Check for updates

Seismic Shot Gather Denoising by Using a Supervised-Deep-Learning Method with Weak Dependence on Real Noise Data: A Solution to the Lack of Real Noise Data

Xintong Dong[1] · Jun Lin[1] · Shaoping Lu[2,3] · Xingguo Huang[1] · Hongzhou Wang[1] · Yue Li[4]

© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

In recent years, supervised-deep-learning methods have shown some advantages over conventional methods in seismic data denoising, such as higher signal-to-noise ratio after denoising, complete separation of signals and noise in shared frequency bands and intelligent denoising without artificial parameter tuning. However, the lack of real noise data matched with raw seismic data has greatly limited its further application. In this paper, we take the surface seismic shot gather as an example to explore the corresponding solutions and propose a novel supervised-deep-learning method with weak dependence on real noise data based on the data augmentation of a generative adversarial network. We utilize the generative adversarial network to augment the pre-arrival noise data acquired from the shot gather itself, thereby obtaining a large amount of synthetic noise data whose probability distribution is extremely similar to that of the real noise in shot gather; the augmented synthetic noise data and sufficient synthetic signal data obtained by forward modeling together form the augmented training dataset. Meanwhile, the dilated convolution and gradual denoising strategy are adopted to construct the basic architecture of denoising convolution neural network. Finally, the above augmented dataset is used to train the network, so as to establish a nonlinear and complex mapping relationship between raw seismic data and desired signals. Both synthetic and real experiments demonstrate that our method can realize the intelligent denoising of different common-shot-point records in shot gather with the help of limited pre-arrival noise data.

Article Highlights

- We introduce the data augmentation strategy into the field of deep-learning-based seismic denoising, thereby alleviating the dependence of supervised-deep-learning methods on real noise data
- We propose a novel denoising network architecture with strong recovery ability for weak desired signals by using the gradual denoising strategy and dilated convolution

---

Extended author information available on the last page of the article

- The augmented synthetic noise data can meet the requirement of supervised-deep-learning methods on the quantity and authenticity of training data, so this data augmentation strategy by using the Generative Adversarial Net (GAN) is a solution to the lack of real noise data

1 Introduction

The suppression of pre-stack seismic background noise is an indispensable step in seismic data processing (Cooper and Cook 1984; Krohn et al. 2008). Effective denoising methods can significantly increase the signal-to-noise ratio (SNR) and resolution of seismic data (Dong et al. 2019a), thereby providing more reliable data support for the velocity analysis, inversion, migration, and interpretation that follow.

In the past few decades, experts have proposed numerous classical seismic denoising methods which can be broadly divided into the following five categories: time–frequency-analysis-based methods, decomposition-based methods, low-rank-based methods, transform-based methods, and dictionary-learning-based methods.

The time–frequency-analysis-based methods mainly leverage the differences between signals and noise in time or frequency domain to distinguish the two, such as band-pass filter, median filter (Tukey, 1974; Bednar, 1983; Duncan and Beresford 1995; Wang et al. 2020a), frequency-offset (f-x) deconvolution (Caneles, 1984; Gulunay 1986, 2017; Harris and White 1997) and predictive filter (Caneles, 1984; Gulunay 2000; Chen and Ma 2014). Caneles (1984) first utilize the predictive filter in f-x domain to suppress the random noise in seismic data. Liu et al. (2012) apply the regularized non-stationary autoregression (RNA) in f-x domain, thereby overcoming the assumption of linearity and stationary of the signals in conventional f-x deconvolution; when facing complex geological structure, this proposed f-x RNA has shown better performance than conventional f-x deconvolution, but its performance is easily affected by two parameters: filter length and smoothing radius of shaping operator. Wang et al. (2020a) propose a robust vector median filtering to process complex seismic data acquired from an area with strong sedimentary rhythmites; this variant of median filtering not only significantly increases the SNR of seismic data, but also effectively preserves the amplitude of signals and the discontinuity of reflections.

The decomposition-based methods, including empirical mode decomposition (EMD; Huang et al. 1998; Battista et al. 2007; Chen and Ma, 2014; Gomez and Velis 2016), variational modal decomposition (VMD; Dragomiretskiy and Zosso, 2014), singular value decomposition (SVD; Andrews and Patterson, 1976; Bekara and van der Baan 2007), and morphological decomposition (Matheron, 1975; Serra, 2011; Huang et al. 2020), decompose raw seismic data into multiple constitutive components, so as to complete the denoising task by retaining the components associated with signals and remove those representing noise (Xue et al. 2019). To copy with short and quickly varying events, Bekara and van der Baan (2007) apply SVD using a local window sliding in space and time; compared with f-x deconvolution and median filtering, this local SVD has advantages in removing noise and disadvantages in enhancing the events with conflicting dips. Chen and Ma (2014) combine EMD with f-x deconvolution to adapt to more complex structures and also alleviate the phenomenon of signal leakage. Han and van der Baan (2015) propose an adaptive-thresholding ensemble EMD workflow to improve the SNR of microseismic data and then extend this approach into f-x domain to attenuate random and coherent noise in post-stack seismic reflection data.

The low-rank-based methods assume that theoretical clean seismic data are a low-rank structure, and noise contamination will increase its rank, so that one can attenuate noise via rank reduction, as well as recovering signals (Hagen 1982; Wang et al. 2020b). Such methods include robust principal component analysis (RPCA; Wright et al. 2009; Chen et al. 2015; Liu et al. 2021), singular spectrum analysis (SSA; Vautard and Ghil, 1989; Oropeza and Sacchi 2011; Lari et al. 2019), and Cadzow filtering (Cadzow 1988; Trickett 2008). Cadzow (1988) propose the Cadzow filtering which is introduced into the field of seismic random noise attenuation by Trickett (2008). Oropeza and Sacchi (2011) utilize the randomized SVD to replace the original SVD adopted in multichannel SSA, so as to alleviate the time cost of rank reduction step; both synthetic and real examples suggest that this improved multichannel SSA is a seismic denoising and reconstruction method that can be considered, but cannot replace mature methods which have been used in modern industrial processing streams. Chen et al. (2015) utilize RPCA in f-x domain to suppress the erratic noise in seismic data; however, there is still some visible residual noise in the denoised result, and the performance can be further enhanced.

The transform-based methods, such as wavelet (Grossman and Morlet, 1984; Mousavi and Langston 2016; Zhao et al. 2020), shearlet (Guo and Labate, 2007; Zhang and van der Baan 2018; Dong et al. 2019b), curvelet (Candes and Guo, 2002; Candes et al. 2006; Herrmann and Hennenfent, 2008; Gorszczyk et al. 2015), seislet (Fomel 2006), and radon (Schultz and Claerbout, 1978; Sacchi and Porsani, 1999), can transform raw seismic data into different sparse domains and then precisely represent the desired signals by using a few number of parameters (Wang et al. 2019; Dong et al. 2019b). Anvari et al. (2017) combine the synchrosqueezed wavelet transform with low-rank matrix approximation to attenuate the seismic random noise; this novel noise attenuation method performs better than three existing methods, but improper rank will degrade its performance. To solve the spectral energy overlapping between signals and ground-roll which disturbs frequency-wave number (f-k) domain dip filtering, Naghizadeh and Sacchi (2018) propose a novel removal workflow for ground-roll by designing a mask function in curvelet domain; this proposed method performs better than f-k domain dip filtering, but it is sensitive to the thresholding constant whose optimal value is obtained via numerical tests and even visual observation. Conventional threshold functions cannot distinguish the signal and noise coefficients of downhole microseismic three-component data with low SNR in three-dimensional (3D) shearlet domain. Therefore, Dong et al. (2019b) utilize a trained back-propagation neural network to act as the threshold function of 3D shearlet transform, which not only enhances the denoising performance, but also brings some additional time–cost.

Different from the fixed basis commonly used in the above transform-based methods, the dictionary-learning-based methods (Mallat and Zhang 1993; Olshausen and Field 1996) adaptively learn the basis from the seismic data itself to make the transform domain optimally sparse, which will in turn suppress noise (Sahoo and Makur 2013; Chen 2020). To sparsely represent seismic data, Kaplan et al. (2009) utilize a data-driven sparse-coding algorithm to adaptively learn basis functions and then achieve denoising in transform domain. Beckouche and Ma (2014) propose a new dictionary-based denoising method on the basis of a variational sparse-representation model. To alleviate huge computational cost caused by thousands of SVD operations, Chen (2020) utilize the sequential generalized K-means algorithm (Sahoo and Makur 2013) to update each dictionary atom; experiment results suggest that this fast dictionary learning exhibits higher processing efficiency at the expense of no observably worse denoising performance, which is beneficial to the denoising of multi-dimensional seismic data.

Numerous real-world problems have demonstrated the effectiveness of the above existing methods, but most of them have some certain limitations. For instance, the denoising performance of transform-based methods largely depends on threshold functions; an inappropriate threshold function will most likely lead to residual noise and attenuation of signal amplitude (Dong et al. 2019b). How to select an optimal filter length has often troubled the predictive-filtering-based methods. EMD-based methods are prone to suffer from the phenomenon of mode aliasing; also, they cannot process large volume of seismic data efficiently because of the large computational cost caused by interpolation step, which is more noticeable in some noise-assisted EMD-based methods (Gomez and Velis 2016). Numerous SVD operations required in the step of dictionary update lead to the heavy computational cost of dictionary-learning-based methods. Besides, almost all of the above methods rely on prior experience, even visual observations, to adjust parameters artificially, thereby obtaining the best possible denoised results (Yu et al. 2019; Zhong et al. 2022), which is not intelligent and also time-consuming.

With the great enhancement of computer software, deep artificial neural networks (or called deep-learning) have developed rapidly since 2010 and been widely used in some fields of data processing, such as image denoising, image super-resolution, pattern recognition, and image fusion (Zhang et al. 2017; Remez et al. 2018; Dian et al. 2021). Typical deep-learning-based methods include convolutional neural network (CNN), recurrent neural network, deep belief network and auto-encoder (Lecun et al. 1998, 2015; Hinton et al. 2006; Vincent et al. 2010; Krizhevsky et al. 2012), among which CNN is a compelling topic in recent years due to its local perception and weight sharing. In general, we utilize supervised, unsupervised, self-supervised, semi-supervised methods (Jing and Tian 2021; Schmarje et al. 2021), or the combination of them (Saad and Chen 2020) to train the network, thereby obtaining an optimal mapping relationship between input and output.

Compared with conventional seismic denoising methods, deep-learning-based methods are a novel data-driven approach and do not rely on accurate assumptions (Yu et al. 2019). With the increasing interest in the use of supervised-deep-learning (SDL) methods (if not specified, SDL methods refer to supervised deep learning with CNN herein) in exploration geophysics, some related denoising approaches for seismic data have been proposed (Dong et al. 2019a; Yang et al. 2021). Yu et al. (2019) utilize CNN to construct a uniform framework for the attenuation of general seismic noise including random noise, ground roll, and multiple; although this uniform framework performs better than traditional methods in both denoising quality and automation (no requirement of parameter tuning), it will fall into some situations: lack of training dataset, improper hyper-parameters, and obvious difference between training and test dataset. Yang et al. (2021) propose an improved residual CNN combined with multiple double-layer convolution blocks; this network shows high training efficiency and encouraging suppression performance for seismic random noise. To conquer the challenge of distributed optical fiber acoustic sensing (DAS) seismic data with low SNR, Dong and Li (2021) propose a convolutional adversarial denoising network (CADN) based on the adversarial training between denoiser and discriminator; this proposed CADN can suppress DAS noise effectively and also recovery the weak up-going reflected signals, but the authenticity and adequacy of training dataset are likely to degrade its denoising performance. To reinforce the preservation of weak signals, Zhong et al. (2022) propose a novel multi-scale CNN combined with hierarchical structure to attenuate seismic random noise; synthetic and real examples suggest that this proposed multi-scale CNN performs better than three existing CNN-based methods in weak signal recovery, but adopting the multi-scale structure leads to an increase in computational cost. Meanwhile, deep-learning-based methods have also been gradually applied in some other fields of

seismic exploration, such as first arrival time picking, full waveform inversion (FWI), fault detection, and interpolation (Pochet et al. 2019; Zhang and Alkhalifah 2019; He and Wang 2020; Guo et al. 2021; Kaur et al. 2021; Kazei et al. 2021).

