



# Attention mechanism-based deep denoiser for desert seismic random noise suppression

Hongbo Lin<sup>1</sup> · Chang Liu<sup>1</sup> · Shigang Wang<sup>1</sup> · Wenhai Ye<sup>2</sup>

Received: 26 June 2022 / Accepted: 20 January 2023 / Published online: 30 March 2023

© The Author(s) under exclusive licence to Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2023

## Abstract

Seismic data collected from desert areas contain a large amount of low-frequency random noise with similar waveforms to the effective signals. The complex noise characteristics make it difficult to effectively identify and recover seismic signals, which will adversely affect subsequent seismic data processing and imaging. In order to recover the complex seismic events from low-frequency random noise, we propose an attention mechanism guided deep convolutional autoencoder network (ADCAE) to assign different importance to different features at different spatial position. In ADCAE, an attention module (AM) is connected to the deep convolutional autoencoder network (DCAE) with soft-thresholded symmetric skip connection that helps to enhance the ability of feature extraction. By combining the global features of the input data and the output local features of DCAE, AM generates an attention weight matrix, which assigns different weights to the features associated with the seismic events and random noise during the training process. In this way, AM can guide the update of the target gradient, thus retains the complex structure of the seismic events in the denoised results and improves the training efficiency of the model. The ADCAE is applied to the synthetic data and field seismic data, and denoised results show that ADCAE has achieved satisfactory denoising performance in signals recovery and low-frequency random noise suppression at the low signal-to-noise ratio.

**Keywords** Seismic exploration · Random noise suppression · Deep convolutional autoencoder network · Attention mechanism · Low signal-to-noise ratio

## Introduction

Seismic exploration is the main method for exploring oil and gas resources in the desert area. Affected by the loose sand layer and variable sand dunes, random noise in seismic data usually has slow-changing waveforms, aliased frequency band with the effective seismic signals, and the similar spatial structure in local areas, which severely hinders the

identification and recovery of weak seismic signals (Zhang et al. 2020), thereby giving the negative effects for the further seismic signal processing and imaging. Therefore, it is an important part of seismic exploration to effectively suppress random noise to increase signal-to-noise ratio (SNR) of seismic data.

Geophysicists and scholars conducted in-depth research on the problem of low-frequency seismic random noise suppression and proposed many effective noise suppression methods. Because of the similarity between low-frequency random noise and the effective seismic signals, it is difficult to effectively suppress low-frequency noise in time domain. Ma et al. (2019a) designed an adaptive bandpass filter according to the spectrum peak of random noise to recover the effective signals from seismic data. Zhang et al. (2020) proposed a structural-adaptive nonlinear complex diffusion method to achieve the object of suppressing seismic random noise collected from different environments along with preserving seismic structures by enjoying the advantages

---

Edited by Prof. Sanyi Yuan (ASSOCIATE EDITOR) / Prof. Gabriela Fernández Viejo (CO-EDITOR-IN-CHIEF).

✉ Hongbo Lin  
hb.lin@jlu.edu.cn

✉ Chang Liu  
1784876148@qq.com

<sup>1</sup> College of Communication and Engineering, Jilin University, Changchun 130012, China

<sup>2</sup> Jilin Province Kewei Traffic Engineering Co., Ltd, Changchun 130000, China

of enhancing-denoising performance of the shock filter and anisotropic behavior of diffusion coefficients.

In addition to these conventional denoising methods, machine learning-based denoising methods are gradually applied to suppress seismic random noise. Linear discriminant analysis based supervised learning is used to project seismic data into low dimensional space where the seismic signals and random noise overlapped in the space–time domain are separated easier, and then the traditional noise reduction method is combined to recover the seismic signal (Ma et al. 2019c). To deal with under-sampled seismic data, Zhao et al. (2020b) proposed a robust data-driven tight frame (DDTF), to transform the nonlinear Huber misfit into a linear operator, which mitigates the influence of erratic noise during the dictionary learning and reconstruction. Sun and Li (2020) proposed an approach that transform the real desert seismic data into time–frequency domain by synchrosqueezing transform, and use classification techniques based on supervised machine learning to identify the coefficients associated with signal and noise.

With the development of computer hardware and the improvement of deep learning theory and technology, autoencoder networks and convolutional neural networks are current mainstream deep learning algorithms. The autoencoder is a kind of deep learning algorithm, which comes from sparse coding and has huge advantages in the feature extraction and data compression (Rumelhart et al. 1986). The deep autoencoder consists of encoder and decoder. Through the backward propagation algorithm, the well trained autoencoder can extract rich and abstract characteristics from the data. Based on the structure of the autoencoder, the denoising autoencoder is proposed by training network through noisy and clean samples, which can extract more powerful features and is robust to the damaged data (Vincent et al. 2008). Ma et al. (2019b) proposed a denoising encoder-decoder networks for seismic random noise suppression. Compared with traditional denoising methods, the denoising effect at extremely low SNR was significantly improved. Considering the lack of high-quality training data set, Feng et al. (2020) proposed a random noise simulation model to expand the data sets, and Sang et al. (2021) generated diverse data through spatial and temporal directions, thus improving the denoising effect of autoencoder. Unsupervised learning-based autoencoders were developed by leveraging patching technique and deep image prior for seismic noise suppression (Saad and Chen 2021; Saad et al. 2021; Yang et al. 2022). The denoising ability and generalization are improved without need labeled data. Some works modified the original autoencoder network by first training the decoder network and then implementing the well-trained decoder as a constraint to train the encoder network, and applied the modified strategy to seismic inversion and porosity prediction, such as porosity prediction using semi-supervised

learning with biased well log data for improving estimation accuracy and reducing prediction uncertainty (Sang et al. 2023) and Double-scale supervised inversion with a data-driven forward model for low-frequency impedance recovery (Yuan et al. 2021).

In contrast, the most important advantage of the convolutional neural network (CNN) model is that it can extract rich information from training set (Krizhevsky et al. 2012), and use fewer parameters to maintain the spatial characteristics of data. CNN has achieved remarkable success in two-dimensional data processing. Recently, many noise reduction methods based on convolutional neural networks were developed for low-frequency seismic random noise. Zhao et al. (2019) improved the denoising convolution neural network (DnCNN) to suppress seismic random noise in desert seismic data in view of the serious aliasing between effective signal and desert noise in the frequency domain. Lin et al. (2021) proposed a branch construction-based denoising network based on CNN. The global features of the early seismic data extracted by the branch network guide the subsequent main network to suppress nonstationary random noise. Due to the complex features of the desert random noise, it is difficult to use limited synthetic data to discover the weak differences between seismic signal features and noise features, which lead to incomplete noise suppression. The combination of convolutional neural networks and traditional noise reduction algorithms can improve the suppression effect of complex random noise (Zhao et al. 2020a; Wang et al. 2020).

Compared with the full connection method of the autoencoder network, the convolutional neural network directly scans the input two-dimensional data through the convolution kernel rather than converting the two-dimensional matrix to one-dimensional vector, which ensures the preservation of the spatial characteristics and fewer parameters. The deep convolutional autoencoder (DCAE) uses convolution operation instead of the matrix internal accumulation in the autoencoder network to improve the feature extraction ability to the two-dimensional data (Masci et al. 2011). In recent years, the deep convolutional autoencoder networks with various structures have been proposed. Mao et al. (2016) proposed a skip convolutional autoencoder, and realized the purpose of image detail feature extraction. Multiscale feature clustering is further combined with the fully convolutional autoencoder method (FCAE) to reconstruct textural background of images, improving the discriminant power of the encoded feature maps (Yang et al. 2019). For the seismic data, Song et al. (2020) utilized convolutional autoencoder neural network to better represent the features of seismic data constrained by L1 regularization.

In summary, the deep convolutional denoising models can learn the local structural features of seismic data from a large number of data sets (Aaditya et al. 2019), but usually ignores the global information of seismic signals. When

processing seismic data, although noise suppression effect can be improved by increasing network depth, the global structures of seismic events can not be preserved well at low SNR. In addition, these denoising models pay same attention to all features at different position during network training. Consequently, the denoising model fails to identify the noise features that are similar with signals features, and thus needs more iterates to make loss function convergence, even fails to obtain satisfactory performance for low-frequency noise suppression.

