ISA Transactions xxx (xxxx) xxx

Contents lists available at ScienceDirect

ISA Transactions

journal homepage: www.elsevier.com/locate/isatrans

Research article

Intelligent fault diagnosis of helical gearboxes with compressive sensing based non-contact measurements

Xiaoli Tang [a], Yuandong Xu [b,*], Xiuquan Sun [c], Yanfen Liu [d], Yu Jia [a], Fengshou Gu [c], Andrew D. Ball [c]

[a] School of Engineering and Technology, Aston University, Birmingham B4 7ET, UK
[b] Dynamics Group, Department of Mechanical Engineering, Imperial College London, London SW7 2AZ, UK
[c] Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield HD1 3DH, UK
[d] College of mechanical and electrical engineering, Qingdao Agricultural University, Qingdao 266109, China

ARTICLE INFO

Article history:
Received 3 February 2022
Received in revised form 25 May 2022
Accepted 16 July 2022
Available online xxxx

Keywords:
Gearbox fault diagnosis
Non-contact measurements
Modulation signal bispectrum
Compressive sensing
Dual-channel CNN

ABSTRACT

Helical gearboxes play a critical role in power transmission of industrial applications. They are vulnerable to various faults due to long-term and heavy-duty operating conditions. To improve the safety and reliability of helical gearboxes, it is necessary to monitor their health conditions and diagnose various types of faults. The conventional measurements for gearbox fault diagnosis mainly include lubricant analysis, vibration, airborne acoustics, thermal images, electrical signals, etc. However, a single domain measurement may lead to unreliable fault diagnosis and the contact installation of transducers is not always accessible, especially in harsh and dangerous environments. In this article, a Compressive Sensing (CS)-based Dual-Channel Convolutional Neural Network (CNN) method was proposed to accurately and intelligently diagnose common gearbox faults based on two complementary non-contact measurements (thermal images and acoustic signals) from a mobile phone. The raw acoustic signals were analysed by the Modulation Signal Bispectrum (MSB) to highlight the coupled modulation components relating to gear faults and suppress the irrelevant components and random noise, which generates a series of two-dimensional matrices as sparse MSB magnitude images. Then, CS was used to reduce the image redundancy but retain key information owing to the high sparsity of thermal images and acoustic MSB images, which significantly accelerates the CNN training speed. The experimental results convincingly demonstrate that the proposed CS-based Dual-Channel CNN method significantly improves the diagnostic accuracy (99.39% on average) of industrial helical gearbox faults compared to the single-channel ones.

1. Introduction

Helical gearboxes are one of the crucial parts of rotating machinery for power transmission. The long-term operation of gearboxes under harsh conditions like heavy loads can lead to various failures and shorten their service life, which can result in significant economic loss and even catastrophic accidents. Online real-time condition monitoring and fault diagnosis of gearboxes can greatly increase reliability, productivity and safety.

The conventional measurements used for gearbox fault diagnosis mainly include lubricant analysis, vibration, airborne acoustics, acoustic emission, strains, electrical signals, encoder signals, temperature, pressure and so on [1,2]. However, most sensing methods are invasive or intrusive, and even require long downtime for installation and testing. In harsh operating environments, the installation of sensors is complicated, inconvenient, inflexible and costly. Therefore, non-contact sensors attract more and more attention in machine condition monitoring. In general, non-contact instruments like high-speed cameras, laser Doppler vibrometers, eddy current sensors, and fibre-optic sensors are of high-costs [3].

Thermal imaging, known as infrared thermography, is an advanced non-contact and non-instructive sensing technique. Thermal cameras detect and capture infrared radiation emitted from the object to create a temperature representation as a thermal image of an object. It has been widely applied for condition monitoring in various fields [4]. As most common gear faults lead to abnormal friction between mating components, it can cause the alteration of heat characteristics such as heat sources, heat transfer paths and temperature distribution. Thus, the infrared radiation from the gearbox captured by the thermal camera

* Corresponding author.
E-mail addresses: x.tang4@aston.ac.uk (X. Tang), y.xu@imperial.ac.uk (Y. Xu), x.sun2@hud.ac.uk (X. Sun), liu_yanfen@163.com (Y. Liu), y.jia1@aston.ac.uk (Y. Jia), f.gu@hud.ac.uk (F. Gu), a.ball@hud.ac.uk (A.D. Ball).

https://doi.org/10.1016/j.isatra.2022.07.020

Please cite this article as: X. Tang, Y. Xu, X. Sun et al., Intelligent fault diagnosis of helical gearboxes with compressive sensing based non-contact measurements. ISA Transactions (2022), https://doi.org/10.1016/j.isatra.2022.07.020.

contains valuable information to support high-performance fault diagnosis of gearboxes [5,6]. However, only a few researchers made contributions to the fault diagnosis of gearboxes with thermal imaging. Li et al. [7,8] respectively fed the infrared thermal images or multi-scale images into the CNN model to detect faults of gearboxes. Kumar et al. [9] selected the strongest features from the 2-D discrete wavelet transform (WT) decomposition of thermal images using Mahalanobis distance criteria to classify different gear faults with a support vector machine (SVM) classifier. In [10], the authors examined the worm gear faults with vibration and sound signals by a multilayer perceptron artificial neural network (ANN) model. Additionally, they observed the heating patterns and predicted the operating speed and oil level with thermal imaging.

Another effective non-contact measurement for gearbox fault diagnosis is the acoustic signal. For instance, Preanesh et al. [11] and Parey et al. [12] extracted features from the time-domain acoustic sequences and angular domain acoustic signals using WTto diagnose gearbox faults, respectively. Hartono et al. [13] proposed an improved Reassigned Smoothed Pseudo Wigner–Ville Distribution and validated its effectiveness with both vibration and acoustic measurements to diagnose the spur gearbox faults. The extracted statistical features of vibration and acoustic signals along with psychoacoustic features of acoustic signals were applied for gear fault diagnosis using intelligent techniques [14]. It was found that the psychoacoustic features of acoustic signals have a better ability to classify faults compared to the statistical features. Vanraj et al. [15] proposed a statistical feature extraction method based on empirical mode decomposition and Teager–Kaiser energy operator to diagnose the severity of gear faults. The comprehensive features of energy, and time and envelope spectrum kurtosis were extracted from acoustic signals for planetary gearbox fault diagnosis which has better performance than the vibration signals in [16]. Fernandez et al. [17] used acoustic pressure spectra and intensity maps to diagnose the planetary gear faults of oil shortage and tooth crack. Lu et al. [18] applied the MSB method to acoustic signals for gear wear diagnosis and demonstrated the MSB magnitude peaks related to the first three harmonics are effective to detect the gradual deterioration of helical gears. Owing to the effectiveness of thermal imaging and acoustic signals, these two non-contact measurements are employed to monitor the health condition of gearboxes in this study.

With the development of deep learning techniques, they are widely applied for gearbox fault diagnosis to automatically handle large amounts of data sets in industrial applications. For example, Jing et al. [19] compared the diagnostic efficiency of three common signals with CNN and demonstrated that frequency data had higher diagnostic accuracy in gearbox faults. Abdul et al. [20] proposed to detect gear faults with the long short-term memory classifier based on the Gamma Tone Cepstral and the Mel-Frequency Cepstral Coefficients (MFCC). Li et al. [21] developed an MFCC based parallel multi-fusion CNN model for bearing and gearbox fault detection in noisy environments because it effectively reduces the noise and retains low-frequency features from vibration signals. Liang et al. [22] investigated a method that generated time–frequency image features and fake samples using WT and generative adversarial net to train a CNN model to achieve stable detection of weak bearing faults and compound gear faults. Huang et al. [23] decomposed vibration signals from gearboxes using the WT technique, then used 1-D CNN to extract multiscale features to achieve high-performance fault diagnosis. Chang et al. [24] applied the CNN on the geographical images of gear surfaces to intelligently diagnose gear wear faults and severity levels. Cao et al. [25] employed variational mode decomposition on image denoising for oil debris images for planetary gearbox fault detection, but they also proposed to combine it with deep neural networks in future work. In [26], a novel attentive kernel residual network was proposed to automatically extract multiscale features and improve their performance from the noisy vibration signals for gearbox fault diagnosis. Yan et al. [27] fused spectrum amplitudes to obtain a health index for synchronised early fault diagnosis and degradation assessment of gears and bearings.

Considering the instability and unreliability of a single source, sensor fusion of multiple measurements has attracted great attention from researchers in gearbox condition monitoring recently. Generally, sensor fusion is performed with the measurements from several sensors of the same type (i.e. multi-sensor data) or several different types of sensors (i.e. multi-source data). The fused signals or features are further combined with deep learning methods for intelligent fault diagnosis [28]. For example, Xia et al. [29] arranged raw data from multiple vibration sensors as a 2-D matrix to preserve both temporal and spatial information, then feed it into CNN to achieve accurate fault diagnosis of bearings and gearboxes. Wang et al. [30] proposed to convert multi-sensor vibration signal segments to 2-D feature maps and improved the CNN model by adding the stacking bottleneck layers to implement classification accuracy of 99.89% of gearbox faults. Azamfar et al. [31] transformed the multi-sensor motor current signals into spectra and stacked them into a 2-D matrix as the inputs of 2D-CNN to diagnose the gearbox faults. Shao et al. [32] proposed a stacked wavelet auto-encoder model to achieve high-performance prediction of gearbox faults in data-, feature-level and decision-level fusion with multi-sensor vibration signals. They also investigated to utilise the feature-fusion covariance matrix to fuse statistical features of multi-sensor vibration signals and built a multi-Riemannian kernel ridge regression model to implement excellent gear fault diagnosis with two cases [33]. In [34], a residual gated dynamic sparse network was investigated based on multi-sensor vibration signals to not only improve feature extraction but also resist the effects of high noise interference. Zhao et al. [35] proposed a novel model-driven deep unrolling method to tackle the “black box” issue of the neural networks and noise attacks of vibration signals. Its effectiveness was evaluated by successfully diagnosing the aero-engine bevel gear faults and helical gear faults with multiple accelerometers. Some researchers focus on the fusion of different types of signals, especially vibrational and other types of signals. Li et al. used the fusion of vibration signals with acoustic emission [36] and motor current signals [37], respectively, combined with improved deep random forest and autoencoder to realise good performance in gearbox fault diagnosis. Ma et al. [38] developed a deep coupling autoencoder model to extract the joint information from vibrational and acoustic signals to effectively classify gears faults and bearing faults, respectively. In [39], the authors investigated a deep coupled dense convolutional model to improve the ability to fuse both self and mutual information of vibration and encoder signals for effective planetary gearbox fault diagnosis. Mao et al. [40] converted vibration signals into spectra and combined them with the thermal images at the data-level fusion. The fused signals were fed into a CNN for intelligent helical gearbox fault diagnosis. Karabacak et al. [41] proposed to merge the extracted statistical time and frequency features of vibration and acoustic signals, together with the extracted statistical features of thermal images at the feature-level fusion, and then detect worm gearbox faults with the ANN and SVM respectively to achieve the accuracy of 99.2% with ANN and 98.7% with SVM. CNN becomes one of the most widely used structures in deep learning because it has a strong ability to automatically learn effective fault features using convolutions from the input signals to avoid manual feature extraction and minimise the requirement for professional knowledge. In addition, CNN has a high accuracy for image recognition

problems, which can fully explore the effective fault signatures from images. Additionally, the weight sharing of CNN reduces the optimisation time and memory resources. To fully exploit the advantages of CNN, acoustic signals are pre-processed by the MSB to form a series of MSB images. The thermal images and MSB images generated from acoustic signals are utilised as the dual-channel inputs of CNN module and properly fused at the feature level to effectively diagnose gearbox faults in this study.

