Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.058051

ARTICLE

Engine Misfire Fault Detection Based on the Channel Attention Convolutional Model

Feifei Yu[1], Yongxian Huang[2,*], Guoyan Chen[1], Xiaoqing Yang[2], Canyi Du[2,*] and Yongkang Gong[2]

[1]School of Mechatronic Engineering, Guangdong Polytechnic Normal University, Guangzhou, 510665, China

[2]School of Automotive and Transportation Engineering, Guangdong Polytechnic Normal University, Guangzhou, 510665, China

*Corresponding Authors: Yongxian Huang. Email: yongxian@mail.ustc.edu.cn; Canyi Du. Email: ducanyi@gpnu.edu.cn

Received: 03 September 2024 Accepted: 23 October 2024 Published: 03 January 2025

ABSTRACT

To accurately diagnose misfire faults in automotive engines, we propose a Channel Attention Convolutional Model, specifically the Squeeze-and-Excitation Networks (SENET), for classifying engine vibration signals and precisely pinpointing misfire faults. In the experiment, we established a total of 11 distinct states, encompassing the engine's normal state, single-cylinder misfire faults, and dual-cylinder misfire faults for different cylinders. Data collection was facilitated by a highly sensitive acceleration signal collector with a high sampling rate of 20,840 Hz. The collected data were methodically divided into training and testing sets based on different experimental groups to ensure generalization and prevent overlap between the two sets. The results revealed that, with a vibration acceleration sequence of 1000 time steps (approximately 50 ms) as input, the SENET model achieved a misfire fault detection accuracy of 99.8%. For comparison, we also trained and tested several commonly used models, including Long Short-Term Memory (LSTM), Transformer, and Multi-Scale Residual Networks (MSRESNET), yielding accuracy rates of 84%, 79%, and 95%, respectively. This underscores the superior accuracy of the SENET model in detecting engine misfire faults compared to other models. Furthermore, the F1 scores for each type of recognition in the SENET model surpassed 0.98, outperforming the baseline models. Our analysis indicated that the misclassified samples in the LSTM and Transformer models' predictions were primarily due to intra-class misidentifications between single-cylinder and dual-cylinder misfire scenarios. To delve deeper, we conducted a visual analysis of the features extracted by the LSTM and SENET models using T-distributed Stochastic Neighbor Embedding (T-SNE) technology. The findings revealed that, in the LSTM model, data points of the same type tended to cluster together with significant overlap. Conversely, in the SENET model, data points of various types were more widely and evenly dispersed, demonstrating its effectiveness in distinguishing between different fault types.

KEYWORDS

Channel attention; SENET model; engine misfire fault; fault detection

1 Introduction

The engine constitutes a vital component of the automotive construction system, serving as the heart and power source of the automobile. Currently, the evolution of automotive engines towards increased complexity and high automation has led to more intricate and demanding operating

conditions. Consequently, the likelihood of encountering various faults is gradually rising. Among these, engine misfire fault diagnosis stands out as a focal point of research in engine fault diagnosis. Engine misfire occurs when the air-fuel mixture in the cylinder fails to undergo normal combustion due to internal component malfunctions during engine operation [1]. This incomplete combustion not only diminishes energy utilization efficiency but also results in the emission of substantial quantities of harmful gases, thereby contaminating the air environment [2]. Additionally, insufficient power can readily precipitate traffic accidents, posing a threat to human life. Therefore, systematically diagnosing engine misfire faults is of paramount importance. Over the past three decades, fault diagnosis methods have transformed from simplistic and uniform approaches to increasingly sophisticated and varied ones. Based on research mechanisms, these methods can be broadly categorized into five common diagnostic techniques [3]: expert systems, fault diagnosis grounded in analytical models, fault diagnosis relying on signal analysis, artificial intelligence methods, and statistical analysis methods. Presently, machine learning is the most prevalent diagnostic method employed by researchers in the realm of artificial intelligence.

In recent years, deep learning, as a pivotal branch of machine learning, has been extensively applied to various domains such as image recognition, facial recognition, speech recognition, computer vision, signal processing, and intelligent control [4]. During the evolution of deep learning, numerous researchers have employed neural network models for diagnosing and analyzing engine faults. Wang leveraged wavelet analysis for signal denoising and neural networks for diagnosing and identifying diesel engine faults [5]. Wang introduced a novel approach by adding a state feedback to the output layer of the backpropagation (BP) neural network, enhancing the accuracy of engine misfire fault diagnosis [6]. Zheng and colleagues integrated Fault Tree Analysis (FTA) with Support Vector Machine (SVM) algorithms to elevate the efficiency and precision of vehicle engine fault identification [7]. Wang et al. proposed the Parallel Online Sequential Regularized Extreme Learning Machine (POS-RELM) model, which is well-suited for online monitoring of engine faults [8]. Wang developed a fault diagnosis system for automotive engine misfires based on a probabilistic neural network. Experimental results validated that the trained Principal Component Analysis-Genetic Algorithm-Product-based Neural Networks (PCA-GA-PNN) method can precisely diagnose and locate single-cylinder and double-cylinder misfires, boasting simplicity, economy, efficiency, and high accuracy [9]. Gao et al. introduced a vehicle engine misfire fault diagnosis system leveraging wavelet packet correlation coefficient and Extreme Learning Machine (ELM), tailored for the non-stationary characteristics of cylinder head vibration signals. This method effectively captures fault-induced differences and accurately identifies single-cylinder misfires, characterized by high accuracy and short training time [10]. Han et al. experimentally demonstrated that utilizing the optimal wavelet packet basis function for feature extraction yields excellent results, and the Particle Swarm Optimization-Support Vector Machine (PSO-SVM) approach is also effective for recognition and diagnosis [11]. Chen proposed a diesel engine anomaly detection and fault diagnosis method based on Autoencoder depth feature extraction, surpassing other traditional methods in accuracy [12]. Gao et al. through diesel engine bench testing, proved that the Convolutional Neural Network (CNN)-based diesel engine misfire real-time diagnosis system achieves high diagnostic accuracy across a wide range of speed and load conditions [13].

Apart from that, some researchers have proposed an engine misfire diagnosis method grounded in torsional vibration and neural network analysis, capable of accurately diagnosing engine misfires [14]. Others have introduced the Single-Valued Neutrosophic Sets (SVNS) method, which precisely identifies the misfire fault state of engines [15]. Suda et al. discovered that combination classifiers can be employed for automated diagnosis of engine misfire faults [16]. Additionally, some have established

an engine misfire fault diagnosis model based on Probabilistic Neural Network (PNN), using the vibration acceleration signal of the engine cylinder block surface as the diagnostic parameter, yielding a highly accurate network model [17]. It is evident that current engine misfire fault diagnosis and recognition trends towards deep learning, achieving remarkable results and contributing significantly to this field.

Through a comprehensive literature review, we found that there has been a significant amount of research in the field of diesel engine fault diagnosis. However, the research on machine learning diagnosis and detection technology for engine misfire faults is still lacking. In addition, there are few studies on using the SENET model with high generalization and efficiency to analyze engine misfire faults. After a thorough review of relevant literature, it is clear that the SENET model also holds a pivotal position within the field of deep learning. Ma et al. introduced a capacitance tomography image reconstruction algorithm that harnesses the dual-path multi-scale feature fusion capabilities of the SENET model [18]. This algorithm tackles the intricate challenge of extracting complex and deep capacitance feature tensors from a solitary neural network. By adeptly capturing multi-scale detailed features and deep features following feature response redistribution, the algorithm exemplifies its proficiency in extracting bidirectional features. Huang et al. integrated the SENET attention mechanism into a three-dimensional Convolutional Neural Network (3D CNN) [19]. By modeling the interdependencies among feature map channels, they significantly enhanced the representational quality of the 3D CNN, empowering it to generate optimal traffic signal control actions. Chen et al. presented a SENET-optimized network model that demonstrates remarkable precision in segmenting landslide edge details, with a notable reduction in misrecognition and missed recognition instances [20]. In comparison to other models, this model exhibits superior recognition performance. It facilitates rapid screening of geological hazards in power line corridors, thereby mitigating the risk of landslides and other geological hazards in mountainous regions, and ensuring the vigilant monitoring and protection of power grid safety along these corridors. Dong et al. proposed a Dynamic Normalized Supervised Contrastive Network (DNSCN) that incorporates a multi-scale composite attention mechanism for identifying unbalanced gearbox faults [21]. DNSCN achieved impressive accuracies of 91.58% and 90.96% on two gearbox datasets characterized by extreme imbalance ratios, further validating the superiority of this innovative approach.