Although these proposed SDL denoising methods for seismic data have shown remarkable performance, they often require sufficient paired clean and noise data with high authenticity to train the network. In other words, insufficient training samples and obvious difference between training and test dataset are likely to result in the impaired performance of these methods. Some studies (Yang et al. 2021; Dong and Li 2021) have demonstrated the feasibility of utilizing forward modeling to obtain the clean data, but how to obtain sufficient noise data for training and also ensure that the distribution of the noise data for training is similar to that of the real noise in raw seismic data?

Previous approaches mainly include: (1) real recorded ambient noise data matched with raw seismic data; (2) utilizing some other kinds of noise instead, such as Gaussian noise; (3) filtering some noise from raw seismic data. Obviously, approach (1) is the best choice; approaches (2) and (3) are likely to degrade the generalization ability and denoising performance of SDL methods. Unfortunately, matched real noise data mentioned in approach (1) is usually limited mainly due to two reasons: on the one hand, there are numerous kinds of seismic background noise (Yu et al. 2019; Dong and Li 2021); on the other hand, even noise of the same kind has different properties (Zhong et al. 2015). To alleviate the dependence on real noise data, some experts turn to unsupervised and self-supervised methods; some relevant studies have demonstrated the potential and feasibility of these methods (Saad and Chen 2020, 2021; Birnie et al. 2021; Meng et al. 2022). However, SDL denoising methods are still an essential topic of seismic data processing, so it is still meaningful to explore the corresponding solutions to the problem of how to obtain a large amount of noise data for paired training.

The Generative Adversarial Net (GAN) proposed by Goodfellow et al. (2014) contains two modules: generator and discriminator. The generator captures the potential features of real data, thereby generating numerous similar synthetic data; next, the discriminator is utilized to distinguish between real and synthetic data (Lu and Mak 2020; Moreno-Barea et al. 2020). Through the continuous confrontation between the two, the generator tries to trick the discriminator with synthetic data; whereas the discriminator want to distinguish between synthetic and real data. Ultimately, the discriminator will determine that the synthetic data produced by generator is true, indicating the probability distribution of synthetic data is extremely similar to that of real data (Dong and Li 2021). Both speech data (Lu and Mak 2020) and medical image (Chaudhari et al. 2020) have demonstrated the feasibility of data augmentation strategy by using GAN.

Obviously, GAN has strong potential to alleviate the dependence of SDL methods on real noise data. In this paper, taking the surface shot gather as an example, we introduce the data augmentation strategy of GAN into seismic denoising and propose a novel CNN-based denoising workflow with weak dependence on real noise data ($\text{CNN}_{\text{WDRND}}$). For the training dataset, we utilize GAN to augment the limited pre-arrival noise data extracted from shot gather to be processed, so as to obtain sufficient synthetic noise data which is used to construct the augmented noise dataset. There are two main reasons for the use of pre-arrival noise data. On the one hand, it is relatively easy to obtain the pre-arrival noise data from shot gather; on the other hand, the probability distribution and statistical properties (gaussianity, stationarity, linearity, etc.; Zhong et al. 2015) of pre-arrival noise are extremely similar to those of post-arrival noise (here refers to random noise, excluding the source-related coherent noise) that contaminates the desired signals. Meanwhile, numerous forward models with variable parameters are activated

by some artificial seismic wavelets with different central frequencies, thereby obtaining a variety of clean seismic data with high authenticity; then, we utilize the clean data to construct the signal dataset of CNN_{WDRND}. For network structure, the dilated convolution and gradual denoising strategy are adopted to construct the basic structure of denoising CNN (D-CNN). The dilated convolution increases the size of receptive field without increasing the number of network parameters, which improves the feature capture ability of D-CNN (Yu and Koltun 2016); the gradual denoising strategy takes into account both low and high dimensional features of input data to suppress noise layer-by-layer, enhancing the recovery ability of D-CNN to weak signals. Both synthetic 2D and real 3D shot gathers demonstrate that, with the help of limited real noise data (i.e., pre-arrival noise data) and the data augmentation by using GAN, the proposed CNN_{WDRND} can effectively suppress seismic background noise and completely recover signals. The denoising performance of CNN_{WDRND} is superior to that of traditional methods and a existing deep-learning method: feed forward denoising CNN (DnCNN). In addition, CNN_{WDRND} can intelligently process different common-shot-point (CSP) records of shot gather without artificial parameter tuning.

2 The Adopted Generative Adversarial Network

In this paper, we adopt the Wasserstein GAN (WGAN; Arjovsky et al. 2017) to augment the pre-arrival noise data. The WGAN utilizes the Earth-Mover distance to measure the difference in probability distribution between generated synthetic data and input real data, rather than the cross entropy used in original GAN. Compared with original GAN, the training of WGAN is more stable, and the generated synthetic data are more diverse (Arjovsky et al. 2017; Yang et al. 2018). As shown in Fig. 1, WGAN contains a generator and a discriminator. The generator can produce a large amount of synthetic noise data whose probability distribution is similar to that of the pre-arrival noise data; then, the discriminator is used to distinguish between the generated synthetic noise data and the pre-arrival noise data. Finally, the discriminator cannot distinguish the synthetic noise data from the pre-arrival noise data, indicating that the probability distributions of the two kinds of noise are extremely close.

The loss function of WGAN is shown in Eq. (1):
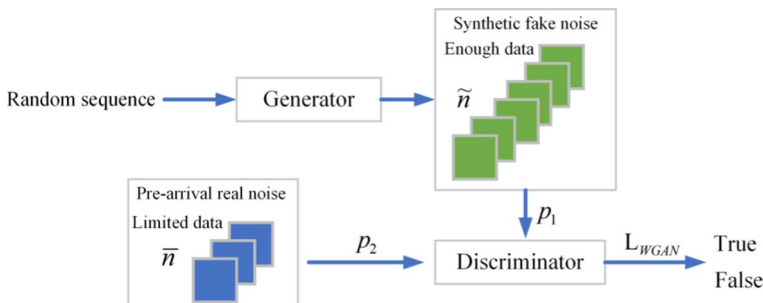
Fig. 1 The workflow of WGAN

$$L_{WGAN} = \underbrace{\underset{\tilde{n}\sim p_1}{E}[D(\tilde{n})] - \underset{\overline{n}\sim p_2}{E}[D(\overline{n})]}_{\text{Wasserstein distance}} + \underbrace{\lambda \underset{\hat{n}\sim p_3}{E}\left(\left\|\nabla_{\hat{n}}D(\hat{n})\right\|_2 - 1\right)}_{\text{Gradient penalty term}} \tag{1}$$

This loss function is composed of a Wasserstein distance and a gradient penalty term. $\tilde{n}$ represents the synthetic noise data produced by generator, and p_1 is its probability distribution. $\overline{n}$ stands for the pre-arrival noise data whose probability distribution is p_2. D is the mapping relationship established by the discriminator. λ represents the penalty coefficient whose value is 10 in this paper; the rationality of this value is testified by a variety of architectures and datasets (Gulrajani et al. 2017).

In the gradient penalty term, the gradient term $\left\|\nabla_{\hat{n}}D(\hat{n})\right\|_2$ is with respect to the points $\hat{n}$ whose probability distribution is p_3. p_3 is implicitly defined as a distribution sampling uniformly from p_1 and p_2 (Gulrajani et al. 2017):

$$\hat{n} = \eta\tilde{n} + (1-\eta)\overline{n} \tag{2}$$

$$\eta \sim U[0,1], \tilde{n} \sim p_1, \overline{n} \sim p_2 \tag{3}$$

Gulrajani et al. (2017) have demonstrated the effectiveness of penalizing the gradient over the distribution p_3 by several experiments.

The basic structures of the adopted generator and discriminator are shown in Fig. 2. In this paper, we utilize the stochastic gradient descent method (SGD) to train the WGAN. This network training for data augmentation is carried in Pytorch framework (Paszke et al. 2019). Through some experience and many experiments, the corresponding parameters of WGAN are set as shown in Table 1.

3 The Augmented Training Dataset

In general, the SDL denoising method entails a training dataset with high authenticity to optimize its network parameters. The proposed CNN_{WDRND} adopts the paired training, and its training dataset contains a signal dataset and a noise dataset.
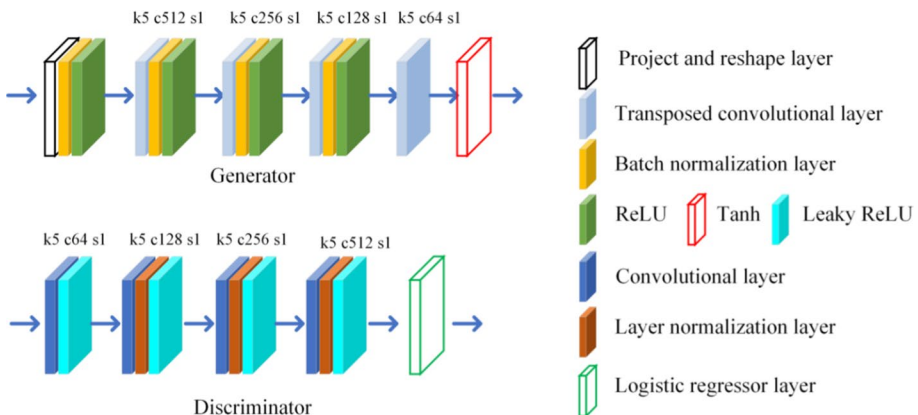
Fig. 2 The basic structures of generator and discriminator (k, c, and s represent the size of convolutional kernel, the number of mapping, and stride)

Table 1 The parameter settings of WGAN

Parameters	Specifications
The size of input random sequence	1×64
Patch size	64×64
Optimizer	Adam
Batch size	128
Learning rate	10^{-4}
Number of epochs	500
The size of convolutional kernel	5×5

3.1 The Augmented Noise Dataset

For noise dataset, the real recorded ambient noise data matched with raw seismic data is often the best choice. However, matched real noise is often limited, and SDL methods require a large amount of noise data to train the network, limiting the further application of SDL methods. To address this problem, we utilize WGAN to augment some pre-arrival noise data acquired from the raw shot gather itself, thereby obtaining sufficient similar synthetic noise data. This data augmentation strategy ensures that: (1) the augmented synthetic noise data are similar to the real noise in raw shot gather and (2) the diversity and volume of augmented synthetic noise data can meet the training requirements of D-CNN that follows. Specifically, we extract about 1000 64×64 pre-arrival noise patches from the raw shot gather to be processed. Then, the WGAN is utilized to augment these pre-arrival noise patches, thereby obtaining about 10,000 synthetic noise patches. These augmented synthetic noise patches form the augmented noise dataset of $\text{CNN}_{\text{WDRND}}$.