In general, 2-D images occupy a large transmission bandwidth and storage space and critically slow down the computing and training speed, which will affect the efficiency of intelligent online condition monitoring. It is necessary to compress the images but retain key fault information before feeding them into the training network. Therefore, compressive sensing is introduced to compress the substantial training images to increase the computational efficiency of the CNN. The CS method is a novel sample technique that can break through the Nyquist–Shannon sampling theorem and achieve pre-acquisition compression with cost-effective hardware directly, which is the unique merit of CS-based fault diagnosis approaches. As an advanced compression algorithm, CS implements random sampling instead of the uniform one and can reconstruct the original signal with few samples if the original signal is sparse or compressible. It has been widely applied in signal processing to reduce the number of samples for machine condition monitoring by some researchers [42,43]. Although CS can achieve pre-acquisition compression before data transmission at the data acquisition end, it is still a crucial challenge to practically implement machine condition monitoring currently [44]. Lu et al. [45] proposed a novel approach to achieve physics-based CS with fewer sensors and limited collected data, which was applied in manufacturing process monitoring with fewer costs and high efficiency.

With the rapid development of the smart industry, non-contact measurements can dominate the machine fault diagnosis in the industry because the non-contact sensors can be carried by robots or drones to inspect large quantities of machines, for instance, wind farms. In this article, a CS-based Dual-Channel CNN method is proposed to intelligently diagnose faults of gearboxes through low-cost non-contact measurements, i.e. thermal images and acoustic signals. The main contributions of this article include:

(1) Using two complementary non-contact measurements from a mobile phone to monitor gearbox health conditions effectively and cost-effectively, which not only overcomes the instability and inaccuracy of a single domain signal but also avoids sensor installation issues in harsh environments.
(2) Reducing the redundancy of thermal and MSB images to accelerate the training speed through CS, which can break through the Nyquist–Shannon sampling theorem and achieve pre-acquisition compression with hardware in the future.
(3) The proposed CS-based Dual-Channel CNN method fuses two complementary measurements to achieve the accurate and efficient diagnosis of various common gearbox faults under wide operating conditions.

The rest of this article is organised as follows. Section 2 introduces the proposed CS-based Dual-Channel CNN method. Section 3 describes the experimental studies. The fault diagnosis results are discussed and compared in Section 4. Finally, the conclusions are remarked.

2. Compressive sensing-based Dual-Channel CNN

2.1. Basic theories for signal pre-processing

2.1.1. Modulation signal bispectrum

A conventional second order spectrum analysis method, named bispectrum, is defined as [46]

$$B_{CB}(f_1, f_2) = E\left\langle X(f_1)X(f_2)X^*(f_1+f_2)\right\rangle \tag{1}$$

in which $(f) = \int_{-\infty}^{+\infty} x(t)\, e^{-i2\pi ft} dt$ is the Fourier transform of a time-domain signal $x(t)$; f_1 and f_2 are two frequency components in $X(f)$; $X^*(f)$ is the complex conjugate of $X(f)$; and $E\langle\cdot\rangle$ denotes the ensemble averaging. The conventional bispectrum is a complex third order measurement of the coupled components. The method of MSB developed by Gu et al. [47–49] originated from the conventional bispectrum by introducing both lower and upper sidebands in the modulation signals. MSB is an efficient method to extract modulation characteristics from dynamic signals of the rotating machinery. Its performance on motor current signals for gear wear detection has been validated in [50]. Airborne acoustic signals of gearboxes are easily contaminated by strong random noise and the MSB method can specifically extract nonlinear modulation features from acoustic signals by reducing the nonrelevant components and random noise simultaneously [18]. Therefore, it is employed to process the acoustic signals from gearboxes, which generates a series of 2-D matrices of sparse modulation characteristics in this study. The extracted modulation characteristics are enhanced but still challenging for straightforward fault diagnosis, which requires sufficient professional background knowledge to understand the useful information. The MSB is defined by the formula

$$B_{MS}(f_x, f_c) = E\left\langle X(f_c+f_x)\, X(f_c-f_x)\, X^*(f_c)\, X^*(f_c)\right\rangle \tag{2}$$

where, f_x is the modulation frequency, f_c is the carrier frequency, $f_c + f_x$ and $f_c - f_x$ are the upper and lower sideband frequencies respectively, $E\langle\cdot\rangle$ represents the ensemble averaging of MSB slices. Each MSB slice $X(f_c+f_x)\, X(f_c-f_x)\, X^*(f_c)\, X^*(f_c)$ can be calculated using a short segment of the raw vibroacoustic signals to highlight the modulation components in the acoustic signals. Alternatively, Eq. (2) can be re-written by the magnitude and phase, yielding

$$B_{MS}(f_x, f_c) = E\left\langle |X(f_c+f_x)|\, |X(f_c-f_x)|\, \left|X^*(f_c)\right|\, \left|X^*(f_c)\right|\, e^{j\phi_{MS}}\right\rangle \tag{3}$$

where the total phase of each MSB slice is

$$\phi_{MS} = \phi(f_c+f_x) + \phi(f_c-f_x) - \phi(f_c) - \phi(f_c) \tag{4}$$

with

$$\begin{aligned} \phi(f_c+f_x) &= \phi(f_c) + \phi(f_x) \\ \phi(f_c-f_x) &= \begin{cases} \phi(f_c) - \phi(f_x) & , AM\ signals \\ \phi(f_c) - \phi(f_x) + \pi & , PM\ signals \end{cases} \end{aligned} \tag{5}$$

Therefore, the total phase ϕ_{MS} of modulation frequencies in the MSB is constant for each slice. Consequently, the ensemble averaging enhances the modulation characteristics and suppresses the noncoupled components and background noise for fault detection and diagnosis.

The dynamic responses of the gearboxes with defects show strong modulation phenomena, including meshing frequency modulation and resonant modulation. In conventional fault detection and diagnosis analysis, the fault signatures are obtained via demodulation analysis. The fault features are difficult to extract in the time domain for achieving effective fault diagnostics. Consequently, the fault signatures in the frequency domain are preferable for the health monitoring of gearboxes. The MSB is a reputational approach in demodulation analysis based fault

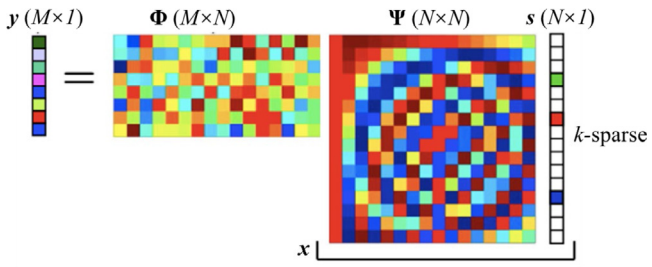

Fig. 1. Matrix representation of the signal compression with CS.

detection and diagnosis of rotating machines. However, due to the intricate modulation mechanisms, exploration of effective fault signatures in a wide frequency range is always a challenging task. The machine learning technique allows searching for the optimal fault signatures from a series of informative MSB bispectrum results in a wide bifrequency range. The MSB results are multiple two-dimensional matrices that can be easily represented by images for intelligent fault diagnosis. In addition, the understanding of the extracted modulation bispectrum requires sufficient professional knowledge of signal processing, gear dynamics and acoustic signals, which is unable to form a readable representation for general users. The machine learning based approach can intelligently analyse the results and generate an easy-understanding and concise diagnosis result.

2.1.2. Compressive sensing

Compressive sensing implements randomly sampling instead of the uniform one. It can reconstruct the raw signals with far fewer samples than required by the Nyquist–Shannon sampling theorem through optimisation if the raw signals are sparse or compressible [51]. Moreover, it has the potential to achieve pre-acquisition compression before data transmission to reduce the data storage and capacity in the field of machine condition monitoring.

Suppose $x(n), n = 1, 2, \ldots, N$ is a raw discrete time signal with the length of N. It can be represented by a basis matrix of $\Psi = [\psi_1, \psi_2, \ldots, \psi_N]$, i.e.

$$\boldsymbol{x} = \sum_{i=1}^{N} s_i \psi_i \text{ or } \boldsymbol{x} = \boldsymbol{\Psi s} \tag{6}$$

where $\psi_i, i = 1, 2, \ldots, N$ are the basis vectors with the size of $N \times 1$; $s_i, i = 1, 2, \ldots, N$ are the weighting coefficients and $\boldsymbol{s}$ is the coefficient vector, which only contains $k\,(k \ll N)$ nonzero values, called k-sparse. Ψ is also called the transformation matrix, commonly adopting the discrete cosine transform, the discrete Fourier transform, discrete wavelet transform and so on. This is the process of sparse representation of the raw signal.

The sparse signal can be compressed by a random matrix $\boldsymbol{\Phi}$ with the size of $M \times N$ to get the compressed signal $\boldsymbol{y}$ only with the length of M, which can be expressed as

$$\boldsymbol{y} = \boldsymbol{\Phi x} = \boldsymbol{\Phi \Psi s} \tag{7}$$

To intuitively observe the compression process of a discrete-time signal using CS, a matrix representation example is displayed in Fig. 1. A raw signal $\boldsymbol{x}$ is represented with a sparse signal $\boldsymbol{s}$ with the length of N, then compressed to be the signal $\boldsymbol{y}$ with the length of M. The compression ratio (CR) is defined as the ratio of the length or size of the raw signal $\boldsymbol{x}$ to that of the compressed signal $\boldsymbol{y}$ as shown in Eq. (8).

$$CR = N/M \tag{8}$$

In this study, since both thermal images and MSB images are sparse and compressible due to their redundancy, CS can be applied to significantly reduce the image storage space and speed up numerical calculations in the process of machine learning.

2.2. Proposed CS-based Dual-Channel CNN

A convolutional neural network typically consists of an input layer, a hidden layer and an output layer, which will continuously extract features from local to global through various convolutional filters to implement object recognition or classification [52]. It is especially suitable for image processing because it has high fault tolerance, parallel processing capabilities and self-learning abilities. Generally, 2-D or 3-D image sequences can be the input signals fed into the input layer. The hidden layers typically consist of one or multiple convolutional layers, pooling layers and fully connected layers. The convolutional layers extract features, such as edges, curves and so on, from the input images through different filters. After the convolution, a nonlinear activation function called rectified linear unit (ReLU) can follow to maintain the nonlinear properties of the extracted features. Additionally, a batch normalisation layer can be added between the convolutional layer and the ReLU layer to normalise a mini-batch of data across all observations for each channel independently. This action can speed up training and produce a reliable model. Pooling is a dimension reduction process to effectively reduce the computational complexity and only preserves important information in the extracted features in the pooling layer. Max pooling and average pooling layers are two types of commonly used pooling layers. Dropouts are the regularisation technique to randomly switch some percentage of neurons of the network to prevent overfitting issues. Afterwards, the fully connected layer reassembles the obtained features through a weight matrix, and then a SoftMax function normalises the reassembled results to corresponding probabilities with the sum of one. Finally, the output layer generates classification labels with probability based on extracted features.

Although the CNN extracts the features by training the network instead of manual extraction, the original image size is still too large which slows down the calculation speed and consumes many computing resources. Fortunately, both the thermal grayscale images and the generated acoustic MSB images are sparse and compressible because the background areas are redundant. Therefore, the images can be compressed using CS before feeding them into the input layer of the CNN. With two complementary non-contact measurements, a CS-based dual-channel CNN method is proposed to classify different gearbox faults as shown in Fig. 2.

The proposed CS-based dual-channel CNN method can be detailed as follows.

