13th International Conference on
Structural Health Monitoring of Intelligent Infrastructure
DOI: 10.3217/978-3-99161-057-1-093

Deep generative models to mitigate data scarcity in bridge structural health monitoring

Sasan Farhadi[1, 0000-0002-9562-9199], Mauro Corrado[1], Danilo Acquesta Nunes[1], Giulio Ventura[1]
[1]Institute of Engineering Geodesy and Measurement Systems, Graz University of Technoclogy, Graz, Austria
email: sasan.farhadi@polito.it, mauro.corrado@polito.it, danilo.acquesta@polito.it giulio.ventura@polito.it

ABSTRACT: Structural health monitoring is essential for ensuring the safety, reliability, and longevity of infrastructural assets. However, conventional monitoring measurements face significant challenges, such as being labor-intensive, costly, and time-consuming. In recent years, the rise of machine learning and deep learning has data analysis frameworks, offering a promising solution to these challenges. Despite this, developing reliable and robust approaches that generalize well to unseen scenarios often requires large amounts of training data. This presents a challenge, mainly with regulatory constraints and difficulties in collecting data, particularly for rare events. To address the issue of data scarcity, this study proposes a generative data augmentation approach using a Wasserstein Generative Adversarial Network (WGAN). This approach generates high-quality short-time Fourier transform (STFT) spectrograms, which are transformed into image-like data, from in-situ acceleration signals for model training. The collected signals, recorded from real-world bridges during various events such as hammering, drilling, environmental noise, and, most importantly, the rare event of wire breakage in prestressed concrete bridges, are processed and fed into the WGAN model to synthesize additional data. This improves the diversity and robustness of training datasets. Evaluation of the generated spectrograms using various performance metrics, such as Structural Similarity Index Measure, Peak Signal-to-Noise Ratio, and Fréchet Inception Distance, demonstrates that the proposed method offers a scalable and cost-effective solution for enhancing the training dataset, particularly in scenarios where event data is sparse, such as prestressing wire breakage.

1 INTRODUCTION

Bridges are critical infrastructures for transportation and economic development, but are increasingly vulnerable to deterioration caused by aging, traffic loads, and environmental impacts [1]. Structural Health Monitoring (SHM) is essential for the early detection of damage, as highlighted by failures such as the Reale Viaduct and Fossano Bridge collapses in Italy [2, 3]. However, conventional SHM methods, such as visual inspections, are often costly, labor-intensive, and ineffective at detecting subtle or internal damage. Although sensor-based methods offer improvements, they generate complex and noisy data, making manual analysis difficult [4]. As a result, recent advances in Machine Learning (ML) and Deep Learning (DL) have emerged as promising approaches, demonstrating strong performance across various domains, including structural damage detection [5, 6]. Nevertheless, real-world implementation of these methods still faces challenges, particularly data scarcity and class imbalance, which are especially essential for rare events like wire breakage, an internal form of structural damage that is difficult to capture.

Data augmentation (DA) techniques have been proposed to address these issues, but conventional methods like Mixup and time-shifting are insufficient for replicating the complexity of real-world events. In this study, Generative Adversarial Networks (GANs) are proposed as a powerful solution for generating realistic synthetic data. Prior work, mainly in speech and audio generation, has demonstrated the effectiveness of spectral-domain representations, such as STFT spectrograms, for improving generative models. However, applications of GANs in SHM, particularly for STFT-based augmentation, remain limited.

To address this gap, this study utilized Wasserstein GAN using gradient penalty (WGAN-GP), a GAN model designed to generate single-channel STFT spectrograms specifically for SHM applications. Unlike traditional three-channel approaches, the single-channel input maintains the spectral information of structural vibration signals. This study also provides a unique real-world dataset from two operational bridges in Italy, offering realistic and challenging data for model training and validation. By enhancing dataset diversity and improving model robustness, the proposed approach aims to tackle key SHM challenges, advancing scalable, adaptive, and reliable monitoring frameworks. The methodology, experimental setup, evaluation strategies, and detailed analysis are presented to support future developments and replication steps.

