Article

A Strong Noise Reduction Network for Seismic Records

Tong Shen [1,*,†], Xuan Jiang [1,†], Wenzheng Rong [2,†], Lei Xu [1] , Xianguo Tuo [2] and Guili Peng [3]

1 School of Information Engineering, Southwest University of Science and Technology, Mianyang 621010, China; jx13106099316@163.com (X.J.); 18328320446@163.com (L.X.)
2 School of Automation and Information Engineering, Sichuan University of Science and Engineering, Yibin 644000, China; rwz0910@126.com (W.R.); tuoxg@cdut.edu.cn (X.T.)
3 School of Control and Mechanical Engineering, Tianjin Chengjian University, Tianjin 300384, China; planepeople678@sina.com
* Correspondence: sht@swust.edu.cn
† These authors contributed equally to this work.

Abstract: Noise reduction is a critical step in seismic data processing. A novel strong noise reduction network is proposed in this study. The network enhances the U-Net architecture with an improved inception module and coordinate attention (CA) mechanism, suppressing noise and enhancing signal clarity. These enhancements improve the network's capability to distinguish between signal and noise in the time–frequency domain. We trained and tested our model on the STEAD dataset, which eliminated noise across various frequency bands, improved the signal-to-noise ratio (SNR) of seismic records, and reduced the waveform distortion significantly. Comparative analyses against U-Net, DeepDenoiser, and DnRDB models, using signals with SNRs ranging from −14 dB to 0 dB, demonstrated our model's superior performance. At the same time, we demonstrated that the Inception Conv Block has a significant impact on the denoising ability of the network. Furthermore, validation using the "Di Ting" dataset and real noisy signals confirmed the model's generalizability. These results show that the proposed model significantly outperforms the comparative methods in terms of the SNR, correlation coefficient (r), and root mean square error (RMSE), delivering higher-quality seismograms. The enhanced phase-picking accuracy underscores the potential of our approach to advance in geophysics applications.

Keywords: denoise; low SNR; deep learning; U-Net; inception; channel attention; seismic records

1. Introduction

The quest for accurate seismic data is fundamental to the field of geophysics, underpinning a broad spectrum of applications from mineral exploration to the understanding of the Earth's tectonic activity [1,2]. However, seismic records are often plagued by noise that obscures signals of interest, thereby complicating subsequent analyses. These noises, including geological activities, human interference, and equipment failures, significantly degrade the quality of seismic data and the accuracy of interpretations [3]. Consequently, the noise reduction of seismic records is essential for the subsequent analysis and processing.

Traditional seismic data denoising methods predominantly utilize mathematical models and signal processing techniques, such as wavelet threshold denoising [4–9], Fourier transform [10–12], and empirical mode decomposition (EMD) [13–18]. Wavelet threshold denoising achieves denoising by processing noisy signals in the wavelet domain. Fourier transform improves the signal clarity and quality by filtering noise in the frequency domain. EMD can effectively handle nonlinear and non-stationary signals without the need for preset basis functions, and it can adaptively decompose different frequency components, making signal analysis more flexible and accurate. Despite their utility, these methods often yield suboptimal results and are time-consuming when applied to complex seismic data [19]. They are also susceptible to the subjective selection of feature functions and

threshold parameters, particularly when noise overwhelms the signal, resulting in a low SNR [20].

The advent of deep learning technology, with convolutional neural networks leading advancements in image and speech denoising, has demonstrated the power of feature learning and nonlinear mapping capabilities [21–27]. Although deep learning techniques require a large amount of data for models to be trained, they have significant advantages over traditional methods because the models are able to autonomously learn and differentiate the features of complex seismic signals and noise without the need for the process of human extraction of features [28–35]. Consequently, an increasing number of researchers are employing deep learning for seismic data denoising [36,37]. In 2018, Jin et al. addressed seismic data denoising using deep residual networks to learn intricate data and noise features, achieving more effective denoising [38]. Zhu et al.'s DeepDenoiser in 2019 employed an encoder–decoder structure to learn a set of sparse features, effectively separating signal from noise [39]. However, the model's encoder down-sampling, achieved through convolution, faced challenges in extracting sufficient seismic signal features. In 2021, Gao et al. introduced a seismic signal denoising model (DnRDB) that incorporated a residual dense block (RDB), deepening the model's structure with multiple RDB modules to extract more relevant feature information [20]. Novoselov et al.'s SEDENOSS method in 2022 utilized a two-path recurrent neural network for effective seismic signal separation and denoising, thereby enhancing the signal quality and accuracy [40]. In 2023, Cai et al. merged atrous convolution with RDB within the U-Net framework, creating the ARDU network, which improved the feature extraction capabilities without increasing the network parameters, reduced the waveform distortion, and preserved the effective signals [41]. Nonetheless, when noise and signal are in the same frequency band [42] or when excessive noise results in a signal-to-noise ratio below 0 dB, the denoising performance of these models is unsatisfactory. This is attributed to the excessively low signal-to-noise ratio and the overwhelming interference of noise, which hamper the model's signal feature extraction.

In conclusion, this paper introduces a model tailored for denoising seismic signals with extremely low signal-to-noise ratios. Building upon the U-Net [43], we have integrated an inception module and an attention mechanism, which collectively enhance the network's feature extraction in the time–frequency domain and its ability to differentiate between signals and noise. To assess our model's denoising efficacy, we utilized the STEAD dataset [44] for training and testing. The experimental outcomes confirm that our model can effectively eliminate diverse noise types and maintain the denoising performance even with signal-to-noise ratios below 0 dB. When compared to the U-Net, DeepDenoiser, and DnRDB networks, our model demonstrates superior denoising effects on seismic signals with very low signal-to-noise ratios and significantly improves the accuracy of the initial phase pickup in denoised signals. Lastly, the "Di Ting" dataset [45] was employed to validate our model's generalization capability.

Our main contributions are summarized as follows.

(1) An improved U-Net model is proposed for denoising noisy signals with low signal-to-noise ratios. This model effectively suppresses various types of noise, achieving successful denoising for seismic signals with signal-to-noise ratios below 0 dB. The denoised results significantly enhance the accuracy of the pickup of the phase arrival.

(2) Considering that seismic signals and noise may coexist across different frequency channels or within specific frequency bands, the CA mechanism is introduced during feature extraction. This mechanism directs the network's attention to key channels or frequency bands, enabling better differentiation between localized signal features and noise features. Consequently, the model more effectively removes noise that shares the same frequency band as the signal.

(3) To mitigate the information loss during down-sampling, we incorporate an inception module in the down-sampling stage. This module enables the network to capture features of different sizes from the seismic signal during feature extraction, ensuring the diversity of the features extracted by the network.

2. Methodology

Time–frequency domain representation is a way that concurrently analyzes signals in terms of both time and frequency, offering insights into signal variations across these dimensions. Consequently, converting seismic signals from the time domain to the time–frequency domain is advantageous for models to harness frequency information in learning the distinct features of seismic signals and noise. In this study, we employ the short-time Fourier transform (STFT) [46] for this conversion to the time–frequency domain, which is particularly suitable for analyzing non-stationary signals. The STFT divides the signal into shorter time periods and applies the Fourier transform (FT) to these time periods to obtain the time–frequency representation of the signal. This allows us to observe the spectral characteristics of the signal at different time periods. The mathematical expression for the STFT is as follows:

$$X(t, f) = \int_{-\infty}^{\infty} x(\tau)w(t - \tau)e^{-j2\pi f\tau}d\tau \tag{1}$$

where $X(t, f)$ is the result of the STFT, $x(\tau)$ is the input discrete signal, $w(t - \tau)$ is the window function, and f is the frequency.

