Rail Suspension System Fault Detection using Deep Semi-Supervised Feature Extraction with One-class data

Xiaomeng Peng[1], Xiaoning Jin[2]

[1,2]*Northeastern University, Boston, MA, 02115, United States*
Peng.x@northeastern.edu
xi.jin@ northeastern.edu

ABSTRACT

In this paper we propose a novel semi-supervised fault detection methodology for a vehicle suspension system with one-class multi-sensor data. Supervised data-driven methods have been applied in fault detection successfully in recent studies. However, it is difficult and expensive to collect data under faulty condition for supervised learning while data collection under normal condition is much easier and cheaper. Fault detection under such situation is a one-class classification problem that requires classification models to identify the positive class when the negative class is either absent or poorly sampled. The efficiency of classifiers is constrained by defining the normal class boundary only with the knowledge of positive class as well as the problem of biased or incorrect feature extracted from the positive class. In proposed method, A deep semi-supervised learning method integrated with physical-based domain knowledge is applied for feature extraction. The suspension system for a full car model is modeled using a simulation tool, SIMPACK to generate the synthetic multi-sensor data. Our results show the effectiveness of the proposed in fault detection and diagnostics with one-class data.

1. INTRODUCTION

Condition monitoring with online fault detection and identification(FDI) has become a domain of interest in the few last decades in many applications, such as aerospace, automotive, nuclear and chemical engineering. Accurate fault detection in early stage assists systems to retain higher safety and to provide time and useful information for maintenance crew to schedule the maintenance, avoiding unplanned down-time and unscheduled maintenance.

In general, two major categories of these approaches have been developed for online fault detection and identification:

model-based and data-driven. Each category of methods has its own advantages and limitations and consequently they are often combined in practical applications called hybrid model. Model-based techniques are applicable when accurate mathematical process or dynamic model is available. In this work, we narrow ourselves down to the data-driven method because of the non-linearity and parameters unknown for system.

Intensive data-driven approaches have been employed recently, for example, PCA-T^2 (Qin,2003), Self-Organizing Map (SOM) (Lapia, Brisset, Ardakani, Siegel and Lee, 2012) and Neural Network (Sobie, Freitas, Nicolai,2018). Detailed reviews of data-driven faulty assessment in rolling bearing can be found in Cerrada, Sánchez, Li, Pacheco d, Cabrera, and Oliveira et al. (2018). The performance of these machine learning approaches depends heavily on the representation of the data that are given. Statistical learning and traditional machine learning methods are applied to predict and classify the faults based on the features extracted from the raw data. Types of features in industrial application involve time domain, frequency domain and time-frequency domain from raw sensor measurements utilizing proper signal processing techniques. Kimotho and Sextro (2014) summaries some possible time domain and frequency features for vibration signals. However, the natural of nonstationary and nonlinearity of a system requires techniques to capture the local time-frequency properties of the vibration. Variety of approaches are used to extract time-frequency domain knowledge. Wavelet transform(WT) is widely used in feature extraction. Loutas and Kostopoulos (2012) reviews the utilization of wavelet transform in condition-based maintenance. Hilbert-Huang Transform (HHT), Empirical Mode Decomposition(EMD), short time Fourier transform(STFT) and Wavelet Packet Transform (WPT) are also common tools for time-frequency features (Loutas et al. 2012). In practical fault detection problem, extracted representation of the data is predesigned with domain knowledge and manual decision. As such bias and uncertainties are usually introduced by hand-craft feature extraction, possibly leading to inaccurate classification

results. Moreover, unlike working on certain known mechanical systems with sufficient information, hand-crafted feature extraction techniques do not surely recognize discriminative features when working on a new or different complex system.

To resolve the feature selection bias problem, deep neural network-based classification and feature extraction are in the interest of research to improve the classifier performance. Compared to traditional machine learning methods, deep neural network (DNN) attempts to explore the high-level abstractions in data using multiple processing layers with complex structures, resulting in better representations of the inputs by learning from input examples (Lu, Wang, Qin and Ma, 2017). Nowadays, deep neural networks are successfully applied in fault detection area. For example, continuous wavelet transform (CWT) and convolutional neural network (CNN) are collaborated to detect earth fault in resonant grounding distribution systems (Guo, Zeng, Chen and Yang, 2018). He and He (2017) developed a deep learning LAMSTAR network to diagnostic the bearing faults. Verstraete, Ferrada, Droguett, Meruane and Modarres (2017) used CNN for rolling element bearing fault diagnosis. Zhang, Peng, Wu, Yao and Guan (2017) presented a deep neural networks based fault diagnosis for raw bearing data. Lu et al. (2017) proposed a rotary machinery components fault diagnosis based on features extracted by stacked denoising autoencoder.

The success of DNN-based feature extraction relies on abundance and sufficiency of labeled historical data. In practice, labeling by expertise is highly time consuming and expensive. In addition, faults of the system are not always analyzed exhaustively. Scarcity of samples from faulty conditions will unbalance the data. For example, mission-critical assets such as airplane engine and nuclear power plane are highly reliable and have few data from failure scenarios. Thus, the labeled historical data set is highly unbalanced as it contains data only from nominal condition, namely, a one-class classification or abnormal detection problem. In the classical supervised multi-class classification, features are learned with the objective of maximizing inter-class distances between classes and minimizing intra-class variances within classes solely utilize labeled data. However, in the absence of multiple classes such a discriminative approach is not possible (Perera and Patel 2018). In the lack of sufficient fault status information, the traditional supervised deep learning techniques are not applicable in this situation.

Semi-Supervised Learning (SSL) has attracted attention because of its utilization not only from the labeled training data, but also from the structural information in unlabeled data(Zhu,2008). Semi-Supervised feature extractor aims to extract information from both labeled data and unlabeled dataset. Thus, classifier using these extracted features will be more separable. Manifold regularization (MR), a

conventional technique in SSL, is first introduced to fault detection area by Yuan and Liu (2013). The improved predictive performance implies the effective of techniques. Jiang, Xuan and Shi (2013) proposed a features extraction method based on semi-supervised kernel marginal fisher analysis. The low-dimension features then fed into simple classifier to isolate the bearing fault.

