DOI: https://doi.org/10.14256/JCE.4500.2026

Građevinar 8/2026

Primljen / Received: 17.1.2026.
Ispravljen / Corrected: 11.5.2026.
Prihvaćen / Accepted: 18.5.2026.
Dostupno online / Available online: 10.9.2026.

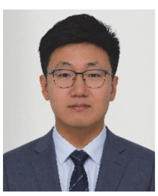
Three-dimensional reconstruction of damaged concrete: Neural radiance field versus explicit methods

Authors:

Shengpeng Zuo, MCE
University of Science and Technology Liaoning, China
School of Civil Engineering
zuoshengpeng@stu.ustl.edu.cn

Xixian Chen, PhD. CE
Ministry of Housing and Urban-Rural Development
Research Institute of Standards and Norms, China
chenxixian1996@163.com

Assoc.Prof. **Bowen Chen**, PhD. CE
Guangxi University, China
School of Civil Engineering and Architecture
zltchenbowen@sina.com



Jiaqi Li, PhD. CE
University of Science and Technology Liaoning, China
School of Civil Engineering
lijiaqi@ustl.edu.cn
Corresponding author

Research Paper

Shengpeng Zuo, Xixian Chen, Bowen Chen, Jiaqi Li

Three-dimensional reconstruction of damaged concrete: Neural radiance field versus explicit methods

To explore the applicability of implicit modelling methods in the three-dimensional (3D) reconstruction of building materials, the neural radiance field (NeRF) algorithm is introduced in this paper to study the effectiveness of 3D reconstruction of images of damaged concrete surfaces, and a comparative analysis is conducted with two traditional explicit reconstruction methods, namely structure from motion and multi-view stereo. Forty-two images of concrete blocks were collected using a smartphone, and four datasets were constructed. Three-dimensional reconstruction and visual rendering were performed on these images. The results show that the peak signal-to-noise ratio of the NeRF model based on instant neural graphics primitives reaches 24.13 dB, outperforming the 17.94 dB of traditional methods. The model also exhibits higher performance in terms of modelling speed and reconstruction details. In addition, this method has obvious advantages in mitigating the surface holes and edge aliasing. The research results verify the feasibility and advantages of NeRF in the visualisation of building materials and provide a new technical path for the 3D modelling and intelligent identification of surface damage in building materials.

Key words:

neural radiance field (NeRF), visualization, 3D reconstruction, intelligent recognition, deep learning

Prethodno priopćenje

Shengpeng Zuo, Xixian Chen, Bowen Chen, Jiaqi Li

Trodimenzionalna rekonstrukcija oštećenog betona: Neuronsko polje zračenja naspram eksplicitnih metoda

Kako bi se ispitala primjenjivost metoda implicitnog modeliranja u trodimenzionalnoj (3D) rekonstrukciji građevnih materijala, uveden je algoritam neuronskog polja zračenja (*Neural Radiance Field* - NeRF) radi procjene učinkovitosti 3D rekonstrukcije slika oštećenih betonskih površina. Provedena je usporedna analiza dvjema tradicionalnim eksplicitnim metodama rekonstrukcije, odnosno metodama strukture iz gibanja (*Structure from Motion* - SfM) i višepogledne stereorekonstrukcije (*Multi-View Stereo* - MVS). Pomoću pametnog telefona prikupljene su 42 slike betonskih blokova te su formirana četiri skupa podataka. Na temelju tih slika provedeni su trodimenzionalna rekonstrukcija i renderiranje modela. Rezultati pokazuju da vršni omjer signal-šum (PSNR) NeRF modela temeljenog na tehnologiji trenutačnih neuronskih grafičkih primitiva (instant neural graphics primitives, Instant-NGP) doseže 24,13 dB, što nadmašuje vrijednost od 17,94 dB ostvarenu tradicionalnim metodama. Model također pokazuje bolje performanse u pogledu brzine rekonstrukcije i kvalitete rekonstrukcije detalja. Osim toga ta metoda ima izražene prednosti u smanjenju diskontinuiteta površine i nazubljenosti rubova. Rezultati istraživanja potvrđuju izvedivost i prednosti NeRF-a u vizualizaciji građevnih materijala te pružaju novi tehnološki pristup za 3D modeliranje i automatiziranu identifikaciju površinskih oštećenja građevnih materijala.

Ključne riječi:
neuronsko polje zračenja (Neural Radiance Field, NeRF), vizualizacija, 3D rekonstrukcija, inteligentno prepoznavanje, duboko učenje

1. Introduction

As the service life of buildings increases, various types of damage, such as wall spalling and crack propagation, inevitably occur on building surfaces. Various intelligent detection and assessment methods have been proposed to monitor the development of apparent damage in a timely manner.

Among the various intelligent technologies, computer vision technology based on deep learning has been developed rapidly in recent years and used in numerous applications in the fields of construction, medicine, transportation, and safety [1-4]. However, traditional two-dimensional image analysis suffers from single perspectives and insufficient information, which makes it difficult to comprehensively assess the spatial morphology and damage distribution of buildings. Three-dimensional (3D) reconstruction restores structural forms with high precision through multi-view fusion, supports observation and damage assessment from arbitrary perspectives, and provides a global analytical basis for safety diagnoses.

The development of 3D reconstruction technology can be traced back to the structure from motion (SfM) differential mathematical framework in Horn's [5] seminal book, "Robot Vision", published in 1986, which laid the foundation for many vision algorithms. Currently, 3D reconstruction methods primarily include structured light [6], stereo vision [7], SfM [8], multi-view stereo (MVS) [9], laser scanning [10], and the implicit reconstruction algorithm that has emerged in recent years, namely, neural radiance field (NeRF) technology [11].

