Slab track condition monitoring based on learned sparse features from acoustic and acceleration signals

Baorui Dai¹, ², Gaëtan Frusque², Qi Li¹, and Olga Fink²

¹Department of Bridge Engineering, Tongji University, Shanghai, 200092, China
dbr@tongji.edu.cn
liqi.bridge@tongji.edu.cn

²Laboratory of Intelligent Maintenance and Operations Systems, EPFL, Lausanne, 1015, Switzerland
gaetan.frusque@epfl.ch
olga.fink@epfl.ch

ABSTRACT

The implementation of concrete slab track solutions has been recently increasing particularly for high-speed lines. While it is typically associated with low periodic maintenance, there is a significant need to detect the state of slab tracks in an efficient way. Data-driven detection methods are promising. However, collecting large amounts of labeled data is particularly challenging since abnormal states are rare for such safety-critical infrastructure. To imitate different healthy and unhealthy states of slab tracks, this study uses three types of slab track supporting conditions in a railway test line. Acceleration sensors (contact) and acoustic sensors (contactless), are installed next to the three types of slab track to collect the acceleration and acoustic signals as a train passes by with different speeds. We use a deep learning framework based on the recently proposed Denoising Sparse Wavelet Network (DeSpaWN) to automatically learn meaningful and sparse representations of raw high-frequency signals. A comparative study is conducted among the feature learning/extraction methods, and between acceleration signals and acoustic signals, by evaluating the detection effectiveness using a multi-class support vector machine. It is found that the classification accuracy using acceleration signals can reach almost 100%, irrespective which feature learning/extraction method is adopted. Due to the more severe noise interference in acoustic signals, the performance of using acoustic signals is worse than of using acceleration signals. However, it can be significantly improved by learning meaningful features with DeSpaWN.

1. INTRODUCTION

The implementation of concrete slab track solutions has been recently increasing particularly for high-speed lines because of their advantages in operation and maintenance [1]. Influenced by the environmental and operating conditions such as dynamic train load and temperature fluctuations, the mortar layer, which is serving as connection between slab track and foundation, is prone to degradation. The degradation of mortar layer will weaken the supporting condition of slab track, thus, affecting the comfort of passengers and even the safety of trains [2, 3].

Acceleration sensors have been one of the preferred choices in structural health monitoring systems to monitor the condition of railway infrastructure [4, 5]. The technology has the advantage of directly reflecting the vibration characteristics of the measured objects. However recently, some concerns have been emerging about the application of accelerometers due to the potential risks to the safety of railway operations stemming from the close proximity to railway clearance placement of sensors [6].

Recently, monitoring the condition of industrial and infrastructure assets with acoustic signals has been gaining importance since acoustic sensors are non-intrusive and easy to install or retrofit [7-9], which is crucial for safety-critical infrastructure like railways. Although promising results have been obtained on the general industrial assets monitoring [10, 11], there have been few studies addressing the application of acoustic monitoring in railways. For example, Pieringer et al. [12] adopted a noise measurement car to collect wheel-rail noise and trained a logistic regression classifier to identify squats in the German railway network. In another research study, Wang et al. [13] used a smartphone to collect acceleration and sound signals inside subway trains, and applied the extreme gradient boosting algorithm to identify squeal and rumble as a reflection of rail defection. Meng et al. [14] used an optical fiber sensing system to pick up the acoustic signals along the fiber line and achieved a high-precision detection of railway perimeter intrusions. In general, research and application of acoustic condition monitoring for
railways has been recently evolving. However, the condition monitoring of slab tracks based on acoustic signals has remained unaddressed.

In this research, we aim to perform a proof of concept and to evaluate the detectability and distinguishability of the different potential degradation states of the mortar supporting layer. Moreover, we aim to compare the results obtained with acoustic monitoring signals and with accelerometers.

