



NTScatNet: An interpretable convolutional neural network for domain generalization diagnosis across different transmission paths

Chao Liu ^a, Xiaolong Ma ^b, Tianyu Han ^a, Xi Shi ^{a,*}, Chengjin Qin ^a, Songtao Hu ^a

^a State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China

^b Shanghai Institute of Aerospace System Engineering, Shanghai, PR China

ARTICLE INFO

Keywords:

Intelligent fault diagnosis
Interpretable convolutional neural network
Wavelet scattering transform
Domain generalization
Transducers across different transmission paths

ABSTRACT

The past decade has witnessed the convolutional neural network(CNN)'s significant progress in intelligent fault diagnosis research. Nevertheless, CNN's weak interpretability and poor domain generalization capability have been preventing its application in industrial practice. The current work develops an interpretable domain generalization diagnosis model, *i.e.*, the normalized wavelet scattering convolutional network (NTScatNet), to remedy these gaps. The architecture of NTScatNet is similar to a standard CNN, while NTScatNet's distinctiveness is that it takes Morlet wavelet convolutional kernel, modulo activation function, and moving averaging pooling layer. This article gives a detailed physical interpretation of each layer of NTScatNet and shows that NTScatNet's feature extractor well characterizes the multi-scale cyclostationarity information of the fault signals. Besides, this work theoretically illustrates the normalized scattering feature's invariance to a linear time-invariant system, indicating that NTScatNet holds the domain generalization diagnosis capability across different transmission paths. Finally, NTScatNet's domain generalization ability across transmission paths is experimentally verified on the transfer diagnosis tasks across transducers and the tasks of detecting foreign objects on the escalator guide rail. Incidentally, this manuscript is a prelude to the second manuscript entitled "Scattering Moment Matching-based Interpretable Domain Adaptation for Transfer Diagnosis Tasks across Bearing specifications and Transducers", which proposes a novel interpretable domain adaptation method.

1. Introduction

Applying machine learning to machine fault diagnosis, referred to as the intelligent fault diagnosis(IFD) technology, has been a research hot in vibration-based machine condition monitoring [1]. IFD replaces diagnosis tasks with a classification problem based on the idea that diagnosis experience from the training dataset could generalize to new testing tasks. According to Antoni's overview paper [2], the IFD technique follows a "Filterbank-Feature-Decision" methodology. The "Filterbank" phase aims to enhance diagnostic information, which is conveyed by the cyclostationarity of the fault signals, with advanced signal processing techniques. Then the "Feature" phase codes diagnostic information into low-dimensional features. Finally, the "Decision" phase maps the handcrafted features to the health conditions of the monitored machine with a machine learning classifier. The rapidly developing convolutional neural network(CNN) technique skips the exhausting handcraft feature stage and fuses the "Filterbank-Feature-Decision" methodology into an end-to-end pipeline, providing an approach to automatically abstract fault representation from training

data. However, the weak interpretability and poor domain generalization capability are two main obstacles preventing applying CNN-based diagnosis techniques into industrial practice.

The standard CNN architecture consists of a hierarchical parameterized feature extractor and a softmax layer. The softmax layer is an interpretable linear classifier. However, the representation learning process is not transparent to the diagnosis engineers due to complex architecture and training data-dependent filters' characteristics. Enhancing the interpretability of the feature representation modules helps enhance engineers' confidence in CNN's prediction and has attracted increasing research interest. One solution is post hoc explaining the trained CNN through weights visualization. For example, Lei et al. [3] found that the trained sparse filtering layer behaves as the Gabor wavelet. Jia et al. [4] work revealed that the convolutional layer in a deep network behaves similarly to multi-channel filterbanks. The other solution is enhancing CNN's intrinsic interpretability by restricting the complexity of network architecture. Li and Yan [5] introduced a continuous wavelet convolutional layer(CWConv) to replace CNN's first random initialization layer to simplify the representing learning

* Correspondence to: Shanghai Jiao Tong University, China.

E-mail address: xishi@sjtu.edu.cn (X. Shi).

process. The CWConv layer is physically interpretable and outputs fault-related semantic information. However, the above work [3–5] focuses only on the first layer of CNN yet fails to whiten the other layers. Recently, Liu and Shi [6] introduced Mallat's wavelet scattering theory into IFD and developed a time scattering convolutional network(TSNet) model for bearing diagnosis [7–12]. TSNet's distinction from WaveletKernelNet is that all convolutional layers are wavelet kernels, not just the first layer. Since both Morlet wavelet transform and softmax layer are intrinsic interpretable, TSNet is a fully interpretable convolutional network.

Besides the interpretability, CNN's poor domain generalization capability is the other limitation. Two concepts, i.e., interpolation and extrapolation, are borrowed from the statistics to explain the IFD model's generalization performance. When new test samples are within the range of the training set, the prediction stage is called interpolation, while when the testing data falls outside the training set, the prediction stage is called extrapolation. Numerous works have witnessed IFD models work well under the interpolation assumption, i.e., test samples are sampled from the same distribution as the historical training samples [13,14]. However, the interpolation hypothesis requires training and testing samples collected from identical working conditions and the same vibration transducer, which is hard to meet in industrial practice. It is practically significant to enhance the IFD model's extrapolation capability to promote the IFD methodology's application to real industries. Domain adaptation has shown the potential to improve the IFD model's generalization capability [1,15–17] and has been applied to address several kinds of transfer diagnosis tasks, e.g., transfer diagnosis across working conditions [18–20], and across transducers at different positions [21,22]. However, the above research generally requires unlabeled or semi-supervised target domain data to align data distribution between source and target domain during the training phase. Although recent research [23–33] developed advanced deep learning-based diagnosis methods for domain generalization tasks across different working conditions, these methods are uninterpretable due to complex deep network architecture.

Benefiting from the L_2 stability and time-warping stability of wavelet scattering, the interpretable TSNet model proposed by Liu and Shi holds excellent domain generalization diagnosis performance across working loads, operation speeds, and additive noise interference. However, TSNet [6] and Ref. [23–32] generally focus on domain generalization study across working conditions and do not concern with domain generalization tasks across different transmission paths. In fact, it is of practical significance to develop domain generalization diagnosis methods across different transmissions. In the security-sensitive industry scenario, multiple transducers are generally mounted to monitor the machine's health condition. Once one transducer malfunctions, the adjacent transducer could substitute the malfunctioned transducer immediately, improving the reliability margin of the health monitoring system. In the cost-sensitive scenario, e.g., the tasks of detecting foreign objects on the escalator guide, the domain generalization diagnosis method could detect the abnormal state at different locations with fewer transducers, saving the transducer cost.

The current work improves TSNet into a normalized wavelet scattering convolutional network(NTSNet) to enhance the IFD model's generalization capability across transmission paths and gives a detailed interpretation of each layer of NTSNet. The contributions of this article are summarized below.

- (1) An interpretable normalized wavelet scattering convolutional network is developed to address domain generalization diagnosis tasks across different transmissions.
- (2) The physical interpretation of each layer of NTSNet is described in detail. Specifically, the Conv1 layer conducts a multi-scale Hilbert demodulation operation on the input signals and outputs a wavelet scalogram. The Conv2 layer analysis the cyclic frequency of the demodulated signals in a constant-Q way and

outputs the “scalogram of scalogram”. Besides, the pooling layer and the global average pooling layer output the wavelet scattering features and global averaged wavelet scattering features, respectively, characterizing multi-scale cyclostationary information of the fault signals.

- (3) This work theoretically illustrates that the normalized scattering features output by the scattering normalization layer is invariant to a linear time-invariant system and experimentally verifies that NTSNet holds the domain generalization diagnosis capability across different transmission paths.

The rest of this article is organized as follows. Section 2 reviews a standard CNN architecture applied to intelligent fault diagnosis. Section 3 develops a novel NTSNet model, describes the operation and physical meaning of each layer of NTSNet in detail, and theoretically illustrates the normalized wavelet scattering feature's invariance to different transmission paths. Section 4 experimentally verifies NTSNet's domain generalization ability across transmission paths through the transfer diagnosis tasks across transducers and the tasks of detecting foreign objects on the escalator guide rail. Section 5 concludes the current article.

2. A standard CNN architecture applied to intelligent fault diagnosis

The convolutional neural network(CNN) is a special deep feed-forward neural network with a local connection and weight-sharing properties. Fig. 1 shows a standard CNN architecture applied to IFD tasks, which contains an input layer, several convolutional layers, pooling layers, two fully connected layers, and a Softmax output layer. The operation of each layer of CNN is described briefly below.

(1) *Input Layer*: The input layer receives a time-domain vibration signal x of length L .

(2) *Convolutional layer*: The convolutional layer(Conv layer) slides a group of learnable convolution kernels on the original input signal (or the intermediate feature map) to extract a new set of feature maps through the convolution operation. The convolution process can be expressed as:

$$x_m^l = \text{ReLU}\left(\sum_n w_{m,n}^l * x_n^{l-1} + b_m^l\right), \quad (1)$$

where x_n^{l-1} is n th feature input of the l th layer, $w_{m,n}^l$ is the weights of the n th channel of the m th convolution kernel in the l th layer, b_m^l is the bias vector of the m th kernel, x_m^l is the m th feature vector output by the l th convolutional layer, and $\text{ReLU}(\cdot)$ is the ReLU activation function.

(3) *Pooling layer*: The pooling layer conducts dimensionality reduction operation on the feature map output by the Conv layer. The outputs of the pooling layer are features with local time-shift invariance and small deformation stability. The max pooling and the average pooling are two commonly used pooling operations. The dimensionality reduction with max pooling could be expressed as:

$$x_m^l = \max\{x_m^{l-1}, r\}, \quad (2)$$

where $\max\{\cdot\}$ is the maximum pooling function, x_m^{l-1} is the m feature map before pooling, x_m^l is the feature vector after pooling, and r represents the pooling parameters, e.g., size of pooling and steps.

(4) *Batch normalization*: The batch normalization layer(BN) aims to reduce internal covariate shifting and speed up the convergence of CNN's training process. Let $B = \{(x^i, y^i)\}_{i=1}^n$ be a batch of training data, μ and σ denote the mean value and standard deviation of B , respectively. Eq. (3) gives the operation in the BN layer:

$$\hat{x}^i = \gamma \left(\frac{x^i - \mu}{\sigma} \right) + \beta \quad (3)$$

where x^i and \hat{x}^i are the input and output of the BN layer, respectively, β and γ are learnable statistics of the BN layer.

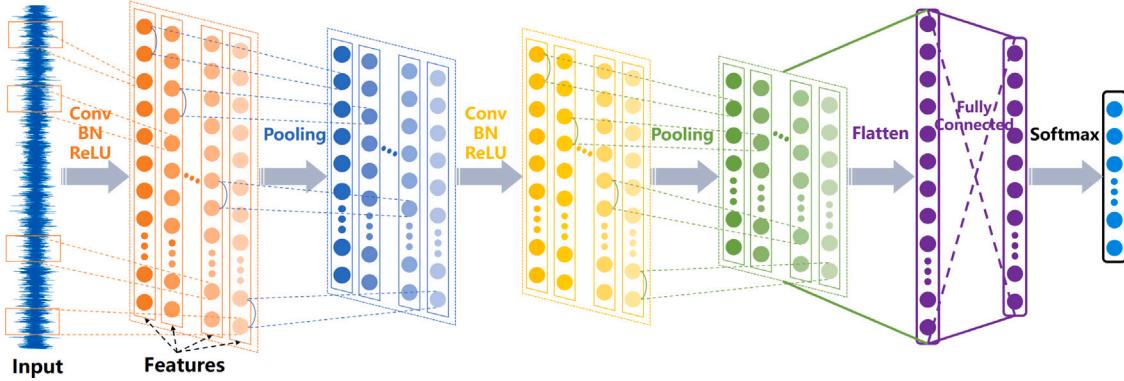


Fig. 1. A standard CNN applied to IFD.

(5) *Fully connected layer*: The operation in the l th fully connected(FC) layer could be expressed as:

$$x^l = \sigma(\mathbf{W}^T x^{l-1} + \mathbf{b}^l), \quad (4)$$

where x^{l-1} and x^l are the input and output of the l th FC layer, respectively; \mathbf{W} and \mathbf{b} are the weight matrix and bias vector of the l th FC layer; $\sigma(\cdot)$ is the activation function.

(6) *Softmax Layer*: The softmax layer maps a feature vector to probability values that the feature belongs to the c th category:

$$p_c(x^l) = \frac{e^{\mathbf{w}_c^T x^l + b_c}}{\sum_{i=1}^C e^{\mathbf{w}_i^T x^l + b_i}}, \quad (5)$$

where x^l is the input of the softmax layer, \mathbf{w}_c and b_c are the weights and the bias of the softmax layer, respectively, $p_c(x^l)$ is the probability value that x^l belongs to the c th category.

3. Proposed interpretable CNN

This section develops an interpretable CNN, i.e., the normalized wavelet scattering convolutional network(NTScatNet), to enhance standard CNN's interpretability. Section 3.1 presents NTScatNet's architecture and describes the operation in each layer of NTScatNet. The physical interpretation of each layer of NTScatNet is given in Section 3.2 by visualizing how NTScatNet operates a simulated bearing outer race fault signal. Section 3.3 illustrates NTScatNet's domain generalization capability across transmission paths through theoretical derivation. Section 3.4 summarizes the training and testing processes of the NTScatNet-based IFD method and analyzes the computational complexity.

3.1. Architecture of NTScatNet

As Fig. 2 presents, the proposed NTScatNet consists of an input layer, two convolutional layers, a pooling layer, a global averaging pooling layer(GAP), a scattering normalization layer(SN), and a softmax layer. Despite being similar to conventional CNN architecture, NTScatNet's distinct from standard CNN is that all convolutional kernels are pre-defined Morlet wavelets as opposed to learnable kernels, the activation function is a modulo operator, and the pooling layer takes a moving averaging operation. The operation of each layer of NTScatNet is described below.

(1) *Input Layer*: The input layer receives a raw time-domain signal \mathbf{x} of length L , where L is a hyperparameter.