The explicit 3D reconstruction algorithms proposed in earlier years have significantly promoted the development of the discipline and laid an important foundation for the field of 3D vision. In the field of construction engineering, 3D reconstruction can capture the spatial characteristics of concrete structural surfaces, including the geometric morphology of cracks and volumetric changes in spalling areas, more comprehensively compared to 2D image analysis, providing a more intuitive basis for damage assessment and subsequent repair. In recent years, many studies have demonstrated that 3D reconstruction technology plays a significant role in the automatic detection and volumetric quantification of concrete damage. Three-dimensional reconstruction technology based on multiview images generates sparse point clouds through SfM and MVS, which can provide basic geometric data for damage detection. For example, Kim et al. [12] proposed an automated concrete crack assessment method combining dual-focal-length stereo vision, which demonstrated significant advantages in depth image processing. Beckman et al. [13] combined a Faster R-CNN with depth cameras to achieve automatic quantification of concrete spalling volume. Liu et al. [14] used smartphones to capture multiple photos around concrete members and employed MVS methods for 3D reconstruction. The reconstruction quality of this method was superior to that of depth camera methods. However, explicit 3D reconstruction algorithms still have several limitations [15]: first, explicit representations struggle to handle complex topological structures in geometric expressions and are prone to holes or distortions when modelling thin-walled or fine features; second, reconstruction quality is sensitive to the integrity of input data and noise, easily leading to missing geometries when perspectives are insufficient or occlusions exist; third, high-precision explicit reconstruction is resource-intensive, affecting reconstruction efficiency. To overcome the limitations of traditional explicit methods, researchers have recently begun studying 3D reconstruction methods based on deep learning. Among these, NeRF, a 3D reconstruction algorithm based on implicit neural networks, has brought about revolutionary changes in the field. The NeRF algorithm is based on fully connected neural networks and parameterises a scene as a continuous radiance field. This method trains a multilayer perceptron (MLP) network through multi-view images to implicitly learn the colour and volume density of 3D points using volume, rendering technology to synthesise new views. Its implicit expression can not only accurately restore fine textures and light-shadow variations in complex scenes but also avoid common problems such as the holes and aliasing found in traditional mesh reconstruction methods. NeRF technology has been successfully applied in multiple fields. Peng et al. [16] provided robust geometric and semantic support based on improved NeRF algorithms, thereby providing a powerful technical solution for automated operations in smart agriculture. Wen et al. [17] utilised view-enhancement modules and deep supervision strategies to effectively overcome the problem of blurred geometric features in traditional NeRF models for large-scale scenes, providing a solution that is superior to traditional explicit modelling for building high-fidelity digital campuses and geographic environments.

In the field of structural engineering, the potential of NeRF models for 3D reconstruction and structural inspection remains unexplored. We expect that the high-precision and high-rendering quality characteristics of the NeRF algorithm can be applied to enhance the 3D reconstruction quality of building engineering structures, further improving the identification and visualisation of damaged areas and providing more intuitive and accurate data support for structural health monitoring and subsequent maintenance. To this end, this study focuses on damaged concrete blocks as the research object, employs the neural radiance field method for 3D reconstruction, and introduces reconstruction quality evaluation metrics, including the peak signal-to-noise ratio (PSNR) [18], structural similarity index (SSIM) [19], and learned perceptual image patch similarity (LPIPS) [20] to systematically compare and analyse the performance of the NeRF and traditional explicit 3D reconstruction algorithms in building material reconstruction. The evaluation is conducted based on three aspects: reconstruction accuracy, detail restoration capability, and model completeness.

2. Method overview

In this study, two types of NeRF algorithms (based on NeRF-PyTorch and instant neural graphics primitives (Instant-NGP) [21]) and a traditional algorithm based on explicit image reconstruction (utilising the SfM and MVS frameworks with the ContextCapture Center Master tool) were adopted to perform high-precision 3D reconstruction and comparative analysis of damaged concrete specimens. A comprehensive performance evaluation was conducted using metrics such as the PSNR, SSIM, and LPIPS. The complete technical workflow is illustrated in Figure 1 and described below.

Images of the damaged concrete specimens were acquired using a Xiaomi 12 smartphone. These multi-view RGB images served as the common input for all three reconstruction branches. In the explicit 3D reconstruction branch illustrated in the upper section of Figure 1, multi-view images were processed using the ContextCapture Center Master platform. This workflow commenced with feature extraction and matching to identify corresponding feature points from different viewpoints. Subsequently, the SfM algorithm was utilised for camera pose estimation and sparse point-cloud generation, followed by the application of the MVS algorithm to compute a dense point cloud. Finally, mesh reconstruction and texture mapping were performed to obtain a 3D model with realistic surface textures of the damaged concrete specimen. In the NeRF-PyTorch branch, as shown in the lower left and lower right of Figure 1, COLMAP [22] was first employed for feature extraction, camera pose estimation, and sparse point-cloud generation. Subsequently, the imgs2poses.py script was used to convert the generated pose information into a standard local light field fusion (LLFF) format for network ingestion. During the model training phase, a 360° inward-facing training strategy was adopted, and an MLP was used for the end-to-end fitting of the 5D neural radiance field. Using spatial coordinates and viewing directions as inputs to predict the volume density and view-dependent colour, continuous and high-resolution rendered views of the damaged concrete surface were generated through volume rendering. In the Instant-NGP branch, illustrated in the lower right of Figure 1, the COLMAP outputs were parsed and converted into the transforms.json file format, which is recognisable by the Instant-NGP framework. The input 3D spatial coordinates and viewing directions were first processed via multiresolution hash encoding for efficient feature mapping and subsequently fed into a lightweight MLP network to predict the volume density and RGB colour of arbitrary points within the scene. Compared to the standard NeRF-PyTorch baseline, this branch accelerated the training speed significantly while achieving high-fidelity real-time rendering.

Finally, by rendering the 3D model results from identical viewpoints, we systematically compared the performances of the two types of NeRF algorithms against traditional explicit 3D reconstruction methods using three metrics: PSNR, SSIM, and LPIPS. The comparison focused on reconstruction accuracy, detailed restoration capability, and adaptability to complex surface morphologies.

2.1. Datasets

To ensure the reliability and generalisability of the algorithm evaluation, we selected the official LLFF dataset [23], which features high-frequency details, as a benchmark platform for verifying the three algorithms systematically. The Fern dataset was captured using a camera moving along a circular trajectory, forming perspectives surrounding the target object. It contained 20 images with a resolution of 4032×3024 pixels and a file size of approximately 3 MB. Further details are shown in Figure 2.