In the time–frequency domain, the noisy seismic signal $Y(t, f)$ consists of the summation of the clean seismic signal $S(t, f)$ and the noise $N(t, f)$, expressed by the subsequent equation. This representation facilitates the separation and analysis of the signal and noise components within the seismic data.

$$Y(t, f) = S(t, f) + N(t, f) \tag{2}$$

The objective of denoising is to maximize the removal of noise $N(t, f)$ so that the predicted clean seismic signal $\widehat{S}(t, f)$ closely approximates $S(t, f)$. The equation to obtain $\widehat{S}(t, f)$ is as follows:

$$\hat{S}(t, f) = \widehat{M_S}(t, f)Y(t, f) \tag{3}$$

where $\widehat{M_S}(t, f)$ represents the time–frequency domain mask of the predicted signal. The network takes the real and imaginary components of the time–frequency coefficients of the noise-containing seismic signals as inputs and outputs the masks of the signal and noise in the time–frequency domain. These predicted time–frequency masks can more accurately estimate the distribution of seismic signals and noise across the frequency spectrum, thereby enhancing the network's ability to discern between signal and noise, resulting in more effective denoising. This study adopts a magnitude–time–frequency mask, influenced by those commonly employed in speech noise reduction algorithms [47,48]. The formulas for the signal's mask $M_S(t, f)$ and the noise's mask $M_N(t, f)$ are as follows:

$$M_S(t, f) = \frac{1}{1 + \frac{|N(t,f)|}{|S(t,f)|}} \tag{4}$$

$$M_N(t, f) = \frac{\frac{|N(t,f)|}{|S(t,f)|}}{1 + \frac{|N(t,f)|}{|S(t,f)|}} \tag{5}$$

The size of each mask is the same as the input time–frequency domain seismic signal $Y(t, f)$, where $M_S(t, f)$ represents the proportion of the clean seismic signal $S(t, f)$ in the noise-containing seismic signal $Y(t, f)$ in the range of 0–1 [49].

After obtaining the predicted clean signal $\widehat{S}(t, f)$ in the time–frequency domain, the predicted seismic signal $\widehat{S}(t)$ in the time domain can be obtained by performing a short-time Fourier inversion [46] on it with the following equation:

$$\widehat{S}(t) = STFT^{-1}\left\{\widehat{S}(t, f)\right\} \tag{6}$$

In this paper, we integrate an inception module and CA module into the U-Net, with the network structure depicted in Figure 1a. We use a convolutional layer with a 3×3 size

convolutional kernel with a step size of 2 and padding of 1 to supplant the maximum pooling layer for down-sampling, thereby mitigating the loss of high-frequency seismic information typically incurred during pooling. The reason for using a convolutional kernel of 3×3 is that it has better feature extraction capabilities and can extract finer-grained features [50]. The inception module introduced in this paper facilitates the encoder's ability to efficiently learn and represent the intricate features of the input data, enabling the network to capture features of different scales within the seismic signal during feature extraction. The configuration of the inception module is illustrated in Figure 1b. The CA mechanism [51] introduced in this paper is shown in Figure 1c. Seismic signals and noise may coexist across various frequency channels or within specific frequency bands. The CA mechanism directs the network to focus its attention on these pivotal frequency channels or bands, consequently amplifying the network's capacity to perceive seismic signals. With the integration of the CA mechanism, the network achieves a more refined distinction between seismic signals and noise and assigns higher weights to seismic signals, retaining the details and structure of the seismic signals more effectively throughout the denoising process and elevating the denoising performance.
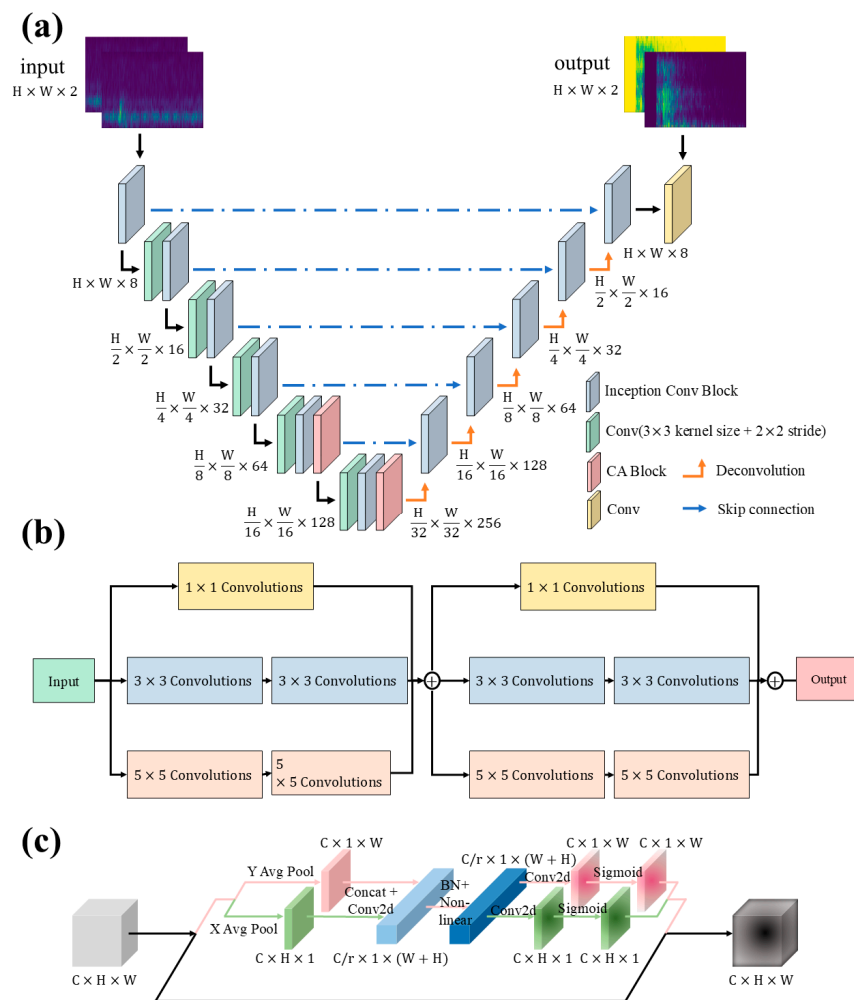

Figure 1. The detailed structure diagram of the proposed network. (**a**) The structure of our network. The inputs are the real and imaginary parts of the noise data after the STFT, and the outputs are the masks corresponding to the signal and noise. The different colored squares represent different modules in the network and the numbers represent the size of the feature map after the output of each layer of modules. (**b**) The inception module used in our network. (**c**) The CA module used in our network.

