IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 24 August 2024, accepted 17 October 2024, date of publication 23 October 2024, date of current version 31 October 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3485075

RESEARCH ARTICLE

A Time-Frequency Depth Convolutional Recurrent Network for Seismic Waveform Automatic Classification

FU LI, DIQUAN LI, YANFANG HU, YUNQI ZHU, YECHENG LIU, ZHE WANG, AND HANYU ZHU

Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring, Ministry of Education, Central South University, Changsha 410083, China
Key Laboratory of Non-Ferrous and Geological Hazard Detection, Changsha 410083, China
School of Geosciences and Info-Physics, Central South University, Changsha 410083, China

Corresponding author: Diquan Li (lidiquan@csu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 42474170, in part by the Science and Technology Innovation Program of Hunan Province under Grant 2023RC1014, in part by the Natural Science Foundation of Hunan Province under Grant 2023JJ40222, and in part by the Research Foundation of Education Bureau of Hunan Province under Grant 22A0457.

ABSTRACT Seismic monitoring has been instrumental in various domains such as natural earthquake early warning, mineral mining safety assessment, and hydraulic fracturing impact evaluation. However, the monitoring data often exhibit low signal-to-noise ratio (SNR) and large volume. Developing an efficient, high-precision, and universally applicable seismic waveform automatic classification network model becomes significant and practical. We propose a physical interpretable time-frequency deep convolutional recurrent neural network (TF-DCRNN) model which consists of an integration of a time-frequency convolutional (TFconv) layer and a convolutional recurrent neural network (CRNN). Subsequently, we evaluate the classification performance by comparing five network models, including convolutional neural network (CNN) and long short-term memory (LSTM), using Ricker wavelet datasets with varying SNR levels ($-15 \sim 0$ dB). Our findings verify the superiority of the TF-DCRNN model in the classification of strong interference environment from both numerical and physical simulation. Moreover, integrating multiple network models or incorporating a TFconv layer can moderately enhance the classification performance, which provides the direction for network model optimization.

INDEX TERMS TFconv layer, TF-DCRNN, Ricker wavelet, waveform classification, physical simulation.

I. INTRODUCTION

Seismic monitoring is fundamental activity in the domains of natural earthquake early warning, mining safety assessment, and hydraulic fracturing monitoring [1], [2]. The huge number and variable quality of seismic monitoring data makes it a challenging task to quickly and effectively differentiate seismic signals from other interference signals. Seismic waveform classification is currently done using a range of automatic or semi-automatic methods. The most common ones are the short term averaging / long term averaging algorithm (STA/LTA) [4], the Akaike Information Criterion method (AIC) [5] and the waveform autocorrelation and cross-correlation methods based on waveform similarity [6], etc. These methods have some limitations in terms of recognition accuracy and timeliness, and the quality of the data greatly influences how well they work.

In recent years, deep learning [7], [8] based on feature learning and matching, such as convolutional neural networks (CNNs) [9], recurrent neural networks (RNNs) [10] and generative adversarial networks (GANs) [11], have been widely applied in the data processing field of seismic monitoring with the rapid development of big data and artificial intelligence. Perol et al. proposed ConvNetQuake model based

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

on convolutional neural network, which realized the classification of a single seismic waveform, with several orders of magnitude higher efficiency than traditional methods, and it was successfully applied to induced seismic monitoring in Oklahoma, USA [12]. Ross et al. used approximatively 300,000 seismic records in Southern California to build a data set and trained them with CNN model. The obtained model has strong generalization ability and has been effectively tested in Kumamoto region of Japan [13]. Zhao et al. realized automatic classification of seismic waveforms with a 95% accuracy based on deep convolutional neural network [14]. Zheng et al. applied deep recurrent neural networks to the identification and extraction of microseismic or acoustic emission events, and studied the accuracy and robustness of the method when the SNR was higher than −5 dB [15]. Li et al. used a generative adversarial network to learn the characteristics of seismic P-waves based on 300,000 seismic data set and the experimental results showed that the network model could identify 99.2% of seismic P-waves and 98.4% of noise [16].

The traditional deep learning methods, such as CNNs and RNNs, are typical black box models, although they have strong feature extraction ability and high efficiency and have been successfully applied in seismic data classification [17]. They are uninterpretable and difficult to find the logical basis for feature extraction and classification, which reduce the credibility of results. Seismic signals are complex waveforms, which makes it extremely challenging to extract feature variables that can fully represent seismic waveforms. It is worth noting that seismic signals possess unique time and frequency domain features. And these features can be extracted by applying the time-frequency transformation algorithm to convert 1D time domain seismic signals into 2D time-frequency spectral data. The 2D time-frequency spectral identification is more suitable for the feature extraction of seismic signals and has the physical significance of time-frequency features compared to the direct identification of seismic signals in the 1D time domain. Therefore, many researchers have integrated time-frequency transformation techniques into deep learning network models to boost interpretability and enhance network performance. Zhang et al. proposed CWT-CNN signal classifier by combining continuous wavelet transform (CWT) and CNN, and the experimental results show that the classification performance of CWT-CNN is significantly better than that of basic depth feedforward neural network through synthetic microseismic data set and field microseismic data set (SNR in −5∼5 dB) [18]. Dokht et al. integrated deep convolutional neural network and wavelet transform for model training aiming at incomplete seismic datasets with low SNR, and verified the network performance with 99% recognition accuracy rate based on the data set of more than 4900 earthquakes recorded in western Canada [19]. Bi et al. proposed an interpretable time-frequency convolutional neural network (XTF-CNN) that allows the model to capture seismic signal characteristics from both time and frequency features. Meanwhile, they also evaluated network performance tests using low SNR data and obtained better experimental results [20]. Not only in seismic monitoring, but also in other fields such as mechanical system fault detection, sound detection and so on, there are a lot of relevant studies. Li et al. proposed a wavelet-driven WaveletKernelNet (WKN), where the network is designed with a physically meaningful continuous wavelet convolution (CWConv) layer to replace the first convolutional layer of the standard CNN, and the results of mechanical diagnostic experiments showed that the WKN's accuracy is improved by more than 10% over the CNN [21]. Wei et al. constructed a time-frequency convolutional neural network based on Hilbert-Huang transform and applied it to the automatic classification of single-channel EEG sleep stage signals with an average accuracy of 84.5% [22]. Chen et al. [23] proposed a class of explainable time-frequency network (TFN) taking into account three typical time-frequency transformation methods of short time Fourier transform (STFT), chirplet transform (CT) [24] and wavelet transform (WT) to develop the time-frequency convolutional layer and proved the network effectiveness through three mechanical fault diagnosis experiments. In addition, many researchers have carried out multi-network fusion research in order to improve network performance. Lim et al. combined the advantages of CNNs and RNNs and proposed a convolutional recurrent neural network, which won the first place in the Task 2 challenge of DCASE2017 and verified the effectiveness of the method [25]. Liu et al. designed the convolution recurrent neural network blind equalizer (CRNNBE) by integrating CNN and RNN. Simulation results showed that CRNNBE has faster convergence and higher accuracy compared with the method based on RNNs or CNNs [26]. Zhang et al. proposed a multi-scale time-frequency convolutional recurrent neural network (MTF-CRNN) for sound event detection, and verified the network with strongly competitive in the DCASE2017 Task2 dataset and the DCASE2019 Task3 dataset (the SNR of data is between −6 and 6 dB) [27]. Previous studies have shown that the accuracy and reliability of signal classification can be improved by integrating multi-network model or embedding signal time-frequency characteristic information.

In summary, deep learning is widely used in automatic classification of seismic signals and related fields. However, previous studies have mostly focused on seismic data with high SNR (>−5 dB), and the measured seismic signals are often severely affected by interference, resulting in low SNR [28], [29]. There is an urgent need to develop an efficient, high-precision, and universally applicable network model, and discussing the effectiveness of the model classification in the context of low SNR data has practical significance.