Secondly, image acquisition of damaged concrete for the 3D reconstruction dataset was conducted using a smartphone via an overhead circular image acquisition method above the damaged concrete block. A total of 42 images were captured. Each image had a file size of approximately 3 MB and a resolution of 3072×4096, and the images comprehensively covered the full-range damage characteristics of the concrete specimen. Additional details are shown in Figure 3.

Adopting an arithmetic extraction strategy, four datasets of different scales were constructed by extracting 42, 32, 22, and 12 images, respectively, from the 42 images. During the extraction process, images were selected at regular intervals to ensure a uniform angular distribution across different datasets, thereby guaranteeing a comprehensive representation of the damage conditions of the block. Upon completion of the extraction, each dataset was systematically organised, the image format was converted to PNG, and the image files were renamed.

The two datasets served different purposes; therefore, they were acquired using different tactics, image counts, and resolutions. The Fern dataset served as a public benchmark to validate the algorithmic pipelines; thus, its original image settings under a circular camera trajectory were strictly retained to prevent preprocessing bias and ensure a fair comparison. Conversely, the damaged concrete dataset contained images representing real-world structural scenes and served as the core engineering dataset in this study to evaluate the applicability of implicit and explicit reconstruction methods to concrete surface damage. An overhead circular trajectory with denser view sampling was adopted to guarantee sufficient viewpoint overlap and to fully

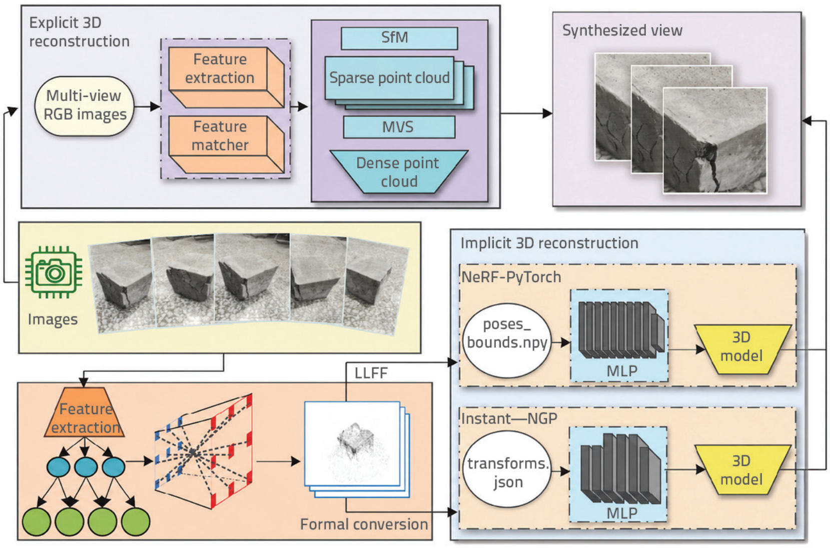


Figure 1. Overall pipeline of multi-view image-based 3D reconstruction and neural rendering

1. Camera Pose Estimation: By analysing the spatial displacement of matched feature points across multiple images, the algorithm computes the exact position, orientation, and intrinsic lens parameters of the camera when each image is captured.
2. Triangulation: This step determines the actual 3D spatial coordinates of the corresponding feature points by calculating the intersections of the viewing rays originating from different camera poses.
3. Sparse Point Cloud: The 3D coordinates obtained through triangulation are aggregated to construct an initial sparse point cloud. However, because of inherent inaccuracies in feature matching, minor positional errors are inevitable at this stage.

capture complex surface damage topologies. Furthermore, image resolution directly affects both detail restoration and computational overhead. While higher resolutions preserve crucial high-frequency damage textures that are vital for accurate reconstruction, they significantly increase memory consumption and training time. Therefore, the resolution of 3072 × 4096 for the concrete dataset was chosen to strike an optimal balance between retaining high-fidelity damage details and maintaining computational feasibility.

2.2. 3D Reconstruction algorithms

2.2.1. Explicit 3D reconstruction algorithms

In multi-view stereo reconstruction tasks, explicit modelling methods typically employ a two-stage processing workflow: SfM and MVS. A comprehensive visual representation of this step-by-step framework is shown in Figure 4. As illustrated by the numbered modules in the diagram, this process involves four steps.

4. Bundle Adjustment: To address these initial errors, a global nonlinear optimisation mechanism called bundle adjustment is applied. This process relies on the camera forward projection model, which maps a 3D spatial point to a 2D image plane and is expressed as

$$\pi_i : \mathrm{R}^3 \rightarrow \mathrm{R}^2, \mathrm{X}_j \rightarrow \pi_i\left(\mathrm{X}_j\right) = \mathrm{normalize}\left(\mathrm{K}_i\left(\mathrm{R}_i\,\mathrm{X}_j + \mathrm{t}_i\right)\right) \tag{1}$$

where π_i denotes the projection function of i camera; $\mathbf{K}_i$ is the camera intrinsic matrix; $\mathbf{R}_i$ and $\mathbf{t}_i$ are the extrinsic rotation matrix and translation vector, respectively; and $\mathbf{X}_j$ represents the j 3D point. The bundle adjustment simultaneously refines the estimated camera poses and 3D coordinates of the sparse points to minimise the reprojection errors, thereby ensuring the overall geometric consistency of the model.