In this paper, we proposed a fault detection method with deep semi-supervised feature extraction to improve the accuracy of fault detection and apply the methods in rail suspension system for demonstration and validation. The main contribution of this work is that we propose a semi-supervised deep learning-based feature extraction method to extract distinct high-level features. The extracted features are then used for detection purpose. Compared with supervised deep learning, this work uses DNN only as feature extractor not a classifier since it is impossible to use multi-class classifier in a one-class classification problem directly. Different from unsupervised deep learning, i.e. autoencoder, proposed method utilizes the label information from labeled data which increases the accuracy for the detection. Thus, the extracted features are more distinct by using proposed semi-supervised feature extraction.

The remainder part of this paper organized as follow. In section 2, problem is introduced. Proposed methodology is outlined in section 3. In section 4, a simulation-based rail suspension system case study is performed. The result of the implementation is stated. The main conclusions of this study are described in section 6.

2. PROBLEM STATEMENT

In this paper, a deep semi-supervised feature extraction method combined with domain knowledge is proposed to improve the accuracy of fault detection. We demonstrate the efficacy of the method on a rail vehicle suspension system considered as time-varying dynamic system. Details of the physical model are described in section 4.1. Experiments are generated as training dataset and test dataset. Each experiment has 18 acceleration sensor measurements with corresponding label. Let x_i^j denote the i_{th} experiment value of sensor j, in which $x_i^j \in \mathbb{R}^m$, $i = 1, ..., n$ and $j = 1, ..., m$. n is the number of experiments. m is the number of sensors. Furthermore, let $\mathrm{x}_i = [x_i^1 \quad x_i^2 \quad ... \quad x_i^m]$. The sensor measurement matrix is $\mathbf{x} = \{\mathrm{x}_1 \quad \mathrm{x}_2 \quad ... \quad \mathrm{x}_n\}$, $\mathrm{x}_i \in \mathbb{R}^n$. Corresponding label set $\mathbf{y} = \{\mathrm{y}_1, \mathrm{y}_2, ... \mathrm{y}_n\}$ are the label of the experiments. The complete dataset is shown as $D = \{(\mathrm{x}_i, \mathrm{y}_i)\}_{i=1,...n}$.

In our problem setting, 200 experiments are conducted under nominal condition and *labeled* as "healthy", while 100 experiments are operated under faulty conditions with different failure modes are *unlabeled*.

The objective of this paper is to collaboratively use semi-supervised deep learning for feature extraction and one-class classification for fault detection. The one-class classification problem is to recognize instances of a concept by only using examples of the same concept (Perera et.al ,2018). In such scenario, instances of only a single object class, i.e. experiments with healthy condition, are labeled. During testing, the classifier may encounter instances from other classes, i.e. experiments with different failure mode. The goal of the classifier is to distinguish objects of the known class from objects of other classes. The proposed fault detection method with semi-supervised feature extraction aims to learn the non-linear mapping from original data to feature representation and further achieve more accuracy prediction given unlabeled data.

3. PROPOSED METHODOLOGY

In this section, a fault detection method with semi-supervised learning feature extraction is presented in detail.

As stated in section 1, deep neural network is a promising way to extract the representative features. However, for the scenario discussed in this paper, the labels of the labeled datasets are one-class that consisting of nominal conditions. Therefore, it is impossible to adopt a supervised CNN architecture used for classification directly. Unsupervised learning technique such as autoencoder and variants of autoencoder is commonly used in feature extraction for one-class classification. It could not take use of the known labels during the learning process. To maximize the utilization of the information from both labeled and unlabeled data, a feature extraction method in semi-supervised fashion is considered.

The Semi-supervised feature extraction method aims to learn an accurate latent space representation. This representation is not only an indicator for nominal/healthy scenario, but also a sufficient description for various fault scenarios.

A fault detection method with semi-supervised learning feature extraction is proposed as follows. The layout procedure is provided in Figure. 1.

1) Data preparation: Generate the time-frequency domain features. The raw acceleration measurements are pre-processed to extract time-frequency features in our specific case. Given a data set which contains labeled data and unlabeled data, pre-processing is performed using short time Fourier transform (STFT). The output of the preprocessing is the spectrogram of frequency response.

2) High-level feature extraction: Use deep neural network to extract high-level features from obtained time-frequency features. The selection of neural networks is based on application. In our case, since time-frequency features are used, Convolutional

Autoencoder(CAE) and Convolutional Neural Network(CNN) are selected and used in semi-supervised fashion. Then CAE captures the data representation from features of both labeled and unlabeled data set. CAE passes the initial weights to CNN. CNN is then fine-tuned using features only from labeled data set. The trained CNN as a feature extractor is used to extract feature for data set. The activations for the fully connected are high-level features. The reason for selection and detailed procedure will be discussed in section 3.2.

3) Fault Detection: Train one-class classifiers using high-level features. Predict the unlabeled experiment using trained one-class classifiers.
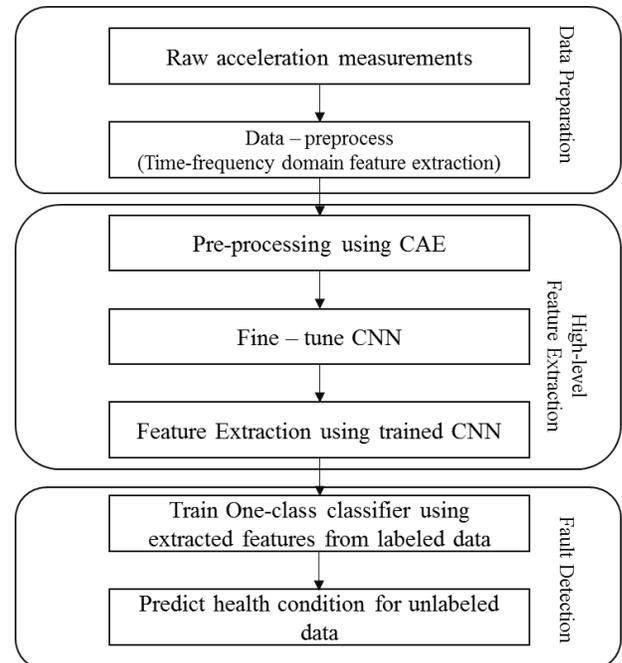

Figure 1. General procedure of proposed method

3.1. Time-frequency domain Features

Since frequency domain features are the indicator of faults and system is time-varying, spectrum matrices are needed to capture the non-stationary characteristic. Time-frequency features are changing along with time, which provides helpful information for further analysis.

