IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 14, 2020, accepted December 25, 2020, date of publication January 5, 2021, date of current version January 13, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3049159

Use AF-CNN for End-to-End Fiber Vibration Signal Recognition

SAISAI RUAN[1], JIAQING MO[1], LIANG XU[2], GANG ZHOU[1], YAJUN LIU[1], AND XIN ZHANG[1]

[1]Key Laboratory of Signal Detection and Processing, College of Information Science and Engineering, Xinjiang University, Ürümqi 830046, China

[2]School of Electrical and Electronic Engineering, Tianjin University of Technology, Tianjin 300384, China

Corresponding author: Jiaqing Mo (2226966386@qq.com)

This work was supported by the Natural Science Foundation of Xinjiang Uygur Autonomous Region under Grant 2019D01C072.

ABSTRACT Traditional optical fiber vibration signal (OFVS) recognition research focuses on signal endpoint detection and feature extraction. These two aspects directly determine the success of OFVS recognition. The traditional method relies on artificially designed features and has a strong pertinence to the classification target, resulting in poor stability and flexibility. In response to the above problems, this paper combines the traditional OFVS recognition ideas (time-frequency analysis and feature extraction) and the characteristics of deep learning automatic learning parameters to construct an end-to-end adaptive filtering convolutional neural network (AF-CNN), which can directly get the classification results through the iterative update of the network. In modeling the original signal, the following steps were taken to make the network interpretable. First, we use a one-dimensional (1-D) convolution on the original OFVS. The convolution kernel can adaptively treat the original signal perform filtering to obtain filtered signals of different frequency bands. Second, using a general convolutional neural network (CNN) to extract the filtered signal features. Finally, a multi-layer perceptron (MLP) is used for classification. This paper compares the AF-CNN network with three traditional pattern recognition methods and proves that the AF-CNN network's accuracy is better than traditional pattern recognition methods. The average accuracy can reach 96.7%, and it can effectively distinguish OFVS with weaker energy and similar waveforms.

INDEX TERMS End-to-end, adaptive filtering, AF-CNN, 1-D convolution, MLP.

I. INTRODUCTION

The fiber perimeter security system mainly includes the Mach-Zehnder (MZ) type, Michelson type, and Sagnac type. Because it has the advantages of long transmission distance, lack of power supply requirement, strong corrosion-resistance, anti-electromagnetic interference, and low cost, it has been widely used in scenarios such as tunnel detection, oil pipeline monitoring, and border security [1]–[4]. The Sagnac type optical fiber interferometer has zero optical path difference and will not cause additional noise when the two sensing arms' lengths are inconsistent. Moreover, it does not require a reference fiber, has a simple structure, and a long propagation distance. It is very suitable for the distributed deployment of OFVS detection [5]–[7]. This article only studies the vibration signal generated by the Sagnac type optical fiber vibration sensor.

The recognition process of the OFVS is divided into denoising, endpoint detection, feature extraction, and classification recognition [8]. In the denoising stage, Qin *et al.* used a wavelet denoising method to denoise the vibration signal. The wavelet denoising method can effectively eliminate the interference caused by excessive background noise [9]. Subsequently, Xu *et al.* used spectral subtraction to reduce broadband background noise to enhance vibration signals' time-frequency features. The denoising effect is better than the wavelet method [10]. In the endpoint detection stage, Wang *et al.* proposed a method based on the threshold crossing rate to detect whether a vibration signal is generated. However, the detection algorithm is too single, and the detection rate for weak signals is low [11]. Later, Tabi Fouda *et al.* used short-term energy and over-threshold value to detect the vibration signal's endpoint [12]. In the feature extraction stage, time-frequency analysis is performed on the signal first, such as short-time Fourier transform, wavelet decomposition, EMD decomposition, and other

The associate editor coordinating the review of this manuscript and approving it for publication was Prakasam Periasamy.

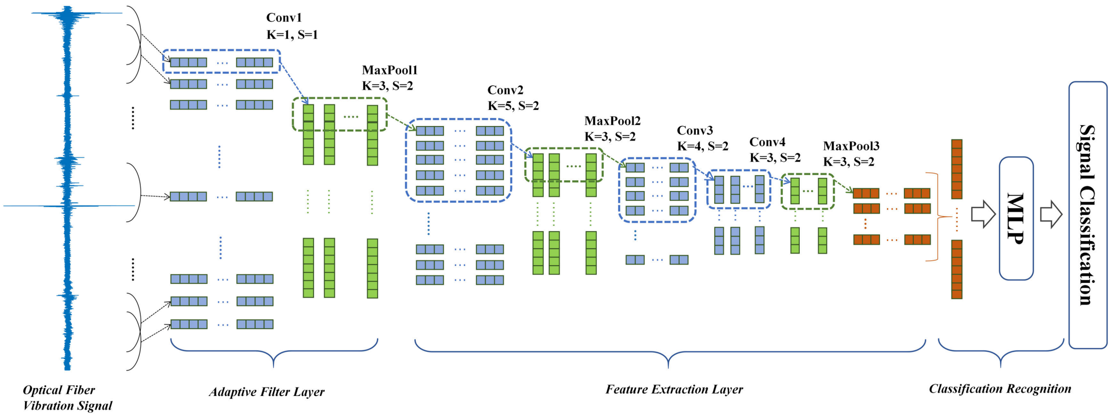


FIGURE 1. **The overall structure of AF-CNN is divided into three parts: adaptive filter layer, feature extraction layer, and classification recognition.**

means. Then, features are extracted from the processed signal. At present, the effective features are the energy ratio, kurtosis, skewness, and spectral entropy of the signal, etc. The final classification and recognition usually use the SVM classifier [10]–[17].