(1) Image preparing

In channel 1 (denoted as CH1), the captured thermal images are the RGB images with the dimensions of $1080 \times 1440 \times 3$. They are converted into grayscale images with the size of 1080×1440 to reduce the channel number. Then, the histogram equalisation is calculated to enhance the contrast and strengthen the edge information of the images. In channel 2 (denoted as CH2), the acoustic signals separated from the recorded thermal videos are divided into segments. The number of the sequence segments for each acoustic signal sequence is the same as the thermal frames, i.e. 800. Since the acoustic signals of gearboxes are easy to be affected by the ambient interference noise, it is hard to diagnose the gearbox faults with the acoustic sequence directly. The MSB method has been verified to be effective in gearbox fault diagnosis by obtaining nonlinear modulation fault features from the acoustic signals. Therefore, the acoustic MSB magnitudes of

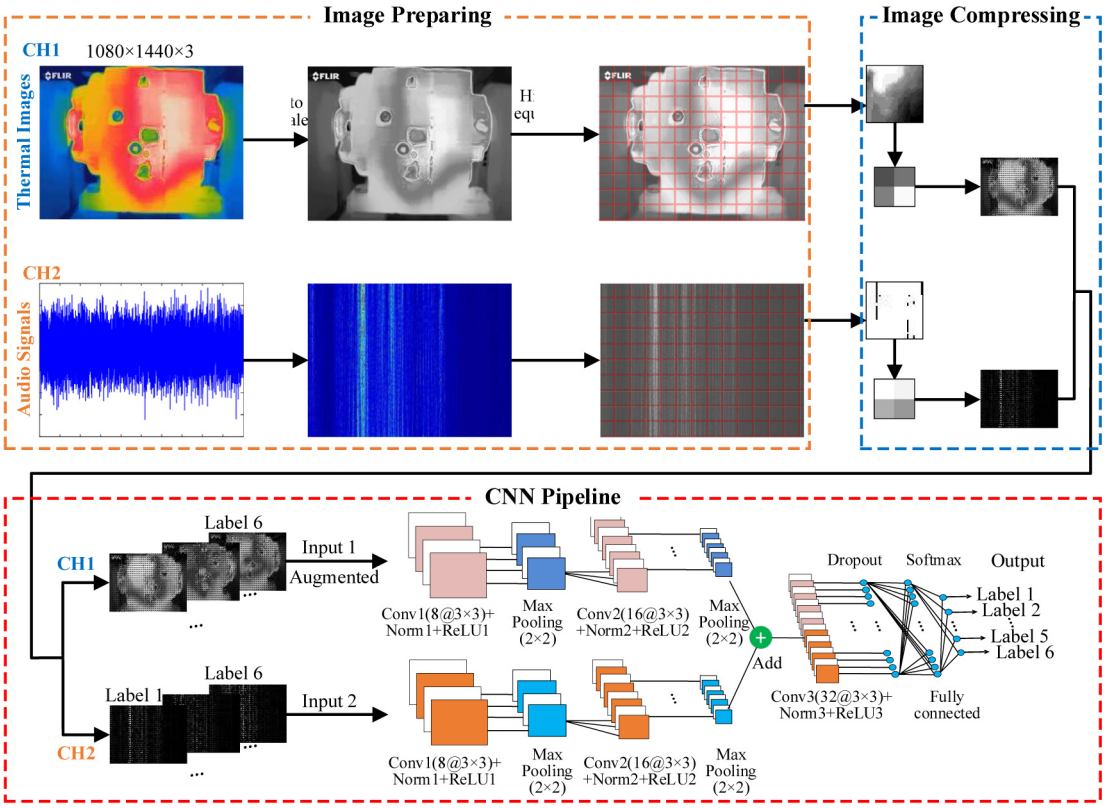

Fig. 2. Flow chart of the proposed CS-based dual-channel CNN method.

each signal segment are calculated and then the MSB magnitude matrices were saved as RGB images. Therefore, the number of the generated MSB frames and their dimensions are the same as the thermal frames. The colourful MSB images are converted to grayscale images to reduce the size to 1080 × 1440 as well.

(2) Image compressing

Because of the large size of the original grayscale images, the network training has a significantly high requirement on the computing resources and inevitably consumes more computing time. CS as an advanced compression method has the potential to break through the Nyquist–Shannon sampling theorem and achieve pre-acquisition compression before data transmission if the signal is sparse or compressible. Therefore, CS is applied to compress the grayscale images to speed up the calculation of network training in this study for the sake of the future CS-based sensing technique.

Each grayscale image from both channels is divided into k small blocks of $n \times n$ pixels. Every block can be reshaped to a column with the length of n^2. Then, the column is compressed to a very short column only with the length of m^2 using a random matrix with the size of $m^2 \times n^2$ ($m < n$) as shown in Eq. (7). Therefore, the compression ratio CR = n^2/m^2. The compressed column is reshaped to a compressed block with the size of $m \times m$. It is followed by the block reassembly to reconstruct a compressed image using these k compressed blocks based on the original image size and its division rules.

In order to determine the optimal block size and the CR, several different sets of parameters were selected empirically to compare the compression performance as shown in Fig. 3. The first row, i.e. Fig. 3(a), displays the compressed images at CR = 18 × 18 and the block size of 18 × 18, 36 × 36, 72 × 72, 90 × 90, 180 × 180 from left to right, respectively. t_{CS} marked below the image is the compression time of the corresponding image. It can be seen the compression time decreases as the block size increases because the number of the blocks decreases. But it increases when the block size is too big which is caused by the operation of large matrices. The full-scale compression (block size and CR are both 18 × 18) and too big block size (block size is 180 × 180) are ineligible because the image is critically distorted after compression and the compression speed is also slow. Although the compression costs less time when the block size is 72 × 72 or 90 × 90, the edge sharpness of the compressed images is much lower than the 36 × 36 block due to the relatively large block size. Therefore, 36 × 36 is an optimal block size in this study. The second row, i.e. Fig. 3(b), shows the compressed images at the block size of 36 × 36 and CR = 2 × 2, 3 × 3, 6 × 6, 9 × 9, 18 × 18 from left to right, respectively. It is apparent that the compression time decreases with the increase of the CR. Although the texture of the image is finer when the CR is small, its brightness and contrast are a little worse than the image compressed at CR = 18 × 18. Therefore, 18 × 18 was selected as the compression ratio in this study.

Consequently, all images from these two channels are compressed as follows. The original grayscale image with the size of 1080 × 1440 can be divided into 1200 blocks in size of 36 × 36 pixels. Each block is reshaped to a column with the length of 1296, then compressed by 324 times to get a new column only with the length of 4 which will be further reshaped to a compressed 2 × 2 block. All 1200 compressed blocks are applied to reassemble a compressed grayscale image with the size of 60 × 80 according to the pixel aspect ratio of the original grayscale image.

(3) CNN pipeline

All the reassembled thermal and acoustic MSB images are labelled according to six different cases. Additionally, the position or direction of the thermal camera may change for different scenarios, which leads to motion, scaling and other forms of distortion of images. Some augmenters including random rotation

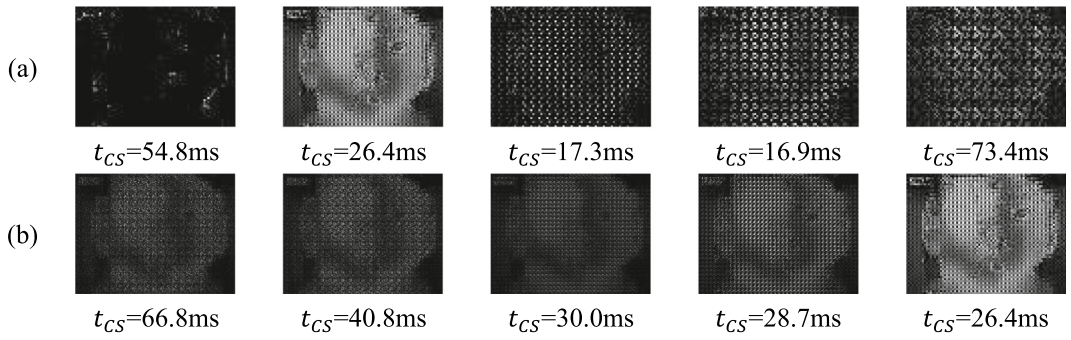

Fig. 3. Comparison of the compressed images using CS at different block sizes and CR: (a) compressed image row at CR = 18 × 18 and different block sizes (left to right: 18×18, 36×36, 72×72, 90×90, 180×180), and (b) compressed image row at a block size of 36 × 36 and different CR (left to right: CR = 2×2, 3×3, 6×6, 9×9, 18×18).

of [−30 30], random translation of [−5 5] and random reflection in both directions are added to the thermal images in CH1 before training to tackle the overfitting issues and accommodate the location variation of thermal cameras. Then, the thermal images and the acoustic MSB images are fed into the input layers from two channels, respectively. Each channel has two sets of a convolution layer with a filter size of 3 × 3 and a stride of 1 × 1, a batch normalisation layer, a ReLU layer, and a max pooling layer with a pool size of 2 × 2 and a stride of 2 × 2. The numbers of filters for the first and second convolution layers are 8 and 16, respectively. After that, all extracted features are fused into one channel in an addition layer. A final convolution layer with a filter size of 3 × 3 and a stride of 1 × 1 is applied for further feature extraction from the fusion signal. The number of filters is 32. Similarly, a batch normalisation layer and a ReLU layer follow. A dropout layer with a rate of 0.5 is used to reduce overfitting. Finally, all the features are sent to the fully connected layer for the final classification of different types of gearbox faults. The parameters used in each layer are summarised in Fig. 2.

3. Experimental studies

3.1. Experimental setup

To validate the effectiveness and efficiency of the proposed CS-based dual-channel CNN in machine condition monitoring, a series of experiments were conducted on a two-stage helical gearbox test rig at the Centre for Efficiency and Performance Engineering. Fig. 4(a) shows the schematic diagram of the test rig, mainly consisting of a 15 kW induction motor, a DC load generator, two helical gearboxes installed back-to-back (the left one, regarded as GB1, is the speed reducer and the right one, regarded as GB2, is the speed increaser) and a sensorless variable speed drive control system.

A commercial Integrated Electronics Piezo-Electric (IEPE) microphone YG-201 with high performance was fixed about 20 cm above GB1. It collected the acoustic signals and sent them to a computer via the data acquisition system YE6232B. Moreover, an infrared thermal imaging camera, named FLIR ONE Pro, was set in front of the test object GB1 at a distance of about 25 cm from the gearbox housing surface. It was connected to an Android mobile phone to record the thermal videos during testing. As the mobile has an integrated microelectromechanical systems (MEMS) microphone, the acoustic signals were recorded in the videos as well. Fig. 4(b) illustrates the photograph of the test rig. The specifications of the test facilities are listed in Table 1.

3.2. Fault status setting

(1) Oil shortage

In a harsh working environment, oil loss caused by leakage and evaporation is inevitable in gearboxes, especially those applied in heavy-duty machinery. Therefore, the oil shortage of the gearbox was simulated in this study. The standard oil volume of the tested gearbox recommended by the manufacturer is 2600 mL. Its level inside the gearbox is approximately shown in Fig. 5(a). This is a healthy working condition of the gearbox, as well as the baseline of the tests which is denoted as BL2600 in Fig. 5(b). Then, 600 mL oil was taken out from the tested gearbox, i.e. GB1, which is the first oil shortage faulty case and regarded as OS2000. Furthermore, another 500 mL oil was released, leaving only 1500 mL oil in the gearbox, which is denoted as OS1500. The oil level and gear immersed states for these three cases are illustrated from various views in the schematic diagram plotted in Fig. 5(b) and (c). It is obvious that the input and output gear sets lack sufficient lubrication because the oil levels are too low under the shortage conditions.

(2) Tooth breakage

The drive gear is connected to the power source, and the driven gear meshes with the drive gear. In general, gear tooth breakage is caused by overload, fatigue or chemical attack. In this study, 50% and 100% of the tooth surface were artificially removed from the two driven gears in the width direction, as two faulty cases of tooth breakage and regarded as TB050 and TB100. They are depicted in Fig. 5(e) and (f), and compared with the healthy gears in Fig. 5(d). Please note that the helical gears used in this study have a total ratio of 4.559. Even if one tooth is completely broken, the gears can have at least three teeth undertake the load and operate smoothly i.e. without very noticeable changes in dynamics such as vibration and angular speeds like that of a spur gear.