If the labeled dataset is a balanced dataset, it is possible to select relevant frequency bands using feature selection techniques, such as, sequential forward and backward selection, statistical test and heuristic method. Usually, the selection is based on optimization of the classifier performance. However, under our assumption, labeled dataset is no longer balanced and corresponding problem is no longer a supervised multi-class classification problem.

Feature selection techniques are not applicable to choose optimal frequency bands.

To mitigate this problem, spectrum matrix can be used directly as a time-frequency feature. These spectrum matrices are converted into grey-scale images and then used as the input of high-level feature extraction.

3.2. High Level Feature Extraction

With physical domain knowledge, system fault will change the spectrum matrix. Images which are converted from spectrum matrices will then change accordingly. An efficient way to detect the abnormality of images is the Convolutional Neural Network (CNN). However, as stated in the beginning of section 3, semi-supervised feature extraction can fully utilize the information to learn the non-linear mapping to lower dimension space.

Unsupervised learning is used to map input $\mathbf{x}$ to lower dimension with labeled and unlabeled samples. Erhan, Bengio, Courville, Manzagol and Vincent (2010) proved the effectiveness of the role of unsupervised pre-training. In traditional way, unsupervised pre-training is used to improve the generalization error of trained deep architectures. From 2010, it is more common to train a deep neural network since the size of data has increased drastically. However, if the problem is becoming complex with little available labeled training data but plenty of unlabeled training data, pre-training is still effective.

Since the training data set is limited in the scenario discussed in this paper, pretraining of data is necessary and effective. The most challenging part of the problem is the absence of a discriminative feature for the labeled training data. Without knowing the information of faulty conditions, the feature detectors may fail to extract the useful information solely from healthy data to detect the anomaly

In this light, unsupervised convolutional autoencoder with labeled and unlabeled data set is used to pretrain the convolutional neural network. The aim of pre-training is to extract the distinct representation from normal condition. The weights of CAE are the initialization of CNN. Training CNN with labeled data fine-tunes the weights of layers. The final trained CNN is the feature extractor to learn latent space.

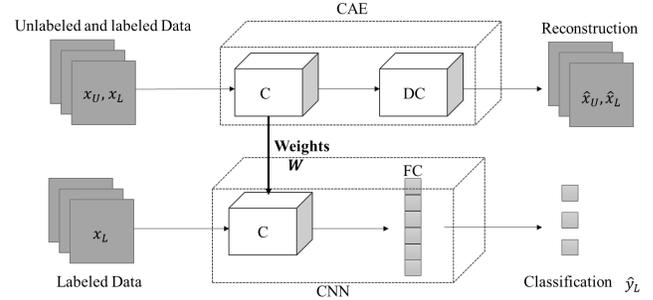A lay-out of high-level feature extraction process is shown in Figure 2.

Figure 2. High-level feature extraction process of proposed method

3.2.1. Unsupervised Pre-training using Convolutional Autoencoder (CAE)

An auto-encoder (AE) model is constructed based on an encoder-decoder paradigm. An encoder transforms an input from a high-dimension efficient feature space into the lower dimension one. Then a decoder reconstructs the data from resulting feature space. Encoder and decoder are symmetric in architecture. Autoencoder performs as a power feature decoder and can be used for unsupervised learning. The objective of autoencoder is to minimize the difference between reconstructions $\hat{\mathbf{x}}$ and the inputs $\mathbf{x}$. The objective function can be written as

$$\mathcal{L}(\mathbf{x}, \hat{\mathbf{x}}) = \|\mathbf{x} - \hat{\mathbf{x}}\|$$

The success of CNN in supervised image classification inspired Masci, Meier, Ciresan and Schmidhuber (2011) to propose a new architecture for autoencoder - Convolutional autoencoder(CAE). CAE learns the visual features by reconstructing images through convolutional layers. Traditional AE used fully connected layer as encoder and decoder whereas CAE changed into convolutional layers.

In general, CAE is a special type of CNN. However, the objectives of two neural networks are different. CNN is supervised learning classifier which aims to classify the input while CAE is an unsupervised learning method to extract the high-level representation of the images. Figure 3 indicates the general architectural of a convolutional autoencoder.
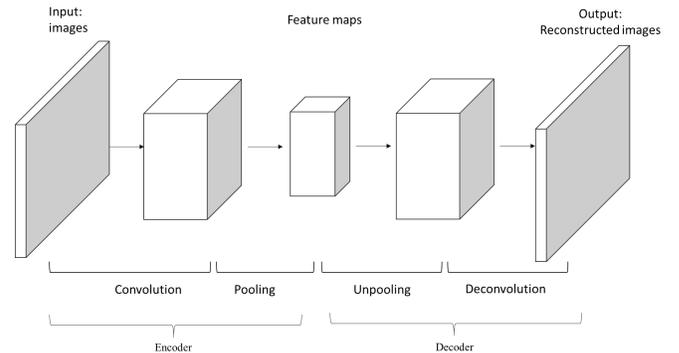

Figure 3. General architectural of CAE

The encoder of ACE is exactly the same with CNN (see section 3.2.2). During the decoder process, unpooling and decovolution are often used. Unpooling is denoted reverse max pooling. In the convolutional network, the max pooling operation is non-invertible. However, by recording the locations of the maxima within each pooling layer, the unpooling operation are able to place the reconstructions from the layer into appropriate locations, preserving the structure of the stimulus (Zeiler and Fergus, 2013). Deconvolution in convolutional neural network means the inverse process of convolutional, which is the same with convolutional layer.