(2) *Conv1 Layer*: The convolution kernel of the Conv1 layer is a group of Morlet wavelets $\{\psi_{\lambda_1}\}_{\lambda_1}$. The Morlet wavelet filterbank's quality factor is Q_1 , and the center frequency is $\lambda_1 = 2^k/Q_1$, $k \in \mathbb{Z}$, where Q_1 is a hyperparameter that denotes the number of wavelet filters per octave. Eq. (6) gives the operation in the Conv1 layer, which

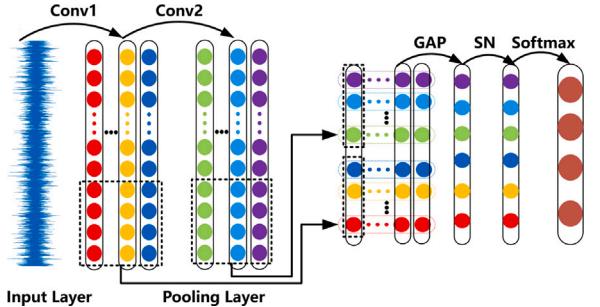


Fig. 2. Architecture of NTScatNet.

presents that the input signal \mathbf{x} is first convolved with the Morlet wavelet filter and then activated by the modulo operator.

$$U_1(\mathbf{x}) = \text{Conv1}[\mathbf{x}] = |\mathbf{x} * \psi_{\lambda_1}| \quad (6)$$

(3) *Conv2 Layer*: The convolution kernel of the Conv2 layer is the other group of Morlet wavelets $\{\psi_{\lambda_2}\}_{\lambda_2}$, the quality factor and the center frequencies of which are Q_2 and $\lambda_2 = 2^k/Q_2$, $k \in \mathbb{Z}$, respectively. Conv2 layer performs Morlet wavelet convolution and modulo activation operations on each subband signals output by Conv1 layer.

$$U_2(\mathbf{x}) = \text{Conv2}[|\mathbf{x} * \psi_{\lambda_1}|] = ||\mathbf{x} * \psi_{\lambda_1}| * \psi_{\lambda_2}| \quad (7)$$

(4) *Pooling Layer*: The pooling layer conducts moving average pooling operation on $U_1(\mathbf{x})$ and $U_2(\mathbf{x})$. The downsampling operator takes the scale filter ϕ , which is complementary to the Morlet wavelet filterbank. The recommended window width and the step size are $L/2$ and $L/8$, respectively. The output of the pooling layer is the first- and the second-order wavelet scattering features.

$$S_1(\mathbf{x}) = U_1(\mathbf{x}) * \phi = |\mathbf{x} * \psi_{\lambda_1}| * \psi_{\lambda_1} * \phi \quad (8)$$

$$S_2(\mathbf{x}) = U_2(\mathbf{x}) * \phi = ||\mathbf{x} * \psi_{\lambda_1}| * \psi_{\lambda_2}| * \phi \quad (9)$$

(5) *GAP Layer*: The GAP layer performs global average pooling operation on the $S_1(\mathbf{x})$ and $S_2(\mathbf{x})$. The output of the GAP layer is the globally averaged wavelet scattering features.

$$\bar{S}_1(\mathbf{x}) = \text{GAP}(S_1(\mathbf{x})) \quad (10)$$

$$\bar{S}_2(\mathbf{x}) = \text{GAP}(S_2(\mathbf{x})) \quad (11)$$

(6) *SN Layer*: The SN layer conducts a scattering normalization operation on the global averaged wavelet scattering features and outputs the normalized wavelet scattering features.

$$\tilde{S}_1(\mathbf{x}) = \bar{S}_1(\mathbf{x}) / (|\mathbf{x}| * \phi) \quad (12)$$

Table 1
Information of NTScatNet.

Layer	Input	Output	Hyperparameters
Input layer	x	x	L
Conv1 layer	x	$U_1(x)$	Q_1
Conv2 layer	$U_1(x)$	$U_2(x)$	Q_2
Pooling layer	$U_1(x), U_2(x)$	$S(x) = \{S_1(x); S_2(x)\}$	$L/2, L/8$
GAP layer	$S(x)$	$\tilde{S}(x) = \{\tilde{S}_1(x); \tilde{S}_2(x)\}$	–
SN layer	$\tilde{S}(x)$	$\tilde{S}(x) = \{\tilde{S}_1(x); \tilde{S}_2(x)\}$	–
Softmax layer	$\tilde{S}(x)$	predicted label	C

$$\tilde{S}_2(x) = \overline{S}_2(x)/\overline{S}_1(x) \quad (13)$$

The normalized wavelet scattering features should be flattened and cascaded before being fed to the softmax layer.

$$\tilde{S}(x) = \left\{ \tilde{S}_1(x); \tilde{S}_2(x) \right\} \quad (14)$$

(7) *Softmax Layer*: Let $D_S : \{x_S^i, y_S^i\}_{i=1}^{N_S}$ the source domain dataset and $D_T : \{x_T^j, y_T^j\}_{j=1}^{N_T}$ the target domain dataset. During the training phase, the Softmax layer maps the normalized wavelet scattering features of x_S^i to the probability values that x_S^i belongs to each fault category:

$$y_{\text{score}}^i = \frac{1}{\sum_{k=1}^C e^{(w_k^T \tilde{S}(x_S^i) + b_k)}} \begin{bmatrix} e^{(w_1^T \tilde{S}(x_S^i) + b_1)} \\ e^{(w_2^T \tilde{S}(x_S^i) + b_2)} \\ \dots \\ e^{(w_C^T \tilde{S}(x_S^i) + b_C)} \end{bmatrix}, \quad (15)$$

where w_k and b_k are the weight vector and bias value that corresponds to the k th category, respectively, and C is the total number of categories. The data loss of the source domain could be calculated according to Eq. (16), and the softmax layer's parameters could be optimized with the mini-batch gradient descent algorithm according to Eq. (17).

$$\text{Loss} = -\frac{1}{N_S} \left[\sum_{i=1}^{N_S} \sum_{k=1}^C I[y^i = k] \log \frac{e^{(w_k^T \tilde{S}(x_S^i) + b_k)}}{\sum_{k=1}^C e^{(w_k^T \tilde{S}(x_S^i) + b_k)}} \right] \quad (16)$$

$$W^* = \underset{W}{\operatorname{argmin}} \text{Loss}(W), \quad W \leftarrow W - \varepsilon \frac{\partial \text{Loss}}{\partial W} \quad (17)$$

During the testing phase, the trained softmax classifier predicts the label of the target domain sample according to Eq. (18).

$$y_{\text{pre}}^j = \underset{k}{\operatorname{argmax}} \frac{e^{(w_k^* T \tilde{S}(x_T^j) + b_k^*)}}{\sum_{k=1}^C e^{(w_k^* T \tilde{S}(x_T^j) + b_k^*)}} \quad (18)$$

The input, output, and hyperparameters of each layer of NTScatNet are summarized in Table 1.

3.2. Interpretation of each layer of NTScatNet

One limitation of the conventional CNN is that its feature extraction process is not transparent to diagnosis engineers due to the complex architecture. The proposed NTScatNet simplifies the standard CNN architecture and improves model interpretability. This subsection gives the physical interpretation of each layer of NTScatNet by visualizing how NTScatNet operates a simulated bearing outer race fault signal $x(t)$ presented in Fig. 3.

(1) *Input Layer*: The input layer receives the simulated bearing outer race fault signal $x(t)$.

(2) *Conv1 Layer*: Conv1 layer's convolution kernel is a family of Morlet wavelets, the frequency domain counterpart of which is a constant-Q bandpass filterbank. Fig. 4 presents the Morlet wavelet filterbank with $Q_1 = 4$. The Conv1 layer conducts the Morlet wavelet

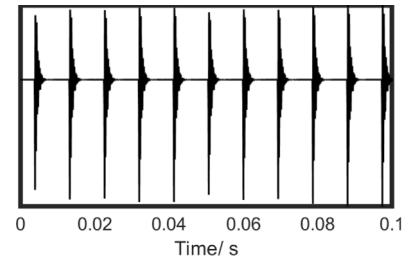


Fig. 3. The simulated bearing outer race fault signal $x(t)$.

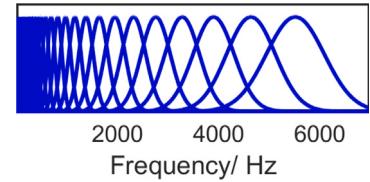


Fig. 4. Morlet wavelet filter bank with $Q_1 = 4$.

transform on $x(t)$ and decomposes $x(t)$ into P1 subband wavelet signals, where P1 is the number of Morlet wavelet filters of the filterbank shown in Fig. 4.

$$\begin{aligned} x(t) * \psi_{\lambda_1}(t) &= w_R(t, \lambda_1) + j w_I(t, \lambda_1) \\ &= w_R(t, \lambda_1) + j \mathcal{H}[w_R(t, \lambda_1)] \end{aligned} \quad (19)$$

In Eq. (19), $w_R(t, \lambda_1)$ and $w_I(t, \lambda_1)$ are the real and imaginary part of the wavelet coefficients, respectively, $\mathcal{H}[\cdot]$ is the Hilbert transform operator. The complex Morlet wavelet coefficients are then activated by the modulo operator, which extracts the Hilbert envelope at each wavelet scale.

$$\begin{aligned} U_1 x(t, \lambda_1) &= |x(t) * \psi_{\lambda_1}(t)| \\ &= \sqrt{w_R(t, \lambda_1)^2 + H[w_R(t, \lambda_1)]^2} \end{aligned} \quad (20)$$

This article terms the Conv1 layer ‘‘Demodulation Filterbank’’ since it performs multiscale Hilbert demodulation on the input signal. Fig. 6 presents the multiscale demodulated envelope signals output by the Conv1 layer.

(3) *Conv2 Layer*: Conv2 layer is the other constant-Q bandpass filterbank, which contains P2 Morlet wavelets. Fig. 7 presents a Morlet wavelet filterbank with $Q_2 = 3$. Conv2 layer conducts Morlet wavelet transform on $U_1 x(t, \lambda_1)$ and then activates the output with the modulo operator.

$$U_2 x(t, \lambda_1, \lambda_2) = \|x(t) * \psi_{\lambda_1}\| * \psi_{\lambda_2} \quad (21)$$

As Fig. 5 presents, the Conv1 layer decomposes $x(t)$ into a wavelet scalogram $U_1 x(t, \lambda_1)$, while the Conv2 layer transforms each demodulated Hilbert envelope into a new scalogram $U_2 x(t, \lambda_1, \lambda_2)$. This article terms $U_2 x(t, \lambda_1, \lambda_2)$ ‘‘Scalogram of Scalogram’’ and the Conv2 layer ‘‘Analysis Filterbank’’ since it analyzes the cyclic frequency of each demodulated envelope with a constant bandwidth filterbank.

(4) *Pooling Layer*: As Fig. 5 shows, the pooling layer downsamples the wavelet scalogram and ‘‘Scalogram of Scalogram’’ with a moving average operator ϕ to obtain local time shift invariant representation. The outputs of the pooling layer are the first- and second-order wavelet scattering feature maps.

$$S_1 x(t, \lambda_1) = U_1 x(t, \lambda_1) * \phi(t) \quad (22)$$

$$S_2 x(t, \lambda_1, \lambda_2) = U_2 x(t, \lambda_1, \lambda_2) * \phi(t) \quad (23)$$

For example, let $L/2$ the window length of $\phi(t)$, and $L/8$ the step size of the moving average. Then the size of $S_1 x(t, \lambda_1)$ and $S_2 x(t, \lambda_1, \lambda_2)$ are $5 \times P_1$ and $5 \times P_1 \times P_2$, respectively.

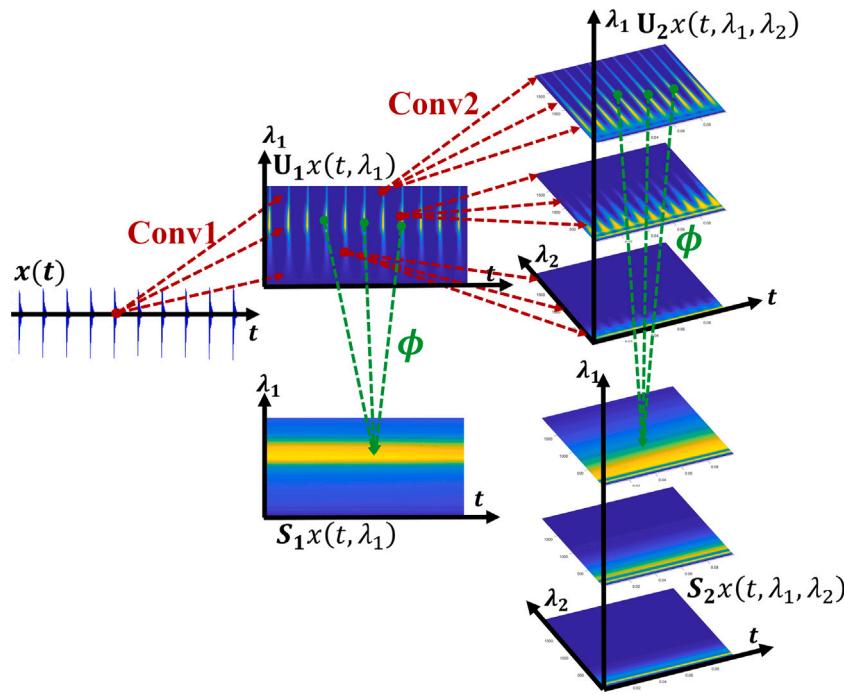


Fig. 5. Interpreting the feature extractor of NTScatNet as Demodulation-Analysis Filterbanks.

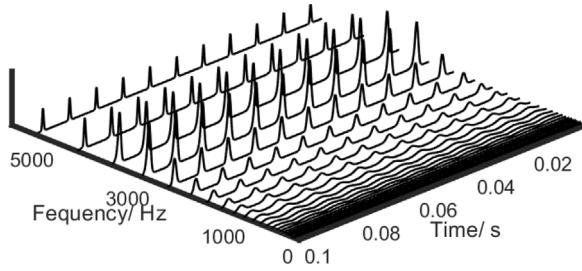


Fig. 6. Multiscale demodulated envelope signals.

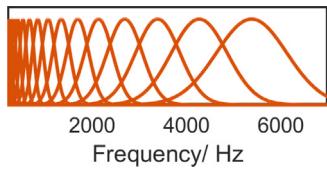


Fig. 7. Morlet wavelet filter bank with $Q_2 = 3$.

(5) GAP Layer: The GAP layer conducts global average pooling on $S_1x(t, \lambda_1)$ and $S_2x(t, \lambda_1, \lambda_2)$ to obtain global time-shift invariant representation.

$$\bar{S}_1x(\lambda_1) = \text{GAP}(S_1x(t, \lambda_1)) \quad (24)$$

$$\bar{S}_2(\lambda_1, \lambda_2) = \text{GAP}(S_2x(t, \lambda_1, \lambda_2)) \quad (25)$$

Fig. 8(a) shows $x(t)$'s Fourier spectrum, and Fig. 8(b)–8(d) present the globally averaged first-order wavelet scattering feature maps of $x(t)$ with $Q_1 = 4, 8, 16$, respectively. Fig. 8 illustrates that the globally averaged first-order wavelet scattering feature map could be considered an approximation of the Fourier spectrum, whose frequency resolution improves with the increase of the quality factor Q_1 . In other words, $\bar{S}_1x(\lambda_1)$ captures the formant of $x(t)$ and the resonance of the monitored machine structure.