For suppressing low-frequency seismic random noise, we propose a deep convolutional autoencoder network by integrating attention mechanism, named ADCAE. An attention module (AM) is embedded at the end of DCAE denoising model to combine the global information of the noisy data through long path connection and the local features extracted by DCAE. As a result, a weight matrix is obtained to dynamically adjust the attention to the features in different position, guiding the training process to improve the efficiency and efficacy of the proposed denoising model. In addition, a symmetric skip connection based on soft-thresholding is designed to transport the thresholded details features extracted by the shallow layers of DCAE to the deeper layer in the decoder. In this way, the ADCAE is able to alleviate the structure distortion of seismic events while completely suppressing low-frequency random noise.

## Attention mechanism

The attention mechanism of deep learning comes from the study of human visual attention mechanism. When a person pays attention to a certain goal or scenario, they quickly scan the area where they are interested, assign different levels of attention to different parts and concentrate on observing the points of interest in detail to obtain more detailed information. The attention mechanism mainly includes two aspects. One is to determine the part that needs to be followed, and the other is to allocate limited information resources to the important part.

From the perspective of application, attention can be divided into soft attention and hard attention (Zhao et al. 2017; Hu et al. 2020; Yang et al. 2020; Wilterson and Graziano 2021; Niu et al. 2022). Hard attention is mainly to randomly cut a certain feature region, so that the network only pays attention to the key region and completely ignores the unselected region, which can effectively reduce the number of network parameters. However, hard attention mechanism is a non-differentiable process, which usually relies on reinforcement learning to obtain weights and cannot be embedded into the network to train and learn with the whole model. Soft attention mainly relies on the relationship between features to learn weights, and then assigns different weights to

different features. Soft attention mechanism is differentiable and can be embedded into the network to learn by loss function convergence.

Spatial attention is a kind of soft attention mechanism. According to the importance of features in different positions, spatial attention learns the dynamic weight coefficients for the feature maps by exploiting the correlation of the features in different spatial positions (Woo et al. 2018). Let the deep neural network contain  $n$  channels and  $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n\}$  be the input of the attention module where  $\mathbf{X}_i$  is the feature map of the  $i$ th channel with the size of  $K_1 \times K_2$ . The weight matrix  $\mathbf{M} = \{\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_n\}$  is generated by the spatial attention mechanism and its dimension is the same as the input feature map. Based on the weight matrix, the important feature  $\mathbf{G} = \{\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_n\}$  of each channel is extracted as follows

$$\mathbf{G}_i = \mathbf{M}_i \cdot \mathbf{X}_i, \quad (1)$$

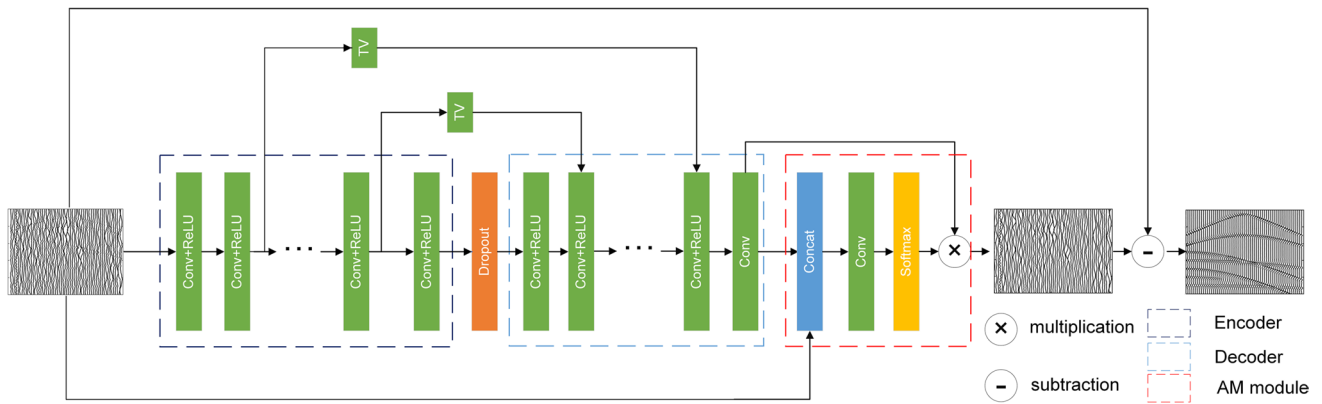
where ‘ $\cdot$ ’ denotes dot product. Spatial attention mechanism focuses on the important feature by assigning different weights for the features at different positions to achieve effective feature.

## Deep convolutional autoencoder network guided by attention mechanism

Low-frequency random noise in the seismic data is similar to the characteristics of the effective seismic signals, which makes it difficult for training deep convolutional network and the important information extraction from complex background noise. This problem is particularly prominent at low SNR. Aiming at this problem, we propose an attention mechanism based deep convolutional autoencoder network model to suppress low-frequency random noise. Through the spatial attention module, ADCAE integrates the local characteristics extracted by the DCAE network and the global characteristics of seismic data to generate a weight matrix. By dynamically adjusting weight values according to the importance of different features, the redundant features are attenuated and the important features are passed through the network, leading to efficient random noise prediction.

## Network architecture

The ADCAE denoising network consists of a deep convolutional autoencoder network and attention module (AM). As shown in Fig. 1, DCAE is divided into encoder module and decoder module. The encoder module contains 15 convolutional layers. Each convolutional layer has 64 convolutional units (Conv) with size of  $3 \times 3$ , and ReLU activation units. The convolutional layer adaptively



**Fig. 1** Architecture of ADCAE

learns useful information from input data according to task requirements. The features extracted by the  $i$ th convolutional kernel can be expressed as:

$$\mathbf{X}^i = \sigma(\mathbf{W}\mathbf{X}^{i-1} + \mathbf{b}), \quad (2)$$

where  $\mathbf{W}$  is the weight matrix,  $\mathbf{b}$  is bias,  $\sigma(\cdot)$  is the activation function. The decoder of DCAE consists of 15 convolutional layers. The first 14 convolutional layers are all composed of 64 convolutional units with the size of  $3 \times 3$ , ReLU activation units, and the last layer contains a  $3 \times 3$  convolutional unit, which makes the dimension of the output data to be the same as the dimension of the input data. The DCAE has the ability to extract the hidden representation of clean data from the noisy data through the encoder module and thus predicts noise by reconstructing the extracted hidden representation. The convolutional network adopted here ensures that the structural characteristics of seismic events can be preserved. In addition, we embed a dropout layer between the encoder and decoder. The dropout layer not only reduces the number of hidden features, but also decreases the correlation among the features, thereby improving the generalization capacity of the denoising model. Furthermore, the DCAE also adopts the soft-thresholding symmetric skip connection that contains a threshold shrinkage module named TV. Therefore, the detail information contained in the shallow layer is passed into the decoder, so as to better predict low-frequency random noise.

The AM module contains a Concat layer, a  $1 \times 1$  Conv unit and a Softmax layer. The AM module integrates the global characteristics of the input data through the long path and the local characteristics of reconstruction data by DCAE to generate the importance weight matrix, which dynamically adjusts features in different position during network training and thus guides the training process to pay more attention to the effective features.

## Denoising principle

The ADCAE with residual learning predicts random noise  $\tilde{\mathbf{N}}$  from the seismic data  $\mathbf{Y} = \mathbf{S} + \mathbf{N}$ , and then recovers the effective seismic signal  $\hat{\mathbf{S}} = \mathbf{Y} - \tilde{\mathbf{N}}$ . The residual learning strategy not only solves the problem of decreased performance as the network depth increasing, but also improves denoising performance of the network. The encoder of the ADCAE uses the mapping function  $F_e$  to extract the hidden representation  $\mathbf{H}$  from noisy data, which can be expressed as:

$$\mathbf{H} = F_e(\mathbf{Y}) = C_{15}(C_{14}(\dots C_1(\mathbf{Y}))), \quad (3)$$

where  $C_k$  indicates the mapping function of the  $k$ th convolutional layer. The hidden representation is mapped to seismic noise through the decoder function  $F_d$  as follows:

$$\tilde{\mathbf{N}} = F_d(\mathbf{H}) = D_{15}(D_{14}(\dots D_1(\mathbf{H}))), \quad (4)$$

where  $D_k$  represents the mapping of the  $k$ th convolutional layer in the decoder.

