Research Article

Discrete Cosine Transformation and Temporal Adjacent Convolutional Neural Network-Based Remaining Useful Life Estimation of Bearings

Yu Pang¹ ,¹ Limin Jia,¹ and Zhan Liu²

¹State Key Lab of Rail Traffic Control & Safety, Beijing Jiao Tong University, Beijing 100044, China

²Beijing Neco Automation Technology Co., Ltd., Beijing 100044, China

Correspondence should be addressed to Yu Pang; pangyu_bjtu@163.com

Received 17 January 2020; Accepted 27 March 2020; Published 9 June 2020

Academic Editor: Anil Kumar

Copyright © 2020 Yu Pang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, several time-frequency representation (TFR) and convolutional neural network- (CNN-) based approaches have been proposed to provide reliable remaining useful life (RUL) estimation for bearings. However, existing methods cannot tackle the spatiotemporal continuity between adjacent TFRs since temporal proposals are considered individually and their temporal dependencies are neglected. In allusion to this problem, a novel prognostic approach based on discrete cosine transformation (DCT) and temporal adjacent convolutional neural network (TACNN) is proposed. Wavelet transform (WT) is applied to effectively map the raw signals to the time frequency domain. Considering the high load and complexity of model computation, bilinear interpolation and DCT algorithm are introduced to convert TFRs into low-dimensional DCT spectrum coding matrix with strong sparsity. Furthermore, the TACNN model is proposed which is capable of learning discriminative features for temporal adjacent DCT spectrum coding matrix. Effectiveness of the proposed method is verified on the PRONOSTIA dataset, and experiment results show that the proposed model is able to realize automatic high-precision estimation of bearings RUL with high efficiency.

1. Introduction

As a key component widely used in various types of rotating machinery, fault diagnosis and failure prediction of bearings are of great significance in industrial production and daily life. [1, 2]. The development of mechanical equipment fault maintenance strategy has evolved from initial postfault and periodic maintenance to the current equipment-based maintenance and gradually evolved into a smart predictive maintenance strategy [3]. Bearings will undergo a process from normal to failure during service and experience a series of degradation states. When fault diagnosis is carried out, its performance is often seriously degraded. In order to provide maintenance solutions as soon as possible, it is necessary to track the degradation states of bearings. Remaining useful life (RUL) estimation method can continuously monitor mechanical components before the bearing failure occurs,

track the degradation state of bearings throughout the life cycle, and establish models to predict the occurrence of failure. So, research on the RUL estimation method of bearings is of great significance for improving safety and reliability of equipment [4].

It is generally believed that the structural information contained in vibration signals is abundant (including a variety of information such as amplitude, frequency, and phase) and is able to reflect mechanical states of systems [5]. With continuous improvement of diagnostic capabilities, the prognostic and health management (PHM) technology under vibration can initially provide early detection and monitoring capabilities for component RUL estimation. After decades of exploration, fruitful theoretical research achievements have appeared in the academic field, and they also have a broad application in industry [6–9]. Existing methods for RUL prediction can be classified into three

kinds, namely, physics-based methods, data driven methods, and fusion methods. Due to the rapidly increasing complexity of engineering equipment, it is extremely difficult to obtain the physical model of failure mechanism in advance or expensive to obtain through experiments. On the contrary, data driven methods can extract information related to reliability from condition monitoring data, which can save time and reduce expenditure effectively. Therefore, data driven methods have become the mainstream methods in the field of RUL prediction.

Due to the complexity of mechanical equipment structure and operating status, bearing vibration signals often have obvious nonlinear and nonstationary time-varying characteristics. Analysis only in time or frequency domain is mostly based on stationarity of signals and is difficult to fully describe the nonstationary time-varying characteristics [10–12]. In comparison, time-frequency representation (TFR) methods have provided us effective solution for nonstationary signal processing. In recent years, Lou and Loparo [13] used the wavelet transform to process the signals and generate feature vectors for bearing fault diagnosis. Shi et al. [14] proposed a bearing instantaneous frequency measured method based on short-time Fourier transform (STFT). Huang et al. [15] used STFT distribution combined with generalized demodulation for variable speed bearing fault diagnosis. Zhang et al. [16] proposed a novel time-frequency analysis method termed CMQGW on the basis of based on continuous wavelet transform (CWT) and multiple Q-factor Gabor wavelets (MQGWs). Brkovic et al. [17] used wavelet transform (WT) for early fault detection and diagnosis. Chen et al. [18] used empirical wavelet transform (EWT) to extract inherent modulation information by decomposing signal into monocomponents under an orthogonal basis, which is seen as a powerful tool for mechanical fault diagnosis. Generally, TFR-based bearing fault diagnosis approaches have achieved fruitful results, which verify its capability to characterize the running state of bearings. While effective applications of TFR in bearing RUL estimation still need to be explored in depth.

With the in-depth development of information theory, artificial intelligence (AI) technology, known as one of the three cutting-edge technologies of the 21st century, has been proved to be more suitable for solving the performance degradation state tracking and failure prediction of complex equipment [19–21]. The advanced deep learning algorithms can effectively analyse a large amount of data and establish the mapping relationship between data and features. Jahromi et al. [22] combined wavelet analysis with dynamic fuzzy neural network for fault prediction of triaxial rotor system. Ali et al. [23] combined empirical mode decomposition (EMD) with artificial neural network (ANN) for automatic extraction of bearing fault characteristics. Kia-kojori and Khorasani [24] used nonlinear autoregressive neural network and Elman neural network to capture two main degradation dynamics of gas turbine compressor fouling and turbine wear. As a typical deep learning method, convolutional neural network (CNN) enjoys big advantages in automatically learning features from TFRs and optimizing parameters of convolutional kernels through training

[25, 26]. Ren et al. [27] proposed a CNN-based method for RUL prediction of bearings. Zhu et al. [28] realized bearing RUL estimation based on WT and multiscale CNN. Li et al. [29] applied a novel directed acyclic graph network combined with CNN and LSTM for bearing RUL prediction. Yang et al. [30] pointed out the great potential of CNN for RUL prediction. In general, these CNN-based methods undoubtedly played an important role in promoting the research of RUL estimation.

