Leveraging Denoising Diffusion Probabilistic Models for Generating Synthetic Datasets Pertaining to Condition Evaluation of Physical Assets

PEDRAM BAZRAFSHAN, BRIAN WISNER
and ARVIN EBRAHIMKHANLOU

ABSTRACT

This study investigates the problem of data scarcity in the structural health monitoring (SHM), nondestructive evaluation (NDE), and condition evaluation of physical assets by presenting a comparative analysis of denoising diffusion probabilistic models (DDPMs), each trained independently on a structurally distinct dataset. The models are evaluated to assess their image generation performance and their ability to reproduce domain-specific visual characteristics observed in real data. Five distinct datasets were used to train separate DDPMs: 1) surface crack images typical of concrete structures and asphalt pavements (448×448 pixels), 2) their masked versions (448×448 pixels), 3) phased array ultrasonic linear scan images along weld lines in a stainless steel structure (256×256 pixels), 4) two-dimensional (2D) microstructural images of powder-bed fusion laser (PBF-L) additive-manufactured (AM) metals from light optical microscopy (LOM) imaging (patches 256×256 pixels), and 5) ground-penetrating radar (GPR) images (patches 256×256 pixels). The training process for each model was conducted using an NVIDIA H200 graphics processing unit (GPU), requiring approximately six days per dataset, with image generation taking approximately 45 seconds per image. The Fréchet Inception Distance (FID) was used to evaluate the quality of the synthetic images, yielding scores of 13.7 to 34.6 for the datasets, indicating high fidelity and diversity across all cases. By leveraging DDPMs, this research not only overcomes the limitations posed by limited datasets but also provides valuable contributions to advancing AI applications in diverse domains, including structural health monitoring, material characterization, and subsurface imaging.

Pedram Bazrafshan, Civil, Architectural, and Environmental Engineering (CAEE), Drexel University, 3141 Chestnut St., Philadelphia, PA, 19104, U.S.A.
Brian Wisner, Center for Advanced Materials Processing, Mechanical Engineering, Russ College of Engineering and Technology, Ohio University, 28 W Green Dr, Athens, OH, 45701, U.S.A
*(Corresponding Author) Arvin Ebrahimkhanlou, Civil, Architectural, and Environmental Engineering (CAEE)/Mechanical Engineering and Mechanics (MEM), Drexel University, 3141 Chestnut St., Philadelphia, PA, 19104, U.S.A.

INTRODUCTION

GenAI models have gained attention as a solution to data scarcity in SHM, NDE, and condition evaluation of physical assets [1,2]. However, there is limited understanding of how these models behave across different imaging tasks. This paper addresses the problem by presenting a comparative analysis of separately trained DDPMs, each developed for a distinct SHM-related dataset, to evaluate image generation performance and domain-specific characteristics.

The significance of this study lies in addressing the persistent problem of data scarcity in structural health monitoring (SHM) and nondestructive evaluation (NDE), where collecting large, labeled datasets across diverse operational conditions is often costly, time-consuming, or infeasible. This limitation impedes the development of robust data-driven models [3–6], such as computer vision [7–10], robotics-based [11–17], and graph-based models [18]. While traditional data augmentation techniques such as rotation, flipping, and scaling can marginally expand datasets, they do not introduce new semantic content or structural variability and, therefore, fail to sufficiently capture the diversity required for generalizable learning. As such, there is a pressing need for generative methods capable of synthesizing realistic, domain-specific data to support the training and evaluation of advanced diagnostic models.

Generative adversarial networks (GAN) [19–21] and variational autoencoders (VAEs) [22,23] have been explored to address data scarcity. Although DDPMs have been applied in individual studies for tasks such as crack detection, generation, and inpainting [24–26] or stress map prediction [27], little-to-no research has presented a comparative analysis of DDPM performance across multiple SHM datasets. Existing literature tends to focus on single-domain applications, offering limited perspective on how dataset characteristics, such as resolution, feature complexity, or modality, influence training outcomes and image fidelity.

The novelty of this research lies in training and evaluating five independent DDPMs—each tailored to a specific SHM dataset—and presenting the results in a unified framework. This manuscript contributes a cross-domain synthesis of generative performance using consistent metrics and architecture, offering practical insights into dataset-specific challenges and opportunities for applying DDPMs in SHM-related image generation.