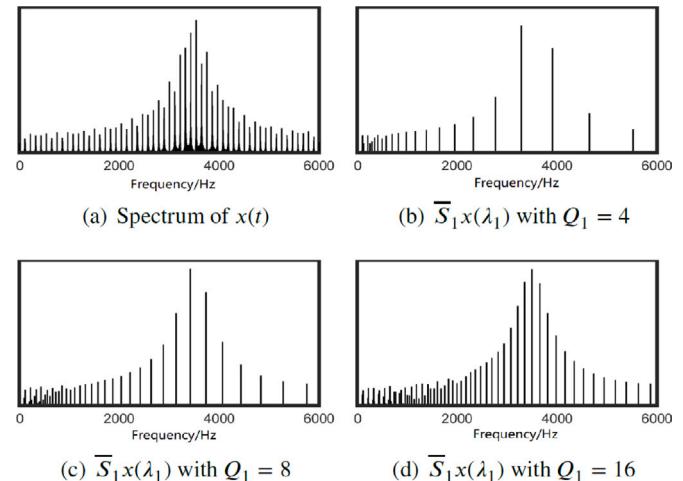


Fig. 8. The globally averaged first-order wavelet scattering feature maps.

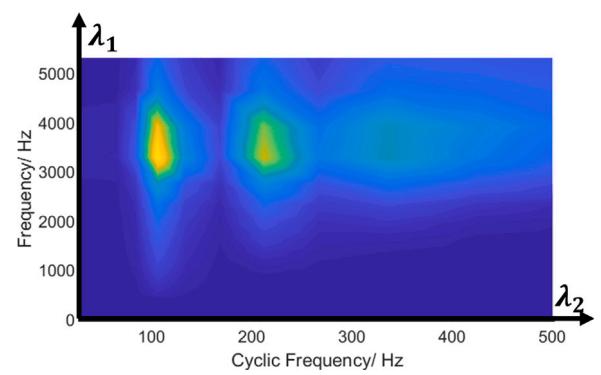


Fig. 9. Globally averaged second-order wavelet scattering feature.

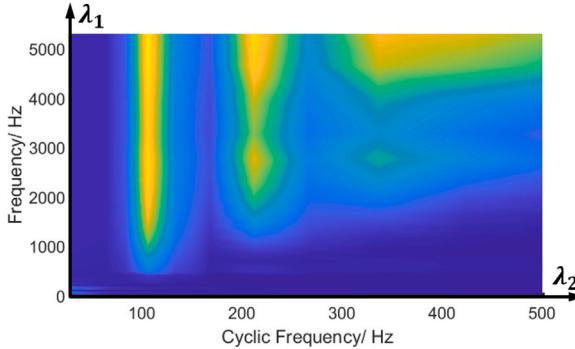


Fig. 10. Normalized second-order wavelet scattering feature.

The globally averaged second-order wavelet scattering feature map, i.e., $\bar{S}_2(\lambda_1, \lambda_2)$ is presented in Fig. 9. As Fig. 9 shows, $\bar{S}_2(\lambda_1, \lambda_2)$ is a bispectral map, where λ_1 is the “spectral frequency” axis, characterizing the resonance of the monitored machine, while λ_2 is the “cyclic frequency” axis, indicating $x(t)$ ’s cyclostationarity at the λ_1 scale. In other words, the globally averaged second-order wavelet scattering features $\bar{S}_2(\lambda_1, \lambda_2)$ characterizes the multiscale cyclostationarity of $x(t)$.

(6) *SN Layer*: The SN layer performs scattering normalization on the globally averaged wavelet scattering features to reduce the influence of the energy at different wavelet scales.

$$\tilde{S}_1 x(\lambda_1) = \bar{S}_1 x(\lambda_1) / (|x(t)| * \phi(t)) \quad (26)$$

$$\tilde{S}_2 x(\lambda_1, \lambda_2) = \bar{S}_2 x(\lambda_1, \lambda_2) / \bar{S}_1 x(\lambda_1) \quad (27)$$

The normalized second-order wavelet scattering feature of $x(t)$ is shown in Fig. 10, which illustrates that the normalized wavelet scattering-based cyclostationarity representation becomes insensitive to the energy at different wavelet scales.

(7) *Softmax Layer*: The softmax layer maps the normalized wavelet scattering features to the health condition category of the monitored machine structure. The softmax layer is interpretable since it is just a simple linear classifier. So far, the specific physical explanation of each layer of NTScatNet has been detailed, which indicates that the proposed NTScatNet is a fully interpretable convolutional neural network.

3.3. NtScatNet’s generalization across transmission paths

This subsection verifies NTScatNet’s domain generalization capability across transmission paths through a brief theory derivation. As is well known to the vibration-based condition monitoring research community, an impact excitation will be produced when a bearing defect strikes another surface, and the impact will excite one or more resonances of the monitored machine structure. The impulse response oscillates at the natural frequencies and decays rapidly due to structural damping, characterizing the characteristics of the transmission path between the impact point and the transducer. In practice, multiple transducers are generally mounted to monitor the machine’s health condition. As shown in Fig. 11, the impulse response excited by the bearing defect is observed by two transducers at two positions, where $x(t)$ is the impact signal, $h_1(t)$ characterizes the transmission path between the impact point and transducer1, and $h_2(t)$ characterizes the transmission path between impact point and transducer2. Despite corresponding to the same bearing fault, significant data distribution discrepancy exists between $x(t) * h_1(t)$ and $x(t) * h_2(t)$ due to different transmission path characteristics.

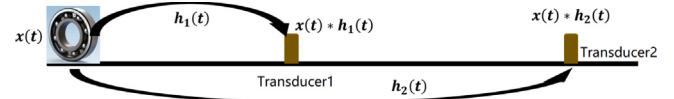


Fig. 11. Transmission path variability.

3.3.1. Linear time-invariant system’s response to morlet wavelet excitation

Let $h(t)$ the impulse response function of a linear time-invariant system, and $H(\omega)$ is the Fourier transform of $h(t)$, i.e., the linear time-invariant system’s frequency response function. As Eq. (28) presents, $h(t)$ ’s response to a complex exponential signal is a complex exponential signal with gain $H(\omega)$.

$$h(t) * e^{j\omega t} = H(\omega)e^{j\omega t} \quad (28)$$

In other words, the characteristic function of a time-invariant system is a group of complex exponential functions $e^{j\omega t}$. Morlet wavelet $\psi_\lambda(t)$ is a complex exponential signal modulated by Gaussian window, and the frequency counterpart of $\psi_\lambda(t)$ is a bandpass Gaussian filter $\Psi_\lambda(\omega)$. For a real mechanical system, the amplitude frequency response $H(\omega)$ of the system generally vary gently with frequency ω . Therefore, $H(\omega)$ could be approximated by a constant $H(\lambda)$ in a narrow wavelet frequency band support $[\lambda - \lambda/Q, \lambda + \lambda/Q]$:

$$H(\omega) \approx H(\lambda), \quad (29)$$

where Q and λ are the quality factor and center frequency of the Morlet wavelet filter, respectively. The linear time-invariant system’s response to a Morlet wavelet excitation could be approximately derived as:

$$\begin{aligned} h(t) * \psi_\lambda(t) &= F^{-1}[H(\omega)\Psi_\lambda(\omega)] \approx F^{-1}[H(\lambda)\Psi_\lambda(\omega)] \\ &= H(\lambda)F^{-1}[\Psi_\lambda(\omega)] = H(\lambda)\psi_\lambda(t) \end{aligned} \quad (30)$$

Eq. (30) implies that the linear time-invariant system’s response to a Morlet wavelet is a Morlet wavelet with gain $H(\lambda)$.

3.3.2. Wavelet scattering features of signals observed by transducers at different positions

According to Eqs. (8), (22), and (30), the first-order wavelet scattering features of $x(t) * h_1(t)$ could be derived as:

$$\begin{aligned} S_1[x(t) * h_1(t)](t, \lambda_1) &= |x(t) * h_1(t) * \psi_{\lambda_1}(t)| * \phi(t) \\ &\approx |x(t) * H_1(\lambda_1)\psi_{\lambda_1}(t)| * \phi(t) \\ &= H_1(\lambda_1)|x(t) * \psi_{\lambda_1}(t)| * \phi(t) \\ &= H_1(\lambda_1)S_1[x(t)](t, \lambda_1) \end{aligned} \quad (31)$$

The second-order wavelet scattering features of $x(t) * h_1(t)$ could be represented as Eq. (32) according to Eqs. (9), (23), and (30).

$$\begin{aligned} S_2[x(t) * h_1(t)](t, \lambda_1, \lambda_2) &= \|x(t) * h_1(t) * \psi_{\lambda_1}(t) * \psi_{\lambda_2}(t)\| * \phi(t) \\ &\approx \|x(t) * H_1(\lambda_1)\psi_{\lambda_1}(t) * \psi_{\lambda_2}(t)\| * \phi(t) \\ &= H_1(\lambda_1)\|x(t) * \psi_{\lambda_1}(t) * \psi_{\lambda_2}(t)\| * \phi(t) \\ &= H_1(\lambda_1)S_2[x(t)](t, \lambda_1, \lambda_2) \end{aligned} \quad (32)$$

Similarly, the first-order and the second-order wavelet scattering features of $x(t) * h_2(t)$ could be derived as Eqs. (33) and (34), respectively.

$$S_1[x(t) * h_2(t)](t, \lambda_1) = H_2(\lambda_1)S_1[x(t)](t, \lambda_1) \quad (33)$$

$$S_2[x(t) * h_2(t)](t, \lambda_1, \lambda_2) = H_2(\lambda_1)S_2[x(t)](t, \lambda_1, \lambda_2) \quad (34)$$

Since $H_1(\lambda_1) \neq H_2(\lambda_1)$, the first- and the second-order scattering feature are not invariant to transmission path.

$$\begin{aligned} S_1[x(t) * h_1(t)](t, \lambda_1) &\neq S_1[x(t) * h_2(t)](t, \lambda_1) \\ S_2[x(t) * h_1(t)](t, \lambda_1, \lambda_2) &\neq S_2[x(t) * h_2(t)](t, \lambda_1, \lambda_2) \end{aligned} \quad (35)$$

3.3.3. Normalized wavelet scattering features of signals observed by transducers at different positions

The normalized second-order scattering feature of $x(t) * h_1(t)$ could be presented as Eq. (36).

$$\begin{aligned} \tilde{S}_2[x(t) * h_1(t)](t, \lambda_1, \lambda_2) &= \frac{S_2[x(t) * h_1(t)](t, \lambda_1, \lambda_2)}{S_1[x(t) * h_1(t)](t, \lambda_1)} \\ &= \frac{H_1(\lambda_1)S_2[x(t)](t, \lambda_1, \lambda_2)}{H_1(\lambda_1)S_1[x(t)](t, \lambda_1)} = \frac{S_2[x(t)](t, \lambda_1, \lambda_2)}{S_1[x(t)](t, \lambda_1)} \end{aligned} \quad (36)$$

Similarly, the normalized second-order scattering feature of $x(t) * h_2(t)$ could be formulated as Eq. (37).

$$\begin{aligned} \tilde{S}_2[x(t) * h_2(t)](t, \lambda_1, \lambda_2) &= \frac{S_2[x(t) * h_2(t)](t, \lambda_1, \lambda_2)}{S_1[x(t) * h_2(t)](t, \lambda_1)} \\ &= \frac{H_2(\lambda_1)S_2[x(t)](t, \lambda_1, \lambda_2)}{H_2(\lambda_1)S_1[x(t)](t, \lambda_1)} = \frac{S_2[x(t)](t, \lambda_1, \lambda_2)}{S_1[x(t)](t, \lambda_1)} \end{aligned} \quad (37)$$

Comparing Eqs. (36) and (37) reveals that the normalized second-order scattering feature is invariant to a linear time-invariant system. In other words, NTScatNet could extract features invariant to transmission path variability and thus holds the domain generalization capability across different transmission paths.

3.4. NtScatNet-based IFD algorithm

3.4.1. Training and testing process of NTScatNet

The training and testing processes of NTScatNet are summarized in Algorithm 1 and Algorithm 2, respectively.

Algorithm 1: The training process of NTScatNet

Input: $\{x_S^i, y_S^i\}_{i=1}^{N_S}; L, Q_1, Q_2$; Optimizer; Training epochs; Learning rate
Output: The trained NTScatNet model

- 1 **Initialization:** Initializing NTScatNet;
- 2 Calculate $\bar{S}(x_S^i)$ of the source domain according to Eqs. (6)–(9);
- 3 Calculate the normalized wavelet scattering features, i.e., $\tilde{S}(x_S^i)$ according Eqs. (10)–(14);
- 4 **for** Training epochs < Pre-defined epochs **do**
- 5 Calculate the data loss of the source domain according to Eqs. (15) and (16);
- 6 Parameters optimization according to (17);
- 7 **end**
- 8 Output the trained NTScatNet model;

Algorithm 2: The testing process of NTScatNet

Input: Trained NTScatNet; $\{x_T^j\}_{j=1}^{N_T}$;
Output: The predicted label: y_{pre}^j

- 1 Calculate $\bar{S}(x_T^j)$ of the target domain samples according to Eqs. (6)–(9);
- 2 Calculate the normalized wavelet scattering features $\tilde{S}(x_T^j)$ according Eqs. (10)–(14);
- 3 Label prediction according to (18);
- 4 Output the predicted label y_{pre}^j ;

3.4.2. Computational complexity analysis

Let x the input sample, L the dimension of x , $N = L/2$ the length of moving averaging widow, and $s = L/8$ the size of steps. The number of time frame is $(L - N)/s + 1 = 5$. Besides, let Q_1, Q_2 are the number of wavelets per octave of the first filterbank and second filterbank, respectively and F_s the sampling frequency of x . For each moving averaging window, the number of first-order wavelets ψ_{λ_1} is about $Q_1 \log_2 N$, and there are about $Q_1 \log_2 N$ first-order wavelet scattering coefficients to be computed. For each first-order wavelet scale, the number of second-order wavelets ψ_{λ_2} is about $Q_2 \log_2 N$, while the number of non-negligible second-order wavelet scattering

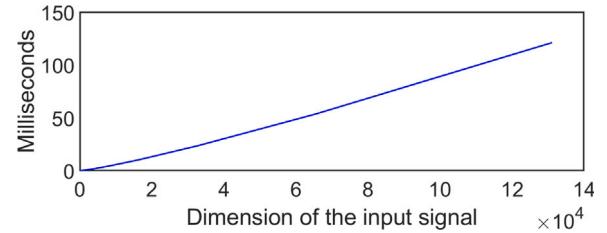


Fig. 12. NTScatNet's calculation time under various dimensions of the input signal.

coefficients is $Q_2 \log_2 N / 2$ since the second-order wavelet scattering coefficients with $\lambda_2 > \lambda_1$ are trivial. In total, the number of wavelet scattering coefficients to be computed is: $5(Q_1 \log_2 N + Q_1 Q_2 (\log_2 N)^2)$. In order to compute the wavelet scattering coefficients, the modulus wavelet coefficients should be first calculated and then averaged with ϕ . The computation complexity of modulus wavelet coefficients is the order of $O(N \log_2 N)$ with a fast Fourier transform(FFT) algorithm. The computation complexity of each wavelet scattering coefficient is also $O(N \log_2 N)$. The number of operations for computing all non-negligible wavelet scattering coefficients is:

$$\text{Complexity} = O\left(5Q_1 N(\log_2 N)^2 + 5 \frac{Q_1 Q_2 N(\log_2 N)^3}{2}\right) \quad (38)$$

We reasonably assume that the rotation speed range of the concerned machine is from 60 rpm/min to 6000 rpm/min, so the period range is from 0.01 s to 1 s. Besides, let $F_s = 12000$ Hz, $Q_1 = 16$, $Q_2 = 4$, each time frame of x contains 4–6 cycles to calculate its scattering coefficients, and the CPU calculates $5 * 10^9$ complex addition operation per second. Fig. 12 plots the calculation time as the function of the dimension of the input signal. As Fig. 12 presents, the wavelet scattering coefficients of a signal with a dimension below 144 000 could be calculated within 0.15 s, demonstrating NTScatNet's computational feasibility from an industrial perspective.