In order to alleviate the problem of the weak similar characteristics of low-frequency random noise, we add soft threshold-based symmetrical skip connection at every symmetric even-numbered layer between encoder and decoder. The threshold shrinkage module in the skip connection contains a pooling layer and sigmoid function. The pooling layer averages the features to reduce the interference, and then normalizes the pooling results by using sigmoid function to obtain the adaptive threshold for each feature. The output feature of the  $i$ th layer in encoder is thresholded as:

$$\mathbf{Q} = \mathbf{T} \cdot \mathbf{X}^i, \quad (5)$$

where  $\mathbf{T}$  is the adaptive threshold matrix, which has the same dimensions as  $\mathbf{X}^i$ . In the decoder, the output of  $l+1$ -i layer of ADCAE is expressed as:

$$\mathbf{X}^{l+1-i} = \mathbf{X}^i - \mathbf{Q} + D_i(\mathbf{X}^{l-i}), \quad (6)$$

where  $l$  represents the number of layers of the entire network. It indicates that symmetric skip connections can transfer the details features extracted in the encoder to the convolutional layers in decoder to help the network recover the signal better. In addition, symmetric skip connection is also important for backward propagating which solves the gradient vanishing during training. The threshold shrinkage module embedded in the skip connections further reduces the interference to ensure the accuracy of the predicted seismic noise  $\tilde{\mathbf{N}}$  by DCAE.

Since the DCAE assigns the same attention to the feature map in each position, it is very easy to confuse signal features with noise features during training, resulting in the residual signal in the predicted noise  $\tilde{\mathbf{N}}$ . It is necessary to pay more attention to the noise features during network training when random noise is similar to the effective signals. To this end, we design the AM module embed in the end ADCAE. The AM module utilizes the Concat layer to connect the predicted noise data with DCAE and the input data through a long path and obtain:

$$\mathbf{P} = \text{Concat}(\mathbf{Y}, \tilde{\mathbf{N}}). \quad (7)$$

Then, the weight coefficient matrix is generated through the convolution unit and Softmax and is expressed as:

$$\mathbf{I} = \text{Softmax}(C_{1 \times 1}(\mathbf{P})), \quad (8)$$

where  $C_{1 \times 1}$  denotes the mapping of  $1 \times 1$  Conv. The Softmax mapping that normalizes the weight into  $[0,1]$  is defined as:

$$\text{Softmax}(\mathbf{v}_i) = \frac{\exp(\mathbf{v}_i)}{\sum_j \exp(\mathbf{v}_j)}, \quad (9)$$

where  $\mathbf{v}_i$  represents the  $i$ th element of the matrix  $\mathbf{v}$ .

The AM module makes full use of the global features of the input data and the local features extracted from the DCAE to mine the correlation between the input data and the prediction noise, so that the weight matrix has different values at the corresponding noise position and signal position. The weight is large at the noise position and small at the signal position. Therefore, ADCAE uses the weight matrix to weight noise  $\tilde{\mathbf{N}}$  predicted by DCAE and the noise estimation is obtained by:

$$\hat{\mathbf{N}} = \mathbf{I} \cdot \tilde{\mathbf{N}}. \quad (10)$$

Compared with DCAE, ADCAE can more accurately predict low-frequency noise by focusing on the important information through the important weight of the AM module, and thereby the effective seismic signal, that is:

$$\hat{\mathbf{S}} = \mathbf{Y} - \hat{\mathbf{N}}, \quad (11)$$

can be better recovered in structural characteristic and details preservation.

## Network training

Training ADCAE model includes forward and backward propagation. During the forward process, the ADCAE model predicts the random noise ADCAE( $\mathbf{Y}_j; \Theta$ ) for the input noisy data  $\mathbf{Y}_j$  under the parameter set  $\Theta = \{\theta, \phi\}$ , where  $\phi$  is the parameter set of the AM module and  $\theta$  is the parameter set of the DCAE network. In this process, AM assigns different weights to features to select important features. Then parameters are updated by minimizing the mean square error loss function as follows:

$$\min_{\Theta} L = \frac{1}{2U} \sum_{j=1}^U \left\| \text{ADCAE}(\mathbf{Y}_j; \Theta) - \mathbf{N}_j \right\|^2, \quad (12)$$

with ADAM optimization algorithm based on gradient descent, where  $\{\mathbf{Y}_j, \mathbf{N}_j\}_{j=1}^U$  denotes  $U$  noisy-noise training pairs. In the backward propagation process, parameters are updated from the last layer to the first layer according to the derivative of the loss. Since the AM module is set at the last layer of the network, the parameters of the AM module  $\phi$  are updated before updating the parameters of DCAE network. When updating the parameters  $\theta$  of DCAE by forms of:

$$\frac{\partial \text{ADCAE}(\mathbf{Y}_j; \Theta)}{\partial \Theta} = \text{AM}(\phi) \cdot \frac{\partial \text{DCAE}(\mathbf{Y}_j; \theta)}{\partial \theta}, \quad (13)$$

AM module acts as a gradient filter to DCAE( $\mathbf{Y}_j; \theta$ ). Generally, network parameters are updated through both error gradients associated with the effective signals and correct gradients associated with random noise that are similar with the signals, leading to unsatisfactory denoising performance and large computation cost. Thanks to the AM module that integrates the global features of input and the local features obtained by DCAE, weight coefficient matrix allocates small weight value to the area dominant by signals and gives large value to the area without signals, thereby the correct gradients are selected and propagated by chain rule. The above analysis indicates that in the training process, the attention weight matrix can promote the ADCAE model to learn more robust noise characteristics, so as to better identify the signal from random noise even in the areas where the noise and signal are similar. Furthermore, AM makes the parameters updating in the right direction, and thus greatly reduce the number of iterations required for training.

## Training set

Because the open seismic training set does not meet the characteristics of the low-frequency random noise, we construct a training set suitable for the noise characteristics. The principle of training set construction is high quality, diversity, and the training data has similar characteristics with field seismic data.

The effective seismic signals are simulated by Ricker wavelet, zero-phase wavelet, and mixed-phase wavelet defined as:

$$g(t) = A \left[ 1 - 2\pi^2 f_0^2 (t - t_0)^2 \right] \exp \left[ -(\pi^2 f_0^2 (t - t_0)^2) \right], \quad (14)$$

$$g(t) = A \cos \left[ 2\pi f_0 (t - t_0) \right] \exp \left[ \frac{-(4\pi^2 f_0^2 (t - t_0)^2)}{r_1^2} \right], \quad (15)$$

$$g(t) = A \sin \left[ 2\pi f_0 (t - t_0) \right] \exp \left[ \frac{-(4\pi^2 f_0^2 (t - t_0)^2)}{r_2^2} \right], \quad (16)$$

where  $t_0$  is the starting time,  $r_1$  and  $r_2$  are the coefficients for adjusting the zero-phase wavelet and the mixed-phase wavelet, respectively. The dominant frequency is specified from 15 to 30 Hz. The apparent velocity ranges from 600 to 4000 m/s. The normalized amplitudes  $A$  attenuate from 1 to 0.1. We generate 64 synthetic seismic records and each seismic record includes 240 traces with 2000 samples in every trace.

For noise training sets, we utilize the random noise model (Li et al. 2017) to simulate seismic random noise and generate 64 synthetic seismic records including 240 traces and 2000 samples per trace. In addition, field noise data collected from desert area is used to build the training set. The low-frequency random noise data have 800 traces and 3000 samples per trace with the sampling rate of 500 Hz. The amplitudes of noise records are normalized to  $[-1, 1]$ .

All the synthetic signal records and noise records are cut into patches with size of  $50 \times 50$  according to the experience, and the overlapped rate is 50%. The noise data patches randomly multiply the noise level coefficient between 1 and 7, and are added to the signal patches to obtain noisy data patches. Finally, we obtain 58,624 noisy-synthetic noise patches pairs and 55,936 noisy-field noise patches pairs. The constructed training data set are used to the synthetic seismic data for training and testing while the ratio of training data and test data is 10:1.

The transfer learning strategy is adopted to train the networks. The synthetic seismic data are first added to train the network to learn the features pattern of the synthetic data, and then the field desert noise data as a label to fine-tune the parameters of the pretrained model to migrate the

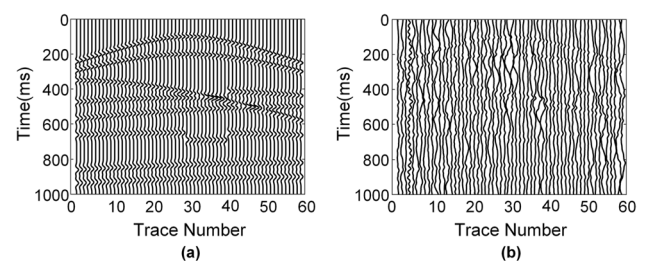
denoising ability of the pretrained denoising model to the low-frequency random noise suppression scenario.

