Lightweight Modal Super-Resolution CNN and Fourier Neural Operator for High-Resolution Damage Identification

Jinghao Fei a, Chul-Woo Kim a,\*, Debao Chen a, Hongkang Pan a, Rongxiu Chen a, Jiaji Wang b    
a Department of Civil and Earth Resources Engineering, Kyoto University, Kyoto, Japan

b Department of Civil Engineering, The University of Hong Kong, Pokfulam, Hong Kong, China

\*Corresponding author. E-mail address: kim.chulwoo.5u@kyoto-u.ac.jp (C-W. Kim).

Abstract

Accurate damage identification is essential for ensuring the safety and reliability of civil infrastructure. This study proposes a synergistic deep learning framework that integrates a Modal Super-Resolution Convolutional Neural Network (MoSRNet) with a Modal–Stiffness Fourier Neural Operator (MS-FNO) for high-resolution damage identification in bridge structures. The MoSRNet reconstructs fine-meshed modal data from coarsely measured mode shapes, while the MS-FNO establishes an operator mapping from modal responses to stiffness fields, enabling end-to-end prediction of spatially continuous damage. Training data are generated using one-dimensional Gaussian Random Fields (GRFs) to simulate realistic, continuous stiffness loss patterns and enhance model generalization. The proposed framework is validated through extensive numerical simulations and laboratory experiments on a simply supported steel beam and is benchmarked against a ResNet baseline. Results demonstrate that the framework achieves superior accuracy, robustness, and physical interpretability, successfully identifying complex damage patterns even when using only three mode shapes and seven sensors, thereby highlighting its effectiveness and practicality for real-world structural health monitoring applications.

Keywords

Fourier Neural Operator, Convolutional Neural Network, Mode shape, Damage Identification, Gaussian random field, Structural Health Monitoring.

1 INTRODUCTION

Civil infrastructure, such as roads, bridges, and railways, constitutes the backbone of modern society. Throughout their service life, these structures inevitably deteriorate due to aging, natural disasters, and human-induced factors. Ensuring the safety of civil infrastructure is essential, as any major failure can lead to severe social and economic consequences [1]. For instance, the Morandi Bridge in Genoa, Italy, collapsed as a result of long-term inadequate maintenance, causing 43 fatalities [2]. Similarly, the Fern Hollow Bridge in Pittsburgh, USA, collapsed after 49 years of service, with investigations revealing that insufficient periodic maintenance was a major contributing factor [3]. To prevent such incidents and to ensure the safe and efficient operation of social and economic systems, the implementation of effective structural health monitoring (SHM) is of critical importance.

In general, SHM approaches can be categorized into two types: manual inspection and automated monitoring. At the current stage, manual inspection remains the mainstream practice. Conventional manual inspections include routine checks, examinations of fracture-critical components, and underwater assessments. These are typically conducted periodically, at intervals ranging from one to six years, depending on national regulations [4]. However, most manual inspections rely heavily on visual assessments of structural surfaces, which often fail to capture the actual internal condition of the structure. A study on 31 types of bridges reported a weak correlation between visual condition ratings and actual structural integrity [5]. To address this limitation, several advanced non-destructive testing (NDT) techniques—such as acoustic emission [6–8] and guided wave methods [9–11] —have been developed to assist engineers. These techniques provide high sensitivity and early-stage damage detection capabilities. Nevertheless, their application is often limited by their localized detection range and high deployment costs, which hinder large-scale or long-term implementation.

In recent years, machine learning (ML) has emerged as a powerful alternative to traditional numerical methods in civil engineering. For forward problems, data-driven and physics-informed approaches can serve as efficient surrogate models that approximate complex structural responses with significantly reduced computational costs [12–16]. For inverse problems, the strong nonlinear representation capability of ML models enables the inference of system parameters and damage distributions directly from large volumes of measurement data [17–20]. These advancements have made ML-based automated SHM systems a rapidly developing research frontier.

A typical automated SHM system consists of sensors, data acquisition modules, processing units, and damage detection algorithms [21]. Among these components, the type of data collected plays a decisive role in determining the system’s effectiveness in identifying diverse damage patterns. Acceleration data is commonly used owing to installation convenience and its sensitivity to dynamic characteristics. However, it is highly susceptible to environmental noise, which poses challenges in real-world applications [22–25]. Strain data is effective in detecting local damages but cannot represent the global structural condition, and strain gauges often lack durability for long-term field monitoring [26,27]. Displacement data is ideal in theory but remains difficult to measure, especially for large-scale structures such as bridges [12,28]. Image and video data have gained popularity with the progress of computer vision techniques, which enable the detection of surface deterioration. However, these methods are ineffective for identifying internal or hidden damage [29–32].

Among the various data sources, modal parameters exhibit distinct advantages because they are directly related to the structure’s dynamic properties and possess clear physical interpretations. In practice, modal parameters are typically obtained through system identification using acceleration data. By processing long-term monitoring data, the influence of noise can be statistically quantified and mitigated, enabling more reliable assessments of structural health [33].

Modal parameters have been extensively utilized in ML-based automated SHM systems. Various machine learning architectures have been explored to address this problem, including autoencoders [17], convolutional neural networks [34], deep residual networks [35], attention mechanisms [36] and physics-informed neural networks [19,37]. These advanced models have demonstrated the feasibility of localizing and quantifying structural damage by extracting latent features from modal parameters. However, most existing approaches still face limitations in terms of efficiency, economic feasibility, and robustness, which restrict their broader practical deployment. The details of these representative methods are summarized in Table 1.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 1** Summary of previous modal parameter- and ML-based SHM methods | | | | | | | |
| Reference | Target structure | Model architecture | Nm | Ns | Ne | Ns / Ne | Dataset descriptions |
| Pathirage et al. [17] | Seven-story steel frame | Autoencoder | 8×1DOF | 8 | 8 | 1.0000 | Single damage: 25400  Multiple damage: 25200 |
| Guo et al. [34] | Beam model | Convolutional neural network | 1×1DOF | 39 | 40 | 0.9750 | Randomly choose 1 to 3 damages; training: 1.5×106 , validation: 5×104, test: 104 |
| Wang et al. [35] | Concrete bridge beam | Deep residual network | 3×1DOF | 7 | 16 | 0.4375 | Randomly add damage to 16 elements; 20000 samples |
| Guo & Fang [37] | Five-story steel frame | Physics-informed neural networks | 3×3DOF | 5 | 5 | 1.0000 | Physics-informed; 30 samples |
| Lei et al. [19] | Aluminum beam | Physics-informed convolutional neural network | 4×1DOF | 13 | 12 | 1.0833 | Single damage: 360  Multiple damage: 660 |
| Wang et al. [36] | Steel beam | Attention mechanisms | 5×1DOF | 7 | 8 | 0.8750 | One to three damages;  711120 samples |
| Note: Nm denotes the number of modes; DOF denotes the degree of freedom; Ns denotes the number of sensors; Ne denotes the number of elements in the model predictions. | | | | | | | |

The drawbacks of these existing solutions are evident from Table 1. Pathirage et al. [17] proposed an autoencoder-based model that requires the first eighth modes as input, which is hard to obtain in real-world engineering practice. The approach developed by Guo et al. [34] relies on densely measured mode shapes, rendering it economically impractical due to the high sensor cost. Wang et al. [35] introduced a deep residual network-based method; however, the experimental validation data were generated by applying damage to specific elements, without ground truth information of the actual stiffness loss. This limitation raises concerns regarding the adaptability of the method to other structural scenarios. Guo and Fang [37] proposed a physics-informed neural network (PINN) framework, but they also showed that the training process lacks stability. Lei et al. [19] developed another PINN-based approach capable of handling unseen damage patterns, yet their experimental validation required training separate models on different datasets for each damage case, compromising the model’s generalization. More recently, Wang et al. [36] incorporated attention mechanisms into their framework, but their dataset contained an excessive number of samples, including experimental cases, making the claimed generalization capability questionable.

Furthermore, in most of these studies, the number of predicted elements is nearly equal to the number of sensors, without exploring how to effectively exploit limited sensor information to enhance the spatial resolution of stiffness prediction. In addition, the datasets used in these methods generally assume that damage occurs in specific elements while others remain intact, which deviates from real-world conditions where structural deterioration tends to be spatially continuous. This oversimplification significantly limits the model’s generalization capability and practical applicability.