(3) Oil degradation

Lubricants in gearboxes degrade mainly due to oxidation, particles and some improper use like adding a lower viscosity fluid. It affects the production and lubrication of the oil film which leads to server gear wear. The lubricant with low viscosity, MILLGEAR 100 EP, was used to simulate the oil degradation. Its key characteristics are listed and compared with the recommended standard MILLGEAR 320 EP in Table 2. This is the last faulty case named Vis100.

3.3. Test procedure

According to the previous description, six different cases were performed with a two-stage helical gearbox test rig, including the healthy case (BL2600), two oil shortage cases (OS2000 and OS1500), two tooth breakage cases (TB050 and TB100) and a low-viscosity case (Vis100). 2600 mL MILLGEAR 320 EP lubricant and healthy gears were used for the health tests as the baseline.

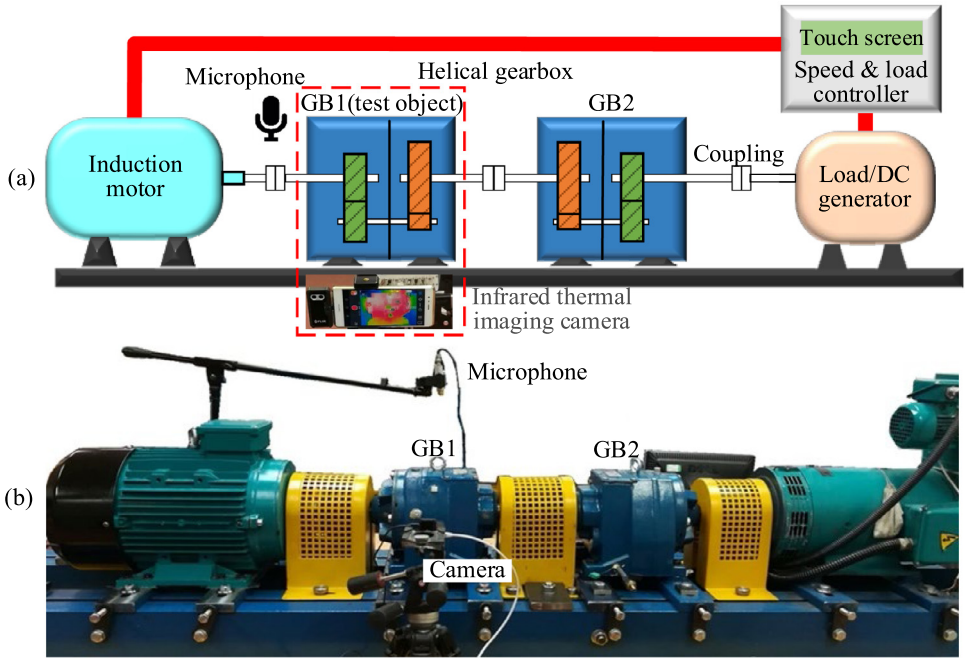

Fig. 4. Two-stage helical gearbox test rig: (a) schematic diagram, and (b) photograph.

Table 1
Key specifications of the test facilities.

Facilities	Parameters	Values
AC motor	Power Rated speed	15 kW 1460 rpm
Gearbox (M07223.6BRC-1)	First stage teeth number Second stage teeth number Ratio Oil type Oil volume	58/47 13/59 3.678/1 EP320 2600 mL
Microphone (YG-201)	Frequency response Sensitivity	16 Hz to 100 kHz 47.7 mV/Pa
DAQ (YE6232B)	A/D bits Signal frequency Range Sample rate	24 bits DC-30 kHz Max. 96 kHz/CH, parallel
FLIR ONE Pro	Thermal resolution Visual resolution Thermal sensitivity Frame rate Focus	160 × 120 pixels 1440 × 1080 pixels 70 mK 8.7 Hz Fixed 15 cm–Infinity

Table 2
Specifications of MILLGEAR 320 EP and MILLGEAR 100 EP.

Model	Specific gravity @ 15 °C	Kinematic viscosity @ 100 °C/cSt	Kinematic viscosity @ 40 °C/cSt	Viscosity index
EP320	0.901	23.5	320	92
EP100	0.885	10.9	100	93

Since thermal characteristics of the gearbox will be captured, the test rig was warmed up at 75% of the full speed and 80% of the rated load until the oil temperature of GB1 was measured by an inserted K-type thermocouple reached 32 °C (the ambient temperature was about 21 °C). Then, all tests were carried out at a constant speed, i.e. the full speed of the driving motor, and four different loads (0%, 30%, 70% and 100% of the rated load of the DC generator), which are controlled by the speed and load control system. The timeline recording of the experiment process of each case is displayed in Fig. 6. After changing the load, it continues to warm up the test rig at 100% speed and the corresponding load for 20 min. Both acoustic signals and thermal videos were captured for 4 min immediately after warming up. Afterwards, 1 minute was reserved to change the load through the control system. Under each load condition, the entire process takes 25 minutes.

Although the frame rate on the camera specification is 8.7 frames/s, the actual stored video frame rate is much lower than the specified value due to the limitation of the storage speed. The length of the video recorded under different working conditions is the same, but the total number of frames is different. As a result, the first 800 frames of the thermal images were selected to represent the characteristics of each set (24 sets, 6 healthy and faulty cases × 4 loads). Simultaneously, the acoustic signals corresponding to the length of time are intercepted for feature extraction of each set. Table 3 lists dataset grouping for CNN

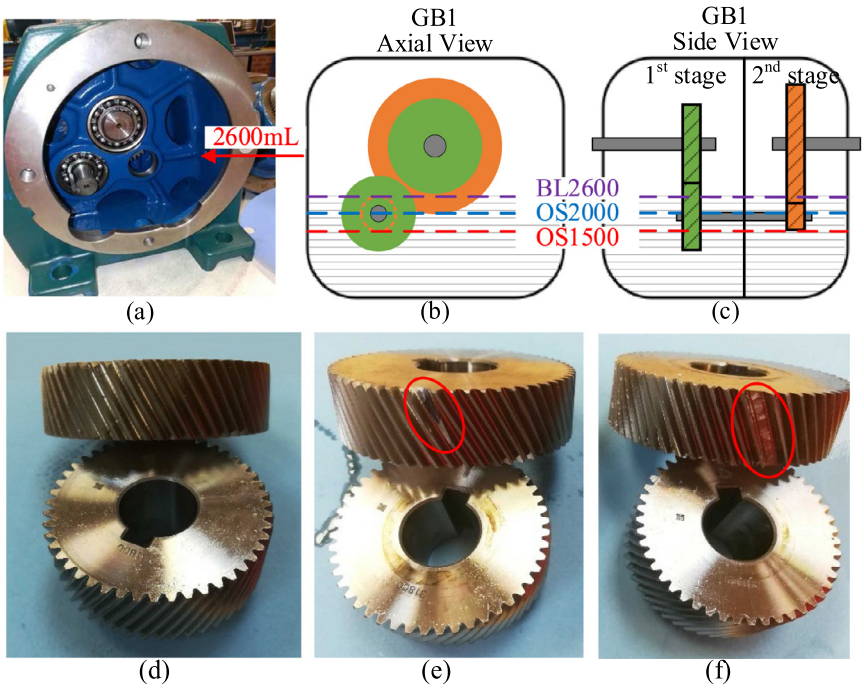

Fig. 5. Different gearbox fault simulation: (a) insider view of the lubricant level, (b) axial view, (c) side view, (d) healthy pair, (e) 50% tooth breakage, and (d) 100% tooth breakage.

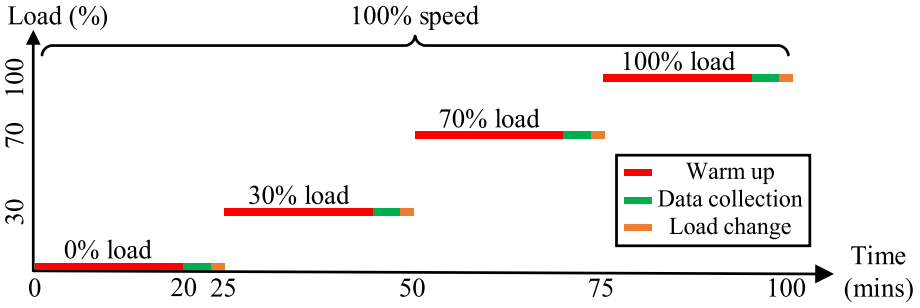

Fig. 6. Timeline recording of the experiment process.

training, validating and testing. In most studies, the signals from all working conditions are used for training and the remaining data is then used for testing. To avoid the overfitting problem, only the data sets under three load conditions are used for network training. The 0% and 100% loads are the extreme working conditions for gear transmission systems, which commonly generate the lowest and highest temperature for thermal images, the smallest and largest modulation features for acoustic signals. Reasonably the extreme conditions that represent the limit of fault signatures are selected as the essential conditions for training. Either one of the medium load conditions (30% and 70% loads) can be used for training and the other one can be used for testing. The selection of the training and testing data sets is in consideration of the practical machine operating conditions and the indispensable overfitting problem in this machine learning study.

4. Results and discussion

4.1. Acoustic signal comparison

Since the recorded thermal videos contain acoustic signals captured by the integrated MEMS microphone of a mobile phone, they are compared with the acoustic signals collected by the commercial IEPE omnidirectional microphone. Fig. 7 displays their spectrum with the key information of the modulation characteristics relating to gear faults.

The low-frequency signals are easily contaminated by strong background noise and are not reliable for machine condition monitoring. Hence, modulation analysis is preferred in fault detection and diagnosis of rotating machines. The modulation mechanism carries the fault information from the low-frequency range to a high-frequency range, which enlarges the fault signatures accordingly. In general, the amplitude between 4000 Hz and 8000 Hz in the spectrum is relatively high and it can be empirically selected as the potential candidate frequency band for demodulation analysis. Consequently, the comparison of the two signals is focused on the frequency band 4000 Hz to 8000 Hz. The amplitude in this frequency band is pronounced in both microphone signals, which shows the high similarity in the spectrum. The employed MEMS microphone is a much cheaper and less-capable sensor compared with the IEPE microphone. The frequency responses and resolutions of MEMS and IEPE microphones cannot be equivalent in a very wide frequency range. In addition, the distance and location of the two microphones are different during the tests, which also contributes to the differences between the two microphones. The gear fault induces the amplitude variation of the shaft rotating frequency and harmonics, which are usually selected as the fault signature. To further show the

Table 3
Dataset grouping for CNN training and testing.

Labels	Training and validation groups		Testing group	
	Speed	Load	Speed	Load
BL2600 OS2000 OS1500 TB050 TB100 Vis100	100%	0%, 30% (or 70%) and 100% (randomly select 70% for training and 30% for validating in each dataset)	100%	70% (or 30%) (use all for testing to avoid overfitting issues)

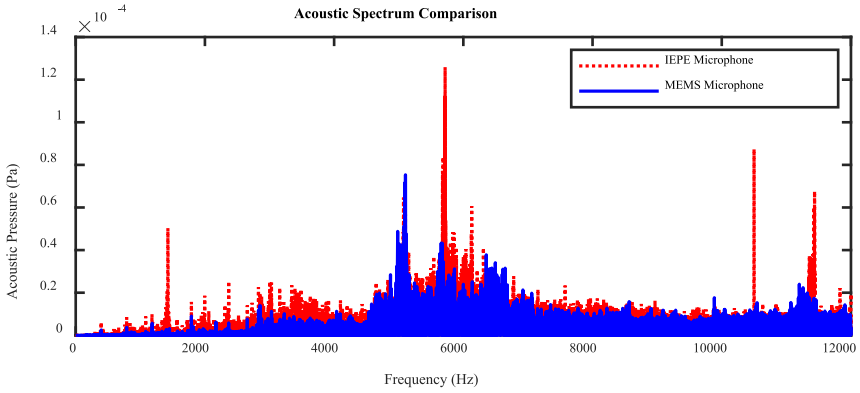

Fig. 7. Comparison of the acoustic spectra of the commercial IEPE microphone and the MEMS microphone of the mobile phone.