3. Results and Discussion

3.1. Dataset

The data for this study were sourced from the STEAD dataset, with a sampling rate of 100 Hz and a length of 60 s. The STEAD dataset contains over two million three-channel waveform recordings covering seismic activity and environmental noise on a global scale. We randomly selected 25,000 instances of high signal-to-noise ratio seismic signals and various types of noise signals from them. These selected signals and noise were normalized by subtracting the mean and dividing by the standard deviation. Subsequently, the noise was randomly scaled within a predefined range and combined with the seismic signals to generate 25,000 low SNR seismic signals, all with SNRs below 0 dB. These noisy signals were then randomly allocated to a training set, validation set, and test set in an 8:1:1 ratio. Figure 2 illustrates a set of clean signals (Figure 2a), noise (Figure 2b), and noisy signals (Figure 2d). Additionally, it presents the denoised signal (Figure 2e) and the noise (Figure 2f) output by the model. Each subfigure's left image depicts the time domain plot of the respective signals, while the right image shows the spectrograms produced by applying a short-time Fourier transform to the signals. Figure 2c shows the masks for the clean signal and the noise.
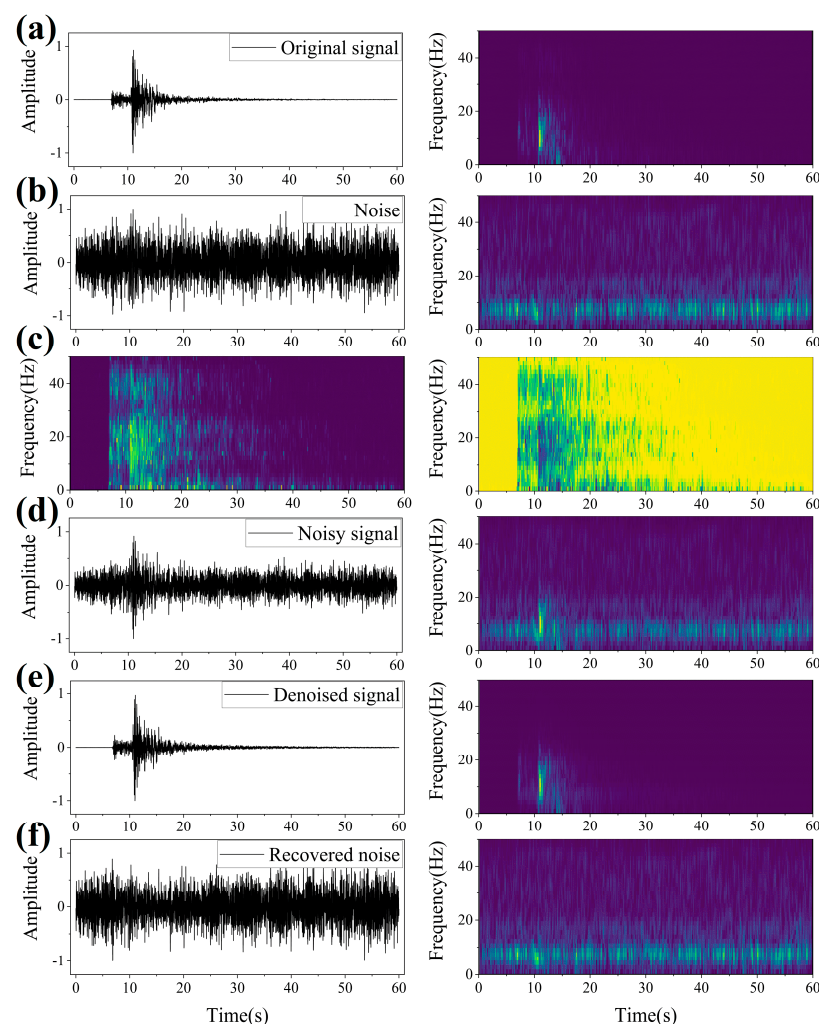

Figure 2. The left panels in (**a**,**b**,**d**–**f**) display the clean signal, noisy signal, denoised signal, and recovered noise, respectively, in the time domain. The right panels present the corresponding spectrograms. (**c**) The mask of the signals and noises depicted in (**a**,**b**).

3.2. Evaluation Index

In this paper, we employ three metrics to assess the denoising performance: signal-to-noise ratio (SNR), correlation coefficient (r), and root mean square error (RMSE) [41]. The definitions of these metrics are as follows.

(1) The SNR represents the ratio of clean signal to noise in decibels (dB). A higher SNR indicates a reduced noise component. The corresponding formula is delineated as follows:

$$\text{SNR} = 10\log_{10}\left(\frac{\sigma_s}{\sigma_n}\right) \tag{7}$$

where σ_s denotes the energy of the clean signal and σ_n denotes the energy of the noise.

(2) The r represents the correlation coefficient that measures the degree of similarity between the denoised and the reference signals. A value approaching one indicates a stronger correlation, suggesting that the denoised seismic signal closely resembles the waveform of the reference seismic signal. The corresponding formula is delineated as follows:

$$\text{r}(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}[X]\text{Var}[Y]}} \tag{8}$$

where X denotes the reference signal and Y denotes the denoised seismic signal predicted by the model. $\text{Conv}(X,Y)$ denotes the covariance of the reference signal and the denoised signal, $\text{Var}[X]$ denotes the variance of the reference signal, and $\text{Var}[Y]$ denotes the variance of the denoised signal.

(3) The RMSE represents the root mean square error at each sample point between the reference signal and the denoised signal predicted by the model. A lower RMSE value indicates the higher resemblance of the denoised seismic signal to the reference signal. The extent of the waveform distortion after signal denoising is characterized by the deviation of each sampling point. The corresponding formula is delineated as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left|X'(i) - X(i)\right|}{n}} \tag{9}$$

where $X'(i)$ denotes the denoised seismic signal output by the model prediction, $X(i)$ denotes the reference seismic signal, and n is the number of signal sampling points.

3.3. Model Training

In this experiment, the network was constructed utilizing the PyTorch (1.12.1) framework. The model underwent training, validation, and testing on an NVIDIA GEFORCE RTX3090 (Nvidia, Santa Clara, CA, USA). The training process was conducted with a learning rate of 0.0001 and a batch size of 32. Parameter optimization was achieved through the Adam optimization algorithm [52]. The mean square loss function (MSELoss) served as the training loss function, for which the formula is as follows. The model completed 100 epochs of training, resulting in the acquisition of the training and validation loss function descent curves, as depicted in Figure 3.

$$\text{MSELOSS} = \frac{1}{n}\sum_{i=1}^{n}\left(y_{i1} - y_{i2}\right)^2 \tag{10}$$
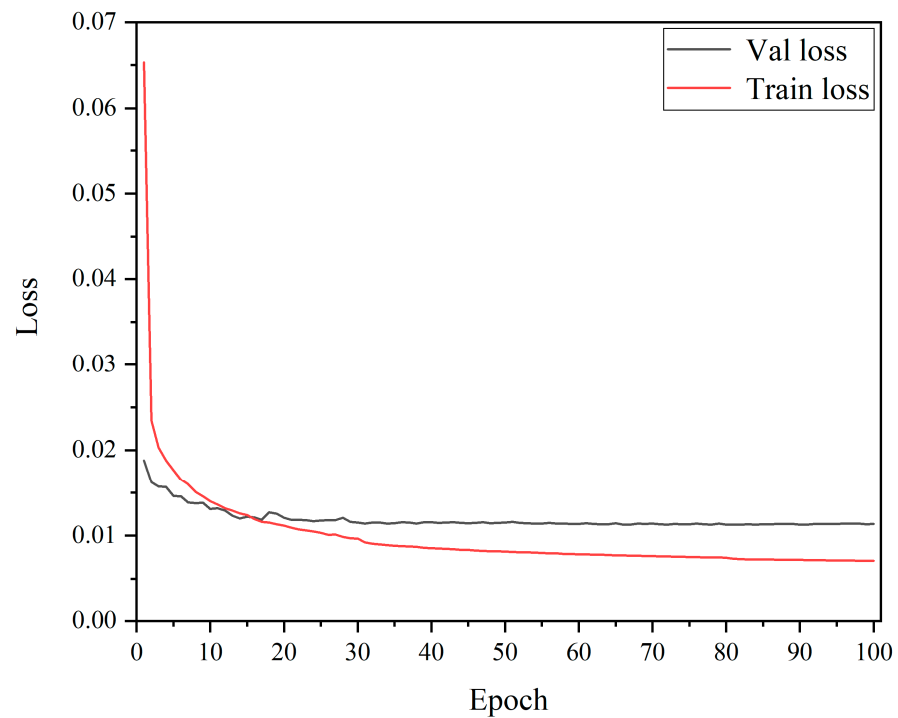
Figure 3. Training and validation loss during the training process.