2 METHODS

2.1 Generative Adversarial Networks

GANs are a type of generative model that learns patterns in data and generates new samples that resemble the original dataset [7]. GANs consist of two neural networks, competing in a zero-sum game: the generators, which create the synthetic data, and the discriminator, which distinguishes between the real and generated data (Figure 1). The generator aims to minimize the discriminator's ability to differentiate between real and fake samples, while the discriminator tries to maximize its ability to classify data correctly.

GANs' training can be challenging due to issues like mode collapse and gradient instability. To address this challenge, Wasserstein GAN (WGAN) was introduced [8]. In this approach, the Wasserstein distance (Earth Mover's distance) was utilized to measure the difference between the real and generated data. The WGAN replaces the discriminator with a critic network, which provides more stable gradients, avoiding the problem of vanishing gradients in traditional GANs.

To further improve training stability, WGAN with gradient penalty (WGAN-GP) was introduced [9], which added a gradient penalty in the loss function. This term makes the critic's gradient more stable and smooth and improves convergence. The WGAN-GP loss functions are:

$$L_D = \frac{1}{N}\sum_{n=1}^{N}[d\big(g(z_n)\big) - d(x_n)] + \lambda.E[(|\nabla \mathrm{d}(\hat{x})|_2 - 1)^2]$$

DOI: 10.3217/978-3-99161-057-1-093

In this equation $d(x_n)$ and $d\big(g(z_n)\big)$ are the critic's outputs for real and generated images, respectively. N is the batch size and λ is controlling the strength of the gradient penalty.

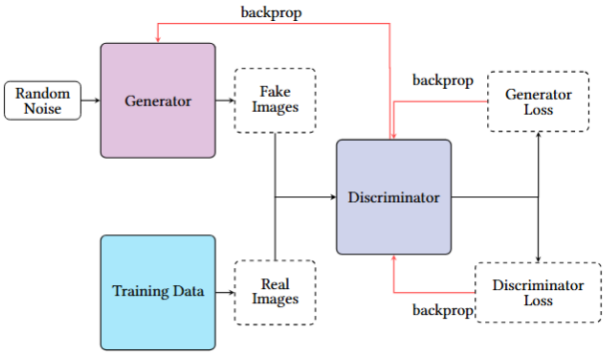

Figure 1. Overview of Generative Adversarial Networks.

2.2 Evaluation Metrics

2.2.1 Structural Similarity Index Measure (SSIM)

SSIM evaluates the similarity between two images by considering structural information, such as luminance, contrast, and structure. The SSIM score ranges from 0 to 1, with 1 being perfect similarity [10].

$$SSIM = \left(\frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}\right).\left(\frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}\right).\left(\frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}\right)$$

In this equation, x and y are image patches. $\mu_x$and μ_y are the mean intensities of the image x and y. σ_x and σ_y are the standard deviations and σ_{xy} is the covariance, and c is the constant value to prevent division by zero.

2.2.2 Peak Signal-to-Noise Ratio (PSNR)

PSNR measures the similarity between two images by comparing the maximum pixel value to the noise in the image. Higher PSNR value indicates better quality [11].

$$PSNR(R,G) = 10log_{10}(\frac{MAX_I^2}{MSE})$$

R and G refer to the real and generated images, respectively. MAX_I is the maximum possible pixel value (set to +1 in this study due to normalization), and *MSE* is the mean squared error between the real and generated images.

2.2.3 Fréchet Inception Score (FID)

FID measures the distance between the feature distributions of real and generated images [12]. A lower FID score indicates better image quality.

$$FID = \left\|\mu_r - \mu_g\right\|^2 + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{\frac{1}{2}})$$

$\mu_r$and μ_g are the mean feature vectors of the real and generated images, respectively. Σ_r and Σ_g are the covariance matrices of the real and generated images. *Tr* refers to the trace of the matrix, which is the sum of its diagonal elements.

2.3 Acquisition Setup and Data Collection

The experimental framework was conducted on two post-tensioned concrete bridges in Italy, Le Pastena and Cerqueta. The data collection phase was performed during the deconstruction and maintenance phase with a collaboration between Politecnico di Torino and Strada dei Parchi S.p.A. Due to the bridges' height and limited access, a Mobile Elevated Work Platform (MEWP) was used for instrumentation setup. The wire cutting process, performed on twisted tendons, was carried out using an electric trimmer in a controlled setup after the prestressing tendons were exposed.