To overcome the aforementioned limitations, this study proposes a synergistic framework that integrates a Modal Super-Resolution Convolutional Neural Network (MoSRNet) with a Modal–Stiffness Fourier Neural Operator (MS-FNO) for high-resolution damage identification in bridge structures. The MoSRNet is designed to enhance the spatial resolution of low-resolution measured mode shape data, thereby generating high-resolution modal fields for subsequent analysis. Building upon the Fourier Neural Operator (FNO) architecture proposed by Li et al. [38], the MS-FNO learns an inverse mapping operator from modal response fields to the stiffness field, enabling efficient and accurate identification of structural damage.

The key innovations and contributions of this study are summarized as follows:

1. **Application of FNO in SHM:** This study pioneers the application of the Fourier Neural Operator (FNO) algorithm to modal-parameter-based structural health monitoring (SHM). An end-to-end framework is established to directly predict the stiffness loss field from modal response fields, providing a physically interpretable and computationally efficient solution for inverse problems.
2. **Development of MoSRNet:** A Modal Super-Resolution Network (MoSRNet) is developed to enhance the spatial resolution of low-resolution mode shape data. By fully exploiting the available sensor information, MoSRNet enables high-resolution modal field reconstruction and facilitates more precise damage localization.
3. **Generation of realistic training data using Gaussian Random Fields (GRFs):** A one-dimensional Gaussian Random Field is adopted to generate training samples that simulate continuous and spatially correlated stiffness loss patterns. Compared with conventional segment-based damage assumptions, this approach better reflects the stochastic nature of real structural deterioration and improves the generalization capability of the trained model.
4. **Utilization of only the first three mode shapes for damage identification:** The proposed framework achieves reliable stiffness field prediction using only the first three mode shapes, which are relatively easy to obtain in practice. This demonstrates the efficiency and practicality of the proposed method under realistic sensing constraints.

The remainder of this paper is organized as follows. Section 2 presents the theoretical background of the proposed framework. Section 3 describes the model architecture and implementation details. Section 4 reports the numerical validation results, while Section 5 focuses on the experimental validation. Finally, Section 6 concludes the study by summarizing the main findings and outlining future research directions.

|  |
| --- |
|  |
| **Figure 1** End-to-end framework from data preparation to implementation for damage identification. |

2 FRAMEWORK FOR HIGH-RESOLUTION DAMAGE IDENTIFICATION

2.1 Framework Overview

To overcome the limitations of previous studies, this study proposes a synergistic framework that integrates a Modal Super-Resolution Network (MoSRNet) and a Modal–Stiffness Fourier Neural Operator (MS-FNO) for high-resolution damage identification using only three coarsely measured mode shapes, as illustrated in Figure 1. In this framework, the MoSRNet first reconstructs fine-meshed mode shapes from coarse modal measurements, which are subsequently mapped by the MS-FNO to fine-meshed stiffness fields. This enables detailed and spatially continuous representation of structural damage.

The end-to-end pipeline consists of two branches: a numerical branch and a physical branch.

In the numerical branch, a finite element (FE) model is constructed based on the target structure. Stiffness fields are generated using Gaussian Random Fields (GRFs), and the corresponding modal shapes are simulated through the FE model. These simulations form the numerical dataset, BeamDI-Num Set, in which each sample contains a paired set of modal shapes and stiffness fields. To facilitate the training of MoSRNet, a down-sampled version, BeamDI-Num(DS) Set, is generated to provide coarse–fine modal data pairs. MoSRNet is trained using BeamDI-Num(DS) Set, while MS-FNO is trained using BeamDI-Num Set, as illustrated in Figure 2 (a).

In the physical branch, the target structure is instrumented with accelerometers (seven in the present experiment), and acceleration responses are collected under manual excitation. The modal shapes are extracted using Bayesian Operational Modal Analysis (BAYOMA) [39], resulting in the experimental dataset, BeamDI-Phy Set. Since the ground-truth stiffness field of the physical structure is unknown, a base state is selected from BeamDI-Phy Set. The predicted stiffness fields of other scenarios are subtracted by the prediction of the base state for bias calibration, yielding the final estimation of structural damage.

The detailed implementation of each module and process is described in the following sections.

2.2 Assumptions and Simplifications

To define the scope of this study and ensure the interpretability and reproducibility of the proposed framework, the following assumptions and simplifications are adopted.

First, the target structure is modeled as a beam, which is simplified into a two-dimensional representation. The material properties of each cross-section are assumed to be homogeneous and are represented by a single finite element.

Second, the FE model of the beam is established based on the Euler–Bernoulli beam theory, inheriting all the classical assumptions associated with this formulation.

Third, structural damage is introduced at the element level by reducing the bending stiffness while keeping the element mass constant. This assumption captures typical deterioration mechanisms such as cracking and corrosion, which primarily manifest as local stiffness reductions.

|  |
| --- |
|  |
| **Figure 2.** Overview of data flow and model architectures. (a) Relationship between data and models. Purple (solid) arrows indicate the data usage for model training and inference, while grey (dashed) arrows denote the data generation process. (b) Architecture of MoSRNet, consisting of three independent subnetworks to process each modal input. (c) Architecture of the MS-FNO, composed of three Fourier layers to learn the mapping from modal shapes to the stiffness field. (d) Architecture of ResNet, employed as the baseline model in this study. |

2.3 Modal Super-Resolution Network

The Modal Super-Resolution Network (MoSRNet) is designed to perform the upstream task of the proposed framework, namely super-resolving coarsely measured mode shapes into fine-meshed representations. MoSRNet comprises three independent subnetworks, each dedicated to processing one of the first three modes. Within each subnetwork, the architecture consists of a one-dimensional convolutional layer, followed by a Gaussian Error Linear Unit (GELU), another one-dimensional convolutional layer, a GELU activation, a batch normalization layer, a flattening operation, and a fully connected linear layer, as illustrated in **Figure 2(b)**.

Let the coarse input of the -th mode, , can be denoted as

where is the number of coarse spatial nodes. The subnetwork begins with a one-dimensional convolution to extract local spatial correlations from the coarse modal input. For the output at the -th location of the -th feature map, the operation is as follows:

where is the number of input channels, is the kernel size, is the padding size, and and are trainable kernel weights and bias respectively. The convolutional output is then passed through the Gaussian Error Linear Unit (GELU), which introduces smooth nonlinearity and probabilistic gating:

After the first convolutional transformation and GELU activation, a second convolutional layer followed by another GELU is applied to further enrich the receptive field from local to midrange scales. Batch normalization is then applied to stabilize training and accelerate convergence by normalizing each activation:

where and are the batch mean and variance respectively, and are learnable affine parameters, and is a small constant to prevent division by zero. The normalized multi-channel features are flattened to a vector:

where is the product of the spatial and channel dimensions. Finally, a fully connected linear layer projects this vector to the fine spatial resolution of the k-th mode, :

with and as trainable parameters, and *N* denoting the number of fine-meshed nodes. The three subnetworks operate independently on the first three modes, and their outputs are concatenated to produce the enhanced modal input for the downstream operator:

The design of this shallow architecture is grounded in the spectral characteristics of low-order structural modes. For a simply supported Euler–Bernoulli beam with uniform flexural stiffness and mass per unit length, the governing eigenvalue problem yields sinusoidal solutions.

where, denotes the span length. These first three modes are smooth and dominated by low spatial frequencies. Let the coarse sampling interval be and the maximum wavenumber of interest be . When the sampling interval satisfies the strict Nyquist constraint:

the coarse measurements preserve all essential low-frequency content required to represent the first three modes. Under this band-limited setting, the mapping from coarse to fine meshes can be written as:

where is the coarse input, is the fine target, and is the residual associated with high-frequency components. donates a simple linear operator which maps and . The fundamental reason of using the simple linear operator is that the structural mode shapes are naturally smooth and dominated by low-frequency components; the high-frequency part is negligible in energy and can be ignored in practical reconstructions.

Consequently, the two Conv1d layers act as learnable low-pass filters that capture the interpolation characteristics, while the final linear layer performs the resolution lift from to . This shallow configuration achieves accurate reconstruction of the dominant low-frequency modal content without incurring the optimization complexity or overfitting risks associated with deeper networks.

