High-Definition 3D Digital Twin for Damage Mapping for Bridge Systems Using Advanced NeRF and STRNet

GEONTAE KIM[1] and YOUNGJIN CHA[2,*]

ABSTRACT

This research presents an innovative approach to generating three-dimensional (3D) digital twin models for damage assessment in large-scale civil structures. By refining the Nerfacto model, which builds upon Neural Radiance Fields (NeRF), the study achieves highly accurate 3D reconstruction with detailed, pixel-level damage visualization. To enhance precision, STRNet with Test Time Augmentation (TTA) was utilized for pixel-wise crack segmentation, ensuring seamless incorporation of damage data into the digital twin model. Extensive case studies validated the effectiveness of this methodology for large-scale infrastructure, highlighting its potential for proactive structural health monitoring. Future investigations will focus on real-world outdoor applications, integrating UAV-based imaging while addressing challenges such as scale variations and lighting inconsistencies. This framework provides a reliable foundation for automating infrastructure inspection and maintenance.

Keywords: Damage detection, computer vision, pixelwise, digital twin, 3D cloud points model, deep learning

INTRODUCTION

In Canada, aging bridges illustrate the urgent need for SHM due to underinvestment, especially at the provincial level [1]. Structural Health Monitoring (SHM) is essential for assessing structural condition and extending infrastructure lifespan [2]. In Canada, aging bridges illustrate the urgent need for SHM due to underinvestment, especially at the provincial level [1]. SHM enables safer, more cost-effective maintenance [3], yet traditional methods like vibration sensing face challenges such as sensor malfunctions and environmental noise [4].

Cracks, often resulting from environmental and mechanical stresses, are key indicators of damage and must be addressed promptly. Vision-based systems offer a more reliable, automated alternative to manual inspections [5-6], though early image-based methods struggled with complex backgrounds [7]. Deep learning, especially CNNs, has significantly improved crack detection [8-9] and its derivatives [10-12] achieving high accuracy.

Advanced models like SDDNet [13] and STRNet [14] integrate modules for feature refinement and real-time processing. However, CNNs require large datasets for training, which is impractical for many SHM applications [15]. Test Time Augmentation (TTA) offers a lightweight alternative that improves prediction accuracy without retraining [16].

[1]MSc, Dept. of Civil Engineering, University of Manitoba, Winnipeg, MB, Canada
[2]Professor, Dept. of Civil Engineering, University of Manitoba, Winnipeg, MB, Canada
[*]Corresponding author: Email: young.cha@umanitoba.ca

Single-image detection limits context and localization of damage, prompting interest in mapping 2D data onto 3D models through autonomous flight of unmanned aerial vehicles (UAVs) [16-19]. 3D reconstruction creates digital twins for intelligent SHM [20], with photogrammetry and Structure from Motion (SfM) being commonly used [21]. CNN-based segmentation enhances 3D crack mapping, although photogrammetry is computationally intensive and struggles with planar surfaces.

Neural Radiance Fields (NeRF) offer a flexible alternative by modeling scenes through neural networks [22], enabling photorealistic 3D rendering. Despite its promise, NeRF is limited by high computational demands and sensitivity to hyperparameters, making it unsuitable for large-scale infrastructure. To address this, Nerfacto [23] was modified for 3D reconstruction and the pixelwise damage segmentation was done through STRNet and TTA. This integrated approach [24-26] aims to enable accurate 3D damage mapping from 2D images for efficient and intelligent SHM. Therefore, in this conference paper, some parametric studies are conducted to investigate this new approach [26].

METHOLOGY

The developed model [26] integrates a deep learning-based pixelwise damage segmentation method with an advanced 3D image reconstruction framework. As illustrated in Figure 1, the workflow initiates with the acquisition of 2D RGB images of the target structure. SfM is utilized to extract the necessary intrinsic and extrinsic camera parameters. These parameters are subsequently employed within the Nerfacto model [23], a cutting-edge NeRF-based rendering technique. Concurrently, the 2D images undergo pixel-level damage segmentation using the newly proposed TTA-STRNet architecture [26]. The segmented images, together with the computed camera parameters, are fed into the Nerfacto model to facilitate 3D rendering and the generation of a spatially accurate 3D damage map.