One of the typical deterioration states is caused by degraded mortar layer, which changes the supporting condition of the slab track. In other words, the connection between track slab and foundation becomes weak. The different degrees of mortar layer degradation will lead to different decision-making of maintenance, thus the classification of different slab track states is beneficial. Since abnormal states of slab tracks in railway lines are rare in real applications, we aim to collect data from a railway test line by imitating the degraded condition of the mortar layer by using different types of support layers. More concretely, this study uses three types of slab track with different supporting conditions, mortar, rubber and spring support, in a railway test line, corresponding to health, intermediate degradation, and severe degradation conditions, respectively. Acceleration sensors (contact) and acoustic sensors (contactless), are installed next to the three types of slab track to collect the acceleration and acoustic signals as a train passes by with different speeds.

Due to the random dynamic excitation and environmental factors, the acceleration and acoustic signals measured from the field monitoring are non-stationary [7]. Time-frequency spectrograms and wavelet coefficients are considered to be effective in capturing meaningful features of non-stationary signals [15, 16]. However, careful feature extraction with the right hyperparameter selection is required to extract a compact representation of the raw high-frequency signals. Moreover, the effectiveness of the extracted features for condition identification might be affected by unexpected noise or changing operating conditions. Since the train speed may be unstable during the train operation, decomposing the collected signals with fixed time resolution is hard to adapt to the speed change. To address these limitations, we use a deep learning framework based on the recently proposed DeSpaWN to automatically learn meaningful and sparse representations of raw high-frequency signals. A comparative study is conducted among the feature learning and the different feature extraction methods, and also between the acceleration signals and acoustic signals, by evaluating the detection effectiveness using a multi-class support vector machine (SVM).

2. Methods

A flow chart of the content of this study is illustrated in Fig. 1.

2.1. Data collection

We use three types of slab track with different supporting conditions in a railway test line to imitate different healthy and degraded states of slab tracks with different levels of degradation, as shown in Fig. 2. The connections between the three track slabs and foundations are mortar (concrete), rubber, and discrete spring support, respectively. We consider mortar layer as the healthy condition of the support layer, rubber layer as intermediate degradation level of the support layer and the spring support as degraded support layer. This results, therefore, in three classes of health conditions: one healthy class and two classes of degraded conditions with two different severity levels.

The train operated on the railway test line is a metro train, with six vehicles and a total length of 140 m. Acceleration sensors and acoustic sensors are installed next to the three types of slab track to collect the acceleration and acoustic signals as the train passes by with different speeds (20, 40, 60, and 80 km/h), as shown in Fig. 3. The sampling frequencies of the acceleration sensors and acoustic sensors are 20 kHz. The collected dataset of each track type consists of 24 samples (the train passes six times under each speed) for both acceleration and acoustic signals.
To facilitate the subsequent feature extraction, the effective duration of each signal, during the time that the slab track is between the first and last wheel sets of the moving train, is intercepted. The average durations of intercepted signals under different train speeds are presented in Table 1. Slab Track 1, 2, and 3 represent the healthy condition, intermediate degradation level, and severe degradation level of support layers, respectively.

| Speed (km/h) | Number of train passes | Average duration of intercepted signals (s) |
|--------------|------------------------|--------------------------------------------|
|              |                        | Slab Track 1 | Slab Track 2 | Slab Track 3 |
| 20           | 6                      | 21.84       | 22.34       | 22.04       |
| 40           | 6                      | 11.41       | 11.47       | 11.43       |
| 60           | 6                      | 7.94        | 7.85        | 7.90        |
| 80           | 6                      | 6.11        | 6.12        | 6.20        |

2.2. Denoising Sparse Wavelet Network

Recently, several architectures that combine the interpretability advantages of signal processing and the learning capabilities of neural networks have shown promising results [17-20]. Thus, we consider the DeSpaWN [8], a recently proposed deep learning framework inspired by fast discrete wavelet transform (FDWT). This method seems to be particularly adapted for our task as it decomposes the input signal in different and adapted time-frequency resolutions.

Fig. 4 illustrates the cascade algorithm related to the FDWT, which forms the basic architecture of DeSpaWN. A low-pass and a high-pass filter both followed by a sub-sampling step by a factor of two respectively decompose an input signal into detail and approximation coefficients. Moreover, the approximation coefficients of the previous layer are decomposed in a similar way. The detail coefficients have accurate time-frequency resolution varying according to the layer. By including the approximation coefficients of the last layer, they form the time-frequency representation of the input signal. From the obtained representation it is possible to perfectly reconstruct the input signal via inverse FDWT.

