

Methods comparison for neural network-based structural damage recognition and classification

Chang Che^{1,*}, *Junchi Tian*²

¹ Mechanical Engineering, The George Washington University, DC, 20052 U.S

² Data Science, The George Washington University, DC, 20052 U.S

* cche57@gwmail.gwu.edu

Abstract. Machine learning has brought significant advancements to the field of structural health monitoring, providing flexible and efficient solutions for detecting both local and global damage in various infrastructures. Local damage detection focuses on identifying issues such as cracks and spalling in specific areas of concrete structures, including bridges, highways, and tunnels. Techniques like artificial neural networks (ANNs) and deep neural networks (DNNs) have proven effective in surface defect recognition, demonstrating their versatility across different structural environments. Additionally, cost-effective methods leveraging devices such as smartphones have been explored for rapid road integrity assessments, offering practical and affordable solutions. On a larger scale, global damage detection involves classifying structural collapse modes and types of damage, utilizing feature extraction and deep learning models to improve the accuracy of identifying large-scale failures. These advancements highlight the growing importance of machine learning and computer vision in enhancing the resilience and real-time monitoring of critical infrastructure systems.

Keywords: machine learning, structural health monitoring, local damage detection, computer vision, crack detection

1. Introduction

Structural health monitoring (SHM) and rapid damage assessment following natural hazards have become critical focal points in civil engineering research and practice (e.g., [1], [2]). Within this domain, image-based damage recognition is gaining prominence, particularly in vision-based SHM and structural reconnaissance, which have traditionally relied on manual visual inspection and expert judgment. However, the integration of cutting-edge deep learning (DL) technologies offers significant potential to automate and enhance the accuracy of these assessments.

This paper focuses on classifying various structural damage types. Specifically, the paper trying to detect the pictures with no damage, flexural damage, shear damage, and combined damage, respectively. Furthermore, the model is trying to classify the picture with various spalling conditions and collapse modes. These tasks have traditionally been performed through human expertise, but DL models present an opportunity to streamline and improve these evaluations.

Our study leverages a dataset provided by PEER Hub ImageNet [18], which contains over 10,000 images depicting various structural types and damage states. We preprocess and standardize these images, utilizing multiple convolutional neural network (CNN) architectures to predict structural damage characteristics. This method offers significant potential to enhance the field of Structural Health Monitoring (SHM) by improving both the efficiency and accuracy of damage assessments. Additionally, our research seeks to explore performance differences across these models, providing a comprehensive evaluation of their effectiveness in identifying structural damage.

2. Related work

Machine learning applications in civil engineering have been developed over several years (e.g., [3]-[14]). Machine learning for structural health monitoring is one of the important areas. Research in this field is typically categorized into local and global damage detection. Local damage refers to damage affecting specific parts of a building structure, with detection methods focusing on identifying cracks or spalling in concrete.

[15] applied two artificial neural network (ANN) based algorithms—backpropagation (BP) and self-organizing maps (SOM)—for recognizing surface defects in images of bridges. Crack detection techniques are not limited to bridges and buildings but can also be used to assess highways.

[16] investigated crack detection on roads, demonstrating the potential of low-cost tools such as DSLR cameras and smartphones. Their experiments showed that smartphones offer a practical and affordable way to quickly assess road conditions. Similarly, [17] employed deep neural networks to detect linear cracks in tunnels. These studies highlight the versatility of machine learning and computer vision in structural health monitoring.

Global damage refers to damage visible from the exterior of structures, encompassing the classification of collapse modes and damage types. [18] utilized feature extraction and fine-tuning techniques to optimize model parameters for global damage classification, while [19] employed region-based deep learning to detect multiple types of damage on concrete surfaces. Furthermore, [20] and [21] utilized computer vision to locate both local and global structural damages using smartphones and drones.

3. Dataset and features

Our application focuses on three distinct tasks aimed at improving structural damage assessment: classifying damage types, evaluating spalling conditions, and identifying collapse modes. Each of these tasks is trained and tested using independent datasets, each consisting of approximately 2,500 images. While this dataset size is relatively modest for deep learning applications, it remains suitable for demonstrating proof-of-concept in our models. To optimize training and ensure robust evaluation, the datasets are split into three subsets: 2,000 images are allocated for training, 300 for validation, and 200 for testing. All images are standardized to a 224x224x3 RGB format to ensure consistency across the tasks.