Since the convolutional encoder is used as pre-training process, it should have same architecture with CNN by natural. The proposed CAE architecture is shown in Figure 4.
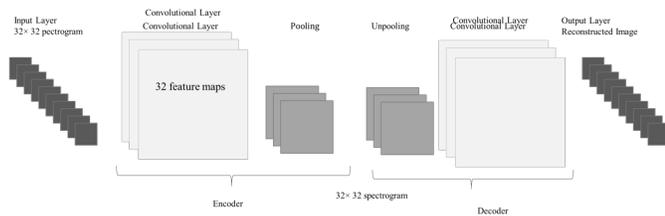

Figure 4. Proposed CAE architecture

3.2.2. Fine – tuning Weights using Convolutional Neural Network(CNN)

Convolutional neural network(CNN) is a specialized category of neural network for processing data that has a known grid-like topology. CNN has been succeeded being applied in image classification, speech recognition and DNA sequences classification. Convolution leverages three important ideas that can help to improve a machine learning system: sparse interactions, parameter sharing and equivariant representation (Goodfellow, Bengio and Courville, 2016). A regular CNN architecture contains three main types of layers: Convolutional Layer, Pooling Layer, and Fully-Connected Layer. A full CNN shown in Figure 5 is constructed by stacking these layers.
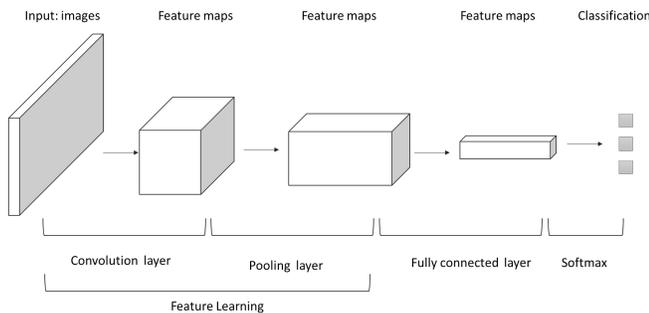

Figure 5. General architecture of CNN

The input of CNN is an image which represents as an array of pixel value. Convolutional layers utilize the filter as feature detectors to capture the pixel structure of the original input image.

The following layer is a pooling layer which effectively reduces the number of parameters and amount of computational burden, but it retains the important information in the work. The convolutional layer is similar in respect to feature construction. the pooling layer in a CNN could be related to a feature selection layer (Verstraete et al.,2017).

The output of the convolutional layer and pooling layers is the high-level feature representation of the original image. Final layer, fully connected layer, uses these features to classify the input data.

In this work, as stated before, CNN as a supervised multi-class classifier cannot be used directly for one-class data. As an alternative, time-frequency features, i.e. spectrograms from different subsystem are considered as different classes. Thus, the whole system is partitioned into multi-subsystems. The one-class problem is therefore transferred into multi-classes problem.

The CNN architecture we proposed in this work is shown in Figure 6. CNN is first initialized using weights from CAE. Then labeled data is used to fine-tune the weights in CNN.
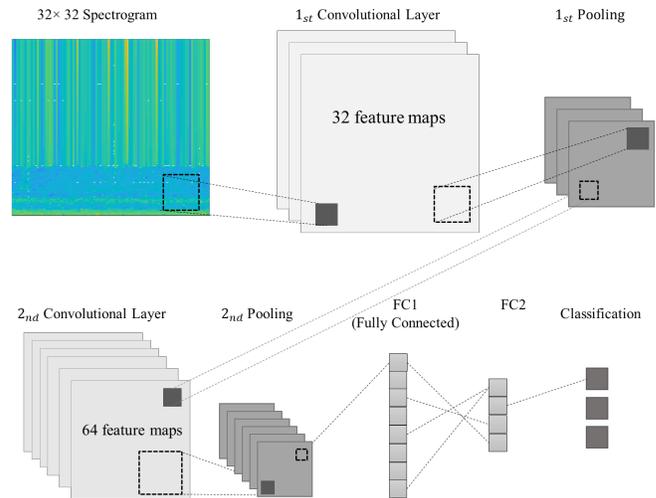
Figure 6. Proposed CNN architecture

3.3. Training Process and Fault detection

The proposed testing procedure involves two phases-feature extraction and detection. First, during feature extraction phase, A set of features G are extracted from labeled dataset and unlabeled data. The extracted features are stored as representation and will be used in the detection phase.

One-class classifier, such as one-class Support Vector Machine(SVM) and K-nearest neighbor can be used to

trained using representation. One-class SVM classifier is used in this paper. The classifier will detect whether one specific subsystem is fault or not. Thus, multiple one-class classifiers are trained for each subsystem. If no subsystem has predicted as fault, the testing experiment is considered as healthy. Otherwise, this experiment is predicted as fault.

Figure 7. Testing process for proposed method

4. SIMULATED CASE STUDY: RAIL VEHICLE SUSPENSION SYSTEM

In this section, a simulated case study will be outlined to validate the proposed method. The physical model of suspension system is first introduced, followed by the setup of simulation and preparation of dataset. After data preprocessing, results are provided and analyzed.

4.1. Physical Model of Suspension System

Suspension system, consisting of tire, spring, dampers and linkages, plays a critical role in rail vehicle. The suspension system supports the car body and bogie, to isolate the forces generated by the track unevenness at the wheels and to control the altitude of car body (Wei, Jia, Liu, 2013). It is an important component for comfortable riding as well as easily handling. The fault occurred at spring or dampers would decrease the level of system stability while passing through curves. What's more, it would also cause a potential loss of contact between the vehicle to the road.