The majority of research on machine learning diagnosis and detection of engine misfire faults employs data-driven methodologies. A primary constraint in advancing such research stems from the scarcity of data. Indeed,acquiring extensive measured data from diverse fault samples presents considerable challenges. To mitigate this issue, researchers have devised effective strategies for data augmentation. Gao et al., for instance, utilize a combination of numerical simulation and generative adversarial networks to augment gear fault sample data [22]. Xiang et al. construct a finite element model, decompose the vibration signal into multiple components using wavelet packet transform (WPT), and compute specific time-domain characteristic parameters for all signal components to generate training samples [23]. To secure sufficient research sample data, this article adopts an experimental approach, wherein different engine misfire faults are artificially induced and the corresponding data is recorded. Compared to acquiring actual measurement data, this experimental methodology ensures the availability of a sufficient number of samples of various types, thereby circumventing the issue of imbalanced samples during the training of machine learning models. Furthermore, the sample data obtained through experiments is more comprehensive and controllable.

Given the pressing need for prompt and precise detection of engine misfires, this study primarily concentrates on the detection and identification of vehicle engine misfires, particularly in scenarios involving misfires in different cylinder bodies of multi-cylinder engines. Prior research efforts typically

leaned on traditional machine learning methods or more rudimentary neural network models to execute engine misfire fault detection tasks. While these approaches have yielded some results,they frequently rely heavily on manually crafted feature extraction processes, such as the extraction of specific parameters like vibration frequency and amplitude. This process is not only cumbersome and prone to subjective biases but also constrains the model's capacity to autonomously learn intricate data patterns. In response to this, this article introduces an innovative end-to-end deep learning solution: the SENET model. It demonstrates that the channel self-attention convolution model achieves higher detection accuracy in engine misfire faults compared to other baseline models. We have dispensed with the cumbersome feature engineering steps inherent in traditional methods and directly fed the originally collected vibration sequence data into a meticulously designed deep learning model. This model is adept at automatically extracting hierarchical and abstract feature representations from raw data, followed by in-depth analysis and learning. This process not only streamlines the workflow but also significantly bolsters the model's generalization ability and adaptability, empowering it to more accurately capture subtle changes during engine misfires and delivering definitive classification and recognition results. This end-to-end approach transcends the limitations of traditional methods, ushering in more efficient and intelligent solutions to the realm of engine misfire fault detection.

2 Channel Self Attention Mechanism

2.1 Attention Mechanism

The attention mechanism finds its origins in the study of human vision. In Cognitive Science, faced with the bottleneck of information processing, humans selectively concentrate on certain information while disregarding other visible data. To efficiently utilize limited visual processing resources, humans prioritize specific regions of their visual field.

The attention mechanism lacks a precise mathematical definition, and traditional techniques such as local image feature extraction and sliding window methods can be seen as forms of attention. In neural networks, the attention mechanism often takes the form of an auxiliary neural network that either rigidly selects segments of the input or assigns varying weights to different parts of the input. This mechanism effectively filters crucial information from vast datasets.

Multiple approaches exist for integrating the attention mechanism into neural networks. Considering convolutional neural networks as an example, the attention mechanism can be incorporated into the spatial dimension, and specifically, the Squeeze-and-Excitation (SE) mechanism can be added in the channel dimension. Additionally, there are hybrid dimensions, such as Convolutional Block Attention Module (CBAM), which combines both spatial and channel dimensions to introduce attention. This paper primarily focuses on the mechanism of enhancing attention in the channel dimension.

2.2 SENET Attention Mechanism

Prior to 2017, researchers primarily concentrated on enhancing model performance in the spatial domain. However, in 2017, the team headed by Hu et al. from Momenta Autonomous Driving Company introduced SENET, which is grounded in the channel attention mechanism [24]. The core novelty of this research was its emphasis on the interplay between channels and the establishment of a channel attention mechanism, aiming to automatically ascertain the significance of each channel feature via the model [25]. To accomplish this, SENET formulated the Squeeze and Excitation (SE) module. This module captures global features at the channel level through squeezing operations, learns the interdependencies among channels via excitation operations, computes the weights for

each channel, and ultimately multiplies these weights with the original feature map to derive the final attention-enhanced features [26]. Essentially, the SE module applies gating and attention mechanisms along the channel dimension, allowing the model to prioritize channels with substantial information while dampening less critical ones [27].

SE attention mechanisms, or Squeeze and Excitation Networks, integrate attention mechanisms into the channel dimension. The pivotal operations within these mechanisms are Squeeze and Excitation. Through automatic learning, a novel neural network determines the importance level of each channel within the feature map and subsequently assigns a weight value to each feature based on this importance. This enables the neural network to focus on specific feature channels, bolstering those feature maps that are beneficial to the current task while suppressing those that are not.

The SE module mainly consists of two operations:squeezing and excitation, which can be applied to any mapping. For example, in convolution,the convolution kernel is $\mathrm{V} = [v_1, \mathrm{v}_2, ..., \mathrm{v}_c]$, where V_c represents the c-th convolution kernel, and the output $U = [u_1, u_2, \ldots, u_c]$:

$$\mathrm{u}_c = \mathrm{v}_c * \mathrm{X} = \sum_{s=1}^{C'} \mathrm{v}_c^s * \mathrm{x}^s \tag{1}$$

Among them, $*$ denotes the convolution operation, s represents the channel of the two-dimensional convolution kernel, and the spatial feature x^s on the input channel will automatically learn the relationships within the feature space. However, since the convolution results of each channel have been summed, the channel feature relationship is intermingled with the spatial relationship learned by the convolution kernel. The purpose of the SE module is to isolate and extract channel features from these intertwined relationships, thereby enabling the model to directly learn the inter-channel feature relationships [28].

As illustrated in Fig. 1 below, the process begins with the Squeeze part, which utilizes a global pooling layer to compress the two-dimensional features of the Convolutional Neural Network (CNN) into a single real number per feature channel, maintaining the number of feature channels unchanged. Following this compression, the width and height of each feature are reduced to 1×1. Subsequently, the excitation part is employed to generate corresponding weight values for the multiple feature channels. This is accomplished through a combined network structure comprising two fully connected layers, followed by the RELU activation function and the sigmoid activation function. Notably, the input and output features have the same number of weight values. Finally, the Scale part multiplies the generated normalized weights by the features of each channel, effectively establishing the SENET attention mechanism for channel dimensions [19].

3 Engine Misfire Fault Experiment and Data Preprocessing

3.1 Engine Misfire Fault Experiment

The testing system for this experiment consists of a Camry Sport vehicle and ECON acceleration signal acquisition equipment, encompassing an ECON vibration analyzer and a PC platform, as depicted in Fig. 2. The experimental setting is situated within the Training Room of the School of Automobile and Transportation at Guangdong University of Technology. The subject of the test is the engine of the Camry Sport vehicle.