Traditional OFVS recognition needs to design different features for different signals, so it has strong pertinence for classification targets and is not universal. In denoising and endpoint detection, it is easy to remove and ignore vibration signals with weaker energy, and it is challenging to distinguish vibration signals with similar waveforms. Given the problems encountered in the above traditional OFVS recognition and the great success of deep learning in image and speech recognition in recent years [18], [19], this paper decided to adopt an end-to-end AF-CNN for OFVS recognition.

Inspired by Muckenhirn *et al.* and Ravanelli *et al.* on speaker recognition research [20], [21], this paper conducts interpretability research on building a CNN model for OFVS recognition.

- We use a 1-D convolution to filter the original signal.
- Using a general CNN to extract the features of the filtered signal.
- Using MLP with a hidden layer to classify the extracted features.
- Achieve the original signal's end-to-end direct modeling.

All parameters are updated iteratively through the CNN network.

II. THE CONSTRUCTION METHOD OF AF-CNN

The overall framework of AF-CNN is shown in Figure 1. This paper is influenced by the traditional OFVS recognition method and combined with the ideas proposed in the speaker recognition research in the papers [20]–[22], an end-to-end recognition method for fiber vibration signals is constructed. Its architecture has three steps:

A. ADAPTIVE FILTER (AF) LAYER

In this layer, firstly, the original signal is divided into frames. The divided signals are stacked into a two-dimensional (2-D) matrix, facilitating 1-D convolution operation and the extraction of frame-level features. Then the convolution filtering operation is performed using 80 different convolution kernels (convolution kernel $k = 1$, stride $s = 1$), which is equivalent to 80 output channels. Each convolution kernel is equivalent to a filter, through the iterative update of the neural network to adjust the filter parameters. Finally, this layer achieves the purpose of filtering the signal's different frequency bands, providing useful information for distinguishing different OFVSs. In the following experimental section, we will describe the role of the AF layer in detail.

B. FEATURE EXTRACTION (FE) LAYER

Before the filtered signal matrix is input to the FE layer, it is pooled to reduce the feature dimension and the compressed data's parameter amount. The FE layer uses a general CNN for convolution operation, and its internal structure has three convolution layers and two pooling layers. The purpose is to extract the features of the filtered signal.

C. CLASSIFICATION RECOGNITION

The output of CNN is a 2-D feature matrix. First, the 2-D feature matrix is converted into a 1-D vector matrix. Then the feature vector is used as the input of the MLP classifier. The final output directly distinguishes different OFVS.

In the framework, this paper uses the cross-entropy loss function and adaptive moment estimation (Adam) to optimize the network. Through several iterations to update the network parameters, the end-to-end OFVS identification is finally realized.

III. DATA COLLECTION AND PREPROCESSING

The experimental data comes from Xinjiang Meite Intelligent Security Engineering Co., Ltd. The OFVS is generated by the Sagnac interferometric perimeter security early warning

TABLE 1. Details of the AF-CNN architecture.

Layers	Kernel /Stride	Input size	Output size	Remarks
Conv1	k=1/s=1	395×1200	395×80	bias=False
Maxpool1	k=3/s=2	395×80	197×80	-
Conv2	k=5/s=2	197×80	97×128	bias=True
Maxpool2	k=3/s=2	97×128	48×128	-
Conv3	k=4/s=2	48×128	23×80	bias=True
Conv4	k=3/s=2	23×80	11×80	bias=True
Maxpool3	k=3/s=2	11×80	5×80	-
FC1	neurons:600	400	600	bias=True/dropout=0.5
FC2/Output	-	600	4	-

system, and the vibration signal is collected by using two layouts of hanging net and buried. The sampling frequency of the system is 8MHz.

In practical applications, according to the preset parameters of the device, 80k data points are collected in 1s, and the duration of each sampled data is 3∼4s. The length of the collected raw data is different, and there are many silent signals. The signal preprocessing must be carried out before constructing the data set.

The preprocessing steps are shown in Figure 2. First, the system performs endpoint detection on the original OFVS, using the fusion feature of short-term logarithmic energy and short-term spectral entropy [23]. Second, if the vibration signal's starting position is detected, the data of 8w sampling points (time is 1s) length is intercepted from the starting position. In the selection of signal length here, through empirical knowledge, we found that the length of 1s can contain a complete vibration signal, and multiple short signals of walking and running are regarded as one vibration signal. Finally, continue the endpoint detection after the second step is completed and continuously repeat the first and second steps.