The primary dataset was acquired using two accelerometers (Model 805M1, DSPM Industria srl), placed 4.5 meters from the cutting point on the sides of the beams. This setup was chosen to ensure the signal could be reliably recorded, considering the effects of signal attenuation and dispersion along the propagation path. These propagation effects can influence the measured acceleration signals and, consequently, the generated spectrograms. These accelerometers recorded structural vibrations during wire cutting, with a high sampling rate of 96 kHz and a frequency response of 0.4 to 12 kHz. In addition to the rare event of wire breakage signals, other vibrational events such as hammering, electric trimmer, and traffic were captured. This comprehensive dataset ensures the development of a robust model for generating sound events and further development of more generalized automated systems (Figure 2).

Figure 2. Sensor positions on La Patenda Bridge.

DOI: 10.3217/978-3-99161-057-1-093

2.4 Experiments

2.4.1 Feature Extraction

The dataset used in this study comprises four signal classes: wire breakage (202 samples), hammering (264 samples), electric trimmer (459 samples), and traffic (415 samples). As it is evident, there is a class imbalance, with critical events like wire breakage being underrepresented, which may affect model performance. The signals were captured at a sampling frequency of 96 kHz, with each signal having 1000 samples over 0.0104 seconds (Figure 2). To ensure consistency and reduce bias from varying amplitudes, all signals were normalized before further processing.

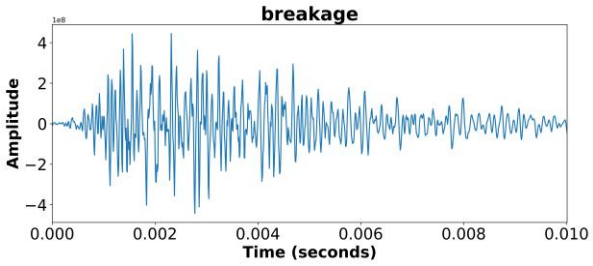

Figure 3. Time-domain representation of the wire breakage acceleration signal (units: m/s²).

This study uses Short-Time Fourier Transform (STFT) for feature extraction from in-situ acceleration signals to analyze event patterns. STFT transforms signals from the time domain to the two-dimensional time-frequency domain, keeping both temporal and spectral information. One important parameter in STFT analysis is the window size selection, which can affect the trade-off between time and frequency resolution. Smaller window sizes (e.g., 64) offer high time resolution but low frequency resolution, while larger windows (e.g., 512) provide better frequency resolution but less temporal precision. In this study, the optimal balance was achieved using window sizes of 128 and 256 (Figure 3). The extracted STFT spectrograms were used as single-channel images for model training, optimizing feature extraction and computational efficiency for event detection and structural health monitoring applications.

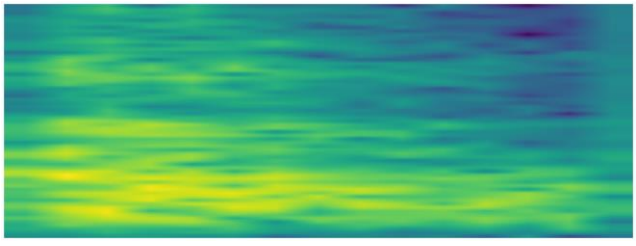

Figure 4. STFT spectrogram of a wire breakage signal. The x-axis represents time (seconds), the y-axis represents frequency (Hz).

2.4.2 Model Developments

The WGAN-GP was implemented as one of the most successful models for image generation. It improves on the standard GAN and WGAN models by using the Wasserstein loss function and adding a gradient penalty. The generator starts with a latent noise vector of size 100, a typical setting in GAN studies. This vector is reshaped into small feature maps and upsampled through transposed convolutional layers to generate spectrogram-like images. In this study, the generator consists of four transposed convolutional layers with kernel sizes of 5×5 and strides of 2, progressively reducing the feature map depth from 512 to 256, 128, and 64 before producing the final spectrogram. To ensure stable training and avoid vanishing or exploding gradients, batch normalization is applied after each layer. Leaky ReLU activation is used throughout the layers to add non-linearity, except in the final layer, where a tanh activation function normalizes the output values to the range [-1, 1].