## Synthetic and field data processing

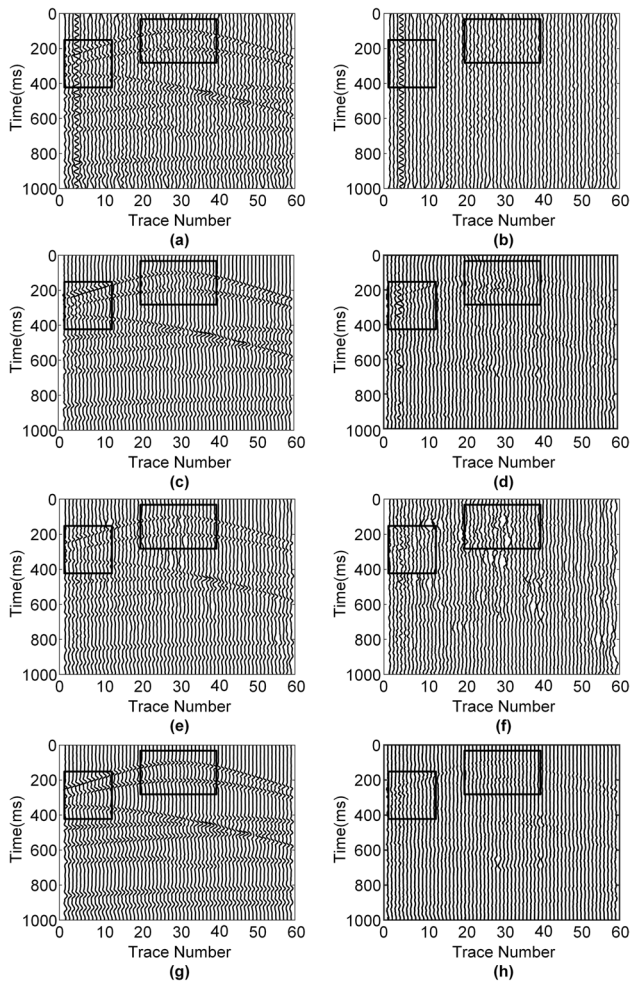
We investigate the efficacy and efficiency of the ADCAE denoising network on the synthetic seismic data and field seismic data, and compare the denoising performance with the F–K filter, DCAE denoising network and DnCNN (Zhao et al. 2019) in time domain and frequency domain. The denoised results are further evaluated by the SNR, MSE and training efficiency. In the synthetic and field examples, the frequency offset of the F–K filter is set to 9. Three deep denoising networks are trained by using the residual learning strategy on the simulated training sets. The learning rate starts from 0.001 and the number of training iterations is 50.

### Example on synthetic seismic data

The synthetic data used in this paper, as shown in Fig. 2a, includes 60 traces with 500 samples per trace and the sampling frequency is 500 Hz. The synthetic seismic data contains eight seismic events that are generated by the Ricker wavelet with the dominant frequencies of 25 Hz, 24 Hz, 22 Hz, 20 Hz, 19 Hz, 18 Hz, 15 Hz and 15 Hz, respectively. Low-frequency random noise simulated by homogeneous medium wave equation is added to obtain the noisy data as shown in Fig. 2b. The SNR is  $-9.88$  dB. As can be seen from the synthetic noisy record, random noise changes slowly and many effective seismic signals become unrecognizable. The effective seismic signals from the 3th to 7th traces even are seriously distorted. The F–K filter, DCAE denoising network, DnCNN and ADCAE denoising network are applied to the synthetic data. The denoised results and the difference data between the noiseless data and the denoised data are shown in Fig. 3a–h. Comparing the denoised results of the four methods, we can see that the F–K filter is able to suppresses most the background noise. However, some low-frequency random noise is still seen in the denoised data as



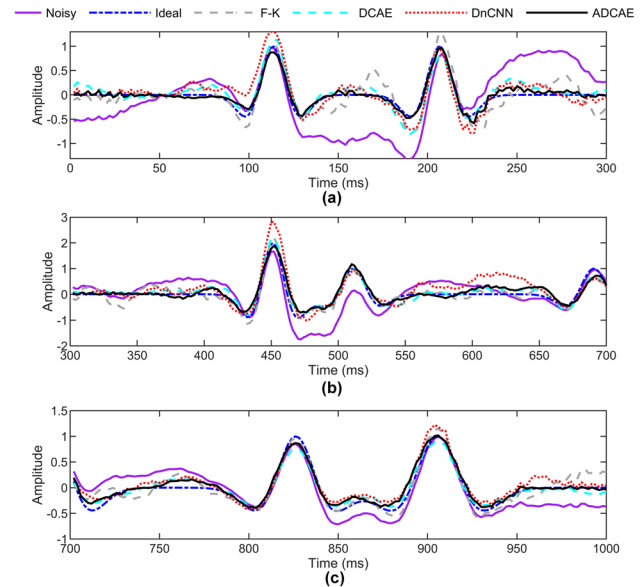
**Fig. 2** Synthetic seismic record: **a** synthetic seismic data, **b** synthetic noisy data



**Fig. 3** Denoised synthetic data: **a** F–K filter, **b** difference data of F–K filter, **c** DCAE, **d** difference data of DCAE, **e** DnCNN, **f** difference data of DnCNN, **g** ADCAE, **h** difference data of ADCAE

indicated by the rectangle frames and fails to well recover the seismic events.

By contrast, DCAE and DnCNN denoising network have the cleaner background and clearer seismic events in the denoised results. However, the effective signals in the area with low SNR are still visible near 0.2 s of the 30th trace in the difference data, indicating that the two methods are unable to well preserve the seismic effective signals from low-frequency random noise with similar characteristics to the effective seismic signals. Compared with DCAE and DnCNN, the proposed ADCAE obtains the minimum signal distortion and maximum noise reduction on the synthetic seismic data illustrated in the denoised result and the difference data. Even under the extremely low SNR, the weak seismic events can be well recovered from random noise with similar waveforms by the proposed denoising network guided by the attention module. To further illustrate the improvement of denoising performance in detail,

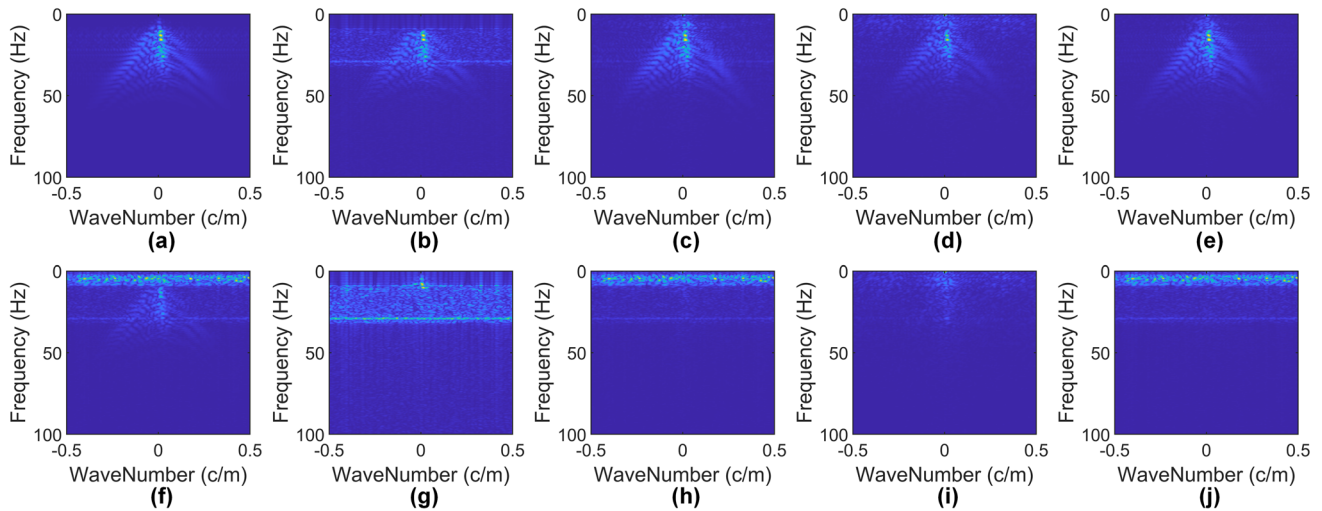


**Fig. 4** Waveform comparison of the 37th trace: **a** signal waveform of 0–300 ms, **b** signal waveform of 300–700 ms, **c** signal waveform of 700–1000 ms