STFT and WT are the most prevalently utilized time-frequency analysis approaches. STFT boasts high processing efficiency and acquires time-frequency spectra with uniform time-frequency resolution. WT features time-frequency windows of varying sizes at different times and

frequencies. However, subject to the constraints of the principle of uncertainty, the time resolution and the frequency resolution cannot be optimally utilized in both directions. Moreover, WT has a lower processing efficiency and cannot obtain more information compared to STFT. Seismic signal is a type of signal characterized by a continuous frequency band, and the frequency range is distributed from a few Hz to several thousand Hz (Requiring high resolution at both low and high frequencies over a wide frequency range). At the same time, seismic monitoring is a long-term monitoring process that generates large amounts of data in the time domain (Requiring algorithms with high efficiency). The fundamental objective of this study is to screen the data and improve the efficiency of data processing. Considering the above influencing factors, the application of STFT is more appropriate for the processing of seismic signals.

We integrate CRNN and STFT algorithms to propose a TF-DCRNN architecture for seismic signal classification in strong interference environments. We simulate seismic signals using Ricker wavelet to establish seismic datasets affected by different random noise interferences. Then, five network models, including CNN, LSTM, CRNN, TF-CNN and TF-LSTM, are used to compare the classification effect. The reliability of TF-DCRNN is further verified by physical simulation experiments of seismic monitoring in order to provide an efficient and reliable automatic classification model architecture of seismic signals.

The main contributions of this study are summarized as follows:

1) A novel TF-DCRNN architecture, which exhibits physical interpretability and is specifically designed for efficient automatic classification of time series data, is presented.
2) The classification effect of 6 kinds of network models under different intensity noise interference is compared and analyzed.
3) The reliability of TF-DCRNN is verified from both numerical and physical simulation.

II. PRELIMINARY

A. RICKER WAVELET

Seismic wavelet is a signal with limited energy, short duration and fixed time-frequency features. It can also be understood as the real waveform record that is excited by the source and received by the detector during the actual data collection. Presently, geophysicists extensively agree that the Ricker wavelet exhibits a high degree of similarity to the seismic wavelet, rendering it a suitable and prevalent choice for use as a seismic simulation source function.

The mathematical expressions of the time domain and frequency domain of the Ricker wavelet are shown in (1) ∼ (2).

$$r(t) = \left(1 - 2\pi^2 f_m^2 t^2\right) \cdot \exp\left[-(\pi f_m t)^2\right]. \quad (1)$$

$$R(f) = \left[2f^2 / \left(\sqrt{\pi} f_m^2\right)\right] \cdot \exp\left[-(f/f_m)^2\right]. \quad (2)$$

where $r(t)$ is the time domain, $R(f)$ is the frequency domain, t is the time, f is the frequency, f_m is the signal main frequency.

The Ricker wavelet, characterized by its single peak and brief temporal extent, can be divided into minimum phase wavelet (wavelet energy is concentrated in the front), mixed phase wavelet (wavelet energy is concentrated in the middle), maximum phase wavelet (wavelet energy is concentrated in the tail) and zero phase wavelet (a special mixed phase wavelet, symmetric at the time origin, phase spectrum is zero) according to the phase of the wavelet. The zero-phase Ricker wavelet is frequently employed in seismic numerical simulations. TABLE 1 shows the time-domain waveform and spectral characteristics of a zero-phase Ricker wavelet with a main frequency of 60 Hz.

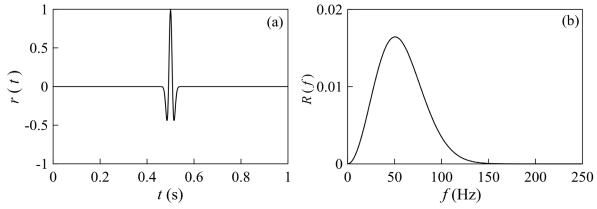

FIGURE 1. Time domain waveform and frequency spectrum of 60Hz Ricker wavelet [30]. (a) Time domain waveform. (b) Frequency spectrum.

It is readily observable that the wavelet main frequency, maximum peak amplitude, and wavelet width are the key feature variables representing seismic wavelet. For instance, in Fig. 1, the main frequency of the Ricker wavelet is 60 Hz, the maximum peak amplitude is 1, and the wavelet width is 0.1 s. Nevertheless, the seismic wavelet is a kind of complex waveform, and it is extremely challenging to extract the feature variables that can fully represent the seismic wavelet waveform. This is mainly because seismic wavelets possess both time and frequency domain features, and also exhibit the features of amplitude and frequency decay in the time domain. Utilizing the three feature variables of "wavelet main frequency, maximum peak amplitude, and wavelet width" can generally describe the wavelet waveforms, but it is not feasible to finely compound the wavelets.

B. TIME-FREQUENCY TRANSFORM AND TFconv LAYER

Most signals in nature are non-stationary, such as seismic signals, which are typical non-stationary signals. Fourier transform (FT) has a unique advantage in analyzing stationary signals. However, FT can only transform the time domain signal to the frequency domain as a whole for non-stationary signals. The transformed signal acquires frequency resolution but lacks time resolution, thus failing to accurately represent the signal's temporal features at different frequencies. Therefore, it becomes essential to conduct a comprehensive time-frequency analysis of the signals to effectively capture the time-frequency features of non-stationary signals.

The time-frequency analysis methods for non-stationary signals mainly include STFT, WT, etc. STFT is one of

the most commonly used time-frequency analysis methods, as shown in (3), where a function is multiplied by a window function and then a 1D Fourier transform is performed, and a series of Fourier transform results are obtained by sliding the window function, which are lined up vertically to obtain a 2D spectrum. WT is actually still a sliding window STFT, except that the window size is taken to be shorter at high frequencies and longer at low frequencies, making high frequencies more accurate in the time domain and low frequencies more accurate in the frequency domain. We can't measure both the frequency and time domains of a signal, either the frequency domain is inaccurate or the time domain is inaccurate, both of which are constrained by the principle of inaccuracy. Thus, WT doesn't really get more accurate information than Fourier.

$$\mathrm{STFT}(t,\omega)=\int_{-\infty}^{\infty}h(\tau)g(\tau-t)e^{-i\omega\tau}d\tau. \tag{3}$$

where $\mathrm{STFT}(t,\omega)$ can be regarded as the spectrum at time t, $h(t)$ is the analyzed time domain signal, $g(t)$ is the window function and $e^{-i\omega t}$ is the trigonometric basis function.

Fig.2 illustrates the time-frequency windows of a time-domain signal obtained through FT, STFT, and WT, respectively. In this representation, each small square symbolizes a distinct time-frequency window, with the side length along the time axis indicating time resolution and the side length along the frequency axis denoting frequency resolution. The time domain signal lack frequency resolution and the frequency domain signal derived through FT can offer superior frequency resolution but lacks time resolution. STFT provides a certain level of resolution in both time and frequency domains, with consistent time-frequency resolution across the entire spectrum. WT employs time-frequency windows of varying sizes at different time and frequency points.

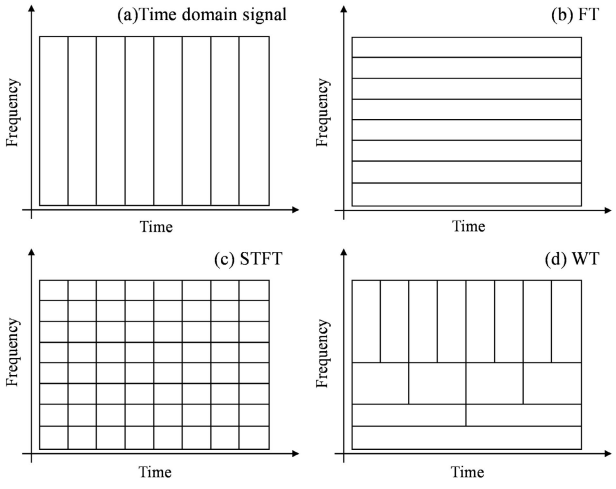

FIGURE 2. Time-frequency resolution of FT, STFT and WT [31], [32]. (a) Time domain signal. (b) FT. (c) STFT. (d) WT.

As is well known, the feature extraction effect of the convolutional layer has a great influence on the result of the whole model. In the CNNs model, a set of time domain convolution between the input signal and the convolution kernel of randomly initialized weights can automatically extract high-dimensional features from the original sample and efficiently complete classification and prediction. However, the convolution operation cannot accurately extract the key influential components of the signal because the decision logic is not clear enough and the feature extraction does not have practical physical significance. While the time-frequency transformation method allows for leveraging time-frequency information for interpreting physical signals, it may face limitations in actively filtering and adaptively extracting essential time-frequency features.