However, wide use of deep neural networks also brings new problems. CNN-based RUL estimation approaches usually consist of two major steps: frame-level TFRs generation and association of proposals across frames [31]. Also, most of these methods only employ two-stream CNN framework to handle spatial feature and cannot tackle the spatiotemporal continuity between adjacent TFRs since temporal proposals are considered individually and temporal dependencies are neglected. In order to further develop the method, it is necessary to consider how to learn the spatiotemporal representation with CNN while maintaining strong feature extraction capability and low computational complexity. To do so, a novel-bearing RUL estimation method based on discrete cosine transformation (DCT) algorithm and temporal adjacent convolutional neural network (TACNN) model is proposed. Considering the high load of model computation and the complexity of CNN parameter training brought by the high dimension of time-frequency images, DCT algorithm is introduced to convert the WT-TFRs into 2-dimensional coding matrix with strong sparsity. Furthermore, a novel TACNN model is proposed that is capable of learning discriminative features for temporal adjacent DCT spectrums. Effectiveness of the proposed method is verified on the PRONOSTIA dataset, RUL of bearings is taken as the output value directly, and the mapping relationship between DCT spectrums and RUL is obtained effectively. Experiment results show that the proposed model is able to realize automatic high-precision estimation of bearing RUL with high efficiency.

2. Theoretical Background

In this section, the proposed model for bearing RUL estimation is presented in detail. Flowchart of the proposed method is shown in Figure 1. There are mainly 3 stages: time-frequency representation of raw signals by wavelet transform, dimensionality reduction and discrete cosine transform coding of TFRs, and regression estimation of RUL based on TACNN. To represent the nonstationary property effectively, TFR by means of wavelet transform is used for raw degradation signals. Since TFRs are usually regarded as high-dimensional features, bilinear interpolation-based dimensionality reduction, and discrete cosine transform are applied to convert the WT-TFRs into low-dimensional sparse DCT feature maps. To effectively characterize and extract the spatiotemporal continuity between adjacent TFRs, new data form of temporal continue DCT spectrum clips were proposed and applied. Then, the spectrum clips together with their assigned labels are sent to train the TACNN model. Since TACNN is a supervised learning

approach, we can assign target RUL value as the label of the constructed DCT spectrum clips. More details of the proposed RUL estimation method are presented as follows.

2.1. Time-Frequency Representation. When bearings begin to degrade, the measured vibration signals will exhibit non-stationary characteristics. Under this situation, both time domain and frequency domain analysis methods fail to provide the time-varying feature information [28]. However, wavelet transform is more suitable for analysis of this kind of signals. In addition, the wavelet function has the property of tight support and the sensitivity of local maximum modulus to singularity, which makes it widely used in condition monitoring of bearings [32, 33].

Considering $\psi(t) \in L^2(R)$, which is the square integrable function, if the Fourier transform of $\psi(t)$ meets the admissible condition

$$\int_{-\infty}^{+\infty} \frac{|\psi(\omega)|^2}{\omega} d\omega < \infty, \quad (1)$$

then $\psi(t)$ can be seen as basic wavelet or generating wavelet function. After the original wavelet $\psi(t)$ is scaled or expanded a times and translated b steps, we can obtain

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right), \quad (2)$$

where a refers to the scale factor and b refers to the shift factor. The function is basis of two-dimensional space generated by the wavelet function after stretching and translation. Morlet wavelet is chosen as the basis wavelet since it is similar to bearing impulse signals [34, 35]. Calculation of WT is to convolute a signal with wavelet basis, decompose the signal into various components in different frequency bands and time bands, and then analyse and process it in the next step. Now, assume the signal $x(t)$, and perform the wavelet transform to obtain

$$\text{WT}_f(a, b) = \frac{1}{\sqrt{a}} \int_0^{+\infty} x(t) \psi^*\left(\frac{t-b}{a}\right) dt, \quad (3)$$

where $\psi^*(t)$ represents the conjugate function of the fundamental wavelet $\psi(t)$. The TFR solution result can be regarded as a Euclidean space with “time” as the abscissa and “frequency” as the ordinate, which can be used to describe the physical state of the signal energy distribution.

2.2. Dimensionality Reduction and Discrete Cosine Transform. Instead of using common image compressing methods, such as principle component analysis (PCA) or singular value decomposition (SVD), bilinear image interpolation technique is used here for dimensionality reduction of TFRs. Bilinear interpolation is the most popular interpolation method for its simplicity in image-based applications. It interpolates a high resolution pixel using weighted average of four surrounding pixels, as shown in Figure 2. Assume that $F(s, t)$ is the pixel to be interpolated and can be calculated by

$$F(s, t) = (1-t) \left[(1-s) \cdot f_{i,j} + s \cdot f_{i+1,j} \right] + t \left[(1-s) \cdot f_{i,j+1} + s \cdot f_{i+1,j+1} \right], \quad (4)$$

where the four nearest neighbours, $f_{i,j}$, $f_{i+1,j}$, $f_{i,j+1}$, and $f_{i+1,j+1}$ are surrounding pixels and s and t represent the locations of $F(s, t)$.

Discrete cosine transform (DCT) [36] is widely used in block signal coding, since it performs closely to the statistically optimal *Karhunen-Loeve* transform for a wide class of signals. Both DCT and FFT transform belong to compression transform. In image processing, it is generally considered that the amount of information in low-frequency part is greater than that in the high-frequency part, but the amount of data in the low-frequency part is much smaller. DCT can concentrate highly correlated data information, transforms images in spatial domain into frequency domain, and has good performance of decorrelation. The DCT transformation is lossless, which creates good conditions for the subsequent quantization in fields of image coding [37].