we compare the denoised trace and the clean seismic trace in the time domain and frequency domain. Figures 4 and 5 show the waveforms of the 37th trace and F–K spectra of the synthetic data before and after denoising, respectively. From waveforms comparison, we observe that the seismic signals recovered by the F–K filter is preserved well and noise is incompletely suppressed. DnCNN and DCAE have similar performance in signal preservation. In contrast, the proposed ADCAE obtains least deviation between the denoised trace and the ideal in waveforms in the four methods. In Fig. 5, low-frequency energy of the denoised signal by the F–K filter and DnCNN is higher than these of DCAE and our method, indicating that desert random noise is suppressed incompletely. The F–K spectra of the DCAE and ADCAE are closer to the ideal one than other two methods. In addition, we analyze the denoising effectiveness of the four methods under different noise intensities with SNR and the mean squared error (MSE) before and after denoising. The SNR of the noisy data varies from  $-2.81$  to  $-10.12$  dB.

As can be seen in Table 1, ADCAE gives the highest SNR and lowest MSE in the four methods under various noise intensity, which verifies the good denoising performance of ADCAE.

Above results come to the conclude that the ADCAE outperforms the other three methods in terms of noise reduction performance and quantitative evaluation, which benefits from the attention module embedded in the ADCAE. Next, we further analyze the role of attention mechanism on improving noise reduction performance and training efficiency. Figure 6b shows the weight matrix generated by AM

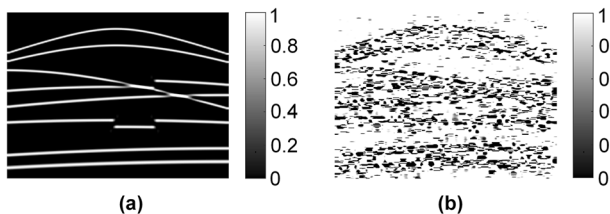


**Fig. 5** Comparison of the F–K spectra for synthetic data: clean data (a), noisy data (f), denoised data of F–K filter (b), DCAE (c), DnCNN (d), ADCAE (e) and the corresponding difference data (g, h)

**Table 1** SNRs and MSEs of results via four denoisers under different noise intensity

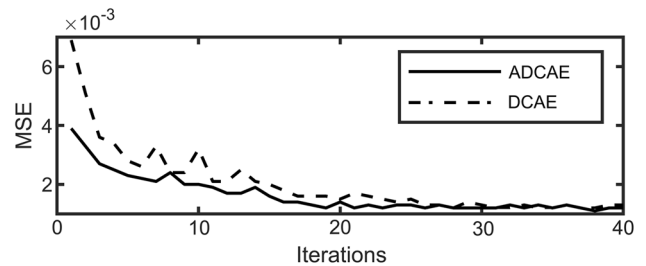
Input SNR (dB)	F–K		DCAE		DnCNN		ADCAE	
	SNR (dB)	MSE	SNR (dB)	MSE	SNR (dB)	MSE	SNR (dB)	MSE
–2.81	9.63	0.014	11.21	0.009	10.37	0.012	<b>12.5</b>	<b>0.007</b>
–4.34	8.40	0.018	10.68	0.011	9.03	0.016	<b>12.06</b>	<b>0.008</b>
–6.24	6.78	0.026	9.63	0.014	6.91	0.025	<b>11.25</b>	<b>0.009</b>
–8.86	4.40	0.045	7.57	0.022	3.62	0.054	<b>9.72</b>	<b>0.013</b>
–10.12	3.20	0.059	6.37	0.028	2.02	0.078	<b>8.80</b>	<b>0.016</b>

Bold indicates the highest value in all SNRs and the smallest value in all MSEs



**Fig. 6** Output weight matrix of AM module: **a** signal location diagram, **b** output weight matrix

module in the 10th iteration during the training ADCAE. Compared with the signal position of the synthetic seismic data shown in Fig. 6a, the weight coefficients at the signal position are mostly 0 (black) or small values (gray), while the weight coefficients at the background noise position are close to 1 (white). Thus, the weight matrix gives different attention to different features. When using residual learning to train the ADCAE model, the AM module pays more attention to the noise position and works like a gradient filter during the backward propagation process, reducing the gradient related to signal and focusing on the characteristics of the noise. This correct attention of the AM module on



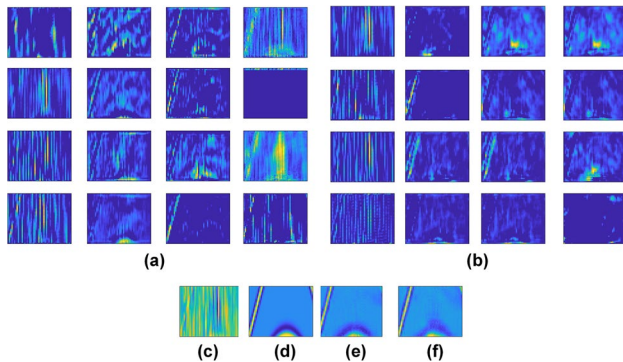
**Fig. 7** Comparison on training efficiency of DCAE and ADCAE

effective characteristics comes from the integration of the correlation between the global characteristics of the seismic data and the local characteristics of the convolution autoencoder network. In the iterative process, the global features reflecting the seismic events structure are strengthened and the wrong features caused by local weak similar interference are weakened.

In addition, the gradient filtering mechanism can effectively improve the training efficiency of the model. Figure 7 depicts the MSEs between the clean data shown in Fig. 2a and the denoised data by using the DCAE and ADCAE at



different training iterations under the same conditions. As can be seen, the MSEs of the ADCAE are lower than that of the DCAE and the two networks finally reach the optimal denoising effect on the synthetic seismic data. It is worth noting that the ADCAE approximates to the minimum value at the 18th iteration, while DCAE reaching the minimum