3.2 The Signal Dataset

The desired seismic signals cannot be obtained due to the constant existence of seismic background noise. Thus, we utilize a large amount of synthetic clean seismic data obtained by forward modeling to construct the signal dataset. Firstly, we construct 50 2D velocity forward models by changing numerous parameters including wave velocity, stratum thickness, reflection interface morphology, density, trace interval and source location; Fig. 3 displays four of the 50 constructed forward models. Secondly, some artificial seismic wavelets with different central frequencies are used to activate these 50 forward models, so as to obtain numerous synthetic clean seismic records by using the elastic wave equation solved by finite difference method. Ultimately, we extract 8940 64×64 signal patches from the synthetic clean seismic records obtained by forward modeling, 35 of which are shown in Fig. 4. The parameter settings of forward modeling are listed in Table 2.
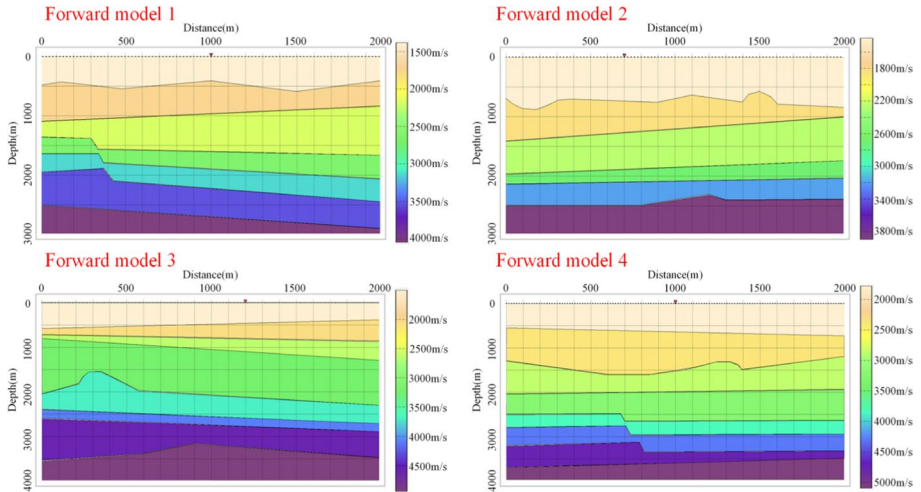
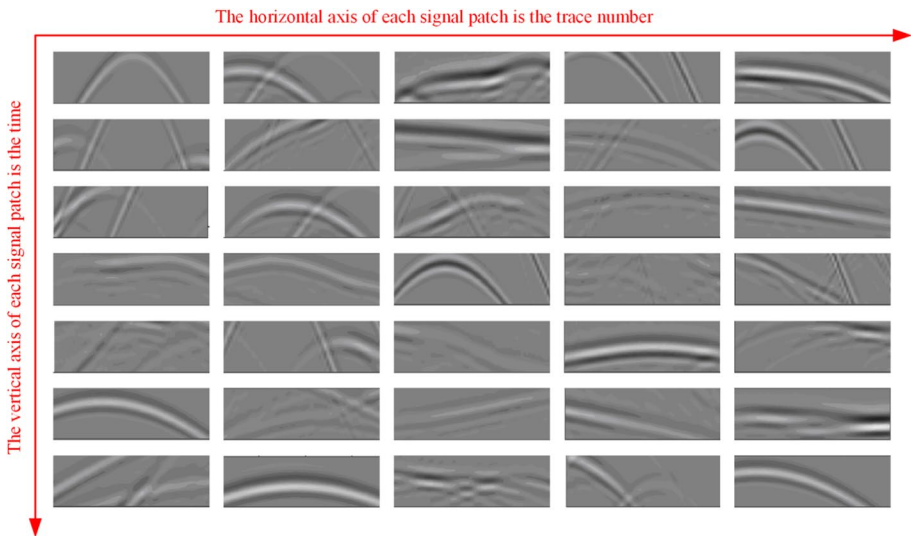**Fig. 3** Four of the 50 forward models

Fig. 4 35 of the 8940 signal patches (each signal patch contains 64 traces, and each trace contains 64 sampling points)

4 The Basic Structure of D-CNN and its Denoising Principle

To adapt to lower SNR and more complicated background noise, it is necessary to explore a D-CNN with stronger feature extraction ability, thereby better realizing the separation of signals and noise.

Table 2 The parameter settings of forward modeling

Parameters	Specifications
Artificial seismic wavelet	Ricker, single, symmetrical
Central frequency (Hz)	15–35
Trace interval (m)	10, 20, 30
Wave velocity (m/s)	1300–5200
Sampling frequency (Hz)	500
Density (kg/m^3)	1900–2800
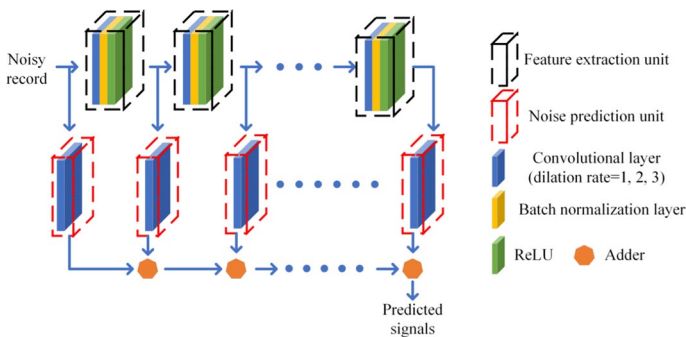
Fig. 5 The basic structure of D-CNN

In this paper, as shown in Fig. 5, we utilize the dilated convolution (Zhang et al. 2019; Liu et al. 2020) and gradual denoising strategy (Romano and Elad 2015; Remez et al. 2018; Lemarchand et al. 2020) to construct the basic structure of D-CNN. In this network, multiple sequential feature extraction units gradually construct the high-dimensional features by extracting the low-dimensional features of input data (it is generally believed that front-end feature extraction units mainly extract low-dimensional features of data, and the back-end is high-dimensional features); meanwhile, numerous noise prediction units perform noise prediction operations on the output of each feature extraction unit in turn. With the cooperation of the two kinds of units, this proposed D-CNN can suppress noise and recover signals layer-by-layer, which is beneficial to the recovery of weak signals. Different from conventional CNN, this proposed D-CNN considers both low-and high-dimensional features, not just the high-dimensional features.

Furthermore, we adopt the dilated convolution in 3rd, 6th, 9th, 12th feature extraction units; the dilated convolution enlarges the receptive field without increasing the number of weights and bias. Considering that the receptive field should be less than patch size, the dilated rates are set as 2, 3, 3, 2, successively.

As shown in Eq. (4), the raw seismic data Y can be expressed as the linear superposition of signals X and noise N.

$$Y = X + N \tag{4}$$

By optimizing numerous network parameters, the D-CNN can establish a nonlinear and complex mapping relationship between the raw seismic data and signals, the specific process is shown in Eqs. (5) and (6):

$$N_p = P(Y;\theta), \tag{5}$$

$$X_d = Y - N_P, \tag{6}$$

where P represents the mapping relationship established by D-CNN; $\mathbf{N_P}$ is the predicted noise by using P; θ stands for the network parameters mainly include weight and bias; $\mathbf{X_d}$ is the denoised seismic data. In this paper, we utilize the loss function based on mean square error (MSE) to optimize the network parameters of D-CNN, thereby obtaining the optimal θ. The specific expression of MSE loss function is shown in Eq. (7):

$$L_{MSE}(\theta) = \frac{1}{2B} \sum_{k=1}^{B} \left\| P(n_k + x_k;\theta) - n_k \right\|_F, \tag{7}$$

where B represents the batch size, $\mathbf{n_k}$ is a batch of noise patches in the augmented noise dataset, $\mathbf{x_k}$ is a batch of signal patches in the signal dataset, and $\|.\|_F$ stands for the Frobenious norm. In the training process, we utilize SGD to optimize the network parameters of D-CNN; in this way, the global gradient is replaced by local gradient, which accelerates the convergence rate of network training. The specific training process is shown in Algorithm 1.

Algorithm 1 Training process of D-CNN

Require: B, batch size; E, epoch number; K, the number of iterations in one epoch; D_1, the signal dataset; D_2, the augmented noise dataset.

1. For i=1, 2, 3, ······, E do

2.   For j=1, 2, 3, ······, K do

3.   Sample $\left\{\mathbf{x}^{(k)}\right\}_{k=1}^{B} \in D_1$, a batch of signal patches in the signal dataset .

4.   Sample $\left\{\mathbf{n}^{(k)}\right\}_{k=1}^{B} \in D_2$, a batch of noise patches in the augmented noise dataset.

5.   Number $\left\{a^{(k)}\right\}_{k=1}^{B}$, a batch of random numbers which are used to adjust the amplitude of noise patches; the value ranges from 0.5 to 3.

6.   Sample $\left\{\mathbf{m}^{(k)}\right\}_{k=1}^{B} = \left\{a^{(k)}\right\}_{k=1}^{B} \times \left\{\mathbf{n}^{(k)}\right\}_{k=1}^{B}$, a batch of noise patches after amplitude adjustment.

7.   Sample $\left\{\mathbf{y}^{(k)}\right\}_{k=1}^{B} = \left\{\mathbf{x}^{(k)}\right\}_{k=1}^{B} + \left\{\mathbf{m}^{(k)}\right\}_{k=1}^{B}$, a batch of noisy patches with different SNRs.

8.   Input $\left\{\mathbf{y}_{nr}^{(k)}\right\}_{k=1}^{B} = \left\{\mathbf{y}^{(k)}\right\}_{k=1}^{B} \Big/ \left\{\max\left(\mathbf{y}^k\right)\right\}_{k=1}^{B}$, a batch of noisy patches after normalization.

9.   Label $\left\{\mathbf{m}_{nr}^{(k)}\right\}_{k=1}^{B} = \left\{\mathbf{m}^{(k)}\right\}_{k=1}^{B} \Big/ \left\{\max\left(\mathbf{m}^k\right)\right\}_{k=1}^{B}$, a batch of noise patches after normalization.

10. $\theta \leftarrow \nabla_\theta \left[\dfrac{1}{B} \sum_{k=1}^{B} \left\| P\left(\mathbf{y}_{nr}^{(k)}\right) - \mathbf{m}_{nr}^{(k)} \right\|_F^2 \right]$, the calculation of MSE loss function.

11.   End for

12. End for

Both the training and testing of D-CNN are performed in MATLAB environment on a server with E5-2600 v4 processor, Windows 10 64-bit operating system, 64-GB memory, and NVIDA GeForce GTX 1080.

In addition to the training dataset, network hyper-parameters are also likely to affect the performance of trained denoising model. Through repeated experiments and comprehensive consideration of denoising performance, training time, and over-fitting, the hyper-parameters of D-CNN are set as shown in Table 3.

In order to be more intuitive, we give the basic workflow of the proposed CNN_{WDRND} in Fig. 6.

Table 3 The hyper-parameters of D-CNN

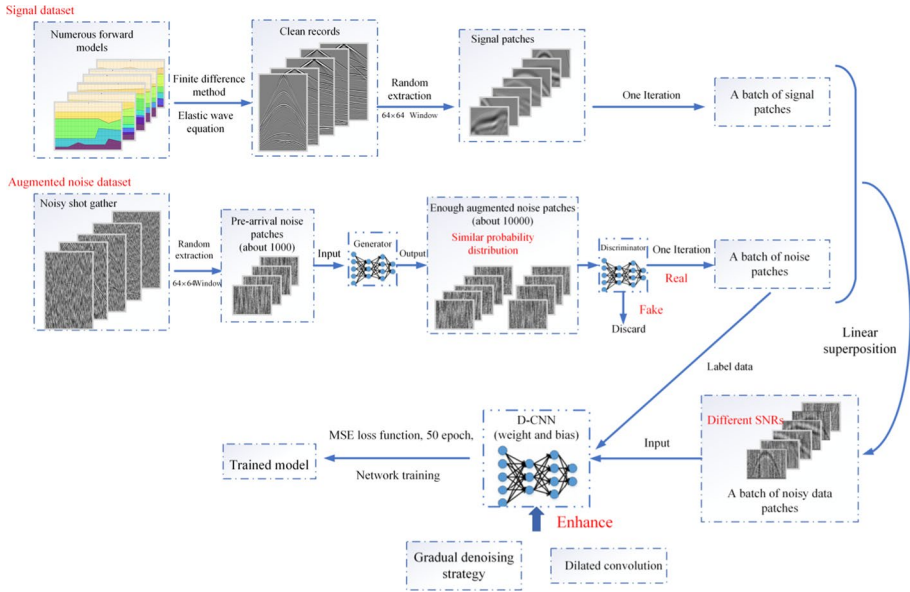| Hyper-parameters	Specifications
Network depth	14
Dilated rate	1,2,3
Patch size	64×64
Batch size	128
Learning rate	$\left[10^{-3}, 10^{-5}\right]$
Number of epochs	50
The size of convolutional kernel	3×3
The number of mappings	128
Stride	1

Fig. 6 The workflow of CNNWDRND

5 Synthetic Example

In this paper, we utilize SNR and root MSE (RMSE) to measure the noise intensity of noisy seismic records and the denoising performance by using different denoising methods. Definition of the two measurements is shown in Eqs. (8) and (9).