Finally, the workflow generates a dense point cloud that serves as the final 3D reconstruction result. Although the

Figure 2. Fern dataset

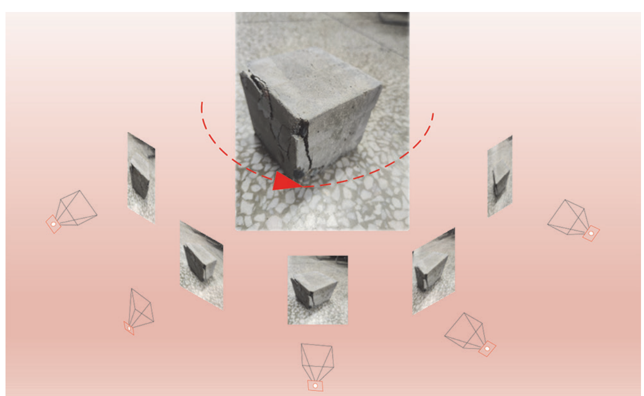


Figure 3. Damaged concrete block dataset

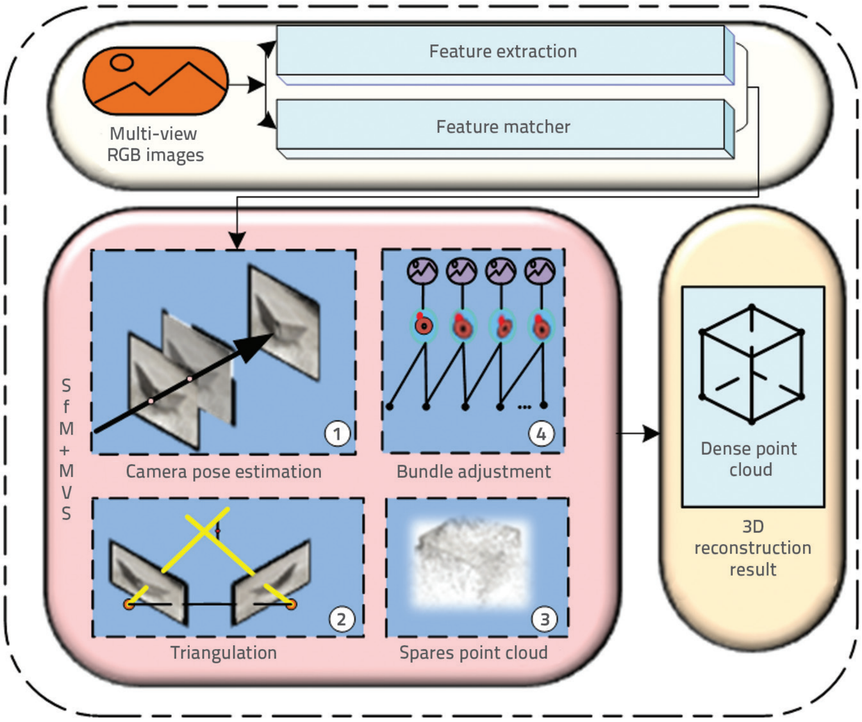


Figure 4. Traditional SfM-MVS reconstruction framework

sparse point cloud provides a spatial framework, it lacks the surface resolution necessary for thorough material inspection. The MVS adds continuous surface details to this skeletal model by conducting depth searches and calculating the depth of uniform concrete surfaces. This effectively fills the structural gaps between sparse points, converting the basic skeleton into a highly detailed dense point cloud that faithfully captures the specific microtextures, volume losses, and intricate morphologies of the damaged surface.

2.2.2. Implicit 3D reconstruction algorithms

Implicit reconstruction utilises implicit functions to render 3D models, eliminating the need to construct meshes or geometric structures directly while demonstrating flexibility and robust data adaptability. The input for the NeRF consists of sparse multiview RGB images, along with the camera's intrinsic and extrinsic parameters, which describe the camera's viewing direction. Model reconstruction requires capturing images from multiple angles to ensure sufficient overlap of details in every region. The necessary camera parameters, including the focal length, optical centre, and camera position and orientation, were obtained using the COLMAP software. Rays are generated by back-projecting the image pixels into a three-dimensional space using the intrinsic and extrinsic parameters of the camera. Each ray represents a 3D spatial trajectory corresponding to a pixel. The ray equation is as follows:

$$\mathbf{r}(t) = \mathbf{o} + t\mathbf{d} \tag{2}$$

where "o" is the optical centre of the camera, "d" is the unit direction vector of the ray, and "t" is the distance from o to r(t) along the viewing direction "d".
N points are uniformly sampled along each ray to obtain the three-dimensional coordinates (x, y, z). To enhance the high-frequency representation capability of the model, positional encoding is performed on the coordinates (x, y, z) of the sampled points to map them onto a high-dimensional space:

$$\gamma(x) = [\sin(2^0\pi x), \cos(2^0\pi x), \ldots, \sin(2^{L-1}\pi x), \cos(2^{L-1}\pi x)] \tag{3}$$

where L is the frequency series of the high-dimensional space.

Positional encoding enables the MLP to capture high-frequency spatial details such as surface continuity and texture [24]. The MLP predicts the scene volume density and point distribution using three-dimensional coordinates (x, y, z). The output volume density is subsequently used in volume rendering to calculate the transparency and weights, thereby determining the final pixel colour contribution of the rays passing through these points. To explicitly map the network architecture illustrated in Figure 5, the standard NeRF MLP employs a decoupled architecture to sequentially process the geometry and appearance. First, 3D spatial coordinates (x, y, z) are mapped onto a 63-dimensional high-frequency feature vector via positional encoding. This vector is fed into the primary network, which consists of eight fully connected layers, each with 256 channels. To preserve the spatial information and prevent feature degradation, a skip connection is introduced in the 5th layer, concatenating the initial encoded input with intermediate features. The output of this 8-layer network branch provides the view-independent volume density (σ) and an intermediate feature vector.
Subsequently, to generate realistic, view-dependent rendering results, the viewing directions (θ, ϕ) are mapped into a 24-dimensional vector through positional encoding. This direction vector is concatenated with an intermediate feature vector and processed by a final fully connected layer with 128 channels. This specific structural design ensures that the volume density (σ) relies solely on spatial coordinates, while the final predicted RGB colour can change dynamically based on the viewing angle, rendering surface textures and view-dependent appearance with precision.