To gather data on various engine misfire faults, we deliberately induced different types of misfires. The experimental plan is outlined in Table 1. During the experiment,we conducted 10 fault experiments, including 4 types of single-cylinder misfires (each involving a separate cylinder), 6 types of dual-cylinder misfires (where 2 out of the 4 cylinders misfire simultaneously), and a normal state,

totaling 11 types of conditions. A highly sensitive 3D acceleration sensor with a high sampling rate was attached to the engine cylinder to capture vibration data during the experiment. This sensor, part of the ECON vibration signal acquisition instrument, has a sampling rate of up to 20,840 Hz.

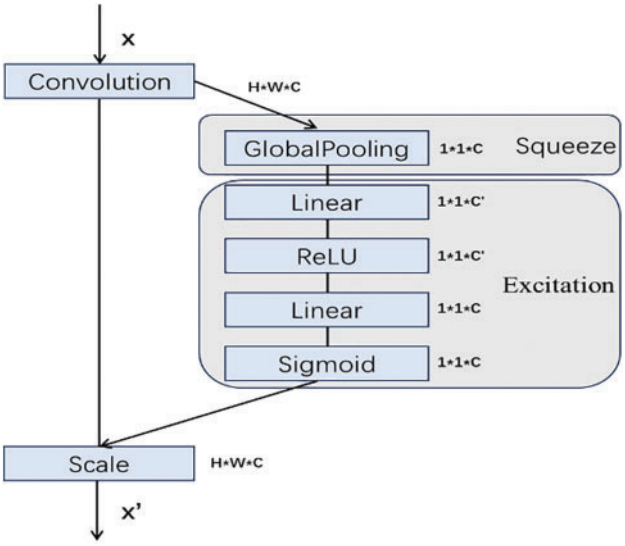

Figure 1: SENET attention mechanism

Figure 2: Engine misfire data collection site

Table 1: Engine misfire test scheme

No.	Status	Engine speed	Sampling duration	Remark
1	Normal 1st cylinder misfire 2nd cylinder misfire 3rd cylinder misfire	1. 1000 rpm 2. 1500 rpm 3. 2000 rpm 4. 2500 rpm	3 times each, approximately 25 s each time	Sampling rate: 20,840 Hz

(Continued)

Table 1 (continued)

No.	Status	Engine speed	Sampling duration	Remark
	4th cylinder misfire	5. Slow and uniform acceleration from idle to 3000 rpm		
2	1st and 2nd cylinders misfire 1st and 3rd cylinders misfire 1st and 4th cylinders misfire	1. 1000 rpm		
3	2nd and 3rd cylinders misfire 2nd and 4th cylinders misfire	2. 1500 rpm		
4	3rd and 4th cylinders misfire	3. 2000 rpm		

After setting a specific state, the experimenter started the engine and adjusted the fuel pedal to reach the designated value, maintaining it for 25 s. The vibration signal was then collected using the ECON vibration signal acquisition instrument. To ensure comprehensive data collection, we set five different speed scenarios for each experimental situation: 1000, 1500, 2000, 2500 rpm, and from idle to 3000 rpm. For each speed scenario under each misfire fault, we conducted three tests,each lasting approximately 25 s.

Throughout the entire experiment, the position and orientation of the sensor remained constant and unaltered. This configuration allowed us to consistently obtain two sets of vibration feature sequences (each comprising three XYZ directions) in each experiment,amounting to six independent feature sequences in total. Fig. 3 illustrates the acceleration data of one sensor in the XYZ directions under the condition of a 3-cylinder misfire at 1500 rpm. It is evident that the range of acceleration for engine vibration in all three directions is approximately the same, primarily falling between −40 and 40 m/s^2.

3.2 Data Preprocessing

3.2.1 Data Standardization

Prior to identifying engine misfire fault signal data, we undertake the following preprocessing steps: data standardization and the partitioning of training and testing sets.

Standardization is a crucial data preprocessing technique aimed at transforming input data into a specific distribution to better accommodate the training of deep learning models. In the realm of deep learning, standardization typically involves converting data into a distribution with a mean of 0 and a standard deviation of 1, a process also known as z-score normalization.

Specifically, for a given numerical vector $X = (x_1, x_2, \ldots, x_n)$, where x_i represents the value of the i-th feature, the normalization calculation method is as follows: for each feature x_i, Min-Max normalization (also known as dispersion normalization) is performed, which scales the value range of

feature x_i to the range of [0, 1]. The calculation formula is:

$$\hat{x}_i = \frac{x_i - \min(X)}{\max(X) - \min(X)} \tag{2}$$

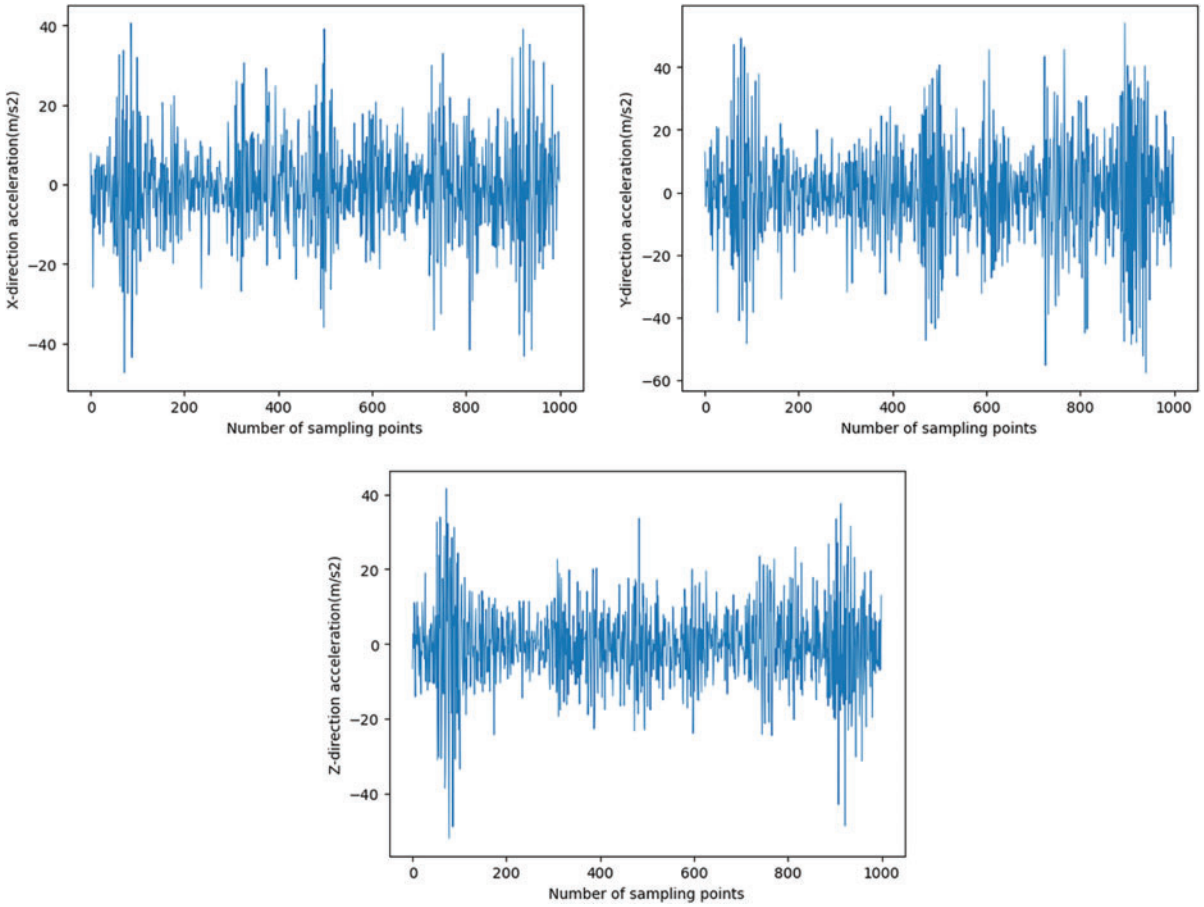

Figure 3: Acceleration data in three directions at a speed of 1500 rpm when the 3rd-cylinder misfires

Among them, $\hat{x}_i$ represents the i-th feature value after Min-Max normalization, and $\min(X)$ and $\max(X)$ represent the minimum and maximum values in vector X.

