Advanced NeRF (ABM-Nerfacto) for High-Definition Digital Twin and Damage Mapping

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

ABSTRACT

Since 2017, extensive damage detection using advanced deep learning models and computer vision techniques has been actively explored. However, efficiently mapping detected damage in 3D digital twin model remains a challenge, as few studies have successfully integrated deep learning and computer vision for precise damage representation. To address this gap, this study investigates an enhanced NeRF-based model, ABM-Nerfacto [1], designed for high-definition and efficient damage mapping in 3D digital twin. This advancement facilitates more effective structural health monitoring, infrastructure maintenance, and a comprehensive pixel-level overview of damage distribution. The Nerfacto model was extensively modified and integrated with an advanced attention mechanism to improve its ability to learn structural features and damage patterns. When applied to a bridge system, the developed model demonstrated exceptional accuracy in pixel-wise damage mapping, successfully generating a high-fidelity 3D digital twin.

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

INTRODUCTION

Traditional image processing techniques for detecting damage in civil infrastructure, as highlighted in prior studies [2-7], suffer from significant limitations, including reliance on manual feature extraction, restriction to single-damage-type detection, and degraded performance under suboptimal image quality. These shortcomings have driven the adoption of deep learning methods [8-11], which have revolutionized damage detection through advancements in image classification [12], object detection [9], autonomous UAV integration [13], and pixel-wise segmentation [14]. Among these, segmentation approaches have garnered particular interest due to their capability for precise damage quantification [11]. Notable convolutional neural network (CNN) architectures such as U-Net [15-16], DenseNet [17], ResNet [18], Mask R-CNN [19], DeepLabV3+ [20-21], and hybrid models [22] have been widely employed for segmentation tasks. However, these general-purpose networks are not optimized for structural damage segmentation, often requiring extensive labeled datasets and incurring high computational costs, which limit their practical applicability [10].

[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

To address these challenges, specialized models tailored for structural damage detection, such as SDDNet [23] and STRNet [24], have been developed. STRNet, in particular, leverages attention mechanisms and optimized loss functions to achieve superior performance, with a mean Intersection over Union (mIoU) of 92.6% and real-time processing capabilities (49 FPS at 1200×800 resolution) [24]. These advancements underscore the potential of domain-specific models to overcome the limitations of general-purpose architectures, offering higher accuracy and efficiency for structural health monitoring (SHM).

Despite these advances in damage detection, effective localization remains a critical challenge. Geotagging techniques have been used to approximate the locations of captured images [11-13], but 2D mapping inherently fails to capture the spatial complexity of structural damage across an entire asset. To address this, 3D reconstruction techniques—such as radar-based methods [25], structure-from-motion (SfM) [26], optical laser triangulation [27], and stereo vision [27]—have been explored. While these methods have shown promise, their application has largely been confined to small-scale scenarios, limiting their scalability for large infrastructure systems.

The emergence of Neural Radiance Fields (NeRF) [28] has marked a significant leap forward in 3D reconstruction, offering the ability to model complex scenes from sparse viewpoints while accounting for intricate lighting conditions. NeRF's ability to generate high-fidelity 3D models has positioned it as a transformative tool for SHM applications. However, its practical deployment is hindered by high computational costs, limited scalability to large scenes, and sensitivity to hyperparameter tuning [28]. To mitigate these issues, the Nerfacto model [29], an optimized variant of NeRF, was introduced to enhance computational efficiency and rendering quality. Recent adaptations of Nerfacto have further demonstrated its potential for generating 3D reconstruction digital twins tailored specifically for structural damage mapping [30].