3.4. Denoising Analysis of Different Types of Noise

The denoising test was conducted using our model on noisy seismic signals contaminated with low-frequency noise, high-frequency noise, mixed-band noise, and spike noise to ascertain the model's efficacy in noise abatement across various frequency bands. The outcomes are depicted in Figures 4–7. Figure 4 illustrates the impact of low-frequency noise, which induces significant fluctuations in the seismic signal. This noise is heavily superimposed on the seismic signal band, resulting in poor differentiation and complicating the denoising process. Our model successfully suppresses these fluctuations, enhancing the SNR from −9.46 dB to 12.85 dB, increasing the correlation coefficient by 96%, and reducing the RMSE by 80.19%. Figure 5 illustrates the impact of high-frequency noise, which has a higher frequency and minimal overlap with the seismic signal band. This noise detrimentally affects the event detection accuracy. Our model successfully removes it, improving the SNR from −6.87 dB to 21.13 dB, raising the correlation coefficient by 110%, and lowering the RMSE by 80.9%. Figure 6 illustrates the impact of mixed-frequency noise, which comprises multiple frequency components that severely overlap with the seismic signal band, posing noise reduction challenges. After the denoising process, the SNR is improved from −8.65 dB to 10.25 dB, the correlation coefficient is enhanced by 66.66%, and the RMSE is decreased by 69.71%. Figure 7 illustrates the impact of spike noise, which impacts the seismic phase and arrival time extraction. Following denoising, the SNR is boosted from −8.5 dB to 14.5 dB, the correlation coefficient is increased by 108.5%, and the RMSE is reduced by 70.66%. In summary, our model effectively eliminates various noise types, augments the SNR, preserves the effective signal, and minimizes the waveform distortion.
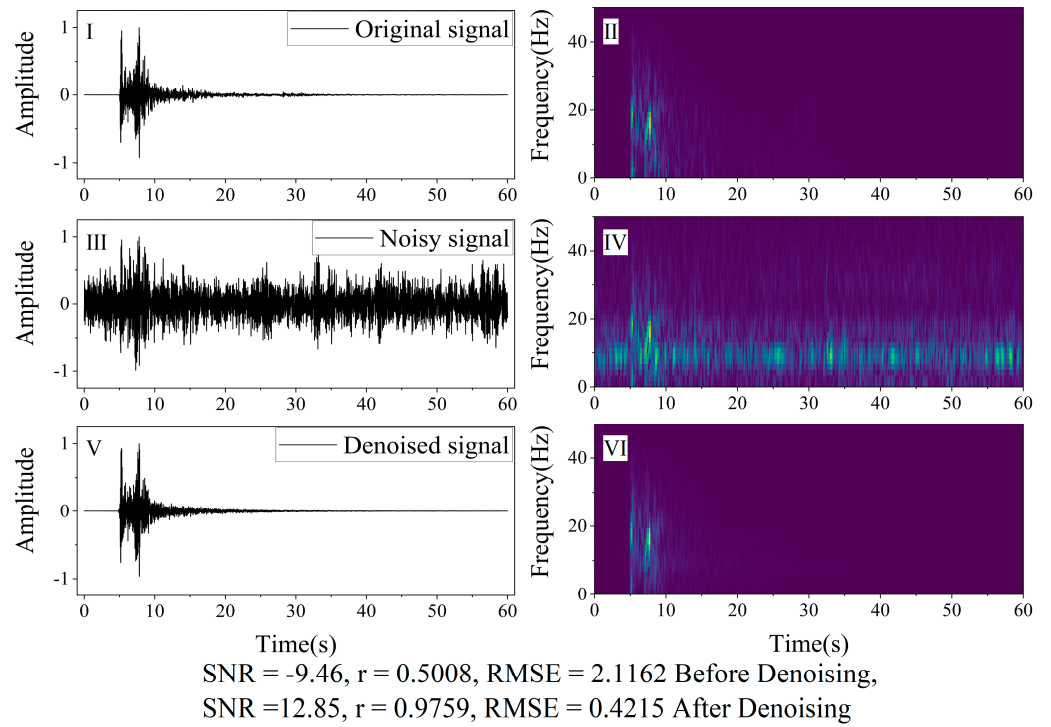
SNR = -9.46, r = 0.5008, RMSE = 2.1162 Before Denoising,
SNR =12.85, r = 0.9759, RMSE = 0.4215 After Denoising

Figure 4. Denoising performance of low-frequency noise: (**I,III,V**) are the original signal, the noise signal, and the denoised signal in the time domain; and (**II,IV,VI**) are the time–frequency domain data of (**I,III,V**).

SNR = -6.87, r = 0.4690, RMSE = 1.9999 Before Denoising,
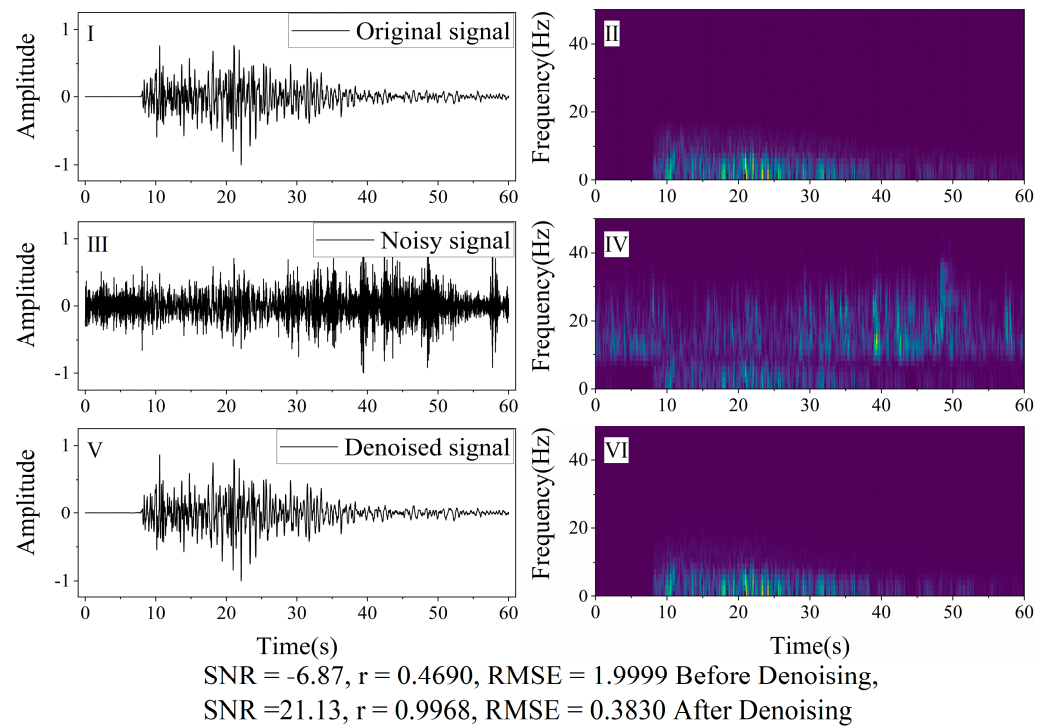SNR =21.13, r = 0.9968, RMSE = 0.3830 After Denoising

Figure 5. Denoising performance of high-frequency noise: (**I,III,V**) are the original signal, the noise signal, and the denoised signal in the time domain; and (**II,IV,VI**) are the time–frequency domain data of (**I,III,V**).