Basis vector of DCT transformation core is independent of the image content, and the transformation core can be separated. So, the two-dimensional DCT can be completed with two one-dimensional DCT transformations, which greatly simplifies the difficulty of mathematical operation. The application of DCT in TFR matrix data compression can reduce the digital information of brightness level on behalf of the image and achieve the purpose of data compression. If there is a TFR matrix block with size $N \times N$ in the airspace, $f(x, y)$ is the amplitude value of a pixel whose coordinate is (x, y) in the matrix, $F(u, v)$ is the coefficient value after transformation, and (u, v) is the position coordinate of the pixel after transformation; then, the corresponding DCT and IDCT are, respectively, as follows:

$$F(u, v) = c(u)c(v) \frac{2}{N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \cos \left(\frac{2x+1}{2N} \pi u \right) \cos \left(\frac{2y+1}{2N} \pi v \right), \quad (5)$$

where $x, y, u, v = 0, 1, \dots, N-1$ and $c(u) = c(v) = \begin{cases} 1/\sqrt{2} & u = 0, v = 0 \\ 1 & \text{else} \end{cases}$.

$$f(x, y) = \frac{2}{N} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} c(u)c(v) F(u, v) \cos \left(\frac{2x+1}{2N} \pi u \right) \cos \left(\frac{2y+1}{2N} \pi v \right). \quad (6)$$

DCT compresses the matrix data according to statistical characteristics of signals in frequency domain. It converts the original TFR block into a set of coefficients representing different frequency components, concentrating most of its energy in a small range of the frequency domain so that only a small number of bits are needed to describe the unimportant components. Under the premise of maximally retaining equipment fault status information, sparsity of the

FIGURE 1: Flowchart of the proposed method.

FIGURE 2: Pixel arrangement of bilinear interpolation.

input data can be increased, which is beneficial to greatly reduce the training time and storage of the network.

2.3. Temporal Adjacent Convolutional Neural Network. Signal analysis in time domain, frequency domain, or amplitude domain alone cannot represent its local and global characteristics in time-frequency joint distribution domain.

In comparison, WT analysis can showcase the time-varying characteristics of each frequency component in signals and bring better RUL prediction results by reasonable application. However, from the current application of TFR in RUL prediction, dependence between current observation and previous states is often ignored. Over time, it is a gradual process for bearings to degrade from normal operation to failure. The state of a bearing at each moment must have relationship with states of previous period. This relationship needs us to obtain with abundant data analysis and signal processing means. To effectively characterize and extract the spatiotemporal continuity between adjacent TFRs, new data form of temporal continue streams were proposed and applied. However, compared with the time-domain signal, the TFRs have already greatly increased the calculation amount of the entire process. The transformation into temporal continue streams will bring a greater computing burden and seriously affect the promotion and application of such methods, especially when the data quantity of vibration signal is large. Bilinear interpolation is a relatively simple dimensionality reduction method; more distribution information will be lost during the process if it is used to reduce the dimensionality by a large margin. The DCT method can achieve almost lossless information compression by converting images from the spatial domain to the frequency domain. Therefore, after a certain degree of

dimension reduction, DCT spectrum stream is constructed to enhance the sparsity of time-frequency image matrix and compress the information to a greater extent in this paper. As an important branch of deep learning approaches, CNN is more suitable for learning and expressing image features than other neural network methods [38–41]. In order to represent the adjacent relations of DCT spectrum stream and improve the calculation efficiency, TACNN is proposed here.

Figure 3 showcases the schematic diagram of the generated continue DCT spectrum stream. For an input temporal continue DCT spectrum stream, we first segment it into small spectrum clips. Each clip s_i consists of P frames. Then, a fixed-interval sampling is performed over the spectrum clips and N clips are obtained, denoted as $S = \{s_i\}_{i=0}^{N-1}$. In order to ensure the continuity of training and data, different clips overlap with frames. While extracting spatial time-frequency features of TFRs, the temporal overlapping method can effectively represent the temporal continuity of DCT spectrums corresponding to adjacent signal intervals, which is more conducive to extracting rich degradation information in vibration signals. A new variable $P - 1$ is introduced here, and its value has great effect on the result. When P value is small, time-adjacent information cannot be represented effectively. However, when P value is large, the data difference corresponding to adjacent labels will decrease, which is not conducive to regression estimation. Selection of P value will be discussed later. Labels of the training data are defined based on the RUL computed by the starting moments of the time segments. By the distance of the coordinates of adjacent time segments, we can define the proximity relationship between these segments. With these relationships, the temporal dependencies can be modeled.

To get a more compact representation, the extracted feature is passed through a fully connected layer with ds output channels. The final representation of a sampled video clip is represented as $f_s \in \mathbb{R}^{ds}$, where ds is the feature dimension. The sampled N clips serve as the basic elements for moment candidate construction. Thus, the feature map of moment candidates by the clip features $\{f_v\}_{v=0}^{N-1}$ is built up. As a result, features of moment candidates are constructed.

A sparse strategy is applied to fast the model. Donate the i th convolutional layer with input channels/height/width of $n_i/h_i/w_i$, and input $x_i \in \mathbb{R}^{n_i \cdot h_i \cdot w_i}$ can be transformed into $x_{i+1} \in \mathbb{R}^{n_{i+1} \cdot h_{i+1} \cdot w_{i+1}}$ by convolutional layers. In this process, one filter generates one feature map by applying n_{i+1} 3D filters on the n_i input channels. And each filter is composed by n_i 2D kernels. Regardless of the bias, the parameter dimension is $n_i \times n_{i+1} \times k_h \times k_w$. The kernels that were applied on the removed feature maps from the filters of the next convolutional layer are also removed, which saves $n_{i+2}k_h^2h_{i+2}w_{i+2}$ operations. Pruning m filters of layer i will help reduce $m = n_{i+1}$ of the computation cost for both layers i and $i + 1$. Calculate the $L1$ weights for each filter and remove the smallest number, and this value gives an expectation of the magnitude of the output feature map. Filters with smaller kernel weights tend to produce feature maps with weak activations as compared to the other filters in that

layer. Due to the strong sparsity of DCT spectrum stream, we delete the filter with the weight of 0 to accelerate the network training.