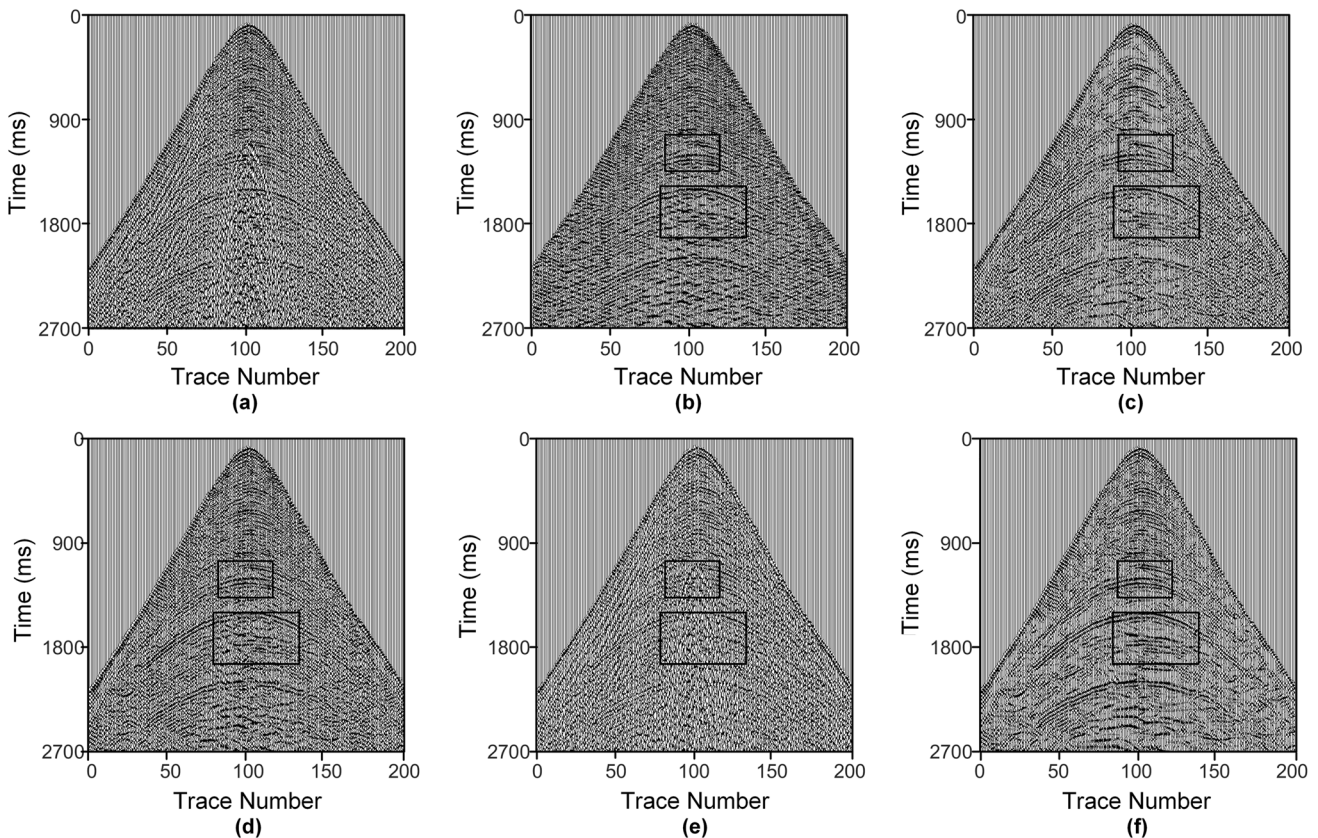


**Fig. 8** Visualization of the denoising process: columns from left to right indicates the feature maps extracted by layer 1, 14, 17 and 29 in ADCAE (a) and DCAE (b) with the noisy patch input (c); d clean data, e denoised data of ADCAE, f denoised data of DCAE

MSE at the 28th iteration. The iterations number for training of ADCAE to achieve the optimal denoising performance is reduced by 10 times compared with DCAE. We visualize the extracted features after convolution with and without AM module to directly understand the function of AM module in denoising process as shown in Fig. 8. The extracted feature maps in the shallow layers are noisy due to the low SNR of the input, and the signal features can be identified in the extracted features of the layer 14 and layer 17. In the layer 29, the feature maps in ADCAE better characterize the feature of desert random noise added in the input patch than DCAE without AM module. Therefore, the structural feature of the effective signal in the denoised data is better preserved by ADCAE than DCAE.

### Example on field seismic data

This section is designed to evaluate the denoising performance of the ADCAE model on the field seismic data. The selected field seismic data as shown in Fig. 9a is collected from Tarim Basin in western China. The field data are composed of 200 traces and the sampling frequency is 500 Hz. It can be seen from the noisy seismic data that random noise is

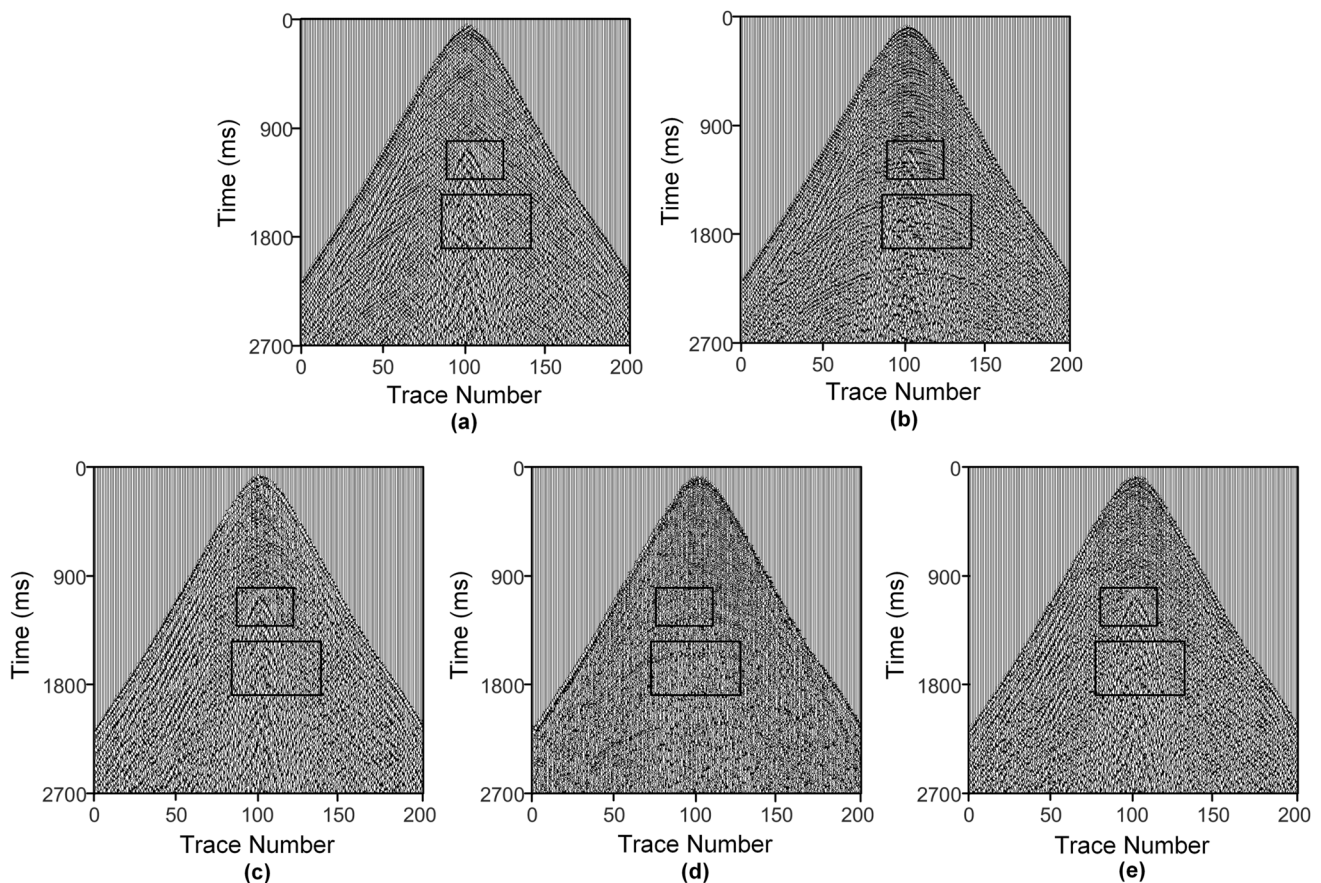


**Fig. 9** Field seismic data and denoised results of five denoising methods: a field seismic data, b result by using F–K filter, c result by using DCAE, d result by using DnCNN, e result by using GAN, f result by using ADCAE

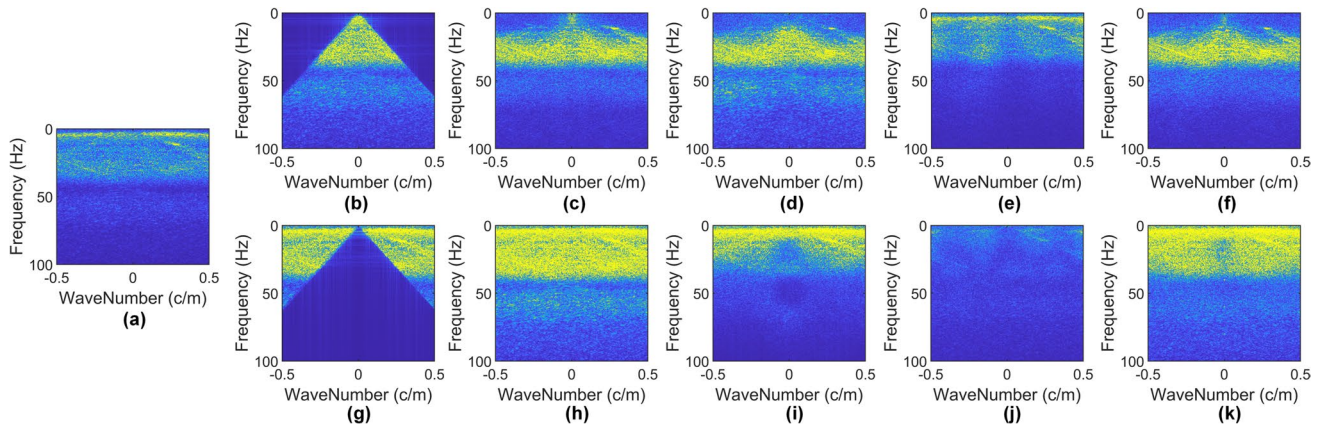
low-frequency and has high intensity that changes over time and space. The effective seismic signals have been severely disturbed by desert random noise, and the seismic events in some areas are damaged and difficult to identify.