2.4 Modal-Stiffness Fourier Neural Operator

In this study, Fourier Neural Operator is used to learn the mapping from mode shapes to stiffness field. The model architecture is derived from the FNO architecture proposed by Li et al. [38], as shown in Figure 2 (c). Generally, the FNO consists of a lifting layer, several Fourier layers with iterative kernel operators, and a projection layer. The mapping operator is learned through the iteration of multiple kernel operators. In each iterative kernel operator, the data is updated from the input to the output via the composition of a non-local integral operator and a local, nonlinear activation function , as outlined in Equation (12).

where, is a linear transformation via convolution; is a kernel operator parameterized by input coefficient functions . and learnable parameters ; is a non-linear activation function whose action is defined component-wise; is the -dimensional spatial domain for the PDE.

Furthermore, the general form of the kernel operator can be defined as Equation (13). Here is the kernel function which can be learned from data: . The integral aggregates global contributions of input function across the domain , weighted by the kernel function , to capture non-local dependencies and encode global information into the output at .

By applying the convolution theorem, the integration in the spatial domain can be replaced with element-wise multiplication in the frequency domain. Consequently, Equation (13) is transformed into Equation (14), where and represent the Fourier transform and inverse Fourier transform, respectively, as defined in Equations (15) and (16). is the Fourier transform of the kernel function , which is directly parameterized in Fourier space. By learning the of multiple kernel operators, the FNO can establish the mapping between function spaces. d

The present problem aims to identify the stiffness field of a beam from its first to third mode shapes . This relationship can be formulated as a nonlinear mapping operator acting on function spaces:

where represents the underlying physical operator that links the spatially distributed mode shapes to the stiffness field, denotes the spatial domain of the beam and represents the space of square-integrable functions defined on the domain . The operator is implicitly determined by the eigenvalue problem of the Euler–Bernoulli beam,

where and denote the mass per unit length and the modal frequency, respectively. Under fixed boundary conditions and for a bounded stiffness field , corresponding mode shapes are smooth and bounded functions over the compact domain . Because the eigenfunctions depend continuously on the structural parameters and , small perturbations in correspond to proportionally small variations in . Consequently, the operator is continuous in the -norm. This property ensures the stability of the inverse mapping, meaning that small perturbations in the modal information will not lead to unbounded changes in the stiffness field.

According to the universal approximation theorem of the FNO [40,41], any continuous operator between infinite-dimensional Banach spaces can be approximated to arbitrary accuracy by a properly parameterized FNO. Therefore, the continuous mapping can be learned by the proposed FNO framework.

2.5 Gaussian Random Field

Previous modal parameter– and machine learning–based methods have typically relied on idealized assumptions to generate training data and validate their models. In most of these studies, the structures are divided into several discrete segments, and damage is simulated by reducing the stiffness of one or several segments while keeping the others intact. However, such simplifications rarely occur in real engineering practice, where local deterioration often affects the surrounding regions due to stress redistribution and material continuity. In realistic conditions, damage should be represented as a continuous field to better reflect actual deterioration patterns and to improve model feasibility and robustness.

In this study, one-dimensional GRFs are employed to simulate the continuous stiffness loss field of the target structure, as illustrated in Figure 2(a). GRFs enable smooth transitions between adjacent elements while maintaining control over the statistical characteristics of the dataset, thereby providing a more realistic and diverse representation of structural damage.

The GRF is characterized by a covariance structure that governs its spatial correlation and smoothness. A GRF over the domain is defined by its covariance operator :

|  |  |  |
| --- | --- | --- |
|  |  | (19) |

where denotes the standard deviation, controlling the magnitude of the random field. The term represents the Laplace operator, responsible for introducing spatial correlation. The parameter is a regularization term that ensures numerical stability by preventing high-frequency components from dominating the field. Finally, determines the decay rate of the spectral components, controlling the smoothness of the GRF. Larger values of lead to faster decay of high-frequency components, resulting in smoother fields.

The covariance operator can be decomposed into eigenfunctions and eigenvalues for efficient representation. Under periodic boundary conditions, the eigen functions of the operator are given by:

|  |  |  |
| --- | --- | --- |
|  |  | (20) |

where represents the frequency index. These eigenfunctions form a complete orthonormal basis on the domain . The corresponding eigenvalues are:

|  |  |  |
| --- | --- | --- |
|  |  | (21) |

where represents the spectral contribution of the Laplace operator, and adds a stabilizing offset.

To approximate the GRF numerically, the field is represented using a truncated Fourier series. Retaining the first Fourier modes, the GRF can be expressed as:

|  |  |  |
| --- | --- | --- |
|  |  | (22) |

where is the mean value of the GRF, representing the average stiffness across the truss members. The coefficients and are independent Gaussian random variables that determine the contributions of cosine and sine terms, respectively. These coefficients are sampled based on the eigenvalues of the covariance operator, ensuring that the GRF satisfies the desired statistical properties.

The variances of and are determined by the eigenvalues as:

|  |  |  |
| --- | --- | --- |
|  |  | (23) |

This relationship ensures that lower-frequency components (smaller ) dominate the field, while higher-frequency components are suppressed, reflecting the smoothness imposed by .

The coefficients and are sampled as:

|  |  |  |
| --- | --- | --- |
|  |  | (24) |
|  |  | (25) |

where and are independent standard normal random variables, i.e.,

|  |  |  |
| --- | --- | --- |
|  |  | (26) |

Finally, the GRF is constructed by summing the weighted cosine and sine terms. This process ensures that the generated damage field adheres to the prescribed statistical properties, such as smoothness, mean value, and spatial correlation. The parameters , , and allow precise control over the magnitude, correlation length, and smoothness of the generated field, making the GRF an effective tool for modeling damage in truss structures.

2.6 Baseline Model

The one-dimensional Residual Network (ResNet) is selected as the baseline model in this study, as illustrated in Figure 2(d). ResNet is a convolutional neural network (CNN) architecture proposed by He et al. [42] to mitigate the gradient degradation problem in deep networks. By introducing residual (skip) connections, it facilitates the efficient training of deeper architectures and has been successfully applied in various fields, including image recognition, signal processing, and time-series analysis.

The one-dimensional variant of ResNet (1D ResNet) replaces two-dimensional convolutional kernels with one-dimensional ones to process sequential or spatially distributed data. This structure effectively captures local dependencies and multi-scale features in vibration or sensor signals, offering strong representational capability for regression and classification tasks based on one-dimensional measurements.

In the field of structural health monitoring (SHM), 1D ResNet has been employed for damage detection and dynamic response reconstruction [43–45]. Owing to its proven ability to handle one-dimensional structural response data and its frequent use in SHM studies, 1D ResNet is adopted as the baseline model in this research. The comparison between the ResNet and the proposed FNO-based framework enables a fair evaluation of how operator learning can outperform conventional convolutional architectures in capturing global physical dependencies and mapping modal parameters to stiffness fields.

3 DATA AND IMPLEMENTATION

3.1 Target Structure and Data Collection

|  |  |  |
| --- | --- | --- |
| **Table 2** Properties of beam FE model | | |
| **Properties** | **Values** | **Units** |
| Length | 5.4 | m |
| Young’s Modulus | 2.1×1011 | N/m2 |
| Mass Density | 7850 | kg/m3 |
| Moment of Inertia (Intact) | 5.709×10-7 | m4 |
| Cross-sectional area (Intact) | 6.542×10-3 | m2 |

The proposed framework is trained using numerically simulated data generated from a FE model of a simply supported beam. The FE model is constructed in two dimensions based on the Euler–Bernoulli beam theory and consists of 540 elements and 541 nodes. The mechanical and geometric properties of the beam are listed in Table 2.

Gaussian Random Fields (GRFs) are employed to simulate diverse stiffness distributions for model training and validation. The stiffness field of each sample,, is defined as follows:

where denotes the element index, and represent the elastic modulus and moment of inertia of the intact section, respectively, and is the Gaussian random field with zero mean, as defined in Equation (22). The field is normalized such that its maximum absolute value equals one. The scaling factor is randomly sampled from the range [0.25, 0.75] and used to control the degree of stiffness reduction.

