Kabeyasawa, Susumu Kono, Minehiro Nishiyama (Japanese Research Team) and John Wallace, Wassim Ghannoum, Jack Moehle, Richard Sause, Wesley Keller, Zeynep Tuna (U.S. Research Team). June 2011.

PEER 2011/103 *In-Situ Monitoring of the Force Output of Fluid Dampers: Experimental Investigation.* Dimitrios Konstantinidis, James M. Kelly, and Nicos Makris. April 2011.

PEER 2011/102 *Ground-Motion Prediction Equations 1964–2010.* John Douglas. April 2011.

PEER 2011/101 *Report of the Eighth Planning Meeting of NEES/E-Defense Collaborative Research on Earthquake Engineering.* Convened by the Hyogo Earthquake Engineering Research Center (NIED), NEES Consortium, Inc. February 2011.

PEER 2010/111 *Modeling and Acceptance Criteria for Seismic Design and Analysis of Tall Buildings.* Task 7 Report for the Tall Buildings Initiative - Published jointly by the Applied Technology Council. October 2010.

PEER 2010/110 *Seismic Performance Assessment and Probabilistic Repair Cost Analysis of Precast Concrete Cladding Systems for Multistory Buildlings.* Jeffrey P. Hunt and Božidar Stojadinovic. November 2010.

PEER 2010/109 *Report of the Seventh Joint Planning Meeting of NEES/E-Defense Collaboration on Earthquake Engineering. Held at the E-Defense, Miki, and Shin-Kobe, Japan, September 18–19, 2009.* August 2010.

PEER 2010/108 *Probabilistic Tsunami Hazard in California.* Hong Kie Thio, Paul Somerville, and Jascha Polet, preparers. October 2010.

PEER 2010/107 *Performance and Reliability of Exposed Column Base Plate Connections for Steel Moment-Resisting Frames.* Ady Aviram, Božidar Stojadinovic, and Armen Der Kiureghian. August 2010.

PEER 2010/106 *Verification of Probabilistic Seismic Hazard Analysis Computer Programs.* Patricia Thomas, Ivan Wong, and Norman Abrahamson. May 2010.

PEER 2010/105 *Structural Engineering Reconnaissance of the April 6, 2009, Abruzzo, Italy, Earthquake, and Lessons Learned.* M. Selim Günay and Khalid M. Mosalam. April 2010.

PEER 2010/104 *Simulating the Inelastic Seismic Behavior of Steel Braced Frames, Including the Effects of Low-Cycle Fatigue.* Yuli Huang and Stephen A. Mahin. April 2010.

PEER 2010/103 *Post-Earthquake Traffic Capacity of Modern Bridges in California.* Vesna Terzic and Božidar Stojadinović. March 2010.

PEER 2010/102 *Analysis of Cumulative Absolute Velocity (CAV) and JMA Instrumental Seismic Intensity (I_{JMA}) Using the PEER–NGA Strong Motion Database.* Kenneth W. Campbell and Yousef Bozorgnia. February 2010.

PEER 2010/101 *Rocking Response of Bridges on Shallow Foundations.* Jose A. Ugalde, Bruce L. Kutter, and Boris Jeremic. April 2010.

PEER 2009/109 *Simulation and Performance-Based Earthquake Engineering Assessment of Self-Centering Post-Tensioned Concrete Bridge Systems.* Won K. Lee and Sarah L. Billington. December 2009.

PEER 2009/108 *PEER Lifelines Geotechnical Virtual Data Center.* J. Carl Stepp, Daniel J. Ponti, Loren L. Turner, Jennifer N. Swift, Sean Devlin, Yang Zhu, Jean Benoit, and John Bobbitt. September 2009.

PEER 2009/107 *Experimental and Computational Evaluation of Current and Innovative In-Span Hinge Details in Reinforced Concrete Box-Girder Bridges: Part 2: Post-Test Analysis and Design Recommendations.* Matias A. Hube and Khalid M. Mosalam. December 2009.

PEER 2009/106 *Shear Strength Models of Exterior Beam-Column Joints without Transverse Reinforcement.* Sangjoon Park and Khalid M. Mosalam. November 2009.

PEER 2009/105 *Reduced Uncertainty of Ground Motion Prediction Equations through Bayesian Variance Analysis.* Robb Eric S. Moss. November 2009.

PEER 2009/104 *Advanced Implementation of Hybrid Simulation.* Andreas H. Schellenberg, Stephen A. Mahin, Gregory L. Fenves. November 2009.

PEER 2009/103 *Performance Evaluation of Innovative Steel Braced Frames.* T. Y. Yang, Jack P. Moehle, and Božidar Stojadinovic. August 2009.

PEER 2009/102 *Reinvestigation of Liquefaction and Nonliquefaction Case Histories from the 1976 Tangshan Earthquake.* Robb Eric Moss, Robert E. Kayen, Liyuan Tong, Songyu Liu, Guojun Cai, and Jiaer Wu. August 2009.

PEER 2009/101 *Report of the First Joint Planning Meeting for the Second Phase of NEES/E-Defense Collaborative Research on Earthquake Engineering.* Stephen A. Mahin et al. July 2009.

PEER 2008/104 *Experimental and Analytical Study of the Seismic Performance of Retaining Structures.* Linda Al Atik and Nicholas Sitar. January 2009.

PEER 2008/103 *Experimental and Computational Evaluation of Current and Innovative In-Span Hinge Details in Reinforced Concrete Box-Girder Bridges. Part 1: Experimental Findings and Pre-Test Analysis.* Matias A. Hube and Khalid M. Mosalam. January 2009.

PEER 2008/102 *Modeling of Unreinforced Masonry Infill Walls Considering In-Plane and Out-of-Plane Interaction.* Stephen Kadysiewski and Khalid M. Mosalam. January 2009.

PEER 2008/101 *Seismic Performance Objectives for Tall Buildings.* William T. Holmes, Charles Kircher, William Petak, and Nabih Youssef. August 2008.

PEER 2007/101 *Generalized Hybrid Simulation Framework for Structural Systems Subjected to Seismic Loading*. Tarek Elkhoraibi and Khalid M. Mosalam. July 2007.

PEER 2007/100 *Seismic Evaluation of Reinforced Concrete Buildings Including Effects of Masonry Infill Walls*. Alidad Hashemi and Khalid M. Mosalam. July 2007.

The Pacifi c Earthquake Engineering Research Center (PEER) is a multi-institutional research and education center with headquarters at the University of California, Berkeley. Investigators from over 20 universities, several consulting companies, and researchers at various state and federal government agencies contribute to research programs focused on performance-based earthquake engineering.

These research programs aim to identify and reduce the risks from major earthquakes to life safety and to the economy by including research in a wide variety of disciplines including structural and geotechnical engineering, geology/seismology, lifelines, transportation, architecture, economics, risk management, and public policy.

PEER is supported by federal, state, local, and regional agencies, together with industry partners.

PEER Core Institutions

University of California, Berkeley (Lead Institution)
California Institute of Technology
Oregon State University
Stanford University
University of California, Davis
University of California, Irvine
University of California, Los Angeles
University of California, San Diego
University of Nevada, Reno
University of Southern California
University of Washington

PEER reports can be ordered at https://peer.berkeley.edu/peer-reports or by contacting

Pacific Earthquake Engineering Research Center
University of California, Berkeley
325 Davis Hall, Mail Code 1792
Berkeley, CA 94720-1792
Tel: 510-642-3437
Email: peer_center@berkeley.edu

ISSN 2770-8314
https://doi.org/10.55461/QFKL9711