Perform z-score normalization for each feature $\hat{x}_i$, which converts feature $\hat{x}_i$ to a distribution with a mean of 0 and a standard deviation of 1. The calculation formula is:

$$\hat{x}_i = \frac{\hat{x}_i - \mu}{\sigma} \tag{3}$$

Among them, $\hat{x}_i$ represents the i-th feature value after standardization, while μ and σ represent the mean and standard deviation of vector $\hat{X} = (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n)$, respectively.

By standardizing the input data for deep learning, we can align its distribution more closely with the assumptions of the deep learning model. This enhances the model's generalization ability and robustness, which in turn facilitates faster training speeds and higher accuracy. We apply data

standardization to all components of the signal data, specifically the three data channels comprising vibration acceleration in the XYZ directions.

3.2.2 Divide Training and Testing Sets

We utilized the set-aside method to partition the dataset into training and testing sets. According to the experimental scheme, there are 11 states, each associated with a distinct speed. Therefore, for each state and each speed, one experimental data point was chosen for the testing set, while the remaining two were allocated to the training set. Consequently, the ratio of the training set to the testing set is 2:1, with the training set comprising 2/3 of the data and the testing set making up the remaining 1/3. This partitioning ensures no overlap between the training and testing sets, thereby allowing for a more accurate assessment of the model's generalization ability, preventing overfitting, and facilitating the discovery of optimal adjustment parameters.

4 Comparison between Methods

4.1 The Method of LSTM

4.1.1 Introduction of Model

RNN models enable more effective processing of sequence data. In these models,the output of neurons at a particular time can be reconsidered as input, allowing the RNN's network structure to fully capture dependencies in time series data. However, traditional RNN models are plagued by issues such as gradient vanishing and gradient explosion. To overcome these challenges, Hochreiter et al. [29] introduced the LSTM (Long Short-Term Memory) network, which represents a significant improvement over traditional RNN. Compared to RNN, LSTM models feature more complex hidden units. Furthermore, LSTM have a broader range of applications and are more effective as sequence models. During operation, LSTM can selectively add or delete information through the use of linear interventions. The structural diagram of an LSTM is illustrated in Fig. 4. In this article, we utilized a 2-layer LSTM with 128 cells per layer.

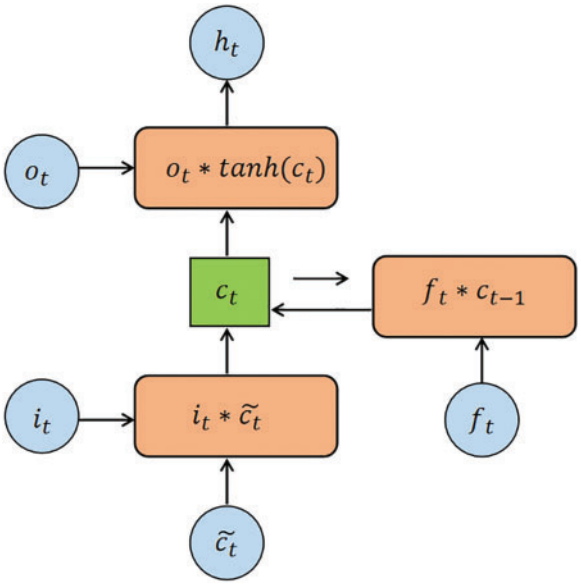

Figure 4: Calculation process of LSTM model

4.1.2 The Analysis Results of Model

The confusion matrix depicting the LSTM model's prediction results on the test set, with an input sequence length of 1000, is illustrated in Fig. 5. It is evident that the detection and recognition rate for single-cylinder misfire in the fourth cylinder is the lowest, at 0.75, while the accuracy rates for other single-cylinder misfires and double-cylinder misfires are below 90%. The overall accuracy rate for all cases stands at 0.84.

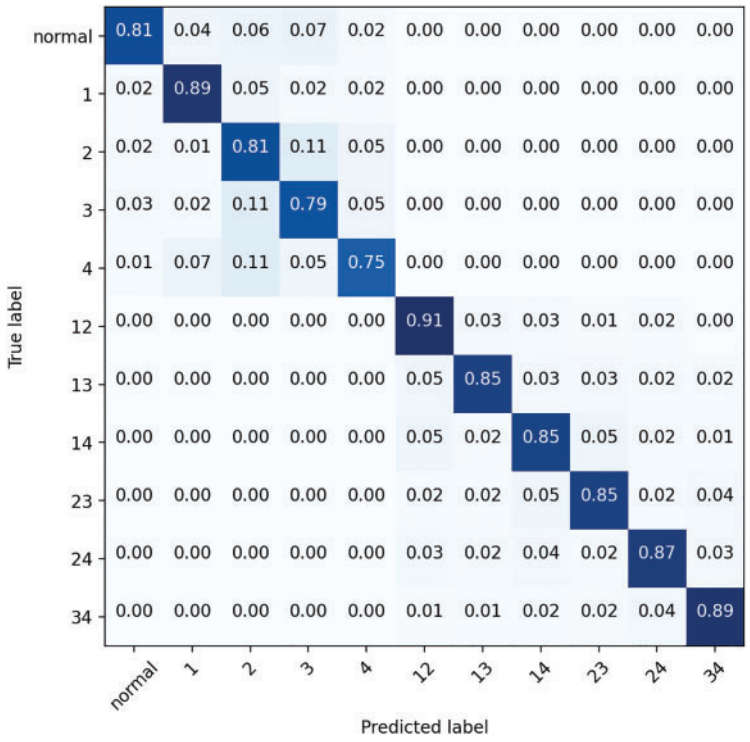

Figure 5: LSTM model predicts the confusion matrix in the test set, where "i" refers to the single cylinder misfire with number i, and "ij" refers to the simultaneous misfire of two cylinders with numbers i and j

Upon analyzing the LSTM model's prediction results, we noted that the recognition accuracy for single-cylinder misfires is relatively low. This is primarily attributed to the frequent misclassification among the four distinct types of single-cylinder misfires and the normal condition. Specifically, 11% of samples misidentified as the second cylinder misfiring were actually the third cylinder misfiring, and vice versa, with an equivalent proportion (11%) of samples experiencing this reciprocal misidentification. Additionally, there were instances of mutual misrecognition among samples with simultaneous misfiring in two cylinders, underscoring the model's challenges in distinguishing between different misfire patterns.

These results highlight significant differences in vibration characteristics between single-cylinder and double-cylinder misfires, which theoretically should facilitate more accurate classifications by the model. However, the current model performance indicates that there is still room for improvement in capturing and distinguishing these subtle yet crucial feature differences.

Furthermore, we computed the F1 scores for 11 distinct categories, as presented in Table 2. The average F1 score for the four types of single-cylinder misfires is 0.79, whereas the average F1 score for double-cylinder misfires is 0.87.

Table 2: Comparison of F1 scores of various models under 1000 time step input

Status				
Models	LSTM	MSRESNET	Transformer	SENET
Normal	0.86	0.96	0.87	0.99
1st cylinder misfire	0.87	0.96	0.73	0.99
2nd cylinder misfire	0.75	0.90	0.64	0.99
3rd cylinder misfire	0.76	0.93	0.70	0.99
4th cylinder misfire	0.79	0.94	0.71	1.00
1st and 2nd cylinders misfire	0.87	0.97	0.87	0.99
1st and 3rd cylinders misfire	0.88	0.97	0.85	1.00
1st and 4th cylinders misfire	0.84	0.96	0.83	0.98
2nd and 3rd cylinders misfire	0.86	0.96	0.81	0.98
2nd and 4th cylinders misfire	0.88	0.97	0.86	0.98
3rd and 4th cylinders misfire	0.89	0.96	0.85	0.98

4.2 The Method of Transformer

4.2.1 Introduction of Model

In 2017, Google introduced the Transformer model in its seminal paper "Attention is All You Need," replacing the conventional RNN network structure in NLP tasks with a Self-Attention mechanism. The primary advantage of the Transformer over RNN architectures is its capability for parallel computing. This deep learning model, rooted in the self-attention mechanism, boasts higher training and inference speeds, as well as a flexible architecture, compared to RNN and LSTM models.