4. Experimental verification

4.1. Introduction to datasets

CWRU datasets are provided by CWRU bearing data center. Fig. 13 presents the experimental platform where the datasets are collected. Although the CWRU dataset is very popular in machine health monitoring, most research concerns motor drive-end bearing's health condition using the acceleration signals measured by the drive-end transducers. This article terms this popular dataset CWRU-DD dataset for convenience. In fact, the bearing data center provided the other three datasets, i.e., CWRU-DF, CWRU-FD, and CWRU-FD datasets.

CWRU-DD dataset concerns the health condition of motor drive-end bearing, the specification of which is SKF-6205. Four bearing health conditions, i.e., Normal(N), ball fault(BF), inner race fault(IF), and outer race fault(OF), are considered. Each bearing fault is machined with a diameter of 7 mils using electrical discharge machining. The transducer is mounted on the drive-end bearing house to monitor the drive-end bearing's health condition. The vibration signals are collected under four working loads(0hp, 1hp, 2hp, and 3hp) at the sampling frequency of 12 kHz. Let $x_D(t)$ the signal excited by drive-end bearing, $h_{DD}(t)$ the transmission path from the drive-end bearing to the drive-end transducer. The signals of CWRU-DD dataset could be presented as $x_D(t) * h_{DD}(t)$. In total, $W_c \times C = 4 \times 4 = 16$ time series are recorded, where W_c is the number of working conditions and C is the number of health conditions. Each recorded signal is segmented into 100 samples of length L with an overlapping trick. The CWRU-DD dataset totally contains 4 sub-datasets, i.e., DD-0hp, DD-1hp, DD-2hp, and DD-3hp, where each sub-dataset contains $100 \times 4 = 400$ samples.

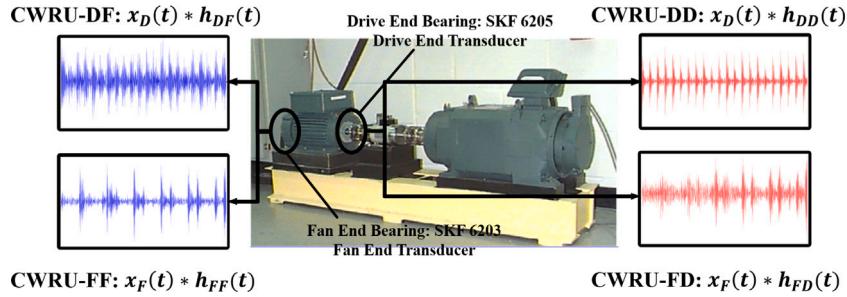


Fig. 13. CWRU experimental platform.

Table 2

Information of CWRU-DD, CWRU-DF, CWRU-FF, and CWRU-FD datasets.

Datasets	CWRU-DD dataset	CWRU-DF dataset	CWRU-FF dataset	CWRU-FD dataset
Bearing position	Motor drive-end	Motor drive-end	Motor fan-end	Motor fan-end
Specification	SKF 6205	SKF 6205	SKF 6203	SKF 6203
Transducer position	Motor drive-end	Motor fan-end	Motor fan-end	Motor drive-end
Working loads	0/1/2/3 hp	0/1/2/3 hp	0/1/2/3 hp	0/1/2/3 hp
Fault categories	N/BF/IF/OF	N/BF/IF/OF	N/BF/IF/OF	N/BF/IF/OF
Sample dimension	L	L	L	L
Dataset size	$4 \times 4 \times 100$			
Signal symbol	$x_D(t) * h_{DD}(t)$	$x_D(t) * h_{DF}(t)$	$x_F(t) * h_{FF}(t)$	$x_F(t) * h_{FD}(t)$
Sub-datasets	DD-0hp/DD-1hp/DD-2hp/DD-3hp	DF-0hp/DF-1hp/DF-2hp/DF-3hp	FF-0hp/FF-1hp/FF-2hp/FF-3hp	FD-0hp/FD-1hp/FD-2hp/FD-3hp

CWRU-DF dataset concerns drive-end bearing's health condition using signals observed by the fan-end transducer. The vibration signals are collected under four working loads(0hp, 1hp, 2hp, and 3hp) at a sampling frequency of 12 kHz. Let $h_{DF}(t)$ the transmission path from the drive-end bearing to the fan-end transducer. The signals of CWRU-DF dataset could be presented as $x_D(t) * h_{DF}(t)$. Each recorded signal is segmented into 100 samples of length L with an overlapping trick. The CWRU-DF dataset totally contains 4 sub-datasets, i.e., DF-0hp, DF-1hp, DF-2hp, and DF-3hp, where each sub-dataset contains $100 \times 4 = 400$ samples.

CWRU-FF dataset concerns the health condition of motor fan-end bearing, the specification of which is SKF-6203. Four bearing health conditions, i.e., Normal(N), ball fault(BF), inner race fault(IF), and outer race fault(OF), are considered. Each bearing fault is machined with a diameter of 7 mils using electrical discharge machining. The transducer is mounted on the fan-end bearing house to monitor the fan-end bearing's health condition. The vibration signals are collected under four working loads(0hp, 1hp, 2hp, and 3hp) at the sampling frequency of 12 kHz. Let $x_F(t)$ the signal excited by fan-end bearing, $h_{FF}(t)$ the transmission path from the fan-end bearing to the fan-end transducer. The signals of CWRU-FF dataset could be presented as $x_F(t) * h_{FF}(t)$. Each recorded signal is segmented into 100 samples of length L with an overlapping trick. The CWRU-FF dataset totally contains 4 sub-datasets,i.e., FF-0hp, FF-1hp, FF-2hp, and FF-3hp, where each sub-dataset contains $100 \times 4 = 400$ samples.

CWRU-FD dataset concerns fan-end bearing's health condition using signals observed by the drive-end transducer. The vibration signals are collected under four working loads(0hp, 1hp, 2hp, and 3hp) at a sampling frequency of 12 kHz. Let $h_{FD}(t)$ the transmission path from the fan-end bearing to the drive-end transducer. The signals of CWRU-FD dataset could be presented as $x_F(t) * h_{FD}(t)$. Each recorded signal is segmented into 100 samples of length L with an overlapping trick. The CWRU-FD dataset totally contains 4 sub-datasets, i.e., FD-0hp, FD-1hp, FD-2hp, and FD-3hp, where each sub-dataset contains $100 \times 4 = 400$ samples. Specific information about the above four bearing datasets are summarized in Table 2.

4.2. Feature visualization

This subsection aims to verify that NTScatNet could signification reduce data discrepancy due to transmission path variability by visualizing the actual shape of normalized scattering features.

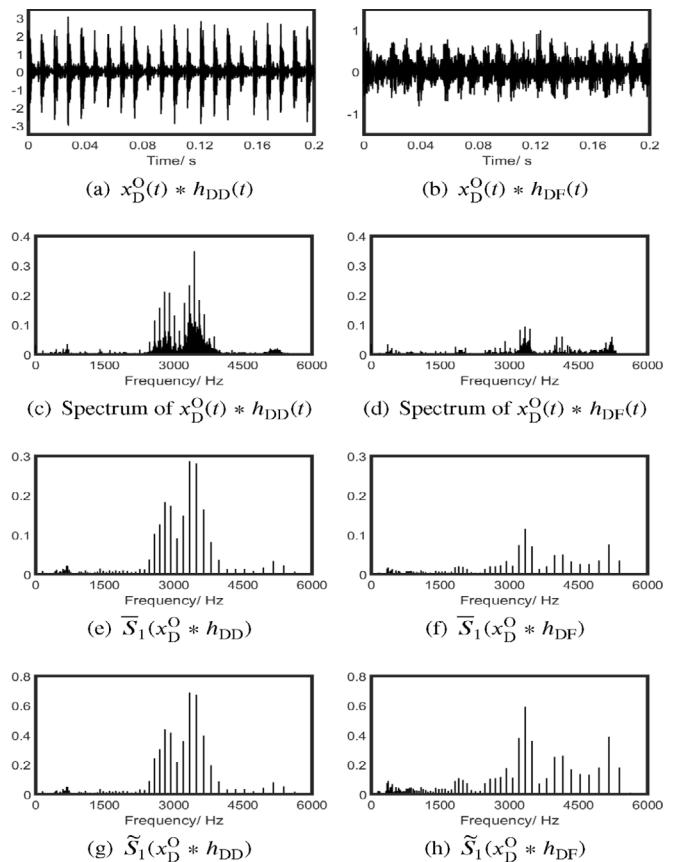


Fig. 14. Visualization of drive-end bearing signal's features.

4.2.1. Feature visualization of drive-end bearing signal

The first case will visualize features of the drive-end bearing with an outer race defect. The experimental working load is 0hp, and two transducers are separately mounted on the drive-end bearing house

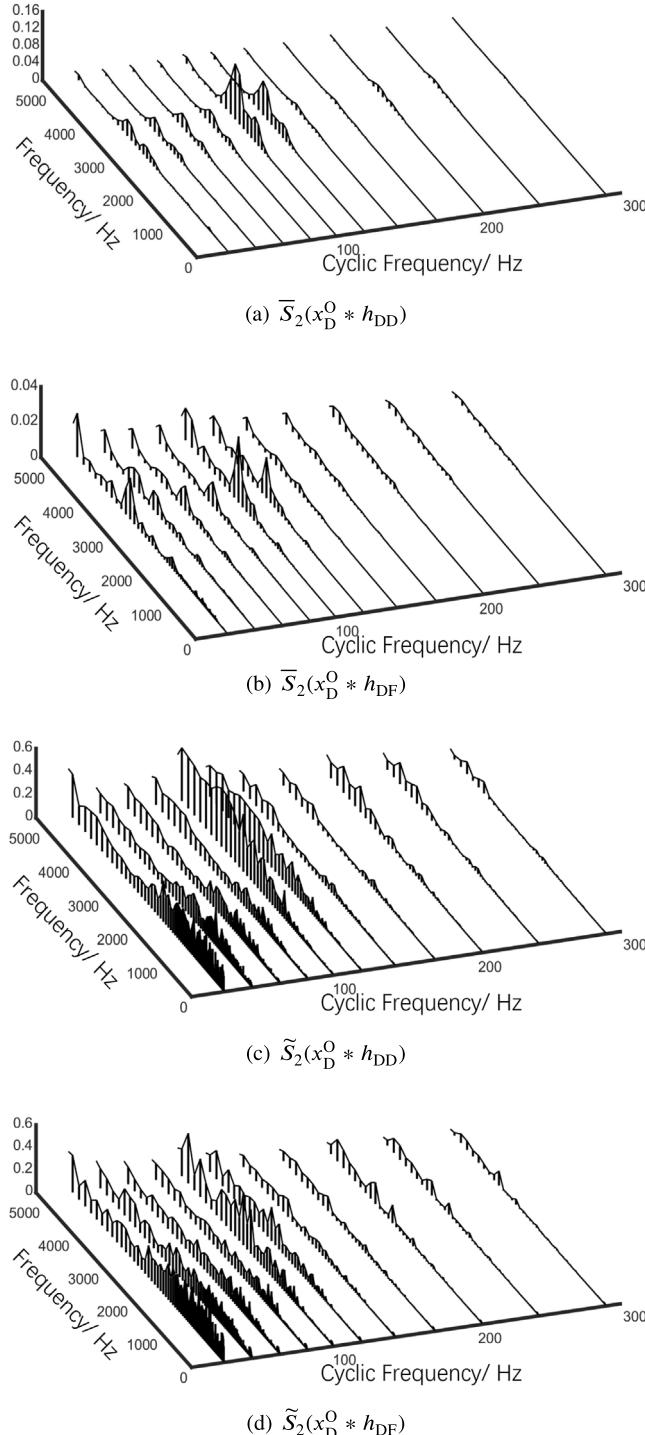


Fig. 15. Visualization of drive-end bearing signal's features.

and the fan-end bearing house to monitor the bearing's health condition. Fig. 14(a)–14(b) present the time domain signals observed by the drive-end transducer and the fan-end transducer, respectively, i.e., $x_D^0(t) * h_{DD}(t)$ and $x_D^0(t) * h_{DF}(t)$. The amplitude of $x_D^0(t) * h_{DD}(t)$ is significantly larger than that of $x_D^0(t) * h_{DF}(t)$ since the drive-end transducer is near the drive-end fault bearing while the fan-end transducer is remote from the drive-end bearing. Fig. 14(c)–14(d) show the amplitude spectrum of $x_D^0(t) * h_{DD}(t)$ and $x_D^0(t) * h_{DF}(t)$, respectively. The amplitude spectrum of $x_D^0(t) * h_{DD}(t)$ and $x_D^0(t) * h_{DF}(t)$ present different formants and amplitude, illustrating that Fourier transform

fails to remove the discrepancy due to transmission path variability. As illustrated in Section 3.2, the globally averaged first-order wavelet scattering feature approximates the Fourier spectrum. Therefore, it fails to remove the discrepancy due to transmission path variability, as Fig. 14(e)–14(f) present. By contrast, the normalized first-order scattering feature effectively reduced amplitude discrepancy, as shown in Fig. 14(g)–14(h).

As Section 3.2 implies, the globally averaged second-order wavelet scattering feature well characterizes the multiscale cyclostationarity of the fault signals. However, the cyclostationarity representation based on the globally averaged second-order wavelet scattering feature is sensitive to the energy of different wavelet scales, as Figs. 15(a) and 15(b) present. By contrast, the cyclostationarity representation based on the normalized second-order wavelet scattering feature become insensitive to the energy of different wavelet scales, as Figs. 15(c) and 15(d) illustrate. Moreover, Eqs. (39) and (40) calculated the normalized Euclidean distance as a metric to quantify transmission path variability in the scattering space and the normalized scattering space, respectively.

$$\frac{\|\bar{S}_2(x_D^0 * h_{DD}) - \bar{S}_2(x_D^0 * h_{DF})\|_2}{\|\bar{S}_2(x_D^0 * h_{DD})\|_2 + \|\bar{S}_2(x_D^0 * h_{DF})\|_2} = 0.56 \quad (39)$$

$$\frac{\|\tilde{S}_2(x_D^0 * h_{DD}) - \tilde{S}_2(x_D^0 * h_{DF})\|_2}{\|\tilde{S}_2(x_D^0 * h_{DD})\|_2 + \|\tilde{S}_2(x_D^0 * h_{DF})\|_2} = 0.22 \quad (40)$$

Comparing Eqs. (39) and (40) reveals that the normalized second-order wavelet scattering feature becomes insensitive to transmission path variability.