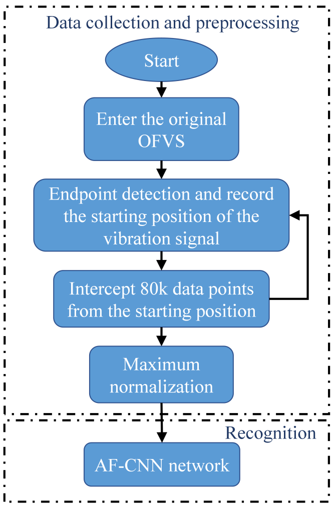

Data collection and preprocessing
Start
Enter the original OFVS
Endpoint detection and record the starting position of the vibration signal
Intercept 80k data points from the starting position
Maximum normalization
Recognition
AF-CNN network

FIGURE 2. Flow chart of data collection and preprocessing.

After preprocessing the original signal, OFVS is cut into a signal segment with a length of 1s. The signal segment is normalized based on the maximum value, and the four types of OFVS preprocessed images are shown in Figure 3.

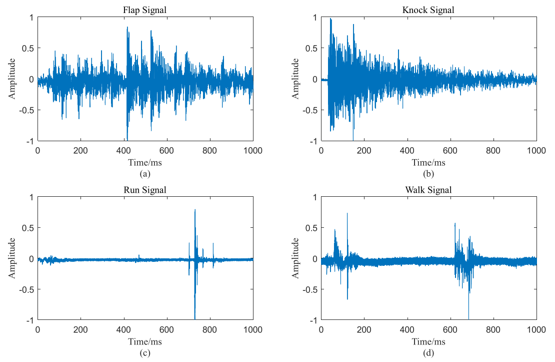


FIGURE 3. Four types of OFVSs after preprocessing: (a) Flap signal, (b) Knock signal, (c) Run signal, (d) Walk signal.

IV. EXPERIMENT AND ANALYSIS

A. NETWORK TRAINING AND TESTING

This article first preprocesses the signal, as shown in Figure 3. After signal preprocessing, a total of 1212 signal fragments were obtained, including 283 flaps, 304 knocks, 300 walks, and 325 runs. This article uses 80% of the data for training the network and 20% for testing.

TABLE 1 is detailed information on AF-CNN architecture. First, the signal is divided into frames (the frame length is 15ms, the frameshift is 2.5ms), and the divided signals are stacked to form a 2-D matrix. Then 1-D convolution is performed on the framed matrix (convolution kernel $k = 1$, stride $s = 1$), and 80 different convolution kernels are used to filter the original signal. The number of output channels is 80. Next, the general CNN network is used to extract the filtered signal features, and three pooling layers are used in the middle to reduce the complexity of the matrix and avoid overfitting. In order to improve the ability of the network to extract features, this paper appropriately increases the number of channels. Finally, the 2-D feature matrix output by the CNN is converted into a 1-D feature vector, and the feature vector is input into the two fully connected (FC) layers for classification. The two FC layers are equivalent to an MLP using a hidden layer. The output of the FC1 layer adds a Dropout layer to avoid data overfitting. The network uses the cross-entropy loss function and Adam optimizer to update.

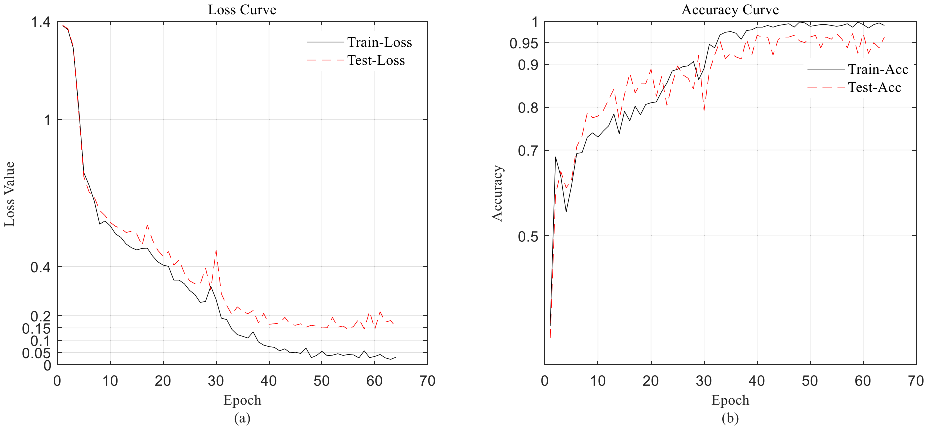


FIGURE 4. The loss value and accuracy curve of AF-CNN.

In this experiment, we build the AF-CNN network to use the PyTorch platform, which has a rich neural network function interface and realizes its rapid construction. The convolutional layer's output and the FC layer (MLP) in the network use the Rectified Linear Unit (ReLU) activation function, the learning rate is adjusted to 10^{-4}, and 64 iteration epochs are performed. Obtain the loss value curve and accuracy curve of the AF-CNN network, as shown in Figure 4.