Building on these developments, the attention-based modified Nerfacto (ABM-Nerfacto) framework [1] represents a novel advancement in 3D damage mapping. By integrating attention mechanisms into the Nerfacto architecture, ABM-Nerfacto enhances the model's ability to focus on critical scene regions, improving both reconstruction accuracy and computational efficiency. When coupled with STRNet's high-accuracy damage segmentation [24], this framework enables precise pixel-wise damage mapping on 3D reconstructed models, offering a holistic view of damage distribution across infrastructure assets. This integrated approach addresses the shortcomings of 2D mapping and traditional 3D reconstruction methods, paving the way for more systematic and automated SHM systems.

However, several challenges remain. The computational demands of NeRF-based models, including ABM-Nerfacto, still pose barriers to real-time applications in large-scale infrastructure monitoring. Additionally, the reliance on high-quality input data and the need for robust hyperparameter optimization continue to limit scalability. Future research should focus on further optimizing computational efficiency, potentially through techniques like model pruning or hardware acceleration, and developing adaptive training strategies to reduce dependency on extensive labeled datasets. Moreover, integrating multi-modal data sources—such as thermal imaging or LiDAR—could enhance the robustness of damage detection and localization under diverse environmental conditions.

The parametric study presented in this conference paper investigates the effectiveness of the ABM-Nerfacto-based 3D damage mapping framework [1], providing valuable insights into its performance across various structural scenarios. By comparing its outcomes with those of traditional methods and other NeRF variants [28-30], this work highlights the potential of attention-based models to transform SHM practices. Ultimately, the continued refinement of such frameworks could lead to fully automated, real-time monitoring systems capable of ensuring the safety and longevity of critical infrastructure assets.

METHOLOGY

This study introduces a 3D damage mapping method that integrates an attention-based modified Nerfacto model for 3D reconstruction with the state-of-the-art STRNet model for pixel-wise damage segmentation. The method aims to provide comprehensive 3D damage distribution visualization for structural health monitoring (SHM). The process, outlined in Figure 1, begins with capturing RGB images of a structure using a handheld camera. These images undergo structure-from-motion (SfM) analysis in Agisoft Metashape (version 2.1) to extract camera parameters. The parameters are then input into the attention-based modified Nerfacto model for 3D reconstruction. For pixel-wise damage mapping, RGB images are first segmented using STRNet to identify damage, and the resulting damage-segmented images are used in the 3D reconstruction. Finally, the Nerfstudio framework [29] renders the pixel-wise damage map onto the 3D model, enabling holistic damage assessment for enhanced SHM automation.

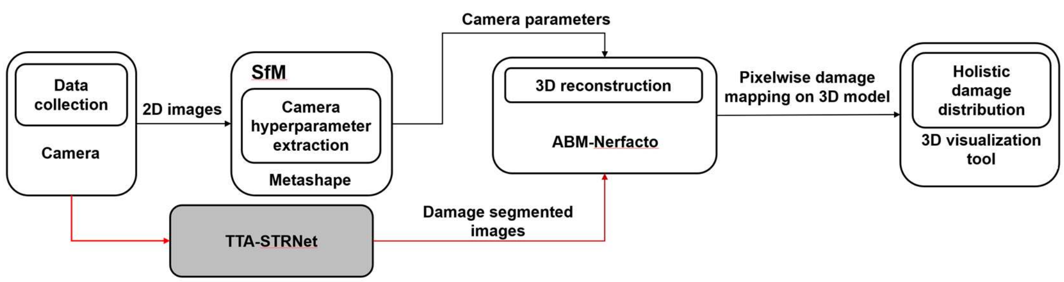

Figure 1: Overall process of the modified Nerfacto based 3D damage mapping method

ABM-NERFACTO

The attention-based modified Nerfacto model builds upon the foundational Neural Radiance Fields (NeRF) framework [28] and the Nerfacto model [29] to enhance 3D reconstruction. The core principle involves training embedded deep attention neural networks (DANNs) to predict opacity and RGB values for 3D sample points along rays corresponding to each pixel in 2D RGB images. These predictions are refined through comparison with ground-truth images during training. As shown in Figure 2, the ABM-Nerfacto model employs proposal sampling [31] to intelligently select 3D sample points, focusing on critical scene regions, unlike the uniform sampling in the original NeRF. Additionally, it incorporates advanced encoding techniques—multiresolution hash encoding [32], spherical harmonics encoding [33], and appearance embedding [34]—to efficiently represent 3D scene data, surpassing the original NeRF's capabilities.