Architecture of the proposed TACNN model for RUL estimation is shown in Figure 4. There are 3 convolution layers and 2 fully connected layers in the proposed network. Input changes from a single TFR to fixed-length DCT spectrum segments. By considering the differences of these time segments as a whole, instead of considering each TFR individually, we can learn more distinguishing features. In the pooling layer, we take advantage of subsampling so that the output feature map becomes invariant to small variance in the input feature map. Furthermore, the computation efficiency is increased owing to the reduced size of feature map. Output of the sparse TACNN model is defined as the actual RUL of the bearing, and it satisfies $RUL_{\text{bearing}} \geq 0$. The predicted results are viewed as a probabilistic model with random variables ω following Gaussian prior distribution. The rectified linear units (ReLU) are used for the activation function of the whole model. For a variate v , the mathematical expression for the ReLU function is

$$\text{ReLU}(v) = \max(0, v). \quad (7)$$

In the process of network training, mean square error (MSE) is used as the loss function. Set $RUL \in \mathbb{R}^{1 \times n}$ as the true output and $\hat{RUL} \in \mathbb{R}^{1 \times n}$ for the corresponding predicted value, and the MSE is defined as follows:

$$\text{MSE}(RUL, \hat{RUL}) = \frac{1}{n} \sqrt{\sum_{i=1}^n (RUL_i - \hat{RUL}_i)^2}. \quad (8)$$

As what the model deals with here is a regression problem, we need to predict not a predefined category, but an arbitrary real number in dealing with the prediction of specific values. Given the training set $D_1 = [(x_i, y_i) | i = 1, 2, \dots, n]$, where x_i refers to the input data and y_i refers to the estimated RUL. Use the rule of structural risk minimization to define the objective function as follows:

$$\arg \min_{\theta} O(D_1; \theta) = \sum_{i=1}^n L(y_i, f(x_i); \theta), \quad (9)$$

where θ represents the whole model parameters, $f(x_i)$ is the function corresponding to the input and output of the model, and $L(y_i, f(x_i); \theta)$ refers to the MSE loss function. We use root mean square prop (RMSProp) algorithm as the adaptive optimization method to minimize the loss function through model training. It normalizes the gradient according to the exponential moving average of the gradient amplitude of each parameter. Its goal is to achieve fast convergence when the algorithm is applied to convex problems. When the algorithm is applied to nonconvex problems, it can pass through different local structures very quickly and finally reach the optimal global minimum. Furthermore, it does not require manual configuration of the learning rate hyperparameters, which is done automatically by the algorithm.

FIGURE 3: Schematic diagram of the generated continue DCT spectrum stream.

FIGURE 4: Architecture of sparse TACNN for RUL estimation. This architecture consists of three convolution layers and two fully connected layers. Detailed descriptions are given in the text.

3. Experimental Verification

3.1. Experimental Setting and Datasets. The bearing degradation data is taken from the PRONOSTIA platform in the IEEE PHM 2012 Data Challenge, provided by FEMTO-ST Institute [42]. Figure 5 shows the overview of PRONOSTIA platform. These accelerometers measure raw vibration signals at an interval of 10 s and with a sampling frequency of 25.6 kHz. The operation condition with constant speed and load (1800 rpm and 4000 N) is considered here, and the datasets are shown in Table 1. The training datasets refer to two run-to-failure bearings, which contain 2803 samples and 871 samples, respectively. The other bearings in this

operating condition are regarded as the testing datasets with censored bearing life data. Signals from the horizontal and vertical directions are considered for more comprehensive information.

We use the provided training datasets to build the prognostics model, estimate the RUL of 5 remaining bearings in the testing dataset. Figure 6 shows the totally 4 whole lifetime vibration signals from horizontal and vertical directions. The vibration signals under the same working condition have a large gap in the degradation trend and length of lifetime. Due to the limited amount and poor stability of data, it is a challenge to predict the RUL of the bearings.

FIGURE 5: Overview of PRONOSTIA platform.

TABLE 1: Bearing dataset with operational conditions.

	Vertical dataset	Horizontal dataset
Training dataset	Bearing 1_1v	Bearing 1_1h
	Bearing 1_2v	Bearing 1_2h
Testing Dataset	Bearing 1_3v	Bearing 1_3h
	Bearing 1_6v	Bearing 1_6h
	Bearing 1_4v	Bearing 1_4h
	Bearing 1_7v	Bearing 1_7h
	Bearing 1_5v	Bearing 1_5h

FIGURE 6: Whole lifetime vibration signal of training datasets from 2 directions.

3.2. RUL Estimation Using Proposed Method. Before the DCT transform, the bilinear interpolation algorithm is also used to reduce the TFR dimension. Figure 7 shows the TFR images after dimension reduction using bilinear interpolation. Figure 7(a) is the original TFR with

dimension of 256×2560 , and Figure 7(b) is the reduced-dimensional image. When the dimension is reduced to 500×500 or 250×250 , the distortion of the image is not high, and the time-frequency component is still clearly visible. When the dimension is reduced to 100×100 , the time-frequency component is already distorted. When the dimensionality is reduced to 50×50 , the distortion is blurred and the boundary of the time-frequency component becomes mosaic. It can be observed that bilinear interpolation may result in loss of high-frequency components of the scaled image. The image edges become blurred to a certain extent, and obvious aliasing and mosaic phenomena will occur when the dimension reduction is too large. In order to improve the calculation efficiency without losing too much information, the reduced image dimension is set as 100.