In this paper, we leverage the Encoder component of the Transformer model to extract features from sequence data. Following this, we append a fully connected layer to serve as the final classification output.

Regarding the hyperparameter configuration of our Transformer model, we have implemented the following settings: We established an embedding dimension of 128, enabling the model to employ 128-dimensional vectors for embedding input feature data during both input and output processing. Additionally, we incorporated eight attention heads, harnessing the multi-head attention mechanism to capture diverse and pertinent information in parallel from the input sequence. To further enhance the model's comprehension and representation of input data, we set the number of encoder layers to four, thereby increasing the model's depth.

4.2.2 The Analysis Results of Model

The prediction outcomes for the Transformer model are illustrated in Fig. 6, which depicts an input sequence length of 1000. The results indicate a relatively low prediction accuracy for both single-cylinder and dual-cylinder misfires. Specifically, the highest accuracy achieved is 0.88 in normal

conditions, whereas the lowest accuracy observed is 0.60 during a single-cylinder misfire involving the second cylinder. Overall, the Transformer model attains an accuracy of 0.79.

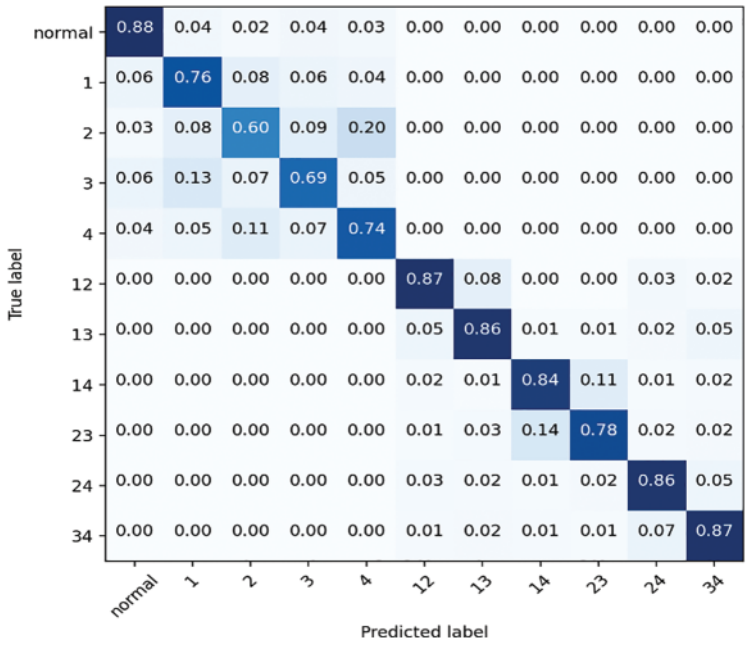

Figure 6: The confusion matrix of Transformer model, where "i" refers to single cylinder misfire with number i, and "ij" refers to simultaneous misfire of two cylinders with numbers i and j

Analogous to the prediction results of the LSTM model, the Transformer model also experiences notable misclassifications in single-cylinder and double-cylinder misfire scenarios, with some normal conditions mistakenly identified as single-cylinder misfires. When compared to double-cylinder misfires, the accuracy for detecting single-cylinder misfires is lower, with an average accuracy of 69.7% for single-cylinder misfires *vs*. 84.7% for double-cylinder misfire recognition.

Table 2 further reveals that, in the prediction results of the Transformer model, the average F1 score for single-cylinder misfires is 0.695, whereas the average F1 score for double-cylinder misfires is 0.845.

4.3 The Method of MSRESNET

4.3.1 Introduction of Model

In the realm of deep learning, enhancing the depth of a network structure can potentially yield superior fitting performance for a model. However, incessantly augmenting the number of network layers not only diminishes the model's generalization capability to unfamiliar data but also introduces challenges such as gradient vanishing or model degradation. As the network expands to a certain number of layers, the overall model accuracy plateaus; further increases in depth lead to the accumulation of training errors, impeding improvements in model accuracy [30]. To address these issues, He et al. [31] introduced the deep residual network, which not only transcends the limitation of layer count in neural networks but also effectively taps into the deep feature information of data by overlaying shallow and deep features, thereby facilitating network convergence.

Traditional CNN fall short in fully harnessing the multi-scale information embedded in facial images [32,33], as each layer is confined to extracting feature information of a single scale. To acquire a richer feature set, the conventional approach involves deepening the network layers, which is prone to overfitting and necessitates substantial computational resources,thereby complicating network training optimization. To mitigate these challenges, multi-scale residual networks (MSRESNET) are employed to streamline network training and optimization [34].

In the MSRESNET model, we initiate the feature extraction process using a convolutional block equipped with 64 filters and a kernel size of 3 (comprising a convolutional layer, a BatchNorm layer, and a RELU layer). Subsequently, the feature maps are distributed into three convolutional blocks of varying scales, featuring kernel sizes of 3, 5, and 7 (with output channel numbers of 64 and 128). These blocks undergo adaptive average pooling and are concatenated. Ultimately, the concatenated features are fed into a fully connected layer for classification and recognition tasks.

4.3.2 The Analysis Results of Model

The prediction outcomes of the MSRESNET model, depicted in Fig. 7 (with an input sequence length of 1000), reveal impressive results. The model attains detection accuracies exceeding 0.87 for the normal state, single-cylinder misfire, and double-cylinder misfire scenarios. Notably, even in the fault state involving the simultaneous misfire of the second and third cylinders, the recognition rate remains as high as 0.98. The lowest accuracy, observed in the case of a "2" cylinder single-cylinder misfire, is still commendable at 0.87. Overall, the MSRESNET model boasts an impressive accuracy of 0.95.

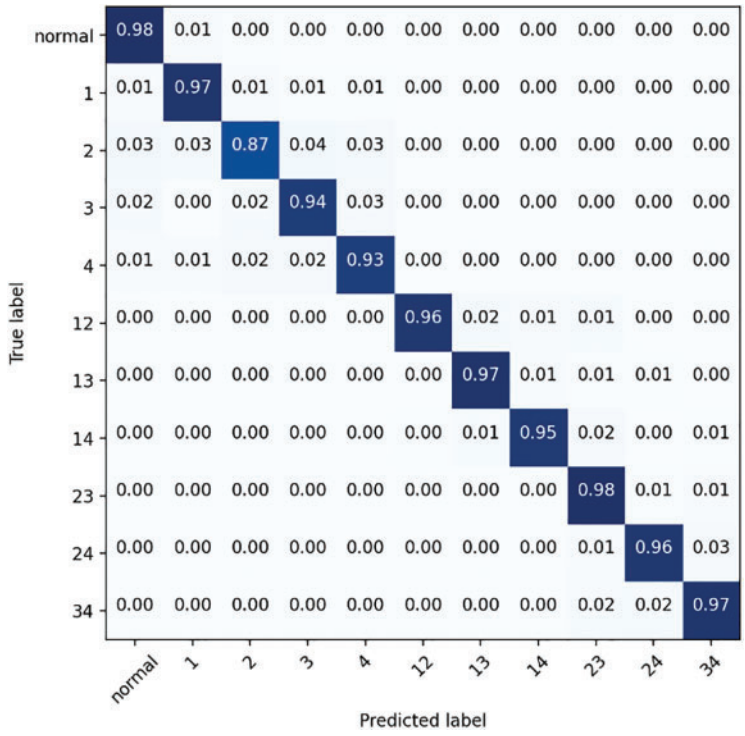

Figure 7: The confusion matrix of MSRESNET model, where "i" refers to the single cylinder misfire with number i, and "ij" refers to the simultaneous misfire of two cylinders with numbers i and j

When compared to LSTM and Transformer models, the MSRESNET model demonstrates a marked improvement in recognition accuracy. However, a discernible difference persists in the distribution of misclassified samples between single-cylinder and double-cylinder misfires.