DENOISING DIFFUSION PROBABILISTIC MODEL (DDPM)

Each model in this study is based on a DDPM [28] framework, which generates synthetic images by learning to reverse a stochastic process that gradually corrupts real data with Gaussian noise. The forward process incrementally adds noise to the original image over a fixed number of steps, modeled as a Markov chain, until the data becomes indistinguishable from white noise. During training, a neural network is optimized to approximate the reverse process by predicting and removing the added noise at each step. The objective is to minimize the mean squared error between the true noise and the model's prediction, enabling the DDPM to sample new images from random noise by progressive denoising through learned conditional distributions.

To implement the reverse denoising process, a U-Net architecture [29] is employed as the neural backbone of each DDPM. U-Net is a convolutional encoder-decoder

network that captures both global context and fine-grained detail through a symmetric structure and skip connection between corresponding layers in the downsampling and upsampling paths. This design allows the network to preserve spatial resolution while maintaining semantic consistency across scales, making it highly effective for structured image generation. The combination of DDPM's iterative sampling and U-Net's spatial sensitivity enables the generation of high-fidelity images tailored to the underlying characteristics of each dataset.

EXPERIMENTAL DATA

This study evaluates the performance of DDPMs trained independently on five datasets spanning diverse domains of condition assessment. These datasets cover various imaging modalities, including surface inspection, subsurface evaluation, and microstructural characterization. The first dataset consists of RGB images of surface cracks observed on the surfaces of concrete structures and asphalt pavements [30]. These images capture a variety of crack geometries, widths, and surface textures. The size of these images is 448×448 pixels. The second dataset includes binary masked versions of the same images, isolating the crack patterns while removing surface texture and color information. The third dataset comprises grayscale phased array ultrasonic linear scan images collected along the weld line of a stainless-steel pipe [31]. Each image is sized 256×256 pixels and represents acoustic reflections used for weld defect characterization in NDE. The fourth dataset includes 256×256 grayscale patches extracted from LOM images of PBF-L AM metals. These patches represent microstructural features such as voids, grains, and melt pools. The fifth dataset contains grayscale images derived from GPR scans [32]. Original large-format GPR images were cropped and processed into 256×256 patches to enhance model training efficiency while preserving critical subsurface features. Table I provides a summary of the dataset specifications used in this study to train the DDPMs.

TABLE I. OVERVIEW OF THE DATASETS USED TO TRAIN THE MODELS

Dataset No.	Domain	Description	Image Type	Original Resolution	Total Images
1	Surface Cracks	Cracks on concrete/asphalt surfaces	RGB	448×448	9,887
2	Masked Surface Cracks	Binary crack masks from the same RGB dataset	Binary	448×448	9,887
3	Phased array ultrasonics	Phased array ultrasonic linear scans of weld lines	Grayscale	256×256	19,810
4	AM Microstructure (LOM images)	2D micrographs of PBF-L metal samples	Grayscale	628×464 to 85676×12489	13
5	Ground-Penetrating Radar (GPR)	Cropped and resized GPR patches from large scans	Grayscale	1480×400	127

IMPLEMENTATION DETAILS

All models in this study were implemented using a U-Net-based DDPM architecture configured for grayscale or RGB image generation at a resolution of 256×256 pixels. To standardize training across datasets, all input images were either collected or resized to this target resolution. In this regard, the surface crack dataset images and their masked version were resized from 448×448 to 256×256. The ultrasonic linear scans were already available in 256×256 resolution. The AM microstructures and GPR images were preprocessed to match 256×256. Specifically, patches of 256×256 resolution were cropped from the larger original-size images. By acquiring the cropped patches from the original images, the number of training images for the AM microstructure images increased from 13 images to 16,811 image patches, and for the GPR images increased from 127 images to 12,700 image patches.

Each DDPM was trained independently using a cosine noise schedule and a U-Net backbone with skip connections, group normalization, and sinusoidal time embeddings. The training objective minimized the mean squared error between the predicted and true noise across diffusion steps. All models were trained separately for approximately six days per model on an NVIDIA H200 GPU, with one image generation taking about 45 seconds due to the iterative sampling process. For simplicity and comparability, the same architectural configuration and training schedule were adopted across all datasets, except for the number of input channels (1 for grayscale and 3 for RGB). Table II summarizes the key hyperparameters used for training. Figure 1 presents two samples per dataset used in this study.