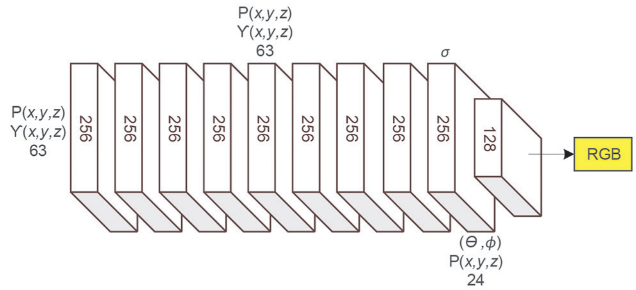


Figure 5. MLP architecture

As a core technology of NeRF, volume rendering derives pixel colours by accumulating the predicted values from all the sampled points. Volume rendering, based on an optical model, integrates these predicted values into pixel colours by simulating the propagation of light through a medium. The physical formula governing this process is given by (4) and (5):

$$C(r) = \sum_{i=1}^{N} T_i \left(1 - \exp\left(-\sigma_i \delta_i\right)\right) c_i \tag{4}$$

$$T_i = \prod_{j=1}^{i-1} \exp\left(-\sigma_j \delta_j\right) \tag{5}$$

where **r** is the direction vector of the ray emitted by the camera; **C**(**r**) is the volume rendering result; $\mathbf{c}_i$ is the colour of sampled point *i* ; T_i is the transparency from the ray origin to point *i*; $\exp(-\sigma_i\delta_i)$ is the light absorption rate at point *i;* δ_i is the distance between adjacent sampled points; and σ_i is the volume density.

Furthermore, to enhance the rendering speed and image quality, the NeRF incorporates a hierarchical sampling strategy that is primarily divided into coarse and fine sampling. In the coarse sampling stage, *N* points are sampled uniformly along the ray to obtain the coarse density and colour distribution values. In the fine sampling stage, based on the weight distribution obtained from the coarse sampling phase, the NeRF concentrates the sampling of *N'* points in the high-density regions, thereby capturing scene details more precisely. To optimise the MLP parameters within the neural radiance field, the NeRF introduces a loss function based on the colour difference to minimise the discrepancy between the predicted and ground-truth colours. Through back-propagation, the network parameters are adjusted continuously to gradually improve the realism and overall synthesis quality of the rendered images.

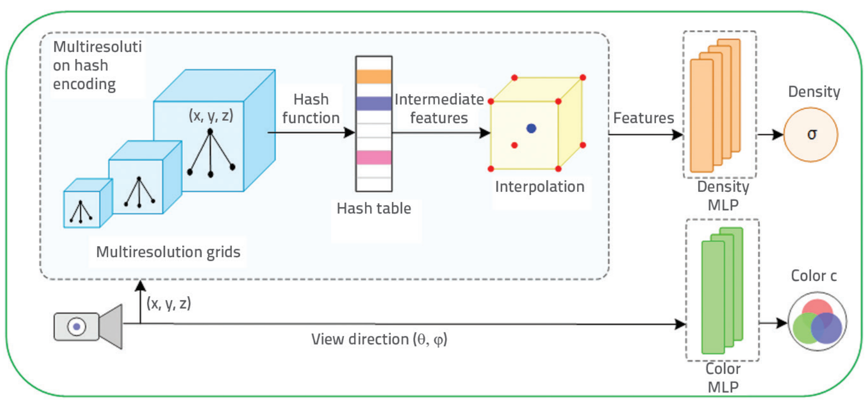


Figure 6. Architecture of Instant-NGP

In this study, two types of NeRF algorithms were used. NeRF-PyTorch is a traditional NeRF algorithm for 3D reconstruction implemented based on the PyTorch framework. Although this method shows excellent performance in terms of detail fidelity and novel view synthesis, it suffers from limitations such as long training times, slow inference speeds, and weak adaptability to large-scale scene reconstruction, leading to the emergence of numerous variants. Subsequently, the Instant-NGP algorithm with multiresolution hash encoding was proposed in 2022. This method utilises multiresolution hash encoding to map the input spatial coordinates into high-dimensional features and introduces a dual-branch lightweight MLP structure. Specifically, it employs a density MLP to predict volume density and geometric features and subsequently decodes RGB colours using a colour MLP in combination with viewing directions. Combined with highly optimised GPU-accelerated volume-rendering techniques, this algorithm reduces the spatial parameter count and computational overhead significantly, achieving real-time high-fidelity reconstruction, as shown in Figure 6. This method possesses advantages such as fast training speed, low memory footprint, and significantly improved inference speed while maintaining rendering quality.

3. Results: comparison and analysis of model performance

3.1. Model training

All the experiments were performed on a workstation running Windows 11, equipped with an NVIDIA GeForce RTX 4090 GPU (24 GB VRAM), an Intel Core i9-13900K processor, and 32 GB of RAM. The software environment included CUDA version 12.6 and driver version 560.94. Both the model training and inference were conducted using the PyTorch 2.6 framework, as detailed in Table 1.

For the NeRF-PyTorch algorithm, the 'fern' scene from the public LLFF dataset, consisting of 20 images, and the self-constructed damaged concrete specimen dataset, comprising

Table 1. Experimental environment and dependency library version information

Environment	Python	OpenCV-python	CUDA	PyTorch	TorchVision	TorchAudio	Matplotlib	tqdm
Version	3.9	4.11	12.6	2.6	0.21	2.6	3.9.4	4.65.2

subsets of 12, 22, 32, and 42 images, were selected. The training parameters were uniformly configured as follows: factor = 8, label = 8, N_rand = 1024, N_samples = 64, N_importance = 64, use_viewdirs = true, and raw_noise_std = 1e0. All datasets were trained for 200,000 iterations, and 120 rendering evaluations were conducted on the test set. The training parameters for the Instant-NGP platform were set as follows: batch size = 262,144, rays/batch = 3,072, samples/ray = 90.83, and seed, 1,337. Using the aforementioned datasets, the models were trained until convergence (approximately 10,000 iterations). The same testing procedure was used to further compare the performance and speed of the two neural rendering frameworks. For comparison, the ContextCapture Center Master (CCCMaster) platform, which is based on explicit reconstruction, was used to perform complete 3D reconstruction on the same datasets. This process included camera calibration, sparse point-cloud generation, dense point-cloud reconstruction, and meshed surface modelling.

3.2. Comparison on public datasets

The performance of the algorithms was analysed based on three aspects: reconstruction quality, training time, and NeRF reconstruction effectiveness under varying numbers of images. Both the public Fern dataset and the self-constructed damaged concrete specimen dataset were used to evaluate the performance of each algorithm.