Prior to input into the convolutional neural network (ConvNet) models, the pixel values of the images, initially ranging from 0 to 255, are rescaled to fall within a normalized 0-1 range. This standardization facilitates more stable model training and enhances generalization. The output labels differ across the tasks: for the damage type classification, there are four distinct classes, spalling condition evaluation is divided into two classes, and collapse mode prediction includes three classes. Sample images representing each task are displayed below to illustrate the dataset.

In terms of feature extraction, our ConvNet models utilize the pixel values of the images as input. These pixel values pass through multiple convolutional layers, where the model learns hierarchical patterns and features critical for each task. After several convolutional layers and subsequent processing, the model predicts the corresponding label based on the task—whether it is determining the type of damage, the spalling condition, or the collapse mode. Despite the relatively small dataset size, our approach leverages powerful deep learning techniques to ensure accurate and reliable predictions, with potential applications in real-time structural health monitoring and disaster response systems.



Figure 1. Damage type classification. (a) no damage (b) flexural damage (c) shear damage (d) combined damage

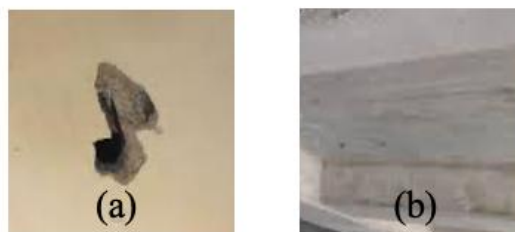


Figure 2. Spalling condition classification. (a) spalling (b) no spalling



Figure 3. Collapse mode classification. (a) Non-collapse (b) partial collapse (c) global collapse

4. Method

In order to predict the label of structural damage, different models are implemented. The three baseline models are implemented including AlexNet, VGG-16 and ResNet50.

4.1. Base models

The structure of AlexNet, as described in [22], consists of a deep convolutional neural network architecture designed to handle large-scale image classification tasks. Initially, the input image passes through a large convolutional layer, which applies multiple filters to extract low-level features such as edges, textures, and color gradients from the image. This convolutional layer is followed by a max pooling layer, which reduces the spatial dimensions of the feature maps, thereby decreasing the computational complexity and helping to prevent overfitting by focusing on the most prominent features. This process is repeated multiple times, allowing the network to progressively capture more abstract and complex features from the input image. Afterward, the image passes through three additional convolutional layers, which are smaller in size but crucial for refining the learned features and detecting high-level patterns such as shapes or objects. These layers are again followed by a final max pooling layer to further downsample the data.

At this stage, the multidimensional feature maps are flattened into a one-dimensional vector, which is then fed into a series of fully connected layers. These layers perform the final classification by combining the high-level features learned during the convolutional stages. The output from the fully connected layers is used to predict the class of the input image. Throughout the network, Rectified Linear Unit (ReLU) activation functions are applied after each convolutional and fully connected layer to introduce non-linearity, enabling the network to learn and model complex relationships in the data.

Compared to AlexNet, the most notable difference in VGG-16 lies in its architectural choices. While AlexNet employs varying kernel sizes, VGG-16 standardizes its convolutional layers by exclusively using 3x3 kernels across all layers. This uniform kernel size enhances the model's ability to capture finer details in the spatial dimensions of the input data while maintaining simplicity in design. Additionally, VGG-16 incorporates 2x2 max pooling layers, which effectively reduce the spatial dimensions of the feature maps while retaining the most prominent features, contributing to more efficient computations. Another key distinction is the significant increase in the depth of the network. This deeper architecture allows VGG-16 to learn more complex patterns and hierarchical representations, making it more powerful in handling tasks that require deeper feature extraction, such as image classification and object recognition. The increased number of layers enables the network to model intricate details and abstractions, but also introduces challenges such as increased computational cost and potential vanishing gradient issues during training. The overall architecture of VGG-16 is shown below, showcasing how the deep network's consistent use of small kernels and pooling layers contributes to its advanced performance in comparison to earlier models like AlexNet.