Furthermore, the MSRESNET model's prediction results showcase average F1 scores of 0.933 for single-cylinder misfires and 0.965 for double-cylinder misfires, respectively, underscoring its robust performance in both scenarios.

4.4 The Method of SENET

The model structure of the SENet utilized in this article is depicted in Fig. 8a. Initially, the model employs a convolutional layer to extract features from the input sequence. Subsequently, it progressively delves deeper into these features through four SE blocks. Ultimately, a fully connected layer serves as the model's final classifier. The configuration of the SE block is illustrated in Fig. 8b. Here, the input feature x traverses two convolutional layers before the channel weights are derived using the squeeze module and extraction model. These weights are then scaled with the convolutional output feature. Lastly, the residual is summed with the downsampled feature map of x' to yield the SE block's output.

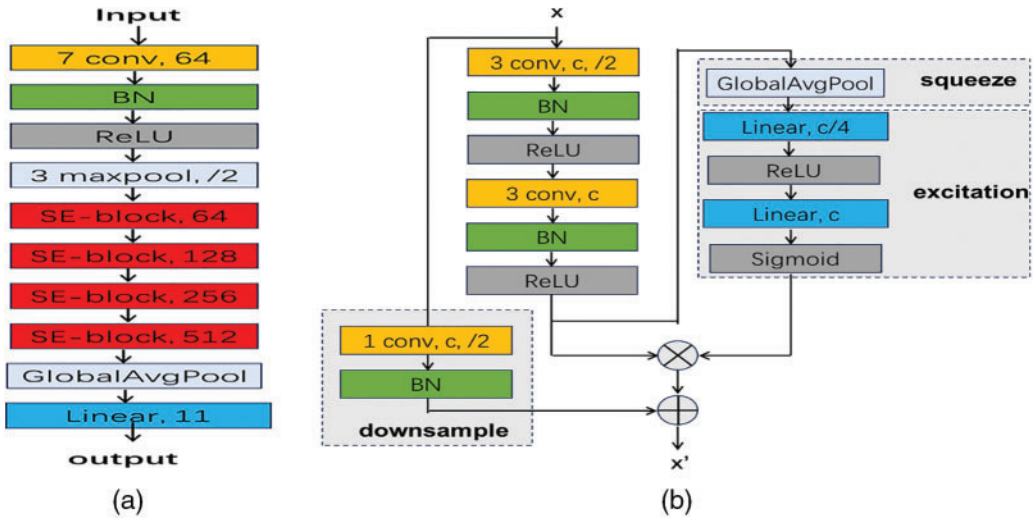

Figure 8: (a) The structure diagram of SENET model, (b) The structure of SE block

When the input sequence spans 1000 units, the model's prediction accuracy is presented in Fig. 9. Across all experimental scenarios, the model's recognition rate surpasses 0.98. The overall accuracy of the model, encompassing all experimental contexts, stands at 0.99.

Remarkably, the SENET model maintains a prediction accuracy exceeding 98% in all instances, thereby precluding significant misidentifications within the individual categories of single-cylinder and double-cylinder misfires. The results further indicate that the F1 scores for all misfire types predicted by the SENET model also exceed 0.98.

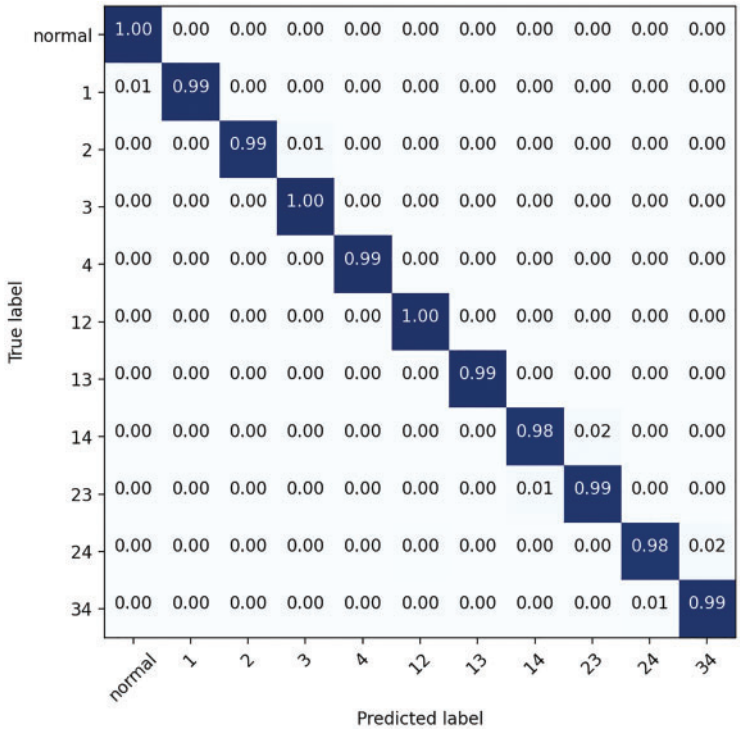

Figure 9: The confusion matrix of SENET model, where "i" refers to single cylinder misfire with number i, and "ij" refers to simultaneous misfire of two cylinders with numbers i and j

5 Discussion

In order to conduct a thorough comparison of the performance capabilities of the SENET model *vs*. the three baseline models, namely LSTM, MSRESNET, and Transformer, we analyzed input sequences of differing lengths. The results of this comparison in terms of accuracy are depicted in Fig. 10. It is evident that, regardless of the length of the input sequences, the SENET model consistently demonstrates superior prediction accuracy compared to the three baseline models. Specifically, for input sequences of 300 in length, the SENET model achieves a fault type recognition accuracy of 0.9, closely followed by MSRESNET with 0.846, while LSTM and Transformer exhibit lower accuracies of 0.674 and 0.680, respectively. As the length of the input sequences increases, the fault type recognition accuracy of both the SENET model and the baseline models experiences gradual improvement. Notably, when the input sequence length reaches 2000, the accuracy of MSRESNET approximates that of SENET, while a notable discrepancy persists between LSTM and Transformer.

Furthermore, we assessed the F1 scores of each model across 11 distinct scenarios with an input length of 1000, as presented in Table 2. In this assessment, we employed precision, recall, and F1 score as the evaluative metrics, which were computed using established formulas.

$$precision = \frac{TP}{TP + FP} \tag{4}$$

$$recall = \frac{TP}{TP + FN} \tag{5}$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \tag{6}$$

where *TP* (True Positives) represents the count of samples that the model correctly predicts as belonging to the positive class, aligning with the actual ground truth. Conversely, *FP* (False Positives) denotes the count of samples that the model erroneously predicts as positive, despite being negative in the actual ground truth. Additionally, *FN* (False Negatives) signifies the count of samples that the model incorrectly predicts as negative, whereas they are actually positive in the ground truth.