DeSpaWN has an encoder-decoder architecture based on the successive use of FDWT and inverse FDWT. It utilizes a fully learnable variation of the cascade algorithm (Fig. 4) by allowing learning the kernel common to both filters at each layer. Moreover, the resulting detail coefficients (plus the approximation coefficients of the last layer) are fed to a specifically designed learnable hard thresholding (HT) activation functions [8]. Independently for each layer, the learnable HT functions operate as an automatic wavelet denoising operation inducing sparsity in the final time-frequency representation. However, due to the HT activation function, it is not possible anymore to recover perfectly the input signal with the decoding part of DeSpaWN.

To achieve firstly a good signal reconstruction and secondly a sparse decomposition, we use the same loss function as proposed in [8], i.e. minimizing the reconstruction error plus a sparse regularization term applied to the time-frequency representation. We initialize the learnable filters with Daubechies-4 wavelets, and set the bias of the HT activation functions to 0.5. The regularization parameter is set to 1. The DeSpaWN is trained using the Adam optimiser with a learning rate of 0.0001.

2.3. Feature learning and extraction

The dataset of either acoustic signals or acceleration signals is divided into training data and test data with a 2:1 ratio. DeSpaWN, FDWT with ‘db4’ basis, Wavelet Packet Transform (WPT) with ‘db4’ basis, Short-time Fourier transform (STFT), and Mel spectrogram are used in this study to extract features. Each data sample is padded with zeros to form an identical length of 512,426 points, which equals to the maximum number of points of raw signals. Therefore, the number of layers that are decomposed by DeSpaWN and FDWT is set to log2512,426 ≈ 18. To maintain consistency
between the different methods, the number of filter bands of the Mel spectrogram are also set to 18. The STFT spectrogram is manually divided into 18 frequency bands with equal width. The window sizes (with overlap ratio of 1/2) for Mel spectrogram and STFT are determined as 1024 to get a frequency resolution of around 20 Hz. Since the number of decomposition layers of WPT can only be set to a power of two, it is determined as 16. Once the transformation is completed, the zeros at the end of each decomposition layer of signals are removed.

In the time dimension, each signal has the characteristic of six periodical fluctuations due to the successive excitation of six vehicles of the train. To enrich the training and test dataset, each spectrogram or wavelet coefficients statistics transformed by aforementioned methods is divided into six time bands with equal width, and the divided signals are regarded as independent sample units. To capture simultaneously the local and global characteristics of sample data, the maximum and average values of the time-frequency or wavelet coefficients (converted to decibels) from each filtered frequency band of divided signals are taken as the features. For DeSpaWN, there are two additional features, namely the maximum and average values of the residuals between the reconstructed signal and original signal.

2.4. Classification

The training and test data for classification are consistent with the training and test data for DeSpaWN. A multi-class SVM with Radial Basis Function kernel is used in this study to classify the slab track conditions. In the training stage, the 5-fold cross validation technique is utilized to obtain the optimal hyperparameters of the SVM. Then, the classification results can be calculated by applying the SVM with optimal hyperparameters to test data.

3. Results and Discussion

3.1. Considered classification tasks

Two classification tasks are implemented to evaluate the effectiveness of slab track condition monitoring based on acoustic and acceleration signals as well as the performance of different feature learning / extraction methods. For the first task, both training dataset and test dataset contain all train speeds. In order to evaluate the generalization ability and to study the ability of identifying slab track conditions under unknown train speed, for the second classification task the collected signals with three different train speeds are assigned to the training dataset, and the signals with the remaining speed form the test dataset.