ResNet is a deeper neural network architecture compared to earlier models like AlexNet and VGG-16. While increasing the depth of a neural network can improve its ability to capture complex patterns, it also introduces a significant challenge known as the degradation problem. As the number of layers grows, the gradients during backpropagation may either explode or vanish, which hampers the model's ability to learn effectively. To mitigate this issue, ResNet introduces a novel solution known as the identity shortcut connection, which allows the model to "skip" one or more layers, essentially passing the input from one layer directly to a later layer without modification. This mechanism, referred to as residual learning, enables the model to preserve essential information through these shortcut connections while allowing deeper layers to focus on learning finer details. By incorporating residual learning, ResNet significantly increases the depth of CNNs, with architectures like ResNet-50, ResNet-101, and even deeper variants, enabling the network to have more layers without falling into the trap of gradient degradation. This innovation has not only improved performance on a wide range of tasks but also opened the door to more advanced, deeper models in the field of deep learning.

4.2. Base results

The baseline results for the predictive models are summarized in Table 1, showing the performance metrics across three key tasks: spalling condition classification, collapse mode detection, and damage type identification. The training accuracy for each task is notably high, with 94.5%, 97.5%, and 94.2% respectively, indicating that the models are learning effectively from the training data. However, there is a significant drop in test accuracy, particularly for the collapse mode and damage type tasks, where the test accuracy falls to 58.1% and 57.7%, respectively, suggesting a potential overfitting issue. This is further illustrated in the training curves (Figure 4) and error trends (Figure 5), where the gap between training and test accuracy is evident, especially in the damage type classification task.

Table 1. Baseline result

Task	Classes	Training Accuracy	Test Accuracy
Spalling Condition	2	94.5%	65.6%
Collapse Mode	3	97.5%	58.1%
Damage Type	4	94.2%	57.7%

The baseline models were trained without fine-tuning or the use of pre-trained parameters to assess the performance of naive models. This approach highlights the limitations of the current setup and points toward the need for improvement strategies. To address the overfitting and enhance generalization to unseen data, several techniques could be applied, including fine-tuning the models with pre-trained weights and incorporating regularization methods such as dropout or weight decay. Moreover, exploring a wider range of model architectures and hyperparameter configurations could help optimize performance. Moving forward, these strategies will be employed to refine the models and achieve better balance between training and test accuracy across all tasks.

4.3. Model improvements

Fine-tuning is employed to enhance the baseline model's performance. This process begins by initializing the model with pre-trained parameters from ImageNet, which are mapped to the corresponding layers of the new model. To accelerate training and extract low-level features such as pixel gradients, texture, and color, the parameters in the lower-level convolutional layers are frozen, meaning they are not updated during training. In contrast, the parameters in the higher-level convolutional blocks and fully connected layers are unfrozen, allowing them to be updated during training to adapt to the new task. This hierarchical approach ensures that the model retains its ability to recognize basic image features while also learning specific patterns related to the new dataset.

In addition to fine-tuning the model, several hyperparameters are optimized to further improve model performance. This tuning process includes adjusting critical elements such as the dropout rate, L1/L2 regularization strength, learning rate, and the number of frozen/unfrozen convolutional blocks. These adjustments are tested across four alternative architectures: VGG19, MobileNet, MobileNetV2, and InceptionV3, to compare performance. The hyperparameter configurations explored include different optimization algorithms such as SGD, RMSProp, Adam, and Momentum, which play a crucial role in improving convergence during training.

The impact of these fine-tuning and hyperparameter adjustments is evident in the model's performance, as reflected in the training curves and prediction results, as shown in Figure 6 and 7. The baseline models, such as AlexNet, VGG16, and ResNet50, exhibit high variance and overfitting, with a large gap between training and test accuracy. By contrast, the improved models, following the fine-tuning process, show significantly reduced overfitting. Dropout and L2 regularization, in particular, help to stabilize the models by introducing a regularization effect, preventing over-reliance on specific features. Despite these improvements, fine-tuning does introduce some bias, as reflected by a slight drop in training set accuracy, but this trade-off leads to improved generalization on the test set.