The TFconv layer represents a fusion of the strengths of CNNs and time-frequency transform methods, where the time-frequency transform is integrated within the convolutional layer. Each convolution kernel comprises two components during the execution of the TFconv layer: a real-part convolution kernel and an imaginary part convolution kernel. These components perform convolutions on the input sample along the length dimension, extracting real-part and imaginary part features, respectively. Subsequently, the extracted real and imaginary features are combined through modulation to generate the feature map, which serves as the output of the TFconv layer, as depicted in (4).

$$\begin{cases} y_{\mathrm{real}}=\psi_{\mathrm{real}}*x \\ y_{\mathrm{imag}}=\psi_{\mathrm{imag}}*x \\ y=\sqrt{y_{\mathrm{real}}^2+y_{\mathrm{imag}}^2} \\ \psi_{\mathrm{real}},\quad \psi_{\mathrm{imag}}\in R \end{cases} \tag{4}$$

where x is the input signal, y is the final output data (instantaneous frequency feature distribution). ψ_{real} and ψ_{imag} are the real kernel and imaginary kernel, respectively, and the internal kernel function is equivalent to the inner product window function in the time-frequency transform. y_{real} and y_{imag} are the feature maps of the real and imaginary parts, respectively.

C. CONVOLUTIONAL RECURRENT NEURAL NETWORKS

CRNNs represent sophisticated deep learning architecture that combine components from CNNs and RNNs. This fusion equips the model with the collective strengths of both CNNs and RNNs. CNNs, a type of feedforward neural network, consist of layers such as convolutional layer, pooling layer, and fully connected layer, etc. CNNs excel in extracting features from input data through the stacking of convolutional and pooling layers, with the fully connected layer facilitating tasks like classification and prediction, particularly in image processing applications. Notably, CNNs typically possess fewer parameters compared to traditional artificial neural networks, and increasing the network's depth can enhance its feature extraction capabilities. RNNs set themselves apart from traditional neural networks with their recurrent cell structure that includes memory capabilities and parameter sharing. This design enables RNNs to store past information

in an internal state and utilize it for current tasks, providing a significant advantage in processing time series data. Traditional RNNs often only consider the most recent input data state, leading to challenges such as gradient vanishing or explosion. Specialized RNN variants like LSTMs address this issue by selectively retaining or forgetting information through the introduction of cell states, enhancing the structural stability of RNNs when handling long sequences and making LSTMs a widely adopted technique in the RNN domain.

CRNNs leverage the strengths of both CNNs and RNNs architectures. In this model, the convolutional layer extracts local features that are then integrated and memorized by the recurrent layer. This integration allows weight sharing among convolutional kernels, reducing the need for extensive hyperparameter training and mitigating overfitting risks. Additionally, CRNNs can process data of various dimensions, effectively capturing the multi-dimensional features of input data and expanding the network's applicability. Fig.3 offers a succinct overview of a typical CRNN structure [33], demonstrating the incorporation of LSTM modules into the CNN architecture, with the flexibility to use single or multiple LSTM units.

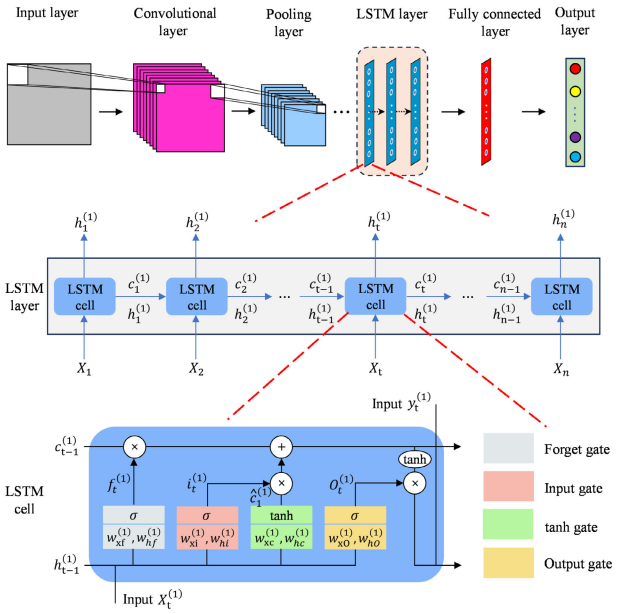

FIGURE 3. The overall structure of CRNN.

III. METHODOLOGY

In traditional networks, including CNNs, RNNs, CRNNs, and others, direct recognition of 1D time series often exhibits inadequate feature extraction, which is compounded by a lack of physical meaning in the extraction process. Such a deficiency can give rise to less-than-optimal performance, particularly in circumstances where time series signals are severely disturbed. CRNN integrates the advantages of CNN and RNN and possesses good universality, nonetheless, it still suffers from the aforementioned problems.

Seismic wavelet constitutes a distinctive type of time series featuring both time and frequency domain features. Capitalizing on this physical information, a TFconv layer is incorporated into the network architecture. This incorporation converts 1D time-domain seismic signals into 2D time-frequency spectral data. Subsequently, a CRNN is utilized to conduct in-depth extraction and classification of the time-frequency data, thereby improving the overall performance of the network.

This study proposes a TF-DCRNN architecture for the fast automatic classification of seismic signals. TF-DCRNN integrates the CRNN and TFconv layer. Since CNN and LSTM network models have many variants (e.g., GRU, BLSTM, etc.), we consider TF-DCRNN model construction using standard CNN and LSTM networks. Therefore, TF-DCRNN is mainly composed of input layer, TFconv layer, standard convolutional layer, pooling layer, flatten layer, LSTM layer, fully connected layer and output layer. The basic network architecture is shown in Fig.4, and the network design follows the following strategies:

1) The input layer is 1D time series data with an unfixed length, which can be used for multi-channel data synchronization input.
2) The TFconv layer is directly connected to the input layer. When considering the existence of seismic monitoring with huge amount of raw data, wide signal bandwidth and high timeliness, it is obvious that it is more suitable to construct TFconv using STFT than other time-frequency transform algorithms. Therefore, we constructed a TFconv layer using STFT. 2D data with time-frequency characteristics is obtained by performing time-frequency convolution on the input time series signal.
3) The standard convolutional layer is a 2D convolution kernel. Convolutional kernels with larger scales have larger local receptive fields, and stacking multiple layers of convolution kernels can improve the non-linear processing ability. Therefore, we can construct multiple standard convolutional layers for signal feature depth extraction, mapping time-frequency feature maps to hidden layer feature spaces. In this study, we set three layers of convolution kernels with sizes of 5×5, 4×4 and 3×3, with 10 kernels per layer.
4) The pooling layer can reduce the feature dimension and computational complexity. Seismic signal features occupy a lower proportion comparing with noise data, so it is more appropriate to use maximum pooling for feature extraction. We constructed four maximum pooling layers with sizes of 5×5, 4×4, 3×3 and 2×2 to further extract signal features.
5) The flatten layer is responsible for flattening the 2D feature data into 1D feature data. In this study, a TFconv layer is employed to transform 1D data into a 2D data, thereby facilitating the extraction of features from the

2D data. Consequently, the flatten layer plays a pivotal role and is deemed indispensable.

6) The LSTM layer receives and outputs 1D data. There can be multiple LSTM layers, but the more the number, the higher the computational complexity. According to previous experience [27], [33], as shown in Fig.4, we set up 2 long short-term memory layers.
7) The fully connected layer introduces the full connection operation at the last layer of the network to transform the feature mapping extracted from the upper layer into the final classification.
8) The output layer is responsible for the output of category labels. This paper set up four categories of labels, "0", "1", "2" and "3" as shown in Fig.4.

In addition, the TF-DCRNN also includes four BN layers, three ReLU layers and three dropout layers and a softmax layer, for a total of twenty-five layers. It is particularly noteworthy that the TF-DCRNN architecture proposed in this paper boasts significant flexibility. The number of convolutional layer, pooling layer, or LSTM layer can be appropriately augmented to facilitate deeper feature extraction. Nonetheless, it is not advisable to simply increase the number of network layers, as this would elevate the model's computational complexity and heighten the risk of overfitting. The design of the network layer count is grounded in the outcomes of extensive prior research in the field.

