Advanced Computer Vision Techniques for Detecting and Segmenting Structural Visible Seismic Damages Under Varied Testing Conditions

---

ENSIEH ALI BAKHSHI[1] and OMID YAZDANPANAH[2,*]

ABSTRACT

A cutting-edge computer vision method, incorporating an attention mechanism, transformer architecture, and a customized U-Net model, is used for pixel-level multicategory detection of visible seismic damage in reinforced concrete (RC) bridge piers. The damage categories include cracks, spalling, reinforcement exposure, crushing, buckling, and structural failure. A semantic segmentation database is created from experimental photos obtained through cyclic tests, shaking table experiments, and real-time hybrid simulations. To ensure seamless reconstruction, smooth blending techniques such as overlapping and mirror padding are applied to the predicted patch masks. The image database undergoes extensive preprocessing, including lens correction, perspective adjustment, labeling, and damage-type balancing using rotation, flipping, Gamma correction, Hue and Saturation adjustments, and blurring effects. Both sample-level and pixel-level data balancing are achieved through hypergeometric distribution and weighted loss functions, respectively, ensuring the desired probability distribution for each damage category. A hybrid loss function optimizes model performance and metrics like Intersection over Union (IoU) and F1 score track training and validation progress. Atrous convolution is integrated for multi-scale feature extraction, enhancing detection accuracy across varying spatial resolutions. The proposed vision-based approach is validated on unseen (out-of-database) RC bridge pier images, demonstrating high accuracy in detecting multicategory seismic damage. Additionally, crack feature extraction is conducted, measuring total crack length, average and maximum crack widths, and angle, while the location of maximum crack width is also investigated. These findings underscore the promise of automated structural health monitoring and post-earthquake safety assessments, enhancing resilience and enabling rapid decision-making.

Keywords: Bridge piers, Attention layer, Transformer, U-Net architecture, Multicategory seismic visible damage, Pixel-level detection

---

[1] Researcher, Industry&Academia Cooperation Foundation, Myongji University, Republic of Korea

[2] Research Professor, Hybrid Structural Testing Center (Hystec), Myongji University, Republic of Korea

*Corresponding author; e-mail: omidyazdanpanah@mju.ac.kr

1 Introduction

After an earthquake, a well-designed retrofitting and reconstruction plan is essential. Initial damage inspections focus on safety but provide limited evaluations due to time constraints. Detailed damage assessments are necessary to guide rehabilitation. Engineers need technical guidelines to ensure safe, efficient, and cost-effective restoration of damaged reinforced concrete (RC) buildings and bridges [1].

Contact-based sensors, such as LVDTs, accelerometers, load cells, and strain gauges, provide valuable real-time data on bridge pier performance. However, they are complex to install, maintain, and measure locally. Non-contact sensors like vision-based systems offer a more efficient solution with simpler deployment, broader coverage, and lower maintenance, making them ideal for comprehensive structural monitoring [2].

Traditional post-earthquake safety assessments of RC buildings and bridges depend on visual inspections by certified engineers. These are effective but time-consuming and prone to human bias. Large-scale disasters can overwhelm experts, delaying evaluations and necessary interventions. However, computer vision and deep learning can automate and improve these assessments, rapidly detecting and categorizing seismic damage from images or video [3]. Despite their potential, challenges remain, such as the need for high-quality annotated datasets and ongoing model evaluation. Advances in deep learning, computer vision, and image processing can help improve these models for better real-world applications [4].

Image data for deep learning-based damage prediction can come from field data, simulations, or laboratory experiments. Cyclic tests, real-time hybrid simulations (RTHS), and shaking table tests provide valuable insights into bridge pier damage [5]. RTHS, which integrates experimental data with real-time simulations, is a cost-effective alternative to shaking table tests. It offers high precision in simulating seismic behavior, particularly for large structures [6-7].

This paper contributes by: (1) Developing a comprehensive database using various test methods to improve deep learning model accuracy in predicting seismic damage. (2) Introducing an advanced deep learning model for multi-category damage semantic segmentation, incorporating innovative techniques such as transformer block including multi-head attention layer and U-Net architecture. (3) Implementing a hybrid loss function to improve predictions on unseen data and applying data balancing strategies. (4) Reducing edge artifacts and improving model performance using smooth blending and reflection padding. (5) Enhancing data through augmentation techniques and utilizing a learning rate scheduler to ensure stable training and better generalization.