Figure 7. Comparison of reconstruction performance of different algorithms on public the benchmark Fern dataset; zoomed-in views of local details are shown in the bottom-right corners: a) Real image; Reconstruction effect of :b) explicit algorithm; c) NeRF-PyTorch; d) Instant-NGP

As shown in Figure 7, to facilitate a clear comparison of fine details, zoomed-in patches are provided in the bottom-right corner of each image. Compared with NeRF algorithms, traditional explicit reconstruction methods (MVS and SfM) exhibit significant artefacts, texture distortion, and partial hollowing on the Fern dataset. Notably, as shown in Figure 7.b, explicit algorithms struggle with textureless or complex regions, resulting in a severe loss of background information. They performed particularly poorly in terms of model completeness; for instance, defects, such as missing edges and artefacts, became significantly more pronounced. In contrast, NeRF-PyTorch and Instant-NGP effectively synthesised the complete background and overcame issues related to model holes, significantly improving model completeness and detail rendering in the zoomed-in areas.

3.3. Comparison of 3D reconstruction of concrete surface damage

Figure 8 illustrates the 3D reconstruction results of various algorithms on the damaged concrete specimen dataset. The results indicate that in scenarios with a small number of images, which leads to limited feature points being available to the algorithms, the NeRF algorithm maximises the preservation of model completeness; however, its performance slightly lags that of the explicit algorithms in capturing the apparent surface damage details of the concrete specimens. As the number of images increased, the 3D restoration fidelity of the NeRF algorithm stabilised earlier than that of the explicit algorithms, and the NeRF algorithm demonstrated excellent photorealism in rendering concrete surfaces. In comparison with the explicit algorithm CCCMaster, which approximates the NeRF algorithm in terms of comprehensive details, integrity, and realism, the NeRF algorithm still exhibits deficiencies in colour reproduction.

Figure 9 shows a comparison of the reconstruction results between the NeRF-PyTorch and Instant-NGP algorithms for damaged concrete specimens under low-illumination conditions. The experimental results demonstrate that Instant-NGP exhibits superior feature extraction capabilities when processing images captured under low light. Compared with NeRF-PyTorch, it preserves the high-frequency texture details of the concrete surface more completely and effectively reduces artefacts induced by environmental noise. This result confirms that NeRF technology maintains robust geometric reconstruction capabilities even in complex low-light scenarios.

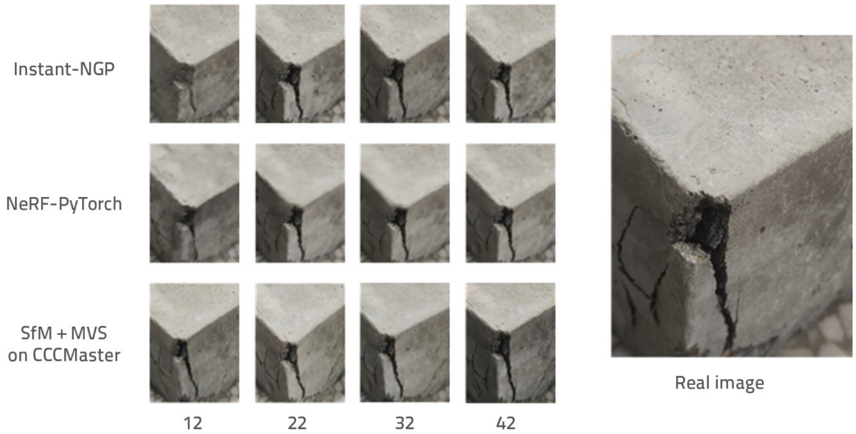


Figure 8. Comparison of reconstruction performance of various algorithms on damaged concrete specimen datasets (Note: 12, 22, 32, and 42 indicate the number of processed images)

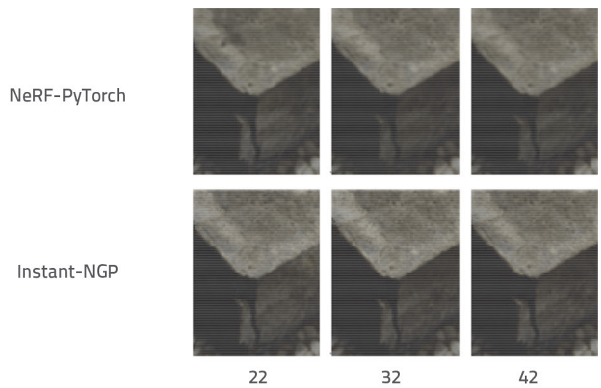


Figure 9. NeRF reconstruction performance under low light conditions (Note: 12, 22, 32, and 42 indicate the number of processed images)

Table 2 presents the performances of various 3D reconstruction algorithms on the concrete specimen dataset, demonstrating that the NeRF algorithms exhibit significant advantages across multiple key metrics. Specifically, the PSNR of the NeRF method showed a substantial improvement of 19.5% over that of traditional explicit methods. With an input of 42 images, the PSNR reached 24.13 dB, verifying its advantages in maintaining structural consistency and preserving texture details. It is particularly noteworthy that the NeRF algorithm excels in high-frequency detail recovery. The LPIPS of Instant-NGP reached 0.42, significantly outperforming the 0.52 of the combined SfM and MVS algorithms, indicating its enhanced capability to preserve fine features. In terms of computational efficiency, Instant-NGP completed the reconstruction in only 1 min 30 s, achieving a speedup of 112 times that of the NeRF-PyTorch method and a 5.5-fold improvement over the combined SfM and MVS algorithms. This characteristic confers significant advantages to the NeRF method for practical applications. Experimental data also revealed that, when the number of input images increased from 22 to 42, the PSNR of Instant-NGP increased from 22.83 dB to 24.13 dB, an increase of 5.4%, whereas NeRF-PyTorch showed an increase of 1.3%. This further confirms the robustness of NeRF with increasing data volume. Conversely, for the combined SfM and MVS algorithms, as the dataset size increased, the overall structure of the reconstructed model may have become more complex, which led to a decrease in PSNR values.