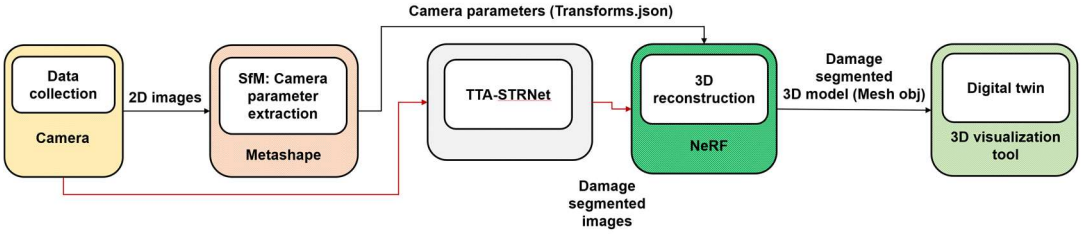


Figure 1: Overall process of the modified Nerfacto based 3D damage mapping method

CAMERA PARAMETER

To employ the NeRF-based Nerfacto model for 3D reconstruction—whether incorporating segmented damage or not—it is imperative to extract the intrinsic and extrinsic camera parameters for each image. The intrinsic matrix encapsulates the camera's internal characteristics, including focal lengths, principal point coordinates, and a skew parameter, which is typically negligible in modern imaging systems. In contrast, the extrinsic matrix defines the spatial transformation between the global coordinate system (GCS) and the local camera coordinate system (LCS), comprising a rotation matrix and a translation vector representing the camera's position in space.

These parameters, along with the corresponding images, are fed into the Nerfacto model. For comprehensive reconstruction, a series of images captured from varying angles and locations—each associated with a unique LCS—is required. A 3D point expressed in camera coordinates is first mapped to normalized image coordinates, which are subsequently transformed into pixel coordinates via the intrinsic matrix. Conversely, spatial transformations between GCS and LCS are governed by the extrinsic matrix, where the rotation matrix is computed as a product of individual axis-aligned rotations.

To obtain these parameters, SfM, implemented through Agisoft Metashape, is employed. Metashape generates an output file containing both intrinsic and extrinsic information. While the focal lengths and principal point can be inferred from image dimensions and metadata, the comprehensive file serves as a critical input for the Nerfacto model, enabling precise ray projection and accurate color sampling during the 3D reconstruction process.

TTA-STRNet

For 2D crack segmentation, this study employs STRNet [14], a high-performance deep learning framework specifically designed for real-time, pixel-level crack detection in large-scale images with complex backgrounds (1280×720 resolution). STRNet has demonstrated state-of-the-art performance, achieving a mean Intersection over Union (mIoU) of 92.6% and a processing speed of 49 frames per second (FPS), affirming its efficacy for structural damage detection tasks.

Due to the constrained variety of testing images from the laboratory-based bridge system utilized in this research, TTA is integrated with STRNet to enhance model robustness and generalizability. Unlike traditional data augmentation applied during training, TTA operates during inference, applying a series of transformations—including resizing, flipping, and rotating—across multiple image variants. This strategy yields augmented test cases, enabling comprehensive evaluation without necessitating model retraining.

High-resolution input images (2992×2000 pixels) are initially downscaled to 1280×720 and 2048×1024 to conform to STRNet's design parameters. Following segmentation, the outputs are upscaled to their original resolution. Functioning analogously to an ensemble approach, TTA aggregates predictions across all augmented instances, thereby improving segmentation precision and minimizing inference bias. The STRNet architecture is composed of an attention-based encoder and a multi-head attention-based decoder, integrated within a series of 11 irregularly repeated STR modules. The network incorporates the Hswish activation function, batch normalization, and a focal-Tversky loss function, all of which contribute to enhancing segmentation accuracy while maintaining computational efficiency.

To preserve critical multi-dimensional features throughout the network, STRNet employs skip connections alongside both coarse and standard up-sampling mechanisms. In the final segmentation stage, outputs from the standard and coarse up-sampling blocks are fused with features from the initial convolution block. This combined feature map is processed through a pointwise convolution (PW) layer, yielding the final segmentation output with high precision at the pixel level. The complete architecture and operational details are elaborated in Kang and Cha [14].