$$SNR(dB) = 10 \log_{10} \sum_{i=1}^{N} s_i^2 \bigg/ \sum_{i=1}^{N} (d_i - s_i)^2, \tag{8}$$
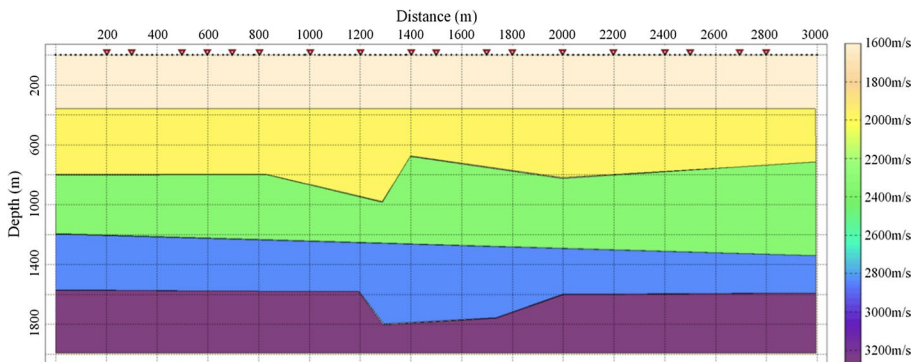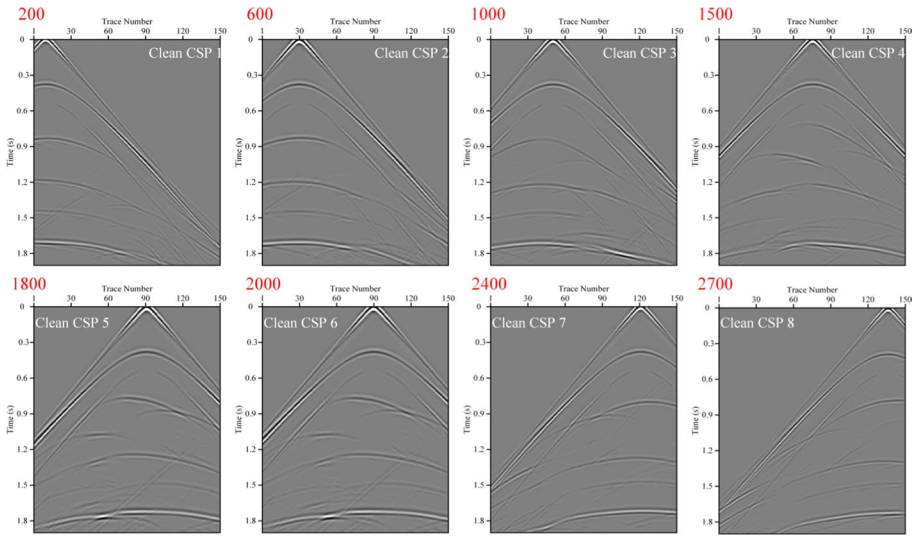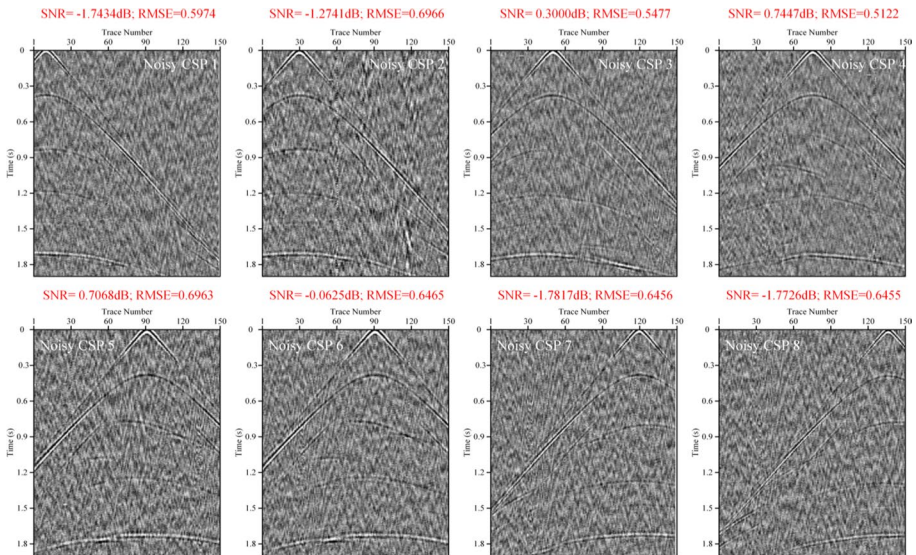

Fig. 7 The velocity forward model adopted in synthetic example

Fig. 8 Eight of the eighteen theoretical clean CSP records

Fig. 9 The eight synthetic noisy CSP records corresponding to the clean records in Fig. 8

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (d_i - s_i)^2}{N}} \qquad (9)$$

where s_i represents the clean seismic record; d_i stands for noisy seismic record or denoised seismic record; N is the number of sampling points. Large SNR and smaller RMSE indicate more complete noise suppression and better preservation for signal amplitude, respectively.

5.1 The Construction of Synthetic 2D Surface Shot Gather

To testify the effectiveness of the proposed CNN_{WDRND}, we firstly construct a velocity forward model shown in Fig. 7. This forward model is not included in the 50 forward models used in the construction of signal dataset, thereby ensuring that test and training are independence of each other. Secondly, we utilize Ricker wavelets whose central frequencies range from 20 to 35 Hz to act as the source; eighteen sources (i.e., red inverted triangles in Fig. 7) with different locations are utilized to activate this forward model, so as to obtain eighteen theoretical clean CSP records by using the elastic wave equation solved by finite difference method. Thirdly, some real random noise extracted from a real seismic ambient noise record is added to the eighteen clean CSP records, thereby obtaining the corresponding eighteen noisy CSP records which together form a 2D noisy synthetic surface shot gather.

As shown in Fig. 8, we display eight of the eighteen clean CSP records whose source locations are given in red numbers. In Fig. 9, we plot the eight noisy CSP records corresponding to the eight clean records shown in Fig. 8; also, SNRs and RMSE of these noisy records are given in red numbers. As shown in Fig. 9, most of the events are seriously contaminated by the random noise with strong energy, showing poor continuity, especially the weak reflected events.

5.2 The Comparison of Denoised Results

We extract 964 64×64 pre-arrival noise patches from the noisy surface shot gather, thereby obtaining 9830 64×64 augmented synthetic noise patches by using WGAN. These 9830 augmented noise patches and the 8940 signal patches are leveraged to train the DnCNN and $CNN_{WDRND.}$ Note that the noise data used for network training only contains the
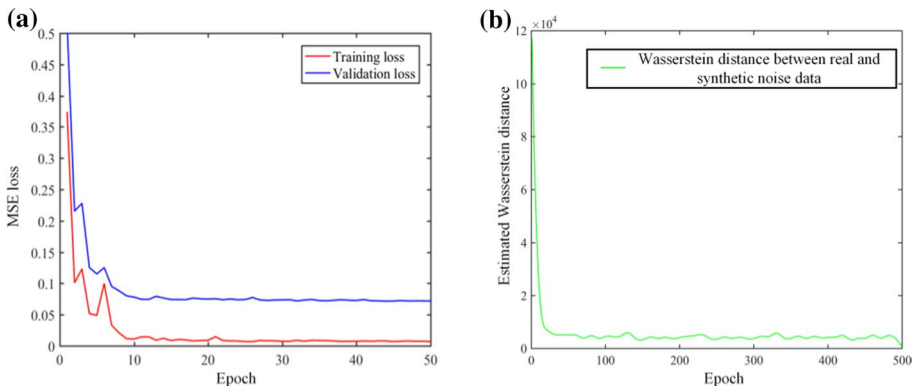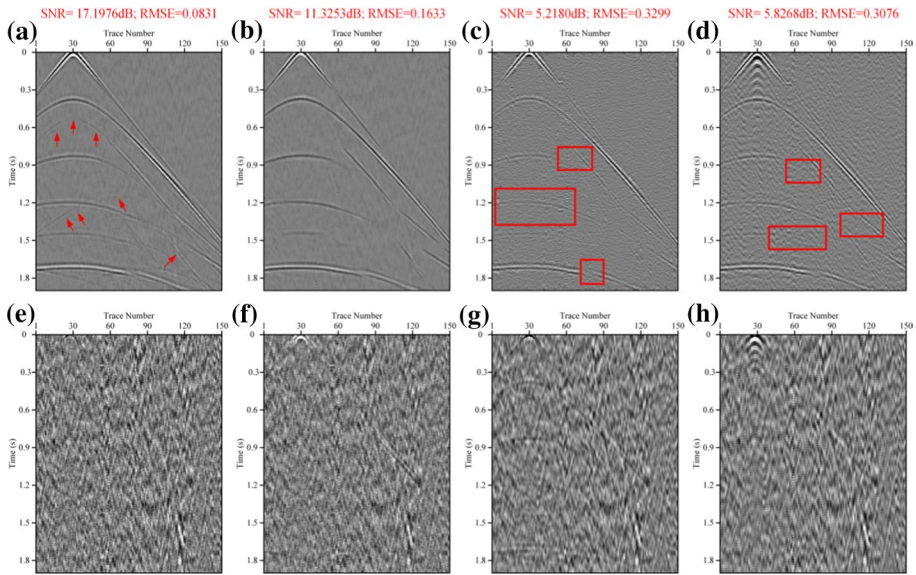

Fig. 10 (**a**) the curve of MSE loss; (**b**) the curve of Wasserstein distance
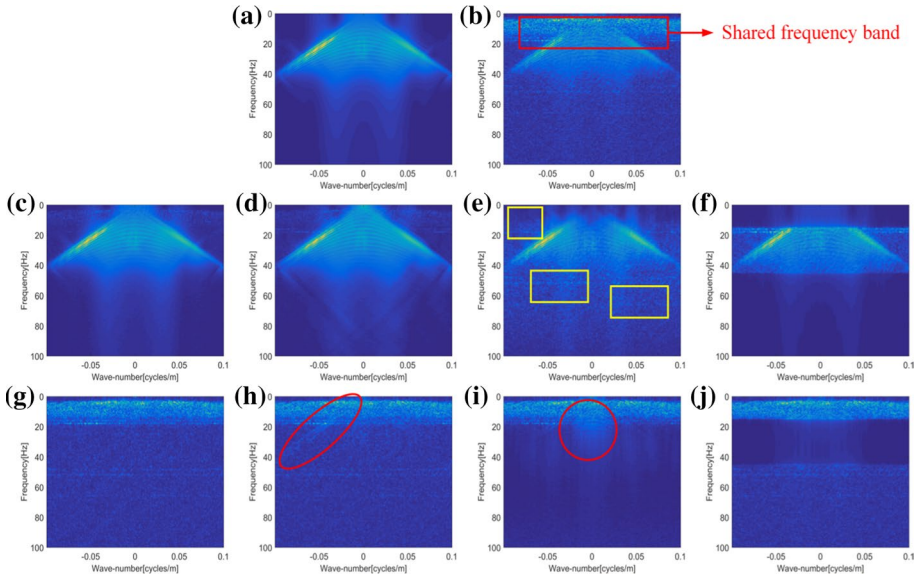
Fig. 11 Denoised results of the noisy CSP record with source location 600. (**a**)-(**d**) are the denoised results by using CNNWDRND, DnCNN, EEMD, and BPF, successively. (**e**)-(**h**) are the corresponding difference records. Corresponding clean and noisy records are the clean and noisy CSP 2 shown in Figs. 8 and ,9 respectively

synthetic noise patches obtained by using the augmentation of WGAN, not the pre-arrival noise patches extracted from the noisy shot gather. The hyper-parameters of DnCNN are as follows: network depth 16, patch size 64×64, batch size 128, the size of convolutional kernel 3×3, the number of mappings 64, number of epochs 50, and learning rate $[10^{-3},10^{-5}]$. These parameters are set under the premise of obtaining the best denoised result. Figure 10 shows the curves of Wasserstein distance of WGAN and MSE loss of CNN_{WDRND} versus number of epochs.

We also compare the proposed CNN_{WDRND} with two conventional denoising methods: band-pass filter (BPF) and ensemble EMD (EEMD). BPF is a classical and high-efficiency seismic denoising method. It retains the frequency bands associated with signals and abandon those with noise, so as to achieve the separation of signals and noise. Compared with the classical EMD, EEMD can alleviate the mode aliasing to a certain extent.