3.2. Same train speeds in training and test dataset (Task 1)

The results of the first task evaluating the classification accuracy are listed in Table 2 and 3. It appears that the best performance of using acoustic signals to identify the slab track condition reaches 92.8%, while each classification accuracy based on acceleration signals is 100%. No matter which method is adopted, the acceleration signals always perform better than acoustic signals in this classification task. Because the acceleration sensors are in direct contact with slab tracks, they can measure the vibration of the track structure without suffering from the environmental interference. On the contrary, the acoustic sensors receive acoustic signals from all directions and are more prone to environmental impacts by noise. In addition to the sound radiated from slab track, the collected acoustic signals consist of various noises produced by wheel-rail contact, machine operation, vibration of irrelevant structures, etc. Therefore, the best classification results of using acoustic signals are slightly worse than of using acceleration signals.

For the classification based on acoustic signals, which have a higher noise level than acceleration signals, DeSpaWN reaches a performance improvement of 3.2% to 23.3% compared to other feature extraction methods. This confirms the observations in a previous publication that DeSpaWN has the advantage of learning a noise-independent representation of signals. WPT and STFT are the two methods that perform the worst in this condition monitoring task. This can be explained by the fact that the dominant frequency range of the slab track vibration is oftentimes below 2 kHz (Fig. 5), while the two methods give equal weights to low and high frequency features within the range of 0-10,000 Hz. On the contrary, the FDWT and the Mel spectrogram, which focus more on the characterization of the low frequency part of signals, achieve a better performance compared to WPT and STFT in this case.

Table 2. Comparative results based on acoustic signals in Classification Task 1

| Slab track condition          | Classification accuracy (%) |
|------------------------------|-----------------------------|
|                              | DeSpaWN | FDWT | WPT | STFT | Mel spectrogram |
| No degradation (Label 1)     | 90.2     | 89.6 | 62.5| 81.2 | 85.4           |
| Intermediate degradation (Label 2) | 92.3     | 79.2 | 68.8| 89.6 | 89.6           |
| Severe degradation (Label 3) | 95.8     | 93.8 | 77.1| 79.2 | 93.8           |
| Average                      | 92.8     | 87.5 | 69.5| 83.3 | 89.6           |
Table 3. Comparative results based on acceleration signals in Classification Task 1

| Slab track condition                  | Classification accuracy (%) |
|---------------------------------------|----------------------------|
|                                       | DeSpaWN | FDWT | WPT | STFT | Mel spectrogram |
| No degradation (Label 1)               | 100.0    | 100.0 | 100.0 | 100.0 | 100.0          |
| Intermediate degradation (Label 2)    | 100.0    | 100.0 | 100.0 | 100.0 | 100.0          |
| Severe degradation (Label 3)          | 100.0    | 100.0 | 100.0 | 100.0 | 100.0          |
| Average                               | 100.0    | 100.0 | 100.0 | 100.0 | 100.0          |

Fig. 5. STFT spectrograms of acceleration signals under the condition of 80 km/h.

3.3. Generalization ability: different train speeds in training and test dataset (Task 2)

In the actual track state monitoring task, the collected dataset may not contain all the train speed conditions. This puts forward the requirement to predict the slab track state under the condition of unknown train speed. To simulate this task, collected signals with train speeds of three different train speeds are used to train and the signals with the remaining speed are used to test the SVM classifier.

1) Training dataset: 20, 40, 60 km/h; Test dataset: 80 km/h

Table 4 and 5 display the classification results on the one hand between different feature leaning / extraction methods, and on the other hand between acoustic signals and acceleration signals. The performance of the feature learning / extraction when using acceleration signals (almost all methods reach 100%) still significantly exceeds the performance of the feature learning / extraction when applied to acoustic signals (the best result is 81.5%). As discussed before, the fact that the acoustic signals are subject to more interference and noise than acceleration signals is the main cause of the classification difference. This situation combined with the inconsistent training speeds in the training dataset and test dataset, leads to a significant decrease in the classification accuracy based on acoustic signals compared to task 1 where samples from all considered speeds were contained in the training and test datasets.