4.2.2. Visualization of fan-end bearing signal's features

The second case will visualize features of the fan-end bearing with an inner race defect. The experimental working load is 3hp, and two transducers are separately mounted on the drive-end bearing house and the fan-end bearing house to monitor the bearing's health condition.

Fig. 16(a)–16(b) present the time domain signals observed by the drive-end transducer and the fan-end transducer, respectively, from which one can find that the amplitude of $x_F^I(t) * h_{FF}(t)$ is significantly larger than that of $x_F^I(t) * h_{FD}(t)$. The amplitude spectrum of $x_F^I(t) * h_{FF}(t)$ and $x_F^I(t) * h_{FD}(t)$ are shown in Fig. 16(c) and Fig. 16(d), respectively. Besides, Fig. 16(e)–16(f) present the globally averaged first-order wavelet scattering features. However, both the Fourier spectrum and the globally averaged wavelet scattering features present significant discrepancies in amplitudes and formants due to different transmission path. In comparison, the normalized first-order scattering feature effectively reduced amplitude discrepancy, as shown in Fig. 16(g)–16(h).

Figs. 17(a) and 17(b) present $\bar{S}_2(x_F^I * h_{FF})$ and $\bar{S}_2(x_F^I * h_{FD})$, respectively, from which one can find a significant discrepancy in formant and amplitudes. By contrast, the cyclostationarity representation based on the normalized second-order wavelet scattering feature becomes insensitive to different transmission paths, as illustrated in Figs. 17(c) and 17(d). In addition, Eqs. (41) and (42) take the normalized Euclidean distance as a metric to quantify transmission path variability in the scattering space and the normalized scattering space.

$$\frac{\|\bar{S}_2(x_F^I * h_{FF}) - \bar{S}_2(x_F^I * h_{FD})\|_2}{\|\bar{S}_2(x_F^I * h_{FF})\|_2 + \|\bar{S}_2(x_F^I * h_{FD})\|_2} = 0.58 \quad (41)$$

$$\frac{\|\tilde{S}_2(x_F^I * h_{DD}) - \tilde{S}_2(x_F^I * h_{DF})\|_2}{\|\tilde{S}_2(x_F^I * h_{DD})\|_2 + \|\tilde{S}_2(x_F^I * h_{DF})\|_2} = 0.21 \quad (42)$$

Comparing Eqs. (41) and (42) further verifies that the normalized second-order wavelet scattering feature becomes less sensitive to transmission path variability.

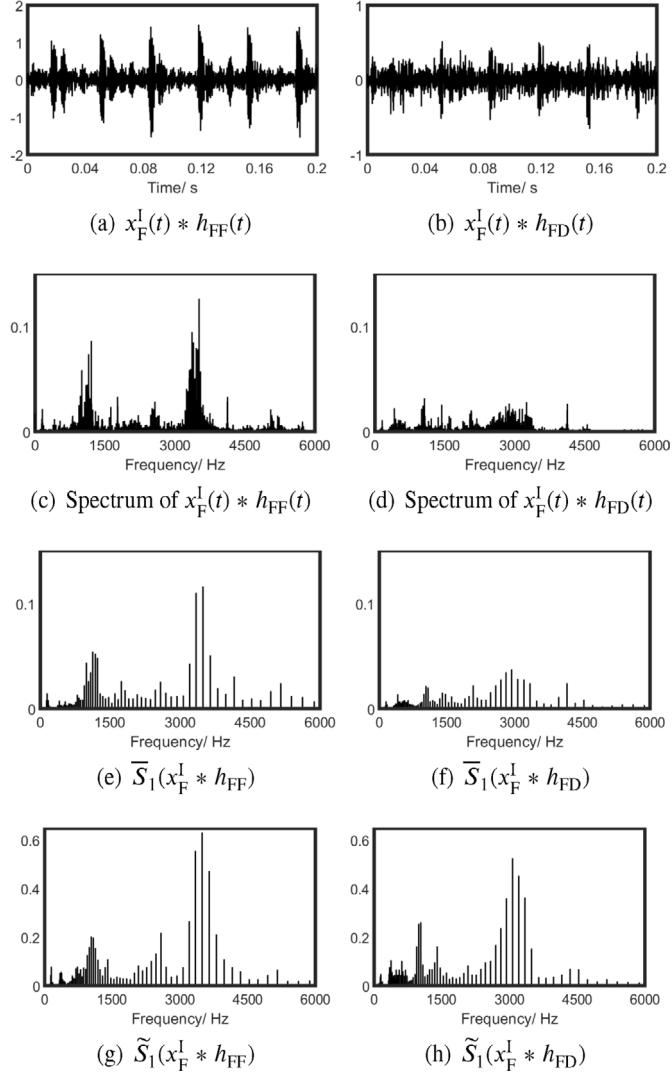


Fig. 16. Visualization of fan-end bearing signal's features.

4.3. Case study 1: transfer diagnosis across transducers at different positions

4.3.1. Transfer diagnosis tasks

The current experiments involve four datasets, *i.e.*, CWRU-DD, CWRU-DF, CWRU-FF, and CWRU-FD, each containing data collected from four working loads(0/1/2/3 hp). As presented in Table 3, 64 transfer diagnosis tasks are constructed based on the above datasets to verify NTScatNet's domain generalization capability across transmission paths. For example, the meaning of task A05 is learning diagnosis experience from the DD-0hp subdataset and applying the learned diagnosis knowledge to the DF-1hp subdataset. The other tasks follow similar conventions. Before introducing the compared methods, we would like to distinguish between the domain adaptation and generalization settings. The domain adaptation setting generally requires semi-supervised or unsupervised target domain data are available during the training phase. In contrast, the domain generalization setting does not require target domain information before the testing phase.

4.3.2. Compared methods

In order to illustrate NTScatNet's technique advantages, the Fourier spectrum methods, TScatNet model, and three recently published deep

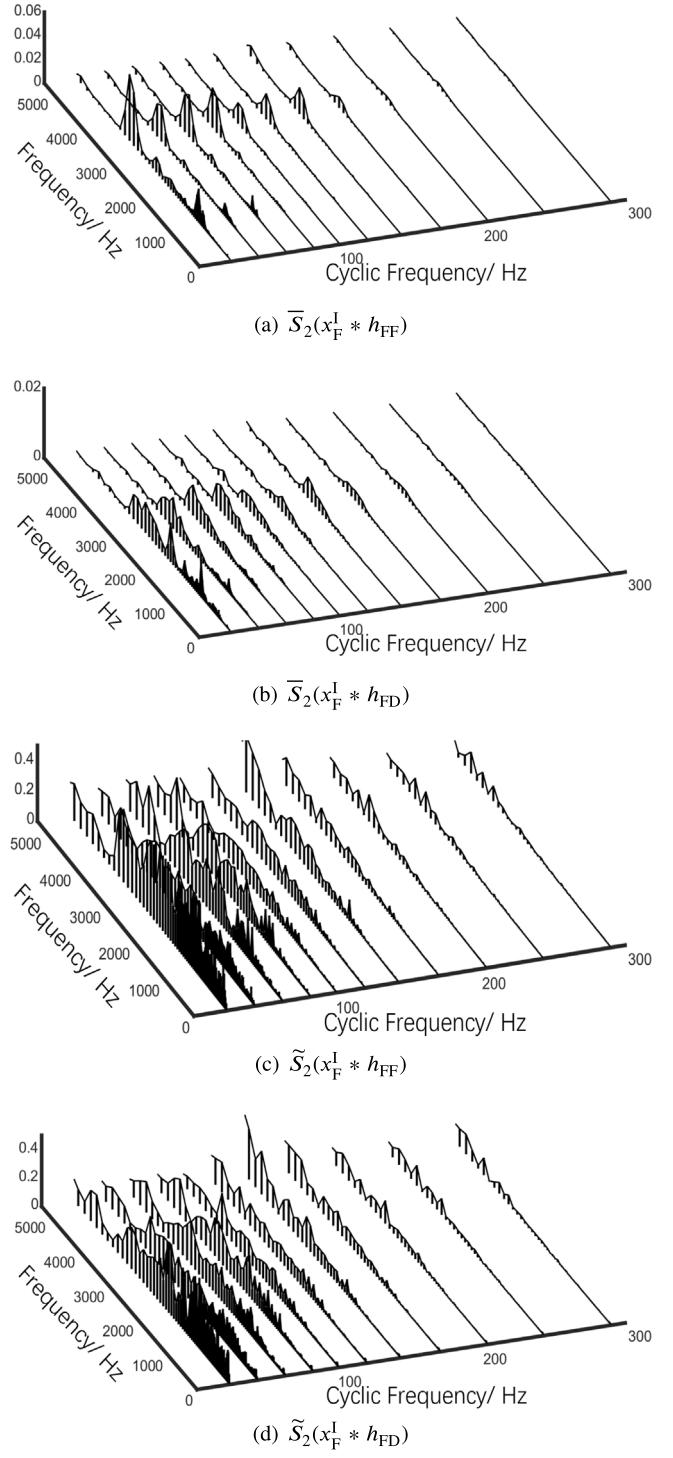


Fig. 17. Visualization of fan-end bearing signal's features.

transfer learning models are also evaluated on the above 64 transfer diagnosis tasks. The compared methods are described below.

(1) *Spectrum + Softmax*: This method first uses a L point fast Fourier transform algorithm to calculate the frequency spectrum of the time-domain signals and then trains a Softmax classifier with the spectrum of the source domain data. The method follows the domain generalization setting.

(2) *WD-DTL* [22]: WD-DTL takes the Fourier spectrum as input and follows the domain adaptation setting. Specifically, it requires labeled source domain data and unlabeled target domain data during the

Table 3

Illustration of the constructed transfer diagnosis tasks.

Index	Source→Target	Index	Source→Target	Index	Source→Target	Index	Source→Target
A01	DD-0hp→DF-0hp	B01	DF-0hp→DD-0hp	C01	FF-0hp→FD-0hp	D01	FD-0hp→FF-0hp
A02	DD-1hp→DF-1hp	B02	DF-1hp→DD-1hp	C02	FF-1hp→FD-1hp	D02	FD-1hp→FF-1hp
A03	DD-2hp→DF-2hp	B03	DF-2hp→DD-2hp	C03	FF-2hp→FD-2hp	D03	FD-2hp→FF-2hp
A04	DD-3hp→DF-3hp	B04	DF-3hp→DD-3hp	C04	FF-3hp→FD-3hp	D04	FD-3hp→FF-3hp
A05	DD-0hp→DF-1hp	B05	DF-0hp→DD-1hp	C05	FF-0hp→FD-1hp	D05	FD-0hp→FF-1hp
A06	DD-0hp→DF-2hp	B06	DF-0hp→DD-2hp	C06	FF-0hp→FD-2hp	D06	FD-0hp→FF-2hp
A07	DD-0hp→DF-3hp	B07	DF-0hp→DD-3hp	C07	FF-0hp→FD-3hp	D07	FD-0hp→FF-3hp
A08	DD-1hp→DF-0hp	B08	DF-1hp→DD-0hp	C08	FF-1hp→FD-0hp	D08	FD-1hp→FF-0hp
A09	DD-1hp→DF-2hp	B09	DF-1hp→DD-2hp	C09	FF-1hp→FD-2hp	D09	FD-1hp→FF-2hp
A10	DD-1hp→DF-3hp	B10	DF-1hp→DD-3hp	C10	FF-1hp→FD-3hp	D10	FD-1hp→FF-3hp
A11	DD-2hp→DF-0hp	B11	DF-2hp→DD-0hp	C11	FF-2hp→FD-0hp	D11	FD-2hp→FF-0hp
A12	DD-2hp→DF-1hp	B12	DF-2hp→DD-1hp	C12	FF-2hp→FD-1hp	D12	FD-2hp→FF-1hp
A13	DD-2hp→DF-3hp	B13	DF-2hp→DD-3hp	C13	FF-2hp→FD-3hp	D13	FD-2hp→FF-3hp
A14	DD-3hp→DF-0hp	B14	DF-3hp→DD-0hp	C14	FF-3hp→FD-0hp	D14	FD-3hp→FF-0hp
A15	DD-3hp→DF-1hp	B15	DF-3hp→DD-1hp	C15	FF-3hp→FD-1hp	D15	FD-3hp→FF-1hp
A16	DD-3hp→DF-2hp	B16	DF-3hp→DD-2hp	C16	FF-3hp→FD-2hp	D16	FD-3hp→FF-2hp

Table 4

Hyperparameters of NTScatNet.

Q_1	Q_2	L	Optimizer	Learning rate	Weight decay	Batch size	Epoch
16	4	2400	Adam	0.001	0.0001	16	500

training phase. The architecture, hyperparameters, and implementation detail are consistent with the original paper [22].

(3) *GANPair-N* [21]: This method takes the Fourier spectrum as input and obeys the domain adaptation setting. GANPair-N requires labeled source domain data and unlabeled target domain data during the training phase. Besides, parallel data in the normal state of both source and the target domain are also required. The architecture, hyperparameters, and implementation detail are consistent with the original paper [21].

(4) *GANPair-all* [21]: The architecture, hyperparameters, and implementation detail are identical to GANPair-N. However, the GANPair-all method requires labeled source domain data, unlabeled target domain data, and parallel data in all possible health conditions of the source and the target domain.

(5) *TScatNet* [6]: TScatNet takes raw time-domain signal as input and follows the domain generalization setting. The hyperparameters adopted in the current experiment are given in Table 4.

4.3.3. Results of the transfer diagnosis experiments

Table 5 reports the six methods' diagnosis accuracies on 64 transfer diagnosis tasks, where each experiment is repeated ten times to reduce the influence of the random factors.

In order to compare the diagnosis performance of six methods intuitively, Fig. 18 presents the experimental results in the box plot. Each box in Fig. 18 involves 64 transfer diagnosis results, comprehensively characterizing a method's transfer diagnosis performance across transducers at different positions. As shown in the first box in Fig. 18, the Fourier spectrum method realizes an average accuracy of 52.5% on 64 transfer diagnosis tasks, indicating a severe data distribution discrepancy between the source and the target domain. Ref. [6] shows that TScatNet possesses excellent domain generalization capability across working load variation. However, as shown in the fifth box in Fig. 18, the average diagnosis accuracy of TScatNet is less than 82%, illustrating that TScatNet fails to address the domain generalization tasks across transducers. As shown in the sixth box in Fig. 18, the proposed NTScatNet achieves an average accuracy of 98.4% on 64 transfer diagnosis tasks. It should be worth noting that the source and target domain samples in tasks A01–A04, B01–B04, C01–C04, and D01–D04, are sampled from different transducers under identical working

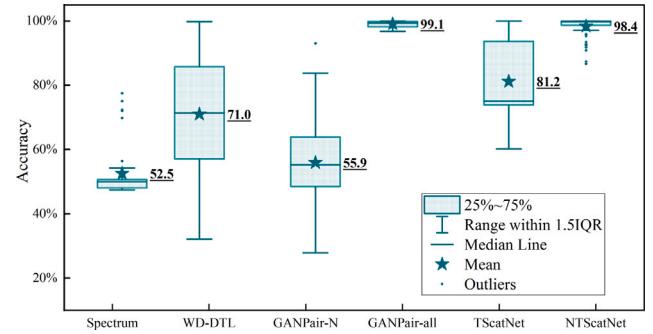


Fig. 18. Results of the transfer diagnosis experiments.

loads. By contrast, the source and target domain samples in tasks A05–A16, B05–B16, C05–C16, and D05–D16, are sampled from different transducers under different working loads. NTScatNet's excellent performance on 64 transfer diagnosis tasks illustrates that the proposed NTScatNet holds the domain generalization capability across different transducers and working loads simultaneously.