We utilize CNN_{WDRND}, DnCNN, BPF, and EEMD to process the synthetic 2D shot gather. Figures 11(a)-(d) show the denoised results of the noisy CSP record with source location 600 (i.e., noisy CSP 2 in Fig. 9) by using the above four methods. The low and high frequencies of BPF are 15 Hz and 45 Hz; EEMD retains the 3rd and 4th modes, and the ratio of standard deviation is 0.55. In Fig. 11(c), EEMD can suppress most of the random noise, but the continuity of events in red rectangles is still poor. As shown in Fig. 11(d), there are obvious residual random noise in the denoised result by using BPF, and some recovered events in red rectangles are still difficult to identify. On the whole, the denoising performance of the two CNN-based methods is obviously superior to that of the two conventional methods. In Figs. 11(a) and (b), CNN_{WDRND} and DnCNN not only suppress most of the random noise and greatly increase the SNR of overall record, but also significantly enhance the continuity of events. The denoised results after applying CNN_{WDRND} and DnCNN indicate that the synthetic noise data obtained by using the WGAN augmentation

Fig. 12 The comparison of f-k spectrum. (**a**) and (**b**) are the f-k spectrum of clean CSP 2 in Fig. 8 and noisy CSP 2 in Fig. 9. (**c**)-(**f**) are the f-k spectrum of the denoised results in Figs. 11(a)-(d) by using CNNWDRND, DnCNN, EEMD, and BPF, successively. (g)-(j) are the f-k spectrum of the difference records in Figs. 11(e)-(h)
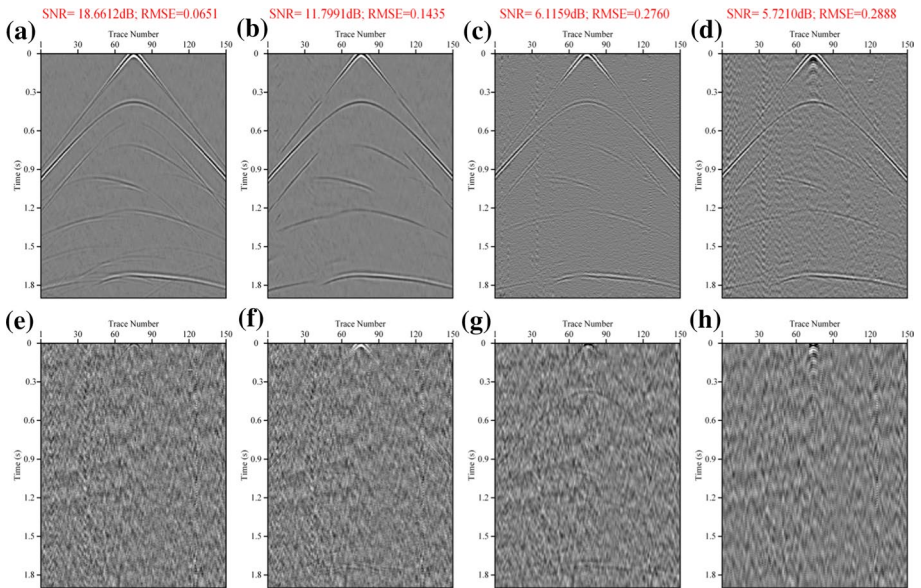
Fig. 13 Denoised results of the noisy CSP record with source location 1500. (**a**)-(**d**) are the denoised results by using CNNWDRND, DnCNN, EEMD, and BPF, successively. (**e**)-(**h**) are the corresponding difference records. Corresponding clean and noisy records are the clean and noisy CSP 4 shown in Fig. 8 and Fig. 9, respectively
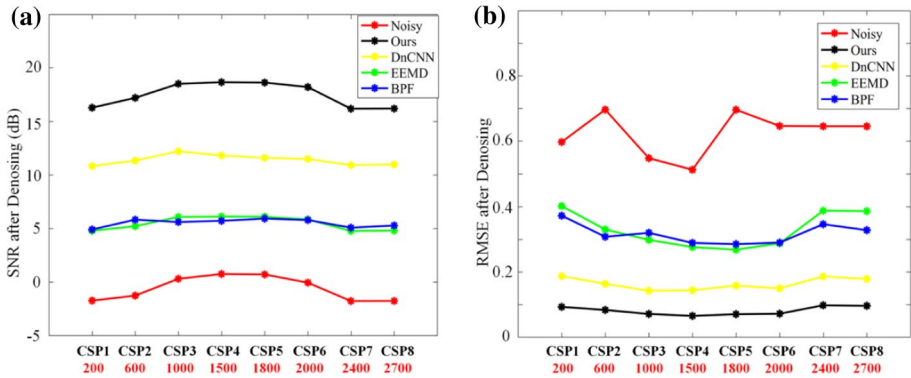
Fig. 14 (**a**) and (**b**) are the curves of SNR and RMSE after denoising, respectively. The red numbers are the source locations

Table 4 Average SNR increase (dB), denoising time–cost (s), and training time (hr)

Method	CNN$_{WDRND}$	DnCNN	EEMD	BPF
Average SNR increase	18.4912	12.0638	6.3017	6.1529
Average time–cost	3.24	3.61	130.52	1.82
Training time of D-CNN	5.27	7.63	N/A	N/A
Training time of WGAN	10.3	10.3	N/A	N/A

of pre-arrival noise is similar to the real noise data contaminating the desired signals; this similarity guarantees the performance of the two trained models derived from CNN$_{WDRND}$ and DnCNN. As shown in the areas indicated by red arrows in Fig. 11(a), CNN$_{WDRND}$ has stronger recovery ability for weak signals in comparison to DnCNN, demonstrating the positive effect of gradual denoising strategy and dilated convolution. Furthermore, SNRs and RMSE of the four denoised results are given in the red numbers in Fig. 11; the proposed CNN$_{WDRND}$ has the largest SNR and smallest RMSE.

Figures 11(e)-(h) show the difference records of CNN$_{WDRND}$, DnCNN, EEMD, and BFP, successively. As shown in Figs. 11(f)-(h), there are some residual signals in the difference records of DnCNN, EEMD, and BFP, demonstrating that the three methods damage the amplitude of signals when suppressing the random noise. Conversely, we hardly observe any residual signal in Fig. 11(e), so the proposed CNN$_{WDRND}$ achieves a excellent trade-off between noise suppression and signal preservation.

To further analyze the denoising performance, as shown in Fig. 12, we plot the f-k spectrum of the clean CSP 2 in Fig. 8, noisy CSP 2 in Fig. 9, four denoised results and their corresponding difference records in Fig. 11. As shown in Figs. 12(a) and (b), the signals and noise co-exist in about 0-20 Hz low-frequency band. The extreme similarity between Figs. 12(a) and (c) prove that the proposed CNN$_{WDRND}$ achieves a complete separation of signals and noise in the shared frequency band; also, we hardly observe signal component in Fig. 12(g). As shown in Fig. 12(h), the signal components in the red ellipse demonstrate that the DnCNN damages partially the signals when suppressing random noise. Figures 12(e) and (i) indicate the residual noise (yellow rectangles) and energy loss of signals (red circle) caused by EEMD. Figures 12(f) and (j) demonstrate the unavoidable signal leakage when using BPF.

Figure 13 displays the four denoised results of the noisy CSP record with source location 1500 (i.e., noisy CSP 4 in Fig. 9) and the corresponding four difference records. The low and high frequencies of BPF are adjusted to 12 Hz and 40 Hz; EEMD still retains the 3th and 4th modes and the ratio of standard deviation is adjusted to 0.75. As shown in Fig. 13, the denoising performance of CNN$_{WDRND}$ is still best in comparison to the other three methods; it can suppress most of the random noise and completely recover the weak signals. The DnCNN is not good at recovering the weak signals, and the continuity of events needs to be further strengthened. Although EEMD can remove lots of random noise, some obvious signals still exist in the difference record shown in Fig. 13(g). BPF only removes partial random noise and numerous events still show poor continuity.

5.3 The Overall Denoising Results of Shot Gather

Figure 14 displays the curves of SNRs and RMSE of the eight noisy CSP records in Fig. 9 after denoising by CNN$_{WDRND}$, DnCNNs, EEMD, and BPF. We discover that the proposed CNN$_{WDRND}$ can boost SNR by about 19 dB and always corresponds to the largest SNR and smallest RMSE.

Table 4 gives the average SNR increase of the eighteen noisy CSP records (i.e., noisy CSP 1–8 in Fig. 9) after denoising, the average denoising time–cost of four denoising methods, and the training time of CNN$_{WDRND}$ and DnCNN. The denoising time–cost of
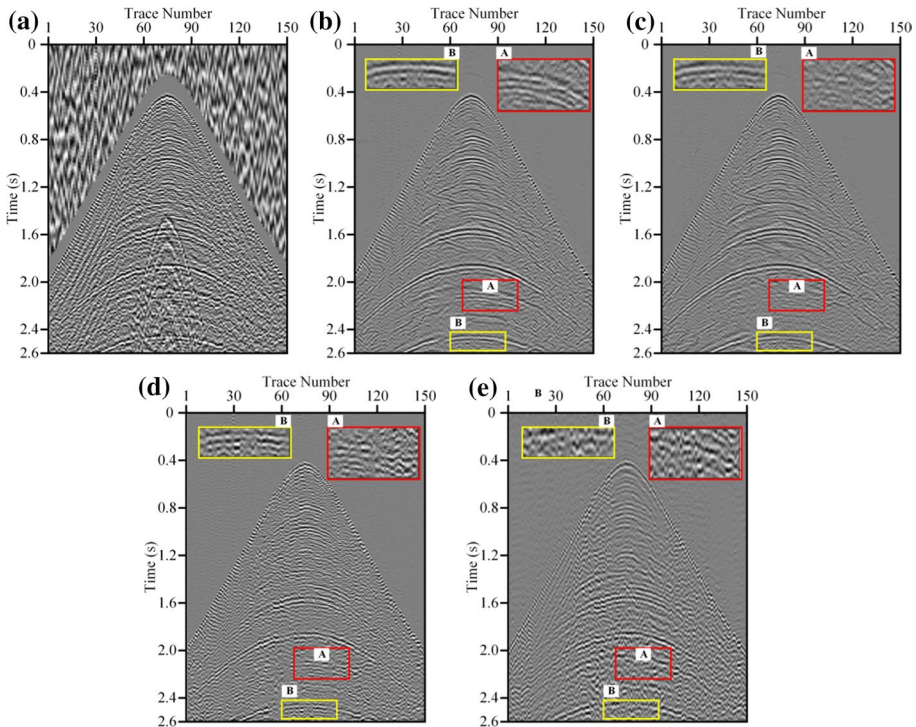
Fig. 15 The denoised records of real record 1. (**a**) Real record 1; (**b**)-(**e**) are the denoised records of real record 1 after applying CNNWDRND, DnCNN, EEMD, and BPF, respectively