2 Damage state definition

The damage evaluation guidelines classify seismic damage in structural components into five levels (see Fig. 1), correlating them with the lateral force-displacement relationship to assess structural performance [1]. Damage State I involves minor cracks (<0.2 mm) without yielding, maintaining a linear behavior. Damage State II (0.2–1.0 mm cracks) introduces some yielding, while Damage State III (1.0–2.0 mm cracks, minor cover concrete loss) exhibits non-linear hardening behavior.

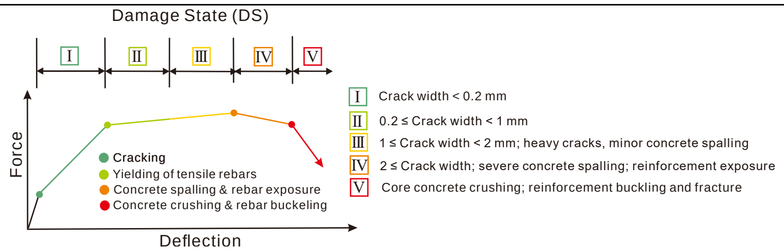


Fig. 1. Linking RC component damage state to visible seismic damages.

More severe damage appears in Damage State IV (cracks >2.0 mm, cover loss, exposed reinforcement), potentially reducing lateral load capacity but preserving gravity load support. The most critical, Damage State V, involves rebar buckling, concrete crushing, vertical deformation, and significant lateral capacity loss [1]. These damage states inform repair and retrofitting strategies. This study classifies visible seismic damage into five categories (classes 1-5): Cracks, Spalling, Exposure, Crushing, and Buckling/Failure, respectively, following standard criteria, while non-damaged areas are assigned to Class 0 as background.

3 Data preprocessing

Effective image data preprocessing is essential for deep learning-based semantic segmentation, ensuring numerical stability, faster convergence, and improved model performance. Key steps include scaling pixel values to [0,1], data augmentation to enhance diversity, and geometric corrections for accurate labeling. To integrate predicted patch masks seamlessly, overlapping and mirror padding are applied, while hypergeometric distribution and weighted loss functions balance damage category probabilities. A patch-based approach is employed to tackle out-of-memory errors when processing large images, with two models developed: (1) Multi-Category Segmentation Network for Classes 0, 2, 3, 4, and 5, processing 256×256 patches, and (2) Crack Segmentation Network for Classes 0 and 1, using 32×32 patches. Both models train with a batch size of 16, with the multi-category network utilizing 1723 patches, reserving 20% (344 patches) as unseen data, and the remaining 1379 patches split 70:30 for training (965 patches) and validation (414 patches). The hypergeometric distribution [8] (see Eq. (1)), defined by total sample size (N), number of samples in category k (N_k), and batch size ($N_b = 16$), determines the probability of each damage category in a batch, ensuring a balanced representation (see Table 1).

$$P(X=i)=\binom{N_k}{i}\binom{N-N_k}{N_b-i}\bigg/\binom{N}{N_b} \tag{1}$$

Table 1 outlines the probability of having no more than 2 samples of category k in a batch, given the limited image patches for Class 3 (exposure category) in the train and validation datasets; notably, there is a 97.586% chance of exceeding this threshold per batch.

Table 1. Number of samples with each damage category and the corresponding probability for multi-category segmentation network

	Class 0	Class 2	Class 3	Class 4	Class 5	Total
Training data	560	820	370	605	582	965
Validation data	225	366	173	285	258	414
Probability for training data	2.102×10^{-4}	1.567×10^{-10}	2.414×10^{-2}	4.508×10^{-5}	1.017×10^{-4}	-

Table 2. Proportion of pixels in each category and the corresponding normalized class weights

	Class 0	Class 2	Class 3	Class 4	Class 5
Training data	0.2867	0.4188	0.0308	0.1738	0.0898
Normalized class weights (*w*)	0.0631	0.0433	0.5881	0.1042	0.2013