The traditional rail vehicle suspension system is under investigation in this paper. As shown in Figure 8, full car model consists of four wheelsets, two bogies (leading and trailing bogie) and a car body. The complete system has 23 degrees of freedom.

Figure 8. Simplified full-car model (PHM 2017 data challenge)

The standard vector matrix form of the full- car state space model is

$$M\ddot{Z} + C\dot{Z} + KZ = C_u\dot{U} + D_uU$$

where Z is the vertical displacement vector of bogie and carbody. U is input of wheelsets. M, C, K are mass, damping and stiffness matrices, respectively. C_u, D_u are damping and stiffness matrices of excitations. (Girstmair, Heigermoser and Rosca, 2017).

Inspired by Li, Liu, Tian, Cui and Wu (2017), the full-car model can be considered as 8 independent subsystems. Each subsystem is a quarter car model with 2 DOF. Detailed subsystem is same with Li et al. (2017). Figure 9 shows the structure of such subsystem.

Figure 9. 2 DOF subsystem

The transfer function of secondary suspension is

$$G_s(s) = \frac{Z_s(s)}{Z_p(s)} = \frac{C_s s + K_s}{m_s s^2 + C_s s + K_s}$$

The transfer function of primary suspension is

$$G_p(s) = \frac{Z_p(s)}{W(s)}$$
$$= \frac{C_p s + K_s}{m_p s^2 + [C_p + C_s(1 - G_s(s))]s + [K_p + K_s(1 - G_s(s))]}$$

Parameter degradation could change both primary and secondary transfer function and therefore change the gain of the system. The gain system, also known as frequency response, is defined as the ratio the steady-state output amplitude to and the steady-state input amplitude. It is used to characterize the dynamics of the system. As a quantitative measure of the output spectrum in response to input, frequency response is widely used to characterize the dynamics of the system.

4.2. Simulation Setup and Dataset Preparation

It is necessary to design a systematic experiment to validate the proposed method stated in section 3. The data set to be used in this paper is simulated by multibody simulation toolkit SIMPACK. SIMPACK is a general purpose 3D Multibody Simulation (MBS) software designed to simulate non-linear mechanical systems, analyze vibrational behavior,

calculate forces and accelerations, and describe and predict the motion of any complex multibody system(Iwnicki,2006).

The model of rail vehicle suspension system has been stated in section 4.1. Eighteen sensors are introduced arbitrarily to measure the vertical acceleration. The locations of these sensors are highlighted in red dots in the Figure 8. The sampling frequency is 200Hz. The vehicle runs on the track at 70 km/h (with randomness of ± 5 km/h). Total simulation duration is 30s, i.e. there will be 6000 measurement sets. The track in the simulation is the normal rail gauge of 1435 mm with UIC60 profile. Three directions (vertical, lateral and rolling) track irregularity are introduced of the fifth-grade track irregularity spectrum of the US railway lines. The simulation is conducted in the track section contains curved section with superelevation. The aerial view of the track is shown in Figure. 10
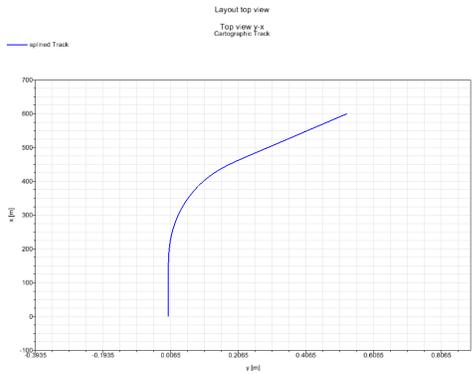
Figure 10. Aerial view of track

Failure modes of the system are simulated as the attenuation of component vertical parameters. Here components refer to the springs and dampers in primary and secondary suspension system. The degradation level is from 70% to 10% of original status.

Labeled training dataset consists of 200 healthy experiments. Testing dataset consists 100 experiments including both healthy and faulty scenarios. Note that for healthy experiments, parameters (stiffness, damping ratio and carbody mass) vary in ± 10% range to increase the level of uncertainty.

4.3. Time-Frequency Feature Extraction

The detailed time-frequency feature extraction procedure is presented in this subsection. As described in subsection 4.1, frequency response can represent the major characteristics of a time-invariant system. However, rail vehicle suspension system is as assumed to be a time-varying system. Besides, the connections between eight subsystems are tight. The roll and pitch motion of vehicle can slightly change the position of the center of gravity. As a result, M_s and M_p of a subsystem are always changing along with the running of the vehicle. This phenomenon is known as the load transfer and shown in Figure 11. To mitigate the impact of such

phenomena, time-frequency features are required to capture the time-varying behaviors of the vehicle.
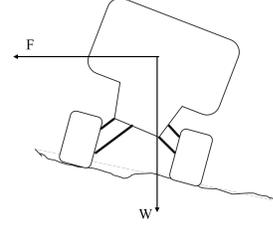

Figure 11. Load transfer of suspension system

During the detection process, 16 frequency responses are captured for 8 subsystems. They are then extracted to spectrograms based on domain knowledge.

The procedure of time-frequency domain feature extraction for each subsystem is provided as follows. A corresponding flowchart is shown in Figure 12:

1) The raw vertical acceleration measurements of input $x(t)$ and output $y(t)$ (e.g. sensor reading from primary suspension and sensor reading from wheelset in the same subsystem) are filtered by a low-pass filter. As discussed by (Mei, 2008), the natural frequencies of the bogies modes are normally found between 7Hz and 20 Hz. To be on the safe side, a lowpass filter 30Hz is applied.

2) Divide filtered raw measurements into N_{seg} segments with L points of each. A fixed overlapping time interval is also introduced between segments. For each segment, system is considered to be time-invariant since the elapsing time is relatively short. The time window is chosen to be around 1 second with half second overlapping.