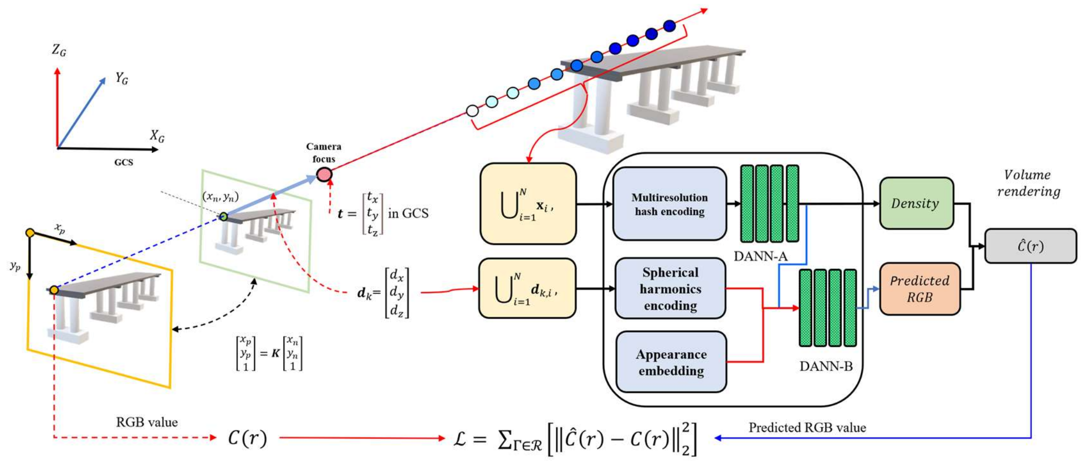

Figure 2: Overall schematic view of ABM-Nerfacto approach

PROPOSAL SAMPLING

The Nerfacto model replaces NeRF's uniform sampling with a two-stage proposal sampling approach, using DANN-α and DNN-β networks to prioritize computationally relevant scene regions, such as object boundaries. From an initial set of 3D points, DANN-α, after multiresolution hash encoding, predicts density values and generates new points via Probability Density Function (PDF) sampling. Subsequently, DANN-β refines these to points, enhancing efficiency compared to NeRF's points sampling.

MULTIRESOLUTION HASH ENCODING

To compactly represent 3D sample point coordinates, ABM-Nerfacto employs multiresolution hash encoding. Voxel grids of varying resolutions are generated, with coordinates hashed into a table using a hash function. Feature vectors are computed via convolution and fed into the deep attention neural network (DANN-A), enabling faster inference while maintaining scene detail, though with potential precision trade-offs for complex lighting.

SPHERICAL HARMONICS ENCODING

Spherical harmonics encoding [33] captures view direction information by converting directional vectors into spherical coordinates and computing harmonic functions. With a degree, this produces a representation of angular dependencies, which is input to DANN-B, enhancing novel view synthesis quality.

APPEARANCE EMBEDDING

Appearance embedding [34] generates a feature vector per image to capture material properties and lighting variations. Combined with spherical harmonics and DANN-A outputs, the resulting vector is processed by DANN-B. A mean squared error loss updates the model pixel-wise, improving adaptability across diverse scene conditions.

DEEP ATTENTION NEURAL NETWORKS

The proposed model introduces DANN-A and DANN-B, each with neurons, four hidden layers, and a multi-head self-attention module (eight heads) at the second layer. DANN-A processes multiresolution-encoded 3D coordinates, producing density and a feature vector. DANN-B integrates this with spherical harmonics and appearance embeddings, outputting RGB values via a sigmoid activation. The resulting 4D vector predicts color and density for each 3D point, enhancing rendering accuracy.