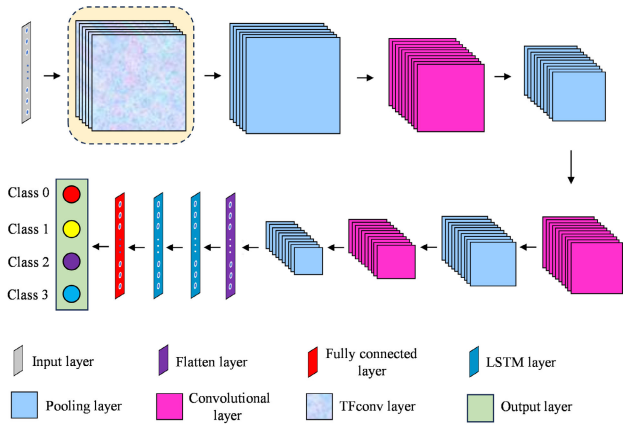

FIGURE 4. The architecture of TF-DCRNN.

IV. DATASET AND COMPARISON MODELS

A. DATASET CONSTRUCTION

In seismic exploration, the high-frequency component of the broad-frequency spike pulse generated by the source is significantly attenuated due to formation absorption, while the medium to low-frequency components are preserved. As the reception distance increases, the frequency component decreases. The seismic signal is inherently unsteady, representing a minute fraction time of the overall lengthy time series, and is frequently plagued by random noise or noise from other frequency bands, resulting in a low SNR in the collected data. Fig.5 illustrates the time-domain waveforms of 60Hz Ricker wavelet under different SNR.

Previous studies have mostly focused on seismic data with high SNR (>-5 dB), and actual seismic signals are often severely affected by interference, resulting in low SNR. Therefore, datasets comprising Ricker wavelet data and pure noise data are generated using Ricker wavelets at different SNR to assess the signal classification performance of the TF-DCRNN network under diverse noise interferences. The dataset constructed in this study includes four types of SNR and each dataset consists of four categories:

1) Category "1": including pure and noise-containing signals, wavelets main frequency 1∼60 Hz, maximum peak amplitude 0.1∼1, wavelets width 0.05∼0.2s.
2) Category "2": including pure and noise-containing signals, wavelets main frequency 61∼150 Hz, maximum peak amplitude 0.1∼1, wavelets width 0.05∼0.2s.
3) Category "3": including pure and noise-containing signals, wavelets main frequency 151∼250 Hz, maximum peak amplitude 0.1∼1, wavelets width 0.05∼0.2s.
4) Category "0": pure random noise, no fixed main frequency, amplitude -0.5∼0.5.

Each type of SNR dataset comprises 80,000 samples, with each label type containing 20,000 samples. Half of the signals are noise-free within categories "1", "2", and "3", while the other half are signals corrupted by varying levels of random noise (with SNR values of 0 dB, -5 dB, -10 dB, and -15 dB).

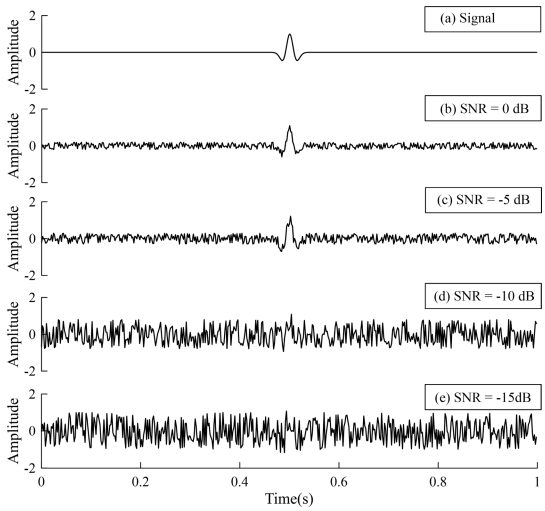

FIGURE 5. The time-domain waveform of 60Hz Ricker wavelet under different SNR. (a) Noiseless signal. (b) SNR = 0 dB. (c) SNR = −5 dB. (c) SNR = −10 dB. (d) SNR = −15 dB.

B. COMPARISON MODELS

In order to conduct a comprehensive evaluation of the network's classification performance, this study introduces five comparative network models with the main architecture shown in Fig. 6, and the networks design is as follows:

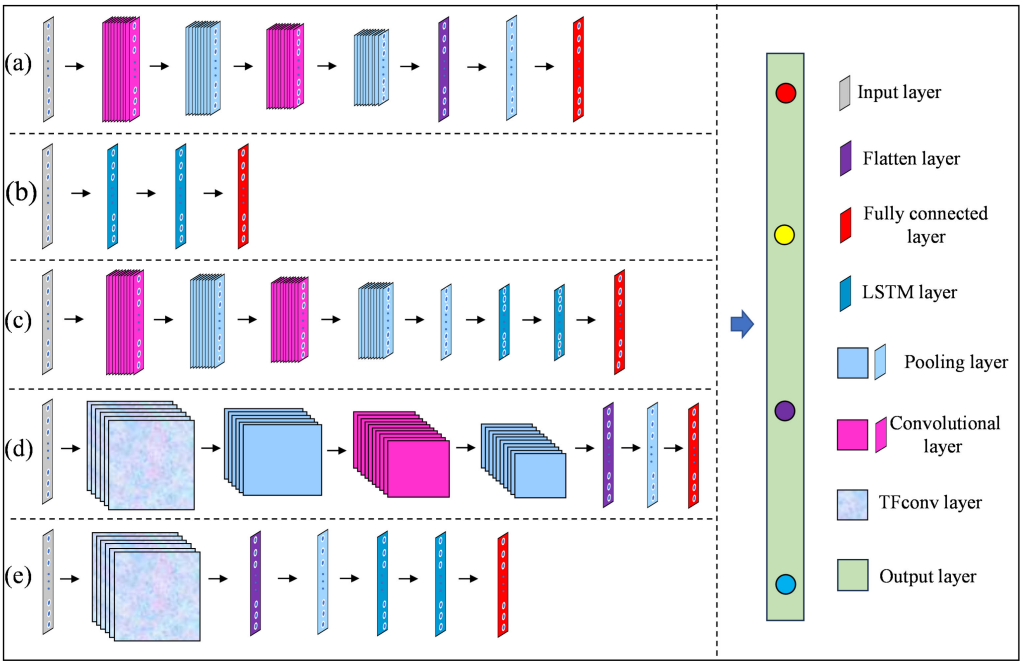

FIGURE 6. The main architecture of five comparative network models. (a) CNN. (b) LSTM. (c) CRNN. (d) TF-CNN. (e) TF-LSTM.

TABLE 1. The mean and variance of classification accuracy of different network models.

Network	SNR=0 dB		SNR=-5 dB		SNR=-10 dB		SNR=-15 dB	
	Accuracy (%)	Variance	Accuracy (%)	Variance	Accuracy (%)	Variance	Accuracy (%)	Variance
CNN	96.33	0.0083	94.35	0.0140	88.43	0.0465	85.40	0.1124
LSTM	93.06	0.2590	90.25	0.2704	80.14	0.8289	75.73	1.5566
CRNN	97.23	0.1094	96.78	0.1210	91.86	0.1806	87.89	0.2964
TF-CNN	96.58	0.0310	95.17	0.0705	89.76	0.0798	86.85	0.1745
TF-LSTM	96.26	0.0964	93.47	0.1108	83.15	0.2364	80.39	0.4614
TF-DCRNN	**98.14**	**0.0182**	**97.05**	**0.0229**	**93.42**	**0.0575**	**90.26**	**0.0985**

1) CNN, a traditional 1D CNN, has 14 network layers, including 2 convolutional layers and 3 maximum pooling layers. The number of convolution kernels in each convolutional layer is 10, and the size of convolution kernels is 5×1 and 4×1, respectively.
2) LSTM, the number of network layers is 10, including 2 LSTM layers.
3) CRNN, a combination of CNN and LSTM, has 18 network layers, including 2 convolutional layers, 3 maximum pooling layers and 2 LSTM layers.
4) TF-CNN, which adds TFconv layer on the basis of traditional CNN, has 16 network layers, including 2 convolutional layers and 3 maximum pooling layers. Since the TFconv layer converts 1D time series data into 2D time-frequency feature maps, the convolution kernels used in the later convolution layers are 2D, with sizes of 4×4 and 3×3, respectively, and the number of convolution kernels in each convolution layer is 10.
5) TF-LSTM, which adds TFconv layer on the basis of traditional LSTM, has 12 network layers, including 2 LSTM layers and 1 maximum pooling layer.