NeRF for 3D Reconstruction

Three-dimensional (3D) reconstruction offers significant advantages over traditional 2D approaches by facilitating immersive visualizations, enabling precise spatial localization of structural damage, and allowing for more accurate damage quantification. Additionally, reconstructed models can be exported as mesh OBJ files, making them compatible with advanced applications such as 3D simulations and finite element analysis (FEA) within digital twin frameworks. The NeRF model, introduced by Mildenhall et al. (2021) [22], represents a paradigm shift in scene reconstruction by leveraging a fully connected deep neural network (DNN) to synthesize 3D scenes from a collection of 2D images. NeRF encodes the scene as a continuous volumetric function, wherein each spatial point is attributed with a volumetric density—indicating transparency or opacity—and a view-dependent RGB color. Through a learning process, the DNN infers these values by regressing the radiance field from pixel-level RGB information, enabling the generation of photorealistic renderings from arbitrary novel viewpoints.

Despite its impressive fidelity, the original NeRF framework is hindered by substantial computational and memory requirements, prolonged training times, and suboptimal rendering performance under inconsistent lighting conditions. The following subsections provide a comprehensive examination of the foundational NeRF model and the more efficient Nerfacto variant, including detailed methodologies for computing 3D spatial coordinates and corresponding ray directions using camera intrinsic and extrinsic parameters.

NeRF

To address the computational and practical limitations of the original NeRF framework, this study adopts the more efficient Nerfacto model, which maintains a similar operational pipeline while offering improved performance. NeRF operates by processing a set of RGB images—such as 2D bridge imagery or segmented damage maps produced via STRNet—alongside a Transforms.json file containing essential camera calibration data. Each pixel is individually trained using backpropagation within a neural radiance field framework. Transformation of the image data from the LCS of each camera view to the GCS is achieved using intrinsic and extrinsic camera parameters. Each pixel corresponds to a camera ray, defined by its origin and direction, which traverses the scene. Along each ray, a sequence of sample points in 3D space is generated. These synthetic sampling points facilitate the learning of radiance values—specifically, color and density—across spatial positions and viewing angles.

NeRF's architecture comprises two fully connected neural networks, denoted as DNN-I and DNN-II. Initially, coarse sampling is performed along each ray, and DNN-I predicts preliminary color estimates. Based on the predicted density distribution, a Probability Density Function (PDF) is constructed to guide refined sampling in regions of high informational relevance. These refined sample points are then processed by DNN-II in the fine phase to generate high-fidelity renderings. The model iteratively compares predicted RGB values against ground-truth image data using a loss function, and the neural networks are optimized through

backpropagation. This multi-stage training strategy enables NeRF—and by extension, Nerfacto—to synthesize photorealistic 3D representations from a sparse set of 2D views, with precise geometric and color fidelity across diverse viewpoints.

Extraction of Sample Points Coordinates and Directional Vectors in GCS

Pixel coordinates are first mapped to normalized image coordinates via the intrinsic camera matrix, which encapsulates parameters such as focal lengths and the principal point. These normalized coordinates are then utilized to construct a unit direction vector within the LCS of the camera. To align this vector with the GCS, it is transformed using the transposed rotation matrix derived from the extrinsic camera matrix. This operation yields a globally consistent ray direction. Subsequently, discrete 3D sample points are uniformly distributed along this ray at predefined intervals, serving as inputs for volumetric rendering and neural radiance field inference.

Nerfacto

To address the inherent limitations of the original NeRF model—namely, its computational inefficiency, excessive memory requirements, prolonged training duration, and limited adaptability—this study adopts the Nerfacto framework. Nerfacto integrates several state-of-the-art enhancements, including:

- Proposal Sampling from Mip-NeRF 360
- Multiresolution Hash Encoding from Instant-NGP for spatial sample point representation
- Spherical Harmonics Encoding from Ref-NeRF for efficient directional encoding
- Appearance Embedding from NeRF-W to account for lighting variability in the secondary network (DNN-B)

The modified Nerfacto model was trained using the number of input images detailed in Table I.