Following this procedure, 9000 independent stiffness fields are generated and used as inputs for FE simulations. The first to third mode shapes are extracted from each simulation and normalized by dividing by their maximum absolute value. The resulting stiffness fields and corresponding mode shapes constitute the BeamDI-Num Set, which is used to train the MS-FNO and ResNet models.

To train the MoSRNet, the mode shapes in BeamDI-Num Set are down-sampled at seven uniformly distributed points (dividing the beam span into eight equal segments) along the beam span and paired with the original fine-meshed data to form a new dataset, BeamDI-Num(DS) Set. Both datasets are divided into two subsets: 8000 samples for training and 1000 samples for validation.

3.2 Network Architecture and Design

The detailed architectural designs of MoSRNet, MS-FNO, and ResNet are summarized in Tables 3, 4, and 5, respectively.

The proposed MoSRNet takes low-resolution mode shape data as input and predicts super-resolved, high-resolution mode shape data. The input data are first separated into three tensors corresponding to the first three mode shapes. Each tensor is then processed individually through a subnetwork consisting of a one-dimensional convolutional layer, a Gaussian Error Linear Unit (GELU) activation, another one-dimensional convolutional layer followed by another GELU activation, a batch normalization layer, a flatten layer, and finally a linear layer. After processing through these subnetworks, all tensors are upsampled to match the resolution of the original fine-meshed mode shape data. The resulting tensors are concatenated along the channel direction to form the reconstructed high-resolution modal field, which is used for subsequent loss computation.

Both MS-FNO and ResNet take fine-meshed mode shape data as input and predict the corresponding stiffness field along the beam span. For both models, an additional positional encoding channel is concatenated with the modal input to embed spatial information.

For the MS-FNO, the concatenated input is first lifted to a higher-dimensional representation along the channel direction. It then passes through three Fourier layers, each containing two branches: a spectral convolution layer that learns the global mapping operator, and a one-dimensional convolutional layer with a kernel size of one that serves as a residual connection. Finally, two linear layers project the output tensor into a one-dimensional stiffness field along the beam span.

For the ResNet, the input data are first embedded into a higher-dimensional space using a linear layer. An initial convolutional layer with a kernel size of seven is then applied to capture global features. Subsequently, a sequence of four residual blocks extracts detailed representations. Each residual block consists of two one-dimensional convolutional layers with a kernel size of three, a batch normalization layer, a ReLU activation, and a residual connection. Afterward, a layer normalization layer and two fully connected layers are used to project the data into a one-dimensional output representing the stiffness field.

|  |  |  |
| --- | --- | --- |
| **Table 3** MoSRNet architecture used in this work | | |
| **Component** | **Operation/Description** | **Output shape** |
| Input | Low-resolution mode shapes | (3, 9) |
| Branching | Channel-wise splitting | 3×(1, 9) |
| Subnetwork 1 | Conv1d(1→16, k=3, pad=1)/GELU | 1×(16, 9) |
| Conv1d(16→32, k=3, pad=1)/GELU/BN | 1×(32, 9) |
| Flatten: (32, 9)→(288) | 1×(288) |
| Linear (288→541) | 1×(541) |
| Subnetwork 2 | Same as above | 1×(541) |
| Subnetwork 3 | Same as above | 1×(541) |
| Concatenation | Stack three subnet outputs along channel dimension | (3, 541) |
| Output | Super-resolved high-resolution mode shapes | (3, 541) |
| Note: “3 × (1, 9)” indicates that the input is split into 3 independent channels, each with shape (1, 9), which are processed by separate subnetworks. “GELU” stands for Gaussian Error Linear Unit. “BN” stands for Batch Normalization. | | |

|  |  |  |
| --- | --- | --- |
| **Table 4** MS-FNO architecture used in this work | | |
| **Component** | **Operation/Description** | **Output shape** |
| Input | Channels of mode shape | (3, 541) |
| Channel of positional encoding | (1, 541) |
| Concatenate | (4, 541) |
| Lifting | Linear (4→128) | (128, 541) |
| FNO layer 1 | Fourier1d(SpectralConv1d 128→128, modes=16) | (128, 541) |
| Conv1d(128→128, k=1) | (128, 541) |
| Add/GELU | (128, 541) |
| FNO layer 2 | Same as above | (128, 541) |
| FNO layer 3 | Same as above | (128, 541) |
| Transpose | Transpose | (541, 128) |
| Projection layer 1 | Linear (128→64) | (541, 64) |
| Projection layer 2 | Linear (64→1) | (541, 1) |
| Output | Damage field | (541, 1) |
| Note: “Fourier1d” denotes a Fourier layer with one-dimensional Fourier transform. “SpectralConv1d” is a one-dimensional spectral convolution layer in the frequency domain. “Conv1d” is a one-dimensional convolution layer in the spatial domain. | | |

|  |  |  |
| --- | --- | --- |
| **Table 5** ResNet architecture used in this work | | |
| **Component** | **Operation/Description** | **Output shape** |
| Input | Channels of mode shapes | (3, 541) |
| Channel of positional encoding | (1, 541) |
| Concatenate | (4, 541) |
| Transpose | Transpose | (541, 4) |
| Embedding | Linear (4→128) | (541, 128) |
| Transpose | Transpose | (128, 541) |
| Initial convolution | Conv1d(128→128, k=7, stride=1, pad=3)/BN/ReLU | (128, 541) |
| Residual Block Group 1 | 2×BasicBlock1D(Conv1d(128→128, k=3, stride=1)/  BN/ReLU/RC) | (128, 541) |
| Residual Block Group 2 | Same as above | (128, 541) |
| Residual Block Group 3 | Same as above | (128, 541) |
| Residual Block Group 4 | Same as above | (128, 541) |
| Transpose | Transpose | (541, 128) |
| Normalization | LN | (541, 128) |
| Projection layer 1 | Linear (128→64) | (541, 64) |
| Projection layer 2 | Linear (64→1) | (541, 1) |
| Output | Damage field | (541, 1) |
| Note: “ReLU” means Rectified Linear Unit. “RC” refers to Residual Connection. “LN” stands for Layer Normalization | | |

3.3 Training Configuration and Hyperparameters

The performance of deep learning models is strongly influenced by their training configuration and hyperparameter settings. This section introduces the training setup, including the loss function, evaluation metrics, optimizer, learning rate scheduler, and other relevant hyperparameters.

All three models are trained purely data driven, with relative L2 norm as loss function. The relative L2 norm is defined as follows:

where, is the batch size, is the number of features (length of the stiffness field) of each sample, is the prediction of *j*-th feather in the *i*-th sample and is the target value of *j*-th feather in the *i*-th sample.

To objectively assess the performance of proposed models, two evaluation metrics are adopted, including mean absolute error (*MAE*) and mean absolute percentage error (*MAPE*), which are expressed as follows:

Other key training hyperparameters are summarized in Table 6. All models are trained with a batch size of 16. The MoSRNet is trained for 120 epochs, while both MS-FNO and ResNet are trained for 170 epochs.

For optimization, all models employ the Muon optimizer, which orthogonalizes updates to hidden-layer weight matrices via Newton–Schulz iterations. This process alleviates ill-conditioned gradient updates and improves convergence stability and training efficiency compared with conventional first-order optimizers [46]. The learning rate is set to 0.001, and the weight decay coefficient is fixed at 0.01.