V. RESULT OF CLARIFICATION

As we all know, the selection of hyperparameters affects the training speed, convergence, generalization ability and other aspects of the model, e.g., the learning rate, the number of iterations, and the regularization parameter are common hyperparameters. The selection of hyperparameters is usually a trial-and-error process, which needs to be adjusted based on experience and relevant data features. We have set the main common hyperparameters of all network models as follows in the previous period through a large number of tests, when we conducted the comparison experiments of this paper: The maximum number of cycles is 30, the MiniBatch-Size of data used in each iteration is 256, and the initial learning rate InitialLearnRate is 0.001. The optimization

algorithm adopted random optimization with high computational efficiency ADAM [34]. And the proportion of training set, validation set and test set data is 60%, 20% and 20%, respectively.

The final classification result was determined by averaging the results from 50 experimental groups during the experiment. Then this study compared the average classification accuracy and variance across different networks, as presented in Table 1. It reveals that CNN outperforms LSTM in terms of classification effectiveness, and integrating CNN with LSTM or adding a TF-conv layer can improve network classification performance. Among the various network models examined, TF-DCRNN demonstrates superior classification effectiveness, exhibiting high accuracy and robust stability. Specifically, all network models achieve classification performance exceeding 90% with higher quality data (SNR ≥ -5 dB). TF-DCRNN achieves classification accuracies of 98.14% and 97.15% respectively for SNR values of 0 dB and -5 dB. However, the classification performance of all network models decreases as the SNR decreases. When SNR $= -10$ dB, only CRNN and TF-DCRNN achieve classification accuracy greater than 90%. When the data is disturbed by strong noise (SNR $= -15$ dB), the classification performance of traditional CNN and LSTM models deteriorates. The accuracy drops to 85.40% for CNN and 75.73% for LSTM, respectively, and the network stability decreases (the variance of LSTM exceeds 1). In contrast, the TF-DCRNN model can still maintain a better classification effect, with an accuracy of 90.26% and a variance of less than 0.1. Then, we calculated the mean and standard deviation of the model accuracy under different SNRs based on the data in the table, and plotted the box plots (Fig. 7) in order to further demonstrate the performance of the different models for the comprehensive classification of the data under different SNRs. The green line in the figure is the median value of accuracy, and the box represents the centralized location of the data, which contains 50% of the intermediate data, and its upper and lower boundaries represent 25% of the high-value data and 25% of the low-value data, respectively. According to the box plot characteristics of different models, we further validate the classification performance of the TF-DCRNN model.

The convergence curve during a training process at SNR $= -10$ dB is presented in Fig. 8. The results distinctly indicate that the TF-DCRNN model exhibits superior convergence characteristics compared to traditional models. Notably, the TF-DCRNN model demonstrates the fastest rate of loss value reduction, achieving the smallest loss value with a stable curve that converges close to the optimal point within 100 iterations. Conversely, the traditional CNN and LSTM models show lower convergence efficiency with unstable curves even after 1,000 iterations, and LSTM model is easy to fall into local optimization. The comparison reveals that the convergence performance of the TF-CNN model surpasses that of the CNN model. Similarly, the TF-LSTM model exhibits superior convergence compared to the LSTM model, and the CRNN model outperforms both the CNN and LSTM models in terms of convergence. Therefore, the network performance can be enhanced through the following approaches: 1) integrating TF-conv layers into conventional networks like TF-CNN, TF-LSTM, etc.; 2) combining multiple network models, such as the CRNN formed by combining CNN and LSTM.

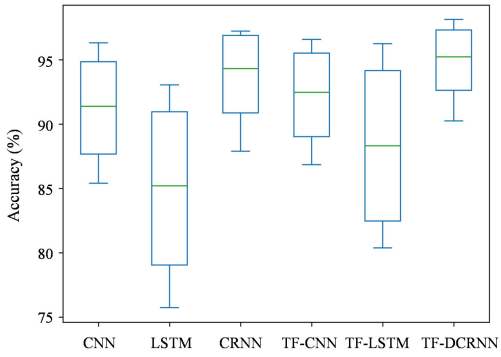

FIGURE 7. Comparison of mean accuracy and standard deviation of six models at different SNRs.

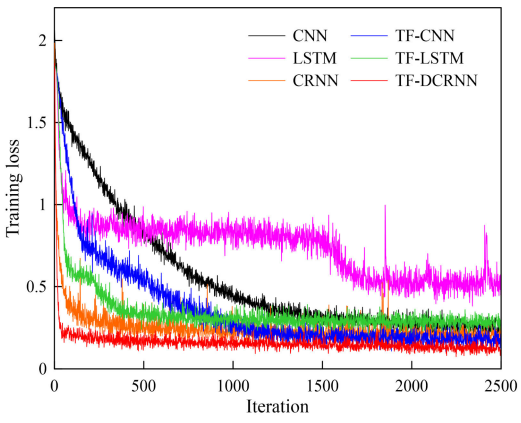

FIGURE 8. The convergence curves of different network models.

In addition, we used confusion matrixes to evaluate the classification performance of the models. The confusion matrixes can visually examine the classification of the test set, with values on the diagonal representing the number of correct classifications of the model for that category, and non diagonal elements representing the number of incorrect classifications for that category. At the same time, we calculated the Precision, Recall and extended F1 scores of each category in order to further demonstrate the effectiveness, as depicted in (4) $\sim$ (6). It is worth noting that the larger the values of parameters Precision, Recall and F1, the better the performance of the network model.

$$\text{Precision} = \frac{TP_i}{TP_i + FP_i}. \tag{5}$$

$$\text{Recall} = \frac{TP_i}{TP_i + FN_i}. \tag{6}$$

$$\mathrm{F1} = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}. \tag{7}$$

where TP_i, FP_i, and FN_i are the number of True Positive, False Positive, and False Negative for category Class_i, respectively.

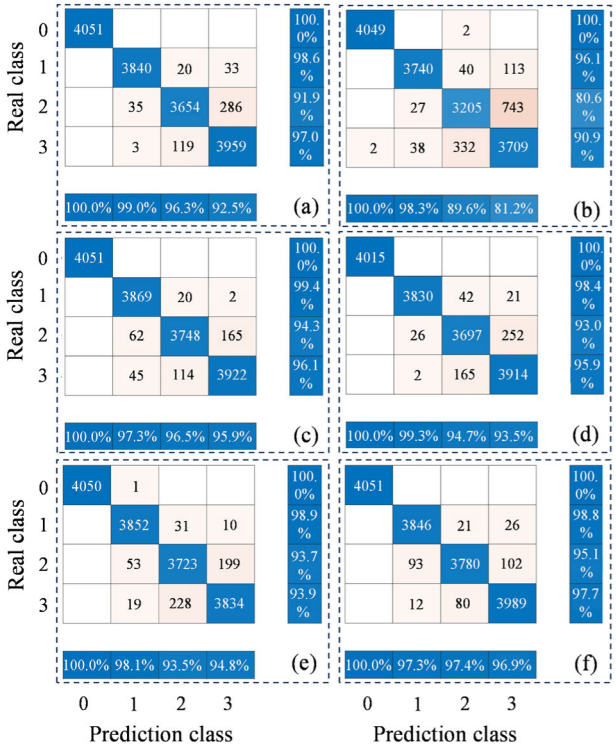

FIGURE 9. Classification effect when SNR = 0 dB. (a) CNN. (b) LSTM. (c) CRNN. (d) TF-CNN. (e) TF-LSTM. (f) TF-DCRNN.