Figure 8 shows the illustration example of signal pre-processing procedure. The focus is using less data to express the operating status of bearings. The raw one-dimensional signal is transformed into two-dimensional time-frequency distribution (for signal), to fully represent the nonstationary time-varying information of the signal. Raw signal and its corresponding time-frequency distributions for 2 samples of bearing 1_1h is shown in the figure. The effectiveness of WT-TFRs and CNN-based bearing RUL estimation method is validated by [28]. However, due to the expansion of data dimension, the amount of computation is greatly increased. So, DCT algorithm is used here to convert the TFRs in the spatial domain (for image) into distribution in frequency domain (for image). It is worth noticing that the frequency domain of the image and the frequency domain of the signal should not be confused. DCT method is then used to convert the image after proper initial dimensionality reduction. Here, the image block number is chosen as 8 in the DCT process, and in the conversion the image information is nearly lossless. Moreover, as can be seen from Figure 8, the converted matrix is highly sparse, and the main information is concentrated in the upper left corner of the matrix, and all

FIGURE 7: Comparison of dimensionality reduction visualized TFR images with bilinear interpolation. (a) Raw WT-TFR of the last sample of bearing 1_1h. (b) The reduced dimensional TFR.

the values of other locations are almost 0, which is beneficial to accelerate the TACNN network.

The topology of the proposed TACNN is shown in Table 2, which has the input DCT distribution from 2 directions (horizontal and vertical) with size of $100 \times 100 \times 2P$. During the calculation, we changed the connection mode between the input layer and the convolutional layer to convert it into the form of clips in Figure 3. DCT transforms time-frequency images from spatial domain to frequency domain without changing the data dimension, so the model input is directly determined by the dimension of the TFR matrix after bilinear interpolation. The 3 convolutional layers have 12 filters, 24 filters, and 48 filters with size of $3 \times 3 \times 2P$ and stride of 2×2 , respectively. For an input temporal continue DCT spectrum stream, segment it into small spectrum clips with 3 frames. The 2 fully connected layers have 48 hidden units and 200 hidden units, and the output layer has 1 unit. The rectified linear units (ReLU) are used for the activation function of the whole model. Root mean square prop (RMSProp) algorithm is used as the adaptive optimization method to minimize the loss function through model training. It does not require manual configuration of the learning rate hyperparameters, which is done automatically by the algorithm. The model is carried out in a NVIDIA 1080Ti GPU with the minibatch of 128 and epoch of 100.

For sake of the influence of frame number P in each DCT clip on prediction, HIs' prediction is implemented on testing bearings with different frame numbers (set to 1, 2, 3, 4, 5), respectively. Mean average error (MAE), mean square error (MSE), and mean absolute percent error (MAPE) are compared to investigate the prediction capability, and average of 5 testing bearings is shown in Table 3. In addition, to compare and analyse the effect of DCT method on the results, we also calculated the above three indicators obtained by directly constructing the TFR clips for RUL prediction. The dimensions of DCT spectrum and time-frequency image are 100×100 . Set $RUL \in \mathbb{R}^{1 \times n}$ as the true output and

$\hat{RUL} \in \mathbb{R}^{1 \times n}$ for the corresponding predicted value, and the MAE and MAPE are defined as follows:

$$MAE(RUL, \hat{RUL}) = \frac{1}{n} \sum_{i=1}^n |RUL_i - \hat{RUL}_i|, \quad (10)$$

$$MAPE(RUL, \hat{RUL}) = \frac{1}{n} \sum_{i=1}^n \left| \frac{RUL_i - \hat{RUL}_i}{RUL_i} \right|.$$

It can be seen from the results that the selection of different frame numbers has a greater impact on the results. When P is selected as 3, there is a minimum value of MAE and MAPE for the DCT-TACNN model, and when P is 4, a minimum MSE is achieved. From the comparison of the data of the three indicators, it is better to estimate the bearing RUL when we convert the time-frequency image into the DCT spectrum. To ensure the prediction accuracy of the results, P is set as 3 in the work.

Figure 9 shows the training and testing error over iterations with different model inputs. For comparison, raw WT-TFRs and DCT spectrums with dimensions of 250×250 , 100×100 , and 50×50 are used, respectively. It can be seen from the results that as the iterations increase, the overall error gradually decreases and stabilizes around 40 iterations, indicating that the model is effective. Separately, the DCT transformation can make the model error decrease faster and enter a stable state after 20 iterations, which shows that using DCT spectrums as a model input has advantages over using TFRs directly. In order to minimize the computational complexity and ensure the accuracy, the selected dimension in the experiment is 100×100 .

Figure 10 shows the estimated RUL of the 5 testing datasets. The grey circles are results of direct output of the proposed model, as can be seen, and there exists a certain volatility. Therefore, we use a moving average filter to smooth the results and set the sliding window length of 100. As we can see, error between the predicted data and the real RUL in the initial stage is relatively large. However, after the

FIGURE 8: Illustration example of the procedure of RUL preprocessing.

TABLE 2: Topology of the sparse TACNN.

Layer	Parameters and output channel size
Input	$100 \times 100 \times 2P$
Convolution	Kernel: $3 \times 3 \times 2P$, channel: 12, ReLU
Convolution	Kernel: $3 \times 3 \times 2P$, channel: 24, ReLU
Convolution	Kernel: $3 \times 3 \times 2P$, channel: 48, ReLU
Fully connected	$48 \times 1 \times 1$, ReLU
Fully connected	$200 \times 1 \times 1$, ReLU
Output	1×1

bearing enters the recession period, trend of the predicted value is consistent with the true value. The figure shows the final predicted value at end of the testing datasets, which is also very close to the real RUL. From the results, it can be concluded that the method can effectively track the trend of bearing performance degradation.