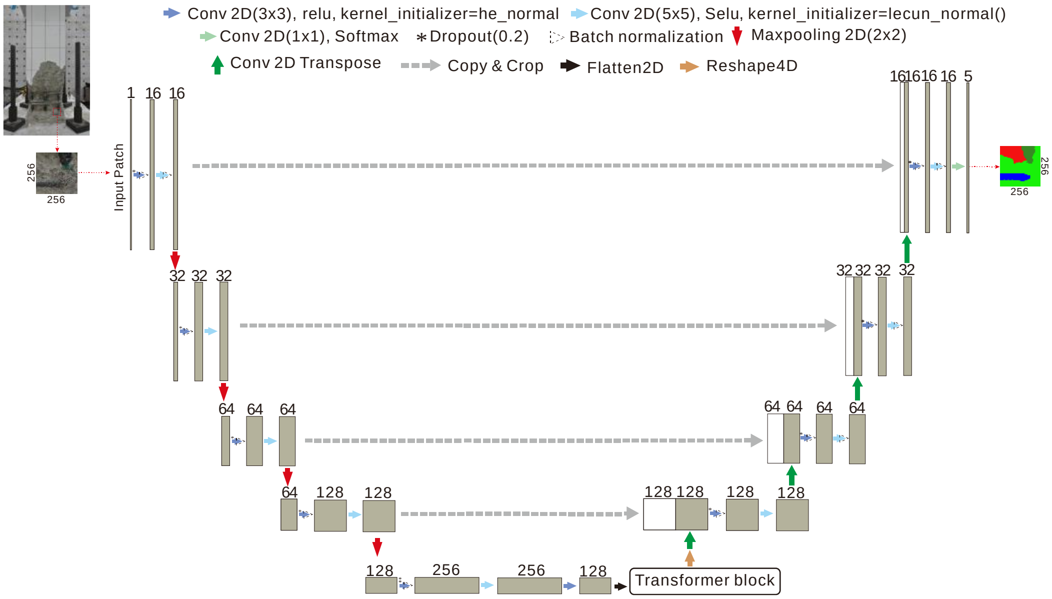
Addressing pixel-level class imbalance is crucial for preventing model bias toward majority classes and ensuring the learning of meaningful features across all categories. Without proper balancing, deep learning models may neglect minority-class pixels, leading to poor generalization, disrupted loss optimization, and misleading evaluation metrics. To mitigate this, techniques like loss function adjustments improve model robustness, computational efficiency, and detection performance for rare features. In this study, we assign higher weights to minority classes in the multicategory segmentation network, as determined by Eq. (2), based on pixel proportions to emphasize underrepresented categories. The class weights are normalized inversely to class proportions, ensuring fair learning across all classes. As shown in Table 2, Class 3 (exposure) is significantly underrepresented at the pixel level, resulting in a higher class weight. A hybrid loss function combining Dice, Focal, Jaccard, and Tversky losses is introduced in Eq. (3) to address the severe class imbalance and optimize training, improving multicategory segmentation network robustness on unseen data. The same procedure is applied to the crack segmentation network.

$$w_i = \left(1/p_i\right) \Big/ \sum_{j=1}^{K} \left(1/p_j\right) \tag{2}$$

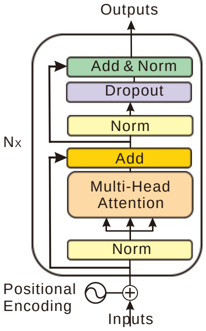
$$L_{Total} = L_{Dice} + L_{Focal} + L_{Jaccard} + L_{Tversky} \tag{3}$$

4 Proposed deep learning model for multi-category pixel-level segmentation

A customized U-Net model is proposed for multicategory seismic damage segmentation in bridge piers under diverse loading and testing conditions. Built on 2D CNNs from the original U-Net, it accepts 256×256 image patches and incorporates Conv2D layers with 3×3 and 5×5 kernels in both the encoder and decoder. To improve performance and reduce overfitting, it includes dropout and batch normalization layers, with a final Conv2D layer and softmax activation (see Fig. 2(a)). The model applies a dilation rate of 2 in block 4 of the encoder and its corresponding block in the decoder (block 6) of the customized U-Net architecture and introduces a 1D transformer block in the bottleneck for enhanced segmentation (see Fig. 2(b)). Fig. 3 illustrates semantic segmentation of seismic damage in unseen bridge piers, assessed via per-class IoU on randomly selected RGB patches, their labels, and predicted masks, later stitched into the original large image. Fig. 4 showcases the model's performance on a large image.