Fig. 9 to Fig. 12 present the classification performance of all network models across different SNR, where the 4×4 grid data is the confusion matrix in each sub-figure, the bottom row of data represents the model's Precision in the corresponding category, and the right-most column of data represents the model's Recall in the corresponding category. Fig. 12 shows the comparison of the F1 scores comparison of the different network models under the same test set. The analysis of the figures reveals that all network models exhibit a commendable classification performance concerning random noise, particularly when the data SNR equals or exceeds -5 dB, enabling precise identification of random noise. And there is a slight decline in recognition accuracy as the noise level escalates, with most models maintaining a rate of 99%. Nevertheless, conventional CNN and LSTM models demonstrate subpar classification efficacy when confronted with noise from neighboring frequency bands. Even at an SNR of -5 dB or higher, the error rate in identification surpasses 10% for the traditional LSTM model, escalating to over 30% at an SNR of -10 dB or lower. It is evident that the F1 score for the LSTM model is the lowest when compared to all other models under evaluation. This score notably diminishes as the level of noise escalates. Specifically, the F1 score drops

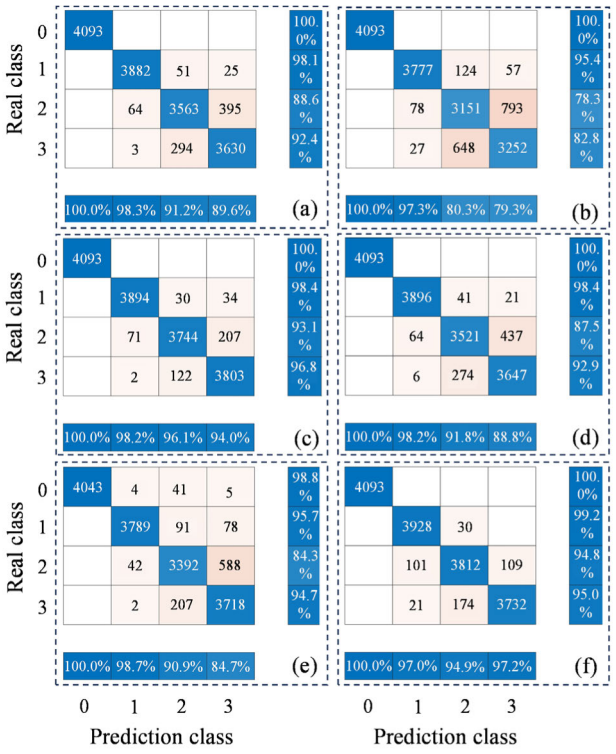

FIGURE 10. Classification effect when SNR = -5 dB. (a) CNN. (b) LSTM. (c) CRNN. (d) TF-CNN. (e) TF-LSTM. (f) TF-DCRNN.

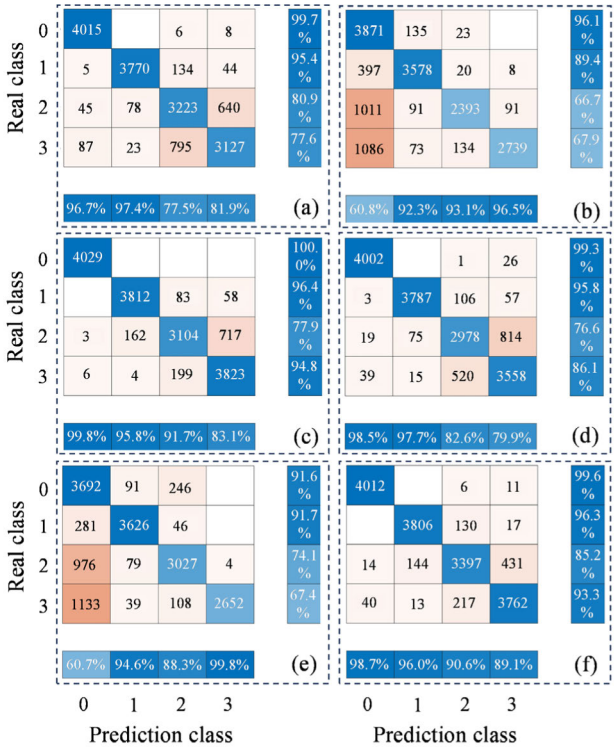

FIGURE 11. Classification effect when SNR = -10 dB. (a) CNN. (b) LSTM. (c) CRNN. (d) TF-CNN. (e) TF-LSTM. (f) TF-DCRNN.

to only 0.81 when the SNR is at -15 dB. In contrast to LSTM, the CNN model demonstrates superior classification

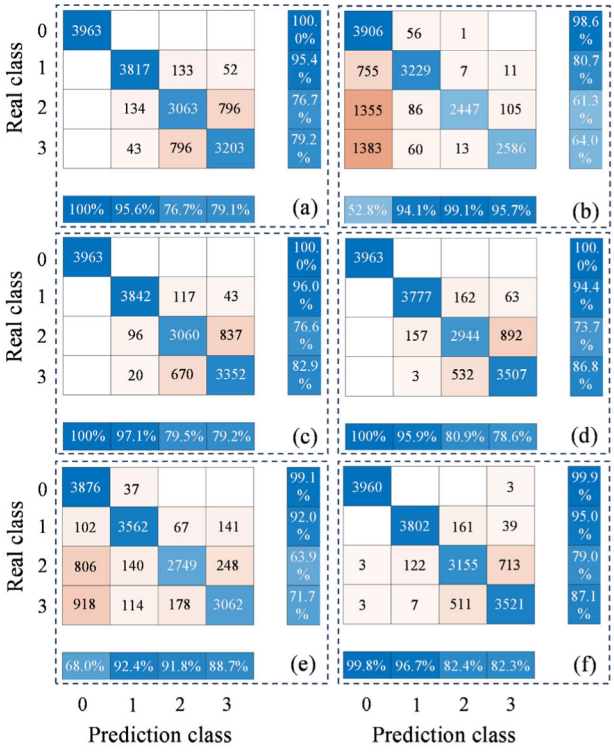

FIGURE 12. Classification effect when SNR = −15 dB. (a) CNN. (b) LSTM. (c) CRNN. (d) TF-CNN. (e) TF-LSTM. (f) TF-DCRNN.

efficacy in handling interference from adjacent frequency bands, with recognition error rates ranging from 10% to 30% and the F1 score approximately ranging from 0.88 to 0.89 under substantial random noise interference (SNR < −5 dB). The CRNN model exhibits enhanced classification performance comparing to the traditional CNN and LSTM model, which showcases a reduction of approximately 5% and 15% in error rates for adjacent frequency band identification and the F1 score approximately rangs from 0.89 to 0.92 at SNRs of −10 dB or −15 dB. The incorporation of the TF-convolutional layer enhances the network model's ability to recognize interference from adjacent frequency bands, resulting in a reduction of 5% to 10% in recognition error rates. Furthermore, this enhancement is clearly evidenced by the F1 score, which demonstrates a marked improvement for the model equipped with the TF-convolutional layer, as compared to the network model without this layer. The enhancement in F1 score varies between 0.01 and 0.05. TF-DCRNN model has the best classification effect compared with CNN, LSTM and other comparison models, especially when the data is interfered by adjacent frequency bands, the model can still maintain a good recognition effect. Notably, the TF-DCRNN model achieves an error rate of approximately 5% for noise recognition in adjacent frequency bands, with the F1 score approximately ranging from 0.97 to 0.98, when SNR is equal to or greater than −5 dB. At SNRs of −10 dB or −15 dB, the error rate for noise recognition in adjacent frequency bands reaches around 15%, the F1 score approximately rangs from 0.90 to 0.94, thereby confirming the reliability of the TF-DCRNN model proposed in this study.

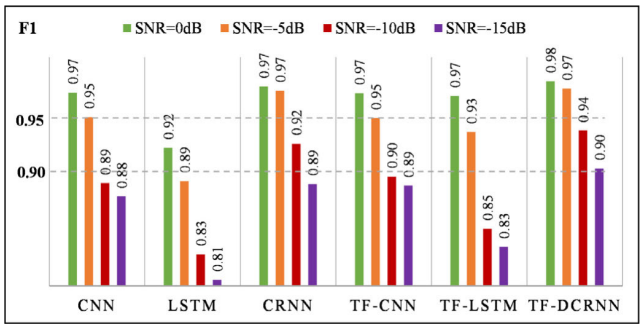

FIGURE 13. F1 scores comparison of the different network models under the same test set.