the two CNN-based methods refers to the time–cost of prediction by using the trained models, excluding training time. From the statistics in Table 4, we can draw a conclusion that although CNN_{WDRND} and DnCNN need a long time to train the network, the two models after training can process different noisy CSP records in shot gather quickly and intelligently, without artificial parameter tuning. In contrast, EEMD and BPF need artificial parameter tuning to obtain the possible best denoising performance (largest SNR and smallest RMSE), in which EEMD retains the 3rd and 4th modes and the ratio of standard deviation ranges from 0.4 to 0.8; the low and high frequencies of BPF range from 8 to 15 Hz and 30 to 45 Hz, respectively. In conclusion, compared with traditional methods, SDL methods not only performs better in denoising, but also is more intelligent.
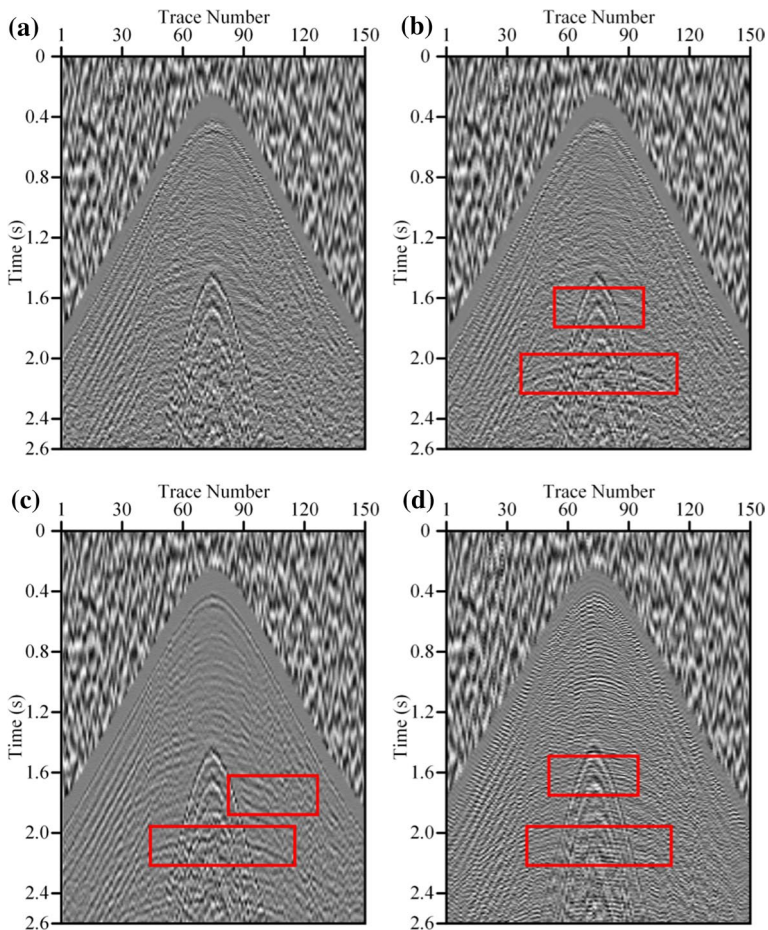

Fig. 16 The comparison of removed noise. (**a**)-(**d**) are the removed noise by using CNNWDRND, DnCNN, EEMD and BPF, successively

6 Real Example

In this paper, we leverage a real 3D shot gather to test the real application effect of CNN$_{WDRND}$. We perform the data augmentation on 1026 64×64 pre-arrival noise patches extracted from the real 3D surface shot gather, thereby obtaining 10,300 64×64 augmented synthetic noise patches. Then, we can obtain the trained models derived from CNN$_{WDRND}$ and DnCNN with the help of these 10,300 augmented synthetic noise patches (the 1026 pre-arrival noise patches extracted from the real 3D surface shot gather are not provided for network training) and the 8940 signals patches. The hyper-parameters of these two CNN-based are consistent with the above synthetic example.

Figure 15(a) shows a real CSP record in the 3D shot gather, named real record 1. The SNR of real record 1 is relatively low, and events are seriously contaminated by both random noise and surface waves. We utilize the CNN$_{WDRND}$, DnCNN, EEMD, and BPF to process the real record 1, and four denoised records are shown in Figs. 15(b)-(e). The denoised record of EEMD is the sum of 2nd, 3rd and 4th modes, and the ratio of standard deviation is 1. The low and high frequencies of BPF are 12 Hz and 28 Hz, respectively.

Next, we compare the denoising performance of the four methods from four aspects: random noise suppression, surface wave suppression, continuity of events, and signal preservation. In terms of random noise suppression, CNN$_{WDRND}$ and DnCNN can attenuate most of the random noise, and the improvement of SNR is visible; on the contrary, there is obvious residual random noise in the denoised records by using EEMD and BPF. For surface wave suppression, CNN$_{WDRND}$, DnCNN, and EEMD can effectively suppress the surface waves; some surface waves still remain in the denoised record by using BPF. In the denoised records by using BPF and EEMD, lots of residual background noise still interferes with the reflected signals, causing the poor continuity of events; whereas CNN$_{WDRND}$ and DnCNN can significantly enhance the continuity of events. As shown in the red and yellow rectangles of Figs. 15(b) to (e), CNN$_{WDRND}$ performs better in noise suppression in comparison to EEMD and BPF; moreover, it can recover some weak signals that cannot be
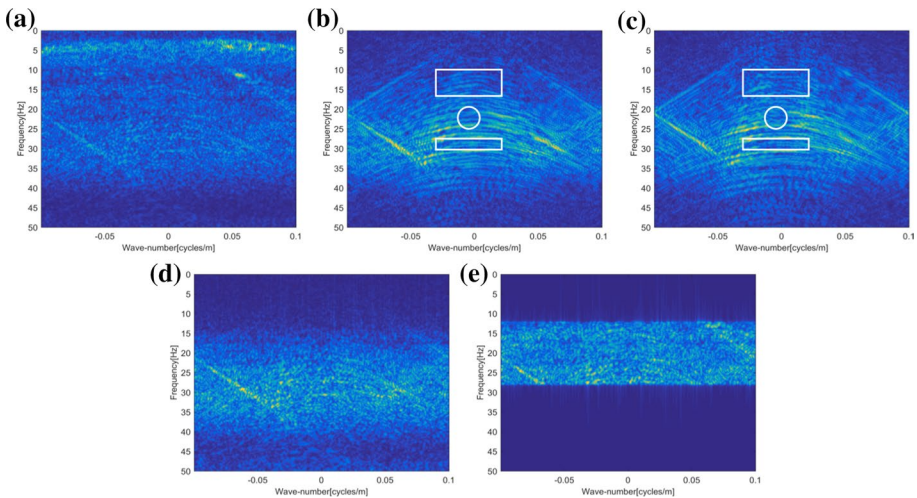
Fig. 17 The f-k spectrum of real record 1 and its four denoised records. (**a**) The f-k spectrum of real record 1; (**b**)-(**e**) are the f-k spectrum of denoised records of real record 1 by using CNNWDRND, DnCNN, EEMD and BPF, successively
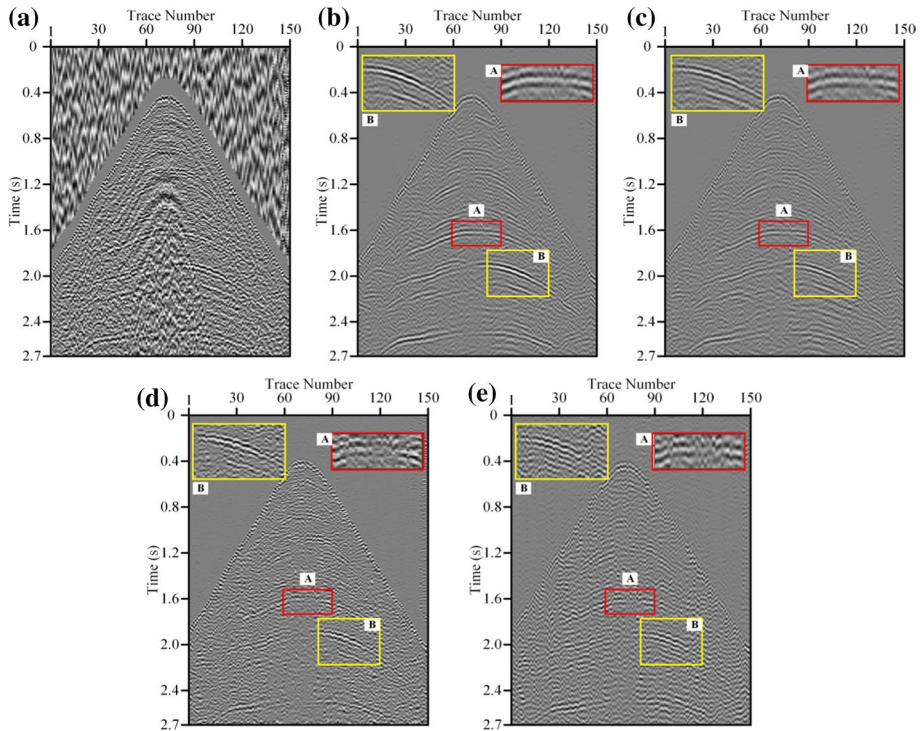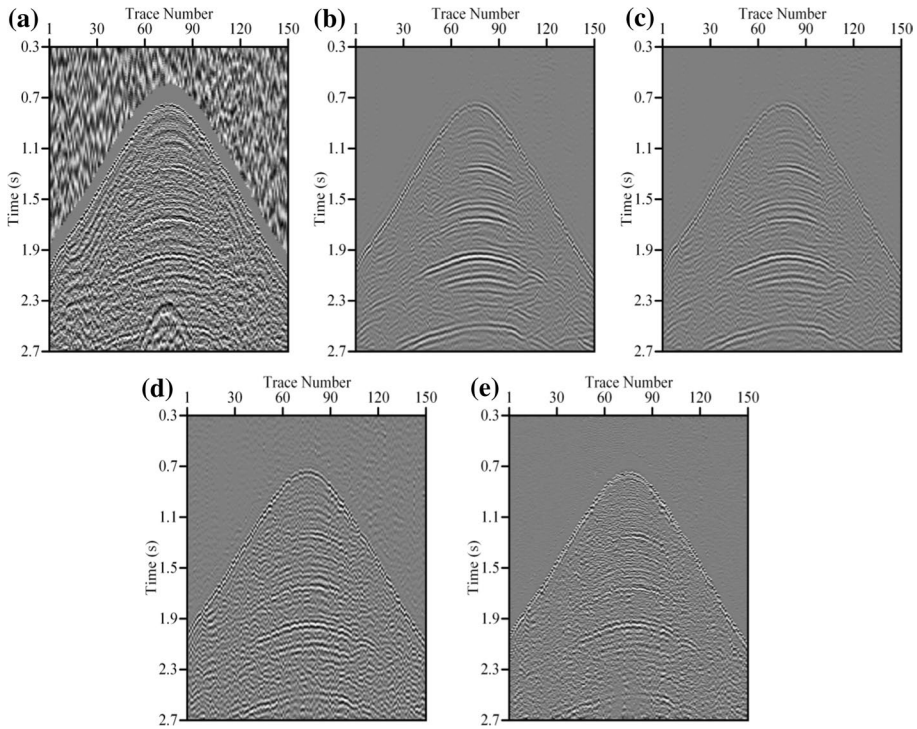
Fig. 18 The denoised records of real record 2. (**a**) Real record 2; (**b**)-(**e**) are the denoised records of real record 2 after applying CNNWDRND, DnCNN, EEMD, and BPF, respectively

recovered by DnCNN, proving that the gradual denoising strategy and dilated convolution can boost the signal recovery ability of the trained model.

To compare the signal preservation ability of different methods, we plot the removed noise by using the four denoising methods (Fig. 15a minus Figs. 15b-e, respectively) in Fig. 16. Obvious residual signals in the red rectangles prove the signal leakage caused by DnCNN, EEMD, and BPF. As shown in Fig. 16(a), the proposed CNN_{WDRDN} significantly reduces the degree of signal leakage, indicating less damage to signals when attenuating the random noise and surface waves.

We also plot the f-k spectrum of real record 1 and its four denoised records in Fig. 17. As shown in Fig. 17(a), the background noise almost completely drowned out signals in frequency domain. As shown in Fig. 17(d), some noise still remains in the denoised result by using EEMD and it cannot recover the low-frequency signals. Figure 17(e) once again demonstrates the unavoidable signal leakage after using BPF. Figures 17(b) and (c) suggest the superior denoising performance of the two deep-learning-based methods; they achieve complete separation of signals and noise in frequency-domain. By comparison, as shown in white rectangles and circles, signals in Fig. 17(b) is more continuous and clearer than those in Fig. 17(c), indicating better signal recovery ability of CNN_{WDRND} in comparison to DnCNN.