A custom exponential learning rate scheduler is used for all models. The first 20 epochs serve as a warm-up phase, during which the learning rate increases linearly from zero to the target value. After the warm-up, the learning rate decays exponentially with a factor of 0.97 for MoSRNet and 0.975 for the other two models.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 6** Hyperparameters of model training | | | |
| **Parameters** | **MoSRNet** | **MS-FNO** | **ResNet** |
| **Epochs** | 120 | 170 | 170 |
| **Batch size** | 16 | 16 | 16 |
| **Optimizer** | Muon | Muon | Muon |
| Learning rate | 0.001 | 0.001 | 0.001 |
| Weight decay | 0.01 | 0.01 | 0.01 |
| **Scheduler** | Custom ExpLR | Custom ExpLR | Custom ExpLR |
| Initial ratio | 0 | 0 | 0 |
| Warmup epochs | 20 | 20 | 20 |
| Decay rate | 0.97 | 0.975 | 0.975 |

3.4 Training Dynamics and Convergence Analysis

|  |
| --- |
|  |
|  |
|  |
| **Figure 3** Training and validation loss curves: (a) MoSRNet, (b) MS-FNO, (c) ResNet |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 7** Training details and evaluation metrics | | | | | | | | |
| **Model** | **Params** | **Training time** | **Relative L2 Norm** | | **MAE** | | **MAPE** | |
| **Train** | **Eval** | **Train** | **Eval** | **Train** | **Eval** |
| MoSRNet | 474135 | 13m 34s | 0.00342 | 0.00442 | 0.00144 | 0.00222 | 0.767% | 1.861% |
| MS-FNO | 844929 | 21m 19s | 0.00351 | 0.00395 | 0.00227 | 0.00256 | 0.245% | 0.277% |
| ResNet | 916865 | 49m 59s | 0.01997 | 0.02113 | 0.01696 | 0.01795 | 1.765% | 1.130% |

Figure 3 illustrates the training and validation loss curves of all models, where subfigures (a)–(c) correspond to the MoSRNet, MS-FNO, and ResNet, respectively. The MoSRNet converges with an evaluation relative L2-norm of 0.00442, while the MS-FNO and the baseline ResNet achieve 0.00395 and 0.02113, respectively. The figure demonstrates that all models converge smoothly during training, showing no evidence of underfitting or overfitting.

Table 7 summarizes the model configurations and training performance. All models are trained on a workstation equipped with an AMD Ryzen™ 9 7950X3D @ 4.2 GHz CPU and an NVIDIA GeForce RTX 4090 GPU. The MoSRNet contains 474135 parameters and requires 13 minutes 34 seconds for training. The MS-FNO includes 844929 parameters with a training time of 21 minutes 19 seconds. The baseline ResNet has 916865 parameters and requires 49 minutes 59 seconds for training.

It is evident that, compared with the baseline ResNet, the proposed MS-FNO achieves higher prediction accuracy across all evaluation metrics while using fewer parameters (0.92×) and significantly reducing the training time (0.43×).

|  |  |
| --- | --- |
|  |  |
| **(a)** Original data (MS-FNO) | **(b)** Original data (ResNet) |
|  |  |
| **(c)** MoSRNet-super-resolved data (MS-FNO) | **(d)** MoSRNet-super-resolved data (ResNet) |
| **Figure 4** Numerical validation results of MS-FNO and ResNet using original high-resolution mode shape data and MoSRNet-super-resolved data after down-sampling | |

4 NUMERICAL VALIDATIONS

4.1 In-Distribution Validation

This section evaluates the performance of the models on in-distribution data. In-distribution testing assesses how well a model performs on data drawn from the same distribution as that used for training, thereby reflecting its fitting capability and generalization performance within known scenarios.

Four testing scenarios are designed. The first two directly evaluate the proposed MS-FNO and the baseline ResNet using high-resolution mode shape data. In the third and fourth scenarios, the original mode shape data are down-sampled at seven uniformly distributed points (dividing the beam span into eight equal segments) and subsequently super-resolved by the MoSRNet. The super-resolved mode shapes are then used as input for validation of the MS-FNO and ResNet, respectively.

Figure 4 presents an example from the validation set. As shown in Figure 4(a) and (b), when using high-resolution mode shape data as input, the MS-FNO prediction closely matches the ground truth, while the ResNet also achieves satisfactory accuracy with only minor discrepancies. Figure 4(c) shows the MS-FNO prediction based on MoSRNet-super-resolved data. The overall trend aligns well with the ground truth; however, noticeable local discrepancies appear near the beam boundaries. In contrast, Figure 4(d) illustrates the ResNet prediction on MoSRNet-super-resolved data, which exhibits sawtooth-like fluctuations. Although the general trend remains comparable to the ground truth, the local deviations are substantial, making it difficult to extract meaningful damage information directly.

The in-distribution numerical validation confirms that the proposed MS-FNO consistently outperforms the baseline ResNet, both when using high-resolution mode shape data and when using MoSRNet-super-resolved inputs.

|  |  |
| --- | --- |
|  |  |
| **(a)** MS-FNO | **(b)** ResNet |
| **Figure 5** Comparison of the mean prediction error on the validation set and the intact scenario prediction using the framework (MoSRNet + downstream model) | |
|  | |  |
| **(a)** MoSRNet + MS-FNO | | **(b)** MoSRNet + ResNet |
| **Figure 6** Numerical validation results of MS-FNO and ResNet using MoSRNet-super-resolved data after down-sampling with bias compensation | | |

4.2 Bias Calibration

Applying the MoSRNet essentially establishes a mapping from a low-dimensional to a high-dimensional space, which inevitably introduces systematic errors. To mitigate this effect, a batch normalization layer is incorporated into each subnetwork to learn and compensate for the overall deviation within the dataset. To verify whether the bias introduced by MoSRNet is consistent across all samples, the following validation procedure was conducted for both MS-FNO and ResNet:

(1) all samples in the validation set were input to obtain the average prediction error along the beam length;

(2) the vibration mode corresponding to the intact state was input to obtain its predicted stiffness field; and

(3) the two results were compared.

The results are illustrated in Figure 5. It can be observed that the average error distribution along the beam span is almost identical to the predicted stiffness field of the intact state, indicating that the bias introduced by MoSRNet remains nearly consistent across different samples. Therefore, the prediction bias can be effectively eliminated by subtracting the prediction of the intact (or any other reference) state from that of each scenario, leading to the bias-calibrated prediction.

Figure 6 presents the same sample as in Figure 4 after bias calibration. It is evident that the performance of both MS-FNO and ResNet improves significantly following calibration. To further quantify the effectiveness of this procedure, the coefficient of determination (R2) is introduced, which is defined as:

The averaged R2 values of both models before and after bias calibration are listed in Table 8. For the MS-FNO, bias calibration increases R2 from 0.920 to 0.950, while for the ResNet, it improves from 0.699 to 0.891, transforming originally less reliable predictions into practically meaningful results. These findings confirm that bias calibration is both essential and effective. All subsequent validation results presented in this study are based on bias-calibrated predictions.

|  |  |  |
| --- | --- | --- |
| **Table 8** Comparison of R2 on the validation set before and after bias compensation | | |
| Models | R2 (before) | R2 (after) |
| MoSRNet + MS-FNO | 0.920 | 0.950 |
| MoSRNet + ResNet | 0.699 | 0.891 |

|  |
| --- |
|  |
| **Figure 7** Example of damage filed used for the Monte Carlo Simulation |

|  |  |
| --- | --- |
|  |  |
| **(a)** Original data (MS-FNO) | **(b)** Original data (ResNet) |
|  |  |
| **(c)** MoSRNet-super-resolved data (MS-FNO) | **(d)** MoSRNet-super-resolved data (ResNet) |
| **Figure 8** Confusion matrixes of Monte Carlo Simulation -based damage localization results | |

4.3 Out-of-Distribution Validation

In real-world applications, one of the major challenges lies in the unpredictability of damage patterns. Although the introduction of Gaussian Random Fields (GRFs) for data generation can partially address this issue, a more comprehensive evaluation of model robustness and generalization capability requires out-of-distribution (OOD) testing.

The OOD numerical test is designed as a damage localization task. First, a random stiffness field containing a single damage region is generated, as illustrated in Figure 7. The damage location is randomly sampled from element No. 10 to No. 530, and its severity is randomly assigned within a range of 15%–60% stiffness loss. On both sides of the damage center, the stiffness varies linearly from the damaged value to 100%, forming a continuous transition zone. The remaining elements are randomly assigned a stiffness perturbation between −5% and 5% to simulate noise conditions (where negative damage represents stiffness gain). The generated stiffness field is then input into the finite element model to compute the corresponding mode shapes. These mode shapes are subsequently fed into both the proposed framework and the baseline ResNet model. After bias calibration using the intact state as the base state, the location of the maximum predicted stiffness reduction is identified as the predicted damage location.