VI. EXPERIMENT

In order to further verify the reliability of the classification effect of TF-DCRNN model, we carried out a physical simulation experiment of source monitoring by using small-scale explosion source. The experimental model is shown in Fig. 14. We used a plastic pipe with a length of 1.5 m and a diameter of 18 cm to simulate the wellbore. Then 24 high-precision single-component (vertical component) seismometers were set up on the surface to receive seismic signals based on the layout of the star observation system, and the measuring stations were spaced 10∼20 m apart. During the experiment, 50 focal points were positioned at the shaft's base, with continuous injection of 80 °C high-temperature water to induce focal point rupture through heat, facilitating seismic signal acquisition.

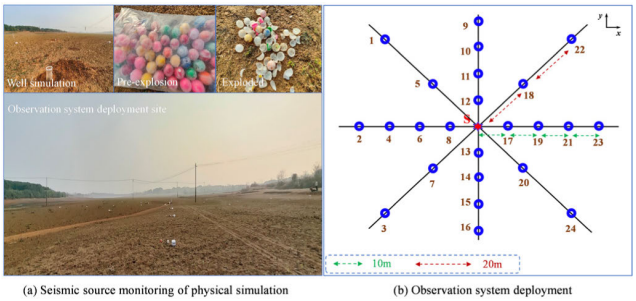

FIGURE 14. Physical model. (a) Seismic source monitoring of physical simulation. (b) Observation system deployment.

The experimental monitoring spanned a total duration of 1 hour. Fig. 14a depicts the schematic illustration before and after the source explosion. As illustrated in Fig. 14b, the distances between the source and measuring stations No.8, 6, 4, and 2 were sequentially 10 m, 20 m, 30 m, and 40 m. The signal strength diminishes with increasing distance, resulting

TABLE 2. Comparison of classification effect of measured data.

Network	Station 8		Station 6		Station 4		Station 2	
	Number	Accuracy (%)	Number	Accuracy (%)	Number	Accuracy (%)	Number	Accuracy (%)
CNN	46	92	43	86	37	74	32	64
LSTM	43	86	37	74	29	58	25	50
CRNN	49	98	46	92	42	84	39	78
TF-CNN	48	96	45	90	40	80	36	72
TF-LSTM	46	92	41	82	33	66	29	58
TF-DCRNN	**50**	**100**	**48**	**96**	**45**	**90**	**41**	**82**

in lower data SNR. Fig. 15 showcases the time-domain waveforms measured at the aforementioned four stations. The SNR of the data obtained at measuring station No.8 is the highest, enabling clear identification of the source signal. Measuring station No.6 exhibits a lower SNR compared to No.8, yet both the source signal and environmental noise remain distinguishable. However, the data SNR collected at measuring stations No.4 and No.2 is considerably low, hindering easy identification of the source signal.

Subsequent to the initial analysis, a time-frequency transform using FT and FFT of measuring station No. 8 was conducted, yielding the time-frequency characteristics of both the source signal and environmental noise, as depicted in Fig. 16. Inspection of the figure reveals that the noise pervades the entire duration, predominantly exhibiting low-frequency characteristics, with energy primarily concentrated within the 0 to 50 Hz range. Conversely, the source signal exhibits brief durations ($<$ 1s) and is characterized by frequencies ranging from 60 to 150 Hz. Consequently, data falling within the frequency bands below 60 Hz and above 150 Hz can be classified as noise.

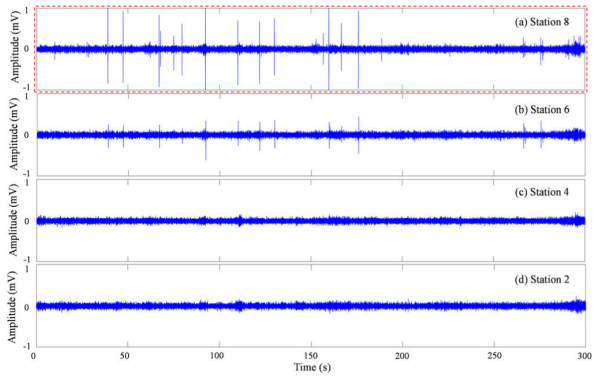

FIGURE 15. Time domain waveforms with different SNR. (a) Station 8. (b) Station 6. (c) Station 4. (d) Station 2.

The one-hour duration of measured data was uniformly segmented into 1-second intervals, resulting in 3600 sub-data sets. Subsequently, signal classification was executed utilizing the previously constructed Ricker wavelet data set and the time-frequency characteristics of the source signal.

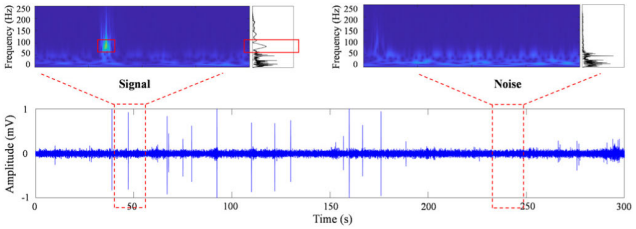

FIGURE 16. Signal and noise time-frequency characteristics.

The classification outcomes for measuring stations No.2, 4, 6, and 8 are presented in Table 2. The results indicate that the TF-DCRNN model demonstrates superior classification performance, achieving an accuracy exceeding 96% when utilizing data from stations No. 8 and No. 6. Moreover, the accuracy rates for stations No. 4 and No. 2 are 90% and 82%, respectively. Conversely, the accuracy of alternative network models falls below 80% at station No. 2. Notably, the TF-DCRNN model exhibits heightened resilience to interference, rendering it particularly suitable for signal recognition within environments characterized by substantial interference, outperforming other network models in this regard.

Utilizing the classification results from TF-DCRNN, the waveforms containing source signals are extracted. Subsequently, the source location algorithm [35], based on waveform stacking, is employed to determine the source location. The results of the source location are illustrated in Fig. 17 (with vertical scaling), where the blue plane represents

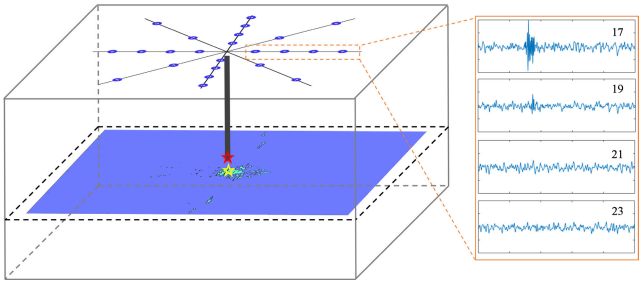

FIGURE 17. Source location based on waveform stacking.

the imaging effect map of the locating. It is evident that the focal point of the imaging plane is centrally positioned. The yellow five-pointed star denotes the source location derived from the location algorithm, while the red five-pointed star signifies the actual source location. The proximity of the source location to the actual position, with an absolute error of less than 5m, serves as validation for the efficacy of the source signal classification.

VII. CONCLUSION

This paper introduces an interpretable TF-DCRNN model that integrates TFconv layer and CRNN. The model's performance is evaluated by constructing Ricker wavelet data sets with different SNR and training it alongside five network models including CNN and LSTM. The classification effectiveness of each network model across different SNR data sets is compared and analyzed. The results of the numerical simulations demonstrate that the TF-DCRNN model exhibits superior classification performance compared to other models. Moreover, the incorporation of multiple network models or the incorporation of TFconv layer can enhance the classification efficacy of the network models to a certain extent. However, the recognition efficacy of network models diminishes in environments characterized by intense random noise interference, with a limited ability to differentiate noise across adjacent frequency bands. The classification accuracy of the TF-DCRNN model exceeds 90% with an approximate 15% recognition error rate for interference signals in adjacent frequency bands when SNR is -15 dB. In contrast, the accuracy rates of the other five networks are below 90%, with recognition error rates for interference signals in adjacent frequency bands ranging from 20% to 40%. These findings underscore the reliability of the TF-DCRNN model. Furthermore, the network model is validated using 3600 measured data samples obtained from physical simulation experiments, affirming the TF-DCRNN model's efficacy in seismic signal classification.