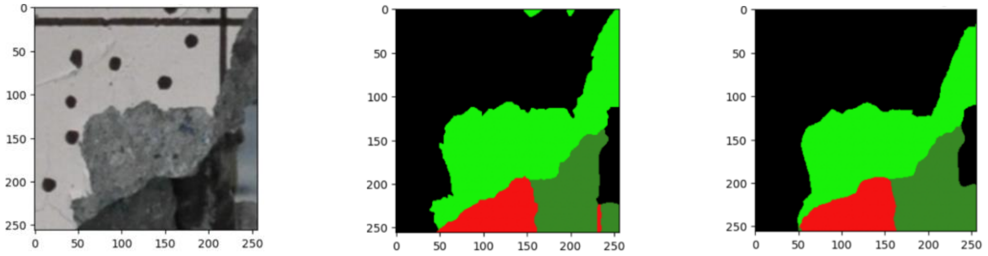


(a) Customized U-Net model



(b) 1D transformer block

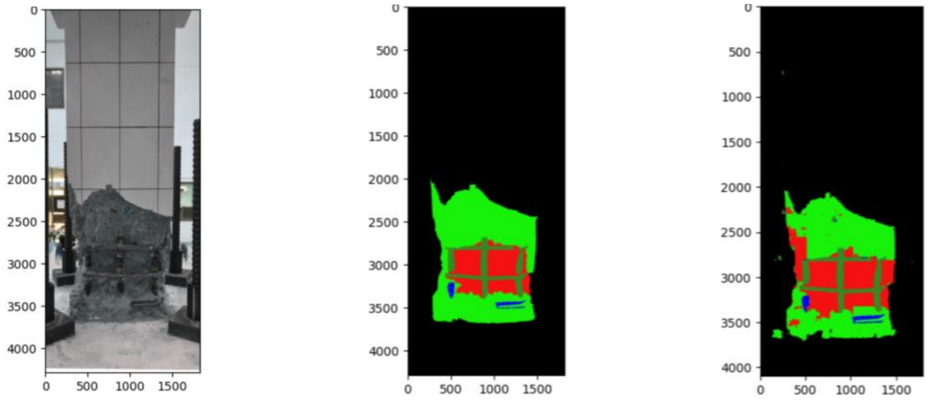
Fig. 2. Proposed multicategory segmentation network architecture.



(a) RGB image patch (b) RGB label patch (c) Predicted mask patch

per-class IoU					mIoU
Class 0	Class 2	Class 3	Class 4	Class 5	
0.9385	0.8733	-	0.8638	0.8854	0.8903

Fig. 3. Patched image, corresponding label, and predicted mask including multicategory damage.



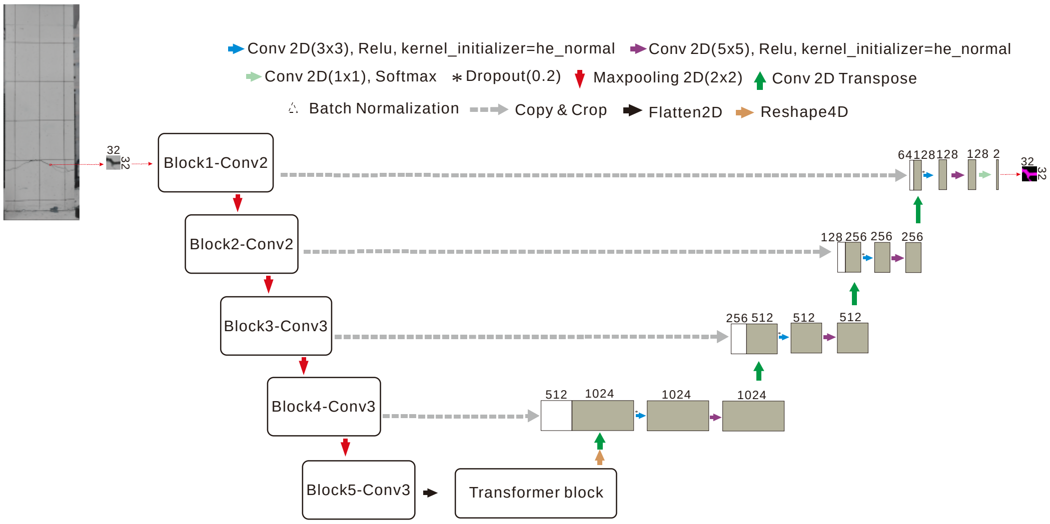
(a) RGB image (b) RGB label (c) Predicted mask