SNR = -8.65, r = 0.5684, RMSE = 2.0824 Before Denoising,
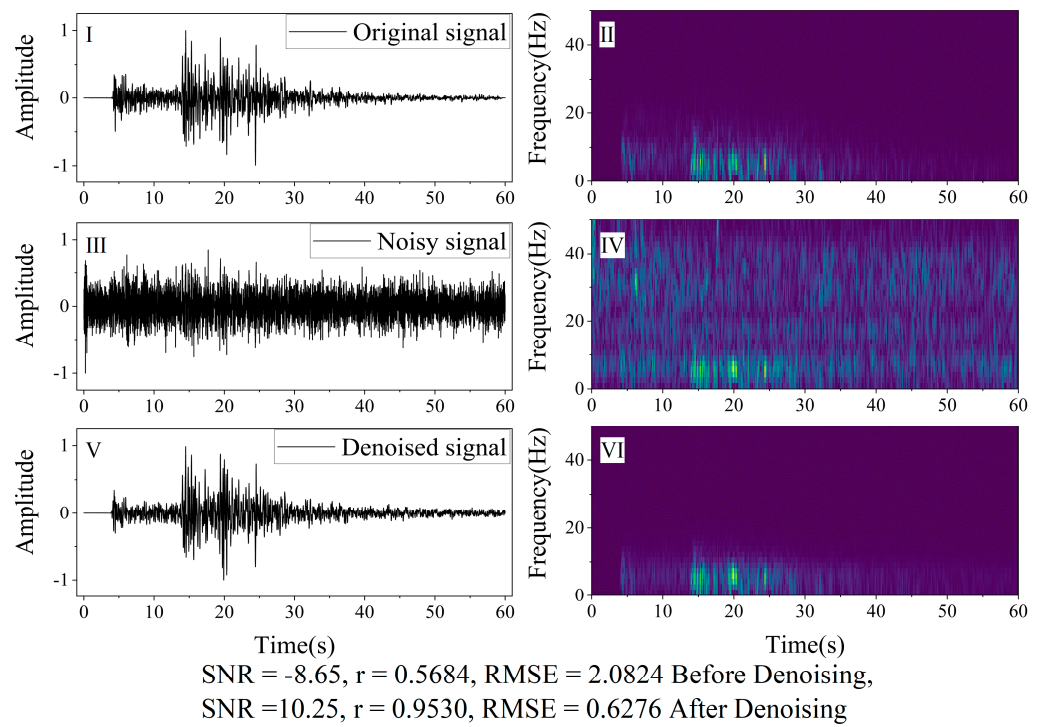SNR =10.25, r = 0.9530, RMSE = 0.6276 After Denoising

Figure 6. Denoising performance of mixed-frequency noise: (**I,III,V**) are the original signal, the noise signal, and the denoised signal in the time domain; and (**II,IV,VI**) are the time–frequency domain data of (**I,III,V**).

SNR = -8.50, r = 0.4740, RMSE = 1.6699 Before Denoising,
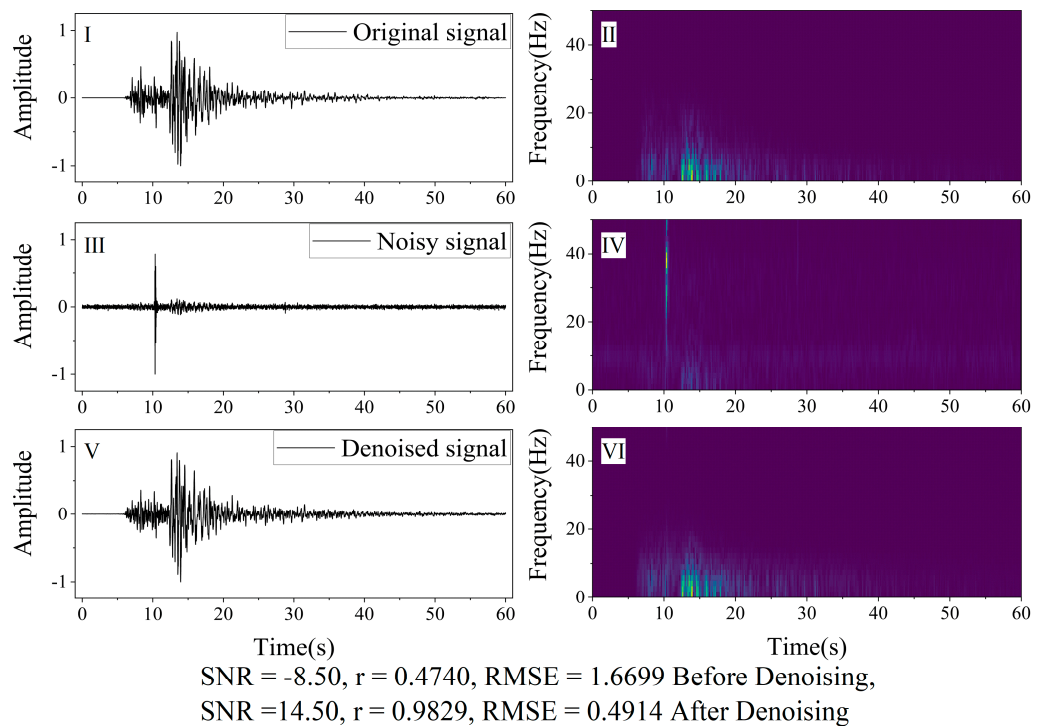SNR =14.50, r = 0.9829, RMSE = 0.4914 After Denoising

Figure 7. Denoising performance of spike noise: (**I,III,V**) are the original signal, the noise signal, and the denoised signal in the time domain; and (**II,IV,VI**) are the time–frequency domain data of (**I,III,V**).

3.5. Comparative Experiments and Analysis

To demonstrate the denoising efficacy of our model, we conducted comparisons with the U-Net, DeepDenoiser, and DnRDB networks. The experimental outcomes were

assessed using the SNR, correlation coefficient, RMSE and arrival error. All the methods utilized data from an identical test set, with the SNR ranging from −14 dB to 0 dB. The results, depicted in Figure 8, reveal that our model consistently achieves higher SNRs and correlation coefficients, alongside lower RMSEs and arrival errors across various noise levels, when contrasted with the U-Net, DeepDenoiser, and DnRDB models. This suggests a superior denoising performance with minimal waveform distortion by our model. Relative to the average SNR, as shown in Table 1, our model exhibits an improvement in the SNR by 1.74 dB, 1.18 dB, and 3.66 dB over the U-Net, DeepDenoiser, and DnRDB models, respectively. Additionally, the average correlation coefficients are enhanced by 0.0050, 0.0044, and 0.0166, while the RMSE values are reduced by 0.0491, 0.0365, and 0.1263, correspondingly. By adding the Inception module to the U-Net, the SNR and r of the denoised signal can be improved more and the RMSE can be reduced more, which proves that adding the Inception module proposed in this paper to the network can improve the ability of the network to distinguish between the signal and the noise and strengthen the denoising performance of the network.
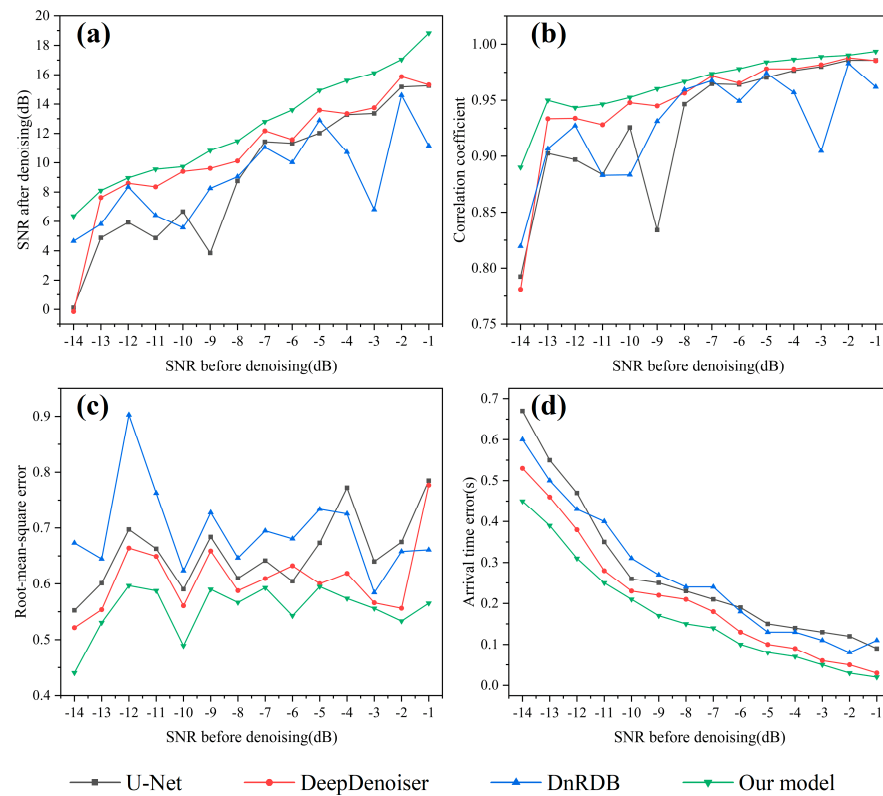