The second and the third boxes in Fig. 18 present the transfer diagnosis accuracies of WD-DTL and GANPair-N, respectively. WD-DTL achieves an average accuracy of 71% on 64 transfer diagnosis tasks. The best performance of WD-DTL is 99.8% (on task A04), while the worst accuracy is 50.7% (on task A13). GANPair-N achieves an average accuracy of 55.9% on 64 transfer diagnosis tasks. The best performance of GANPair-N is $93.0 \pm 11.9\%$ (on task C04), while the worst accuracy is only $27.9 \pm 6.3\%$ (accuracy on task B04). The above experiment results show that WD-DTL and GANPair-N fail to realize stable transfer diagnosis performance. As the fourth box in Fig. 18 presents, the GANPair-all achieves an average accuracy of 99.1% on 64 transfer diagnosis tasks, slightly outperforming NTScatNet by 0.7 percent. It should be noted that GANPair-all requires labeled source domain data, unlabeled target domain data, and parallel data of all source and target domain classes during the training phase. However, in practice, the target domain data is hard to obtain during the training phase. In contrast, the developed NTScatNet follows the domain generalization setting and only requires the labeled source domain data during the training phase, holding better practical application prospects than GANPair-all.

4.3.4. Results of unsupervised manifold embedding

Manifold embedding aims to recover the low-dimensional manifold structure from the high-dimensional data, where t-distribution stochastic neighbor embedding (t-SNE) is one of the most effective algorithms. This sub-section embeds wavelet scattering features and normalized

Table 5
Results of the transfer diagnosis experiments.

Index	Spectrum	WD-DTL	GANPair-N	GANPair-All	TScatNet	NTScatNet
A01	77.5 ± 7.8%	76.4 ± 23.7%	49.6 ± 23.6%	99.8 ± 0.3%	88.5 ± 6.5%	99.8 ± 0.1%
A02	47.5 ± 7.9%	86.3 ± 19.9%	54.8 ± 31.5%	99.0 ± 1.9%	83.2 ± 4.2%	100 ± 0.1%
A03	50.0 ± 0.1%	98.8 ± 2.0%	66.2 ± 30.1%	98.5 ± 1.5%	77.4 ± 2.2%	99.7 ± 0.1%
A04	50.0 ± 0.1%	99.8 ± 0.5%	57.5 ± 33.4%	100 ± 0.0%	79.2 ± 2.5%	100 ± 0.1%
A05	47.5 ± 7.9%	64.4 ± 16.8%	41.9 ± 11.8%	98.8 ± 2.8%	82.2 ± 4.9%	100 ± 0.1%
A06	47.5 ± 7.9%	54.3 ± 9.9%	40.0 ± 12.9%	99.6 ± 1.0%	73.5 ± 3.2%	99.7 ± 0.1%
A07	47.5 ± 7.9%	59.8 ± 12.2%	54.8 ± 28.4%	99.7 ± 0.8%	73.0 ± 4.3%	98.2 ± 5.9%
A08	77.5 ± 7.8%	53.4 ± 6.3%	64.8 ± 31.6%	99.7 ± 0.6%	90.5 ± 4.9%	99.2 ± 0.2%
A09	47.5 ± 7.9%	53.0 ± 18.4%	49.8 ± 22.7%	99.3 ± 1.1%	74.0 ± 3.5%	99.4 ± 0.2%
A10	47.5 ± 7.9%	68.5 ± 25.0%	52.0 ± 28.8%	99.9 ± 0.4%	73.5 ± 4.5%	99.1 ± 0.3%
A11	75.0 ± 0.1%	67.1 ± 20.9%	49.4 ± 23.7%	99.4 ± 1.5%	92.2 ± 6.4%	99.8 ± 0.1%
A12	50.0 ± 0.1%	57.1 ± 15.8%	54.1 ± 28.4%	98.2 ± 3.4%	85.9 ± 6.6%	100 ± 0.1%
A13	75.0 ± 0.1%	50.7 ± 18.4%	55.7 ± 21.4%	100 ± 0.0%	76.8 ± 2.1%	100 ± 0.1%
A14	50.0 ± 0.1%	57.4 ± 11.3%	42.3 ± 26.4%	99.4 ± 1.2%	93.0 ± 6.5%	99.7 ± 0.1%
A15	50.0 ± 0.1%	57.3 ± 15.3%	55.9 ± 29.2%	99.7 ± 0.6%	87.5 ± 7.5%	100 ± 0.1%
A16	50.0 ± 0.1%	64.7 ± 21.5%	50.7 ± 29.0%	98.0 ± 3.2%	80.3 ± 3.9%	99.7 ± 0.1%
B01	52.3 ± 8.0%	77.9 ± 15.7%	51.3 ± 18.7%	97.6 ± 1.2%	98.8 ± 2.2%	100 ± 0.1%
B02	72.3 ± 0.9%	95.9 ± 4.5%	47.3 ± 14.5%	97.4 ± 3.0%	99.9 ± 0.1%	100 ± 0.1%
B03	48.5 ± 1.7%	99.1 ± 1.3%	49.8 ± 25.6%	98.2 ± 0.8%	100 ± 0.1%	99.9 ± 0.2%
B04	54.3 ± 0.1%	97.9 ± 2.3%	27.9 ± 6.3%	97.0 ± 2.9%	99.7 ± 0.2%	100 ± 0.1%
B05	69.8 ± 7.0%	70.3 ± 13.6%	46.8 ± 25.0%	98.1 ± 1.2%	99.7 ± 1.0%	98.0 ± 1.1%
B06	48.2 ± 0.6%	85.2 ± 15.6%	58.7 ± 27.9%	96.8 ± 2.7%	99.9 ± 0.2%	99.1 ± 1.6%
B07	51.3 ± 0.9%	66.8 ± 12.1%	56.2 ± 20.3%	99.0 ± 1.3%	95.5 ± 3.9%	99.7 ± 1.0%
B08	52.3 ± 8.0%	74.1 ± 15.8%	40.0 ± 12.9%	97.9 ± 1.6%	97.4 ± 4.6%	100 ± 0.1%
B09	50.7 ± 0.9%	78.7 ± 16.7%	57.4 ± 25.9%	98.1 ± 2.2%	99.9 ± 0.4%	100 ± 0.1%
B10	56.3 ± 6.6%	73.2 ± 14.0%	50.6 ± 16.4%	98.9 ± 2.7%	94.4 ± 5.5%	100 ± 0.1%
B11	49.8 ± 0.1%	71.4 ± 13.7%	49.8 ± 23.9%	97.8 ± 1.0%	98.3 ± 2.1%	100 ± 0.1%
B12	69.8 ± 7.0%	71.5 ± 16.8%	45.0 ± 21.6%	98.3 ± 1.0%	100 ± 0.1%	99.6 ± 0.3%
B13	53.8 ± 1.3%	78.2 ± 18.5%	57.2 ± 24.4%	97.6 ± 1.2%	95.3 ± 2.2%	100 ± 0.1%
B14	49.8 ± 0.1%	64.1 ± 20.0%	31.5 ± 16.4%	97.5 ± 1.4%	100 ± 0.1%	100 ± 0.1%
B15	72.0 ± 0.1%	70.5 ± 12.0%	41.3 ± 24.4%	97.7 ± 2.5%	100 ± 0.1%	99.9 ± 0.2%
B16	48.0 ± 0.1%	73.3 ± 13.2%	42.3 ± 15.9%	98.5 ± 1.4%	100 ± 0.1%	100 ± 0.1%
C01	49.5 ± 10.6%	45.9 ± 8.4%	65.7 ± 24.4%	99.9 ± 0.1%	74.7 ± 0.6%	95.9 ± 2.5%
C02	47.5 ± 7.9%	97.1 ± 6.4%	79.4 ± 16.1%	99.9 ± 0.4%	75.0 ± 0.1%	93.4 ± 0.7%
C03	47.5 ± 7.9%	99.8 ± 0.2%	80.0 ± 19.9%	100 ± 0.0%	75.0 ± 0.1%	95.8 ± 1.0%
C04	50.0 ± 0.1%	96.6 ± 3.6%	93.0 ± 11.9%	97.7 ± 0.4%	75.0 ± 0.1%	92.7 ± 1.1%
C05	47.5 ± 7.9%	54.2 ± 16.5%	51.7 ± 16.1%	99.4 ± 0.8%	74.9 ± 0.3%	92.5 ± 0.4%
C06	47.5 ± 7.9%	72.1 ± 14.8%	56.3 ± 28.6%	100 ± 0.0%	74.8 ± 0.6%	90.9 ± 1.1%
C07	47.5 ± 7.9%	46.9 ± 14.3%	48.2 ± 23.8%	99.6 ± 0.4%	74.1 ± 2.9%	86.7 ± 0.9%
C08	49.5 ± 10.6%	40.7 ± 8.6%	83.7 ± 16.5%	100 ± 1.1%	75.0 ± 0.1%	99.1 ± 0.2%
C09	47.5 ± 7.9%	71.2 ± 25.6%	83.3 ± 11.7%	100 ± 0.1%	75.0 ± 0.1%	91.8 ± 0.9%
C10	47.5 ± 7.9%	86.5 ± 17.1%	58.3 ± 19.5%	100 ± 0.1%	74.5 ± 1.7%	87.4 ± 1.0%
C11	49.5 ± 10.6%	74.6 ± 20.3%	72.6 ± 11.7%	100 ± 0.1%	74.8 ± 0.1%	100 ± 0.1%
C12	47.5 ± 7.9%	74.5 ± 23.3%	70.8 ± 20.1%	99.4 ± 1.7%	75.0 ± 0.1%	97.4 ± 0.9%
C13	47.5 ± 7.9%	78.1 ± 23.4%	65.2 ± 23.9%	100 ± 0.1%	74.9 ± 0.2%	92.8 ± 1.4%
C14	50.0 ± 0.1%	57.1 ± 26.5%	67.8 ± 24.6%	100 ± 0.0%	74.1 ± 0.6%	99.8 ± 0.1%
C15	50.0 ± 0.1%	91.9 ± 14.9%	43.2 ± 12.8%	100 ± 0.0%	75.1 ± 0.1%	97.1 ± 0.8%
C16	50.0 ± 0.1%	73.9 ± 26.0%	66.0 ± 29.6%	100 ± 0.1%	75.0 ± 0.1%	95.3 ± 0.6%
D01	50.2 ± 0.6%	48.6 ± 12.6%	59.2 ± 10.3%	99.5 ± 1.0%	73.8 ± 10.5%	100 ± 0.1%
D02	50.0 ± 0.1%	97.5 ± 3.9%	50.3 ± 6.6%	99.4 ± 0.5%	77.3 ± 14.5%	100 ± 0.1%
D03	50.0 ± 0.1%	99.3 ± 0.6%	55.6 ± 15.3%	99.5 ± 1.0%	60.2 ± 9.3%	100 ± 0.1%
D04	50.1 ± 0.4%	96.2 ± 7.8%	60.1 ± 16.3%	100 ± 0.0%	71.3 ± 6.0%	100 ± 0.1%
D05	48.5 ± 4.8%	32.1 ± 8.3%	67.9 ± 9.7%	99.3 ± 0.3%	73.7 ± 10.6%	99.4 ± 1.8%
D06	48.4 ± 5.2%	50.1 ± 16.0%	65.4 ± 16.4%	99.4 ± 1.0%	68.8 ± 11.6%	99.8 ± 0.6%
D07	50.7 ± 2.2%	41.9 ± 13.6%	60.6 ± 25.1%	99.0 ± 1.2%	69.8 ± 11.8%	98.5 ± 0.5%
D08	50.0 ± 0.1%	48.7 ± 16.9%	39.2 ± 11.2%	99.7 ± 0.7%	76.2 ± 12.3%	99.9 ± 0.1%
D09	50.0 ± 0.1%	54.5 ± 28.8%	57.4 ± 23.2%	100 ± 0.0%	61.9 ± 12.7%	100 ± 0.1%
D10	50.0 ± 0.1%	86.5 ± 18.7%	48.9 ± 8.8%	99.3 ± 1.2%	64.9 ± 9.3%	100 ± 0.1%
D11	50.0 ± 0.1%	63.0 ± 15.1%	47.7 ± 15.2%	98.1 ± 2.3%	67.2 ± 11.2%	99.0 ± 0.5%
D12	50.0 ± 0.1%	63.2 ± 21.0%	58.6 ± 15.3%	99.7 ± 0.5%	66.0 ± 12.3%	100 ± 0.1%
D13	50.0 ± 0.1%	77.4 ± 19.8%	45.3 ± 14.2%	98.2 ± 2.2%	66.0 ± 8.7%	100 ± 0.1%
D14	49.4 ± 1.8%	43.4 ± 16.1%	54.7 ± 20.1%	99.9 ± 0.1%	74.1 ± 9.4%	99.8 ± 0.1%
D15	49.9 ± 0.5%	95.2 ± 2.1%	68.1 ± 16.3%	99.4 ± 0.1%	71.1 ± 9.1%	100 ± 0.1%
D16	50.6 ± 1.8%	63.3 ± 31.0%	63.0 ± 22.2%	99.5 ± 1.4%	66.9 ± 9.2%	100 ± 0.1%

scattering features of the CWRU dataset into two-dimensional space to intuitively verify that normalized scattering features' stability against transmission path variability.

Fig. 19 presents the t-SNE results of the CWRU-DD and CWRU-DF datasets. As Fig. 19(a) shows, scattering features from different working conditions are gathered together while features of different categories

are separated, illustrating that wavelet scattering features are robust against working load variation while preserving discriminative to different categories. However, wavelet scattering features collected by transducers at different locations are not compact, meaning wavelet scattering fails to remove data distribution discrepancy brought by transmission path variability. In comparison, as Fig. 19(b) presents,

Table 6
NTScatNet's accuracies under 24 hyperparameter settings.