per-class IoU					Mean IoU
Class 0	Class 2	Class 3	Class 4	Class 5	
0.9930	0.8146	0.7944	0.6674	0.8612	0.8261

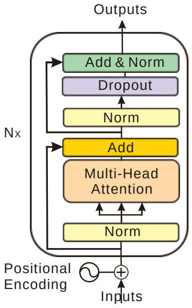
Fig. 4. Large image, corresponding label, and predicted mask including multiple damage categories.

5 Proposed deep learning model for crack pixel-level segmentation

The crack segmentation network architecture employs a VGG16 backbone, pre-trained on ImageNet, as its encoder with frozen layers to retain feature representations. Feature maps from specific convolutional layers (block1_conv2, block2_conv2, block3_conv3, and block4_conv3) are integrated into the U-Net decoder via skip connections, aiding segmentation. The deepest layer, block5_conv3, acts as the bottleneck, capturing high-level features. A transformer block is added to the bottleneck and the key and value vectors of the multi-head attention layer are set to 512 for dimensional compatibility. As depicted in Fig. 5, the model processes 32×32 pixel patches, using the same loss function as the multi-category segmentation network but with adjusted class weights. Training is conducted with a batch size of 16, employing the Adamax optimizer, an exponential learning rate scheduler, and weight decay for stability and performance. Fig. 6 presents a randomly selected RGB image patch sample containing crack pixels. The results suggest that the crack segmentation network effectively segments and detects seismic cracks in unseen bridge piers with an acceptable per-class IoU. Fig. 7 illustrates the performance of the crack segmentation network in blending small patches back to the original large size.

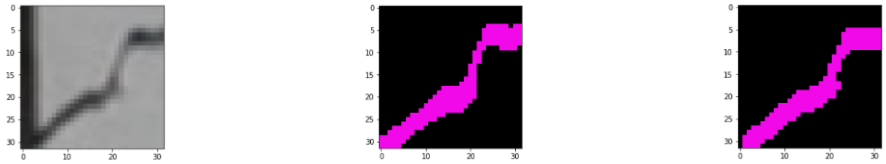


(a) Customized VGG16-U-Net model



(b) 1D transformer block

Fig. 5. Proposed crack segmentation network architecture.



(a) RGB image patch

(b) RGB label patch

(c) Predicted mask patch

per-class IoU

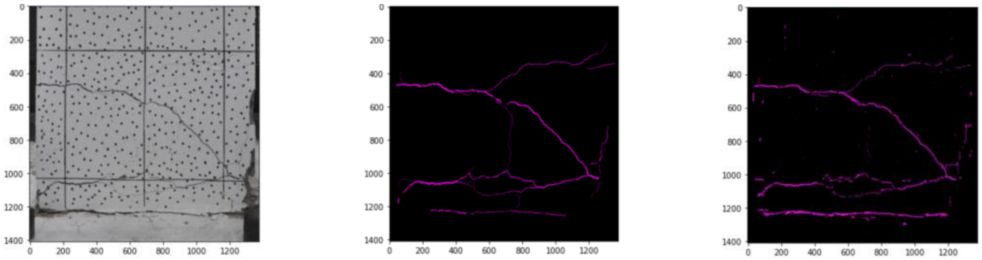
Class 0: 0.9596

Class 1: 0.8187

mIoU

0.8892

Fig. 6. Image patch, corresponding label, and predicted mask patch including crack category damage.