Figure 8. Comparison of the denoising effects of different networks: (**a**) improvement of the SNR; (**b**) correlation coefficient; (**c**) root mean square error; and (**d**) arrival time error.

Table 1. Denoising results of the different networks.

Model	Improvement of SNR	Improvement of r	Reduction of RMSE
U-Net	16.94	0.3096	1.1881
DeepDenoiser	17.49	0.3102	1.2007
DnRDB	15.01	0.2980	1.1109
U-Net + Inception	18.39	0.3140	1.2267
Our model	18.67	0.3146	1.2372

Figure 9 illustrates the waveforms after noise reduction using various models when the SNR of the added noise signal is −10 dB. It can be seen that although U-Net, DnRDB, and DeepDenoiser have removed most of the noise, some of the noise has not been removed

before the signal P-wave arrives at the time point, which means the P-wave first arrivals pickup of the denoised signals has errors. Our model denoises the signal best and can effectively remove the noise before the signal P-wave arrives, which means the P-wave first arrivals of the signal denoised by our model have errors. Our model has the best denoising effect, which can effectively remove the noise before the arrival of the P-wave of the signal, which makes the P-wave first arrivals of the denoised signal from our model more accurate, which is helpful for the subsequent seismic positioning.
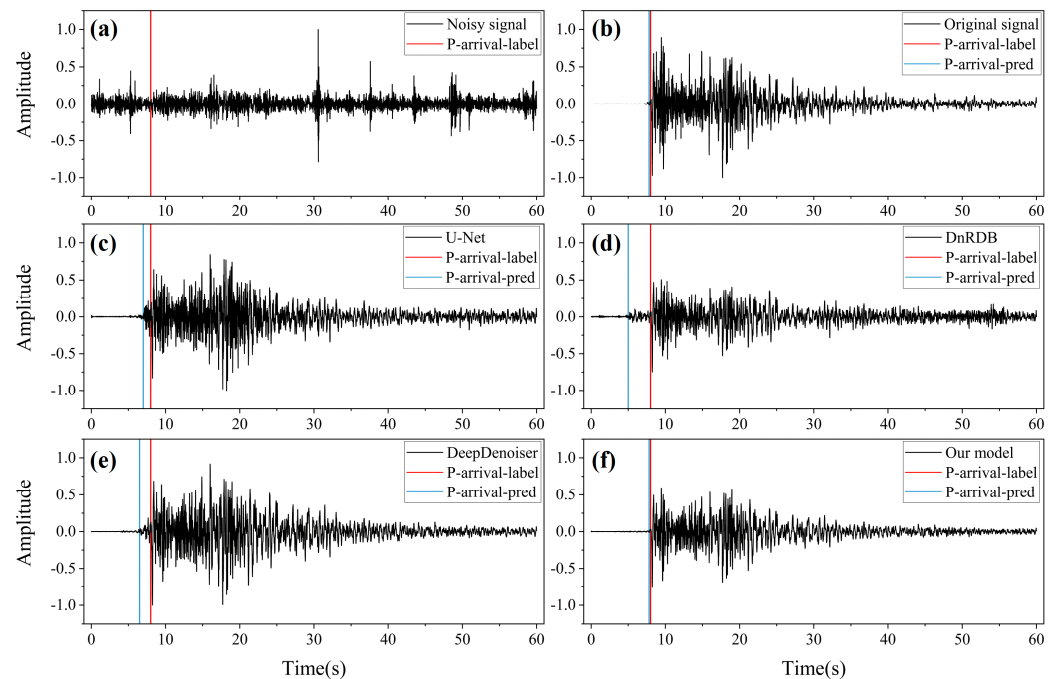

Figure 9. Waveform comparison of the noise reduction results of various methods for an SNR of 10 dB: (**a**) noisy signal; (**b**) original signal; (**c**) U-Net denoising result; (**d**) DnRDB denoising result; (**e**) DeepDenoiser denoising result; and (**f**) our model's denoising result. The red line represents the true first arrival time of the P-wave, whereas the blue line represents the result of picking up the first arrival of the denoised signals from each network using the STA/LTA algorithm [53].

3.6. Generalization

To validate the generalizability of our model, we applied the trained model to the "Di Ting" artificial intelligence seismology training dataset, which was furnished by the National Earthquake Science Data Center, with a sampling rate of 50 Hz and a length of 9000 sample points. For compatibility with our model's input length, we designated waveforms from the initial 6000 points as signal samples and the subsequent 3000 points as noise samples. Following these procedures, we yielded 1500 data points with an SNR below 0 dB. After normalization, we utilized DeepDenoiser, DnRDB, and our model for denoising. Table 2 presents the denoising outcomes for each model, including the SNR and correlation coefficient improvement of the signal before and after denoising and the RMSE reduction. As indicated in Table 2, our model outperforms DeepDenoiser and DnRDB, elevating the SNR by 3.72 dB and 2.63 dB, respectively, augmenting the average correlation coefficients by 0.1492 and 0.0272, and diminishing the RMSE by 0.1346 and 0.1587, respectively. These results underscore our network's significant advancements over other methods and its robust denoising proficiency, thereby corroborating its exceptional generalizability.