Our work provides a direction for network model optimization and an effective network model for seismic signal identification in strongly disturbed environments. However, we have only shown one of the models with a specific number of layers, and there are some differences in the performance of the models with different numbers of layers, which are not sufficiently analyzed in this regard. In the model comparison experiments, the model parameters have not been analyzed in a more detailed way, which is one of the directions for the follow-up work. Furthermore, seismic monitoring poses challenges due to the intricate signal characteristics influenced by source type and propagation medium. The utilization of a simplified Ricker wavelet for seismic source simulation lacks universality. Hence, our future endeavors aim to curate datasets encompassing diverse source types (e.g., natural earthquakes, blasting, rock rupture-induced earthquakes, etc.) and further refine network design. These efforts are geared towards advancing the intelligent evolution and application of seismic signal detection methodologies.

REFERENCES

[1] S. Lü, L. Fang, H. Ren, and B. Zhou, "Overview of the earthquake monitoring status in the United States," *China Earthq. Eng. J.*, vol. 46, no. 2, pp. 431–448, Feb. 2024.

[2] L. Li, J. Tan, D. A. Wood, Z. Zhao, D. Becker, Q. Lyu, B. Shu, and H. Chen, "A review of the current status of induced seismicity monitoring for hydraulic fracturing in unconventional tight oil and gas reservoirs," *Fuel*, vol. 242, pp. 195–210, Apr. 2019.

[3] V. Grechka and W. Heigl, *Microseismic Monitoring*. Tulsa, OK, USA: Society of Exploration Geophysicists, 2017, ch. 1, pp. 1–5.

[4] R. Allen, "Automatic phase pickers: Their present use and future prospects," *Bull. Seismological Soc. Amer.*, vol. 72, no. 6, pp. 225–242, Dec. 1982.

[5] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.

[6] R. J. Skoumal, M. R. Brudzinski, B. S. Currie, and J. Levy, "Optimizing multi-station earthquake template matching through re-examination of the Youngstown, Ohio, sequence," *Earth Planet. Sci. Lett.*, vol. 405, pp. 274–280, Nov. 2014.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[8] G. Cheng, X. Xie, J. Han, L. Guo, and G.-S. Xia, "Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3735–3756, 2020.

[9] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202, Apr. 1980.

[10] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent neural network regularization," 2014, *arXiv:1409.2329*.

[11] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial Nets," in *Proc. Adv. Neural Inf. Process. Syst.* Cambridge, MA, USA: MIT Press, 2014, pp. 1–9.

[12] T. Perol, M. Gharbi, and M. Denolle, "Convolutional neural network for earthquake detection and location," *Sci. Adv.*, vol. 4, no. 2, Feb. 2018, Art. no. e1700578.

[13] Z. E. Ross, M. C. White, F. L. Vernon, and Y. Ben-Zion, "An improved algorithm for real-time S-wave picking with application to the (augmented) ANZA network in Southern California," *Bull. Seismological Soc. Amer.*, vol. 106, no. 5, pp. 2013–2022, Oct. 2016.

[14] M. Zhao, S. Chen, and Y. Dave, "Waveform classification and seismic recognition by convolution neural network," *Chin. J. Geophys.*, vol. 62, no. 1, pp. 374–382, Jan. 2019.

[15] J. Zheng, J. Lu, S. Peng, and T. Jiang, "An automatic microseismic or acoustic emission arrival identification scheme with deep recurrent neural networks," *Geophys. J. Int.*, vol. 212, no. 2, pp. 1389–1397, Feb. 2018.

[16] Z. Li, M. Meier, E. Hauksson, Z. Zhan, and J. Andrews, "Machine learning seismic wave discrimination: Application to earthquake early warning," *Geophys. Res. Lett.*, vol. 45, no. 10, pp. 4773–4779, May 2018.

[17] Y. Bathaee, "The artificial intelligence black box and the failure of intent and causation," *Harvard J. Law Technol.*, vol. 31, p. 889, Sep. 2018.

[18] G. Zhang, C. Lin, and Y. Chen, "Convolutional neural networks for microseismic waveform classification and arrival picking," *Geophysics*, vol. 85, no. 4, pp. 227–240, Jul. 2020.

[19] R. M. H. Dokht, H. Kao, R. Visser, and B. Smith, "Seismic event and phase detection using time–frequency representation and convolutional neural networks," *Seismological Res. Lett.*, vol. 90, no. 2, pp. 481–490, Mar. 2019.

[20] X. Bi, C. Zhang, Y. He, X. Zhao, Y. Sun, and Y. Ma, "Explainable time–frequency convolutional neural network for microseismic waveform classification," *Inf. Sci.*, vol. 546, pp. 883–896, Feb. 2021.

[21] T. Li, Z. Zhao, C. Sun, L. Cheng, X. Chen, R. Yan, and R. X. Gao, "WaveletKernelNet: An interpretable deep neural network for industrial intelligent diagnosis," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 52, no. 4, pp. 2302–2312, Apr. 2022.

[22] L. Wei, Y. Lin, J. Wang, and Y. Ma, "Time-frequency convolutional neural network for automatic sleep stage classification based on single-channel EEG," in *Proc. IEEE 29th Int. Conf. Tools Artif. Intell. (ICTAI)*, Nov. 2017, pp. 88–95.

[23] Q. Chen, X. Dong, G. Tu, D. Wang, C. Cheng, B. Zhao, and Z. Peng, "TFN: An interpretable neural network with time-frequency transform embedded for intelligent fault diagnosis," *Mech. Syst. Signal Process.*, vol. 207, Jan. 2024, Art. no. 110952.

[24] Y. Yang, W. Zhang, Z. Peng, and G. Meng, "Multicomponent signal analysis based on polynomial chirplet transform," *IEEE Trans. Ind. Electron.*, vol. 60, no. 9, pp. 3948–3956, Sep. 2013.

[25] H. Lim, J. Park, and Y. Han, "Rare sound event detection using 1D convolutional recurrent neural networks," in *Proc. DCASE*, May 2017, pp. 80–84.

[26] Q. Liu, W. Sun, and G. Ru, "Short-wave time-varying channel blind equalization algorithm based on convolutional recurrent neural network," *J. Wuhan. Univ.*, vol. 67, no. 3, pp. 241–246, Jun. 2021.

[27] K. Zhang, Y. Cai, Y. Ren, R. Ye, and L. He, "MTF-CRNN: Multiscale time-frequency convolutional recurrent neural network for sound event detection," *IEEE Access*, vol. 8, pp. 147337–147348, 2020.

[28] J. Li, X. Wang, Y. Zhang, W. Wang, J. Shang, and L. Gai, "Research on the seismic phase picking method based on the deep convolution neural network," *Chin. J. Geophys.*, vol. 63, no. 4, pp. 1591–1606, Apr. 2020.

[29] J. Fu, X. Wang, Z. Li, Q. Tan, and J. Wang, "Automatic picking up earthquake's P waves using signal-to-noise ratio under a strong noise environment," *Chin. J. Geophys.*, vol. 62, no. 4, pp. 1405–1412, Apr. 2019.

[30] N. Ricker, "The form and laws of propagation of seismic wavelets," *Geophysics*, vol. 18, no. 1, pp. 10–40, Jan. 1953.

[31] S. Qian and D. Chen, "Joint time-frequency analysis," *IEEE Signal Process. Mag.*, vol. 16, no. 2, pp. 52–67, Mar. 1999.

[32] S. Marple, "Computing the discrete-time 'analytic' signal via FFT," in *Proc. Asilomar Conf. Signals*, 2002, pp. 1322–1325.

[33] Z. Cheng, W. Liao, X. Chen, and X. Lu, "A vibration recognition method based on deep learning and signal processing," *Eng. Mech.*, vol. 38, no. 4, pp. 230–246, 2021.

[34] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, Dec. 2014, pp. 1–15.

[35] T. Zhan, L. Li, and H. Chen, "Research on microseismic interferometric location method based on the instantaneous phase," *Chin. J. Geophys.*, vol. 65, no. 5, pp. 1753–1768, May 2022.