Figure 18(a) shows another CSP record in the real 3D shot gather, named real record 2. Compared with real record 1, the energy of random noise and surface waves in real

Fig. 19 The denoised records of real record 3. (**a**) Real record 3; (**b**)-(**e**) are the denoised records of real record 3 after applying CNNWDRND, DnCNN, EEMD, and BPF, respectively

record 2 is stronger; also, the distribution morphology of events is more diversified. Figures 18(b)-(e) are the denoised records of Fig. 18(a) after applying the four denoising methods. EEMD still keeps the 2nd, 3rd, and 4th modes, and the ratio of standard

**Fig. 20** The effect of network depth
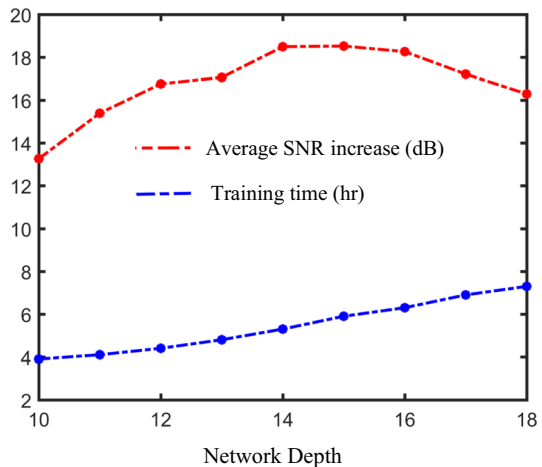
deviation is adjusted to 0.9; the low and high frequencies of BPF are 16 Hz and 28 Hz, respectively. Obviously, the denoising performance of EEMD and BPF is inferior to that of the two CNN-based denoising methods. A large amount of random noise still remains in Figs. 18(d) and (e); moreover, numerous events with weak energy are still hard to identify. As shown in Figs. 18(b) and (c), CNN$_{WDRND}$ and DnCNN remove almost all random noise and surface waves; the background of real record 2 becomes clear after applying the two CNN-based denoising methods. As shown in the partial enlargements (i.e., red and yellow rectangles), compared with DnCNN, CNN$_{WDRND}$ performs better in signal preservation, and some weak reflections recovered by CNN$_{WDRND}$ show better continuity.

Figure 19(a) is also a CSP record (named real record 3) extracted from the real 3D shot gather and the four denoised records are plotted in Figs. 19(b)-(e). The parameter settings of EEMD and BPF are as follows: retaining 2nd, 3rd, and 4th modes, the ratio of standard deviation 0.85, low frequency 13 Hz, and high frequency 31 Hz. The proposed CNN$_{WDRND}$ still performs better than the other three competitive methods. According to the denoised results of the above three CSP records, we can draw three conclusions: (1) the synthetic noise data augmented by GAN can meet the requirements of network training and reduce the dependence of SDL method on real noise data; (2) adopting the gradual denoising strategy and dilated convolution is beneficial to the recovery of weak signals and the continuity of events; (3) the trained model can efficiently process different CSP records in shot gather without parameter tuning, so CNN-based denoising methods are more intelligent than the conventional methods based on prior assumptions.

7 Discussion

7.1 The Effect of Network Depth

Appropriate network depth is crucial to the denoising performance of SDL methods. In general, adopting too small network depth cannot obtain the optimal denoised result and excessive network depth is likely to cause the phenomenon of over-fitting and some unnecessary time–cost. To discuss the effect of network depth (i.e., the number of sparse layers in D-CNN) on CNN$_{WDRND}$, we change the number of sparse layers and then utilize the trained models derived from D-CNN with different network depths to
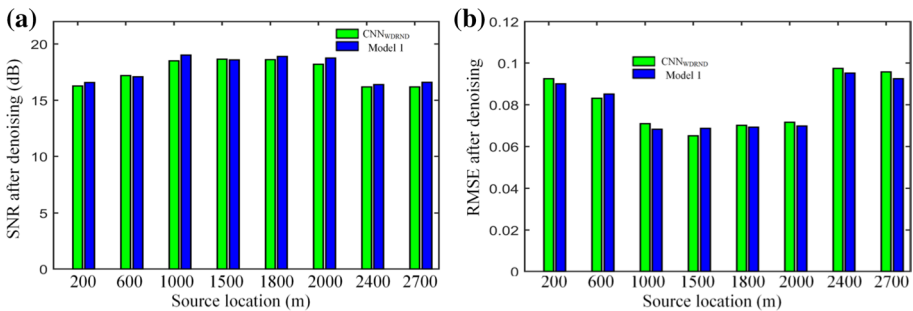
Fig. 21 The denoising performance comparison of CNNWDRND and model 1. (**a**) The comparison of SNR after denoising. (**b**) The comparison of RMSE after denoising
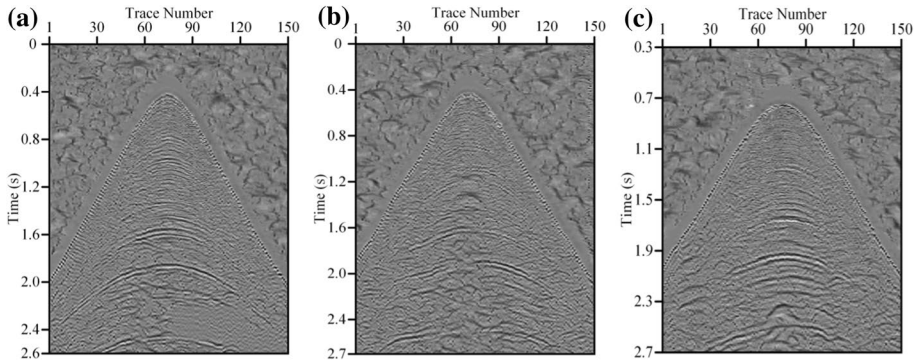
Fig. 22 The denoising performance of model 2. (**a**)-(**c**) are the denoised records of real record 1, 2, 3 after applying model 2, successively

Table 5 SNRs(dB) of synthetic noisy CSP records after denoising by CNN_{WDRND} and Model 3

Label of noisy records	CSP 1	CSP 2	CSP 3	CSP 4	CSP 5	CSP 6	CSP 7	CSP 8
CNN_{WDRND}	16.2768	17.1976	18.5092	18.6612	18.6259	18.2065	16.1782	16.1920
Model 3	6.4912	6.9726	8.4139	9.0617	8.2719	7.8017	6.7013	6.2773

process the eighteen noisy CSP records mentioned in the synthetic example. In Fig. 20, we plot the curve about the average SNR increase of the overall 2D synthetic shot gather and training time versus network depth. As the network depth increases from 10 to 14, the value of average SNR increases gradually, proving the enhancement of denoising performance. When the network depth is larger than 15, the training time–cost is still increasing, but the over-fitting phenomenon leads to the degradation of denoising performance revealed by the descending red curve.

Not just the network depth, some other hyper-parameters are likely to influence the denoising performance and training time–cost of D-CNN, like batch size, learning rate, patch size, and number of epochs. In this paper, the hyper-parameters of D-CNN and WGAN are determined by repeated experiments, which is time-consuming and laborious. This bottleneck will most likely be solved by the hyper-parameter optimization method which is an important part of our further research.

7.2 Can the Augmented Synthetic Noise Replace the Real Noise ?

In the above synthetic example, the random noise added to the clean records is extracted from a real seismic ambient noise record which can be approximated as a real random noise record. To explore whether the synthetic noise augmented by WGAN can replace the real noise, we extract 12,000 real random noise patches from the real seismic ambient record (note that these noise patches are not included in the random noise added to the clean records in the above synthetic example). These real random noise patches and

Table 6 RMSE of synthetic noisy CSP records after denoising by CNN$_{WDRND}$ and Model 3

Label of noisy records	CSP 1	CSP 2	CSP 3	CSP 4	CSP 5	CSP 6	CSP 7	CSP 8
CNN$_{WDRND}$	0.0926	0.0831	0.0709	0.0651	0.0701	0.0716	0.0975	0.0958
Model 3	0.2528	0.2481	0.2098	0.1952	0.2206	0.2326	0.2471	0.2731

the 8940 signal patches are utilized to train the D-CNN in Fig. 5, so as to obtain a trained model named model 1. By comparing the denoising performance of CNN$_{WDRND}$ and model 1, we can judge whether the augmented synthetic noise results in the performance degradation of trained model. We utilize the model 1 and CNN$_{WDRND}$ to process the eight synthetic noisy CSP records shown in Fig. 9; the SNRs and RMSE after denoising are shown in Fig. 20. Approximate SNRs and RMSE after denoising demonstrate that the augmented synthetic noise hardly degrades the denoising performance and indirectly prove that the probability distribution of the augmented synthetic noise data is extremely similar to that of the real random noise data (Fig. 21).

7.3 The Effect of Data Augmentation by Using WGAN

In general, it is optimal to train the network by leveraging the real noise corresponding to raw noisy seismic record. Unfortunately, this kind of noise data is often insufficient to meet the requirement of network training on data quantity, which greatly limits the further application of SDL methods. Thus, we utilize the WGAN to augment the pre-arrival noise, so as to alleviate the above problem. To demonstrate the effect of WGAN, we just use the 1026 pre-arrival noise patches adopted in the above real example to train the D-CNN without data augmentation, thereby obtaining a denoising model named model 2. The denoised records of real record 1, 2, 3 after applying model 2 are shown in Fig. 22. Both numerous discontinuous events and lots of residual background noise demonstrate the performance degradation compared to using the synthetic noise data augmented by WGAN to train the
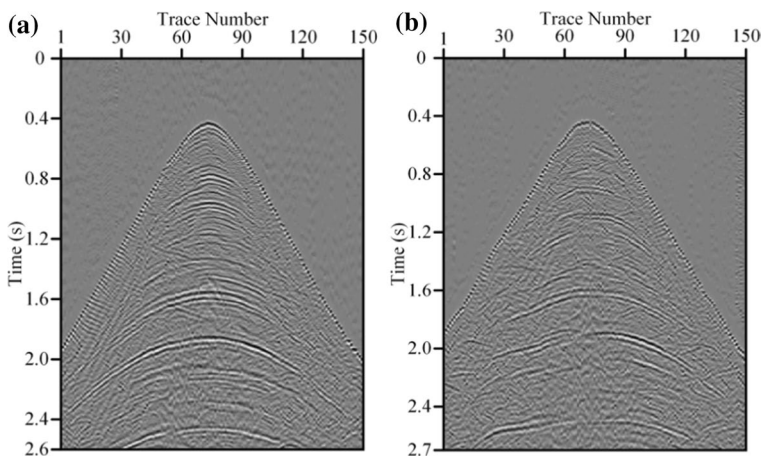
Fig. 23 The denoising performance of model 4. (**a**), (**b**) are the denoised records of real record 1 and 2 after applying model 4, respectively

D-CNN. This experiment also indirectly proves that the augmented synthetic noise data can meet the requirement of SDL methods on the quantity and authenticity of training data and the indispensability of WGAN.

In addition, to further quantify the effect of data augmentation, we just utilize the 964 real noise patches extracted from the 2D noisy shot gather in synthetic example to train the D-CNN, rather than the 9830 synthetic noise patches obtained by using data augmentation. After network training, we can obtain a denoising model named model 3. Then, model 3 and CNN$_{WDRND}$ are utilized to process the eight synthetic noisy CSP records in Fig. 9; the comparison of SNRs and RMSE after denoising is shown in Tables 5 and 6. Both the obvious decrease of SNR in Table 5 and the increase of RMSE in Table 6 demonstrate the necessity of data augmentation by using WGAN.

7.4 Why CNN$_{WDRND}$ can Remove Surface Waves From Real CSP Records

In noisy shot gather, the pre-arrival noise mainly from the surrounding environment can be approximated as ambient noise or called random noise. Thus, the synthetic noise data obtained by the augmentation of pre-arrival noise data is also characterized by random noise.

Since the augmented noise dataset is mainly consisted of random noise, the proposed CNN$_{WDRND}$ should suppress such noise, rather than both random noise and surface waves regarded as a typical kind of coherent noise in seismic data. However, as shown in Fig. 15(b) and Fig. 18(b), CNN$_{WDRND}$ shows good attenuation effect on surface waves. We suppose that this confusing phenomenon is mainly caused by the obvious differences between signals and surface waves in central frequency and wave velocity, especially the latter; specifically, the wave velocity ranges of surface wave and signal are often from 300 to 800 m/s and above 1200 m/s. As shown in Table 2, the range of wave velocity is set as 1300–5200 m/s when utilizing forward modeling to construct the signal dataset, so the trained model will judge that the surface waves are non-signal components and then remove them.