In addition, to verify the denoising performance of the models on real noise-containing signals, we selected 771 signals with SNRs below 5 dB from the STEAD dataset. However, since the corresponding clean signals for these noise-containing signals are unknown, we could not evaluate the denoising effect using the SNR, correlation coefficient (r), or RMSE. Because our goal is to improve the subsequent P-wave picking accuracy, we assessed

the denoising performance by comparing the P-wave picking accuracy before and after denoising, as shown in Table 3. Here, the STA/LTA algorithm is used for P-wave picking. The results show that the P-wave picking accuracy of the denoised signals obtained by U-Net, DeepDenoiser, and DnRDB did not improve, probably because these models remove the P-wave first-arrival information, as shown in Figure 10. The P-wave picking accuracy of the signal after the denoising of our model is improved by 6.75%, and there is no missed picking. This experiment demonstrates that our model effectively denoises real noise-containing signals and enhances the P-wave picking accuracy after denoising.
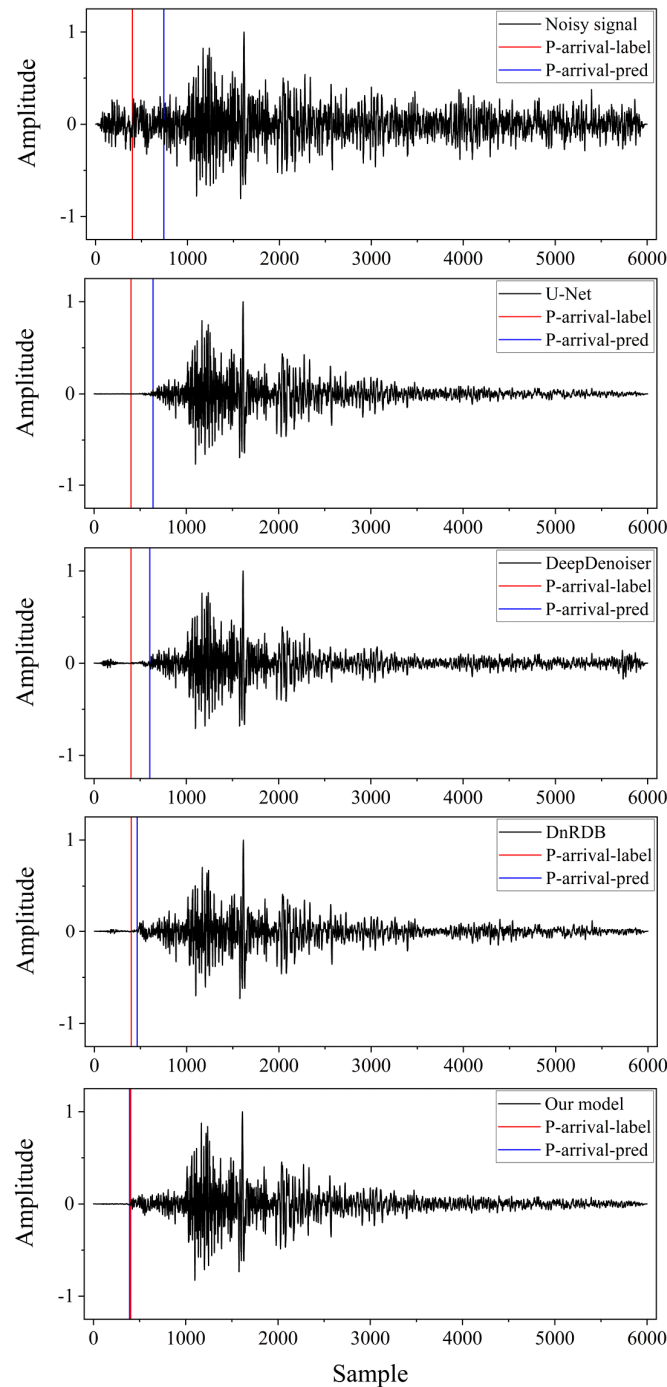
Figure 10. P-wave picking results of various models after denoising real noisy signals.

Table 2. Generalization experiment.

Model	Improvement of SNR	Improvement of r	Reduction of RMSE
DeepDenoiser	10.36	0.0884	1.0484
DnRDB	11.45	0.2104	1.0243
Our model	14.08	0.2376	1.1830

Table 3. Real signal denoising experiment.

Model	Accuracy (%)	Missed Rate (%)
Not denoised	63.42	1.82
U-Net	62.16	0
DeepDenoiser	62.65	0
DnRDB	62.52	0
Our model	70.17	0

4. Conclusions

This paper addresses the suboptimal denoising performance of seismic signal processing techniques on seismic signals with very low SNRs by introducing a novel denoising model. Our model significantly enhances the feature extraction and the ability to distinguish between signal and noise through the integration of Inception modules and a coordinate attention (CA) mechanism within the U-Net architecture. Training and testing using the STEAD dataset have demonstrated the model's superior denoising capabilities across various seismic noise conditions. Compared to conventional methods and other deep learning models, our model achieves a notable increase in the mean SNR and correlation coefficient, while also reducing the RMSE.

Furthermore, the application of our model to real seismic data has effectively improved the precision of identifying the first arrival of seismic phases. Additional testing on the "Di Ting" dataset further confirms the model's robust generalization capabilities. This research not only advances the adaptability and accuracy of earthquake early warning systems but also enhances their practical value in complex real-world settings. We believe that these outcomes will have a profound impact on the field of seismology and provide new directions for the development of future seismic signal processing technologies. In the future, we will continue to deeply analyze the characteristics of various types of noise in seismic data so that our model can remove other types of noise and improve its generalization ability.

Author Contributions: All the authors made significant contributions to this work. Formal analysis, T.S., W.R., X.T. and G.P.; investigation, T.S., X.J., W.R. and X.T.; methodology, T.S., X.J. and W.R.; prepared the training dataset, X.J. and L.X.; writing original draft, T.S., X.J., W.R. and L.X.; editing, T.S., X.J. and W.R. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available at [32,33].

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Chen, G.; Yang, W.; Liu, Y.; Wang, H.; Huang, X. Salt Structure Elastic Full Waveform Inversion Based on the Multiscale Signed Envelope. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4508912. [CrossRef]
2. Chen, G.; Yang, W.; Wang, H.; Zhou, H.; Huang, X. Elastic Full Waveform Inversion Based on Full-Band Seismic Data Reconstructed by Dual Deconvolution. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 8028205. [CrossRef]

3. Zhang, Y.; Lin, H.; Li, Y.; Ma, H. A patch based denoising method using deep convolutional neural network for seismic image. *IEEE Access* **2019**, *7*, 156883–156894. [CrossRef]

4. Cao, S.; Chen, X. The second-generation wavelet transform and its application in denoising of seismic data. *Appl. Geophys.* **2005**, *2*, 70–74. [CrossRef]

5. Gaci, S. The use of wavelet-based denoising techniques to enhance the first-arrival picking on seismic traces. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4558–4563. [CrossRef]

6. Liu, W.; Cao, S.; Chen, Y. Seismic Time–Frequency Analysis via Empirical Wavelet Transform. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 28–32. [CrossRef]

7. Mousavi, S.M.; Langston, C.A. Hybrid Seismic Denoising Using Higher-Order Statistics and Improved Wavelet Block Thresholding. *Bull. Seismol. Soc. Am.* **2016**, *106*, 1380–1393. [CrossRef]

8. Mousavi, S.M.; Langston, C.A.; Horton, S.P. Automatic microseismic denoising and onset detection using the synchrosqueezed continuous wavelet transform. *Geophysics* **2016**, *81*, V341–V355. [CrossRef]

9. Mousavi, S.M.; Langston, C.A. Automatic noise-removal/signal-removal based on general cross-validation thresholding in synchrosqueezed domain and its application on earthquake data. *Geophysics* **2017**, *82*, V211–V227. [CrossRef]

10. Sacchi, M.D.; Ulrych, T.J.; Walker, C.J. Interpolation and extrapolation using a high-resolution discrete Fourier transform. *IEEE Trans. Signal Process.* **1998**, *46*, 31–38. [CrossRef]

11. Mousavi, S.M.; Langston, C.A. Adaptive noise estimation and suppression for improving microseismic event detection. *J. Appl. Geophys.* **2016**, *132*, 116–124. [CrossRef]

12. Zhai, M.Y. Seismic data denoising based on the fractional Fourier transformation. *J. Appl. Geophys.* **2014**, *109*, 62–70. [CrossRef]

13. Liu, Y.; Li, Y.; Lin, H.; Ma, H. An Amplitude-Preserved TimeFrequency Peak Filtering Based on Empirical Mode Decomposition for Seismic Random Noise Reduction. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 896–900. [CrossRef]

14. Chen, Y.; Ma, J. Random noise attenuation by f-x empirical mode decomposition predictive filtering. *Geophysics* **2014**, *79*, V81V91. [CrossRef]

15. Bekara, M.; Van der Baan, M. Random and coherent noise attenuation by empirical mode decomposition. In *Seg Technical Program Expanded Abstracts*; Society of Exploration Geophysicists: Houston, TX, USA, 2008; pp. 2591–2595.