TABLE II. DDPM HYPERPARAMETERS AND TRAINING CONFIGURATION

Hyperparameter	Value
Image size	256 × 256 pixels
Channels	1 (grayscale) or 3 (RGB)
Time steps (T)	1000
Noise schedule	Cosine
Optimizer	AdamW
Learning rate	3e-4
Batch size	8
Activation function	GELU
Normalization	Group normalization
Embedding	Sinusoidal time embedding
Training duration	~6 days per model
GPU used	NVIDIA H200
Generation time per image	~45 seconds

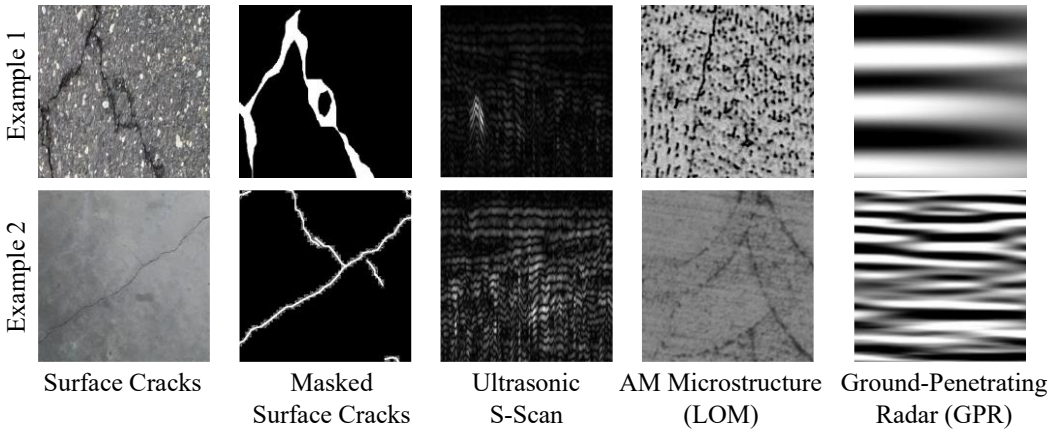

Figure 1. A few samples of 256×256 crops of the datasets used in this study

RESULTS AND DISCUSSIONS

The generative performance of the trained models was evaluated both qualitatively and quantitatively. Qualitatively, the DDPM-generated images were visually consistent with the structural and textural patterns present in their respective training datasets. Figure 2 presents 256×256 representative samples of the generated images for each dataset, illustrating the visual fidelity and dataset-specific characteristics learned by the models. For example, the generated surface crack images captured realistic crack branching and continuity, while the ultrasonic linear scan images exhibited reflection patterns characteristic of defects in weld line inspections. Similarly, the synthetic microstructural patches of AM metals and the GPR outputs preserved domain-specific features such as melt pool boundaries and subsurface reflections. Across all cases, the generated samples visually aligned with the expected content and resolution of the real data.

FID was used to assess the similarity of the synthetic images to real ones quantitatively. FID measures the distance between feature distributions of real and generated images in a pre-trained feature space, providing a representation for both image fidelity and diversity. Lower FID values indicate closer alignment between the two distributions, reflecting better generative performance. Table III summarizes the FID scores for training the DDPMs across the five datasets. The FID values ranged from 13.7 to 34.6 across the five datasets, with the ultrasonic linear scan model achieving the lowest score. This outcome is consistent with the relatively uniform structure of linear scan images, which may facilitate easier learning and generation. On the other hand, higher FID scores for more complex textures, such as AM microstructures, are expected due to their inherent multiscale variability. The FID scores obtained in this study demonstrated satisfactory performance across all datasets, suggesting that the models successfully captured the key statistical characteristics of their respective domains.