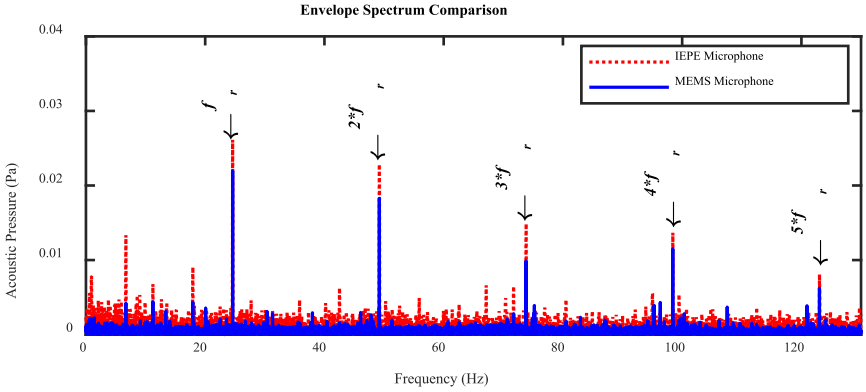

Fig. 8. Comparison of the envelope spectra of the IEPE and the MEMS microphones.

feasibility of the mobile MEMS microphone, the envelope spectra of two acoustic signals are given in Fig. 8. It can be seen that amplitudes at the fundamental frequency and harmonics of the rotating frequency are close to each other in terms of spectral profiles, both of them being able to provide correct diagnostic information. However, the lower amplitudes of the MEMS spectrum are because it was located 5 cm farther away, compared with the IEPE one. In addition, the 3rd harmonic has a larger decrease due to the nonlinearity of the frequency responses.

In this study, the proposed approach used the acoustic signals captured by the non-contact MEMS microphone to achieve an accurate diagnosis of common gearbox faults, which also verifies the effectiveness and feasibility of the MEMS sensor in machine condition monitoring.

4.2. MSB analysis

According to the previous selection of the frequency band from 4000 Hz to 8000 Hz, the MSB results are calculated through the acoustic signals recorded by the MEMS microphone of the mobile phone for further intelligent fault diagnosis. Fig. 9 displays the MSB magnitudes of six cases under 100% loads.

The yellow components indicate the intricate coupling effects between the carrier and modulation signals. It can be seen that the carrier frequencies are similar among various fault cases because of the dynamic transfer properties of the same gearbox. Gear faults mainly change the modulation degree or generate new modulated frequencies and the differences are clear between different cases. The fault cases of OS1500, OS2000 and Vis100 are related to the different tribo-dynamic responses due to various lubrication conditions and the main yellow bifrequency components are quite similar but with substantially distinct magnitudes. When the tooth breakage faults occurred, the additional modulation leads to the increase of the coupled components in the MSB magnitude. The bifrequency components in yellow colour for the fault cases of TB050 and TB100 are much more than the other four cases. The differences in the magnitude and bifrequencies form the significantly important features to distinguish various faults in gearboxes, which increases the accuracy and robustness of the proposed Dual-Channel CNN method in fault detection and diagnosis of gearboxes. The uncoupled components in blue have relatively small magnitudes and only modulated components have large magnitudes, which shows the high sparsity of

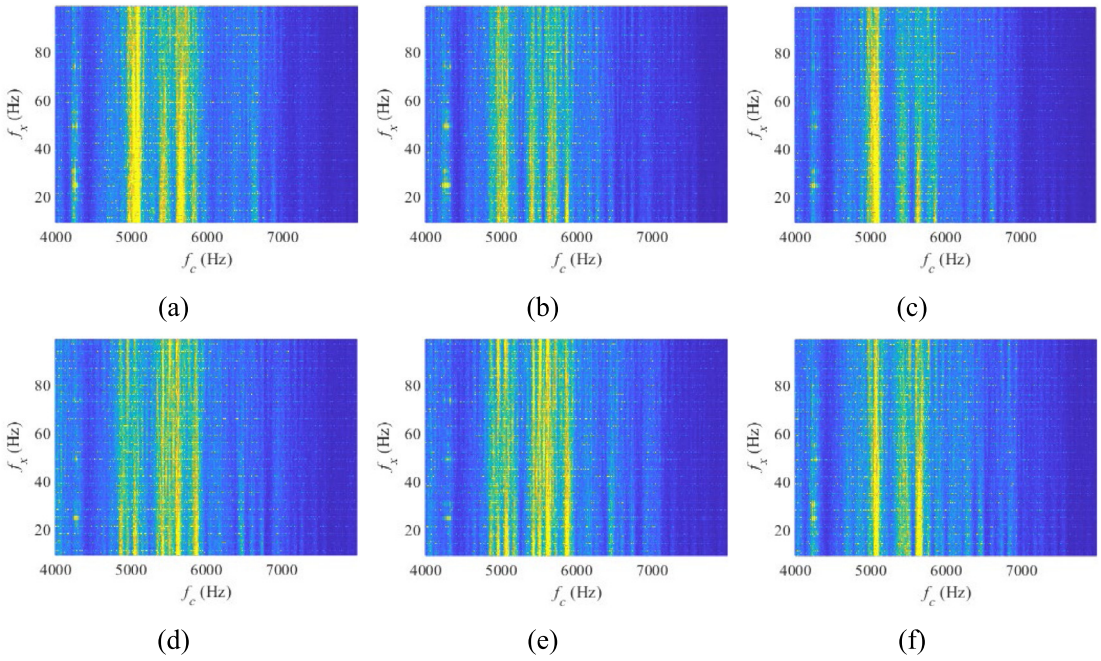

Fig. 9. MSB RGB images for different cases under 100% load: (a) BL2600, (b) OS1500, (c) OS2000, (d) TB050, (e) TB100 and (f) Vis100.. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

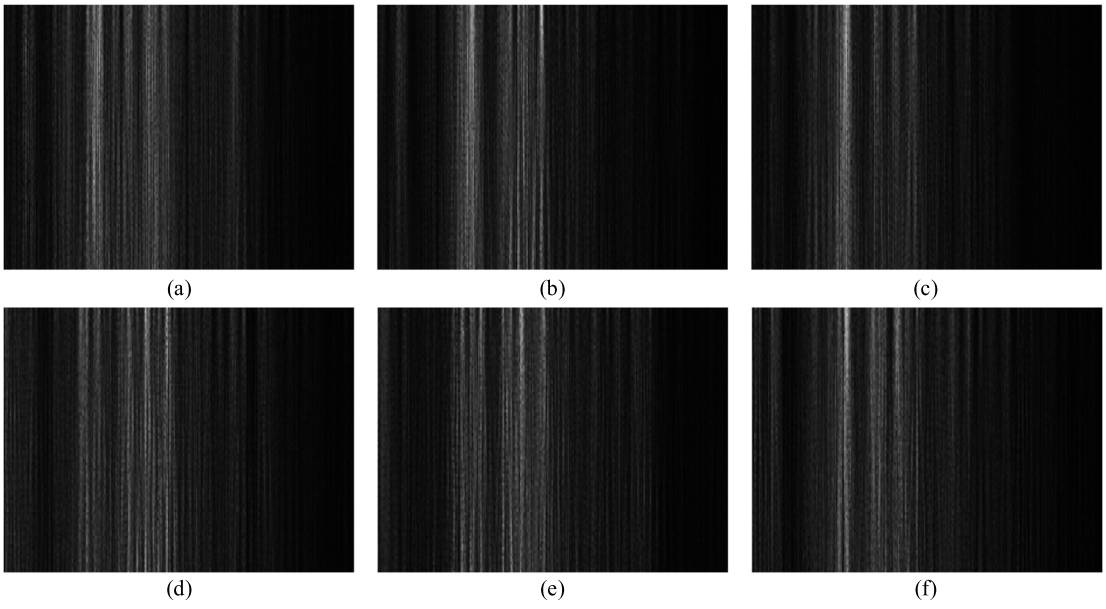

Fig. 10. MSB grayscale images for different cases under 100% load: (a) BL2600, (b) OS1500, (c) OS2000, (d) TB050, (e) TB100 and (f) Vis100.

the MSB results across various test cases. The sparsity allows the efficient compression to speed up the Dual-Channel CNN for efficient condition monitoring.

The MSB RGB images are converted into 2-D grayscale images to reduce the dimensions to 1080 × 1440, which are displayed in Fig. 10. It is noticeable that the MSB grayscale images are sparse and can be further compressed with CS.

Fig. 11 depicts the CS compressed MSB grayscale images with a size of only 60 × 80. It will dramatically speed up the training of the network. In addition, the main fault characteristics are also preserved to a certain extent.

4.3. Temperature distribution analysis

Fig. 12 (a) shows the surface topography of the tested gearbox housing. It is an uneven surface that affects its temperature distribution as presented in Fig. 12(b). But the damage to the surface texture is deterministic therefore it does not affect the application of the thermal images. Note that the recorded thermal images are flipped compared to the test rig.

The heat sources in a gearbox mainly include gear meshing, bearing rolling, and seal friction. Most of the heat dissipated into the lubrication oil and the remaining heat conducts through the shaft and bearing to the gearbox housing. As most of the

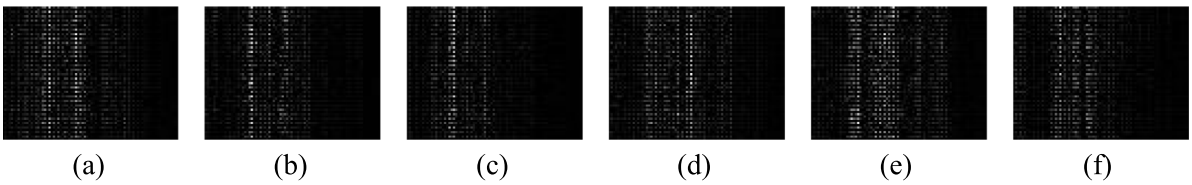

Fig. 11. CS based MSB grayscale images for different cases under 100% load: (a) BL2600, (b) OS1500, (c) OS2000, (d) TB050, (e) TB100 and (f) Vis100.

Fig. 12. Surface topography of the gearbox housing and its thermal image: (a) Surface topography and (b) thermal image.

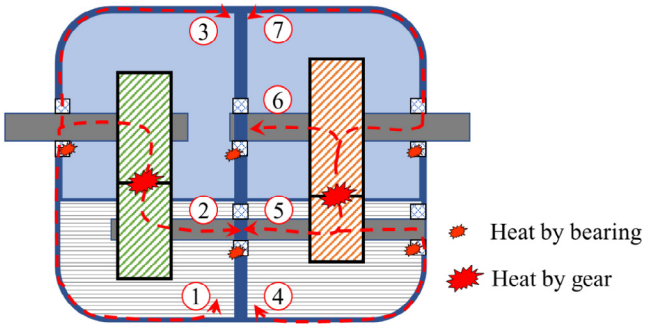

Fig. 13. Heat generation and transmission path of gearboxes.

generated heat goes into the lubrication oil, the lower part of the gearbox has basically a high temperature. The other metal components act as heat conduction ways for dissipating heat to the outside surface for radiating into the air. Fig. 13 shows the heat transfer paths in the metal parts from the gears and bearings to the front surface targeted by the thermal imaging camera. There mainly exist seven paths for heat conduction according to the structure of the tested gearbox. Due to the uneven heat generation and conduction, the high-temperature areas vary in different health conditions.

To visually observe the temperature distribution, one thermal image is selected from each video recorded at 100% load of different faulty states. They are converted into grayscale images and further equalised to enhance contrast and edge information as compared in Fig. 14.