We apply the F–K filter, DCAE, DnCNN, generative adversarial network (GAN) and ADCAE denoising models to this field seismic data, respectively. The structure and hyperparameters of DCAE, ADCAE, and DnCNN are the same as above synthetic example. The architecture and hyperparameters of GAN are set as the method proposed by Li et al. (2021). For the field data denoising, the DCAE and ADCAE trained by the synthetic dataset are migrated to the field desert noise by transfer learning. We set parameters of first seven layers of the trained network as the initial parameters, and then fine-tune the network by using the noisy synthetic data–field noise training set through residual learning with field noise record without shooting as ground-truth labels. This training method not only improves the noise suppression ability and signal recovery effect of the model by using supervised training on synthetic data, but also transfers the denoising ability of the denoising model to the field random noise suppression scenario.

Figures 9 and 10 show the denoised results of the five denoising methods and the difference data between the raw data and the denoised data. The filtering ability of F–K filter to suppress field random noise is limited so that the effective seismic events are unable to recover and there are still signal residues can be seen in the difference data. DCAE can effectively suppress background noise. However, many effective signals are also considered as noise and are eliminated as marked by the rectangle frames. The effective seismic signal between 0 and 900 ms are destroyed. In contrast, DnCNN can achieve cleaner background, clearer and more continuous seismic events, but some artificial signals are introduced in the denoised data. The denoising performance of GAN is not satisfactory in noise reduction. Compared with four methods, the ADCAE more effectively restores the reflected seismic events, and the events become smoother and more continuous. In addition, the signals that are severely damaged by random noise are also recovered in the rectangular frames. The difference data shows that plenty of removed noise can be seen in the difference data of DnCNN and ADCAE, indicating that most of background noise in the



**Fig. 10** The difference data between field seismic data and denoised data via five denoisers: **a** difference data of F–K filter, **b** difference data of DCAE, **c** difference data of DnCNN, **d** difference data of GAN, **e** difference data of ADCAE

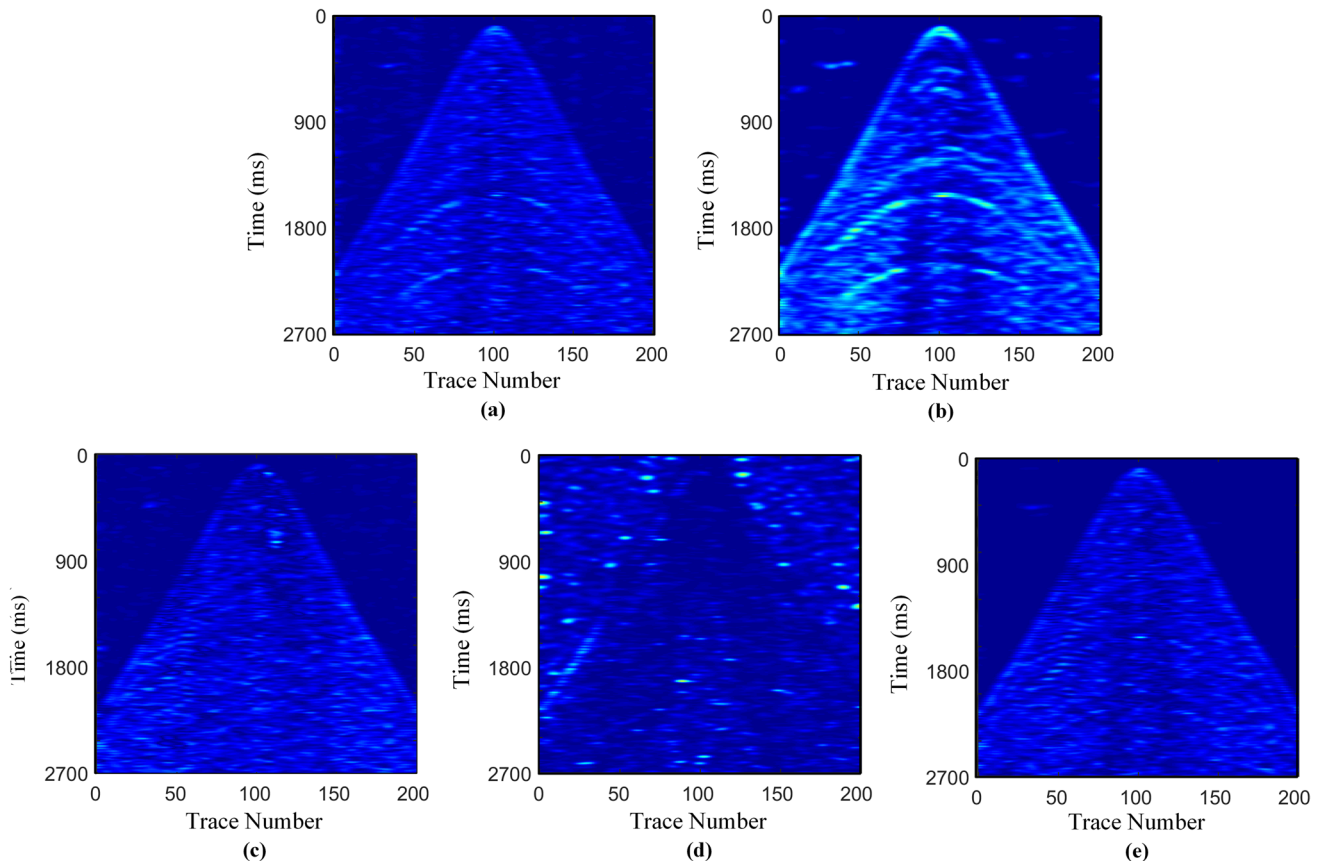


**Fig. 11** The comparison of F–K spectra of the field data: noisy data (a), denoised data by F–K filter (b), DCAE (c), DnCNN (d), GAN (e), ADCAE (f), and their difference data (g–k), respectively

field seismic records can be suppressed. However, there is signal leakage in the results of DCAE and GAN. Because ADCAE can focus on the important characteristics during training, ADCAE achieves less signal leakage in the difference data than other four methods. In summary, the tests on the synthetic and field data demonstrate that ADCAE guided by the AM module outperforms other four methods

in effective seismic signals preservation and weak similar noise suppression.

We also evaluate the denoising performance in frequency-wavenumber domain. The F–K spectra of the field data, denoised data and the difference data are shown in Fig. 11. Comparing the F–K spectra, we can see that the F–K filter not only removes random noise, but also



**Fig. 12** The comparison of the local similarity maps between denoised data and removed noise: a local similarity map by F–K filter, b local similarity map by DCAE, c local similarity map by DnCNN, d local similarity map by GAN, e local similarity map by ADCAE

removes the seismic signals. GAN removes some noise and are unable to recover seismic signals. In contrast, DnCNN can well preserve the seismic signals while can not thoroughly remove random noise. Compared with four methods, ADCAE better recovers seismic signals and removes more random noise.

We also compare the local similarity map of the denoised data and the difference data (Chen and Fomel 2015) to evaluate the signal leakage. From the similarity comparison are shown in Fig. 12, we can see that F–K filter and DCAE lead to obvious leakage of the reflection events. In contrast, the signal leakage of reflection events is almost invisible from the local similarity map of ADCAE and DnCNN denoising. The energy loss (mean of the local similarity map) is 0.2133, 0.3727, 0.2078, 0.1014 and 0.1901 for F–K filter, DCAE, DnCNN, GAN, and ADCAE, respectively. We find that the ADCAE gives less energy loss for seismic signals than F–K filter, DCAE, DnCNN that give satisfactory denoising performance. The comparison of local similarity verifies that the proposed ADCAE can well preserve the effective seismic signals while suppressing field desert random noise processing.

## Conclusion

The deep convolutional autoencoder denoising model based on attention mechanism is proposed to reduce low-frequency seismic random noise in this paper. The theoretical analysis and field seismic data results verifies that ADCAE can allocate the correct attention on the effective features in different position by leveraging the global information and local information of the seismic data. Consequently, the gradient associate with the effective features can be selected and propagated during training and thus improving the training efficiency and denoising ability of ADCAE model. Furthermore, the symmetric skip connection with the threshold shrinkage module can help the ADCAE to filter and then passes the details features in the encoder to the decoder to retain the complex structure of the seismic events in the field seismic data. Simultaneously, the dropout layer and AM module can reduce the number of the network parameters to be learned and thereby further improving the efficiency of the model training. The results on the field seismic data collected from desert areas demonstrate that the ADCAE can thoroughly suppress low-frequency random noise while preserving the complex structures of the seismic events.