TABLE I. TRAINING PERFORMANCE ASSESSMENT

Image size (pixel)	Number of images	PSNR (db)	SSIM	Loss function (MSE)	Training time (Sec)
2992×2000	1,923	28.35	0.8025	0.00284	38235.53

RESULTS

A distinctive aspect of our research is illustrated in Figure 2, wherein a segmented crack is prominently rendered on the 3D bridge model. This visualization exemplifies the method's capability for accurately localizing and quantifying structural damage. The clear delineation of the segmented cracks significantly enriches the interpretability of both the damage location and its severity. Furthermore, Figure 3 offers a comparative perspective between the bridge model with and without crack segmentation. This juxtaposition underscores the value of our segmentation approach,

as the segmented model reveals markedly enhanced detail and diagnostic clarity, offering deeper insights into the structural integrity compared to the unsegmented counterpart.

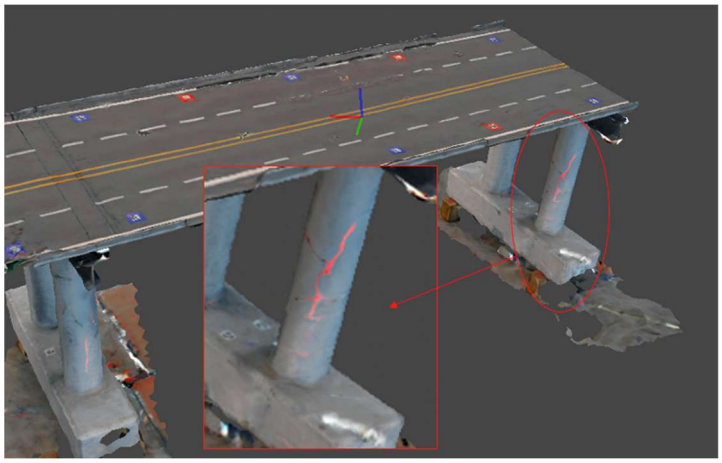
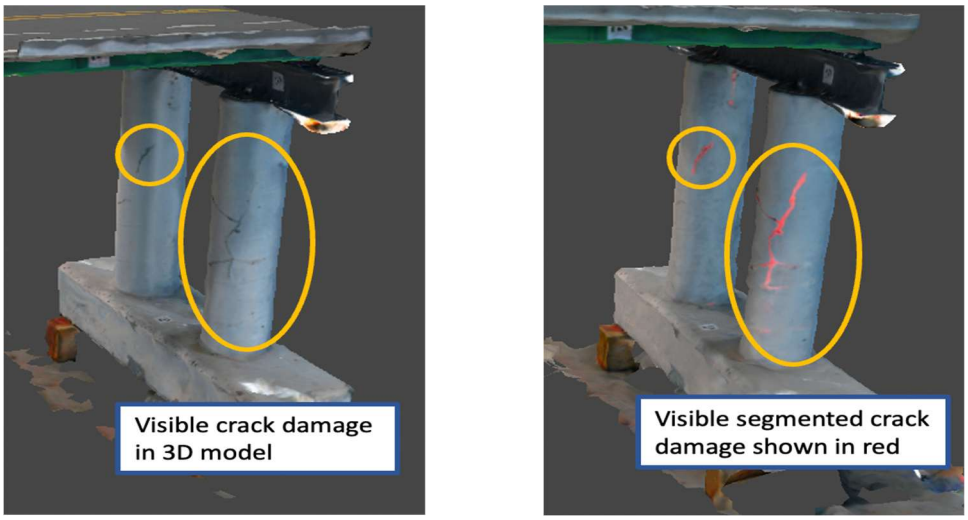


Figure 2. Segmented crack mapped 3d bridge model [25].



(a) (b)

Figure 3. Comparative visualization of an unsegmented crack and segmented crack on the 3d model [25].

REFERENCES

1. Hammad, A. W., Yan, J., & Mostofi, B. (2007). Recent development of bridge management systems in Canada. In 2007 Annual Conference and Exhibition of the Transportation Association of Canada: Transportation-An Economic Enabler (Les Transports: Un Levier Economique) Transportation Association of Canada (TAC).
2. Cha, Y. J., Ali, R., Lewis, J., & Büyüköztürk, O. (2024). Deep learning-based structural health monitoring. Automation in Construction, 161, 105328.
3. Giurgiutiu, V. (2015). Structural health monitoring of aerospace composites.
4. Cha, Y. J., & Buyukozturk, O. (2015). Structural damage detection using modal strain energy and hybrid multiobjective optimization. Computer-Aided Civil and Infrastructure Engineering, 30(5), 347-358.