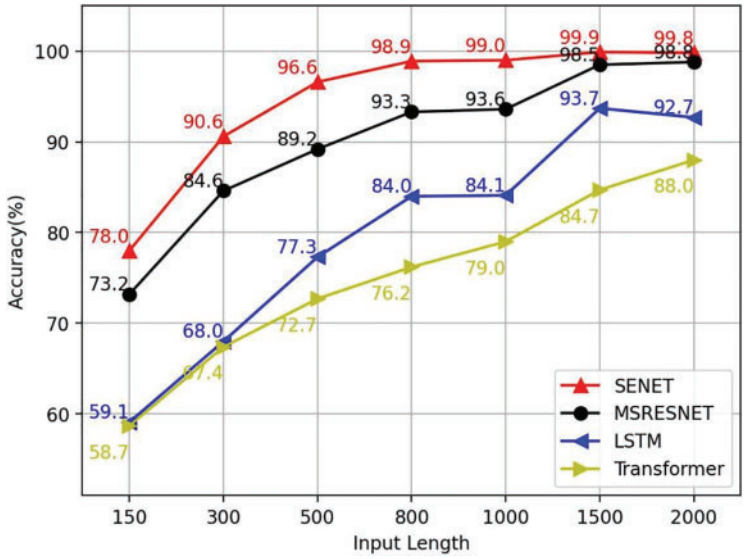

Figure 10: Accuracy of each model under different length input sequences

Our analysis reveals that across all eleven scenarios considered, the F1 score of the SENET model surpasses that of the other models, achieving a remarkable value exceeding 0.98. This indicates the SENET model's superior discriminative ability and robustness in fault-type recognition tasks.

In our prior analysis, we observed a noteworthy phenomenon: the LSTM and Transformer models exhibited relatively low predictive accuracy, particularly in the context of mutual misclassification between the fault categories of single-cylinder misfire and double-cylinder misfire. This finding implies that, while these two models can effectively differentiate between the broader categories of single-cylinder and double-cylinder misfires, they struggle to accurately identify specific instances within each category. Conversely, the SENET model demonstrated exceptional recognition performance, consistently maintaining an accuracy rate above 98% without manifesting the aforementioned misclassification issues, thereby highlighting its robust classification capabilities.

To further investigate the underlying differences in feature representation that account for this phenomenon, we innovatively applied the T-Distributed Stochastic Neighbor Embedding (T-SNE) technique for visually analyzing the key features extracted by the LSTM and SENET models. Specifically, we extracted input features from the classification layers (i.e., the final fully connected layers) of these two models and subsequently utilized the T-SNE algorithm to effectively reduce the high-dimensional feature space to two dimensions. Finally, we presented the insights gained from this dimensionality reduction through graphical visualizations (as depicted in Fig. 11).

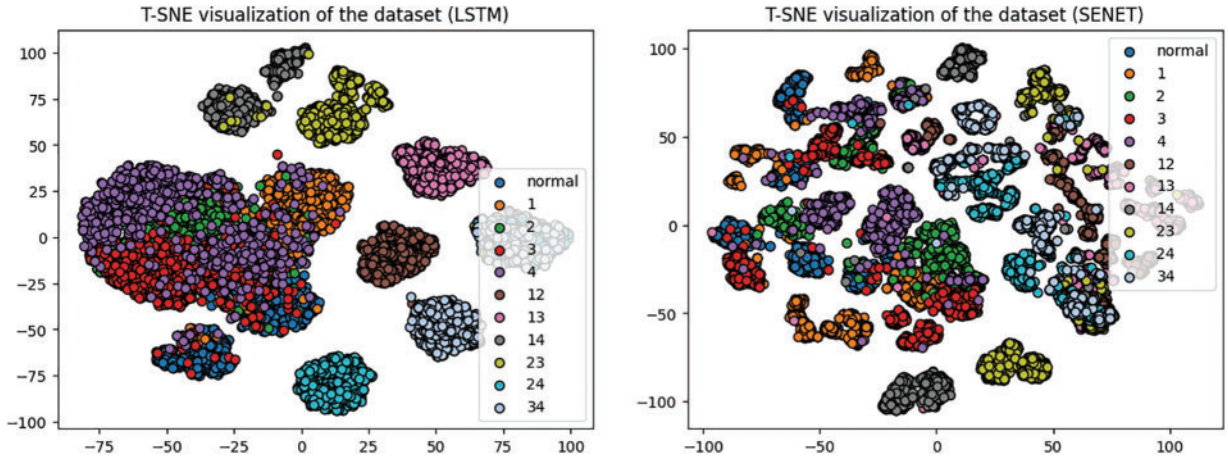

Figure 11: LSTM model and SENET model extract feature T-SNE visualization graph

The T-SNE visualization results for the LSTM model revealed significant overlap in the distribution of data points between normal operating conditions and the four single-cylinder misfire states. This phenomenon directly explains the limited recognition ability of the LSTM model in distinguishing between these five sample types. Conversely, the data points representing double-cylinder misfires exhibited a more independent distribution range, indicating a higher degree of inter-class separability.

In contrast, the T-SNE plot of the SENET model presents a starkly different scenario: the data points of various types are more widely and evenly dispersed, suggesting a richer and more discriminative feature representation. Notably, unlike the LSTM model, the data points of the same type in the SENET model do not solely cluster into a single group but may form multiple relatively independent sub-clusters. This characteristic potentially enables the model to maintain greater flexibility and accuracy when dealing with complex and varied fault patterns. In summary, the SENET model, with its superior feature extraction and representation capabilities, has demonstrated immense potential and advantages in the realm of engine fault diagnosis.

6 Conclusion

This article proposes the application of the SENET model for identifying various types of car engine misfire faults. We artificially induce four types of single-cylinder misfires and six types of dual-cylinder misfires in automotive engines, combined with normal operational conditions, yielding a comprehensive dataset encompassing 11 distinct scenarios. During data acquisition, a high-precision acceleration signal collector with a high sampling rate is employed to obtain vibration acceleration information from the engine in three orthogonal directions across multiple experimental trials. The SENET model, upon analysis, attains a fault type recognition accuracy of 0.99 when the input sequence length is set to 1000 (approximately 50 ms in duration). When compared to baseline models such as LSTM, MSRESNET, and Transformer, the SENET model demonstrates the highest accuracy in fault type identification across varying input sequence lengths. The analytical results further indicate that the SENET model's F1 score surpasses those of the three baseline models. Across all 11 scenarios, the SENET model achieves an F1 score exceeding 0.98. Notably, the LSTM and Transformer models exhibit a significant issue of mutual misclassification between single-cylinder and dual-cylinder misfire fault categories, whereas the SENET model does not encounter this challenge. Visualization of features

extracted by the LSTM and SENET models using the T-SNE technique reveals that in the LSTM model, data points of the same type tend to cluster together, with notable overlap between data points representing normal operational conditions and single-cylinder misfire scenarios. Conversely, the data points of various types in the SENET model are more widely and uniformly dispersed.

This article primarily leverages deep learning technology to develop an efficient and reliable automatic detection system, addressing the limitations of traditional detection methods, which often suffer from insufficient accuracy and slow response speeds. By conducting a thorough analysis of engine misfire vibration characteristics, we employ an end-to-end approach utilizing the SENET model. This approach involves acquiring vibration characteristics of the engine through acceleration signal acquisition devices and utilizing these characteristics to train a deep learning model, thereby achieving accurate classification and recognition of engine misfires.

The objective of this article is to collect vibration sequence data under various misfire scenarios by installing vibration sensors in specific engine components. This data is then utilized to achieve precise identification of misfire fault types using deep learning models. However, the successful application of this technology in practical engineering fields necessitates further exploration of several key issues. Primarily, there is a need to identify optimal sensor placement points that are both easy to install repeatedly and ensure relatively stable vibration characteristics. Additionally, while this article covers several types of engine misfires, the research scope must be expanded to encompass more potential misfire scenarios and other types of engine failures, to comprehensively enhance the diagnostic system's generalization ability. Furthermore, future research will extend to additional Camry models and other vehicle brands, deeply analyzing the changes in vibration characteristics induced by engine misfires, thereby laying a solid foundation for the widespread application of this technology.

Acknowledgement: I express my sincere gratitude to all authors who have contributed to this paper. Their dedication and insights have been invaluable in shaping the outcome of this work.