4. Discussions

4.1. Contribution of this study to the body of the knowledge

1. A quantitative evaluation benchmark is established for the performance of various models for the 3D reconstruction of damaged concrete. To investigate the

Table 2. Quantitative evaluation of reconstruction quality on the damaged concrete block dataset

Method	Evaluation Metrics	42 Multi-view Images	32 Multi-view Images	22 Multi-view Images
SfM + MVS	PSNR↑	17.94	18.87	18.93
	LPIPS↓	0.36	0.41	0.39
	SSIM↑	0.60	0.64	0.63
NeRF-PyTorch	PSNR↑	20.20	20.68	20.47
	LPIPS↓	0.53	0.51	0.52
	SSIM↑	0.71	0.72	0.71
Instant-NGP	PSNR↑	24.13	22.56	22.83
	LPIPS↓	0.43	0.43	0.42
	SSIM↑	0.69	0.70	0.68

differences between explicit and implicit 3D reconstruction algorithms, the comprehensive performance of implicit methods (Instant-NGP and NeRF-PyTorch) and the explicit method (MVS-SfM) in the 3D reconstruction of damaged concrete surfaces was systematically evaluated. The results demonstrate that, on the same image dataset, Instant-NGP achieved PSNR, LPIPS, and SSIM of 24.13 dB, 0.43, and 0.69, respectively. Furthermore, it completed the training and rendering process in 1 min and 30 s, thereby outperforming other algorithms in terms of both reconstruction quality and computational efficiency.

2. This study validated the reconstruction performance of implicit 3D algorithms for damaged concrete surfaces. Compared with the model holes and edge-aliasing artefacts commonly observed in traditional explicit methods, the NeRF algorithm effectively restores high-frequency details, such as surface pores and crack trajectories, within the damaged areas. In particular, aided by multiresolution hash encoding technology, Instant-NGP successfully captures complex physical textures while achieving rapid minute-level reconstruction. It also demonstrates robust performance when addressing low-light and data-constrained scenarios.
3. This study provides a feasible technical approach for the refined quantitative evaluation and digital twin modelling of structural damage. Research indicates that high-precision 3D models generated by the implicit NeRF rendering framework not only meet the practical demands for rapid on-site feedback in engineering but also provide reliable data support for subsequent automated crack width measurement and volumetric quantification. This study validated the feasibility and advantages of NeRF in representing surface damage to building structures, thereby establishing a reliable theoretical foundation and technical pathway for refined civil engineering digital twin modelling [25] and structural health monitoring.

4.2. Limitations and possible future research directions

Although this study validated the advantages of implicit neural representations for reconstructing damaged concrete surfaces, several limitations and future research directions remain:

1. This study evaluated the 3D reconstruction performance based on metrics such as PSNR, demonstrating the feasibility of the NeRF algorithm. Meanwhile, a geometric quantification of damaged regions in the physical space is necessary. Therefore, future research will further analyse 3D models to measure the physical dimensions of the damage. Additionally, high-precision laser-scanning equipment will be introduced as a ground-truth benchmark to verify the accuracy of the NeRF algorithm in geometric quantification tasks.
2. This paper proposes a 3D reconstruction method for damaged concrete specimens based on the NeRF algorithm, verifying the effectiveness of implicit neural representations in model rendering. Future research will extend beyond the laboratory to construct field datasets of actual scenes to enable this method to better reflect the physical scale and environmental complexity of real-world infrastructures. Concurrently, to address the high hardware requirements of the current NeRF algorithms, future work will focus on developing lightweight NeRF architectures to facilitate their application in engineering inspections.
3. This study verified the effectiveness of the NeRF method in restoring the surface details of damaged concrete blocks. To achieve automated evaluation and quantification of defects from a 3D spatial perspective, future work will focus on integrating concrete damage instance segmentation algorithms with the NeRF reconstruction framework. The planned approach involves capturing multiview images of actual engineering scenes and utilising instance segmentation algorithms to extract 2D damage masks. These masks will then be introduced as prior information into the NeRF model to construct a 3D representation that fuses geometric and semantic features, ultimately realising 3D visual recognition and the quantitative measurement of structural damage.

5. Conclusions

We conducted a comparative study of the 3D reconstruction of damaged concrete surfaces. The advantages of the NeRF implicit representation in 3D model rendering were validated through a systematic comparison with the traditional explicit methods (SfM and MVS). When addressing complex concrete specimen textures and data-constrained scenarios, traditional explicit methods exhibit surface blurring, aliasing artefacts, and loss of background information. In contrast, the implicit algorithm successfully overcame these limitations, effectively eliminating artefacts and achieving high-fidelity restoration of surface pores and crack propagation, thereby demonstrating better performance in terms of model completeness and visual fidelity. In the comprehensive evaluation of reconstruction accuracy and efficiency, the Instant-NGP model based on multiresolution hash encoding demonstrated a favourable balance. Using the same image dataset, this model not only achieved reconstruction quality with a PSNR of 24.13 dB, an LPIPS of 0.43, and an SSIM of 0.69 but also completed the entire training and rendering process in 1 min 30 s.

This capability, which combines high-precision detail rendering with rapid, minute-level inference, can satisfy the practical requirements for rapid 3D modelling feedback during on-site engineering inspections.

Although the implicit reconstruction method proposed in this study effectively restores the geometric details of damaged concrete specimens, the current evaluation is primarily based on the quality of the visual synthesis. Absolute geometric quantification in the physical space has not yet been conducted, nor has this method been effectively integrated into automated detection algorithms. Therefore, future research should focus on developing instance segmentation models specific to concrete surface damage and tightly integrating them with the NeRF reconstruction framework. This method aims to achieve automatic extraction and precise quantitative assessment of structural defects from a 3D spatial perspective.

Acknowledgments

This research was supported by the Liaoning Provincial Joint Science and Technology Program (2025-BSLH-185) and the National Natural Science Foundation of China (Grant No. 52408163). The contributions of all the participants in this study are gratefully acknowledged.