3) Calculate the frequency response for each subsystem.

$$H(e^{j\omega}) = \frac{S_{xy}(e^{j\omega})}{S_{xx}(e^{j\omega})}$$

where $S_{xy}(e^{j\omega})$ is the estimated cross spectral density in the frequency domain of input $x(t)$ and output $y(t)$. $S_{xx}(e^{j\omega})$ is the estimated autocorrelation of input $x(t)$ in frequency(or equivalently the power spectral density of input)

4) Time-frequency feature is the amplitude of $H(e^{j\omega})$

5) Scaling the spectrograms into grey-scale 32×32 pixel images

The resulting images are then used as an input to a deep neural network. Since the calculated time-frequency feature is a good representation of the estimated frequency response and the transfer function, the extracted features are capable to detect the degradation of components.
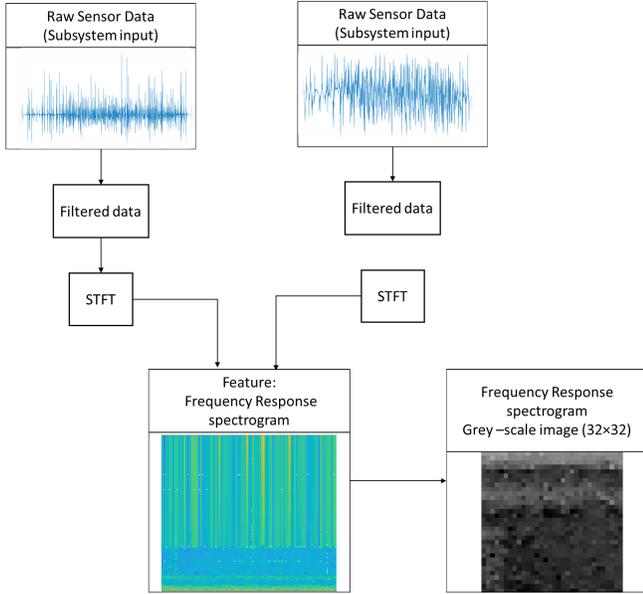
Figure 12. Flowchart of data pre-processing.

4.4. Simulation Results

The results of the proposed semi-supervised feature extractor are obtained and analyzed in this subsection.   As a comparison, results from other 3 methods are carried out.

1) Low-level features (LLF). After generating the time-frequency features (i.e. Data-preparation step in proposed method), the one-class SVM is used directly to detect the fault directly without further feature extraction

2) High-level features from CAE (HLF-CAE). The input of CAE is grey scale images. CAE is trained using labeled and unlabeled data. The high-level features are solely extracted by unsupervised CAE.

3) High-level features from CNN (HLF-CNN). Labeled healthy data are used to fine tune the supervised CNN.   Trained CNN as a feature extractor can be used to extract high-level features for final detection.   One-class SVM is used to identify healthy and faulty experiments. The trained weights of CNN are identical to those of CAE. The high-level features are only extracted by CNN.

The metric to assess the performance of the model is

$$Accuray = \frac{\text{The number of experiments correctly predicted}}{\text{Total number of testing experiments}}$$

The comparison of prediction accuracies among several methods is provided in Table 1. It is shown that the proposed deep semi-supervised method can predict the most number of experiments correctly.

Table 1. Prediction accuracies with different features

Model	Accuracy
LLF	0.69
HLF_CAE	0.74
HLF_CNN	0.71
HLF_SSL	0.85

The confusion matrices calculated based on all the four methods are shown in Figure13, Figure14, Figure15 and Figure 16 respectively.
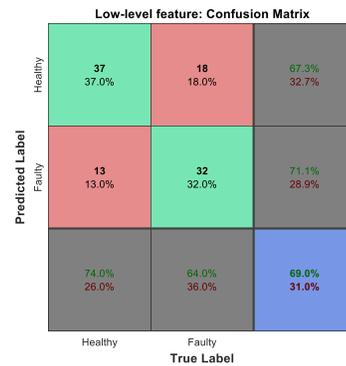

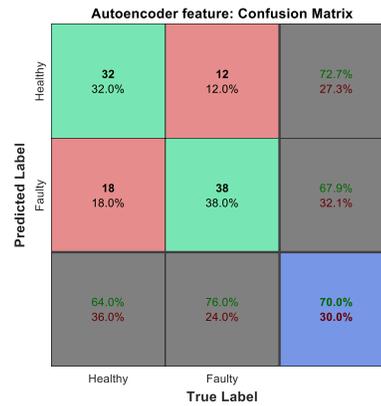Figure 13. Confusion matrix with low level features
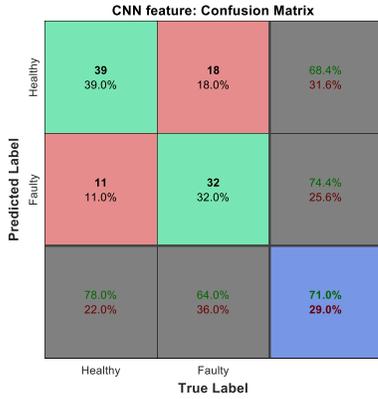

Figure 14. Confusion matrix with CAE features

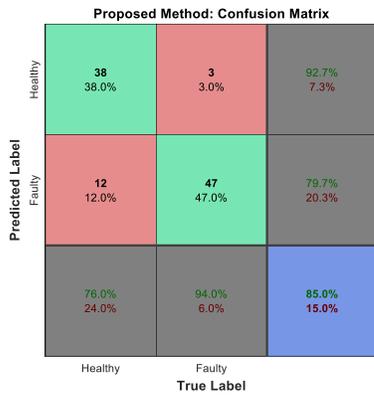Figure 15. Confusion matrix with CNN features

Figure 16. Confusion matrix with proposed semi-supervised features

From the confusion matrices we can see that, the proposed method achieves best result of accuracy, especially in detecting faults. One explanation is that the proposed method maximizes the use of available information.