The aforementioned process is defined as one run. In this test, 8000 runs are performed as a Monte Carlo Simulation (MCS) to evaluate the damage localization performance of each model on out-of-distribution data. The beam is divided into several equal segments. If the predicted damage location and the ground truth fall within the same segment, the prediction is considered a positive sample; otherwise, it is treated as negative. The number of segments is defined as mesh density.

Specifically, true positives (TP) refer to correctly localized damage within the same segment as the ground truth. False positives (FP) correspond to cases where a damaged segment is incorrectly predicted in an intact region. False negatives (FN) occur when the model fails to identify an actual damaged segment, and true negatives (TN) denote correctly recognized intact regions. These four quantities are used to compute the metrics of accuracy, precision, and recall for quantitative evaluation.

To assess classification performance, three commonly used metrics—accuracy (Acc), precision (Pre), and recall (Rec)—are adopted.

Accuracy represents the overall proportion of correctly predicted samples, precision indicates the proportion of correctly predicted positive samples among all predicted positives, and recall measures the proportion of correctly predicted positives among all actual positives. Their mathematical formulations are expressed as:

The confusion matrices for ten-segment localization are shown in Figure 8. Figures 8(a)–(b) indicate that when high-resolution mode shape data are used as input, both MS-FNO and ResNet achieve high accuracy—97.5% and 93.2%, respectively. Figure 8(c) presents MS-FNO’s predictions using MoSRNet-super-resolved data, achieving an overall accuracy of 80.4%. The main errors are concentrated near the beam boundaries, while most central segments still exhibit high precision and recall. In contrast, Figure 8(d) shows ResNet’s predictions based on MoSRNet-super-resolved data, with a total accuracy of 64.1%. The prediction errors are nearly uniformly distributed across all segments, making the output difficult to interpret for practical applications.

Table 9 summarizes the macro accuracy and recall for different mesh densities, including results calculated only for the central 60% of the beam. When using high-resolution input, the accuracy of MS-FNO remains relatively stable with increasing mesh density, whereas ResNet performance degrades significantly—from 93.32% to 83.59% as the mesh density increases from 10 to 20. When using MoSRNet-super-resolved data, MS-FNO achieves accuracies of 80.39%, 72.13%, and 67.13% for mesh densities of 10, 15, and 20, respectively, while ResNet attains 67.06%, 47.75%, and 41.73%. When considering only the central 60% of the beam, MS-FNO further improves to 93.67%, 90.27%, and 82.31% for the same mesh densities, whereas ResNet shows no significant improvement.

In summary, the proposed framework demonstrates strong robustness and generalization on out-of-distribution data, outperforming the baseline ResNet in all cases. When focusing on the central 60% region and setting a recall threshold of 80% as the criterion for effective damage identification, the proposed framework maintains satisfactory performance even at a mesh density of 20 using only seven sensors—exceeding the capabilities of most existing studies, where the number of identifiable segments typically matches the number of sensors.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table 9** Performance metrics of Monte Carlo Simulation-based damage localization results | | | | | |
| Mesh Density | Metrics | MS-FNO (Ori) | ResNet (Ori) | MS-FNO (SR) | ResNet (SR) |
| 10 elements | Acc | 97.46% | 93.23% | 80.39% | 64.06% |
| Rec | 97.49% | 93.18% | 80.76% | 64.15% |
| Acc (C60) | 98.86% | 89.76% | 93.67% | 70.02% |
| Rec (C60) | 98.89% | 92.67% | 94.22% | 88.41% |
| 15 elements | Acc | 96.46% | 87.65% | 72.13% | 47.75% |
| Rec | 96.50% | 87.65% | 72.38% | 47.56% |
| Acc (C60) | 98.24% | 85.32% | 90.27% | 54.92% |
| Rec (C60) | 98.26% | 89.85% | 91.92% | 62.66% |
| 20 elements | Acc | 94.94% | 83.59% | 67.13% | 41.73% |
| Rec | 95.17% | 85.47% | 70.41% | 53.83% |
| Acc (C60) | 97.97% | 79.97% | 82.31% | 49.59% |
| Rec (C60) | 97.99% | 84.92% | 84.85% | 67.03% |
| Ori: original data, SR: MoSRNet super-resolved data, Acc: accuracy, Rec: recall rate, C60: performance metrics of central 60% span | | | | | |

|  |  |  |
| --- | --- | --- |
| **Table 10** Experiment scenarios | | |
| **Scenario Names** | **DMG1** | **DMG2** |
| MCUT | ● | ○ |
| WEDG | ○ | ● |
| WCUT | ● | ● |
| REIN | ○ | ○ |
| ● Damaged, ○ Reinforced | | |

5 EXPERIMENTAL VALIDATIONS

5.1 Experiment Design and Platform

The proposed framework is experimentally validated on a test platform established at Kyoto University, using a steel beam specimen with the same mechanical and geometric properties as those listed in Table 2. As shown in Figure 9, the target structure is a simply supported steel beam with an I-shaped cross-section. The weak axis of the beam is oriented along the loading direction. The span length of the beam is 5.4 m, and the boundary conditions are realized through a pin support at one end and a roller support at the other.

Two distinct damage scenarios are designed to simulate different deterioration types. On the left side of the beam, DMG1 consists of three vertical cuts on the bottom flange, while on the right side, DMG2 is a long wedge-shaped cut on the bottom flange. Both damage configurations are equipped with detachable reinforcement plates, allowing the beam to be tested in damaged or reinforced states, as illustrated in Figure 9. The sectional details of the intact region, DMG1, DMG2, and their reinforced configurations are also presented in the figure.

By selectively attaching or removing the reinforcement plates at DMG1 and DMG2, four experimental scenarios are constructed: MCUT (multiple cuts), WEDG (wedged), WCUT (combined wedge and multiple cuts), and REIN (reinforced), as summarized in Table 10. In the experimental validation, the REIN scenario is designated as the base state for bias calibration.

|  |
| --- |
|  |
| **Figure 9** Mode shape data from laboratory experiment |

|  |
| --- |
|  |
| **Figure 10** Mode shape data from laboratory experiment |

5.2 Data Acquisition and Processing

In practical applications, mode shapes can be extracted through system identification using acceleration responses. To collect acceleration data, seven accelerometers (EPSON LP-WS92-EACS01-2) were installed at seven uniformly distributed points along the beam span, dividing it into eight equal segments, as shown in Figure 9. During the experiment, the beam was manually excited, and the resulting damped free-vibration responses were recorded.

For each experimental scenario listed in Table 10, four independent measurements were conducted. The mode shapes for each case were identified from the acceleration data using the Bayesian Operational Modal Analysis (BAYOMA) method, in which modal parameters are represented by the maximum posterior estimates. Details of the BAYOMA method can be found in the study by Au et al. [39]. The averaged mode shape from the four repeated measurements was regarded as the representative result for each scenario.

All averaged mode shapes were normalized such that their maximum absolute amplitude equals one, following the same normalization procedure applied to the numerical data used for model training. Through this process, the experimental dataset (BeamDI-Phy Set) was constructed.

The measured mode shapes for all scenarios are presented in Figure 10. It can be observed that the overall shapes are highly similar across different scenarios, making it difficult to directly identify damage locations from the raw modal data alone.

5.3 Validation under Different Scenarios

|  |  |
| --- | --- |
|  |  |
| **(a)** MCUT (MS-FNO) | **(b)** MCUT (ResNet) |
|  |  |
| **(c)** WEDG (MS-FNO) | **(d)** WEDG (ResNet) |
|  |  |
| **(e)** WCUT (MS-FNO) | **(f)** WCUT (ResNet) |
| **Figure 11** Experimental validation results of MS-FNO and ResNet in different scenarios | |

This section presents the experimental validation results of the proposed framework and the baseline model. The REIN scenario is selected as the base state for bias calibration. The predictions of both models are shown in Figure 11. In each figure, the red regions indicate stiffness loss, the blue regions represent stiffness increase, and the grey regions denote the theoretical damage pattern. The latter is obtained by subtracting the stiffness field of the base-state (REIN) scenario from the theoretical stiffness field. It should be noted that the grey regions only represent the theoretical shape of the damaged areas rather than the exact ground truth of the stiffness field.