This ABM-Nerfacto model leverages targeted sampling and advanced encodings to improve efficiency and rendering quality for complex 3D scenes, as detailed in subsequent sections.

RESULTS

Building on the case studies presented earlier, the ABM-Nerfacto approach proved effective in generating high-resolution 3D reconstructions of a large-scale bridge system. In addition to reconstruction, this study aimed to map structural damage in 3D using deep learning-based segmentation. Specifically, the STRNet with TTA module was applied. Figure 3 highlights our results by showing segmented cracks accurately overlaid on the 3D bridge model, clearly illustrating the location and extent of damage.

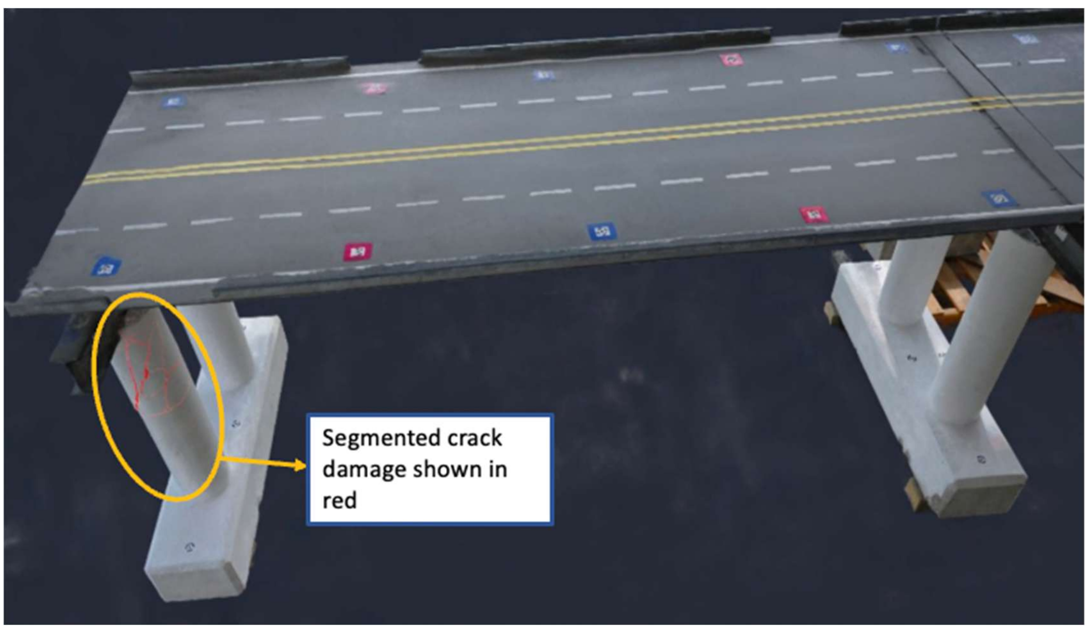

Figure 3: 3D mapping of the segmented cracks

REFERENCES

1. Kim, G., & Cha, Y. (2024). 3D Pixelwise damage mapping using a deep attention based modified Nerfacto. Automation in Construction, 168, 105878.
2. Chen, J. G., Wadhwa, N., Cha, Y. J., Durand, F., Freeman, W. T., & Buyukozturk, O. (2015). Modal identification of simple structures with high-speed video using motion magnification. Journal of Sound and Vibration, 345, 58-71.
3. Shi, J., Xu, X., Wang, J., & Li, G. (2010). Beam damage detection using computer vision technology. Nondestructive Testing and Evaluation, 25(3), 189-204.
4. 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.
5. Cha, Y. J., Chen, J. G., & Büyüköztürk, O. (2017). Output-only computer vision based damage detection using phase-based optical flow and unscented Kalman filters. Engineering Structures, 132, 300-313.