For CAE, autoencoder can learn mapping function from input to lower dimension space. This mapping is not related to the labels. Thus, autoencoder did not show significant improvement compared with low-level feature experiment. The CNN features experiment shows the effectiveness of supervised learning. The high accuracy in predicting healthy experiments is because the trained CNN can extract representation that clusters the normal conditions together. However, it is hard to differentiate healthy and faulty conditions using only CNN high-level features. Overall, the proposed method outperforms the other models.

One contribution of our problem is to learn a representation that can distinguish nominal conditions and faulty conditions. For instance, feature spaces of extracted features from right side secondary suspension system of leading bogie are compared in Figure 17, 18, 19 and 20. In the figures, '×' indicates normal samples and '○' indicates

faulty samples. The 2D visualization is performed using t-SNE (Matten and Hinton, 2008).

Figure 17 is a 2D visualization of the extracted 1024-dimensional low-level features, i.e. 32×32 spectrogram. It is hard to separate the normal and abnormal conditions from low level features directly. Data from one-class are not in the same cluster, indicating that low-level feature are not suitable for detection.
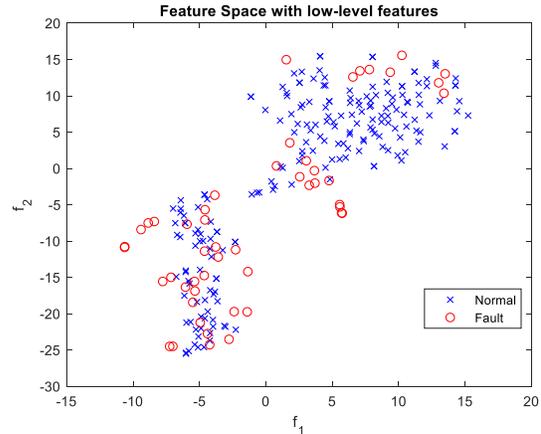

Figure 17. Feature space obtained without high-level feature extraction (Low-level features experiment)

Figure 18 shows the 2D visualization of obtained high-level 50-dimensional features extracted only with unsupervised CAE. Without fine-tune by supervised CNN, data from one class (normal or abnormal class) cannot form a cluster that covers very few data points of another class.
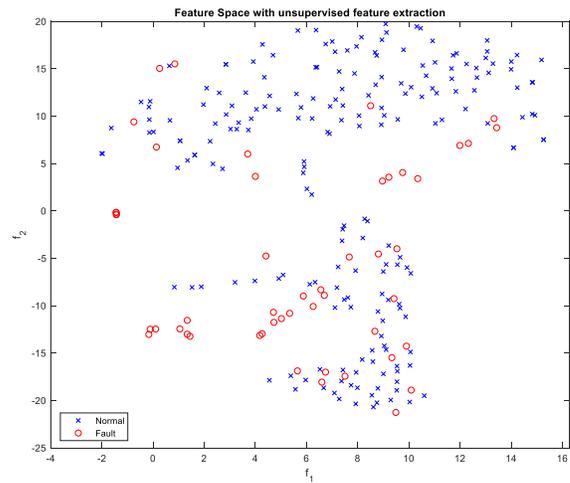

Figure 18. Feature space obtained only using the unsupervised CAE (CAE features experiment)
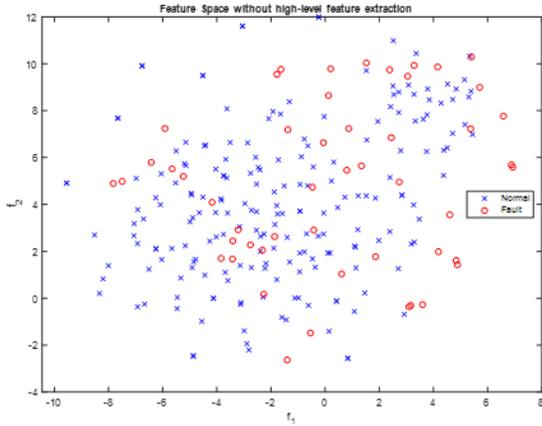
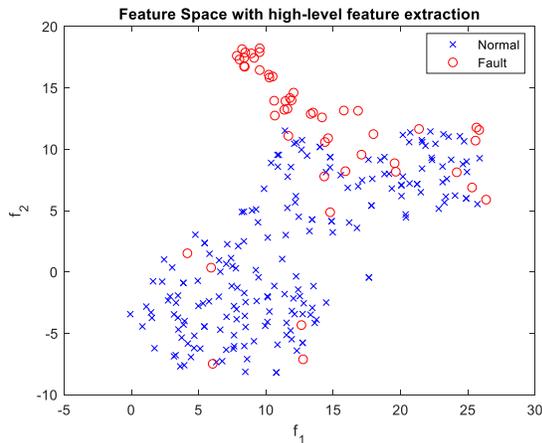Figure 19. Feature space obtained only using CNN (CNN features experiment)

Figure 20. Feature space obtained using the proposed method (Semi-supervised features experiment)

Figure 19 is the 2D visualization of the extracted 30-dimensional high-level features extract solely with CNN. Since the data from one class is sufficient to extract data representation, the data from normal conditions are clustered together in the extracted feature space. However, the data from faulty conditions are in the same cluster too. Normal and faulty samples are not able to be separated.

Figure 20 is the 2D visualization of the extracted 30-dimensional high-level features extracted with proposed method. Samples from normal condition and faulty condition are more separable than Figure 15,16,17. Although it seems some fault samples still fall in the normal boundary, it is because the fault refers to system level, but in this subsystem, the data from this faulty experiment are still within nominal range. The semi-supervised feature extraction did show the ability to separate the nominal condition and faulty condition.