When computing the theoretical stiffness of each section, the Composite Section Method is employed, which treats the composite cross-section as a single equivalent section. This method assumes perfect bonding and linear compatibility between components, which may lead to an overestimation of stiffness due to connection flexibility or interfacial slip. Furthermore, the sharp stiffness discontinuities observed in the grey regions do not occur in practice, as actual stiffness transitions are gradual owing to geometric continuity and material deformation effects.

Figures 11(a), (c), and (e) present the experimental validation results of the proposed framework. For the MCUT and WEDG scenarios, the proposed framework successfully identifies the damaged regions. The predicted damage boundaries are close to the actual sections where stiffness changes occur, and the gradual stiffness transitions are captured as expected. In the WCUT scenario, where multiple damages coexist, the prediction of the wedge-shaped region is partially influenced; nevertheless, the framework still clearly identifies both damaged areas and their approximate extents

Figures 11(b), (d), and (f) show the corresponding results of the baseline ResNet model. For all three scenarios, the predicted stiffness fields exhibit high-frequency sawtooth fluctuations. Although the baseline model can roughly locate the damaged regions, its predictions are less accurate and physically consistent compared with the proposed framework. In the WCUT scenario, where two damages of different severities occur simultaneously, the baseline model fails to distinguish their relative magnitudes.

For both models, localized stiffness increases are observed in intact regions. This phenomenon arises from the statistical properties of the training dataset. Since only mode shapes are used as input features, the global stiffness level remains implicitly fixed, as frequency information—which could indicate overall stiffness variation—is not included. Because all stiffness fields in the training data are normalized to have an average value of one, the trained models tend to produce predictions centered around this mean. This limitation could potentially be mitigated by incorporating modal frequency data as an additional input feature in future studies.

Based on the above results, it can be concluded that the proposed framework demonstrates strong feasibility and robustness in experimental validation. The MoSRNet effectively supports both the MS-FNO and ResNet models in high-resolution damage identification, while the proposed MS-FNO consistently outperforms the baseline model in terms of accuracy and stability.

6 CONCLUSIONS AND OUTLOOK

This study presented a synergistic framework that combines a Modal Super-Resolution Convolutional Neural Network (MoSRNet) and a Modal–Stiffness Fourier Neural Operator (MS-FNO) to achieve high-resolution damage identification in bridge structures. Through comprehensive numerical and experimental validations, the proposed framework demonstrated superior performance compared with the baseline ResNet model, confirming its efficiency, robustness, and potential for practical implementation in SHM systems.

A key advancement of this work lies in the introduction of operator learning into modal-parameter-based SHM. By leveraging the Fourier Neural Operator, the proposed framework establishes a direct mapping between modal response fields and the corresponding stiffness fields. This formulation not only provides a more physically interpretable and computationally efficient approach to solving inverse problems but also enables accurate reconstruction of spatially continuous stiffness loss distributions across structural spans.

In addition, the MoSRNet effectively reconstructs fine-meshed modal information from sparse sensor measurements, thereby maximizing the utility of limited sensor data. The integration of MoSRNet with MS-FNO allows for high-resolution damage localization even when only coarse modal measurements are available, addressing one of the most critical limitations in vibration-based SHM.

Another important contribution is the use of Gaussian Random Fields (GRFs) to generate training data with continuous and spatially correlated stiffness variations. This approach moves beyond conventional segment-based damage assumptions, providing a more realistic simulation of structural deterioration. As a result, the trained model exhibits enhanced generalization capability and improved robustness against out-of-distribution damage patterns.

Furthermore, the framework successfully achieves reliable stiffness field prediction using only the first three mode shapes, which are typically the most easily measurable in practical monitoring scenarios. This highlights the method’s applicability under realistic sensing constraints and its potential to reduce the cost and complexity of SHM systems.

Overall, the proposed MoSRNet–MS-FNO framework demonstrates strong generalization, stability, and interpretability across both numerical simulations and experimental validations. Nonetheless, the model’s sensitivity to minor damage (below approximately 15% stiffness loss) remains limited and requires further improvement. Future developments will therefore focus on enhancing the framework’s capability to detect small-scale damages with higher precision, extending it toward multi-source data fusion and adapting the model to accommodate varying boundary conditions. From an application perspective, the proposed framework shows strong potential for deployment in large-scale continuous beam structures as part of automated SHM systems, and it may also serve as a valuable tool for rapid post-earthquake reliability assessment of bridge infrastructures.

CRediT Authorship Contribution Statement

**Jinghao Fei:** Conceptualization, Methodology, Investigation, Formal Analysis, Data Curation, Software, Validation, Visualization, Writing – Original Draft. **Chul-Woo Kim:** Conceptualization, Supervision, Funding Acquisition, Resources, Project Administration,Writing – Review & Editing. **Debao Chen:** Methodology, Investigation, Validation, Writing – Review & Editing. **Hongkang Pan:** Investigation, Formal Analysis, Visualization, Writing – Review & Editing. **Rongxiu Chen:** Validation, Software, Writing – Review & Editing. **Jiaji Wang:** Methodology, Supervision,Writing – Review & Editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The models and related data shown in this paper will be open sourced on the project webpage in GitHub: https://github.com/Fei-JH/MoSRNet-MSFNO

Acknowledgment

This study is partially supported by a Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Scientific Research (B) under project No. 22H01576, JST SPRING, Grant Number JPMJSP2110 and JSPS Bilateral Joint Research Projects (2024 - 2026). Those financial supports are gratefully acknowledged.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the writing process of this paper, we utilized OpenAI's language model, GPT-5, to enhance the clarity, coherence, and precision of the manuscript. The model was used for tasks such as language polishing, grammar correction, and refining technical descriptions. All content generated or suggested by the model was carefully reviewed and edited by the authors to ensure academic rigor and domain accuracy. The study's core ideas, methodologies, experimental designs, and conclusions were developed entirely by the authors, with the language model serving solely as a tool to support the writing process.

References

[1] Frangopol DM, Liu M. Maintenance and management of civil infrastructure based on condition, safety, optimization, and life-cycle cost∗. Structure and Infrastructure Engineering 2007;3:29–41. https://doi.org/10.1080/15732470500253164.

[2] Morgese M, Ansari F, Domaneschi M, Cimellaro GP. Post-collapse analysis of Morandi’s Polcevera viaduct in Genoa Italy. J Civil Struct Health Monit 2020;10:69–85. https://doi.org/10.1007/s13349-019-00370-7.

[3] National Transportation Safety Board, USA. Collapse of the Fern Hollow Bridge 2022. https://www.ntsb.gov/investigations/AccidentReports/Reports/HIR2402.pdf.

[4] Hearn G. Bridge Inspection Practices. 2007. https://doi.org/10.13140/RG.2.1.5179.5604.

[5] Chupanit P, Phromsorn C. The importance of bridge health monitoring. International Science Index 2012;6:135–8.

[6] Shahidan S, Pulin R, Muhamad Bunnori N, Holford KM. Damage classification in reinforced concrete beam by acoustic emission signal analysis. Construction and Building Materials 2013;45:78–86. https://doi.org/10.1016/j.conbuildmat.2013.03.095.

[7] Li D, Chen Q, Wang H, Shen P, Li Z, He W. Deep learning-based acoustic emission data clustering for crack evaluation of welded joints in field bridges. Automation in Construction 2024;165:105540. https://doi.org/10.1016/j.autcon.2024.105540.

[8] Zhu S, Zhang D, Zhou M, Guan P, Duo S, Huang H. AI-driven non-destructive concrete stress state identification: A novel approach based on time–frequency analysis of active acoustic emission signals. Advanced Engineering Informatics 2025;68:103757. https://doi.org/10.1016/j.aei.2025.103757.

[9] Yu L, Giurgiutiu V. In situ 2-D piezoelectric wafer active sensors arrays for guided wave damage detection. Ultrasonics 2008;48:117–34. https://doi.org/10.1016/j.ultras.2007.10.008.

[10] Yang Y, Ng C-T, Kotousov A, Sohn H, Lim HJ. Second harmonic generation at fatigue cracks by low-frequency Lamb waves: Experimental and numerical studies. Mechanical Systems and Signal Processing 2018;99:760–73. https://doi.org/10.1016/j.ymssp.2017.07.011.