It can be seen from Figure 4(a) that as the number of iterations increases, the loss values of the test set and training set are slowly decreasing. In about 40 iterations, the neural network gradually converges, the loss value of the training set drops to about 0.02, and the test set's loss value drops to about 0.17. As the loss value of the AF-CNN network decreases, the classification accuracy is gradually increasing. As shown in Figure 4(b), it can be seen that the accuracy has stabilized after 40 iteration epochs, and the network has indeed converged. At this time, the accuracy of the training set is close to 99%, and the accuracy of the test set is stable at around 95%.

In this paper, test samples are used to verify the classification accuracy of AF-CNN. The test sample classification result is made into a confusion matrix, as shown in Figure 5. It can be found from Figure 5 that there are 56 flap samples, of which 53 are classified correctly, one is classified as a knock, and two are classified as run, with a recognition accuracy of 94.64%. Among 61 knocking samples, 57 were classified correctly, two were classified as flapping, and two were classified as running, with a recognition accuracy of 93.44%. A total of 60 walking samples, all classified as correct. There are 65 running samples, of which 64 are classified correctly, and one is classified as flapping, with a recognition accuracy of 98.46%. From the confusion matrix, the average accuracy of AF-CNN is 96.69%, which can effectively distinguish four types of OFVSs.

B. AF LAYER

When the traditional pattern recognition classifies the OFVS, the first step is to perform a time-frequency analysis on the signal, and then the feature parameters can be designed according to the features of the time-frequency signal.

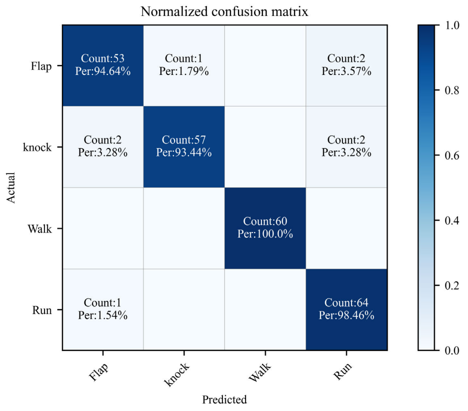


FIGURE 5. The confusion matrix of AF-CNN.

Typical time-frequency analysis is a short-time Fourier transform (STFT) and wavelet decomposition. STFT needs to be windowed in the time-domain; The narrow window has high time resolution and low frequency resolution; The wide window has low time resolution and high frequency resolution. For time-varying unsteady signals, high frequencies are suitable for small windows, and low frequencies are suitable for large windows. However, the window width of the STFT is fixed, and the window width will not change in an STFT, so the STFT cannot meet the demand for analyzing the time-varying frequency of the unsteady signal.

Wavelet decomposition changes the basis function, replacing the infinite trigonometric function basis with a finite length wavelet basis, and its energy decays in the time-domain. Compared with the fixed-window STFT, the wavelet base's size can be scaled, which solves inconsistent resolution in the time-domain and frequency-domain. Besides, the wavelet decomposition can also be seen as filtering different frequency bands of the signal. Wavelet decomposition solves low resolution of STFT, but when performing wavelet decomposition, once the wavelet base is determined, only a fixed bandwidth can be filtered, and the bandwidth cannot be adjusted at any time to deal with different types of vibration signals.

We hope to use a basis function that adjusts the frequency band at any time to filter the signal and only filter the bandwidth that is useful for classification. In other words, it can adjust the filter frequency band adaptively. In response to this problem, inspired by papers [20], [21], we propose an AF-CNN network with an AF layer. Because it combines the CNN neural network, it can realize the automatic update and optimization of the parameters.

The first step of AF-CNN is to perform a 1-D convolution operation on the original waveform, equivalent to filtering the original signal. Here, 80 different convolution kernels are used for filtering. These convolution kernels are equivalent to some finite impulse responses (FIR) filter [24], where each filter will select the frequency band of interest. The definition of each convolution is as follows:

$$y_k[n] = x[n] * h_k[n] = \sum_{l=0}^{L-1} x[l] \cdot h_k[n-l], \quad k = 1, 2, 3, \ldots 80. \quad (1)$$

$x[n]$ is the original OFVS input; $h_k[n]$ is the k-th filter of length L; $y_k[n]$ is the k-th filter's output. The above formula is the convolution operation of the adaptive filter in the time-domain. According to the time-domain convolution formula, the frequency-domain convolution formula is derived:

$$Y_k[\omega] = X[\omega] \cdot H_k[\omega], \quad k = 1, 2, 3, \ldots, 80. \quad (2)$$

$X[\omega]$ is the frequency-domain form of the input signal; $H_k[\omega]$ is the filter's frequency response function; $Y_k[\omega]$ is the frequency-domain form of the filter output. The time-domain convolution corresponds to the frequency-domain is the product. Its purpose is bandpass filtering the different frequencies of the signal.