(a) RGB image

(b) RGB label

(c) Predicted mask

per-class IoU

Class 0: 0.9883

Class 1: 0.5013

mIoU

0.7448

Fig. 7. Large image, corresponding label, and predicted mask focusing on crack regions.

6 Conclusions

This study presents a novel computer-vision approach for pixel-level multi-category segmentation of seismic damage in RC bridge piers. A diverse image database is compiled, covering cyclic tests, shaking table tests, and RTHS, which provide a more realistic representation of seismic damage compared to traditional cyclic tests. The annotation process includes five damage categories: crack, spalling, exposure, crushing, buckling, and failure. To address severe data imbalance, augmentation techniques, hypergeometric distribution, class weighting, and a hybrid loss function are applied at both sample and pixel levels, and two deep learning models are introduced. The Multi-category segmentation network (excluding cracks), is based on a U-Net framework with multi-scale feature extraction, dropout, batch normalization, and a 1D transformer block in the bottleneck to enhance accuracy. The crack segmentation network replaces the U-Net encoder with the VGG16 encoder, leveraging transfer learning from ImageNet to improve segmentation precision while reducing overfitting. Both models utilize a patch-based approach to efficiently process large images and employ advanced blending techniques, such as overlapping and mirror padding, for seamless reconstruction. While previous studies primarily relied on predefined displacement time histories, which can overestimate or underestimate seismic demands, this research lays the foundation for future studies focusing on RTHS-derived images for more realistic damage assessment. Additionally, key structural response indicators, including residual crack width and lateral force-drift hysteresis curves, are particularly sensitive to these improvements. By integrating deep learning with real-world seismic data, this study enhances the accuracy of seismic damage detection, ultimately supporting more reliable post-earthquake assessments and structural resilience strategies for RC bridge piers.

Acknowledgments

The authors gratefully acknowledge the financial support provided by the Creative Challenge Research project funded by the Ministry of Science and ICT through the National Research Foundation of Korea (NRF) (RS-2023-00241807).

References

1. Maeda, M., T. Nishida, K. Matsukawa, M. Murakami. 2017. “Revision of guideline for postearthquake damage evaluation of reinforced concrete buildings in Japan”, In 16th World Conference on Earthquake Engineering, Santiago, Chile.
2. Hamidia, M., M. Sheikhi, A. H. Asjodi, K.M. Dolatshahi. 2024. “Computer vision-based quantification of updated stiffness for damaged RC columns after earthquake,” *Advances in Engineering Software*, 190: 103597.
3. Miao, Z., X. Ji, T. Okazaki, N. Takahashi. 2021. “Pixel-level multicategory detection of visible seismic damage of reinforced concrete components,” *Computer-Aided Civil and Infrastructure Engineering*, 36(5): 620-637.
4. Yazdanpanah, O., M. Park, M. Chang, Y. Chae. 2024. “Mastering seismic time series response predictions using an attention-Mamba transformer model for bridge bearings and piers across varied testing conditions,” *Scientific Reports*, 14(1): 29751.
5. Yazdanpanah, O., M. Chang, E. Ali Bakhshi. 2024. “Attention-based hybrid convolutional-long short-term memory network for bridge pier hysteresis and backbone curves prediction,” *Integrated Computer-Aided Engineering*, 10692509241301240.
6. Park, J., M. Park, Y. Chae, C.Y. Kim. 2023. “Real-time hybrid simulation for investigating the influence of vertical ground motions on the lateral response of RC piers,” *Earthquake Engineering & Structural Dynamics*, 52(10): 2928-2944.
7. Pizarro, D., M. Kovarbašić, B. Stojadinović. 2025. “Hybrid Simulation Tests of Real-Scale Squat Reinforced Concrete Shear Wall Specimens,” *Journal of Structural Engineering*, 151(4): 04025017.
8. Rice, J.A. 2007. “Mathematical Statistics and Data Analysis (Third ed.),” *Duxbury Press*, p 42.