Comparing OS1500, OS2000 and BL2600 in Fig. 14(b), (c) and (a), it can be noticed that the high-temperature area shifts from the upper to the lower of the gearbox housing because of the low oil level. Moreover, the heat generated by gear and bearing friction is conducted more along the shaft to the housing when the gearbox works at lubrication starvation under heavy loads. Therefore, the temperature around the bearing of the output shaft is extremely high at the lowest level for the case of OS1500. For the tooth breakage cases shown in Fig. 14(d) and (e), it is difficult to observe the differences in their temperature distributions compared with the baseline. The tooth breakage faults reduce the meshing stiffness and result in more friction at the other teeth. However, the tooth breakage faults generate limited heat compared with the other faults. It only slightly expands the high-temperature area close to the faulty gear, which is marked by the red rectangles referring to the baseline case. Fig. 14(d) depicts the temperature distribution of the low viscosity case. The low viscous oil cannot build a perfect oil film for ideal lubrication, and hence more heat is widely generated due to increased friction. Therefore, the high-temperature areas were located similarly but obviously expanded compared to the baseline case. In summary, the heat distribution varies between different faults. The lubricity related faults show more pronounced differences in heat distribution, whilst the localised faults result in less variation of heat distribution. It is still hard to detect and classify the faults with human observation.

Fig. 15 illustrates the equalised grayscale images for the case of OS2000 at four different loads. It is noticeable that the temperature distribution is similar to the case of OS1500 when the load is light, except for the shaft connecting the GB2. As the load increases, the high-temperature area gradually expands, and the boundary is much clearer. Therefore, both light and heavy loads are selected to train the neural network, together with either one of the medium load conditions (30% and 70% loads), and the other one is used for testing in this study to prevent the overfitting issue.

Fig. 16 displays the compressed thermal images with a CS compression ratio of 324 for six different faulty cases at 100% load. It is obvious that the valuable information for fault diagnosis, like the edge region and brightness, is preserved after compression.

Therefore, these CS-based thermal and MSB grayscale images are fed into the input layers of the CNN as two channels. The fault features will be further extracted and trained to obtain an efficient network for fault classification.

4.4. Classification results

It has been discussed that some cases can be clearly detected with visualisation by means of temperature distributions and acoustic MSB features. However, sometimes they are affected by variable loads and are difficult to observe manually. Accordingly, intelligent fault diagnosis has the capability to achieve fault detection and classification instead of human observation. In this study, the proposed CS-based Dual-Channel CNN method is carried out by feeding the CS-based thermal and acoustic MSB grayscale images into the input layers from two channels. After twice operations of convolution, batch normalisation, ReLU activation, and max pooling, the extracted features from two channels are fused in the addition layer. The features are then further extracted and compressed by the final convolution and max pooling. To compare the effectiveness of the proposed method, the compressed thermal and acoustic MSB grayscale images are fed as the input signals of two single-channel CNNs to perform cubic convolution and max pooling operations with the same parameters as the dual-channel network, respectively. The architectures of the proposed dual-channel CNN and the compared single-channel CNN are displayed in Fig. 17.

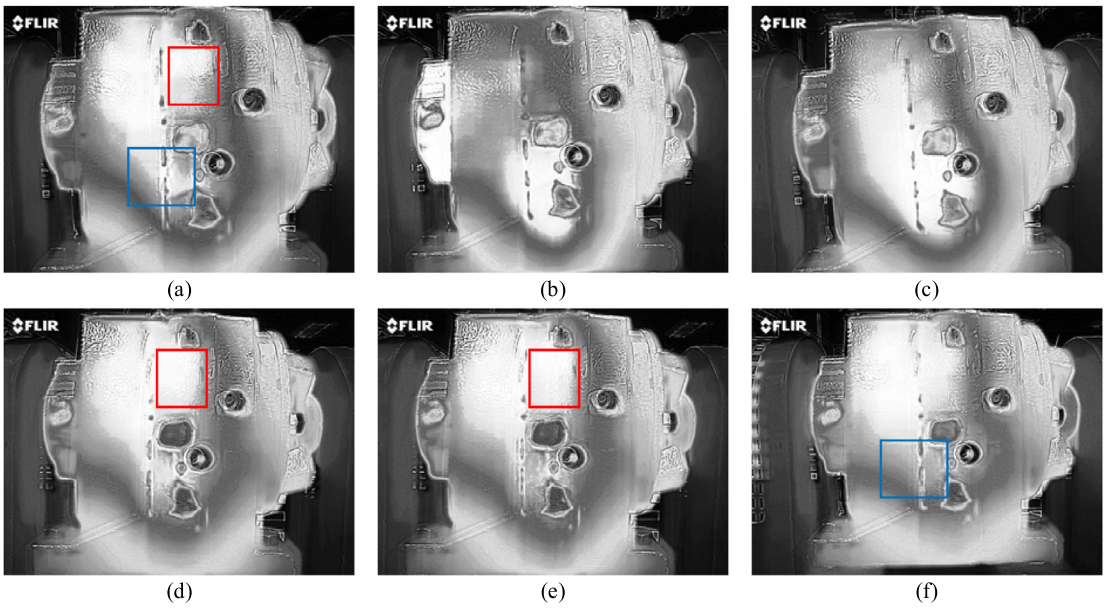

Fig. 14. Equalised grayscale images for different states under 100% load: (a) BL2600, (b) OS1500, (c) OS2000, (d) TB050, (e) TB100 and (f) Vis100.. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

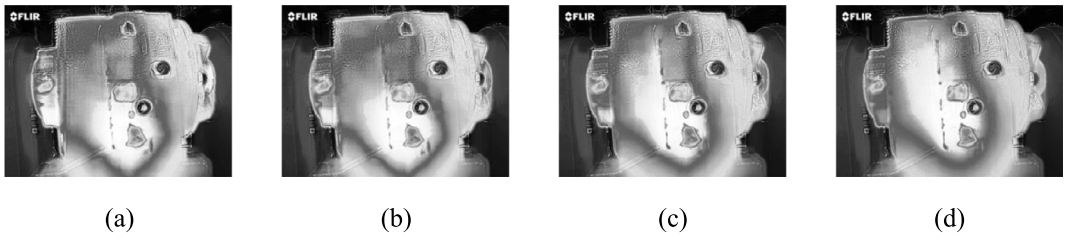

Fig. 15. Equalised grayscale images for OS2000 under different loads: (a) 0%, (b) 30%, (c) 70% and (d) 100%.

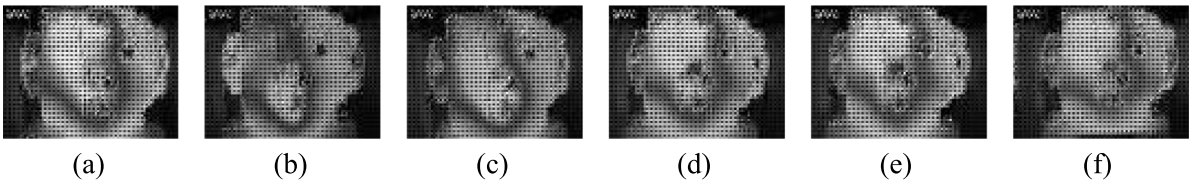

Fig. 16. CS based images for different cases under 100% load: (a) BL2600, (b) OS1500, (c) OS2000, (d) TB050, (e) TB100 and (f) Vis100.

In addition, both the dual-channel and single-channel CNNs use the stochastic gradient descent with momentum algorithm (SGDM) as the optimiser because it is a popular optimisation algorithm to speed up convergence. The initial learning rate is set to 0.001. The maximum number of epochs is 30. A mini-batch has 128 observations at each iteration in this study.

To visualise activations and interpret the prediction of a CNN network, some visualisation techniques are investigated, such as the class activation mapping (CAM), Grad-CAM, locally-interpretable model-agnostic explanation (LIME), t-distributed stochastic neighbour embedding (t-SNE) and so on. Since the proposed method is based on dual-channel signals, CAM [53] can work for it to produce heat maps to highlight the features extracted in the activations. Fig. 18 compares features visualised in the final convolutional layer using CAM in different cases with different training methods. Fig. 18(a), (b) and (c) show the heat maps generated by the single-channel thermal images, single-channel acoustic MSB images, and dual-channel fused images in six different cases, respectively. They indicate that the trained CNN can grab the key features from the thermal images and acoustic MSB images. However, it is difficult to distinguish all faults with a single signal source. For example, the thermal features of TB050 and TB100 are similar and difficult to distinguish. Although MSB features have better discrimination, acoustic signals are inherently dynamic and noisy which may result in low classification accuracy. With the dual-channel signal fusion, the discriminative image regions are more evident for the intelligent fault classification of helical gearboxes.

Furthermore, t-SNE [54] is also conducted to visualise extracted high-dimensional features in the final convolution and the fully connected layers through mapping to the low dimension. Because it is a statistical method to make the high-dimensional similar objects close in the low dimension with a high probability and high-dimensional dissimilar objects far in the low dimension with a high probability. Fig. 19(a) to (c) display the visualisation of final convolutional activations from single thermal channel CNN, single acoustic MSB channel CNN, and proposed dual-channel CNN, respectively. It is noticeable that several clusters are split into two or more pieces in the individual thermal imaging result in Fig. 19(a). For the acoustic MSB results, they

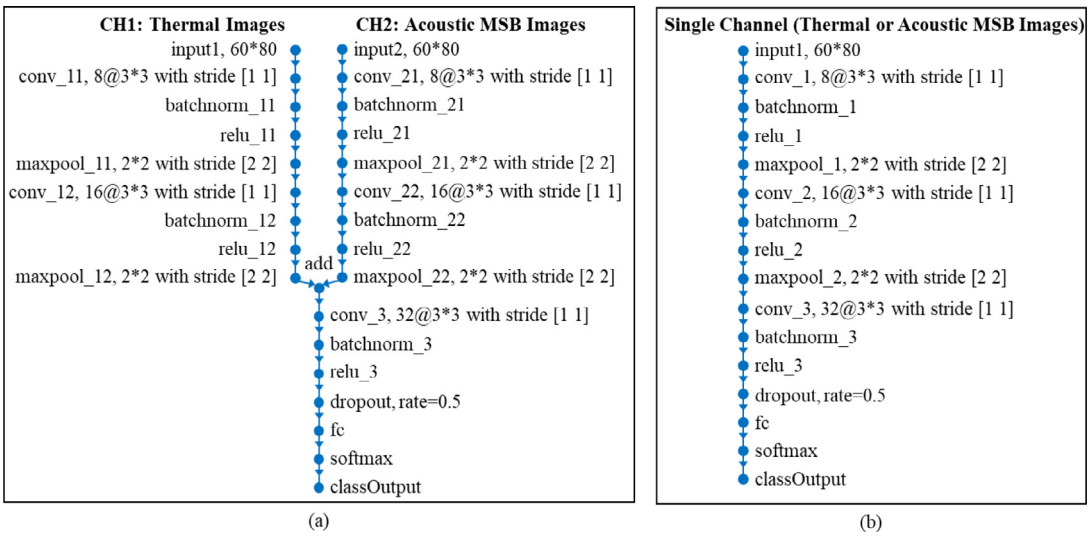

Fig. 17. Diagram of the networks: (a) Dual-Channel, (b) Single-Channel.