The structure of the AF layer is shown in Figure 6. The first step is to frame the original signal. The frame length is 1200 sampling points, and the frameshift is 200 sampling points. After framing, the stack forms a 2-D matrix of 395 × 1200 and then uses a 1-D convolution operation to output a 395 × 80 matrix. As shown in the second box in Figure 6, the AF layer essentially uses FIR filtering on the original signal, and framing the original signal is entirely for the convenience of subsequent 1-D convolution operations. The length of the filter is 1200, which corresponds to the frame length of the framing step, and the number of filters is 80, which corresponds to the number of 1D convolution output channels.

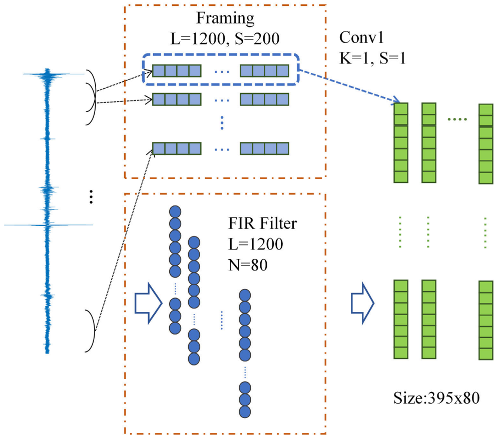


FIGURE 6. In the structure diagram of the AF layer, the upper half is the actual structure of the network, and the lower half is the equivalent structure of the network.

This paper extracts 80 different convolution kernels from the first layer of the trained AF-CNN neural network and analyzes their filtering features. In the research of this paper, it is found that the trained adaptive filter has a different sensitivity to different frequencies. Eighty different filters are numbered, K = 1, 2...... 80, and the frequency response of some filters is shown in Figure 7. Through the analysis of 80 different filters, it can be found that the filters are mainly sensitive to the frequency range of 0∼5kHz, the center frequencies of different filters are very different, and the center frequencies of some filters are as high as 36kHz. The results show that the feature parameters of high frequency and low frequency can affect the classification results.

C. FE LAYER

After the AF layer of the AF-CNN neural network filters the signal, the output is 80 filtered signals of different frequency bands. The output matrix of the AF layer size of 395 × 80 is used as the FE layer input. As shown in the Part1 of Figure 8, a general CNN is used to extract features of the filtered signal. There are three convolutional layers inside the CNN, and two pooling layers are used to reduce network complexity. Two FC layers are used after feature extraction, equivalent to an MLP with a hidden layer to classify and recognize the extracted features. The structure is shown in the Part2 of Figure 8.

In the experiment, the general CNN network is used to extract the features of OFVS. In order to observe the effect of network feature extraction, this paper extracts the output of four OFVS signals after passing CNN.

As shown in Figure 9, this paper compares the features extracted from four different OFVSs through the CNN network. From the feature curves, it can be found that the feature curves of signals with similar energy and similar waveform are very close, such as run and walk signal, flap and knock signal. The original walk and run signals have low energy, so the feature curve graph's amplitude is low. On the contrary, the feature curve amplitude of the flap and knock signals is higher. From the amplitudes of the four OFVSs' feature curves, it can be found that the feature of the four signals has a high degree of discrimination, and even signals with similar waveforms can be clearly distinguished in the feature diagram.

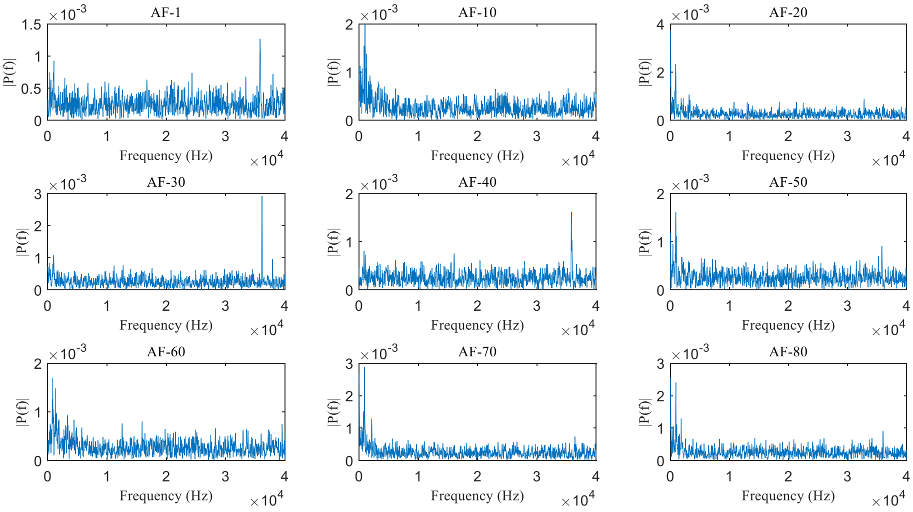


FIGURE 7. This is a partial waveform diagram of the adaptive filter (convolution kernel) frequency-domain.