REFERENCES

[1] Cha, Y., Choi, W., Büyüköztürk, O.: Deep learning-based crack damage detection using convolutional neural networks, Computer-Aided Civil and Infrastructure Engineering, 32 (2017) 5, pp. 361-378, https://doi.org/10.1111/mice.12263/

[2] Kwiatkowska, D., Kluska, P., Reich, A.: Convolutional neural networks for the detection of malignant melanoma in dermoscopy images, Postepy Dermatol Alergol, 38 (2021) 3, pp. 412-420, https://doi.org/10.5114/ada.2021.107927.

[3] Kheder, M.Q., Mohammed, A.A.: Real-time traffic monitoring system using IoT-aided robotics and deep learning techniques, Kuwait Journal of Science, 51 (2024) 1, pp. 100153, https://doi.org/10.1016/j.kjs.2023.10.017.

[4] Li, J.Q., Zhao, X.F., Zhou, G.G., et al.: Standardized use inspection of workers' personal protective equipment based on deep learning, Safety Science, 150 (2022), pp. 105689, https://doi.org/10.1016/j.ssci.2022.105689.

[5] Horn, B.K.P.: Robot Vision, MIT Press, Cambridge, 1986.

[6] Geng, J.: Structured-light 3D surface imaging: a tutorial, Advances in Optics and Photonics, 3 (2011) 2, pp. 128-160, https://doi.org/10.1364/AOP.3.000128.

[7] Zhang, Z.: A flexible new technique for camera calibration, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22 (2000) 11, pp. 1330-1334, https://doi.org/10.1109/34.888718.

[8] Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: exploring photo collections in 3D, ACM Transactions on Graphics, 25 (2006) 3, pp. 835-846, https://doi.org/10.1145/1141911.1141964.

[9] Seitz, S.M., Curless, B., Diebel, J., et al.: A comparison and evaluation of multi-view stereo reconstruction algorithms, 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, 2006: pp. 519-528, https://doi.org/10.1109/CVPR.2006.19.

[10] Besl, P.J., McKay, N.D.: A method for registration of 3-D shapes, IEEE Transactions on Pattern Analysis and Machine Intelligence, 14 (1992) 2, pp. 239-256, https://doi.org/10.1109/34.121791.

[11] Mildenhall, B., Srinivasan, P.P., Tancik, M., et al.: NeRF: Representing scenes as neural radiance fields for view synthesis, Computer Vision - ECCV 2020, Cham, 2020: pp. 405-421, https://doi.org/10.1007/978-3-030-58452-8_24.

[12] Kim, H., Sim, S.H., Spencer, B.F.: Automated concrete crack evaluation using stereo vision with two different focal lengths, Automation in Construction, 135 (2022), pp. 104136, https://doi.org/10.1016/j.autcon.2022.104136.

[13] Beckman, G.H., Polyzois, D., Cha, Y.J.: Deep learning-based automatic volumetric damage quantification using depth camera, Automation in Construction, 99 (2019), pp. 114-124, https://doi.org/10.1016/j.autcon.2018.12.006.

[14] Liu, C.C., Zhou, L., Wang, W.W., et al.: Concrete surface damage volume measurement based on three-dimensional reconstruction by smartphones, IEEE Sensors Journal, 21 (2021) 10, pp. 11349-11360, https://doi.org/10.1109/JSEN.2021.3067739.

[15] Zhu, F.: 3D scene representation: A review of recent achievements in neural radiance fields (NeRF), Journal of Communication University of China (Science and Technology), 29 (2022), pp. 64-77, https://doi.org/10.16196/j.cnki.issn.1673-4793.2022.05.007.

[16] Peng, H.X., Guo, S.K., Zou, X.J., et al.: UAVO-NeRF: 3D reconstruction of orchards and semantic segmentation of fruit trees based on neural radiance field in UAV images, Computers and Electronics in Agriculture, 237 (2025), pp. 110631, https://doi.org/10.1016/j.compag.2025.110631.

[17] Wen, X., Sun, K., Chen, T., et al.: A NeRF-based technique combined depth-guided filtering and view enhanced module for large-scale scene reconstruction, Knowledge-Based Systems, 316 (2025), pp. 113411, https://doi.org/10.1016/j.knosys.2025.113411.

[18] Horé, A., Ziou, D.: Image quality metrics: PSNR vs. SSIM, 2010 20th International Conference on Pattern Recognition, Istanbul, 2010: pp. 2366-2369, https://doi.org/10.1109/ICPR.2010.579.

[19] Wang, Z., Bovik, A.C., Sheikh, H.R., et al.: Image quality assessment: from error visibility to structural similarity, IEEE Transactions on Image Processing, 13 (2004) 4, pp. 600-612, https://doi.org/10.1109/TIP.2003.819861.

[20] Zhang, R., Isola, P., Efros, A.A., et al.: The unreasonable effectiveness of deep features as a perceptual metric, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018: pp. 586-595, https://doi.org/10.1109/CVPR.2018.00068.

[21] Mueller, T., Evans, A., Schied, C., et al.: Instant neural graphics primitives with a multiresolution hash encoding, ACM Transactions on Graphics, 41 (2022) 4, pp. 1-15, https://doi.org/10.1145/3528223.3530127.

[22] Schönberger, J.L., Frahm, J.M.: Structure-from-motion revisited, 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016: pp. 4104-4113, https://doi.org/10.1109/CVPR.2016.445.

[23] Mildenhall, B., Srinivasan, P.P., Ortiz-Cayon, R., et al.: Local light field fusion: practical view synthesis with prescriptive sampling guidelines, ACM Transactions on Graphics, 38 (2019) 4, pp. 1-14, https://doi.org/10.1145/3306346.3322941.

[24] Tancik, M., Srinivasan, P.P., Mildenhall, B., et al.: Fourier features let networks learn high frequency functions in low dimensional domains, Advances in Neural Information Processing Systems, 33 (2020), pp. 7537-7547.

[25] Boje, C., Guerriero, A., Kubicki, S., et al.: Towards a semantic construction digital twin: Directions for future research, Automation in Construction, 114 (2020), pp. 103179, https://doi.org/10.1016/j.autcon.2020.103179.