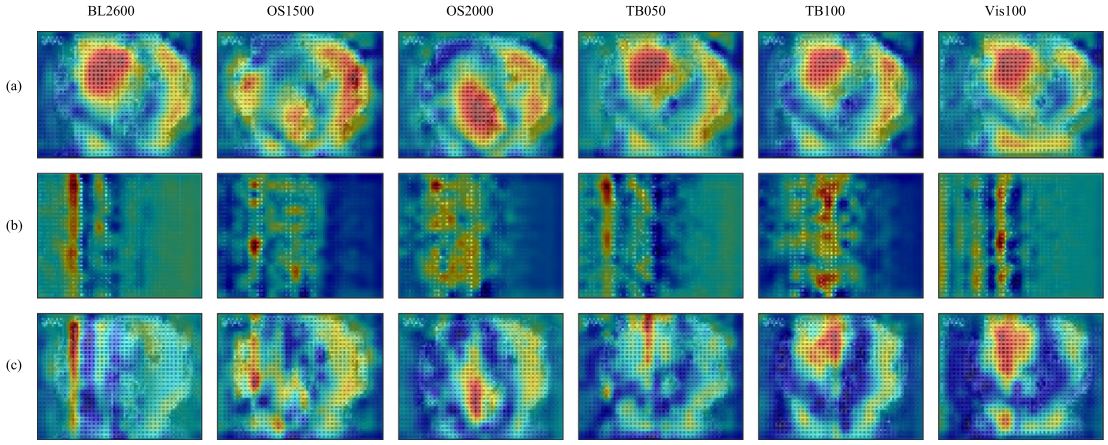

Fig. 18. Feature visualisation of the final convolutional layer with CAM under different cases: (a) row for single-channel thermal images, (b) row for single-channel acoustic MSB images and (c) row for dual-channel images.

are embedded clusters with less space between them as shown in Fig. 19(b). Only the result of the proposed method apparently shows compact clusters in Fig. 19(c). Fig. 19(d) to (f) exhibit the fully connected (FC) activations of these three trained networks. The clustering results are better than that of the final convolutional layer, especially for two single-channel CNNs. Although the clustering results are improved a lot, the split clusters and embedded clusters still exist for the two single-channel CNNs. But for the proposed dual-channel CNN, the clusters in the fully connected layer are clearly divided, which means the different cases are successfully classified with a high probability. The feature visualisation results further prove that the proposed CS-based Dual-Channel CNN method is effective and efficient for fault diagnosis of the helical gearbox.

The classification accuracy of two testing datasets is displayed with confusion matrices as presented in Figs. 20 and 21. The CS-based single thermal channel CNN has high performance for the diagnosis of different types of gearbox failures, but it is hard to distinguish the severity of the same type of failure. The CS-based MSB channel CNN is capable of diagnosing different faults and severities. However, most diagnostic results are less than 95%, making it unable to independently take on the burden of gear fault diagnosis, of which the reason is that the acoustic signals are dynamic and noisy by nature. With the proposed CS-based Dual-Channel CNN method, the classification accuracy is higher to 99.39% on average with smaller fluctuation among various cases, which means that the fused features of thermal imaging and acoustic signals are more reliable and robust for fault diagnosis of the gearbox faults.

Finally, the average testing classification accuracy performed by different trained networks is compared in Table 4. The proposed Dual-Channel CNN method has a high classification accuracy of 99.39% on average. Moreover, the image resolution is compressed from 1080 × 1440 pixels to 60 × 80 pixels with a compression ratio of 324 through CS compression. The training was conducted on a computer equipped with a processor of Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10 Hz and 32 GB RAM. Training times spent for the network training with original and CS-based grayscale images are compared in Table 4. It is clear that if the network is trained with original grayscale images, the training time is extremely long and even cannot be carried out due to the limited memory of this computer. But with CS-based grayscale images, training time is around 375 times less than the original grayscale images, which evidently shows the high performance of the CS in accelerating the training speed of the neural network.

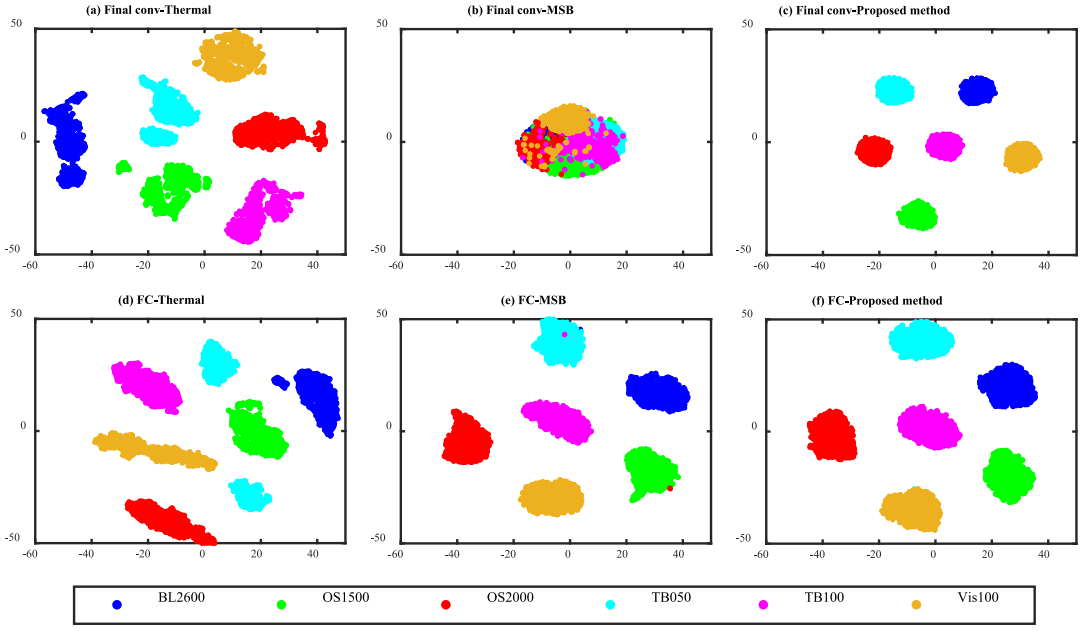

Fig. 19. Feature visualisation with t-SNE: (a) final convolutional activations of thermal images, (b) final convolutional activations of acoustic MSB images, (c) final convolutional activations of the proposed dual-channel method, (d) fully connected activations of thermal images, (e) fully connected activations of acoustic MSB images and (f) fully connected activations of proposed dual-channel method.

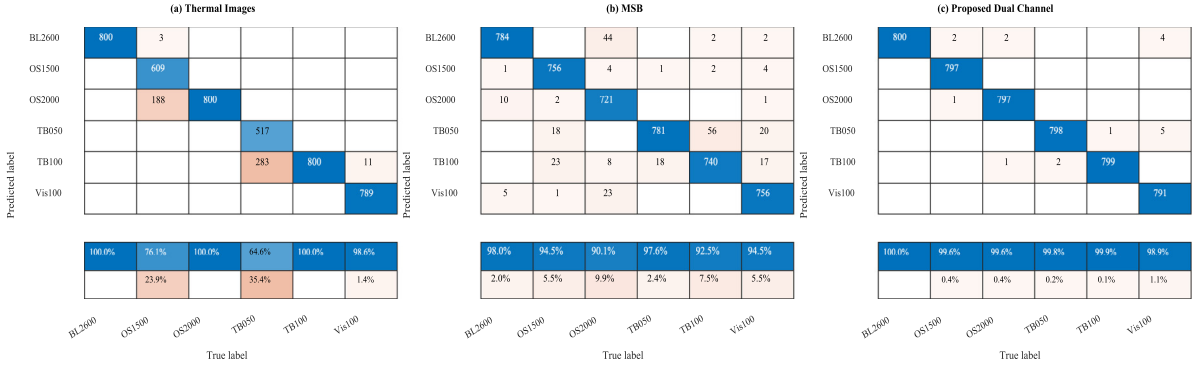

Fig. 20. Confusion chart of the diagnostic accuracy under 70% load: (a) CS-based thermal channel CNN, (b) CS-based MSB channel CNN, and (c) proposed CS-based Dual-Channel CNN.

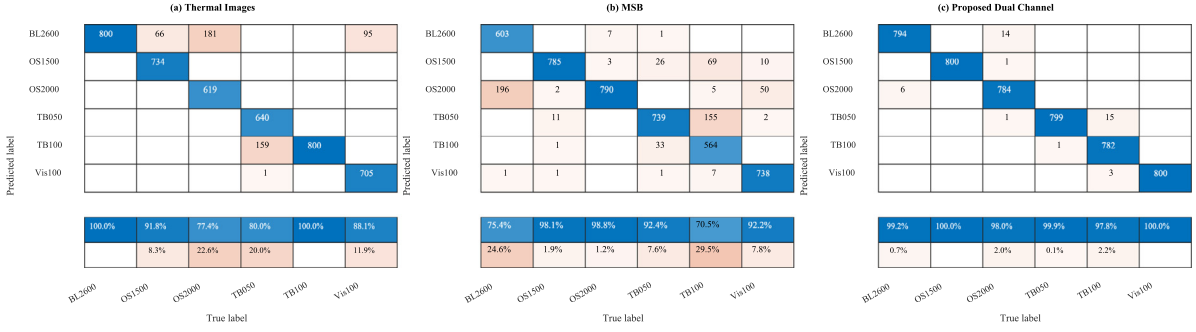

Fig. 21. Confusion chart of the diagnostic accuracy under 30% load: (a) CS-based thermal channel CNN, (b) CS-based MSB channel CNN, and (c) proposed CS-based Dual-Channel CNN.

5. Conclusions

In this article, a CS-based Dual-Channel CNN method was proposed to intelligently diagnose various types of helical gearbox faults with two complementary non-contact measurements, thermal images and acoustic signals from a mobile phone. The comparison results verify that acoustic signals recorded by the MEMS microphone of a mobile phone can effectively detect the faults instead of the high-performance IEPE microphone. The MSB method can highlight the coupling components of acoustic signals during gear meshing and suppress the irrelevant components and random noise, which generate highly sparse acoustic MSB

Table 4
Comparison of the accuracy and training time.

Networks	Mean testing accuracy (%)	Training time	
		Original images	CS-based images
Single thermal channel CNN	89.72	71.20 h	11.35 min
Single acoustic MSB channel CNN	91.22	69.17 h	11.07 min
Proposed Dual-Channel CNN	99.39	Out of memory	24.09 min

images. The compression ratio of CS reaches 324 times, and the compressed images apparently show that CS significantly accelerates the training speed by reducing the image redundancy but preserving the valuable fault information. The proposed CS-based Dual-Channel CNN method was demonstrated to effectively and efficiently overcome the unreliable and unstable shortcomings of using a single domain source. The performance of gearbox fault diagnosis and the diagnostic accuracy (reaching 99.39% on average) are both improved.

Although using dual channels improves the classification accuracy, it consumes more than double the training time of the single-channel network. Consequently, accelerating the training speed of the dual-channel networks will be further investigated in the future. In addition, CS has the potential to implement pre-acquisition compression with hardware like the single-pixel camera [55]. Then, the thermal camera and the mobile phone can be carried with a drone or robots to capture compressed thermal images to save storage space and transmission bandwidth in the future.

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.

Acknowledgement

The authors would like to thank the Centre for Efficiency and Performance Engineering at the University of Huddersfield for supporting the experiments in this research.