5. Cha, Y. J., You, K., & Choi, W. (2016). Vision-based detection of loosened bolts using the Hough transform and support vector machines. Automation in Construction, 71, 181-188.
6. Ramana, L., Choi, W., & Cha, Y. J. (2019). Fully automated vision-based loosened bolt detection using the Viola–Jones algorithm. Structural Health Monitoring, 18(2), 422-434.
7. Kheradmandi, N., & Mehranfar, V. (2022). A critical review and comparative study on image segmentation-based techniques for pavement crack detection. Construction and Building Materials, 321, 126162.
8. Cha, Y. J., Choi, W., & Büyüköztürk, O. (2017). Deep learning-based crack damage detection using convolutional neural networks. Computer-Aided Civil and Infrastructure Engineering, 32(5), 361-378.
9. 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.
10. Fu, H., Meng, D., Li, W., & Wang, Y. (2021). Bridge crack semantic segmentation based on improved Deeplabv3+. Journal of Marine Science and Engineering, 9(6), 671.
11. Liu, Z., Cao, Y., Wang, Y., & Wang, W. (2019). Computer vision-based concrete crack detection using U-net fully convolutional networks. Automation in Construction, 104, 129-139.
12. Kang, D., Benipal, S. S., Gopal, D. L., & Cha, Y. J. (2020). Hybrid pixel-level concrete crack segmentation and quantification across complex backgrounds using deep learning. Automation in Construction, 118, 103291.
13. Choi, W., & Cha, Y. J. (2019). SDDNet: Real-time crack segmentation. IEEE Transactions on Industrial Electronics, 67(9), 8016-8025.
14. Kang, D. H., & Cha, Y. J. (2022). Efficient attention-based deep encoder and decoder for automatic crack segmentation. Structural Health Monitoring, 21(5), 2190-2205.
15. Dryden, N., Maruyama, N., Benson, T., Moon, T., Snir, M., & Van Essen, B. (2019, May). Improving strong-scaling of CNN training by exploiting finer-grained parallelism. In 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS) (pp. 210-220). IEEE.
16. Ayhan, M. S., & Berens, P. (2018, June). Test-time data augmentation for estimation of heteroscedastic aleatoric uncertainty in deep neural networks. In Medical Imaging with Deep Learning.
17. Kang, D., & Cha, Y. J. (2018). Autonomous UAVs for structural health monitoring using deep learning and an ultrasonic beacon system with geo-tagging. Computer-Aided Civil and Infrastructure Engineering, 33(10), 885-902.
18. Ali, R., Kang, D., Suh, G., & Cha, Y. J. (2021). Real-time multiple damage mapping using autonomous UAV and deep faster region-based neural networks for GPS-denied structures. Automation in Construction, 130, 103831.
19. Waqas, A., Kang, D., & Cha, Y. J. (2024). Deep learning-based obstacle-avoiding autonomous UAVs with fiducial marker-based localization for structural health monitoring. Structural Health Monitoring, 23(2), 971-990.
20. Ye, C., Butler, L., Calka, B., Iangurazov, M., Lu, Q., Gregory, A., ... & Middleton, C. (2019). A digital twin of bridges for structural health monitoring.
21. Hartley, R., & Zisserman, A. (2003). Multiple view geometry in computer vision. Cambridge university press.
22. Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2021). Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM, 65(1), 99-106.
23. Tancik, M., Weber, E., Ng, E., Li, R., Yi, B., Wang, T., ... & Kanazawa, A. (2023, July). Nerfstudio: A modular framework for neural radiance field development. In ACM SIGGRAPH 2023 Conference Proceedings (pp. 1-12).
24. Kim, G., & Cha, Y. (2024). 3D Pixelwise damage mapping using a deep attention based modified Nerfacto. Automation in Construction, 168, 105878.
25. Kim G., and Cha, Y.J. (2025). Deep learning-based 3D image reconstruction and damage mapping using neural radiance fields (Nerfacto), Structural Health Monitoring, SAGE, SHM-24-0736R1, accepted.
26. Kim, G, 2024, Master Thesis, University of Manitoba, 3D Damage Mapping and Segmentation Using Neural Radiance Fields and Advanced Deep Learning Techniques.