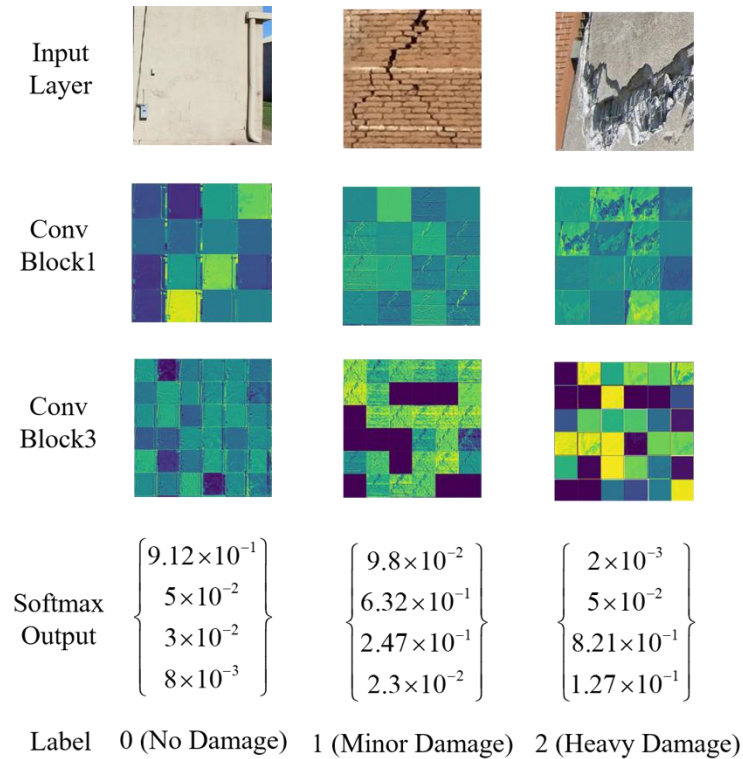
The summary table demonstrates that MobileNetV2 performs the best overall, achieving the highest accuracy across all tasks. Other models, such as VGG19 and InceptionV3, also show marked improvements compared to their baseline versions, though MobileNetV2's balance of speed, accuracy, and generalization makes it the top-performing model in this analysis. The results illustrate the effectiveness of fine-tuning and hyperparameter optimization in significantly boosting the performance of deep learning models for image-based tasks.

Table 2. Identification result summary

	Task	Spalling Condition	Collapse Model	Damage Type
Baseline Model	AlexNet	59.61%	53.30%	56.20%
	VGG16	65.63%	58.10%	57.70%
	ResNet50	57.94%	51.20%	52.10%
Improved Model	VGG16	84.76%	67.40%	64.11%
	VGG19	84.34%	66.50%	62.34%
	ResNet50	79.02%	62.20%	58.42%
	MobileNet	84.75%	68.20%	61.90%
	MobileNetV2	85.45%	69.11%	64.50%
	InceptionV3	80.50%	63.20%	57.14%

4.4. Intermediate output

While the input and output of a neural network are often well-understood, the outputs at intermediate layers receive less attention, despite being equally important. In an ideal convolutional neural network (CNN), the initial convolutional blocks are tasked with extracting low-level features, such as basic shapes, edges, and textures. For instance, in the context of damage type classification, these intermediate layers can identify structural patterns that indicate different damage categories.

**Figure 4.** Intermediate output

As we move deeper into the neural network, the activations transition from recognizing simple, low-level patterns to capturing more abstract and complex features, such as borders, corners, and angles. These higher-level layers are essential because they represent a transformation from concrete visual information to abstract, class-specific representations. By this stage, the network's focus shifts from the specific visual contents of an image (e.g., individual pixels or edges) to more holistic characteristics that help identify the image's overall category, such as the type of damage or object.

This is where techniques like fine-tuning with layer freezing become crucial. In practice, without fine-tuning, a model may perform well during training, but it could fail to generalize effectively on unseen data, such as validation or test sets. This failure often occurs because the network's early layers are not effectively tuned to detect critical low-level features like shapes and edges. By selectively freezing certain layers and fine-tuning others, we can control the network's variance, allowing it to better extract and process relevant features across different layers. This method significantly enhances the model's ability to generalize to new, unseen data, leading to better overall performance in real-world applications.

