NEURAL-VSI: VARIATIONAL SYSTEM IDENTIFICATION OF STRUCTURAL PARAMETER FIELDS IN HIGH-ORDER PDES

Xuyang Li
University of North Carolina
Charlotte, NC 28223, USA

Mahdi Masmoudi, Rami Gharbi, Nizar Lajnef & Vishnu Boddeti
Michigan State University
East Lansing, MI 48823, USA

John Harlim & Romit Maulik
The Pennsylvania State University
University Park, PA 16802, USA

ABSTRACT

Structural parameter identification in high-order partial differential equations (PDEs), such as the Euler-Bernoulli beam equation, remains challenging due to the computation of high-order derivative operators, particularly when structural parameters vary spatially. Current physics-informed machine learning approaches, including Physics-Informed Neural Networks (PINNs) and Universal Differential Equations (UDEs), typically require expensive automatic differentiation (AD) or adjoint calculations. These methods often fail when measurement data is noisy or boundary conditions (BCs) are unknown. This paper proposes Neural-VSI, a variational framework that parameterizes unknown fields with neural networks and reformulates the inverse problem using local variational forms with pre-computed integration weight matrices. Our method bypasses the heavy computational overhead of AD, allowing for the estimation of hard-to-measure structural parameter fields without utilizing boundary condition information. Experiments on the vibration of beams with two unknown distributed parameters demonstrate that this approach achieves a speedup of over 200$\times$ compared to strong-form baselines while maintaining robust identification accuracy.

1 INTRODUCTION

Structural System Identification (SI) (Sirca Jr & Adeli, 2012; Fan & Qiao, 2011; Ghorbani et al., 2020) is vital for Structural Health Monitoring (SHM) (Gharehbaghi et al., 2022; Li et al., 2024a), serving as the foundation for damage detection, prognosis, and remaining useful life estimation. Traditionally, SI relies on model updating (Mottershead et al., 2011) or operational modal analysis (Bajric et al., 2015). However, these methods typically assume linear, time-invariant behavior or lumped parameter models, struggling to capture spatially distributed nonlinearities, such as localized stiffness degradation or heterogeneous damping fields, that are critical for high-fidelity digital twins.

Recent advances in scientific machine learning (SciML) (Cuomo et al., 2022; Rackauckas et al., 2020) have introduced promising avenues for solving inverse parameter estimation problems. A dominant approach employs physics-informed learning, notably Universal Differential Equations (UDEs) (Rackauckas et al., 2020), which leverage adjoint sensitivity analysis to optimize parameters within dynamic models. While computationally efficient for ODEs, extending these approaches to structural dynamics governed by high-order PDEs (e.g., the fourth-order Euler-Bernoulli beam equation) often requires the Method of Lines (MOL) for spatial discretization (Li et al., 2024b; Masmoudi et al.). This dependence on numerical differentiation can lead to significant computational costs and necessitates the explicit incorporation of BCs.

On the other hand, PINNs (Raissi et al., 2019) have seen widespread adoption across SciML and civil structural applications (Li et al., 2022; Bolandi et al., 2022). Despite their success, standard

PINNs rely heavily on AD to compute high-order derivatives, which can become computationally expensive and memory-intensive for large-scale or deep neural networks, particularly in high-order PDE settings. To mitigate the derivative sensitivity of strong forms, weak-form variants, such as Variational PINNs (VPINNs) (Kharazmi et al., 2019), hp-VPINNs (Kharazmi et al., 2021), cv-PINN (Liu & Wu, 2023), and FastVPINNs (Ghose et al., 2024), have been developed. However, these methods typically rely on expensive online numerical quadrature during training. Even with local decompositions, the computational cost of AD scales linearly with the number of integration points, prohibiting scaling to high-resolution spatiotemporal data. Furthermore, the lack of inherent boundary adherence requires explicit enforcement, often causing optimization imbalances.