6. Chaiyasarn, K. (2014). Damage detection and monitoring for tunnel inspection based on computer vision (Doctoral dissertation).
7. 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.
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. Cha, Y. J., Ali, R., Lewis, J., & Büyüköztürk, O. (2024). Deep learning-based structural health monitoring. Automation in Construction, 161, 105328.
11. 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.
12. 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.
13. 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.
14. Arafin, P., Billah, A. M., & Issa, A. (2024). Deep learning-based concrete defects classification and detection using semantic segmentation. Structural Health Monitoring, 23(1), 383-409.
15. 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.
16. Pan, G., Zheng, Y., Guo, S., & Lv, Y. (2020). Automatic sewer pipe defect semantic segmentation based on improved U-Net. Automation in Construction, 119, 103383.
17. Alfaz, N., Hasnat, A., Khan, A. M. R. N., Sayom, N. S., & Bhowmik, A. (2022, March). Bridge crack detection using dense convolutional network (densenet). In Proceedings of the 2nd International Conference on Computing Advancements (pp. 509-515).
18. Wang, R., Zhou, X., Liu, Y., Liu, D., Lu, Y., & Su, M. (2024). Identification of the surface cracks of concrete based on resnet-18 depth residual network. Applied Sciences, 14(8), 3142.
19. Yang, F., Huo, J., Cheng, Z., Chen, H., & Shi, Y. (2023). An improved mask R-CNN micro-crack detection model for the surface of metal structural parts. Sensors, 24(1), 62.
20. Shen, Y., Yu, Z., Li, C., Zhao, C., & Sun, Z. (2023). Automated detection for concrete surface cracks based on Deeplabv3+ BDF. Buildings, 13(1), 118.
21. Pan, Z., Zhang, X., Jiang, Y., Li, B., Golsanami, N., Su, H., & Cai, Y. (2025). High-precision segmentation and quantification of tunnel lining crack using an improved DeepLabV3+. Underground Space, 22, 96-109.
22. He, Y., Huang, Z., Liu, D., Zhang, L., & Liu, Y. (2022). A novel structural damage identification method using a hybrid deep learning framework. Buildings, 12(12), 2130.
23. Choi, W., & Cha, Y. J. (2019). SDDNet: Real-time crack segmentation. IEEE Transactions on Industrial Electronics, 67(9), 8016-8025.
24. 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.
25. Gao, J., Liu, J., & Ji, S. (2023). A general deep learning based framework for 3D reconstruction from multi-view stereo satellite images. ISPRS Journal of Photogrammetry and Remote Sensing, 195, 446-461.
26. Zhao, Sizeng, et al. "Structural health monitoring and inspection of dams based on UAV photogrammetry with image 3D reconstruction." Automation in Construction 130 (2021): 103832.
27. Hua, L., Lu, Y., Deng, J., Shi, Z., & Shen, D. (2022). 3D reconstruction of concrete defects using optical laser triangulation and modified spacetime analysis. Automation in Construction, 142, 104469.
28. 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.

29. 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).
30. 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.
31. Barron, J. T., Mildenhall, B., Verbin, D., Srinivasan, P. P., & Hedman, P. (2022). Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 5470-5479).
32. Müller, T., Evans, A., Schied, C., & Keller, A. (2022). Instant neural graphics primitives with a multiresolution hash encoding. ACM Transactions on Graphics (ToG), 41(4), 1-15.
33. Verbin, D., Hedman, P., Mildenhall, B., Zickler, T., Barron, J. T., & Srinivasan, P. P. (2022, June). Ref-nerf: Structured view-dependent appearance for neural radiance fields. In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 5481-5490). IEEE.
34. Martin-Brualla, R., Radwan, N., Sajjadi, M. S., Barron, J. T., Dosovitskiy, A., & Duckworth, D. (2021). Nerf in the wild: Neural radiance fields for unconstrained photo collections. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7210-7219).