16. Chen, W.; Xie, J.; Zu, S.; Gan, S.; Chen, Y. Multiple-Reflection Noise Attenuation Using Adaptive Randomized-Order Empirical Mode Decomposition. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 18–22. [CrossRef]

17. Han, J.; Van der Baan, M. Microseismic and seismic denoising via ensemble empirical mode decomposition and adaptive thresholding. *Geophysics* **2015**, *80*, KS69–KS80. [CrossRef]

18. Roth, H.R.; Yao, J.; Lu, L.; Stieger, J.; Burns, J.E.; Summers, R.M. Detection of Sclerotic Spine Metastases via Random Aggregation of Deep Convolutional Neural Network Classifications. *arXiv* **2015**, arXiv:1407.5976.

19. Feng, J.; Li, X.; Liu, X.; Chen, C. Multigranularity Feature Fusion Convolutional Neural Network for Seismic Data Denoising. *IEEE Trans. Geosci. Remote Sens.* **2022**, *6*, 05907911. [CrossRef]

20. Gao, Z.; Zhang, S.; Cai, J.; Hong, L.; Zheng, J. Research on deep convolutional neural network time-frequency domain seismic signal denoising combined with residual dense blocks. *Front. Earth Sci.* **2021**, *9*, 681869. [CrossRef]

21. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]

22. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.

23. Mehta, S.; Rastegari, M.; Caspi, A.; Shapiro, L.; Hajishirzi, H. ESPNet: Efficient Spatial Pyramid of Dilated Convolutions for Semantic Segmentation. In Proceedings of the European Conference on Computer Vision, Cham, Switzerland, 8–14 September 2018; pp. 561–580.

24. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. Proceedings of The Thirty-First AAAI Conference on Artificial Intelligence (AAAI), Phoenix, AZ, USA, 12–17 February 2016; pp. 4–16.

25. Huang, G.; Liu, Z.; Geoff, P.; Der Maaten, V.; Kilian, W. Convolutional Networks with Dense Connectivity. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 8704–8716. [CrossRef] [PubMed]

26. Hu, J.; Shen, L.; Samuel, A.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *42*, 7132–7141.

27. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 3431–3440.

28. DeVries, P.M.; Viégas, F.; Wattenberg, M.; Meade, B.J. Deep learning of aftershock patterns following large earthquakes. *Nature* **2018**, *560*, 632–634. [CrossRef] [PubMed]

29. Perol, T.; Gharbi, M.; Denolle, M. Convolutional neural network for earthquake detection and location. *Sci. Adv.* **2018**, *4*, e1700578. [CrossRef]

30. Mousavi, M.; Zhu, W.; Sheng, Y.; Beroza, G.C. CRED: A Deep Residual Network of Convolutional and Recurrent Units for Earthquake Signal Detection. *Sci. Rep.* **2018**, *9*, 10267. [CrossRef]

31. Ross, Z.E.; Meier, M.; Hauksson, E. *P* Wave Arrival Picking and First-Motion Polarity Determination with Deep Learning. *J. Geophys. Res. Solid. Earth* **2018**, *123*, 5120–5129. [CrossRef]

32. Ross, Z.E.; Meier, M.; Hauksson, E.; Heaton, T.H. Generalized Seismic Phase Detection with Deep Learning. *Bull. Seismol. Soc. Am.* **2018**, *108*, 2894–2901. [CrossRef]

33. Ross, Z.E.; Yue, Y.; Meier, M.; Hauksson, E.; Heaton, T.H. PhaseLink: A Deep Learning Approach to Seismic Phase Association. *J. Geophys. Res. Solid Earth* **2019**, *124*, 856–869. [CrossRef]

34. Zheng, J.; Jiren, L.; Peng, S.; Jiang, T. An automatic microseismic or acoustic emission arrival identification scheme with deep recurrent neural networks. *Geophys. J. Int.* **2018**, *212*, 1389–1397. [CrossRef]

35. Zhu, W.; Beroza, G.C. PhaseNet: A Deep-Neural-Network-Based Seismic Arrival Time Picking Method. *Geophys. J. Int.* **2018**, *216*, 261–273. [CrossRef]

36. Yu, S.; Ma, J.; Wang, W. Deep Learning Tutorial for Denoising. *Geophysics* **2018**, *84*, V333–V350. [CrossRef]

37. Dong, X.; Zhong, T.; Li, Y. A deep-learning-based denoising method for multiarea surface seismic data. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 925–929. [CrossRef]

38. Jin, Y.; Wu, X.; Chen, J.; Han, Z.; Hu, W. Seismic data denoising by deep-residual networks. In *SEG Technical Program. Expanded Abstracts*; Society of Exploration Geophysicists: Houston, TX, USA, 2018; pp. 4593–4597.

39. Zhu, W.; Mousavi, S.M.; Beroza, G.C. Seismic signal denoising and decomposition using deep neural networks. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9476–9488. [CrossRef]

40. Novoselov, A.; Balazs, P.; Bokelmann, G. SEDENOSS: SEparating and DENOising Seismic Signals with Dual-Path Recurrent Neural Network Architecture. *J. Geophys. Res. Solid Earth* **2022**, *127*, e2021JB023183. [CrossRef]

41. Cai, J.; Wang, L.; Zheng, J.; Duan, Z.; Li, L.; Chen, N. Denoising method for seismic co-band noise based on a U-Net network combined with a residual dense block. *Appl. Sci.* **2023**, *13*, 1324. [CrossRef]

42. Gao, H. Study of Seismic Data Residual Statics and AI Denoising. Ph D. Thesis, University of Science and Technology of China, Hefei, China, 2018.

43. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv* **2015**, arXiv:1505.04597.

44. Mousavi, S.M.; Sheng, Y.; Zhu, W.; Beroza, G.C. STanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI. *IEEE Access* **2019**, *7*, 179464–179476. [CrossRef]

45. Zhao, M.; Xiao, Z.; Chen, S.; Fang, L. DiTing: A large-scale Chinese seismic benchmark dataset for artificial intelligence in seismology. *Earthq. Sci.* **2023**, *36*, 84–94. [CrossRef]

46. Griffin, D.; Lim, J. Signal estimation from modified short-time Fourier transform. *IEEE Trans. Acoust. Speech Signal Process.* **1984**, *32*, 236–243. [CrossRef]

47. Huang, P.S.; Kim, M.; Hasegawa-Johnson, M.; Smaragdis, P. Deep learning for monaural speech separation. In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014; pp. 1562–1566.

48. Xu, Y.; Du, J.; Dai, L.R.; Lee, C.H. An experimental study on speech enhancement based on deep neural networks. *IEEE Signal Process. Lett.* **2013**, *21*, 65–68. [CrossRef]

49. Donoho, D.L.; Johnstone, I.M. Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **1994**, *81*, 425–455. [CrossRef]

50. Karen, S.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.

51. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13713–13722.

52. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.

53. Allen, R.V. Automatic earthquake recognition and timing from single traces. *Bull. Eismol. Soc. Am.* **1978**, *68*, 1521–1532. [CrossRef]