Another prominent direction lies in the SINDy framework (Brunton et al., 2016), particularly its weak-form extensions like WSINDy (Messenger & Bortz, 2021a;b) and WENDy (Bortz et al., 2023; Rummel et al., 2025). By integrating governing equations against test functions, these methods effectively mitigate noise sensitivity and bypass numerical differentiation. However, their primary objective is sparse structure identification to recover parsimonious models governed by global scalar coefficients. Consequently, they are ill-suited for reconstructing continuous, spatially varying parameter fields.

In this work, we propose Neural-VSI, a variational framework designed for the robust identification of spatially varying structural parameter fields (e.g., stiffness and damping). Leveraging the universal approximation capability of neural networks to parameterize these complex, continuous fields, a strategy established in NeuralSI (Li et al., 2022), Neural-VSI fundamentally reformulates the inverse optimization problem using a local weak formulation with pre-computed integration weight matrices. This shift from strong-form collocation to offline matrix operations yields three distinct contributions:

- *Computational efficiency:* The computationally expensive integral operations are reduced to offline-computed matrix multiplications, achieving a speedup of two orders of magnitude ($> 200\times$) over strong-form baselines.
- *BC independence:* The method relies solely on local conservation laws within the integration window, enabling accurate parameter identification without prior knowledge or reconstruction of global boundary conditions.
- *Parameter estimation accuracy:* The method accurately estimates parameter fields, achieving error levels about 50% lower than those of strong-form baselines.

2 METHOD

2.1 PROBLEM DESCRIPTION

We consider the governing equation for the transverse vibration of a beam:

$$F(t) = \frac{\partial^2}{\partial x^2}\left(E(x)I\frac{\partial^2 u}{\partial x^2}\right) + \rho A\frac{\partial^2 u}{\partial t^2} + C(x)\frac{\partial u}{\partial t} \tag{1}$$

where $u(x,t)$ is the displacement function with respect to space and time. E, I, ρ, and A are related to the material properties and the geometry of the beam. F is the distributed force applied to the beam. C represents damping which is related to energy dissipation in the structure. We define the spatially varying flexural stiffness as $E(x) = P(x)E_0$.

The inverse problem aims to identify the unknown parameter fields $P(x)$ and $C(x)$ given discrete noisy measurements of $u(x,t)$ and the external load $F(t)$.

2.2 LOCAL WEAK FORMULATION IN SPACE AND TIME

While weak formulations have been applied to time-domain ODEs (Li et al., 2025), extending to fourth-order PDEs requires handling extreme derivative sensitivity. We define a local spatio-temporal subdomain $\mathcal{D}_{k,m} = \Omega_k \times \mathcal{T}_m$, where $\Omega_k = [x_k, x_k + L_x]$ represents the local spatial window of length L_x, and $\mathcal{T}_m = [t_m, t_m + L_t]$ represents the local temporal window of duration L_t. Therefore, we construct compactly supported test functions $\Psi(x,t) = \phi(x)\varphi(t)$ using scaled polynomials of the form $\psi(s) = (1-s^2)^p$ mapped to the local intervals. Specifically, for a spatial window

Ω_k, the coordinate is $s(x) = \frac{2(x-x_k)}{L_x} - 1$, and for a temporal window $\mathcal{T}_m$, $s(t) = \frac{2(t-t_m)}{L_t} - 1$. By choosing $p \geq 2$, we ensure that the test functions and their derivatives vanish at the boundaries of $\mathcal{D}_{k,m}$, allowing for valid integration by parts (IBP).

Multiplying the governing equation by $\Psi(x,t)$ and performing IBP twice in space and time yields the following integral formulation:

$$\int_{\mathcal{T}_m}\int_{\Omega_k}\left[P(x)E_0I\frac{\partial^2 u}{\partial x^2}\frac{d^2\phi(x)}{dx^2}\varphi(t) + \rho A u(x,t)\phi(x)\frac{d^2\varphi(t)}{dt^2} - C(x)u(x,t)\phi(x)\frac{d\varphi(t)}{dt}\right]dxdt = \int_{\mathcal{T}_m}\int_{\Omega_k} F(x,t)\phi(x)\varphi(t)dxdt \tag{2}$$

Notably, the fourth-order spatial derivative requirement on data u is reduced to second-order, allowing for stable evaluation via finite differences. The temporal derivatives of displacement are entirely eliminated via IBP.

2.3 Neural network based structural parameter field estimation

Similar to the previous structural identification work (Li et al., 2022), we approximate $P(x)$ and $C(x)$ using two separate fully connected networks with residual connections. The networks take the spatial coordinate x as input, which is rescaled from the physical domain $[0, L_x]$ to the unit interval $[0, 1]$ to facilitate training. To ensure physical validity and numerical stability, we enforce positivity via Softplus activation and employ a custom bias initialization strategy (see Appendix A.1).

Unlike standard PINNs that rely on expensive AD or VPINNs (Kharazmi et al., 2019) requiring numerical quadrature at every iteration, we adopt a discrete weak-form approach (Li et al., 2025). Following this strategy, Lagrange polynomial bases are employed to construct the test functions on the measurement grid. Since the measurement data is assumed to be on a uniform spatiotemporal grid (typical for sensor arrays) and the test functions are fixed, the integrals of the test functions and their derivatives can be computed.

Consequently, these integrals are pre-computed offline and stored as constant integration weight matrices. Let $\mathbf{W}''_\phi$ and $\mathbf{W}_\phi$ denote the weight matrices representing the second derivative and the value of the spatial test functions, respectively, and $\mathbf{W}_\varphi$ be the temporal integration weights.

The continuous integral Eq. (2) is discretized into a vectorized matrix operation. For a batch of displacement data $\mathbf{u}$ (with pre-computed spatial derivatives $\mathbf{u}_{xx}$), the loss function $\mathcal{L}$ is:

$$\mathcal{L}_{(\theta_P,\theta_C)} = \left\| \underbrace{E_0I\mathbf{W}''_\phi(\hat{P}\odot\mathbf{u}_{xx})\mathbf{W}_\varphi^T}_{\text{Stiffness (Spatial IBP)}} + \underbrace{\rho A\mathbf{W}_\phi\mathbf{u}(\mathbf{W}''_\varphi)^T}_{\text{Inertia (Temporal IBP)}} - \underbrace{\mathbf{W}_\phi(\hat{C}\odot\mathbf{u})(\mathbf{W}'_\varphi)^T}_{\text{Damping (Temporal IBP)}} - \underbrace{\mathbf{W}_\phi\mathbf{F}\mathbf{W}_\varphi^T}_{\text{External Force}} \right\|_{\mathrm{F}}^2 \tag{3}$$

where $\odot$ denotes the element-wise Hadamard product, and $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm. Terms $\hat{P}(x)$ and $\hat{C}(x)$ represent the network predictions evaluated at the spatial grid points.

Crucially, this formulation allows the direct use of displacement $\mathbf{u}$, thereby eliminating numerical time differentiation for the inertial and damping terms. Simultaneously, it reduces the spatial differentiation requirement from fourth-order to second-order, significantly mitigating the numerical instability inherent in high-order strong-form solvers.

3 Results

Comprehensive details regarding the reference structural parameter field, training data generation and hyperparameter settings are provided in Appendix A.2 and Appendix A.3, respectively.

We begin our analysis by demonstrating the significant speed advantage of the proposed framework. The computational efficiency is evaluated in Figure 1 (Left). The strong-form baseline scales poorly with mesh size (2494s at spatial grid size $N_x = 21$), while Neural-VSI maintains a nearly constant total wall-clock time (until convergence) of $\sim 3.8s$. This yields a significant speedup of over

200×. This efficiency empirically validates our weak-form strategy: by pre-computing integration weight matrices offline, we replace expensive online AD and adjoint loops with instantaneous matrix multiplications.

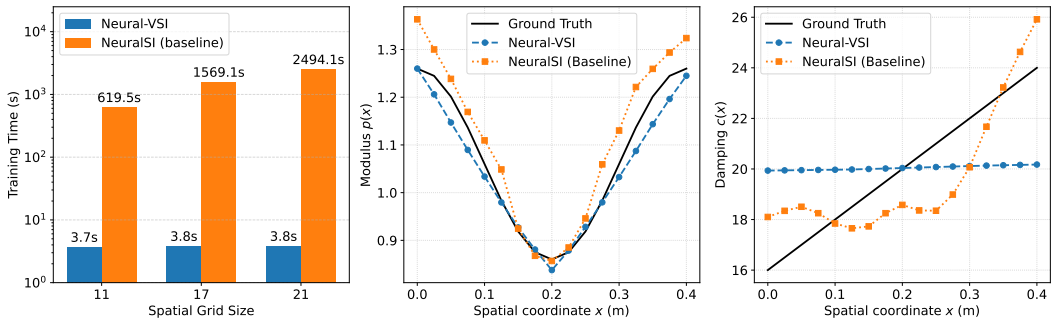

Figure 1: Performance evaluation of Neural-VSI. Left: Computational efficiency comparison across spatial grid sizes N_x. Middle: Stiffness $P(x)$ identification error at $N_x = 17$. Right: Damping $C(x)$ identification error at $N_x = 17$.

Next, we assess the identification accuracy of the spatially varying parameters in Figure 1 (Middle and Right), focusing on the performance at a spatial grid resolution of $N_x = 17$. Neural-VSI demonstrates superior accuracy in estimating the flexural stiffness, or modulus $P(x)$, a critical parameter for structural assessment. The Mean Absolute Error (MAE) for $P(x)$ is reduced to 0.023, which represents an improvement of approximately 50% compared to the strong-form baseline (MAE = 0.047).

Regarding the damping parameter $C(x)$, Neural-VSI yields a slightly higher MAE (1.67) compared to the baseline (1.44). However, this marginal trade-off is acceptable within the context of Structural Health Monitoring (SHM). In structural dynamics, the system response is predominantly governed by the modulus $P(x)$, which dictates natural frequencies and mode shapes, whereas damping forces are typically orders of magnitude smaller and inherently harder to identify. Consequently, the significant gain in stiffness reconstruction accuracy provided by Neural-VSI is of greater practical value for damage detection and prognosis than the slight variance in damping estimation.

We further extended our evaluation to varying noise levels. At the 1% noise level, Neural-VSI maintains its competitive advantage ($P(x)$ MAE: 0.030 vs 0.046). However, at the higher 5% noise level, performance degrades (MAE: 0.106 vs 0.054). This sensitivity stems from computing $\mathbf{u}_{xx}$ via finite differences on raw data, which amplifies high-frequency noise. While standard smoothing techniques could mitigate this, we intentionally omit them here to highlight the framework's intrinsic efficiency and raw baseline performance.

Baselines and Scope. We benchmark solely against NeuralSI (Li et al., 2022), which is the established strong-form baseline for this specific structural identification task. Other weak-form variants (e.g., VPINNs, FastVPINNs) are excluded from this comparison because existing formulations predominantly address second-order spatial problems and do not natively support the coupled spatiotemporal dynamics (i.e., fourth-order spatial derivatives with time integration) required by Euler-Bernoulli beam theory. Furthermore, the necessity of employing deep neural networks for parameter field estimation, verified through extensive ablation studies, has been empirically supported in Li et al. (2024b).

4 Conclusion and Discussion

This paper presented Neural-VSI, a variational framework for identifying spatially varying parameter fields in high-order PDE systems. By reformulating the governing equations into a local weak form, we symmetrically redistribute the differentiation between the trial solution and test functions. This strategy effectively bypasses the heavy computational overhead of AD and adjoint sensitivity analysis, achieving significant speedup and parameter estimation accuracy. Future research will

focus on extending the framework to meshless domains and irregular temporal sampling to accommodate realistic sensor configurations. Furthermore, we aim to integrate advanced smooth fitting techniques to improve parameter field reconstruction accuracy when dealing with extremely sparse data and high-level noise.

REFERENCES

Anela Bajric, Rune Brincker, and Sebastian Thöns. Evaluation of damping estimates in the presence of closely spaced modes using operational modal analysis techniques. In *6th International Operational Modal Analysis Conference*, 2015.

Hamed Bolandi, Gautam Sreekumar, Xuyang Li, Nizar Lajnef, and Vishnu Naresh Boddeti. Physics informed neural network for dynamic stress prediction. *arXiv preprint arXiv:2211.16190*, 2022.

David M Bortz, Daniel A Messenger, and Vanja Dukic. Direct estimation of parameters in ode models using wendy: Weak-form estimation of nonlinear dynamics. *Bulletin of Mathematical Biology*, 85(11):110, 2023.

Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, 113(15):3932–3937, 2016.

Salvatore Cuomo, Vincenzo Schiano Di Cola, Fabio Giampaolo, Gianluigi Rozza, Maziar Raissi, and Francesco Piccialli. Scientific machine learning through physics–informed neural networks: Where we are and what's next. *Journal of Scientific Computing*, 92(3):88, 2022.

Wei Fan and Pizhong Qiao. Vibration-based damage identification methods: a review and comparative study. *Structural health monitoring*, 10(1):83–111, 2011.

Vahid Reza Gharehbaghi, Ehsan Noroozinejad Farsangi, Mohammad Noori, TY Yang, Shaofan Li, Andy Nguyen, Christian Málaga-Chuquitaype, Paolo Gardoni, and Seyedali Mirjalili. A critical review on structural health monitoring: Definitions, methods, and perspectives. *Archives of computational methods in engineering*, 29(4):2209–2235, 2022.

Esmaeil Ghorbani, Oral Buyukozturk, and Young-Jin Cha. Hybrid output-only structural system identification using random decrement and kalman filter. *Mechanical Systems and Signal Processing*, 144:106977, 2020.

Divij Ghose, Thivin Anandh, and Sashikumaar Ganesan. Fastvpinns: A fast, versatile and robust variational pinns framework for forward and inverse problems in science. In *ICLR 2024 Workshop on AI4DifferentialEquations In Science*, 2024.

Ehsan Kharazmi, Zhongqiang Zhang, and George Em Karniadakis. Variational physics-informed neural networks for solving partial differential equations. *arXiv preprint arXiv:1912.00873*, 2019.

Ehsan Kharazmi, Zhongqiang Zhang, and George Em Karniadakis. hp-vpinns: Variational physics-informed neural networks with domain decomposition. *Computer Methods in Applied Mechanics and Engineering*, 374:113547, 2021.

Xuyang Li, Hamed Bolandi, Talal Salem, Nizar Lajnef, and Vishnu Naresh Boddeti. Neuralsi: structural parameter identification in nonlinear dynamical systems. In *European Conference on Computer Vision*, pp. 332–348. Springer, 2022.

Xuyang Li, Hamed Bolandi, Mahdi Masmoudi, Talal Salem, Ankush Jha, Nizar Lajnef, and Vishnu Naresh Boddeti. Mechanics-informed autoencoder enables automated detection and localization of unforeseen structural damage. *Nature Communications*, 15(1):9229, 2024a.

Xuyang Li, Mahdi Masmoudi, Nizar Lajnef, and Vishnu Boddeti. Estimating field parameters from multiphysics governing equations with scarce data. In *ICLR 2024 Workshop on AI4DifferentialEquations In Science*, 2024b.

Xuyang Li, John Harlim, Dibyajyoti Chakraborty, and Romit Maulik. A weak penalty neural ode for learning chaotic dynamics from noisy time series. *arXiv preprint arXiv:2511.06609*, 2025.

Chuang Liu and HengAn Wu. cv-pinn: Efficient learning of variational physics-informed neural network with domain decomposition. *Extreme Mechanics Letters*, 63:102051, 2023.

Mahdi Masmoudi, Xuyang Li, Nizar Lajnef, and Vishnu Boddeti. Parafind: Parameter field inference on non-uniform domains using neural network. In *NeurIPS 2024 Workshop on Data-driven and Differentiable Simulations, Surrogates, and Solvers*.

Daniel A Messenger and David M Bortz. Weak sindy for partial differential equations. *Journal of Computational Physics*, 443:110525, 2021a.

Daniel A Messenger and David M Bortz. Weak sindy: Galerkin-based data-driven model selection. *Multiscale Modeling & Simulation*, 19(3):1474–1497, 2021b.

John E Mottershead, Michael Link, and Michael I Friswell. The sensitivity method in finite element model updating: A tutorial. *Mechanical systems and signal processing*, 25(7):2275–2296, 2011.

Christopher Rackauckas, Yingbo Ma, Julius Martensen, Collin Warner, Kirill Zubov, Rohit Supekar, Dominic Skinner, Ali Ramadhan, and Alan Edelman. Universal differential equations for scientific machine learning. *arXiv preprint arXiv:2001.04385*, 2020.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

Nic Rummel, Daniel A Messenger, Stephen Becker, Vanja Dukic, and David M Bortz. Wendy for nonlinear-in-parameters odes. *arXiv preprint arXiv:2502.08881*, 2025.

GF Sirca Jr and H Adeli. System identification in structural engineering. *Scientia Iranica*, 19(6): 1355–1364, 2012.

A Appendix

A.1 Neural Network Architecture Details

To accurately capture the spatially varying structural parameters, we employ two independent neural networks, $\mathcal{N}_P(x;\theta_P)$ and $\mathcal{N}_C(x;\theta_C)$, to approximate the dimensionless stiffness $P(x)$ and damping coefficient $C(x)$, respectively. Distinct architectures are adopted to address the specific characteristics of each parameter field.

The stiffness network $\mathcal{N}_P$ is designed as a fully connected residual network with a hidden dimension of $h_P = 32$. The 1D spatial coordinate x is first projected into a high-dimensional feature space via a linear layer. The core feature extraction is performed by two stacked Residual Blocks to mitigate the vanishing gradient problem. Let $\mathbf{z}_k$ be the input to the k-th residual block. The output $\mathbf{z}_{k+1}$ is computed as $\mathbf{z}_{k+1} = \mathbf{z}_k + \mathcal{F}(\mathbf{z}_k)$, where the residual mapping $\mathcal{F}(\cdot)$ consists of a sequence of operations: ReLU $\to$ Linear(h_P, h_P) $\to$ ReLU $\to$ Linear(h_P, h_P). Following the residual blocks, a ReLU activation and a linear layer project the features to a scalar value. Crucially, a final Softplus activation is applied to ensure the stiffness remains strictly positive, and a small constant $\epsilon = 10^{-6}$ is added for numerical stability.

In contrast, the damping network $\mathcal{N}_C$ utilizes a compact multilayer perceptron architecture with a reduced hidden dimension of $h_C = 16$. This sub-network consists of an input linear layer, a ReLU activation, and an output linear layer followed by a Softplus activation. To align with the expected physical magnitude of the damping coefficient, the output of $\mathcal{N}_C$ is scaled by a factor of 20.0 after the addition of the stabilization term ϵ.

A critical challenge in structural identification is ensuring that the identified parameters adhere to physical laws. Specifically, stiffness and damping must be strictly positive to prevent non-physical behavior or numerical divergence. To address this, we apply a Softplus activation function, $\sigma(x) = \log(1 + e^x)$, at the final output layer of both networks. Unlike ReLU, Softplus is smooth and strictly

positive, preventing the gradient death problem associated with zero outputs. The final physical predictions are formulated as:

$$\begin{aligned} P_{pred}(x) &= \text{Softplus}(\mathcal{N}_P^{raw}(x)) + \epsilon \\ C_{pred}(x) &= \lambda_C \cdot (\text{Softplus}(\mathcal{N}_C^{raw}(x)) + \epsilon) \end{aligned} \tag{4}$$

where $\mathcal{N}^{raw}(\cdot)$ denotes the unconstrained scalar output of the final linear layer prior to activation. $\epsilon = 10^{-6}$ is a small constant for numerical stability, and $\lambda_C = 20.0$ is a scaling factor for the damping term to balance the magnitude differences between stiffness and damping. Furthermore, instead of standard random initialization which typically yields values near zero, we explicitly initialize the bias of the final linear layer such that the network output starts from a physically plausible mean value. For the stiffness network $\mathcal{N}_P$, the final bias is set to 2.0, ensuring an initial guess of approximately 2.12. Similarly, for the damping network $\mathcal{N}_C$, the final bias is set to 0.55, ensuring an initial pre-scaled output of approximately 1.0, resulting in a physical damping value near 20.0. The weights of the final layer are initialized with a small normal distribution ($\sigma = 0.01$) to ensure the initial prediction is spatially uniform before training begins.

A.2 Numerical Model Setup, Data Generation, and Noise Generation

We consider a clamped-clamped beam of length $L = 0.4$ m. The boundary conditions are defined as $u(0,t) = u(L,t) = 0$ and $\frac{\partial u}{\partial x}(0,t) = \frac{\partial u}{\partial x}(L,t) = 0$. The initial conditions are assumed to be zero, i.e., $u(x,0) = 0$ and $\frac{\partial u}{\partial t}(x,0) = 0$. The physical parameters of the beam are chosen as: thickness $h_0 = 5 \times 10^{-3}$ m, width $w = 5 \times 10^{-2}$ m, Young's modulus $E_0 = 70 \times 10^9$ Pa, and density $\rho_0 = 2700$ kg/m^3. To ensure numerical stability during optimization, we scale the system parameters to a normalized range using E_0. While this scaling facilitates the estimation process, accurate estimation at arbitrary physical scales in real-world scenarios can be achieved by adopting a two-stage parameter estimation strategy (Li et al., 2024b).

The training data are generated using the Method of Lines (MOL) to solve the governing PDE with high spatiotemporal resolution. We employ a high-fidelity reference grid of $N_{ref} = 81$. This resolution is strategically chosen such that the nodes of the coarser evaluation grids ($N_x \in \{11, 17, 21\}$) are spatially coincident with the reference grid. This allows for direct subsampling of the high-resolution data, effectively eliminating potential artifacts arising from spatial interpolation. The simulation spans a total time of 0.45 s with 160 timesteps. Further details of the parameter setup and known PDE information can be referred to Li et al. (2022).

We simulate sensor noise by adding Gaussian noise scaled by the signal's Root Mean Square (RMS):

$$\mathbf{u}_{obs} = \mathbf{u}_{ref} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2) \tag{5}$$

where the standard deviation σ is determined by the noise ratio η (e.g., $\eta = 0.01$ for 1% noise):

$$\sigma = \eta \cdot \text{RMS}(\mathbf{u}_{ref}) = \eta \sqrt{\frac{1}{N} \sum_i (\mathbf{u}_{ref}^{(i)})^2} \tag{6}$$

A.3 Hyperparameter Settings

For the pre-computation of spatial derivatives $\mathbf{u}_{xx}$, we employ a fourth-order accurate 5-point central difference stencil to minimize discretization errors on the coarse grid.

The hyperparameters for the weak formulation are summarized in Table 1. We define the test functions as $\psi(s) = (1 - s^2)^p$. Notably, we employ a significantly large spatial window ($w_x = 13$), covering approximately 76% of the beam length. This semi-global integration range is deliberately chosen to maximize the noise-filtering capability of the weak form. Additionally, for test functions, a higher polynomial degree $p_t = 8$ is selected for the temporal domain (vs. $p_x = 6$) to ensure sufficient smoothness and strictly vanishing derivatives at window boundaries.

Table 1: Key Hyperparameters for Neural-VSI.

Parameter	**Value**	**Parameter**	**Value**
Physical Length L	0.4	Spatial Window w_x	13 (grid points)
Training Grid N_x	17	Temporal Window w_t	21 (time steps)
Generation Grid N_x^{gen}	81	Spatial Poly. Degree p_x	6
Finite Diff. Stencil	5-point (4th order)	Temporal Poly. Degree p_t	8