In future work, we will develop semi-supervised and unsupervised strategy for reducing the need for labels in our supervised denoising model and transfer learning. In addition, we will add physics-based constraints to the training procedure of our denoising framework so that the degrees of freedom can be reduced.

**Acknowledgements** This work is supported by the National Natural Science Foundation of China (41774117).

## Declarations

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

## References

- Aaditya P, James S, Dinei F, et al (2019) RePr: improved training of convolutional filters. In: 2019 IEEE/cvf conference on computer vision and pattern recognition USA
- Chen Y, Fomel S (2015) Random noise attenuation using local signal-and-noise orthogonalization. *Geophysics* 80(6):WD1–WD9
- Feng Q, Li Y, Wang H (2020) Intelligent random noise modeling by the improved variational autoencoding method and its application to data augmentation. *Geophysics* 86(1):19–31. <https://doi.org/10.1190/geo2019-0815.1>
- Hu J, Shen L, Albanie S et al (2020) Squeeze-and-excitation networks. *IEEE Trans Pattern Anal Mach Intell* 42(8):2011–2023. <https://doi.org/10.1109/TPAMI.2019.2913372>
- Krizhevsky A, Sutskever I, Hinton G (2012) ImageNet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 25(2):84–90. <https://doi.org/10.1145/3065386>
- Li G, Li Y, Yang B (2017) Seismic exploration random noise on land: modeling and application to noise suppression. *IEEE Trans Geosci Remote Sens* 55(8):4668–4681. <https://doi.org/10.1109/TGRS.2017.2697444>
- Li Y, Wang H, Dong X (2021) The denoising of desert seismic data based on cycle-gan with unpaired data training. *IEEE Geosci Remote Sens Lett* 18(11):2016–2020. <https://doi.org/10.1109/LGRS.2020.3011130>
- Lin H, Wang S, Li Y (2021) A branch construction-based CNN denoiser for desert seismic data. *IEEE Geosci Remote Sens Lett* 18(4):736–740. <https://doi.org/10.1109/LGRS.2020.2981965>
- Ma H, Qian Z, Li Y et al (2019a) Noise reduction for desert seismic data using spectral kurtosis adaptive bandpass filter. *Acta Geophys* 67(1):123–131. <https://doi.org/10.1007/s11600-018-0232-0>
- Ma H, Yao H, Li Y et al (2019b) Deep residual encoder-decoder networks for desert seismic noise suppression. *IEEE Geosci Remote Sens Lett* 17(3):529–533. <https://doi.org/10.1109/LGRS.2019.2925062>
- Ma H, Yan J, Li Y et al (2019c) Desert seismic random noise reduction based on LDA effective signal detection. *Acta Geophys* 67:109–121. <https://doi.org/10.1007/s11600-019-00250-0>
- Mao X, Shen C, Yang Y (2016) Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In: 30th conference on neural information processing systems (NIPS)
- Masci J, Meier U, Dan C et al (2011) Stacked convolutional autoencoders for hierarchical feature extraction. *Lect Notes Comput Sci* 6791(1):52–59. [https://doi.org/10.1007/978-3-642-21735-7\\_7](https://doi.org/10.1007/978-3-642-21735-7_7)
- Niu R, Sun X, Tian Y et al (2022) Hybrid multiple attention network for semantic segmentation in aerial images. *IEEE Trans Geosci Remote Sens* 60:1–18. <https://doi.org/10.1109/TGRS.2021.3065112>
- Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back propagating errors. *Nature* 323(6088):533–536. <https://doi.org/10.1038/323533a0>
- Saad O, Chen Y (2021) A fully unsupervised and highly generalized deep learning approach for random noise suppression. *Geophys*

- Prospect 69(4):709–726. <https://doi.org/10.1111/1365-2478.13062>
- Saad O, Bai M, Chen Y (2021) Uncovering the microseismic signals from noisy data for high-fidelity 3D source-location imaging using deep learning. *Geophysics* 60(6):ks161–ks173. <https://doi.org/10.1190/GEO2021-0021.1>
- Sang W, Yuan S, Yong X et al (2021) DCNNs-based denoising with a novel data generation for multidimensional geological structures learning. *IEEE Geosci Remote Sens Lett* 18(10):1861–1865. <https://doi.org/10.1109/LGRS.2020.3007819>
- Sang W, Yuan S, Han H et al (2023) Porosity prediction using semi-supervised learning with biased well log data for improving estimation accuracy and reducing prediction uncertainty. *Geophys J Int* 232(2):940–957. <https://doi.org/10.1093/gji/ggac371>
- Song H, Gao Y, Chen W et al (2020) Seismic random noise suppression using deep convolutional autoencoder neural network. *J Appl Geophys* 178:0926–9851. <https://doi.org/10.1016/j.jappgeo.2020.104071>
- Sun X, Li Y (2020) Denoising of desert seismic signal based on synchrosqueezing transform and Adaboost algorithm. *Acta Geophys* 68:403–412. <https://doi.org/10.1007/s11600-020-00408-1>
- Vincent P, Larochelle H, Bengio Y et al (2008) Extracting and composing robust features with denoising autoencoders. In: *Proceedings of the twenty-fifth international conference*
- Wang S, Li Y, Zhao Y (2020) Desert seismic noise suppression based on multimodal residual convolutional neural network. *Acta Geophys* 68(2):389–401. <https://doi.org/10.1007/s11600-020-00405-4>
- Wilterson A, Graziano M (2021) The attention schema theory in a neural network agent: controlling visuospatial attention using a descriptive model of attention. *Proc Natl Acad Sci*. <https://doi.org/10.1073/pnas.2102421118>
- Woo S, Park J, Lee J et al (2018) CBAM: convolutional block attention module. In: *Computer vision. 15th European conference*
- Yang H, Chen Y, Song K et al (2019) Multiscale feature – clustering—based fully convolutional autoencoder for fast accurate visual inspection of texture surface defects. *IEEE Trans Autom Sci Eng* 16(3):1450–1467. <https://doi.org/10.1109/TASE.2018.2886031>
- Yang H, Kim J, Kim H et al (2020) Guided soft attention network for classification of breast cancer histopathology images. *IEEE Trans Med Imaging* 39(5):1306–1315. <https://doi.org/10.1109/TMI.2019.2948026>
- Yang L, Wang S, Chen X et al (2022) Unsupervised 3-D random noise attenuation using deep skip autoencoder. *IEEE Trans Geosci Remote Sens* 60:5905416. <https://doi.org/10.1109/TGRS.2021.3100455>
- Yuan S, Jiao X, Luo Y et al (2021) Double-scale supervised inversion with a data-driven forward model for low-frequency impedance recovery. *Geophysics* 87(2):R165–R181. <https://doi.org/10.1190/geo2020-0421.1>
- Zhao B, Wu X, Peng Q et al (2017) Diversified visual attention networks for fine-grained object classification. *IEEE Trans Multimedia* 19(6):1245–1256. <https://doi.org/10.1109/TMM.2017.2648498>
- Zhao Y, Li Y, Dong X et al (2019) Low-frequency noise suppression method based on improved DnCNN in desert seismic data. *IEEE Geosci Remote Sens Lett* 16(5):811–815. <https://doi.org/10.1109/LGRS.2018.2882058>
- Zhao Y, Li Y, Yang B (2020a) Low-frequency desert noise intelligent suppression in seismic data based on multiscale geometric analysis convolutional neural network. *IEEE Trans Geosci Remote Sens* 58(1):650–665. <https://doi.org/10.1109/TGRS.2019.2938836>
- Zhao Q, Du Q, Li Q et al (2020) Robust dictionary learning for erratic noise-corrupted seismic data reconstruction. *Acta Geophys* 68:687–700. <https://doi.org/10.1007/s11600-020-00433-0>
- Zhang Y, Lin H, Li Y et al (2020) Seismic signal enhancement and noise suppression using structure-adaptive nonlinear complex diffusion. *IEEE Trans Geosci Remote Sens* 58(3):2198–2211. <https://doi.org/10.1109/TGRS.2019.2954949>

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.