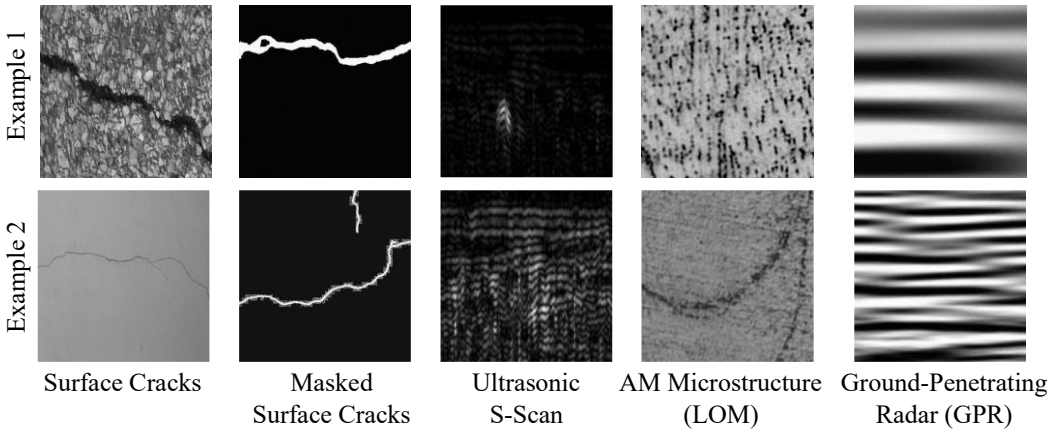

Figure 2. 256×256 samples of synthetically generated images using the trained DDPMs

TABLE III. FID SCORES FOR DDPM-GENERATED IMAGES ACROSS THE FIVE DATASETS

Dataset No.	Domain	FID Score
1	Surface Cracks (RGB)	30.1
2	Surface Cracks (Masked)	22.4
3	Ultrasonic linear scan	13.7
4	AM Microstructure (LOM)	34.6
5	Ground-Penetrating Radar (GPR)	20.9

CONCLUSION AND FUTURE WORK

This study presented a unified evaluation of DDPMs trained independently on five structurally and materially distinct datasets relevant to condition assessment in civil and materials engineering. The qualitative results showed that the generated images successfully reflected the visual characteristics of their respective domains, and the FID scores confirmed satisfactory generative performance across all cases. By standardizing image resolutions and the model architecture, this work enables a direct comparison of DDPM behavior under varying data complexities, offering practical insights for synthetic image generation in SHM, NDE, and subsurface imaging.

While the results demonstrate the capability of DDPMs to produce high-fidelity images across diverse modalities, this work is limited to unconditional generation and fixed image resolution. Future studies should explore conditional DDPMs to incorporate auxiliary information such as damage severity, material type, or spatial metadata into the generation process. In addition, evaluating the utility of the synthetic images for downstream tasks such as detection, segmentation, or anomaly identification would further establish the functional relevance of the generated data. Addressing computational efficiency in sampling and scaling the framework to higher-resolution images are also important directions for expanding the applicability of DDPMs in real-world monitoring systems.

ACKNOWLEDGEMENT

The authors thank Dr. Isabel Morris, Assistant Professor in the Civil Engineering Department at the New Mexico Institute of Mining and Technology, for sharing the dataset of GPR scans.

REFERENCES

[1] S. Cano-Ortiz, L. Lloret Iglesias, P. Martinez Ruiz del Árbol, D. Castro-Fresno, Improving detection of asphalt distresses with deep learning-based diffusion model for intelligent road maintenance, Developments in the Built Environment 17 (2024) 100315. https://doi.org/10.1016/j.dibe.2023.100315.

[2] M. Heesch, K. Mendrok, Z. Dworakowski, Time-Domain Signal Synthesis with Style-Based Generative Adversarial Networks Applied to Guided Waves, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Science and Business Media Deutschland GmbH, 2021: pp. 78–88. https://doi.org/10.1007/978-3-030-87986-0_7.

[3] P. Bazrafshan, T. On, A. Ebrahimkhanlou, A computer vision-based crack quantification of reinforced concrete shells using graph theory measures, in: D. Zonta, Z. Su, B. Glisic (Eds.), Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2022, SPIE, 2022: p. 25. https://doi.org/https://doi.org/10.1117/12.2612359.

[4] T. Ghosh Mondal, M.R. Jahanshahi, Z.Y. Wu, Deep Learning-Based RGB-D Fusion for Multimodal Condition Assessment of Civil Infrastructure, Journal of Computing in Civil Engineering 37 (2023) 04023017. https://doi.org/10.1061/JCCEE5.CPENG-5197.