[11] Rautela M, Gopalakrishnan S. Ultrasonic guided wave based structural damage detection and localization using model assisted convolutional and recurrent neural networks. Expert Systems with Applications 2021;167:114189. https://doi.org/10.1016/j.eswa.2020.114189.

[12] Kaewnuratchadasorn C, Wang J, Kim C-W. Neural Operator for Structural Simulation and Bridge Health Monitoring 2023.

[13] Wang J, Mo YL, Izzuddin B, Kim C-W. Exact Dirichlet boundary Physics-informed Neural Network EPINN for solid mechanics. Computer Methods in Applied Mechanics and Engineering 2023;414:116184. https://doi.org/10.1016/j.cma.2023.116184.

[14] Kaewnuratchadasorn C, Wang J, Kim C-W. Physics-informed neural operator solver and super-resolution for solid mechanics. Computer-Aided Civil and Infrastructure Engineering 2024;39:3435–51. https://doi.org/10.1111/mice.13292.

[15] Ding W, He Q, Tong H, Wang Q, Wang P. Solving coupled differential equation groups using PINO-CDE. Mechanical Systems and Signal Processing 2024;208:111014. https://doi.org/10.1016/j.ymssp.2023.111014.

[16] Chen R, Kim C-W, Wang J. Real-time surrogate model of vehicle-bridge interaction using a multiple-input neural operator. Automation in Construction 2025;180:106576. https://doi.org/10.1016/j.autcon.2025.106576.

[17] Pathirage CSN, Li J, Li L, Liu W, Ni P. Structural damage identification based on autoencoder neural networks and deep learning. Engineering Structures 2018;172:13–28. https://doi.org/10.1016/j.engstruct.2018.05.109.

[18] Hao J, Zhu X, Yu Y, Zhang C, Li J. Damage localization and quantification of a truss bridge using PCA and convolutional neural network. Smart Structures and Systems 2022;30:673–86. https://doi.org/10.12989/SSS.2022.30.6.673.

[19] Lei Y, Li J, Hao H. Physics-guided deep learning based on modal sensitivity for structural damage identification with unseen damage patterns. Engineering Structures 2024;316:118510. https://doi.org/10.1016/j.engstruct.2024.118510.

[20] Fei J, Chen D, Pan H, Kim C-W. Application of Neural Operator for Damage Identification of a Simply Supported Beam. IABSE Symposium Tokyo 2025, 2025, p. 1227–34. https://doi.org/10.2749/tokyo.2025.1227.

[21] Avci O, Abdeljaber O, Kiranyaz S, Hussein M, Gabbouj M, Inman DJ. A review of vibration-based damage detection in civil structures: From traditional methods to Machine Learning and Deep Learning applications. Mechanical Systems and Signal Processing 2021;147:107077. https://doi.org/10.1016/j.ymssp.2020.107077.

[22] Ghiasi R, Torkzadeh P, Noori M. A machine-learning approach for structural damage detection using least square support vector machine based on a new combinational kernel function. Structural Health Monitoring 2016;15:302–16. https://doi.org/10.1177/1475921716639587.

[23] He H, Zheng J, Liao L, Chen Y. Damage identification based on convolutional neural network and recurrence graph for beam bridge. Structural Health Monitoring 2021;20:1392–408. https://doi.org/10.1177/1475921720916928.

[24] Sony S, Gamage S, Sadhu A, Samarabandu J. Vibration-based multiclass damage detection and localization using long short-term memory networks. Structures 2022;35:436–51. https://doi.org/10.1016/j.istruc.2021.10.088.

[25] Le-Xuan T, Bui-Tien T, Tran-Ngoc H. A novel approach model design for signal data using 1DCNN combing with LSTM and ResNet for damaged detection problem. Structures 2024;59:105784. https://doi.org/10.1016/j.istruc.2023.105784.

[26] Li M, Jia D, Wu Z, Qiu S, He W. Structural damage identification using strain mode differences by the iFEM based on the convolutional neural network (CNN). Mechanical Systems and Signal Processing 2022;165:108289. https://doi.org/10.1016/j.ymssp.2021.108289.

[27] Li X, Bolandi H, Masmoudi M, Salem T, Jha A, Lajnef N, et al. Mechanics-informed autoencoder enables automated detection and localization of unforeseen structural damage. Nat Commun 2024;15:9229. https://doi.org/10.1038/s41467-024-52501-4.

[28] Shao Y, Li L, Li J, An S, Hao H. Computer vision based target-free 3D vibration displacement measurement of structures. Engineering Structures 2021;246:113040. https://doi.org/10.1016/j.engstruct.2021.113040.

[29] Kong X, Li J. Vision-Based Fatigue Crack Detection of Steel Structures Using Video Feature Tracking. Computer-Aided Civil and Infrastructure Engineering 2018;33:783–99. https://doi.org/10.1111/mice.12353.

[30] Mei Q, Gül M. A cost effective solution for pavement crack inspection using cameras and deep neural networks. Construction and Building Materials 2020;256:119397. https://doi.org/10.1016/j.conbuildmat.2020.119397.

[31] Ali R, Chuah JH, Talip MSA, Mokhtar N, Shoaib MA. Structural crack detection using deep convolutional neural networks. Automation in Construction 2022;133:103989. https://doi.org/10.1016/j.autcon.2021.103989.

[32] Guo F, Qian Y, Liu J, Yu H. Pavement crack detection based on transformer network. Automation in Construction 2023;145:104646. https://doi.org/10.1016/j.autcon.2022.104646.

[33] Jiang W-J, Kim C-W. On the issue of global variability in long-term SHM: damage detection using SSA-KPCA-enhanced MSPC. Mechanical Systems and Signal Processing 2025;238:113188. https://doi.org/10.1016/j.ymssp.2025.113188.

[34] Guo T, Wu L, Wang C, Xu Z. Damage detection in a novel deep-learning framework: a robust method for feature extraction. Structural Health Monitoring 2020;19:424–42. https://doi.org/10.1177/1475921719846051.

[35] Wang R, Chencho, An S, Li J, Li L, Hao H, et al. Deep residual network framework for structural health monitoring. Structural Health Monitoring 2021;20:1443–61. https://doi.org/10.1177/1475921720918378.

[36] Wang Y, Song X, Li R, Yang F, Xiao Y, Zheng S, et al. Channel mode attention network for structural damage identification. Engineering Structures 2025;325:119389. https://doi.org/10.1016/j.engstruct.2024.119389.

[37] Guo X-Y, Fang S-E. Structural parameter identification using physics-informed neural networks. Measurement 2023;220:113334. https://doi.org/10.1016/j.measurement.2023.113334.

[38] Li Z, Kovachki N, Azizzadenesheli K, Liu B, Bhattacharya K, Stuart A, et al. Fourier Neural Operator for Parametric Partial Differential Equations 2021. https://doi.org/10.48550/arXiv.2010.08895.

[39] Au S-K, Zhang F-L, Ni Y-C. Bayesian operational modal analysis: Theory, computation, practice. Computers & Structures 2013;126:3–14. https://doi.org/10.1016/j.compstruc.2012.12.015.

[40] Chen T, Chen H. Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems. IEEE Transactions on Neural Networks 1995;6:911–7. https://doi.org/10.1109/72.392253.

[41] Kovachki N, Lanthaler S, Mishra S. On Universal Approximation and Error Bounds for Fourier Neural Operators n.d.

[42] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition, 2016, p. 770–8.

[43] Wang R, Li J, Li L, An S, Ezard B, Li Q, et al. Structural damage identification by using physics-guided residual neural networks. Engineering Structures 2024;318:118703. https://doi.org/10.1016/j.engstruct.2024.118703.

[44] Shi J, Shi H, Li J, Yu Z. Train-induced vibration response reconstruction for bridge damage detection with a deep learning methodology. Structures 2024;64:106496. https://doi.org/10.1016/j.istruc.2024.106496.

[45] Lu X, Wei C, Sun L, Xia Y, Zhang W. Transfer Reconstruction from High-Frequency to Low-Frequency Bridge Responses Under Vehicular Loading with a ResNet. Applied Sciences 2024;14:10927. https://doi.org/10.3390/app142310927.

[46] Muon: An optimizer for hidden layers in neural networks | Keller Jordan blog n.d. https://kellerjordan.github.io/posts/muon/ (accessed October 1, 2025).