Funding Statement: Yongxian Huang supported by Projects of Guangzhou Science and Technology Plan (2023A04J0409). The website where this project is in https://gzsti.gzsi.gov.cn/pms/homepage.html, accessed on 01 September 2024.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Feifei Yu, Canyi Du; data collection: Guoyan Chen, Yongkang Gong; analysis and interpretation of results: Yongxian Huang; draft manuscript preparation: Guoyan Chen, Xiaoqing Yang. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

[1] X. Zhao, "Engine misfire fault diagnosis based on model identification," M.S. thesis, Jilin Univ., Changchun, China, 2014.

[2] X. Wang, U. Kruger, G. W. Irwin, G. McCullough, and N. McDowell, "Nonlinear PCA with the local approach for diesel engine fault detection and diagnosis," *IEEE Trans. Control Syst. Technol.*, vol. 16, no. 1, pp. 122–129, 2008. doi: 10.1109/TCST.2007.899744.
[3] D. Zhou and Y. Hu, "Fault diagnosis technology for dynamic systems," *J. Autom.*, vol. 35, no. 6, pp. 748–758, 2009.
[4] R. Natarajan, P. Santosh Reddy, S. C. Bose, H. L. Gururaj, F. Flammini and S. Velmurugan, "Fault detection and state estimation in robotic automatic control using machine learning," *Array*, vol. 19, 2023, Art. no. 100298. doi: 10.1016/j.array.2023.100298.
[5] D. Wang, "Research on diesel engine misfire fault diagnosis method based on wavelet and neural network," M.S. thesis, Huazhong Agric. Univ., Wuhan, China, 2011.
[6] J. Wang, "Application of OFBP neural network in engine fault diagnosis," M.S. thesis, Harbin Univ. Sci. Technol., Harbin, China, 2013.
[7] J. Zheng and Z. Yang, "Fault identification method for vehicle engines based on FTA SVM," (in Chinese), *Ind. Saf. Environ. Prot.*, vol. 6, pp. 33–35+51, 2015.
[8] P. Wang, Y. Shi, J. Liu, and L. Cheng, "Research on online diagnosis of engine composite faults based on POS-RELM," (in Chinese), *Inf. Technol.*, no. 5, pp. 112–114, 2016.
[9] Z. Wang, "Fault diagnosis of engine misfire based on probabilistic neural network," M.S. thesis, Jilin Univ., Changchun, China, 2016.
[10] Y. Gao and Y. Li, "Misfire identification of automobile engines based on wavelet packet and extreme learning machine," *J. Meas. Sci. Instrum.*, vol. 8, no. 4, pp. 384–395, 2017.
[11] J. Han, J. Jia, J. Mei, G. Ren, and X. Jia, "Misfire fault diagnosis based on optimal wavelet packet and PSO-SVM," *Mech. Des. Res.*, vol. 2, pp. 137–141, 2019.
[12] K. Chen, "Research on diesel engine anomaly detection and fault diagnosis technology based on Autoencoder depth feature extraction," M.S. thesis, Beijing Univ. Chem. Technol., Beijing, China, 2020.
[13] W. Gao, Y. Wang, X. Wang, P. Zhang, Y. Li and Y. Dong, "Real time diagnosis of diesel engine misfire fault based on convolutional neural network," (in Chinese), *J. Jilin Univ. (Eng. Technol. Ed.)*, vol. 2, pp. 417–424, 2022.
[14] C. Du, K. Ding, Z. Yang, and C. Yang, "Diagnosis for engine misfire fault based on torsional vibration and neural-network analysis," *Adv. Mater. Res.*, vol. 1566, pp. 433–440, 2012. doi: 10.4028/www.scientific.net/AMR.433-440.7240.
[15] X. Wang, J. Wei, and J. Ye, "Misfire fault diagnosis of gasoline engines using the cosine measure of single-valued neutrosophic sets," *J. New Theory*, no. 10, pp. 39–44, 2016.
[16] J. Suda and D. Kagaris, "Automated diagnosis of engine misfire faults using combination classifiers," *SAE Int. J. Commer. Veh.*, vol. 13, no. 2, pp. 103–113, 2020. doi: 10.4271/02-13-02-0007.
[17] C. Du, W. Li, F. Yu, F. Li, and X. Zeng, "Misfire fault diagnosis of automobile engine based on time domain vibration signal and probabilistic neural network," *Int. J. Perform. Eng.*, vol. 16, no. 9, pp. 1488–1496, 2020.
[18] M. Ma and J. Li, "ECT image reconstruction based on SENet dual path multi-scale feature fusion," *Vib. Shock*, vol. 42, no. 7, pp. 180–186, 2023.
[19] S. Huang, M. Wang, and C. Yang, "Traffic signal control technology based on attention mechanism," *Inf. Technol. Informatization*, vol. 3, pp. 276–296, 2023.
[20] Z. Chen, S. Lu, Z. Qin, and Q. Zhang, "Deeplabv3 + landslide identification optimized by SENet," (in Chinese), *Sci. Technol. Eng.*, vol. 22, no. 33, pp. 14635–14643, 2022.
[21] Y. Dong, H. Jiang, W. Jiang, and L. Xie, "Dynamic normalization supervised contrastive network with multiscale compound attention mechanism for gearbox imbalanced fault diagnosis," *Eng. Appl. Artif. Intell.*, vol. 133, no. 4, 2024, Art. no. 108098. doi: 10.1016/j.engappai.2024.108098.
[22] Y. Gao, X. Liu, and J. Xiang, "Fault detection in gears using fault samples enlarged by a combination of numerical simulation and a generative adversarial network," *IEEE/ASME Trans. Mechatron.*, vol. 27, no. 5, pp. 3798–3805, Oct. 2022. doi: 10.1109/TMECH.2021.3132459.

[23] J. Xiang, Y. Zhong, and A. M. C. Vasques, "A novel personalized diagnosis methodology using numerical simulation and an intelligent method to detect faults in a shaft," *Appl. Sci.*, vol. 6, no. 12, 2016, Art. no. 414.
[24] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.
[25] Y. Han, C. Wei, R. Zhou, and Z. Hong, "Combining 3D-CNN and squeeze-and-excitation networks for remote sensing sea ice image classification," *Math. Probl. Eng.*, vol. 2020, no. 1, pp. 1–15, 2020. doi: 10.1155/2020/8065396.
[26] X. Zhang, G. Ding, J. Li, W. Wang, and Q. Wu, "Deep learning empowered MAC protocol identification with squeeze-and-excitation networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 8, no. 2, pp. 683–693, Jun. 2022.
[27] H. Ma, G. Han, L. Peng, L. Zhu, and J. Shu, "Rock thin sections identification based on improved squeeze-and-Excitation Networks model," *Comput. Geosci.*, vol. 152, no. 7, 2021, Art. no. 104780. doi: 10.1016/j.cageo.2021.104780.
[28] S. Chen, "Research on zero sample learning method based on clustering guidance and semantic extension," M.S. thesis, Beijing Univ. Technol., Beijing, China, 2020.
[29] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735.
[30] J. Liu, F. Chen, and H. Xiong, "Open electrical impedance imaging algorithm for multi-scale residual network models," (in Chinese), *J. Zhejiang Univ. (Eng. Sci.)*, vol. 56, no. 9, pp. 1789–1795, 2022.
[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.
[32] Y. Zhou and J. Zhang, "Product image retrieval based on multiscale deep learning," *Comput. Res. Dev.*, vol. 54, no. 8, pp. 1824–1832, 2017.
[33] C. Zhu, Y. Zheng, K. Luu, and M. Savvides, "CMS-RCNN: Contextual multi-scale region-based CNN for unconstrained face detection," 2016. doi: 10.48550/arXiv.1606.05413.
[34] F. Wang, Y. Zhang, H. Shao, D. Zhang, and Q. Mou, "Research and application of multiscale residual network models," *J. Electron. Meas. Instrum.*, vol. 33, no. 4, pp. 19–28, 2019.