Table 4 shows that the wavelet-based methods (DeSpaWN, FDWT, and WPT) are generally superior to the other two methods (STFT and Mel spectrogram) in this task. A possible explanation is that the wavelet-based methods have more flexible time and frequency resolutions at different decomposition layers compared to the other two methods, while the speed may influence the activation time of the characteristic features indicating the slab track conditions. Surprisingly, WPT, STFT, and Mel spectrogram are not able to classify the non-degraded state correctly. This is because the wavelet coefficients or spectrograms of Label 1 (No degradation) and Label 2 (Intermediate degradation) are quite similar. The WPT, STFT, and Mel spectrogram fail to produce a higher frequency resolution in the low frequency range (dominant frequency range of slab track vibration). Therefore, they lack the ability of effectively distinguishing Label 1 and Label 2 in a robust way.

DeSpaWN shows, a very good performance on all classes and outperforms other feature extraction methods with an accuracy improvement of 2.8% to 45.4%. This indicates that DeSpaWN has the ability to learn effective thresholds for denoising and limiting the impact of the speed on the obtained features and, thus, achieves a robust performance in the classification task for all the classes.

2) Other compositions of training and test datasets

We evaluate three other compositions of training and test datasets as listed in Table 6.
Table 6. Train speed assignment in training and test dataset

| Composition Number | Training dataset | Test dataset |
|--------------------|------------------|--------------|
| C1                 | 40, 60, 80 km/h  | 20 km/h      |
| C2                 | 20, 60, 80 km/h  | 40 km/h      |
| C3                 | 20, 40, 80 km/h  | 60 km/h      |

Table 7 shows the average classification accuracy of classifying the three types of slab track under each composition of training and test datasets. The classification accuracy of using almost all the feature extraction / learning methods based on acceleration signals reaches 100% irrespective of the speed composition in training and test dataset. It demonstrates once again the excellent performance of classification using acceleration signals.

For the classification based on acoustic signals, DeSpaWN performs consistently the best in the three conditions, with an accuracy improvement of 0.3% to 13.9% compared to other feature extraction methods. In addition, the classification tasks related to speed interpolation regime (dataset compositions C2 and C3) achieve a globally better classification performance compared to the tasks related to speed extrapolation regime (dataset composition C1 and the case described in Section C (1)).

It is worth noting that the classification accuracy on the dataset composition C1 based on acoustic signals is the lowest and even hardly reaches 50% accuracy. The test dataset of C1 is formed by acoustic signals with the train speed of 20 km/h. The slab track vibration under this slow train speed is weak, while some noise such as the machine noise produced by the train and the background noise are basically constant regardless of the train speed. This induces a low signal to noise ratio (SNR) of the acoustic signal. The target features covered by severe noise cannot be effectively extracted by the current feature learning / extraction methods, which leads to the low classification accuracy of C1.

Table 4. Comparative results based on acoustic signals in Classification Task 2

| Slab track condition               | Classification accuracy (%) | DeSpaWN | FDWT | WPT | STFT | Mel spectrogram |
|------------------------------------|-----------------------------|---------|------|-----|------|-----------------|
| No degradation (Label 1)            |                             | 94.4    | 97.2 | 0.0 | 0.0  | 0.0             |
| Intermediate degradation (Label 2)  |                             | 80.6    | 61.1 | 83.0| 8.3  | 25.0            |
| Severe degradation (Label 3)        |                             | 69.4    | 77.8 | 77.8| 100.0| 100.0           |
| Average                            |                             | 81.5    | 78.7 | 53.6| 36.1 | 41.7            |

Table 5. Comparative results based on acceleration signals in Classification Task 2

| Slab track condition               | Classification accuracy (%) | DeSpaWN | FDWT | WPT | STFT | Mel spectrogram |
|------------------------------------|-----------------------------|---------|------|-----|------|-----------------|
| No degradation (Label 1)            |                             | 100.0   | 97.2 | 100.0| 100.0| 100.0           |
| Intermediate degradation (Label 2)  |                             | 100.0   | 100.0| 100.0| 100.0| 100.0           |
| Severe degradation (Label 3)        |                             | 100.0   | 100.0| 100.0| 100.0| 100.0           |
| Average                            |                             | 100.0   | 99.1 | 100.0| 100.0| 100.0           |