[5] J. Feng, L. Yang, E. Hoxha, B. Jiang, J. Xiao, Robotic Inspection of Underground Utilities for Construction Survey Using a Ground Penetrating Radar, Journal of Computing in Civil Engineering 37 (2023) 04022049. https://doi.org/10.1061/(ASCE)CP.1943-5487.0001062.

[6] P. Bazrafshan, A. Ebrahimkhanlou, Detection of cracking mechanism transition on reinforced concrete shear walls using graph theory, in: P.J. Shull, T. Yu, A.L. Gyekenyesi, H.F. Wu (Eds.), Nondestructive Characterization and Monitoring of Advanced Materials, Aerospace, Civil Infrastructure, and Transportation XVIII, SPIE, 2024: p. 28. https://doi.org/10.1117/12.3011092.

[7] Q.G. Alexander, V. Hoskere, Y. Narazaki, A. Maxwell, B.F. Spencer, Fusion of thermal and RGB images for automated deep learning based crack detection in civil infrastructure, AI in Civil Engineering 1 (2022) 3. https://doi.org/10.1007/s43503-022-00002-y.

[8] M. Ghyabi, D. Lattanzi, Computer vision-based video signal fusion using deep learning architectures, J Civ Struct Health Monit (2025) 1–16. https://doi.org/10.1007/s13349-024-00891-w.

[9] E. Ichi, S. Dorafshan, Evaluation of Infrared Thermography Dataset for Delamination Detection in Reinforced Concrete Bridge Decks, Applied Sciences 14 (2024) 2455. https://doi.org/10.3390/app14062455.

[10] M. Hamidia, M. Kaboodkhani, H. Bayesteh, Vision-oriented machine learning-assisted seismic energy dissipation estimation for damaged RC beam-column connections, Eng Struct 301 (2024) 117345. https://doi.org/10.1016/j.engstruct.2023.117345.

[11] F. Azizi Zade, A. Ebrahimkhanlou, Point clouds to as-built two-node wireframe digital twin: a novel method to support autonomous robotic inspection, Autonomous Intelligent Systems 4 (2024) 25. https://doi.org/10.1007/s43684-024-00082-w.

[12] A. Ghadimzadeh Alamdari, F.A. Zade, A. Ebrahimkhanlou, A Review of Simultaneous Localization and Mapping for the Robotic-Based Nondestructive Evaluation of Infrastructures, Sensors 25 (2025) 712. https://doi.org/10.3390/s25030712.

[13] P. Bazrafshan, A. Ebrahimkhanlou, A virtual-reality framework for graph-based damage evaluation of reinforced concrete structures, in: P.J. Shull, T. Yu, A.L. Gyekenyesi, H.F. Wu (Eds.), Nondestructive Characterization and Monitoring of Advanced Materials, Aerospace, Civil Infrastructure, and Transportation XVII, SPIE, 2023: p. 5. https://doi.org/10.1117/12.2657736.

[14] P. Bazrafshan, A. Ebrahimkhanlou, A Robotic-Based Framework For Quantifying Surface Cracks Of Concrete Shear Walls, in: Proceedings of the 14th International Workshop on Structural