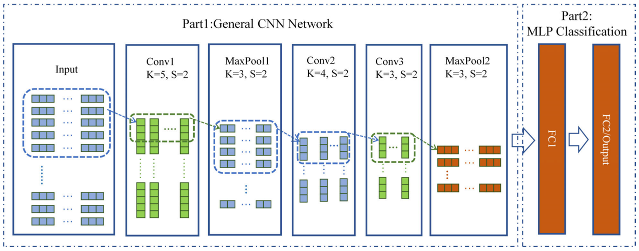


FIGURE 8. Schematic diagram of the FE layer and classification layer structure.

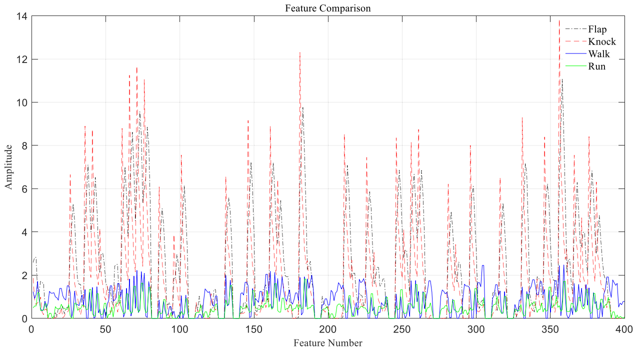


FIGURE 9. Comparison of features extracted from four different OFVSs after passing through the CNN network.

D. EXPERIMENTAL COMPARISON

The three most popular traditional OFVS recognition methods are used to compare with the AF-CNN network; they are wavelet decomposition method, EMD decomposition method, and VMD [23] decomposition method. They all use features such as the kurtosis, skewness, and energy ratio of the decomposed signal, and finally, they combine multiple features and use SVM for classification.

For the experiment's fairness, the traditional SVM classification uses the same training set and test set as the AF-CNN. Since SVM classification does not require many samples, the number of samples here is reduced to half of the original. A total of 600 samples were used, 150 samples of each type, and training and testing were randomly allocated at a ratio of 8:2.

The accuracy of traditional OFVS recognition is affected by multiple factors, such as denoising, endpoint detection, and feature selection, all affecting classification accuracy. Because denoising and endpoint detection maybe eliminate weaker energy vibration signals, it will affect the comparison results of the two identification methods on the effectiveness of features. The traditional pattern recognition here refers to the signal after denoising and endpoint detection, and its classification accuracy is mainly affected by feature parameters.

The comparison experiment of the four OFVS classification methods is shown in Table 2. The experiment shows that the three traditional recognition methods have higher accuracy for flapping and knocking signals. In recognition of flapping signs, the wavelet and VMD decomposition method's accuracy can get 93.3%. When recognizing knocking signs, the accuracy of the three conventional recognition methods can reach over 97%. However, the traditional three methods have low recognition accuracy for walking and running signals. Among them, the recognition accuracy

TABLE 2. Recognition accuracy comparison.

Method Classification	Wavelet	EMD	VMD[23]	AF-CNN
Flap	93.3%	91.0%	93.3%	94.6%
Knock	100.0%	97.7%	99.3%	93.4%
Walk	69.0%	76.7%	86.9%	100.0%
Run	79.9%	80.7%	92.9%	98.5%
Average	81.8%	83.8%	92.2%	96.7%

of walking signals is the worst. The EMD decomposition method and VMD decomposition method can only reach 76.7% and 86.9%. In recognition of running signs, the VMD decomposition method with the highest accuracy can only reach 92.9%.

This paper uses the AF-CNN network to identify the four OFVS. The flap signal's recognition accuracy is slightly higher than the traditional three methods, which is 94.6%. However, the knock signal's recognition accuracy is lower than the traditional three methods, which is 93.4%. This network is more sensitive to walking and running signals, and the recognition accuracy is much higher than the traditional three methods, which are 100.0% and 98.5%. Overall, the average accuracy of the AF-CNN network is 96.7%, and the accuracy of the traditional three methods can only reach 92.2%. Among the four types of OFVS recognition, the traditional three methods have higher recognition accuracy for flapping and knocking signals. The AF-CNN network used in this article not only has a higher recognition accuracy for walking and running signals but also has a higher average recognition accuracy than three traditional methods.

V. CONCLUSION

The recognition accuracy of weak signals, such as walking signals, is usually low in traditional pattern recognition. It is because traditional pattern recognition needs to detect the endpoints of the original vibration signal first and then divide it into very short vibration signal fragments, which duration is about 0.25s. In a very short time scale, it is difficult to distinguish the feature parameters of walking and running signals. Although signals like flapping and knocking are also cut into vibration signal fragments, the vibration signals can be effectively distinguished by using the fusion features, that is because their vibration energy is high, and vibration duration is relatively long, which is about 1s.