Table 7. Average classification accuracy based on acceleration and acoustic signals

| Signal type | Dataset composition | Classification accuracy (%) | DeSpaWN | FDWT | WPT | STFT | Mel spectrogram |
|-------------|---------------------|----------------------------|---------|------|-----|------|-----------------|
| Acceleration| C1                  |                            | 100.0   | 100.0| 100.0| 98.1 | 100.0           |
|             | C2                  |                            | 100.0   | 100.0| 100.0| 100.0| 100.0           |
|             | C3                  |                            | 100.0   | 100.0| 100.0| 100.0| 100.0           |
| Acoustic    | C1                  |                            | 50.0    | 40.7 | 38.9| 49.1 | 38.9            |
|             | C2                  |                            | 83.6    | 82.4 | 74.1| 82.4 | 83.3            |
|             | C3                  |                            | 84.3    | 83.3 | 76.3| 70.4 | 76.8            |
Another possible factor causing the low classification accuracy of C1 based on acoustic signals is that the acoustic signals radiated by slab tracks appear to be highly impacted by different train speeds, resulting in different patterns. However, this possible cause is less significant compared to the influence of low SNR of acoustic signals under the condition of low train speeds. It can be inferred from the fact that the characteristic patterns of structural acoustic signals and acceleration signals are similar while the classification accuracy of C1 based on acceleration signals reaches a remarkably high level (reaching almost 100% using different feature extraction / learning methods).

The comparison of classification results between Task 1 and Task 2 demonstrates the importance of constructing a training dataset with sufficiently representative operating conditions for slab track state monitoring. Detecting the slab track condition under the train speed that has never appeared in the training dataset may result in a low detection accuracy. To alleviate the problem in this case, an effective feature extraction method needs to be carefully determined.

4. Conclusion

In this study, we use three types of slab track with different supporting conditions (mortar, rubber and spring support) in a railway test line to imitate different healthy and unhealthy states of slab tracks. Track-side acoustic and acceleration signals are collected as a train passes by with different speeds. We use a deep learning framework based on the recently proposed DeSpaWN to automatically learn meaningful and sparse representations of raw high-frequency signals. Several other feature extraction methods with two classification tasks (Task 1: same train speeds in training and test dataset, Task 2: different train speeds in training and test dataset) are adopted for the comparison of promoting classification effectiveness of slab track states.

The classification accuracy using acceleration signals can reach almost 100% irrespective of the applied feature learning / extraction method. This is because the acceleration signals hardly suffer from the environmental interference. On the contrary, due to the more severe noise interference in acoustic signals, the classification performance shown in Task 1 and the generalization ability evaluated by Task 2 of using acoustic signals are worse compared to using acceleration signals.

For the classification based on acoustic signals, which have a higher noise level than acceleration signals, DeSpaWN reaches a competitive performance improvement compared to other feature extraction methods. Besides, equipped with flexible time and frequency resolutions at different decomposition layers, DeSpaWN has the ability of interpreting signals in a targeted manner, thus, limits the impact of the train speed on the obtained features and achieves a robust performance in the classification task.

Classifying the slab track conditions in Task 2 is a speed extrapolation regime. A further study on speed interpolation regime will be conducted. The potential conclusion would be constructive for optimizing the training dataset to improve the generalization ability of SVM classifier in the detection of slab track states.

Acknowledgement

This study was supported by the National Natural Science Foundation of China (grant numbers 52178432 and 51878501) and China Scholarship Council (grant number 202106260178).