The critic, unlike the discriminator, does not classify samples as real or fake; instead, it assigns real-valued scores to samples, helping to approximate the Wasserstein distance between real and generated data. The critic consists of several 2D convolutional layers, which progressively reduce the spatial dimensions of the input, followed by batch normalization and Leaky ReLU activations to improve learning stability. To enforce the Lipschitz constraint, a gradient penalty term with a coefficient of λ=12 is added to the loss function. The selection of hyperparameters search space was based on a combination of recommendations from GAN literature and preliminary tuning experiments on dataset to achieve stable and high-quality spectrogram generation. The summary of hyperparameter selection for this model is in Table 1.

Table 1. Hyperparameter Selection for WGAN-GP Model

Parameter	Value	Search Space
Learning Rate (Generator)	2×10^{-5}	2×10^{-5} to 2×10^{-5}
Learning Rate (Critic)	2×10^{-6}	2×10^{-5} to 2×10^{-5}
Batch Size	16	16 to 64
Epochs	1500	1000 to 5000
Optimizer	Nadam	Nadam, Adam, RMSProp
Gradient Penalty Coefficient	12	1 to 25
Activation Function (Generator)	Leaky ReLU, Tanh	Leaky ReLU, ReLU, ELU, Tanh
Activation Function (Critic)	Leaky ReLU	Leaky ReLU, ReLU, ELU, Tanh
Generator Layers	4 Transposed Convolutions	3 to 8
Critic Layers	5 Convolutions	3 to 8
Kernel Size	5x5	3x3 to 5x5

To assess the performance of WGAN-GP and the quality of generated images, the model was evaluated using multiple metrics, including SSIM, PSNR, and FID. The results are shown in Table 2 for both window sizes of 128 and 256, which enable the determination of the effect of window sizes in generating STFT-based images.

Table 2. Performance Metrics of WGAN-GP

Metrics	Class	Window Size:128	Window Size:256
SSIM	Breakage	**0.367**	0.180
	Trimmer	0.342	**0.403**
	Hammer	0.208	0.232
	Traffic	0.170	0.186
PSNR	Breakage	13.409	11.972

DOI: 10.3217/978-3-99161-057-1-093

	Trimmer	**13.424**	**13.621**
	Hammer	10.520	11.289
	Traffic	11.374	12.020
FID	Breakage	0.219	0.331
	Trimmer	**0.173**	0.270
	Hammer	0.221	**0.194**
	Traffic	0.179	0.237

For SSIM, WGAN-GP achieved the highest value for the Trimmer event at window size 256 (0.403), indicating better structural similarity compared to the other events. It is notable that for the breakage event at a window size of 128, the score dropped from 0.367 to 0.18, which suggests that the model is not able to preserve the pattern of more complex signals, such as wire breakage. For Hammer and Traffic events, the SSIM values were relatively lower, indicating that WGAN-GP had difficulty maintaining high structural similarity for these types of events across both window sizes.

In terms of PSNR, which measures image clarity and noise levels, the best performance was obtained for the Trimmer event with a PSNR of 13.424 at window size 128 and 13.621 at window size 256, indicating that WGAN-GP generated spectrograms with minimal distortion for this event. However, the Breakage event exhibited lower PSNR scores, particularly at window size 256 (11.972), highlighting a reduction in image quality for events with more intricate features. Hammer and Traffic events also showed similar trends, with Traffic performing better at window size 256 (12.020). For reference, higher PSNR values indicate greater similarity, with values above 20 generally considered good for images; however, in the context of generated spectrograms, PSNR values are typically lower, and values above 10 are commonly reported as acceptable in the literature for synthetic data with complex structures.

The FID metric, which evaluates the similarity between real and generated data distributions, showed that WGAN-GP performed well for the Trimmer event at window size 128 (FID of 0.173). However, it struggled with Breakage and Hammer, with higher FID scores indicating that the generated spectrograms deviated more from real data. Overall, Traffic also showed relatively low FID values at both window sizes, indicating good model performance for simpler events.