4. Comparison with Other Related Methods

To further verify the effectiveness of the proposed method, we compared the proposed method with several other typical methods using the same dataset. There is a unified evaluation standard in the IEEE 2012 PHM Challenge. And participants are scored based on their RUL results converted into percent errors. The percent error on each experiment is defined by

$$\%Er_i = 100 \times \frac{RUL_i - \hat{RUL}_i}{RUL_i}. \quad (11)$$

Score of RUL estimation for experiment i is defined as follows:

$$A_i = \begin{cases} \exp\left(-\ln(0.5) \cdot \left(\frac{Er_i}{5}\right)\right), & Er_i \leq 0, \\ \exp\left(\ln(0.5) \cdot \left(\frac{Er_i}{20}\right)\right), & Er_i > 0. \end{cases} \quad (12)$$

Figure 11 depicts the evolution of this scoring function. The RUL score function was estimated based on the percentage error between the predicted value and the actual value; as can be seen, underestimation and overestimation will not be considered in the same way; good estimation performance is related to early predictions of RUL (i.e., $Er_i > 0$), deductions to early deletions, and more severe deductions when the RUL estimation exceeds the actual value (i.e., $Er_i \leq 0$).

The final score of all RUL estimates is defined as the mean of all experiment's score:

$$\text{score} = \frac{1}{5} \sum_{i=1}^5 A_i. \quad (13)$$

Table 4 shows the comparison of results between our proposed method and the long short term memory- (LSTM-) based method [43], multi-CNN-based method [28], sparse representation model-based method [44], and deep belief network-diffusion process- (DBN-DP-) based method [45] for RUL estimation of the same dataset. Current time, actual RUL, estimation error, and mean score of the 5 methods are all presented in the table. Our proposed method achieved a mean score of 0.64, which is significantly higher than the scores of the other 4 methods.

Compared with the method in the literature [43], it can be seen that it is difficult to make high-precision prediction using only the raw signals [28]. Effectively, TFR and CNN

TABLE 3: Performance evaluation with different frame numbers.

	DCT-TACNN/TFR-TACNN				
	$P = 1$	$P = 2$	$P = 3$	$P = 4$	$P = 5$
MAE	90.85/121.68	66.03/100.66	45.23/80.35	53.31/72.12	63.69/79.35
MSE	90.45/105.93	41.79/80.59	36.27/75.02	33.15/70.33	50.10/88.54
MAPE	0.3754/0.5214	0.3266/0.5025	0.1893/0.4621	0.2619/0.4244	0.3028/0.4459

FIGURE 9: Training and testing error over iterations.

FIGURE 10: Continued.

FIGURE 10: RUL estimation results of the totally 5 testing datasets. (a) Bearing 1-3. (b) Bearing 1-4. (c) Bearing 1-5. (d) Bearing 1-6. (e) Bearing 1-7.

are combined to improve the prediction accuracy. However, since it is difficult for CNN to extract the relevant dependencies of signals, the accuracy needs to be improved further. The method proposed in [44] considers this kind of temporal dependence, but the improvement of accuracy is still limited only using sparse features of the signal. Method in the literature [45] used 29 kinds of statistical characteristics of signals as input, whose results are greatly affected by subjective factors since it requires artificial screening and fusion of the characteristics. Based on TFR and CNN, the method proposed in this paper uses the DCT spectrum to characterize time-frequency domain characteristics of signals and improve the sparseness of model inputs. And then, TACNN is used to predict RULs directly. This method can

make full use of the spatial-temporal characteristics and time-dependent information of the data and has achieved good prediction results.

Furthermore, Table 5 shows the time consumption of the training TFR-TACNN model and the proposed DCT-TACNN model with different dimensional input images. The pruned percent in TACNN for sparse calculation is also presented in the table. As can be seen, with the image dimensions decrease, the time spent on model training also decreases. However, as mentioned above, excessively large dimensionality reduction will result in great loss of image information, which is not conducive to the identification of the final result. We only discuss the difference in the results with/without the DCT conversion in the same image

FIGURE 11: Evolution of the scoring function.

TABLE 4: RUL estimation results and comparison.

Testing bearing	Current time (s)	Actual RUL (s)	Er (%)				
			LSTM [43]	MCNN [28]	Sparse R [44]	DBN-DP [45]	Proposed method
1-3	18010	5730	93.02	17.43	4.00	2.26	-1.80
1-4	11380	339	191.18	10.67	-10.00	23.31	20.35
1-5	23010	1610	503.11	-148.19	19.00	-83.23	13.66
1-6	23010	1460	542.47	-19.45	81.00	20.34	16.99
1-7	15010	7570	3.43	19.55	60.00	22.93	-2.10
Score			0.25	0.36	0.37	0.46	0.64

TABLE 5: Time consuming in model training (unit: s).

Input size/ $H \times W \times C$	TFR-TACNN	DCT-TACNN
$500 \times 500 \times 2$	Overflow	598 (pruned 26.5%)
$250 \times 250 \times 2$	1021 (pruned 5.9%)	266 (pruned 20.9%)
$100 \times 100 \times 2$	452 (pruned 10.0%)	72 (pruned 17.5%)
$50 \times 50 \times 2$	231 (pruned 9.5%)	40 (pruned 24.0%)

dimension. The advantages of the proposed method in bearing RUL estimation accuracy and time consumption of the results are obvious. In practical applications, a relatively suitable dimension can be determined according to the characteristics of the data itself, and then the method can be used for RUL prediction analysis.