References

[1] L. Gao, B. An, T. Xin et al., “Measurement, analysis, and model updating based on the modal parameters of high-speed railway ballastless track,” Measurement, vol. 161, pp. 107891, 2020.
[2] S. Zhu, and C. Cai, “Interface damage and its effect on vibrations of slab track under temperature and vehicle dynamic loads,” International Journal of Non-Linear Mechanics, vol. 58, pp. 222–232, 01/31, 2014.
[3] H. Gou, R. Xie, C. Liu et al., “Analytical study on high-speed railway track deformation under long-term bridge deformations and interlayer degradation.” pp. 1005-1015.
[4] N. Popovic, G. Feltrin, K. E. Jalsan et al., “Event - driven strain cycle monitoring of railway bridges using a wireless sensor network with sentinel nodes,” Structural Control and Health Monitoring, vol. 24, no. 7, pp. e1934, 2017.
[5] C. Du, S. Dutta, P. Kurup et al., “A review of railway infrastructure monitoring using fiber optic sensors,” Sensors and Actuators A: Physical, vol. 303, pp. 111728, 2020.
[6] J. Zhang, W. Huang, W. Zhang et al., “Train-induced vibration monitoring of track slab under long-term temperature load using fiber-optic accelerometers,” Sensors, vol. 21, no. 3, pp. 787, 2021.
[7] G. Guo, X. Cui, and B. Du, “Random‒Forest Machine Learning Approach for High‒Speed Railway Track Slab Deformation Identification Using Track-Side Vibration Monitoring,” Applied Sciences, vol. 11, no. 11, pp. 4756, 2021.
[8] G. Michau, G. Frusque, and O. Fink, “Fully learnable deep wavelet transform for unsupervised monitoring of high-frequency time series,” Proceedings of the National Academy of Sciences, vol. 119, no. 8, pp. e2106598119, 2022.
[9] S. Mohine, B. S. Bansod, P. Kumar et al., “Single acoustic sensor-based time–frequency spectrum sensing approach for land vehicle detection,” IEEE Sensors Journal, vol. 20, no. 13, pp. 7275-7282, 2020.
[10] Y. Koizumi, Y. Kawaguchi, K. Imoto et al., “Description and discussion on DCASE2020 challenge task2: Unsupervised anomalous sound detection for machine condition monitoring,” arXiv preprint arXiv:2006.05822, 2020.
[11] L. Gantert, M. Sammarco, M. Detyniecki et al., “A supervised approach for corrective maintenance using spectral features from industrial sounds,” in 2021 IEEE 7th World Forum on Internet of Things (WF-IoT), 2021, pp. 723-728.
[12] A. Pieringer, M. Stangl, J. Rothhämel et al., “Acoustic monitoring of rail faults in the German railway network,” in 13th International Workshop on Railway Noise, Ghent, 2019.
[13] Y. Wang, P. Wang, Q. Wang et al., “Using Vehicle Interior Noise Classification for Monitoring Urban Rail Transit Infrastructure,” Sensors, vol. 20, no. 4, 2020.
[14] H. Meng, S. Wang, C. Gao et al., “Research on recognition method of railway perimeter intrusions based on Φ-OTDR optical fiber sensing technology,” IEEE Sensors Journal, vol. 21, no. 8, pp. 9852-9859, 2020.
[15] Y. Zhou, S. Wang, M. Zhou et al., “Percussion-based bolt looseness identification using vibration-guided sound reconstruction,” Structural Control and Health Monitoring, vol. 29, no. 2, pp. e2876, 2022.
[16] J. Wang, L. Huo, C. Liu et al., “A new acoustic emission damage localization method using synchrosqueezed wavelet transforms picker and time-order method,” Structural Health Monitoring, vol. 20, no. 6, pp. 2917-2935, 2021.
[17] M. Ravanelli, and Y. Bengio, “Speaker recognition from raw waveform with sincnet,” in 2018 IEEE Spoken Language Technology Workshop (SLT), 2018, pp. 1021-1028.
[18] H. B. Sailor, D. M. Agrawal, and H. A. Patil, “Unsupervised Filterbank Learning Using Convolutional Restricted Boltzmann Machine for Environmental Sound Classification,” in Interspeech, 2017, pp. 9.
[19] W. Ha, C. Singh, F. Lanusse et al., “Adaptive wavelet distillation from neural networks through interpretations,” Advances in Neural Information Processing Systems, vol. 34, 2021.
[20] G. Frusque, and O. Fink, “Learnable Wavelet Packet Transform for Data-Adapted Spectrograms,” arXiv preprint arXiv:2201.11069, 2022.