Index	Hyperparameters	Accuracy	Index	Hyperparameters	Accuracy
H01	$L = 1200, Q_1 = 8, Q_2 = 1$	$90.3 \pm 8.7\%$	H13	$L = 3600, Q_1 = 8, Q_2 = 1$	$94 \pm 8.2\%$
H02	$L = 1200, Q_1 = 8, Q_2 = 2$	$88.7 \pm 8.4\%$	H14	$L = 3600, Q_1 = 8, Q_2 = 2$	$96.1 \pm 6.6\%$
H03	$L = 1200, Q_1 = 8, Q_2 = 4$	$90.1 \pm 8.5\%$	H15	$L = 3600, Q_1 = 8, Q_2 = 4$	$99 \pm 2.4\%$
H04	$L = 1200, Q_1 = 16, Q_2 = 1$	$89.9 \pm 8.5\%$	H16	$L = 3600, Q_1 = 16, Q_2 = 1$	$97.1 \pm 5.0\%$
H05	$L = 1200, Q_1 = 16, Q_2 = 2$	$85.3 \pm 10.6\%$	H17	$L = 3600, Q_1 = 16, Q_2 = 2$	$98.7 \pm 2.9\%$
H06	$L = 1200, Q_1 = 16, Q_2 = 4$	$88.5 \pm 9.9\%$	H18	$L = 3600, Q_1 = 16, Q_2 = 4$	$98.6 \pm 3.4\%$
H07	$L = 2400, Q_1 = 8, Q_2 = 1$	$92.1 \pm 8.1\%$	H19	$L = 4800, Q_1 = 8, Q_2 = 1$	$94.9 \pm 8.1\%$
H08	$L = 2400, Q_1 = 8, Q_2 = 2$	$94.8 \pm 7.1\%$	H20	$L = 4800, Q_1 = 8, Q_2 = 2$	$96.8 \pm 6.2\%$
H09	$L = 2400, Q_1 = 8, Q_2 = 4$	$96.5 \pm 5.4\%$	H21	$L = 4800, Q_1 = 8, Q_2 = 4$	$99.6 \pm 1.8\%$
H10	$L = 2400, Q_1 = 16, Q_2 = 1$	$96.0 \pm 5.9\%$	H22	$L = 4800, Q_1 = 16, Q_2 = 1$	$98.1 \pm 4.1\%$
H11	$L = 2400, Q_1 = 16, Q_2 = 2$	$95.7 \pm 6.0\%$	H23	$L = 4800, Q_1 = 16, Q_2 = 2$	$99.4 \pm 2.0\%$
H12	$L = 2400, Q_1 = 16, Q_2 = 4$	$98.4 \pm 3.3\%$	H24	$L = 4800, Q_1 = 16, Q_2 = 4$	$99.1 \pm 3.0\%$

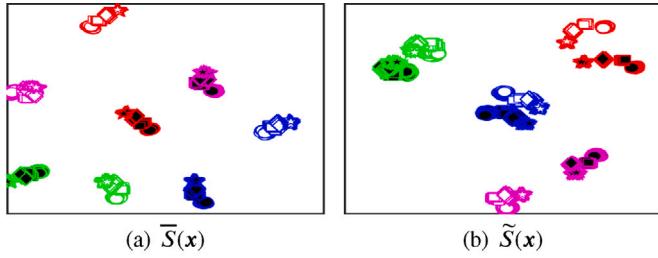


Fig. 19. t-SNE of CWRU-DD and CWRU-DF datasets.

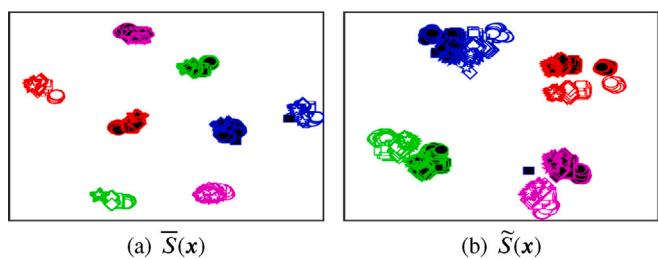


Fig. 20. t-SNE of CWRU-FF and CWRU-FD datasets.

normalized wavelet scattering features collected from different transducers become gathered together, implying that data distribution discrepancy due to transmission path variability has been significantly reduced. Fig. 20 shows the t-SNE results of the CWRU-FF and CWRU-FD datasets, which reports the same finding, *i.e.*, normalized wavelet scattering features successfully reduced domain discrepancy due to transmission path variability.

4.3.5. The effects of NTScatNet's hyperparameters

This subsection studies how hyperparameter settings affects NTScatNet's domain generalization diagnosis performance. Table 6 summarizes NTScatNet's average diagnosis accuracy on 64 transfer diagnosis tasks under 24 hyperparameter settings. As Table 6 reports, NTScatNet's performance is mainly influenced by the dimension of the input signal. NTScatNet's diagnosis accuracies with different input dimensions are presented in Fig. 21, which illustrates that NTScatNet's domain generalization performance improves as L increases. Specifically,

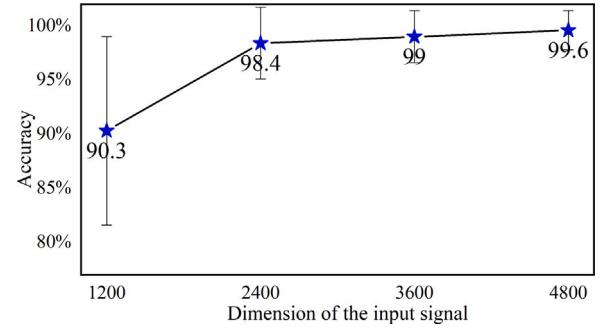


Fig. 21. The effects of L on NTScatNet's performance.

NTScatNet realizes $90.3 \pm 8.7\%$ accuracy at hyperparameters H03, which is the best diagnosis accuracy when $L = 1200$. Besides, the proposed NTScatNet achieves $98.4 \pm 3.3\%$ and $99 \pm 2.4\%$ accuracy at hyperparameters H12 and H15, respectively, where $98.4 \pm 3.3\%$ is the best diagnosis accuracy when $L = 2400$ and $99 \pm 2.4\%$ is the best diagnosis accuracy when $L = 3600$. Finally, NTScatNet realizes $99.6 \pm 1.8\%$ accuracy at hyperparameters H21, which is the best diagnosis accuracy when $L = 4800$. Section 3.4.2 has derived that NTScatNet's computational complexity is approximately the order of $L(\log_2 L)^3$, which grows not fastly as L increases. Therefore, in practice, selecting a large dimension of input signals is recommended to improve NTScatNet diagnosis performance as the computing resources allow.

4.4. Case study 2: detecting foreign objects left on the escalator guide rail

This case study applies the proposed NTScatNet to detect foreign objects at different positions of the escalator guide rail, which provides the other case to verify NTScatNet's domain generalization capability across transmission paths.

4.4.1. Background

The escalator is a vertical transportation system for transporting passengers over a short distance, consisting of a series of end-to-end steps, guide rails, et al. The guide rail is installed on the escalator truss and is used to support the escalator step. The steps are pulled by the step chain and circulate on the guide rails to transport passengers. In the daily use of escalators, some passengers may unintentionally discard foreign objects, and the foreign objects may fall on the escalator guide rails. It will generate a jolt when the escalator step rollers run over the foreign objects left on the escalator guide rail. It is practically significant to develop an IFD method to automatically detect foreign objects left on the escalator guide rail for escalator comfort and safety.

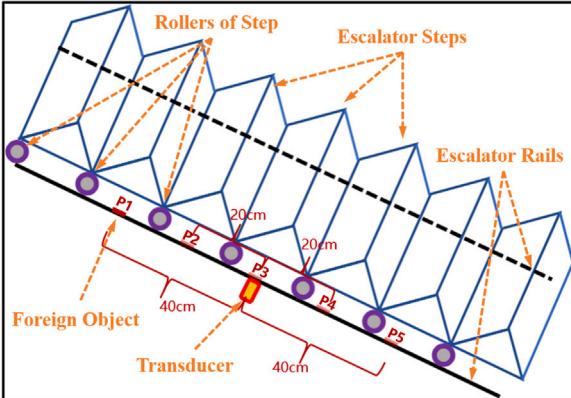


Fig. 22. Diagram of the experimental scheme.

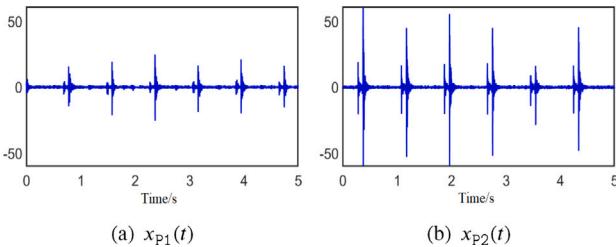


Fig. 23. Time domain signals.

4.4.2. Principles of detecting foreign objects left on the escalator guide rail

The escalator guide rail length is usually more than 1 meter, and the foreign object may be left at various positions of the guide rail. Fig. 22 presents the diagram of the experimental scheme, which shows that the transducer is mounted on the middle of the guide rail, i.e., P3. Besides, the experimental scheme gives five potential positions of the foreign objects, i.e., P1-P5. Since the transmission path characteristics between the foreign object and the transducer vary with the position of the foreign object, a significant distribution discrepancy generally exists between fault signals generated by foreign objects at different positions. Fig. 23(a) and 23(b) present the signal generated by a foreign object at P3 and P1 positions, respectively, which shows a significant amplitude discrepancy. The domain discrepancy due to transmission variability brings difficulties to the IFD method and should be removed before the diagnostic decision.

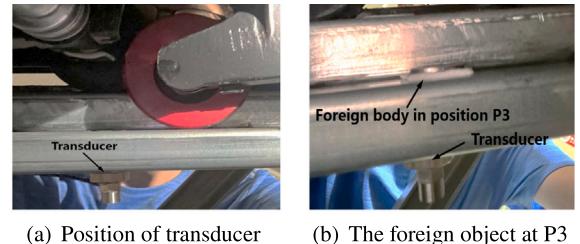
4.4.3. Datasets and transfer diagnosis tasks

This subsection introduces the datasets and diagnosis tasks that are used to verify NTScatNet's domain generalization capability across transmission paths. Fig. 24 presents the escalator experimental platform where the datasets are collected. Fig. 25(a) presents the transducer installation position in the escalator guide rail. The condition monitoring signals of the escalator guide rail are collected at the sampling frequency of 1200 Hz. Firstly, the vibration signals of the escalator guide rail in the normal state are recorded for 5 min. Then a coin is pasted on the rail to simulate the foreign object left on the escalator guide rail. Fig. 22 gives five positions where the coin is pasted, i.e., P1-P5. More specifically, position P3 is nearby the transducer, P2 and P4 are 20 centimeters from the transducer, and P1 and P5 are 40 centimeters from the transducer. Figs. 25(b) and 25(c) show the foreign body at P3 and P4, respectively. The signals of the guide rail in an abnormal state are recorded for 1 min for each foreign object position.

(1) *Datasets:* The vibration signal in the normal state is divided into 500 samples of dimension L with the overlapping trick, while the vibration signal in the abnormal state is divided into 100 samples



Fig. 24. Escalator experimental platform.



(a) Position of transducer
 (b) The foreign object at P3
 (c) The foreign object at P4

Fig. 25. The transducer installation position and the position of the foreign body in the guide.

Table 7

Information of escalator datasets,belowfloat=15pt,abovefloat=15pt.

	ES_{P1}	ES_{P2}	ES_{P3}	ES_{P4}	ES_{P5}
Position of the foreign object	P1	P2	P3	P4	P5
Number of normal samples	100	100	100	100	100
Number of abnormal samples	100	100	100	100	100
Dimension of the sample	L	L	L	L	L

Table 8

Transfer diagnosis tasks on the escalator datasets.

Index	Source→Target domain	Index	Source→Target domain
T01	$ES_{P1} \rightarrow ES_{P2}$	T11	$ES_{P3} \rightarrow ES_{P4}$
T02	$ES_{P1} \rightarrow ES_{P3}$	T12	$ES_{P3} \rightarrow ES_{P5}$
T03	$ES_{P1} \rightarrow ES_{P4}$	T13	$ES_{P4} \rightarrow ES_{P1}$
T04	$ES_{P1} \rightarrow ES_{P5}$	T14	$ES_{P4} \rightarrow ES_{P2}$
T05	$ES_{P2} \rightarrow ES_{P1}$	T15	$ES_{P4} \rightarrow ES_{P3}$
T06	$ES_{P2} \rightarrow ES_{P3}$	T16	$ES_{P4} \rightarrow ES_{P5}$
T07	$ES_{P2} \rightarrow ES_{P4}$	T17	$ES_{P5} \rightarrow ES_{P1}$
T08	$ES_{P2} \rightarrow ES_{P5}$	T18	$ES_{P5} \rightarrow ES_{P2}$
T09	$ES_{P3} \rightarrow ES_{P1}$	T19	$ES_{P5} \rightarrow ES_{P3}$
T10	$ES_{P3} \rightarrow ES_{P2}$	T20	$ES_{P5} \rightarrow ES_{P4}$

Table 9
Hyperparameters of NTScatNet.

Q_1	Q_2	L	Optimizer	Learning rate	Weight decay	Batch size	Epoch
16	4	2400	Adam	0.001	0.0001	16	500

Table 10
Experimental results.

Tasks	Spectrum	TScatNet	NTScatNet
T01	100 \pm 0.0%	50.9 \pm 0.7%	100 \pm 0.0%
T02	91.6 \pm 9.3%	50.0 \pm 0.0%	100 \pm 0.0%
T03	91.0 \pm 7.5%	50.0 \pm 0.0%	100 \pm 0.0%
T04	79.8 \pm 13.7%	81.9 \pm 1.1%	100 \pm 0.0%
T05	83.5 \pm 19.7%	100 \pm 0.0%	100 \pm 0.0%
T06	97.4 \pm 6.9%	55.2 \pm 2.0%	100 \pm 0.0%
T07	99.9 \pm 0.2%	99 \pm 0.0%	100 \pm 0.0%
T08	71.0 \pm 19.1%	100 \pm 0.0%	100 \pm 0.0%
T09	50.5 \pm 1.2%	100 \pm 0.0%	100 \pm 0.0%
T10	81.1 \pm 20.9%	100 \pm 0.0%	100 \pm 0.0%
T11	85.9 \pm 16.2%	100 \pm 0.0%	100 \pm 0.0%
T12	57.5 \pm 14.2%	99.9 \pm 0.3%	100 \pm 0.0%
T13	59.1 \pm 10.9%	99 \pm 0.0%	99.5 \pm 0.5%
T14	99.9 \pm 0.3%	100 \pm 0.0%	100 \pm 0.0%
T15	94.1 \pm 12.7%	88.0 \pm 4.6%	100 \pm 0.0%
T16	69.9 \pm 21.2%	100 \pm 0.0%	100 \pm 0.0%
T17	92.1 \pm 9.9%	100 \pm 0.0%	100 \pm 0.0%
T18	99.7 \pm 0.8%	63.9 \pm 2.2%	100 \pm 0.0%
T19	79.4 \pm 12.7%	50.0 \pm 0.9%	100 \pm 0.0%
T20	99.9 \pm 0.2%	50.5 \pm 0.2%	100 \pm 0.0%

of dimension L . A total of 500 samples in the normal state and 500 samples in the abnormal state are obtained. As Table 7 presents, five datasets are constructed using the above 100 samples, where the ES_{Pi} dataset contains the $(i - 1) * 100 + 1$ st to $(i * 100)$ th normal samples and the abnormal samples of position P_i .