5. Conclusion

In order to make efficient use of the spatiotemporal continuity characteristics of bearing vibration signals for RUL estimation, a novel prognostic approach based on DCT and TACNN model is put forward. A new data form of temporal DCT continue clip is proposed and applied to effectively characterize the spatiotemporal continuity between adjacent TFRs. To improve the computing efficiency without losing as much data as possible, bilinear interpolation and DCT algorithm is used to convert the WT-TFRs from spatial domain into DCT distribution in frequency domain. The model input changes from a single TFR to a fixed-length DCT continue clips, and TACNN is applied to establish the mathematical connection between the input and RUL. Effectiveness of the proposed method is verified on the PRONOSTIA dataset, and relatively good results at present is obtained, which proved the effectiveness of the combined

DCT and TACNN method. To our best knowledge, this study first leverages DCT continue clips for bearing RUL estimation. Furthermore, for the data input with same dimension, calculation efficiency is improved by about 5 times compared with TFR-CNN-based models and advantages of the proposed method in application are highlighted.

Data Availability

The data used to support the findings of this study are available from <https://www.femto-st.fr/en>.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

Rolling bearing data used in this work was taken from PRONOSTIA platform.

References

- [1] F. Zhang, J. Huang, F. Chu, and L. Cui, "Mechanism and method for outer raceway defect localization of ball bearings," *IEEE Access*, vol. 8, pp. 4351-4360, 2020.
- [2] J. Wang, Y. Liang, Y. Zheng, R. X. Gao, and F. Zhang, "An integrated fault diagnosis and prognosis approach for predictive maintenance of wind turbine bearing with limited samples," *Renewable Energy*, vol. 145, pp. 642-650, 2020.
- [3] J. B. Ali, B. Chebel-Morello, L. Saidi, S. Malinowski, and F. Fnaiech, "Accurate bearing remaining useful life prediction based on weibull distribution and artificial neural network,"

- Mechanical Systems and Signal Processing*, vol. 56–57, pp. 150–172, 2015.
- [4] F. Wang, X. Liu, C. Liu, H. Li, and Q. Han, “Remaining useful life prediction method of rolling bearings based on Pchip-EEMD-GM (1, 1) model,” *Shock and Vibration*, vol. 2018, Article ID 3013684, 10 pages, 2018.
 - [5] M. Elforjani and S. Shanbr, “Prognosis of bearing acoustic emission signals using supervised machine learning,” *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5864–5871, 2018.
 - [6] Y. Wang, Y. Peng, Y. Zi, X. Jin, and K.-L. Tsui, “A two-stage data-driven-based prognostic approach for bearing degradation problem,” *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 924–932, 2016.
 - [7] L. Zhang, J. Lin, B. Liu, Z. Zhang, X. Yan, and M. Wei, “A review on deep learning applications in prognostics and health management,” *IEEE Access*, vol. 7, pp. 162415–162438, 2019.
 - [8] J. Chen, H. Jing, Y. Chang, and Q. Liu, “Gated recurrent unit based recurrent neural network for remaining useful life prediction of nonlinear deterioration process,” *Reliability Engineering & System Safety*, vol. 185, pp. 372–382, 2019.
 - [9] L. Guo, N. Li, F. Jia, Y. Lei, and J. Lin, “A recurrent neural network based health indicator for remaining useful life prediction of bearings,” *Neurocomputing*, vol. 240, pp. 98–109, 2017.
 - [10] N. H. Chandra and A. S. Sekhar, “Fault detection in rotor bearing systems using time frequency techniques,” *Mechanical Systems and Signal Processing*, vol. 72–73, pp. 105–133, 2016.
 - [11] R. Liu, B. Yang, X. Zhang, S. Wang, and X. Chen, “Time-frequency atoms-driven support vector machine method for bearings incipient fault diagnosis,” *Mechanical Systems and Signal Processing*, vol. 75, pp. 345–370, 2016.
 - [12] H. Huang, N. Baddour, and M. Liang, “Bearing fault diagnosis under unknown time-varying rotational speed conditions via multiple time-frequency curve extraction,” *Journal of Sound and Vibration*, vol. 414, pp. 43–60, 2016.
 - [13] X. Lou and K. A. Loparo, “Bearing fault diagnosis based on wavelet transform and fuzzy inference,” *Energy*, vol. 18, no. 5, pp. 1077–1095, 2004.
 - [14] J. Shi, G. Du, R. Ding, and Z. Zhu, “Time frequency representation enhancement via frequency matching linear transform for bearing condition monitoring under variable speeds,” *Applied Sciences*, vol. 9, no. 18, p. 3828, 2019.
 - [15] W. Huang, G. Gao, N. Li, X. Jiang, and Z. Zhu, “Time-frequency squeezing and generalized demodulation combined for variable speed bearing fault diagnosis,” *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 8, pp. 2819–2829, 2019.
 - [16] X. Zhang, Z. Liu, J. Wang, and J. Wang, “Time-frequency analysis for bearing fault diagnosis using multiple Q-factor gabor wavelets,” *ISA Transactions*, vol. 87, pp. 225–234, 2019.
 - [17] A. Brkovic, D. Gajic, J. Gligorijevic, I. S. Gajic, O. Georgieva, and S. D. Gennaro, “Early fault detection and diagnosis in bearings for more efficient operation of rotating machinery,” *Energy*, vol. 136, no. 1, pp. 63–71, 2017.
 - [18] J. Chen, J. Pan, Z. Li, Y. Zi, and X. Chen, “Generator bearing fault diagnosis for wind turbine via empirical wavelet transform using measured vibration signals,” *Renewable Energy*, vol. 89, pp. 80–92, 2016.
 - [19] H. D. M. D. Azevedo, A. M. Araújo, and N. Bouchonneau, “A review of wind turbine bearing condition monitoring: state of the art and challenges,” *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 368–379, 2016.
 - [20] B. Wu, W. Li, and M.-Q. Qiu, “Remaining useful life prediction of bearing with vibration signals based on a novel indicator,” *Shock and Vibrations*, vol. 2017, Article ID 8927937, 10 pages, 2017.
 - [21] G.-Q. Hou and C.-M. Lee, “Estimation of the defect width on the outer race of a rolling element bearing under time-varying speed conditions,” *Shock and Vibration*, vol. 2019, Article ID 8479395, 11 pages, 2019.
 - [22] A. T. Jahromi, M. J. Er, X. Li, and B. S. Lim, “Sequential fuzzy clustering based dynamic fuzzy neural network for fault diagnosis and prognosis,” *Neurocomputing*, vol. 196, pp. 31–41, 2016.
 - [23] J. B. Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, and F. Fnaiech, “Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals,” *Applied Acoustics*, vol. 89, pp. 16–27, 2015.
 - [24] S. Kiakojoori and K. Khorasani, “Dynamic neural networks for gas turbine engine degradation prediction, health monitoring and prognosis,” *Neural Computing and Applications*, vol. 27, no. 8, pp. 2157–2192, 2016.
 - [25] L. Jing, M. Zhao, P. Li, and X. Xu, “A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox,” *Measurement*, vol. 111, pp. 1–10, 2017.
 - [26] D. Belmiloud, T. Benkedjouh, M. Lachi, A. Laggoun, and J. P. Dron, “Deep convolutional neural networks for bearings failure prediction and temperature correlation,” *Journal of Vibroengineering*, vol. 20, no. 8, pp. 2878–2891, 2018.
 - [27] L. Ren, Y. Sun, H. Wang, and L. Zhang, “Prediction of bearing remaining useful life with deep convolution neural network,” *IEEE Access*, vol. 6, pp. 13041–13049, 2018.
 - [28] J. Zhu, N. Chen, and W. Peng, “Estimation of bearing remaining useful life based on multiscale convolutional neural network,” *IEEE Transactions on Industrial Electronics*, vol. 66, no. 4, pp. 3208–3216, 2019.
 - [29] J. Li, X. Li, and D. He, “A directed acyclic graph network combined with CNN and LSTM for remaining useful life prediction,” *IEEE Access*, vol. 7, pp. 75464–75475, 2019.
 - [30] B. Yang, R. Liu, and E. Zio, “Remaining useful life prediction based on a double-convolutional neural network architecture,” *IEEE Transactions on Industrial Electronics*, vol. 66, no. 12, pp. 9521–9530, 2019.
 - [31] S. Zhang, H. Peng, J. Fu, and J. Luo, “Learning 2d temporal adjacent networks for moment localization with natural language,” in *Proceeding of The Thirty-Fourth AAAI Conference on Artificial Intelligence*, New York, NY, USA, February 2020.
 - [32] Ł. Jedliński and J. Jonak, “Early fault detection in gearboxes based on support vector machines and multilayer perceptron with a continuous wavelet transform,” *Applied Soft Computing*, vol. 30, pp. 636–641, 2015.
 - [33] T. Benkedjouh, N. Zerhouni, and S. Rechak, “Tool wear condition monitoring based on continuous wavelet transform and blind source separation,” *The International Journal of Advanced Manufacturing Technology*, vol. 97, no. 9–12, pp. 3311–3323, 2018.
 - [34] N. Ahmed, T. Natarajan, and K. R. Rao, “Discrete cosine transform,” *IEEE Transactions on Computers*, vol. C-23, no. 1, pp. 90–93, 1974.
 - [35] W.-H. Chen, C. Smith, and S. Fralick, “A fast computational algorithm for the discrete cosine transform,” *IEEE Transactions on Communications*, vol. 25, no. 9, pp. 1004–1009, 1977.