The deep learning method is very different from the traditional pattern recognition method. The deep learning method uses equal-length data, where the length is 1s. Through the constructed AF-CNN network, the signal's vibration information can be effectively extracted from these equal length data, and various types of OFVS can be distinguished. The main work of this paper is summarized as follows:

- In terms of data acquisition, this article first detects the starting position of the vibration signal, then intercepts the signal of equal length, and finally continues to perform endpoint detection, looping the first and second steps. After the above method processing, the original OFVS is cut into signal fragments with a length of 1s, where the length is determined by empirical knowledge. Each signal segment contains a type of OFVS, and these signal segments are used to form a data set.
- When recognizing OFVS, this article constructs an end-to-end AF-CNN network based on deep learning and explains this network. The AF-CNN network refers to the traditional methods of time-frequency analysis, feature extraction, and classification. First, an adaptive filter layer is constructed for signal time-frequency analysis, then a general CNN network is used to extract features, and finally, MLP is used for classification. Since all the parameters of AF-CNN are automatically updated through the network's iteration, there is no need to manually set the filter bandwidth and feature parameters, which avoids human errors and enhances the adaptability to different environments.
- In the experimental section, this article trains and tests the AF-CNN network. By analyzing the network's loss curve and accuracy curve, it is found that its average accuracy on the test set can reach 96.7%. Using three traditional pattern recognition methods and AF-CNN comparison, the method proposed in this paper is higher than the traditional method in average accuracy, an increase of 4.5%.

The above experimental analysis and comparison show that the AF-CNN network can effectively recognize four common OFVSs of flapping, knocking, walking, and running, which exceeds the traditional pattern recognition methods.

REFERENCES

[1] F. Bi, C. Feng, H. Qu, T. Zheng, and C. Wang, "Harmful intrusion detection algorithm of optical fiber pre-warning system based on correlation of orthogonal polarization signals," *Photon. Sensors*, vol. 7, no. 3, pp. 226–233, Sep. 2017.

[2] W. Liang, L. Zhang, Q. Xu, and C. Yan, "Gas pipeline leakage detection based on acoustic technology," *Eng. Failure Anal.*, vol. 31, pp. 1–7, Jul. 2013.

[3] G. Allwood, G. Wild, and S. Hinckley, "Optical fiber sensors in physical intrusion detection systems: A review," *IEEE Sensors J.*, vol. 16, no. 14, pp. 5497–5509, Jul. 2016.

[4] J. C. Juarez, E. W. Maier, K. N. Choi, and H. F. Taylor, "Distributed fiber-optic intrusion sensor system," *J. Lightw. Technol.*, vol. 23, no. 6, pp. 2081–2087, Jun. 2005.

[5] C. Zhu, D. Zuo, and J. Wang, "Intrusion signal recognition of perimeter security based on sagnac sensor," *Transducer Microsyst. Technol.*, vol. 35, no. 1, pp. 19–21 and 28, 2016.

[6] Q. Zhu and W. Ye, "Distributed fiber-optic sensing using double-loop sagnac interferometer," in *Proc. 9th IEEE Conf. Ind. Electron. Appl.*, Jun. 2014, pp. 499–503.

[7] A. D. McAulay and J. Wang, "A Sagnac interferometer sensor system for intrusion detection and localization," *Proc. SPIE*, vol. 5435, pp. 114–119, Aug. 2004.

[8] B. Wang, Q. Sun, S. Pi, and H. Wu, "Research on the feature extraction and pattern recognition of the distributed optical fiber sensing signal," *Proc. SPIE*, vol. 9193, Sep. 2014, Art. no. 91930C.

[9] Z. Qin, L. Chen, and X. Bao, "Wavelet denoising method for improving detection performance of distributed vibration sensor," *IEEE Photon. Technol. Lett.*, vol. 24, no. 7, pp. 542–544, Apr. 1, 2012.

[10] C. Xu, J. Guan, M. Bao, J. Lu, and W. Ye, "Pattern recognition based on enhanced multifeature parameters for vibration events in φ-OTDR distributed optical fiber sensing system," *Microw. Opt. Technol. Lett.*, vol. 59, no. 12, pp. 3134–3141, Dec. 2017, doi: 10.1002/mop.30886.

[11] L. Wang, Y. Guo, T. Sun, J. Huo, and L. Zhang, "Signal recognition of the optical fiber vibration sensor based on two-level feature extraction," in *Proc. 8th Int. Congr. Image Signal Process. (CISP)*, Oct. 2015, pp. 1484–1488.

[12] B. M. T. Fouda, D. Han, B. An, X. Lu, and Q. Tian, "Events detection and recognition by the fiber vibration system based on power spectrum estimation," *Adv. Mech. Eng.*, vol. 10, no. 11, pp. 1–9, 2018.

[13] Z. Sheng, X. Zhang, Y. Wang, W. Hou, and D. Yang, "An energy ratio feature extraction method for optical fiber vibration signal," *Photon. Sensors*, vol. 8, no. 1, pp. 48–55, Mar. 2018.

[14] J. C. Zhang, Z. Zeng, P. Lai, H. Feng, and S. Jin, "A recognition method with wavelet energy spectrum and wavelet information entropy for abnormal vibration events of a petroleum pipeline," *J. Vib. Shock*, vol. 29, no. 5, pp. 1–4, 2010.

[15] C. Aneesh, S. Kumar, P. M. Hisham, and K. P. Soman, "Performance comparison of variational mode decomposition over empirical wavelet transform for the classification of power quality disturbances using support vector machine," *Procedia Comput. Sci.*, vol. 46, pp. 372–380, Jan. 2015.