5. Conclusions

In this study, multiple deep convolutional neural networks were employed to classify structural damage. While these models demonstrated fast computational performance, they exhibited certain limitations in achieving high accuracy, particularly in multi-class classification tasks. The results suggest that the models' classification performance can be significantly improved through further fine-tuning of the models and hyperparameter optimization, which could enhance their ability to generalize across different damage categories.

Beyond classification, this research highlights possibility for future investigation, such as damage localization and quantification at system and regional level [24-25], which are critical for a comprehensive understanding of structural integrity. These tasks could be achieved by extending the current model to incorporate techniques for identifying the specific locations and extents of damage. Moreover, the integration of drone or satellite imagery could further elevate the model's capability, particularly in post-disaster scenarios where rapid and large-scale assessment of infrastructure is essential for timely disaster response and recovery planning.

However, a notable challenge encountered in this research was the prediction accuracy falling below expectations. This was primarily attributed to the low quality of the input images used during training and testing. One potential solution to this issue would be to apply image segmentation techniques during the pre-processing stage. By isolating relevant structural features and excluding unnecessary background noise, segmentation could help the model focus on critical areas of damage, thereby improving classification accuracy. Overall, the study provides a foundation for further refinement and expansion of CNN-based methods for structural damage assessment, with promising implications for improving post-disaster infrastructure evaluation and maintenance planning.

References

- [1] Liu, T., & Meidani, H. (2023). Physics-informed neural network for nonlinear structural system identification. *Changes*, 10, 11.
- [2] Liu, T., & Meidani, H. (2023). Physics-informed neural networks for system identification of structural systems with a multiphysics damping model. *Journal of Engineering Mechanics*, 149(10), 04023079.
- [3] Yan, Y., Chow, A. H., Ho, C. P., Kuo, Y.-H., Wu, Q., & Ying, C. (2022). Reinforcement learning for logistics and supply chain management: Methodologies, state of the art, and future opportunities. *Transportation Research Part E: Logistics and Transportation Review*, 162, 102712.
- [4] Yan, Y., Deng, Y., Cui, S., Kuo, Y.-H., Chow, A. H., & Ying, C. (2023). A policy gradient approach to solving dynamic assignment problem for on-site service delivery. *Transportation Research Part E: Logistics and Transportation Review*, 178, 103260.
- [5] Ying, C., Chow, A. H., Yan, Y., Kuo, Y.-H., & Wang, S. (2024). Adaptive rescheduling of rail transit services with short-turnings under disruptions via a multi-agent deep reinforcement learning approach. *Transportation Research Part B: Methodological*, 188, 103067.
- [6] Yan, Y., Cui, S., Liu, J., Zhao, Y., Zhou, B., & Kuo, Y.-H. (2024). Multimodal fusion for large-scale traffic prediction with heterogeneous retentive networks. *Information Fusion*, 102695.
- [7] Cheng, X., & Lin, J. (2024). Is electric truck a viable alternative to diesel truck in long-haul operation? *Transportation Research Part D: Transport and Environment*, 129, 104119.
- [8] Cheng, X., Mamalis, T., Bose, S., & Varshney, L. R. (2024). On carsharing platforms with electric vehicles as energy service providers. *IEEE Transactions on Intelligent Transportation Systems*.
- [9] Liu, K., Hu, F., Lin, H., Cheng, X., Chen, J., Song, J., Feng, S., Su, G., & Zhu, C. (2024). Deep reinforcement learning for real-time ground delay program revision and corresponding flight delay assignments. *arXiv preprint arXiv:2405.08298*.
- [10] Guo, G., Li, X., Zhu, C., Wu, Y., Chen, J., Chen, P., & Cheng, X. (2025). Establishing benchmarks to determine the embodied carbon performance of high-speed rail systems. *Renewable and Sustainable Energy Reviews*, 207, 114924.
- [11] You, J., Shi, H., Jiang, Z., Huang, Z., Gan, R., Wu, K., Cheng, X., Li, X., & Ran, B. (2024). V2X-VLM: End-to-end V2X cooperative autonomous driving through large vision-language models. *arXiv preprint arXiv:2408.09251*.
- [12] Su, G., Cheng, X., Feng, S., Liu, K., Song, J., Chen, J., Zhu, C., & Lin, H. (2024). Flight path optimization with optimal control method. *arXiv preprint arXiv:2405.08306*.
- [13] Cheng, X., Nie, Y. M., & Lin, J. (2024). An autonomous modular public transit service. *Transportation Research Part C: Emerging Technologies*, 104746.
- [14] Cheng, X., & Kontou, E. (2023). Estimating the electric vehicle charging demand of multi-unit dwelling residents in the United States. *Environmental Research: Infrastructure and Sustainability*, 3(2), 025012.
- [15] Peng, J., Zhang, S., Peng, D., & Liang, K. (2018). Research on bridge crack detection with neural network-based image processing methods. In *Proceedings of the 12th International Conference on Reliability, Maintainability, and Safety (ICRMS)* (pp. 419–428). IEEE.