- [36] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: a review," *Neurocomputing*, vol. 187, pp. 27–48, 2016.
- [37] S. Liao, J. Wang, R. Yu, K. Sato, and Z. Cheng, "CNN for situations understanding based on sentiment analysis of twitter data," *Procedia Computer Science*, vol. 111, pp. 376–381, 2017.
- [38] H. Valavi, P. J. Ramadge, E. Nestler, and N. Verma, "A 64-tile 2.4 mb in-memory-computing CNN accelerator employing charge-domain compute," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 6, pp. 1798–1799, 2019.
- [39] J. Chen, Q. Zhang, W.-S. Zheng, and X. Xie, "Efficient and switchable CNN for crowd counting based on embedded terminal," *IEEE Access*, vol. 7, pp. 51533–51541, 2019.
- [40] S. Kala, B. R. Jose, J. Mathew, and S. Nalesh, "High-performance CNN accelerator on FPGA using unified winograd-GEMM architecture," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 27, no. 12, pp. 2816–2828, 2019.
- [41] S. T. H. Rizvi, G. Cabodi, and G. Francini, "Optimized deep neural networks for real-time object classification on embedded GPUs," *Applied Sciences*, vol. 7, no. 8, p. 826, 2017.
- [42] P. Nectoux, R. Gouriveau, K. Medjaher et al., "PRONOSTIA: an experimental platform for bearings accelerated degradation tests," in *Proceeding of the IEEE International Conference on Prognostics and Health Management, PHM'12*, pp. 1–8, Denver, CO, USA, June 2012.
- [43] B. Zhang, S. Zhang, and W. Li, "Bearing performance degradation assessment using long short-term memory recurrent network," *Computers in Industry*, vol. 106, pp. 14–29, 2019.
- [44] L. Ren, W. Lv, and S. Jiang, "Machine prognostics based on sparse representation model," *Journal of Intelligent Manufacturing*, vol. 29, no. 2, pp. 277–285, 2018.
- [45] C.-H. Hu, H. Pei, X.-S. Si, D.-B. Du, Z.-N. Pang, and X. Wang, "A prognostic model based on DBN and diffusion process for degrading bearing," *IEEE Transactions on Industrial Electronics*, p. 1, 2019.