[16] H. Wei, M. Wang, B. Song, X. Wang, and D. Chen, "Study on the magnitude of reservoir-triggered earthquake based on support vector machines," *Complexity*, vol. 2018, pp. 1–10, Jul. 2018.

[17] L. H. Jiang, J. Y. Gai, W. B. Wang, X. L. Xiong, S. Liang, and X. Z. Sheng, "Ensemble empirical mode decomposition based event classification method for the fiber-optic intrusion monitoring system," *Acta Optica Sinica*, vol. 35, no. 10, 2015, Art. no. 1006002.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[19] S. Hershey, S. Chaudhuri, D. P. W. Ellis, J. F. Gemmeke, A. Jansen, R. C. Moore, M. Plakal, D. Platt, R. A. Saurous, B. Seybold, M. Slaney, R. J. Weiss, and K. Wilson, "CNN architectures for large-scale audio classification," Jan. 2016, *arXiv:1609.09430*. Accessed: Nov. 12, 2020. [Online]. Available: http://arxiv.org/abs/1609.09430

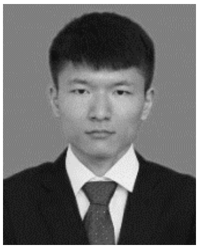
[20] H. Muckenhirn, M. Magimai-Doss, and S. Marcell, "Towards directly modeling raw speech signal for speaker verification using CNNS," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 4884–4888.

[21] M. Ravanelli and Y. Bengio, "Speaker recognition from raw waveform with SincNet," in *Proc. IEEE Spoken Lang. Technol. Workshop (SLT)*, Dec. 2018, pp. 1021–1028.

[22] D. Palaz, M. Magimai-Doss, and R. Collobert, "End-to-end acoustic modeling using convolutional neural networks for HMM-based automatic speech recognition," *Speech Commun.*, vol. 108, pp. 15–32, Apr. 2019.

[23] J. Bao, J. Mo, L. Xu, Y. Liu, and X. Lv, "VMD-based vibrating fiber system intrusion signal recognition," *Optik*, vol. 205, Mar. 2020, Art. no. 163753.

[24] L. Rabiner and R. Schafer, *Theory and Applications of Digital Speech Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 2010.



SAISAI RUAN received the bachelor's degree in electronic information science and technology, in China, in June 2019. He is currently pursuing the master's degree in information and communication engineering with Xinjiang University. He passed the Chinese Graduate Admissions Examination and was admitted to Xinjiang University, in September 2019. His current research interest includes processing of distributed optical fiber vibration signals.

JIAQING MO was born in Yangjiang, Guangdong, China, in 1972. He received the M.S. degree from the Beijing University of Technology. Since 2019, he has been a Professor with the College of Information Science and Engineering, Xinjiang University. He is the author of more than 50 articles, and holds more than ten patents. His research interests include life information detection and analysis, photoelectric information detection and sensor, perceptual intelligence, social public security, and topology photonics.

LIANG XU was born in Xinjiang, China, in 1980. He received the B.S. degree in electronics and information engineering from Xidian University, in 2004, and the Ph.D. degree in control science and engineering from Xi'an Jiao Tong University. He achieved the postdoctoral research in computer science and technology. From 2007 to 2012, he was a Lecturer with the School of Information Science and Engineering, Xinjiang University. Since 2013, he has been an Associate Professor with the School of Electrical and Electronics Engineering, Tianjin University of Technology. He is currently the Under-Secretary-General of System Simulation Committee, China Automation Society. He has presided more than ten science and technology projects, including National Natural Science Foundation, Chinese Postdoctoral Science Foundation, Xinjiang Natural Science Foundation, and major science and technology projects in Tianjin. He is the author of two books, and more than 40 research articles and holds six patents. His research interests include aero-optic imaging and applications, intelligent information processing, and artificial intelligence applications. He is the editorial board member of two chinese academic journals.

GANG ZHOU was born in Ürümqi, Xinjiang, China, in 1981. He received the Ph.D. degree from the School of Information and Communication Engineering, Xi'an Jiaotong University, in 2013. He is currently an Associate Professor with Xinjiang University. His current research interests include computer vision, image processing, and pattern recognition.

YAJUN LIU was born in Shihezi, Xinjiang, China, in 1973. She received the B.S. degree in material physics from Beijing Normal University, the M.S. degree in information and communications engineering from Xinjiang University, in 2005, and the Ph.D. degree in electronic science and technology from Xi'an Jiaotong University, in 2013. Since 2013, she has been a Lecturer with the College of Information Science and Engineering, Xinjiang University. Her research interests include physical characteristic and application of photonic devices, photonic crystal, and topological photonic crystal.

XIN ZHANG was born in Nanjing, Jiangsu, China, in 1994. He received the bachelor's degree in engineering in 2018. He is currently pursuing the master's degree in engineering from Xinjiang University, under the supervision of Prof. Jiaqing Mo. His main research interest includes identification of optical fiber vibration signals.