- [16] O'Byrne, M., Pakrashi, V., Schoefs, F., & Ghosh, B. (2018). Damage assessment of the built infrastructure using smartphones. *Civil Engineering Research in Ireland*.
- [17] Ukai, M. (2019). Tunnel lining crack detection method by means of deep learning. *Quarterly Report of RTRI*, 60(1), 33–39.
- [18] Gao, Y., Li, K., Mosalam, K., & Günay, S. (2018). Deep residual network with transfer learning for image-based structural damage recognition. In *Proceedings of the Eleventh US National Conference on Earthquake Engineering*.
- [19] Cha, Y.-J., Choi, W., Suh, G., Mahmoudkhani, S., & Büyüköztürk, O. (2018). Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types. *Computer-Aided Civil and Infrastructure Engineering*, 33(9), 731–747.
- [20] Che, C., Li, C., & Huang, Z. (2024). The integration of generative artificial intelligence and computer vision in industrial robotic arms. *International Journal of Computer Science and Information Technology*, 2(3), 1-9.
- [21] Tianbo, S., Weijun, H., Jiangfeng, C., Weijia, L., Quan, Y., & Kun, H. (2023, January). Bio-inspired swarm intelligence: A flocking project with group object recognition. In *2023 3rd International Conference on Consumer Electronics and Computer Engineering (ICCECE)* (pp. 834-837). IEEE.
- [22] Xu, J., Jiang, Y., Yuan, B., Li, S., & Song, T. (2023, November). Automated scoring of clinical patient notes using advanced NLP and pseudo labeling. In *2023 5th International Conference on Artificial Intelligence and Computer Applications (ICAICA)* (pp. 384-388). IEEE.
- [23] Zhang, Q., Cai, G., Cai, M., Qian, J., & Song, T. (2023). Deep learning model aids breast cancer detection. *Frontiers in Computing and Intelligent Systems*, 6(1), 99-102.
- [24] Xu, X., Yuan, B., Song, T., & Li, S. (2023, November). Curriculum recommendations using transformer base model with infonce loss and language switching method. In *2023 5th International Conference on Artificial Intelligence and Computer Applications (ICAICA)* (pp. 389-393). IEEE.
- [25] Yuan, B., & Song, T. (2023, November). Structural resilience and connectivity of the IPv6 internet: An AS-level topology examination. In *Proceedings of the 4th International Conference on Artificial Intelligence and Computer Engineering* (pp. 853-856).
- [26] Yuan, B., Song, T., & Yao, J. (2024, January). Identification of important nodes in the information propagation network based on the artificial intelligence method. In *2024 4th International Conference on Consumer Electronics and Computer Engineering (ICCECE)* (pp. 11-14). IEEE.
- [27] Qian, J., Song, T., Zhang, Q., Cai, G., & Cai, M. (2023). Analysis and diagnosis of hemolytic specimens by AU5800 biochemical analyzer combined with AI technology. *Frontiers in Computing and Intelligent Systems*, 6(3), 100-103.
- [28] Cai, G., Qian, J., Song, T., Zhang, Q., & Liu, B. (2023). A deep learning-based algorithm for crop disease identification positioning using computer vision. *International Journal of Computer Science and Information Technology*, 1(1), 85-92.
- [29] Song, T., Zhang, Q., Cai, G., Cai, M., & Qian, J. (2023). Development of machine learning and artificial intelligence in toxic pathology. *Frontiers in Computing and Intelligent Systems*, 6(3), 137-141.