5. CONCLUSION

In this paper, a fault detection method with deep semi-supervised feature extraction is proposed to solve for one-class classification problem. It is better than traditional supervised and unsupervised learning methods in sense of performance and utilizing more information

Semi-supervised feature extraction keeps the information from labeled and unlabeled data to the utmost and improve the performance of the fault detection. The presented method preprocesses the acceleration measurements using short time Fourier transform(STFT) to extract low-level time-frequency domain features. Deep Semi-supervised feature extractor is then introduced to extract high-level representation. The one-class classifier is implemented finally to detect the fault. A rail vehicle suspension system model is constructed, and numerical simulations are conducted. Compared with three other prevailing methods, results indicate the effectiveness and better performance of proposed method.

REFERENCES

Cerrada M. *et al*(2018)., A review on data-driven fault severity assessment in rolling bearings, *Mech. Syst. Signal Process.*, vol. 99, pp. 169–196 ,2018.

Erhan D., Courville A., and Vincent P.(2010), Why Does Unsupervised Pre-training Help Deep Learning ?, *J. Mach. Learn. Res.*, vol. 11, pp. 625–660, 2010.

Girstmair B., Haigermoser A., and Rosca J.(2017), Combination of Data-driven Feature Selection Methods with Domain Knowledge for Diagnosis of Railway Vehicles, *Annu. Conf. Progn. Heal. Manag. Soc.*, pp. 1–10, 2017.

Goodfellow I., Bengio Y., Courville. A.(2016) *Deep Learning*, MIT Press ,2016

Guo. M.-F., Zeng. X.-D., Chen. D.-Y.and Yang. N.-C.(2017) Deep-learning-based Earth Fault Detection using Continuous Wavelet Transform and Convolutional Neural Network in Resonant Grounding Distribution Systems, *IEEE Sens. J.*, vol. 18, no. 3, pp. 1–1, 2017.

Hanxin C., Yang H., Chua P., and Hian L. G. (2010), Fault Diagnosis," vol. 53, no. 2, pp. 0–5, 2010.

Iwnicki S. (2006). Simulation. In Polach O., Berg M. and Iwnicki S. *Handbook of railway vehicle dynamics* (pp.359-423). Boca Raton London New York.

Jiang L., Xuan J., and Shi T(2013)., Feature extraction based on semi-supervised kernel Marginal Fisher analysis and its application in bearing fault diagnosis, *Mech. Syst. Signal Process.*, vol. 41, no. 1–2, pp. 113–126, 2013.

Kimothol J.K., and Sextro W. (2014). An approach for feature extraction and selection from non-trending data for machinery prognosis, *Annual Conference of the Prognostics and Health Management Society*. Sept 29 – Oct 02, Fort Worth.

Lapira E., Brisset D., Ardakani H. Davari, Siegel D., and Lee

J.(2012), Wind turbine performance assessment using multi-regime modeling approach, *Renew. Energy*, vol. 45, pp. 86–95, 2012.

Li. C., Liu. J, Tian. C., Cui. P., and Wu. M. (2017). Similarity-based Fault Detection in Vehicle Suspension System, *Annual Conference of the Prognostics and Health Management Society*. Oct 2 – 5, St. Petersburg, Florida, US.

Loutas T.and Kostopoulos V.(2012), Utilising the Wavelet Transform in Condition-Based Maintenance : A Review with Applications, *Adv. Wavelet Theory Their Appl. Eng. Phys. Technol.*, pp. 273–312, 2012.

Lu C., Wang Z. Y., Qin W. L., and Ma J. (2017), Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification, *Signal Processing*, vol. 130, pp. 377–388, 2017.

Masci J., Meier U., Cireşan D., and Schmidhuber J.(2011), Stacked convolutional auto-encoders for hierarchical feature extraction, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6791 LNCS, no. PART 1, pp. 52–59, 2011.

Perera. P. and Patel. V. M. (2018), Learning Deep Features for One-Class Classification, pp. 1–12, 2018.

Sobie C., Freitas C., and Nicolai M. (2018), Simulation-driven machine learning: Bearing fault classification, *Mech. Syst. Signal Process.*, vol. 99, pp. 403–419, 2018.

Verstraete.D., Ferrada. A., Droguett. E. L., Meruane V., and Modarres M.(2017), Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings," *Shock Vib.*, vol. 2017

Wei.X, Jia.L,Guo. K., and Wu.S.(2014), On fault isolation for rail vehicle suspension systems,*Veh. Syst. Dyn.*, vol. 52, no. 6, pp. 847–873

Yuan J. and Liu X.(2013), Semi-supervised learning and condition fusion for fault diagnosis, *Mech. Syst. Signal Process.*, vol. 38, no. 2, pp. 615–627, 2013.

Zeiler M. D.and Fergus R.(2014), Visualizing and understanding convolutional networks, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8689 LNCS, no. PART 1, pp. 818–833, 2014.

Zhang R., Peng Z., Wu L., Yao B., and Guan Y.(2017), Fault Diagnosis from Raw Sensor Data Using Deep Neural Networks Considering Temporal Coherence, *Sensors*, vol. 17, no. 3, p. 549, 2017.

Zhu.X.(2008), Semi-Supervised Learning Literature Survey Contents, *Sci. York*, vol. 10, no. 1530, p. 10, 2008.

BIOGRAPHIES

Xiaomeng Peng. holds a bachelor's degree on Electrical Engineering from Chongqing University, China. She received M.S Degree in Electrical Engineering from Northeastern University, Boston, USA. Currently, she is a Ph.D. student in the Department of Mechanical and Industrial Engineering at Northeastern University. Her research interests is the fields of Prognostics and Health Management and data analytics solutions for diagnosis and prognosis.

Xiaoning Jin is currently an Assistant Professor in the Department of Mechanical and Industrial Engineering at Northeastern University. She received her Ph.D. degree in Industrial and Operations Engineering from the University of Michigan. Her research interests are predictive analytics and decision support tools for manufacturing system operations and maintenance, manufacturing process monitoring, diagnostics, and prognostics using model-based and data-driven methods. Her works have been applied to a variety of industry applications ranging from automotive manufacturing, roll-to-roll printing process monitoring, smart operations and maintenance decision-making, etc.