References

[1] Shah J, Wang W. An evolving neuro-fuzzy classifier for fault diagnosis of gear systems. ISA Trans 2021. http://dx.doi.org/10.1016/j.isatra.2021.05.019.
[2] Miao Y, Zhao M, Liang K, Lin J. Application of an improved MCKDA for fault detection of wind turbine gear based on encoder signal. Renew Energy 2020;151:192–203. http://dx.doi.org/10.1016/j.renene.2019.11.012.
[3] Xu Y, et al. Orthogonal on-rotor sensing vibrations for condition monitoring of rotating machines. J Dyn Monit Diagn 2021. http://dx.doi.org/10.37965/jdmd.v2i2.47.
[4] Bagavathiappan S, Lahiri BB, Saravanan T, Philip J, Jayakumar T. Infrared thermography for condition monitoring – A review. Infrared Phys Technol 2013;60:35–55. http://dx.doi.org/10.1016/j.infrared.2013.03.006.
[5] Tang X, Wang X, Cattley R, Gu F, Ball AD. Energy harvesting technologies for achieving self-powered wireless sensor networks in machine condition monitoring: A review. Sensors 2018;18(12):4113. http://dx.doi.org/10.3390/s18124113.
[6] Touret T, Changenet C, Ville F, Lalmi M, Becquerelle S. On the use of temperature for online condition monitoring of geared systems – A review. Mech Syst Signal Process 2018;101:197–210. http://dx.doi.org/10.1016/j.ymssp.2017.07.044.
[7] Li Y, Gu JX, Zhen D, Xu M, Ball A. An evaluation of gearbox condition monitoring using infrared thermal images applied with convolutional neural networks. Sensors 2019;19(9):2205. http://dx.doi.org/10.3390/s19092205.
[8] Li Y, Du X, Wang X, Si S. Industrial gearbox fault diagnosis based on multiscale convolutional neural networks and thermal imaging. ISA Trans 2022. http://dx.doi.org/10.1016/j.isatra.2022.02.048.
[9] Anil K, Deepam G, Pabla BS. Thermal image based fault diagnosis of gears using support vector machines. Int J Innov Technol Explor Eng 2019;9(1):155–60. http://dx.doi.org/10.35940/ijitee.A3957.119119.
[10] Waqar T, Demetgul M. Thermal analysis MLP neural network based fault diagnosis on worm gears. Measurement 2016;86:56–66. http://dx.doi.org/10.1016/j.measurement.2016.02.024.
[11] Pranesh SK, Abraham S, Sugumaran V, Amarnath M. Fault diagnosis of helical gearbox using acoustic signal and wavelets. IOP Conf Ser Mater Sci Eng 2017;197:012079. http://dx.doi.org/10.1088/1757-899X/197/1/012079.
[12] Parey A, Singh A. Gearbox fault diagnosis using acoustic signals, continuous wavelet transform and adaptive neuro-fuzzy inference system. Appl Acoust 2019;147:133–40. http://dx.doi.org/10.1016/j.apacoust.2018.10.013.
[13] Hartono D, Halim D, Wyn Roberts G. Gear fault diagnosis using an improved reassigned smoothed pseudo Wigner-Ville distribution. Cogent Eng 2018;5(1):1436928. http://dx.doi.org/10.1080/23311916.2018.1436928.
[14] Kane PV, Andhare AB. Critical evaluation and comparison of psychoacoustics, acoustics and vibration features for gear fault correlation and classification. Measurement 2020;154:107495. http://dx.doi.org/10.1016/j.measurement.2020.107495.
[15] Vanraj S Dhami, Pabla B. Hybrid data fusion approach for fault diagnosis of fixed-axis gearbox. Struct Health Monit 2018;17(4):936–45. http://dx.doi.org/10.1177/1475921717727700.
[16] Yao J, Liu C, Song K, Feng C, Jiang D. Fault diagnosis of planetary gearbox based on acoustic signals. Appl Acoust 2021;181:108151. http://dx.doi.org/10.1016/j.apacoust.2021.108151.
[17] Garcia Fernandez P, de Juan A, Diez-Ibarbia A, Sanchez-Espiga J, Fernandez del Rincon A. Acoustic intensity technique applied to monitor planetary gears. Appl Acoust 2021;172:107621. http://dx.doi.org/10.1016/j.apacoust.2020.107621.
[18] Lu K, Gu JX, Fan H, Sun X, Li B, Gu F. Acoustics based monitoring and diagnostics for the progressive deterioration of helical gearboxes. Chin J Mech Eng 2021;34(1):82. http://dx.doi.org/10.1186/s10033-021-00603-1.
[19] Jing L, Zhao M, Li P, Xu X. A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. Measurement 2017;111:1–10. http://dx.doi.org/10.1016/j.measurement.2017.07.017.
[20] Abdul ZK, Al-Talabani AK, Ramadan DO. A hybrid temporal feature for gear fault diagnosis using the long short term memory. IEEE Sens J 2020;20(23):14444–52. http://dx.doi.org/10.1109/JSEN.2020.3007262.
[21] Li G, Wu J, Deng C, Chen Z. Parallel multi-fusion convolutional neural networks based fault diagnosis of rotating machinery under noisy environments. ISA Trans 2021. http://dx.doi.org/10.1016/j.isatra.2021.10.023.
[22] Liang P, Deng C, Wu J, Yang Z. Intelligent fault diagnosis of rotating machinery via wavelet transform, generative adversarial nets and convolutional neural network. Measurement 2020;159:107768. http://dx.doi.org/10.1016/j.measurement.2020.107768.
[23] Huang D, Zhang W-A, Guo F, Liu W, Shi X. Wavelet packet decomposition-based multiscale CNN for fault diagnosis of wind turbine gearbox. IEEE Trans Cybern 2021;1–11. http://dx.doi.org/10.1109/TCYB.2021.3123667.
[24] Chang H, Borghesani P, Peng Z. Automated assessment of gear wear mechanism and severity using mould images and convolutional neural networks. Tribol Int 2020;147:106280. http://dx.doi.org/10.1016/j.triboint.2020.106280.
[25] Cao W, et al. Image denoising and feature extraction of wear debris for online monitoring of planetary gearboxes. IEEE Access 2021;9:168937–52. http://dx.doi.org/10.1109/ACCESS.2021.3137261.
[26] Ye Z, Yu J. AKRNet: A novel convolutional neural network with attentive kernel residual learning for feature learning of gearbox vibration signals. Neurocomputing 2021;447:23–37. http://dx.doi.org/10.1016/j.neucom.2021.02.055.
[27] Yan T, Wang D, Xia T, Xi L. A generic framework for degradation modeling based on fusion of spectrum amplitudes. IEEE Trans Autom Sci Eng 2022;19(1):308–19. http://dx.doi.org/10.1109/TASE.2020.3029162.
[28] Li W, et al. A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges. Mech Syst Signal Process 2022;167:108487. http://dx.doi.org/10.1016/j.ymssp.2021.108487.

[29] Xia M, Li T, Xu L, Liu L, de Silva CW. Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks. IEEE/ASME Trans Mechatronics 2018;23(1):101–10. http://dx.doi.org/10.1109/TMECH.2017.2728371.

[30] Wang H, Li S, Song L, Cui L, Wang P. An enhanced intelligent diagnosis method based on multi-sensor image fusion via improved deep learning network. IEEE Trans Instrum Meas 2020;69(6):2648–57. http://dx.doi.org/10.1109/TIM.2019.2928346.

[31] Azamfar M, Singh J, Bravo-Imaz I, Lee J. Multisensor data fusion for gearbox fault diagnosis using 2-D convolutional neural network and motor current signature analysis. Mech Syst Signal Process 2020;144:106861. http://dx.doi.org/10.1016/j.ymssp.2020.106861.

[32] Shao H, Lin J, Zhang L, Galar D, Kumar U. A novel approach of multi-sensory fusion to collaborative fault diagnosis in maintenance. Inf Fusion 2021;74:65–76. http://dx.doi.org/10.1016/j.inffus.2021.03.008.

[33] Li X, Zhong X, Shao H, Han T, Shen C. Multi-sensor gearbox fault diagnosis by using feature-fusion covariance matrix and multi-Riemannian kernel ridge regression. Reliab Eng Syst Saf 2021;216:108018. http://dx.doi.org/10.1016/j.ress.2021.108018.

[34] Huang H, Tang B, Luo J, Pu H, Zhang K. Residual gated dynamic sparse network for gearbox fault diagnosis using multisensor data. IEEE Trans Ind Inf 2022;18(4):2264–73. http://dx.doi.org/10.1109/TII.2021.3099060.

[35] Zhao Z, et al. Model-driven deep unrolling: Towards interpretable deep learning against noise attacks for intelligent fault diagnosis. ISA Trans 2022. http://dx.doi.org/10.1016/j.isatra.2022.02.027.

[36] Li C, Sanchez R-V, Zurita G, Cerrada M, Cabrera D, Vásquez RE. Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals. Mech Syst Signal Process 2016;76–77:283–93. http://dx.doi.org/10.1016/j.ymssp.2016.02.007.

[37] Yang S, Wang Y, Li C. Wind turbine gearbox fault diagnosis based on an improved supervised autoencoder using vibration and motor current signals. Meas Sci Technol 2021;32(11):114003. http://dx.doi.org/10.1088/1361-6501/ac0741.

[38] Ma M, Sun C, Chen X. Deep coupling autoencoder for fault diagnosis with multimodal sensory data. IEEE Trans Ind Inf 2018;14(3):1137–45. http://dx.doi.org/10.1109/TII.2018.2793246.

[39] Jiao J, Zhao M, Lin J, Ding C. Deep coupled dense convolutional network with complementary data for intelligent fault diagnosis. IEEE Trans Ind Electron 2019;66(12):9858–67. http://dx.doi.org/10.1109/TIE.2019.2902817.

[40] Mao G, Zhang Z, Qiao B, Li Y. Fusion domain-adaptation CNN driven by images and vibration signals for fault diagnosis of gearbox cross-working conditions. Entropy 2022;24(1):1. http://dx.doi.org/10.3390/e24010119.

[41] Karabacak YE, Gürsel Özmen N, Gümüşel L. Intelligent worm gearbox fault diagnosis under various working conditions using vibration, sound and thermal features. Appl Acoust 2022;186:108463. http://dx.doi.org/10.1016/j.apacoust.2021.108463.

[42] Shao H, Jiang H, Zhang H, Duan W, Liang T, Wu S. Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing. Mech Syst Signal Process 2018;100:743–65. http://dx.doi.org/10.1016/j.ymssp.2017.08.002.

[43] Hu Z, Wang Y, Ge M, Liu J. Data-driven fault diagnosis method based on compressed sensing and improved multiscale network. IEEE Trans Ind Electron 2020;67(4):3216–25. http://dx.doi.org/10.1109/TIE.2019.2912763.

[44] Abo-Zahhad MM, Hussein AI. Compressive sensing algorithms for signal processing applications: A survey. Commun Netw Syst Sci 2015;197–216.

[45] Lu Y, Wang Y. Monitoring temperature in additive manufacturing with physics-based compressive sensing. J Manuf Syst 2018;48:60–70. http://dx.doi.org/10.1016/j.jmsy.2018.05.010.

[46] Kim YC, Powers EJ. Digital bispectral analysis and its applications to nonlinear wave interactions. IEEE Trans Plasma Sci 1979;7(2):120–31. http://dx.doi.org/10.1109/TPS.1979.4317207.

[47] Gu F, Shao Y, Hu N, Naid A, Ball AD. Electrical motor current signal analysis using a modified bispectrum for fault diagnosis of downstream mechanical equipment. Mech Syst Signal Process 2011;25(1):360–72. http://dx.doi.org/10.1016/j.ymssp.2010.07.004.

[48] Xu Y, Fu C, Hu N, Huang B, Gu F, Ball AD. A phase linearisation–based modulation signal bispectrum for analysing cyclostationary bearing signals. Struct Health Monit 2021;20(3):1231–46. http://dx.doi.org/10.1177/1475921720949827.

[49] Xu Y, Feng G, Tang X, Yang S, Gu F, Ball AD. A modulation signal bispectrum enhanced squared envelope for the detection and diagnosis of compound epicyclic gear faults. Struct Health Monit 2022. http://dx.doi.org/10.1177/14759217221098577.

[50] Zhang R, Gu F, Mansaf H, Wang T, Ball AD. Gear wear monitoring by modulation signal bispectrum based on motor current signal analysis. Mech Syst Signal Process 2017;94:202–13. http://dx.doi.org/10.1016/j.ymssp.2017.02.037.

[51] Baraniuk RG. Compressive sensing [lecture notes]. IEEE Signal Process Mag 2017;24(4):118–21. http://dx.doi.org/10.1109/MSP.2007.4286571.

[52] Yang Y, Zheng H, Li Y, Xu M, Chen Y. A fault diagnosis scheme for rotating machinery using hierarchical symbolic analysis and convolutional neural network. ISA Trans 2019;91:235–52. http://dx.doi.org/10.1016/j.isatra.2019.01.018.

[53] Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. In: 2016 IEEE conference on computer vision and pattern recognition. Las Vegas, NV, USA; 2016, p. 2921–9. http://dx.doi.org/10.1109/CVPR.2016.319.

[54] Cevher V, Duarte MF, Baraniuk RG. Distributed target localization via spatial sparsity. In: Signal processing conference, 2008 16th European. 2008, p. 1–5.

[55] Duarte MF, et al. Single-pixel imaging via compressive sampling. IEEE Signal Process Mag 2008;25(2):83–91.