Health Monitoring, Destech Publications, Inc., 2023. https://doi.org/https://www.dpi-proceedings.com/index.php/shm2023/article/view/36868.
[15] A. Ellenberg, A. Kontsos, F. Moon, I. Bartoli, Bridge related damage quantification using unmanned aerial vehicle imagery, Struct Control Health Monit 23 (2016) 1168–1179. https://doi.org/10.1002/stc.1831.
[16] Y. He, Z. Liu, Y. Guo, Q. Zhu, Y. Fang, Y. Yin, Y. Wang, B. Zhang, Z. Liu, UAV based sensing and imaging technologies for power system detection, monitoring and inspection: a review, Nondestructive Testing and Evaluation (2024) 1–68. https://doi.org/10.1080/10589759.2024.2421938.
[17] J. Seo, L. Duque, J. Wacker, Drone-enabled bridge inspection methodology and application, Autom Constr 94 (2018) 112–126. https://doi.org/10.1016/j.autcon.2018.06.006.
[18] P. Bazrafshan, T. On, S. Basereh, P. Okumus, A. Ebrahimkhanlou, A graph-based method for quantifying crack patterns on reinforced concrete shear walls, Computer-Aided Civil and Infrastructure Engineering 39 (2024) 498–517. https://doi.org/10.1111/mice.13009.
[19] S. Li, Y. Le, X. Zhao, Style-Controlled Image Synthesis of Concrete Damages Based on Fusion of Convolutional Encoder and Attention-Enhanced Conditional Generative Adversarial Network, Journal of Computing in Civil Engineering 38 (2024) 04024032. https://doi.org/10.1061/JCCEE5.CPENG-6007.
[20] K. Zhang, Y. Zhang, H.D. Cheng, Self-Supervised Structure Learning for Crack Detection Based on Cycle-Consistent Generative Adversarial Networks, Journal of Computing in Civil Engineering 34 (2020) 04020004. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000883.
[21] C. Han, T. Ma, J. Huyan, Z. Tong, H. Yang, Y. Yang, Multi-stage generative adversarial networks for generating pavement crack images, Eng Appl Artif Intell 131 (2024) 107767. https://doi.org/10.1016/j.engappai.2023.107767.
[22] R.A. Cody, B.A. Tolson, J. Orchard, Detecting Leaks in Water Distribution Pipes Using a Deep Autoencoder and Hydroacoustic Spectrograms, Journal of Computing in Civil Engineering 34 (2020) 04020001. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000881.
[23] Y. Li, S. Liu, M. Wang, S. Li, J. Tan, Teleoperation-Driven and Keyframe-Based Generalizable Imitation Learning for Construction Robots, Journal of Computing in Civil Engineering 38 (2024) 04024031. https://doi.org/10.1061/JCCEE5.CPENG-5884.
[24] E. Aldao, L. Fernández-Pardo, F. Veiga-López, L.M. González-deSantos, H. González-Jorge, Synthetic Data Generation Techniques for Enhancing Crack Detection in Railway Concrete Sleepers, Journal of Computing in Civil Engineering 39 (2025) 04025032. https://doi.org/10.1061/JCCEE5.CPENG-6158.
[25] C. Han, H. Yang, T. Ma, S. Wang, C. Zhao, Y. Yang, CrackDiffusion: A two-stage semantic segmentation framework for pavement crack combining unsupervised and supervised processes, Autom Constr 160 (2024) 105332. https://doi.org/10.1016/j.autcon.2024.105332.
[26] Q. Lei, J. Zhong, M. Dong, K. Ota, Faithful crack image synthesis from evolutionary pixel-level annotations via latent semantic diffusion model, Expert Syst Appl 275 (2025) 126986. https://doi.org/10.1016/j.eswa.2025.126986.
[27] Y. Jadhav, J. Berthel, C. Hu, R. Panat, J. Beuth, A. Barati Farimani, StressD: 2D Stress estimation using denoising diffusion model, Comput Methods Appl Mech Eng 416 (2023) 116343. https://doi.org/10.1016/j.cma.2023.116343.
[28] J. Ho, A. Jain, P. Abbeel, Denoising Diffusion Probabilistic Models, Adv Neural Inf Process Syst 2020-December (2020). http://arxiv.org/abs/2006.11239 (accessed January 17, 2024).
[29] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer, Cham, 2015: pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
[30] Crack Segmentation Dataset, (2020). https://www.kaggle.com/datasets/lakshaymiddha/crack-segmentation-dataset (accessed July 30, 2024).
[31] I. Virkkunen, T. Koskinen, O. Jessen-Juhler, J. Rinta-aho, Augmented Ultrasonic Data for Machine Learning, J Nondestr Eval 40 (2021) 4. https://doi.org/10.1007/s10921-020-00739-5.
[32] P. Bazrafshan, I. Morris, A. Ebrahimkhanlou, Synthetic ground-penetrating radar image generation using denoising diffusion probabilistic models, in: P. Rizzo, Z. Su, F. Ricci, K.J. Peters (Eds.), Health Monitoring of Structural and Biological Systems XIX, SPIE, 2025: p. 57. https://doi.org/10.1117/12.3051096.