(2) Transfer diagnosis tasks: As presented in Table 8, 20 transfer diagnosis tasks are constructed based on the escalator datasets to verify NTScatNet's domain generalization capability across transmission paths. For example, the meaning of task T04 is learning diagnosis experience from the ES_{Pi} dataset and applying the learned diagnosis knowledge to the ES_{Ps} subdataset. The other tasks follow similar conventions.

4.4.4. Experimental results

Table 10 and Fig. 26 report NTScatNet's accuracies on tasks T01–T20, where each experiment is repeated ten times to reduce the influence of random factors. NTScatNet's hyperparameters adopted in the current experiments are given in Table 9. The Fourier spectrum and the TScatNet methods were also evaluated on the above tasks as comparisons. As Fig. 26 shows, the Fourier spectrum method realized an average 84.2% accuracy on tasks T01–T20, which is inferior to the proposed NTScatNet 15.8%. Besides, TScatNet achieved an average 81.9% accuracy on tasks T01–T20, inferior to NTScatNet 18.1%. Specifically, as Table 10 presents, the Fourier spectrum method realized only $50.5 \pm 1.2\%$, $57.5 \pm 14.2\%$, and $59.1 \pm 10.9\%$ accuracies on tasks T09, T12, and T13, respectively. Moreover, the accuracies of the TScatNet method on tasks T01–T03, T06, and T18–T20 are generally below 70%, indicating significant domain discrepancies due to transmission variabilities. In comparison, the developed NTScatNet achieves above $99.5 \pm 0.5\%$ accuracy on all 20 transfer diagnosis tasks. The above experimental results verify that NTScatNet holds excellent domain generalization diagnosis capability across different transmission paths and is practical to detect foreign objects left on the escalator guide rail.

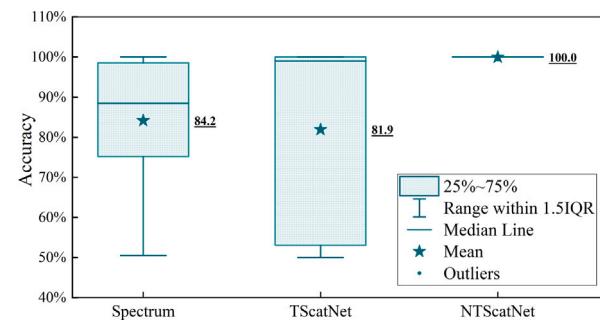


Fig. 26. Results of the transfer diagnosis experiments.

4.4.5. Effects of hyperparameters

Table 11 reports NTScatNet's average diagnosis accuracy on 20 diagnosis tasks under 24 hyperparameter settings. As Table 11 presents, NTScatNet realizes above 99.8% accuracy when the dimension of the input signal takes 1200 and achieves 100% accuracy when the input signal dimension is larger than 1200. The experimental results illustrate that the proposed NTScatNet hold excellent domain generalization capability across transmission paths under various hyperparameters.

5. Conclusion

This article developed an interpretable convolutional neural network, i.e., NTScatNet, for domain generalization diagnosis across transmission paths. NTScatNet's architecture is similar to a standard CNN, while NTScatNet's distinctiveness is that it takes Morlet wavelet convolutional kernel, modulo activation function, and scaling function window-based average averaging pooling layer. This article gave a detailed physical interpretation of each layer of NTScatNet and showed that NTScatNet's feature extractor well characterizes the multi-scale cyclostationarity information of the fault signals. Besides, we theoretically illustrated the normalized scattering feature's invariance to a linear time-invariant system. Finally, we experimentally verified that NTScatNet holds the domain generalization diagnosis capability across different transmission paths through transfer diagnosis tasks across transducers and the tasks of detecting foreign objects on the escalator guide rail.

CRediT authorship contribution statement

Chao Liu: Idea, Methodology, Original draft, Software, Data analysis. **Xiaolong Ma:** Data collecting, Data analysis. **Tianyu Han:** Software, Data analysis. **Xi Shi:** Writing – review & editing. **Chengjin Qin:** Review, Supervision. **Songtao Hu:** Review & language 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 data that has been used is confidential.

Acknowledgment

The authors acknowledge the supported by the National Natural Science Foundation of China under Grant No. 51935007.

Table 11
NTScatNet's accuracies under 24 hyperparameter settings.

Index	Hyperparameters	Accuracy	Index	Hyperparameters	Accuracy
H01	$L = 1200, Q_1 = 8, Q_2 = 1$	$100 \pm 0.1\%$	H13	$L = 3600, Q_1 = 8, Q_2 = 1$	$100 \pm 0.0\%$
H02	$L = 1200, Q_1 = 8, Q_2 = 2$	$99.8 \pm 0.5\%$	H14	$L = 3600, Q_1 = 8, Q_2 = 2$	$100 \pm 0.1\%$
H03	$L = 1200, Q_1 = 8, Q_2 = 4$	$99.9 \pm 0.3\%$	H15	$L = 3600, Q_1 = 8, Q_2 = 4$	$100 \pm 0.1\%$
H04	$L = 1200, Q_1 = 16, Q_2 = 1$	$99.8 \pm 0.6\%$	H16	$L = 3600, Q_1 = 16, Q_2 = 1$	$100 \pm 0.2\%$
H05	$L = 1200, Q_1 = 16, Q_2 = 2$	$99.8 \pm 0.6\%$	H17	$L = 3600, Q_1 = 16, Q_2 = 2$	$100 \pm 0.1\%$
H06	$L = 1200, Q_1 = 16, Q_2 = 4$	$99.8 \pm 0.4\%$	H18	$L = 3600, Q_1 = 16, Q_2 = 4$	$100 \pm 0.1\%$
H07	$L = 2400, Q_1 = 8, Q_2 = 1$	$100 \pm 0.0\%$	H19	$L = 4800, Q_1 = 8, Q_2 = 1$	$100 \pm 0.0\%$
H08	$L = 2400, Q_1 = 8, Q_2 = 2$	$100 \pm 0.0\%$	H20	$L = 4800, Q_1 = 8, Q_2 = 2$	$100 \pm 0.0\%$
H09	$L = 2400, Q_1 = 8, Q_2 = 4$	$100 \pm 0.0\%$	H21	$L = 4800, Q_1 = 8, Q_2 = 4$	$100 \pm 0.0\%$
H10	$L = 2400, Q_1 = 16, Q_2 = 1$	$100 \pm 0.0\%$	H22	$L = 4800, Q_1 = 16, Q_2 = 1$	$100 \pm 0.0\%$
H11	$L = 2400, Q_1 = 16, Q_2 = 2$	$100 \pm 0.0\%$	H23	$L = 4800, Q_1 = 16, Q_2 = 2$	$100 \pm 0.0\%$
H12	$L = 2400, Q_1 = 16, Q_2 = 4$	$100 \pm 0.0\%$	H24	$L = 4800, Q_1 = 16, Q_2 = 4$	$100 \pm 0.0\%$

References

- [1] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, A.K. Nandi, Applications of machine learning to machine fault diagnosis: A review and roadmap, *Mech. Syst. Signal Process.* 138 (2020).
- [2] J. Antoni, A critical overview of the filterbank-feature-decision methodology in machine condition monitoring, *Acoust. Aust.* 49 (2) (2021) 177–184, <http://dx.doi.org/10.1007/s40857-021-00232-7>.
- [3] Y.G. Lei, F. Jia, J. Lin, S.B. Xing, S.X. Ding, An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data, *Ieee Trans. Ind. Electron.* 63 (5) (2016) 3137–3147.
- [4] F. Jia, Y.G. Lei, N. Lu, S.B. Xing, Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization, *Mech. Syst. Signal Process.* 110 (2018) 349–367.
- [5] T. Li, Z. Zhao, C. Sun, L. Cheng, X. Chen, R. Yan, R.X. Gao, Waveletkernelnet: An interpretable deep neural network for industrial intelligent diagnosis, *IEEE Trans. Syst. Man Cybern.: Syst.* (2021).
- [6] C. Liu, C. Qin, X. Shi, Z. Wang, G. Zhang, Y. Han, Tscatnet: An interpretable cross-domain intelligent diagnosis model with anti-noise and few-shot learning capability, *IEEE Trans. Instrum. Meas.*, 0 <http://dx.doi.org/10.1109/TIM.2020.3041905>.
- [7] S. Mallat, Group invariant scattering, *Comm. Pure Appl. Math.* 65 (10) (2012) 1331–1398.
- [8] J. Bruna, S. Mallat, Invariant scattering convolution networks, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (8) (2013) 1872–1886.
- [9] J. Andén, S. Mallat, Deep scattering spectrum, *IEEE Trans. Signal Process.* 62 (16) (2014) 4114–4128, <http://dx.doi.org/10.1109/TSP.2014.2326991>.
- [10] S. Mallat, Understanding deep convolutional networks, *Phil. Trans. R. Soc. A* 374 (2016).
- [11] B.Y.J. Bruna, S.T. Mallat, E. Bacry, J.F.R. Muzy, Intermittent process analysis with scattering moments, *Ann. Statist.* 43 (1) (2015) 323–351, <http://dx.doi.org/10.1214/14-AOS1276>.
- [12] M. Andreux, T. Angles, G. Exarchakis, R. Leonarduzzi, G. Rochette, L. Thiry, J. Zarka, S. Mallat, J. Andén, E. Belilovsky, J. Bruna, V. Lostanlen, M. Chaudhary, M.J. Hirn, E. Oyallon, S. Zhang, C. Cellia, M. Eickenberg, Kymatio: Scattering transforms in python, *J. Mach. Learn. Res.* 21 (2020).
- [13] R. Liu, B. Yang, E. Zio, X. Chen, Artificial intelligence for fault diagnosis of rotating machinery: A review, *Mech. Syst. Signal Process.* 108 (2018) 33–47.
- [14] Z. Zhao, T. Li, J. Wu, C. Sun, S. Wang, R. Yan, X. Chen, Deep learning algorithms for rotating machinery intelligent diagnosis: An open source benchmark study, *ISA Trans.* 107 (2020) 224–255.
- [15] Z. Zhao, Q. Zhang, X. Yu, C. Sun, S. Wang, R. Yan, X. Chen, Applications of unsupervised deep transfer learning to intelligent fault diagnosis: A survey and comparative study, *IEEE Trans. Instrum. Meas.* 70 (2021) <http://dx.doi.org/10.1109/TIM.2021.3116309>.
- [16] H. Zheng, R. Wang, Y. Yang, J. Yin, Y. Li, Y. Li, M. Xu, Cross-domain fault diagnosis using knowledge transfer strategy: A review, *IEEE Access* 7 (2019) 129260–129290.
- [17] W. Li, R. Huang, J. Li, Y. Liao, Z. Chen, G. He, R. Yan, K. Gryllias, A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges, *Mech. Syst. Signal Process.* 167 (2022).
- [18] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, T. Zhang, Deep model based domain adaptation for fault diagnosis, *IEEE Trans. Ind. Electron.* 64 (3) (2017) 2296–2305.
- [19] X. Li, W. Zhang, Q. Ding, J.Q. Sun, Multi-layer domain adaptation method for rolling bearing fault diagnosis, *Signal Process.* 157 (2019) 180–197.
- [20] X. Li, W. Zhang, Q. Ding, A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning, *Neurocomputing* 310 (2018) 77–95.
- [21] X. Li, W. Zhang, N.X. Xu, Q. Ding, Deep learning-based machinery fault diagnostics with domain adaptation across sensors at different places, *IEEE Trans. Ind. Electron.* 67 (8) (2020) 6785–6794.
- [22] C. Cheng, B. Zhou, G. Ma, D. Wu, Y. Yuan, Wasserstein distance based deep adversarial transfer learning for intelligent fault diagnosis with unlabeled or insufficient labeled data, *Neurocomputing* 409 (2020) 35–45.
- [23] L. Chen, Q. Li, C. Shen, J. Zhu, D. Wang, M. Xia, Adversarial domain-invariant generalization: A generic domain-regressive framework for bearing fault diagnosis under unseen conditions, *IEEE Trans. Ind. Inf.* 18 (3) (2022) 1790–1800, <http://dx.doi.org/10.1109/TII.2021.3078712>.
- [24] T. Han, Y.F. Li, M. Qian, A hybrid generalization network for intelligent fault diagnosis of rotating machinery under unseen working conditions, *IEEE Trans. Instrum. Meas.* 70 (2021) <http://dx.doi.org/10.1109/TIM.2021.3088489>.
- [25] J. Li, C. Shen, L. Kong, D. Wang, M. Xia, Z. Zhu, A new adversarial domain generalization network based on class boundary feature detection for bearing fault diagnosis, *IEEE Trans. Instrum. Meas.* 71 (2022) <http://dx.doi.org/10.1109/TIM.2022.3164163>.
- [26] Y. Liao, R. Huang, J. Li, Z. Chen, W. Li, Deep semisupervised domain generalization network for rotary machinery fault diagnosis under variable speed, *IEEE Trans. Instrum. Meas.* 69 (10) (2020) 8064–8075, <http://dx.doi.org/10.1109/TIM.2020.2992829>.
- [27] M. Ragab, Z. Chen, W. Zhang, E. Eldele, M. Wu, C.K. Kwoh, X. Li, Conditional contrastive domain generalization for fault diagnosis, *IEEE Trans. Instrum. Meas.* 71 (2022) <http://dx.doi.org/10.1109/TIM.2022.3154000>.
- [28] H. Wang, X. Bai, S. Wang, J. Tan, C. Liu, Generalization on unseen domains via model-agnostic learning for intelligent fault diagnosis, *IEEE Trans. Instrum. Meas.* 71 (2022) <http://dx.doi.org/10.1109/TIM.2022.3152316>.
- [29] Q. Zhang, Z. Zhao, X. Zhang, Y. Liu, C. Sun, M. Li, S. Wang, X. Chen, Conditional adversarial domain generalization with a single discriminator for bearing fault diagnosis, *IEEE Trans. Instrum. Meas.* 70 (2021) <http://dx.doi.org/10.1109/TIM.2021.3071350>.
- [30] C. Zhao, W. Shen, Adaptive open set domain generalization network: Learning to diagnose unknown faults under unknown working conditions, *Reliab. Eng. Syst. Saf.* 226 (2022) <http://dx.doi.org/10.1016/j.ress.2022.108672>.
- [31] C. Zhao, W. Shen, Adversarial mutual information-guided single domain generalization network for intelligent fault diagnosis, *IEEE Trans. Ind. Inf.* (2022) 1, <http://dx.doi.org/10.1109/TII.2022.3175018>.
- [32] C. Zhao, W. Shen, A domain generalization network combining invariance and specificity towards real-time intelligent fault diagnosis, *Mech. Syst. Signal Process.* 173 (2022) <http://dx.doi.org/10.1016/j.ymssp.2022.108990>.
- [33] Chengjin Qin, Yanrui Jin, Jianfeng Tao, Dengyu Xiao, Honggan Yu, Chao Liu, Gang Shi, Junbo Lei, Chengliang Liu, Dtcnnmi: a deep twin convolutional neural networks with multi-domain inputs for strongly noisy diesel engine misfire detection, *Measurement: Journal of the International Measurement Confederation* 180 (2021) <http://dx.doi.org/10.1016/j.measurement.2021.109548>.