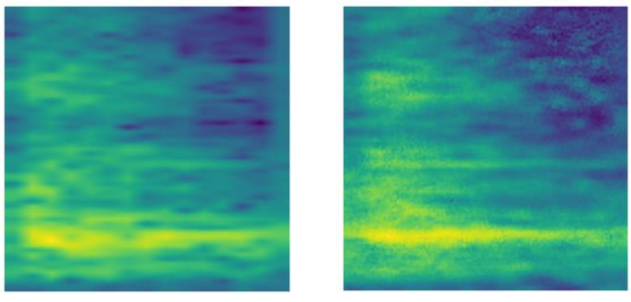

Figure 5. STFT Spectrogram images (window size 128) of wire breakage signal (Left) Real Sample, (Right) Generated by WGAN-GP. The x-axis represents time (seconds), the y-axis represents frequency (Hz).

In summary, WGAN-GP demonstrated strengths in generating synthetic spectrograms for certain events like Trimmer and Traffic, especially at window size 128. However, the model faced challenges in generating high-quality spectrograms for more complex events such as Breakage and Hammer, particularly at larger window sizes. These results highlight the need for further development and utilization of GAN models to better capture fine spectral features and improve consistency across different types of events.

2.5 Conclusion and Future Developments

This study demonstrated the potential of GAN-based data augmentation in the context of structural health monitoring using STFT spectrograms from in-situ acceleration signals. The model showed promising performance, particularly for Trimmer and Traffic events. However, it is important to note that WGAN-GP showed some limitations in generating more complex STFT patterns, such as those associated with Wire Breakage events.

To further improve the model, future work will focus on addressing these limitations by refining and customizing the architecture. Additionally, integrating more advanced models, such as sequential models and attention mechanisms, which are robust for time-series data, will be explored to enhance the model's ability to capture intricate temporal dependencies and improve the generation of complex event patterns. Furthermore, future studies will systematically evaluate the impact of GAN-generated data on downstream event classification and detection models to better quantify the practical benefits of data augmentation for structural health monitoring.

REFERENCES

[1] Ko, J. M., & Ni, Y. Q. (2005). Technology developments in structural health monitoring of large-scale bridges. Engineering Structures, 27(12), 1715–1725.

[2] Ferro, G. A., Restuccia, L., Falliano, D., Devitofranceschi, A., & Gemelli, A. (2022). Collapse of Existing Bridges: From the Lesson of La Reale Viaduct to the Definition of a Partial Safety Coefficient of Variable Traffic Loads. Journal of Structural Engineering, 148(11), 04022181.

[3] Zingoni, A. (Ed.). (2019). Advances in engineering materials, structures and systems: Innovations, mechanics and applications: Proceedings of the seventh International Conference on Structural Engineering, Mechanics, and Computation, 2-4 September 2019, Cape Town, South Africa. International Conference on Structural Engineering, Mechanics and Computation, Boca Raton. CRC Press, Taylor & Francis Group.

[4] Saleem, M. R., Park, J.-W., Lee, J.-H., Jung, H.-J., & Sarwar, M. Z. (2021). Instant bridge visual inspection using an unmanned aerial vehicle by image capturing and geo-tagging system and deep convolutional neural network. Structural Health Monitoring, 20(4), 1760–1777.

[5] Farhadi, S., Corrado, M., & Ventura, G. (2024a). Acoustic Event-Based Prestressing Concrete Wire Breakage Detection. Procedia Structural Integrity, 64, 549–556.

DOI: 10.3217/978-3-99161-057-1-093

[6] Farhadi, S., Corrado, M., & Ventura, G. (2024b). Automated acoustic event-based monitoring of prestressing tendons breakage in concrete bridges. Computer-Aided Civil and Infrastructure Engineering, 39(24), 3700–3720.

[7] Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. "Generative adversarial networks." Communications of the ACM 63, no. 11 (2020): 139-144.

[8] Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein GAN (No. arXiv:1701.07875). arXiv.

[9] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. (2017). Improved Training of Wasserstein GANs (Version 3). arXiv.

[10] Li, Y., Gan, Z., Shen, Y., Liu, J., Cheng, Y., Wu, Y., Carin, L., Carlson, D., & Gao, J. (2018). StoryGAN: A Sequential Conditional GAN for Story Visualization (Version 2).

[11] Smith, S. W. (1997). The scientist and engineer's guide to digital signal processing (1. ed). California Technical Publ.

[12] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium.