engineering proceedings

Proceedings

Attention Mechanism-Driven Sensor Placement Strategy for Structural Health Monitoring †

Joo Wang Kim, Matteo Torzoni *, Alberto Corigliano and Stefano Mariani

Dipartimento di Ingegneria Civile e Ambientale, Politecnico di Milano, Piazza L. da Vinci 32, 20133 Milano, Italy; E-mail1 (J.W.K.); E-mail2 (A.C.); E-mail3 (S.M.)

* Correspondence: matteo.torzoni@polimi.it

† Presented at the 9th International Electronic Conference on Sensors and Applications, 1–15 Nov 2022; Available online: https://ecsa-9.sciforum.net/.

Abstract: Automated vibration-based structural health monitoring (SHM) strategies have been recently proven promising in the presence of aging and material deterioration threatening the safety of civil structures. Within such a framework, ensuring high quality and informative data is a critical aspect, highly dependent on the deployment of the sensors in the network and on their capability to provide damage-sensitive features to be exploited. This paper presents a novel data-driven approach to the optimal sensor placement, devised to identify sensor locations that maximize the information effectiveness for SHM purposes. The optimization of the sensor network is addressed by means of a deep neural network (DNN) equipped with an attention mechanism, a state-of-the-art technique in natural language processing (NLP) useful to focus on a limited number of important components in the information stream. The trained attention mechanism eventually allows to quantify the relevance of each sensor in terms of the so-called attention scores, and therefore enables to identify the most useful input channels to solve the relevant downstream SHM task. With reference to the damage localization task, framed here as a classification problem handling a set of predefined damage scenarios, the DNN is trained to locate damage on labeled data, that have been formerly simulated to emulate the effects of damage under different operational conditions. The capabilities of the proposed method are demonstrated by referring to an eight-story shear building, characterized by damage states possibly located at any story and of unknown severity.

Keywords: attention mechanism; optimal sensor placement; sensor networks; structural health monitoring; deep learning; damage identification

Citation: Kim, J.W.; Torzoni, M.; Corigliano, A.; Mariani, S. Attention Mechanism-Driven Sensor Placement Strategy for Structural Health Monitoring. *Eng. Proc.* **2022**, *4*, 0. https://doi.org/

Academic Editor: Francisco Falcone

Published: 1 November 2022

1. Introduction

Civil structures such as buildings, highways, tunnels and bridges are a backbone of our modern society [1]. Aging and ever-increasing extreme loading conditions threaten such systems, stressing the need of SHM strategies to detect and identify any deviation from the damage-free state, ultimately allowing to reduce maintenance costs and avoid potential tragic events.

Traditionally, the condition assessment of civil structures has been carried out through nondestructive testing and visual inspection, which can provide only local health assessment and highly depend on the personal expertise. Nowadays, in-service remote vibration-based SHM is a standard and widespread approach for continuous and automated global health monitoring, see e.g., [2,3], allowing to assess damage from the vibration response in terms of, e.g., acceleration or displacement multivariate time series acquired with pervasive sensing systems [4,5]. As these SHM techniques rely on their capability to extract damage-sensitive features from the raw sensor recordings, ensuring a satisfactory quality and informativeness of recorded data is a critical aspect. Beside the limited amount of available sensors due to installation costs, the optimization of the sensors deployment in the network is a key aspect in order to maximize the information effectiveness for SHM purposes.

The optimal sensor placement (OSP) problem has been systematically addressed in the literature; for an overview, interested readers can refer to e.g., [6]. Notable contributions in this field have been obtained by means of the Fisher information matrix and its related metrics [7,8], information entropy [9,10] and value of information [11,12].

This work proposes a novel approach to the OSP, leveraging on data-driven methods empowered by deep learning (DL) algorithms. Its key component is the use of an attention mechanism [13,14] in a neural network, trained in a supervised fashion to solve an SHM task by exploiting the structural response data from a set of feasible sensors locations. Beside allowing to address the considered SHM task, the trained DNN also enables to identify the most useful input channels by assigning an attention score to each sensor.

The use of DL in SHM has been shown very effective to automatize the feature engineering stage, required to improve the effectiveness of a damage detection strategy. Indeed, DL allows to automatize the selection and extraction of optimized damage-sensitive features through an end-to-end learning processes, to ultimately relate them with the corresponding structural states. Nevertheless, supervised techniques require labeled data referred to the possible damage states of the structure, which can not be obtained for real civil structures. To cope with this, we resort to a simulation-based approach, see e.g., [15,16], by adopting a physics-based model of the structure to be monitored, allowing to systematically simulate the effect of damage on the structural response under different operational conditions.

The proposed methodology is investigated through the virtual monitoring of an eight-story shear building, with reference to the damage localization task. The latter is framed as a classification problem involving a set of predefined damage states, possibly located at any story. The obtained results confirm the capabilities of the proposed approach, both in terms of damage localization and of optimal sensor placement.

2. SHM Methodology

The proposed methodology is detailed as follows: the composition of the training set is specified in Section 2.1; the working principle of the attention mechanism is described in Section 2.2; the setup of the proposed OSP approach is finally explained in Section 2.3.

2.1. Datasets Definition

Considering an observation time window $(0, T)$, short enough to assume invariant operational and damage conditions, the training set $\mathbf{D}$ is assembled by collecting vibration data from a virtual sensing network deployed to feature N_u feasible sensor locations, and a sampling period Δt. The training set $\mathbf{D}$ is built from the assembly of I instances as follows:

$$\mathbf{D} = \{(\mathbf{U}_i, y_i)\}_{i=1}^{I} \, , \tag{1}$$

with each instance consisting of vibration time-histories $\mathbf{U}_i = \mathbf{U}(\mathbf{x}_i, y_i, \delta_i) = [\mathbf{u}_1, \ldots, \mathbf{u}_{N_u}]_i \in \mathbb{R}^{N_u \times L}$ shaped as N_u arrays of $L = 1 + T/\Delta t$ measurements. This is here obtained from a numerical model of the structure to be monitored for the corresponding $\mathtt{N}_{\mathtt{par}}$ input parameters $\mathbf{x}_i \in \mathbb{R}^{\mathtt{N}_{\mathtt{par}}}$ defining the operational conditions (for instance the loadings acting on the structure), and for the relevant damage state characterized by y_i and δ_i according to: $y_i \in \{0, \ldots, Y\}$ labels the specific damage scenario undergone by the structure while collecting the i-th instance, among a set of predefined Y damage states, each referring to a different damage location, and with $y_i = 0$ identifying the damage-free baseline. Damage is modeled as a selective reduction of the material stiffness of amplitude δ_i, taking place within the pre-designated region associated to y_i. In this work, $\mathbf{x}_i$ and δ_i are not considered part of the label, as only the damage localization task is addressed. To populate $\mathbf{D}$, the parametric input space is assumed to display a uniform probability distribution for each parameter, and it is sampled via the latin hypercube rule. Unless if necessary, index i will be dropped in the reminder of the paper for ease of notation.

2.2. Attention Mechanism for Data Analytics in SHM

In the neural networks community, *attention* is a mechanism to mimic the cognitive attention behavior useful to adaptively focus on few, but important, components of the data stream. This is achieved by means of learnable weights, optimized through gradient descent algorithms, which can change at runtime as a function of the input data. Originally proposed for neural machine translation problems [13], attention is nowadays a state-of-the-art technique in NLP. The main reason behind its popularity is that it allows to code the data stream into a series of embeddings and learn how to adaptively choose a subset of them, thus preventing early information from getting lost, as is often the case when processing long sequences with sequence-to-sequence recurrent encoder-decoders.

The corresponding working principle can be described as mapping a query and a set of key-value pairs to an output, computed as a weighted sum of the values, and weights assigned by a compatibility function of the query with the corresponding key. Queries, keys and values can be obtained in several ways, and most often are the output of previous layers in the neural network. In this work, the scaled dot-product attention introduced in [14] is employed as an effective and efficient form of self-attention. The input consists of a set of m_Q queries $\mathbf{Q} \in \mathbb{R}^{m_Q \times d_Q}$ of length d_Q, of a set of m_K keys $\mathbf{K} \in \mathbb{R}^{m_K \times d_K}$ of length d_K, and of a set of m_K values $\mathbf{V} \in \mathbb{R}^{m_K \times d_V}$ of length d_V. The output of the scaled dot-product attention is computed as:

$$\mathbf{A}(\mathbf{QW}_Q, \mathbf{KW}_K, \mathbf{VW}_V) = \texttt{Softmax}\left(\frac{\overline{\mathbf{Q}}\overline{\mathbf{K}}^\top}{\sqrt{s_K}}\right)\overline{\mathbf{V}}\,, \quad \mathbf{A}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) \in \mathbb{R}^{m_Q \times s_V}\,, \tag{2}$$

where: $\mathbb{R}^{m_Q \times s_K} \ni \overline{\mathbf{Q}} = \mathbf{QW}_Q$, $\mathbb{R}^{m_K \times s_K} \ni \overline{\mathbf{K}} = \mathbf{KW}_K$, and $\mathbb{R}^{m_K \times s_V} \ni \overline{\mathbf{V}} = \mathbf{VW}_V$ are the projection of queries, keys and values, onto different subspaces, respectively spanned by the learnable matrices $\mathbf{W}_Q \in \mathbb{R}^{d_Q \times s_K}$, $\mathbf{W}_K \in \mathbb{R}^{d_K \times s_K}$ and $\mathbf{W}_V \in \mathbb{R}^{d_V \times s_V}$; the scaled dot-product in brackets is the previously mentioned compatibility function, measuring the alignment of each query with each key; the `Softmax` function serves to obtain a set of weights on the values, which are the so-called attention score, summing to one for each query.

There are only a few contributions in the SHM literature exploiting attention techniques, see e.g., [17–19]. However, to the best of our knowledge, this is the first application explicitly using the attention scores to address the OSP problem. In particular, each attention score is exploited to assess the informativeness of the corresponding sensor for the downstream damage location task. That is, attention is applied across a fictitious sensor dimension, comprised by the set of feasible sensor locations, deprived of any geometrical notion of spatial location.

2.3. Attention Mechanism-Driven Sensor Placement

The DNN adopted to address the OSP problem for damage localization purposes is reported in Figure 1. The architecture is made of two main branches, namely the query branch and the key/value branch. The former takes vibration recordings $\mathbf{U}$ from all the available channels and run them through three one-dimensional (1D) convolutional units, each comprised by a `ReLU`-activated 1D convolutional layer and a max pooling layer. The resulting output is then passed through a fully-connected layer to obtain a query $\mathbf{Q} \in \mathbb{R}^{1 \times d_Q}$, representing the current structural response. The key/value branch instead consists of a stack of N_u sub-neural networks operating in parallel, each receiving input data $\mathbf{u}_j$ from the corresponding j-th sensor, with $j = 1, \dots, N_u$, but all sharing the same set of tunable parameters. The base sub-neural network features three `ReLU`-activated 1D convolutional layers and a normalization layer. The output of each subnetwork is a key $\mathbf{k}_j \in \mathbb{R}^{d_K}$, with $m_K = N_u$, which coincides with the associated value $\mathbf{v}_j \in \mathbb{R}^{d_V = d_K}$, and accounts for sensor-specific damage-sensitive features extracted by acting separately on each input channel.

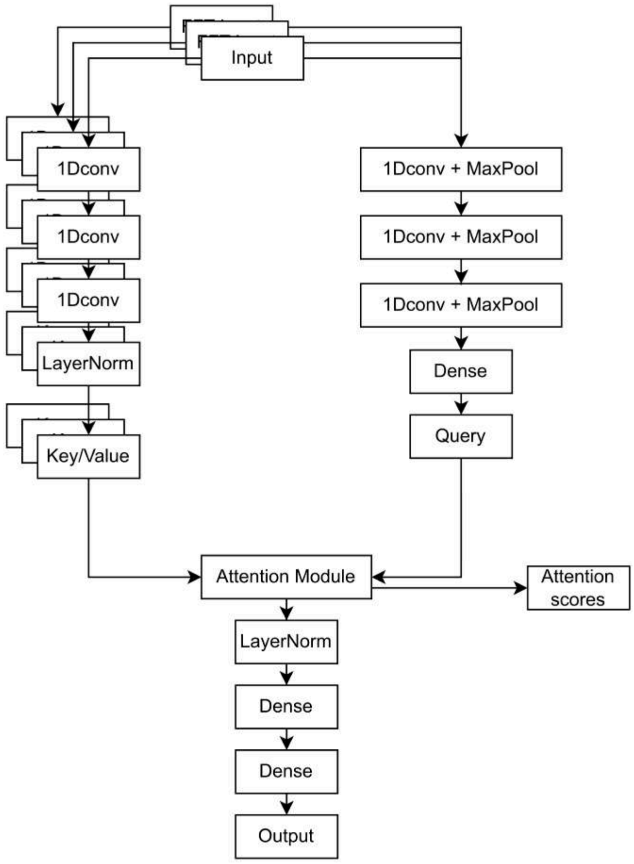

Figure 1. Scheme of the DNN architecture adopted to address the OSP problem.

The query $\mathbf{Q}$ and keys $\mathbf{K}$ then feed the scaled dot-product attention, to compute the attention scores and the attention module output $\mathbf{V} \in \mathbb{R}^{1 \times s_V}$, according to Equation (3). Each query can be interpreted as: "where should I look for to most sensitive answer to the damage localization problem, given the current structural behavior?". Similarly, we can look at the keys as: "the possible answers to the query, i.e., sensor locations, with good answers aligned to the question, i.e., high attention score, and bad answers orthogonal to it, i.e., low attention score".

The remainder of the DNN architecture simply addresses the downstream damage localization task, and consists of a normalization layer and of two fully-connected layers; the first one is `ReLU`-activated, while the second one is activated by a `Softmax` function, which is the standard choice for classification problems.

During training, the set of tunable weights parametrizing the DNN is optimized by minimizing the categorical cross-entropy between the predicted and target label classes, using the Adam algorithm, which is a first-order stochastic gradient descent optimizer. Once the DNN is trained, the OSP is addressed by processing a testing set not seen during training, and by extracting the corresponding attention scores from the attention module. These attention scores can then be used in several ways to rank the sensors according to their relevance to locate the damage. In the present work, we simply compute the mean attention score for each channel under different operational conditions and look for the channels featuring the highest values.

3. Results: Eight-Story Shear Building

The proposed approach is assessed on the eight-story shear building model depicted in Figure 2a, adapted from [20]. Each story features a mass $m = 625$ t with an inter-story stiffness $k_{sh} = 106$ kN/m. Structural damping is introduced through a Rayleigh model, accounting for a 1% damping ratio on each vibration mode. By neglecting the axial deformability of the elements, only the horizontal degrees of freedom are considered. The structure is excited by harmonic loads, acting on each floor with the same frequency and phase according to:

$$p_j(t) = \frac{j}{8} P_0 \sin(2\pi f t), \quad j = 1, \ldots, 8, \tag{3}$$

where: $P_0 \in [2, 3]$ kN is the load amplitude; $f \in [0, 13]$ Hz is the load frequency, sampled in a range including all the natural frequencies of the structure; the factor $\frac{j}{8}$ shapes a triangular load distribution along the building elevation, with j growing from the bottom. Therefore, the parametrization ruling the operational conditions is based on $\mathbf{x} = \{P_0, f\}^\top$.

The possible damage states are defined by a $\delta = 25\%$ reduction of the corresponding inter-story stiffness, with associated labels $y = 1, \ldots, 8$ from the ground inter-story to the roof one, and with $y = 0$ labeling the undamaged case.

Displacement time histories $\mathbf{U}(\mathbf{x}, y, \delta) = [\mathbf{u}_1, \ldots, \mathbf{u}_8]$ are recorded from a virtual sensing system made of $N_u = 8$ sensors, placed at each floor. The recordings are provided for a time interval characterized by $T = 5$ s and with a sampling period of $\Delta t = 0.01$ s, thus consisting of $L = 501$ measurements each.

The dataset $\mathbf{D}$ is assembled from $I = 9999$ instances generated for different values of the parameters selected via the latin hypercube sampling rule. Before training, data are polluted by adding an independent, identically distributed Gaussian noise, featuring a signal-to-noise ratio equal to 100. Moreover, the data are preprocessed via discrete Fourier transform and subsequently standardized, to improve the damage localization performance of the DNN.

In terms of damage localization capabilities, the classifier provides a satisfactory 88.6% classification accuracy against a testing set, without showing any particular misclassification trend. The obtained results are summarized by the confusion matrix in Figure 2b, characterized by high values along the main diagonal.

Given the good damage localization capability of the DNN, the corresponding attention scores can be considered optimized. The average attention score for each input channel is reported in Figure 2c, showing a clear trend with increasing values from the ground floor to the top one. This is reasonable, as the response of upper floors is expected to be more sensitive to a damage on a floor below it, than to a damage on a floor above it.

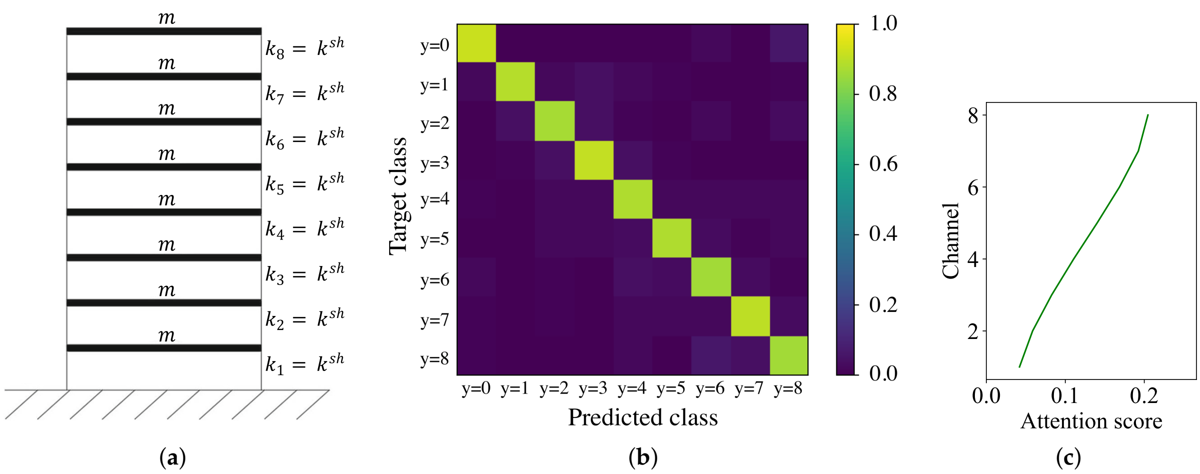

Figure 2. Eight-story shear building case study: (**a**) physics-based digital twin; (**b**) confusion matrix relevant to the classifier testing; (**c**) average attention scores for each monitored channel.

4. Conclusions

This paper has presented an approach to the optimal sensor placement for structural health monitoring purposes. By relying on deep neural networks, the strength of the procedure stems from the interpretability of the attention scores associated to a set of feasible sensor locations. The method rests on a numerical model of the structure, useful to obtain labeled data pertaining to specific damage conditions. With reference to a damage localization case study, the obtained results show the capability of the attention mechanism to identify the most informative input channels to locate the damage.

The future studies will investigate the proposed method while exploiting multiple attention heads, as dealing with features from different representation subspaces is expected to improve the overall performance. Moreover, the effect of a strong L^1 regularization will be also analyzed, with the aim of inducing sparsity in the attention score vector.

Author Contributions: xxx.

Funding: xxx.

Institutional Review Board Statement: xxx.

Informed Consent Statement: xxx.

Data Availability Statement: xxx.

Acknowledgments: The authors would like to thank Luca Rosafalco for having provided the numerical model of the shear frame. M.T. acknowledges the financial support by Politecnico di Milano through the interdisciplinary Ph.D. Grant "Physics-Informed Deep Learning for Structural Health Monitoring".

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Torzoni, M.; Manzoni, A.; Mariani, S. Structural health monitoring of civil structures: A diagnostic framework powered by deep metric learning. *Comput. Struct.* **2022**, *271*, 106858. https://doi.org/10.1016/j.compstruc.2022.106858.
2. Rosafalco, L.; Torzoni, M.; Manzoni, A.; Mariani, S.; Corigliano, A. Online structural health monitoring by model order reduction and deep learning algorithms. *Comput. Struct.* **2021**, *255*, 106604. https://doi.org/10.1016/j.compstruc.2021.106604.
3. Torzoni, M.; Manzoni, A.; Mariani, S. A Deep Neural Network, Multi-fidelity Surrogate Model Approach for Bayesian Model Updating in SHM. In *European Workshop on Structural Health Monitoring*; Springer International Publishing: Berlin/Heidelberg, Germany, 2023; pp. 1076–1086. https://doi.org/10.1007/978-3-031-07258-1_108.
4. Rosafalco, L.; Torzoni, M.; Manzoni, A.; Mariani, S.; Corigliano, A. A Self-adaptive Hybrid Model/data-Driven Approach to SHM Based on Model Order Reduction and Deep Learning. In *Structural Health Monitoring Based on Data Science Techniques*; Springer International Publishing: Berlin/Heidelberg, Germany, 2022; pp. 165–184. https://doi.org/10.1007/978-3-030-81716_9.
5. García-Macías, E.; Ubertini, F. Integrated SHM Systems: Damage Detection Through Unsupervised Learning and Data Fusion. In *Structural Health Monitoring Based on Data Science Techniques*; Springer International Publishing: Berlin/Heidelberg, Germany, 2022; pp. 247–268. https://doi.org/10.1007/978-3-030-81716_12.
6. Ostachowicz, W.; Soman, R.; Malinowski, P. Optimization of sensor placement for structural health monitoring: A review. *Struct. Health Monit.* **2019**, *18*, 963–988. https://doi.org/10.1177/1475921719825601.
7. Shi, Z.Y.; Law, S.S.; Zhang, L.M. Optimum Sensor Placement for StructuralDamage Detection. *J. Eng. Mech.* **2000**, *126*, 1173–1179. https://doi.org/10.1061/(ASCE)0733-9399(2000)126:11(1173).
8. Penny, J.E.T.; Friswell, M.I.; Garvey, S.D. Automatic choice of measurement locations for dynamic testing. *AIAA J.* **1994**, *32*, 407–414. https://doi.org/10.2514/3.11998.
9. Capellari, G.; Chatzi, E.; Mariani, S. Structural Health Monitoring Sensor Network Optimization through Bayesian Experimental Design. *ASCE-ASME J. Risk Uncertain. Eng. Syst.* **2018**, *4*, 04018016. https://doi.org/10.1061/ajrua6.0000966.
10. Capellari, G.; Chatzi, E.; Mariani, S. Cost-benefit optimization of structural health monitoring sensor networks. *Sensors* **2018**, *18*, 2174. https://doi.org/10.3390/s18072174.
11. Malings, C.; Pozzi, M. Value-of-information in spatio-temporal systems: Sensor placement and scheduling. *Reliab. Eng. Syst.* **2018**, *172*, 45–57. https://doi.org/10.1016/j.ress.2017.11.019.
12. Kamariotis, A.; Chatzi, E.; Straub, D. Value of information from vibration-based structural health monitoring extracted via Bayesian model updating. *Mech. Syst. Signal Process.* **2022**, *166*, 108465. https://doi.org/10.1016/j.ymssp.2021.108465.
13. Bahdanau, D.; Kyung Hyun, C.; Bengio, Y. Neural machine translation by jointly learning to align and translate. In Proceedings of theInternational Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015; Volume 3. Available online: https://arxiv.org/pdf/1409.0473.pdf (accessed on).
14. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All you Need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30. Available online: https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf (accessed on).
15. Taddei, T.; Penn, J.D.; Yano, M.; Patera, A.T. Simulation-based classification; a model-order-reduction approach for structural health monitoring. *Arch. Comput. Methods Eng.* **2018**, *25*, 23–45. https://doi.org/10.1007/s11831-016-9185-0.

16. Torzoni, M.; Rosafalco, L.; Manzoni, A.; Mariani, S.; Corigliano, A. SHM under varying environmental conditions: An approach based on model order reduction and deep learning. *Comput. Struct.* **2022**, *266*, 106790. https://doi.org/10.1016/j.compstruc.2022.106790.
17. Lei, X.; Xia, Y.; Wang, A.; Jian, X.; Zhong, H.; Sun, L. Mutual information based anomaly detection of monitoring data with attention mechanism and residual learning. *Mech. Syst. Signal Process.* **2023**, *182*, 109607. https://doi.org/10.1016/j.ymssp.2022.109607.
18. Li, G.; Ma, B.; He, S.; Ren, X.; Liu, Q. Automatic Tunnel Crack Detection Based on U-Net and a Convolutional Neural Network with Alternately Updated Clique. *Sensors* **2020**, *20*, 717. https://doi.org/10.3390/s20030717.
19. Pan, Y.; Ventura, C.E.; Li, T. Sensor placement and seismic response reconstruction for structural health monitoring using a deep neural network. *Bull. Earthq. Eng.* **2022**, *20*, 4513–4532. https://doi.org/10.1007/s10518-021-01266-y.
20. Rosafalco, L.; Manzoni, A.; Mariani, S.; Corigliano, A. Fully convolutional networks for structural health monitoring through multivariate time series classification. *Adv. Model. Simul. Eng. Sci.* **2020**, *7*, 38. https://doi.org/10.1186/s40323-020-00174-1.