To testify the above viewpoint, we just change the range of wave velocity in forward modeling to 600–5200 m/s, and other situations remain the same. The trained model is named model 4. The denoised results of real record 1 and 2 (Fig. 15a and Fig. 18a) after applying model 4 are shown in Fig. 23(a) and (b), respectively. Some residual surface

Fig. 24 Visualization of 2000 noise patches adopted in real example (1000 pre-arrival noise patches and 1000 augmented synthetic noise patches) by using t-SNE
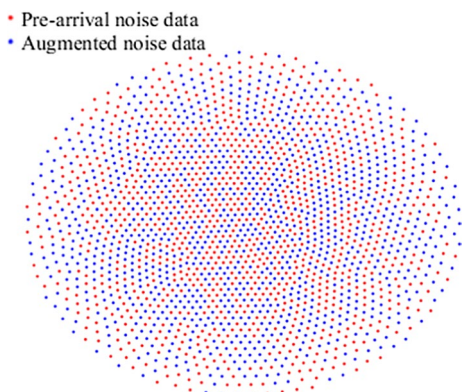
waves in Fig. 23 indicate the performance degradation of CNN$_{\text{WDRND}}$ on surface wave attenuation and the correctness of above viewpoint.

Although the proposed CNN$_{\text{WDRND}}$ effectively attenuates the surface waves in the real example, this augmentation strategy of pre-arrival noise is mainly applied to random noise suppression due to the fact that pre-arrival noise extracted from surface shot gather can be roughly regarded as random noise. The attenuation of coherent noise is also an essential topic in seismic data processing, so how to expand the application scope of data augmentation needs further consideration and exploration.

7.5 The Visualization of Probability Distribution

The premise of excellent denoising performance by using CNN$_{\text{WDRND}}$ is the extreme similarity between pre-arrival noise data and augmented synthetic noise data. The excellent denoising performance of CNN$_{\text{WDRND}}$ in synthetic and real examples have demonstrated this similarity.

t-SNE (Van der Maaten and Hinton 2008) is a good way to visually compare the probability distributions of different data. To further demonstrate the aforementioned similarity, we utilize t-SNE to compare the probability distributions of pre-arrival noise data and augmented noise data. Specifically, we randomly select 1000 noise patches from the 1026 pre-arrival noise patches and 10,300 augmented synthetic noise patches used in real example, respectively. Then, we perform visualization on the selected the 1000 pre-arrival noise patches and the 1000 augmented synthetic noise patches by using t-SNE; the corresponding result is shown in Fig. 24. The large overlap between red and blue dots once again demonstrates the similarity between pre-arrival noise data and augmented synthetic noise data in probability distribution.

8 Conclusions

In this paper, we develop a novel SDL method with weak dependence on real noise data, called CNN$_{\text{WDRND}}$. This proposed method utilize the data augmentation strategy of GAN to produce sufficient synthetic noise data used to train the D-CNN, which can alleviate the dependence on real noise data. Meanwhile, we adopt the gradual denoising strategy and dilated convolution to construct a novel architecture of D-CNN, so as to boost the recovery ability of trained models for weak desired signals. Both synthetic and real examples demonstrate that our method can effectively suppress the background noise and completely recover the weak reflections. Furthermore, the proposed CNN$_{\text{WDRND}}$ is driven by a large amount of training data and does not require the parameter fine-tuning when processing different CSP records in one shot gather. Therefore, the CNN$_{\text{WDRND}}$ is more intelligent than conventional methods.

References

Andrews HC, Patterson CL (1976) Singular value decomposition and digital image processing. IEEE Trans Acoust Speech Signal Process 24:26–53

Anvari R, Siahsar MAN, Gholtashi S, Roshandel Kahoo A, Mohammadi M (2017) Seismic random noise attenuation using synchrosqueezed wavelet transform and low-rank signal matrix approximation. IEEE Trans Geosci Remote Sens 55(11):6574–6581

Arjovsky M, Chintala S, Bottou L (2017) Wasserstein gan, arXiv preprint arXiv:1701.07875.

Battista BM, Knapp CC, McGee T, Goebel V (2007) Application of the empirical mode decomposition and Hilbert-Huang transform to seismic reflection data. Geophysics 72(2):H29–H37

Beckouche S, Ma J (2014) Simultaneous dictionary learning and denoising for seismic data. Geophysics 79(3):A27–A31

Bednar JB (1983) Applications of median filtering to deconvolution, pulse estimation and statistical editing of seismic data. Geophysics 48:1598–1610

Bekara M, van der Baan M (2007) Local singular value decomposition for signal enhancement of seismic data. Geophysics 72(2):V59–V65

Bekara M, van der Baan M (2009) Random and coherent noise attenuation by empirical mode decomposition. Geophysics 74(5):V89–V98

Birnie C, Ravasi M, Liu S, Alkhalifah T (2021) The potential of self-supervised networks for random noise suppression in seismic data. Artif Intell 2:47–59

Cadzow J (1988) Signal enhancement-a composite property mapping algorithm: IEEE transactions on acoustics. Speech and Signal Process 36:49–62

Canales L (1984) Random noise reduction. 54th Annual international meeting, SEG, expanded abstracts, 525–527.

Candes E, Guo F (2002) New multiscale transforms, minimum total variation synthesis: applications to edge-preserving image reconstruction. Signal Process 82(11):1519–1543

Candes E, Demanet L, Donoho D, Ying L (2006) Fast discrete curvelet transforms. Multiscale Model Simul 5(3):861–899

Chaudhari P, Agrawal H, Kotecha K (2020) Data augmentation using MG-GAN for improved cancer classification on gene expression data. Soft Comput 24(15):11381–11391

Chen YK (2020) Fast dictionary learning for noise attenuation of multidimensional seismic data. Geophys J Int 222(3):1717–1727

Chen YK, Ma JT (2014) Random noise attenuation by f-x empirical-mode decomposition predictive filtering. Geophysics 79(3):V81–V91

Chen YK, Zhou C, Yuan J, Jin ZY (2014) Applications of empirical mode decomposition in random noise attenuation of seismic data. J Seism Explor 23(5):481–495

Cheng J, Chen K, Sacchi MD (2015) Application of robust principal component analysis (RPCA) to suppress erratic noise in seismic records. SEG Tech Progr Expand Abstr 34:4646–4651

Cooper HW, Cook RE (1984) Seismic data gathering. Proc IEEE 72(10):1266–1275

Dian R, Li S, Kang X (2021) Regularizing hyperspectral and multispectral image fusion by CNN denoiser. IEEE Trans Neural Netw Learn Syst 32:1124–1135

Dong XT, Li Y (2021) Denoising the optical fiber seismic data by using convolutional adversarial network based on loss balance. IEEE Transactions on Geoscience and Remote Sensing. 59(12): 10544–10554

Dong XT, Jiang H, Zheng S, Li Y, Yang BJ (2019a) Signal-to-noise ratio enhancement for 3C downhole microseismic data based on the 3D shearlet transform and improved back-propagation neural networks. Geophysics 84(4):V245–V254

Dong XT, Li Y, Yang BJ (2019b) Desert low-frequency noise suppression by using adaptive DnCNNs based on the determination of high-order statistic. Geophys J Int 219(2):1281–1299

Dragomiretskiy K, Zosso D (2014) Variational modal decomposition. IEEE Trans Signal Process 62(3):531–544

Duncan G, Beresford G (1995) Median filter behaviour with seismic data. Geophys Prospect 43(3):329–345

Fomel S (2006) Towards the seislet transform. SEG Technical Program Expanded Abstracts 25(1):2847–2851

Gomez JL, Velis DR (2016) A simple method inspired by empirical mode decomposition for denoising seismic data. Geophysics 81(6):V403–V413

Gomez JL, Velis DR, Sabbione JI (2020) Noise suppression in 2D and 3D seismic data with data-driven sifting algorithms. Geophysics 85(1):V1–V10

Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. Adv Neural Info Process Syst 27:2672–2680

Gorszczyk A, Malinowski M, Bellefleur G (2015) Enhancing 3D post-stack seismic data acquired in hardrock environment using 2D curvelet transform. Geophys Prospect 63(4):903–918

Grossman A, Morlet J (1984) Decomposition of Hardy functions into square integrable wavelets of constant shape. SIAM J Math Anal 15(4):723–736

Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville A (2017) Improved training of wasserstein GANs. Statistic 22(3):1467–1477

Gulunay N (2000) Noncausal spatial prediction filtering for random noise reduction on 3-D poststack data. Geophysics 65:1641–1653

Gulunay N (2017) Signal leakage in f-x deconvolution algorithms. Geophysics 82(5):W31–W45

Gulunay N (1986) FX decon and complex Wiener prediction filter: 56th Annual international meeting, SEG, Expanded abstracts 279–281.

Guo K, Labate D (2007) Optimally sparse multidimensional representation using shearlets. SIAM J Math Anal 39(1):298–318

Guo C, Zhu T, Gao Y, Wu S, Sun J (2021) AEnet: automatic picking of P-wave first arrivals using deep learning. IEEE Trans Geosci Remote Sens 59(6):5293–5303

Hagen DC (1982) The application of principal components analysis to seismic data sets. Geoexploration 20:93–111

Han J, van der Baan M (2015) Microseismic and seismic denoising via ensemble empirical mode decomposition and adaptive thresholding. Geophysics 80(6):KS69–KS80

Harris PE, White RE (1997) Improving the performance of f-x prediction filtering at low signal-to-noise ratios. Geophys Prospect 45(2):269–302

He HQ, Wang WY (2020) Reparameterized full-waveform inversion using deep neural networks. Geophysics 86(1):V1–V13

Herrmann F, Hennenfent G (2008) Non-parametric seismic data recovery with curvelet frames. Geophys J Int 173:233–248

Hinton GE, Osindero S, Teh Y (2006) A fast learning algorithm for deep belief nets. Neural Comput 18:1527–1554

Huang NE, Shen Z, Long SR, Wu ML, Shih HH, Zheng Q, Yen NC, Tung CC, Liu HH (1998) The empirical mode decomposition and hilbert spectrum for nonlinear and non-stationary time series analysis. Proc R Soc Lond A 454:903–995

Huang WL, Wang RQ, Zu SH, Chen YK (2020) Low-frequency noise attenuation in seismic and microseismic data using mathematical morphological filtering. Geophys J Int 222(3):1728–1749

Jing L, Tian Y (2021) Self-supervised visual feature learning with deep neural networks: a survey. IEEE Trans Pattern Anal Mach Intell 43(11):4037–4058

Kaplan ST, Sacchi MD, Ulrych TJ (2009) Sparse coding for data-driven coherent and incoherent noise attenuation. 79th Annual international meeting, SEG, Expanded abstracts, 3327–3331.

Kaur H, Pham N, Fomel S (2021) Seismic data interpolation using deep learning with generative adversarial networks. Geophys Prospect 69(2):307–326

Kazei V, Ovcharenko O, Plotnitskii P, Peter D, Zhang X, Alkhalifah T (2021) Mapping full seismic waveforms to vertical velocity profiles by deep learning. Geophysics 86(5):R711–R726

Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 25:1097–1105

Krohn C, Ronen S, Deere J, Gulunay N (2008) Introduction to this special section: seismic noise. Lead Edge 27(2):163–165

Lari HH, Naghizadeh M, Sacchi MD, Gholami A (2019) Adaptive singular spectrum analysis for seismic denoising and interpolation. Geophysics 84(2):V133–V142

Lecun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86:2278–2324

Lecun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444

Lemarchand F, Findeli T, Nogues E, Pelcat M (2020) NoiseBreaker: gradual image denoising guided by noise analysis. IEEE 22nd International workshop on multimedia signal processing, MMSP https://doi.org/10.1109/MMSP48831.2020.9287095.

Liu G, Chen X, Du J, Wu K (2012) Random noise attenuation using f-x regularized nonstationary autoregression. Geophysics 77(2):V61–V69

Liu Q, Kampffmeyer M, Jenssen R, Salberg A (2020) Dense dilated convolutions' merging network for land cover classification. IEEE Trans Geosci Remote Sens 58(9):6309–6320

Liu XY, Chen XH, Li JY, Chen YK (2021) Nonlocal weighted robust principal component analysis for seismic noise attenuation. IEEE Trans Geosci Remote Sens 59(2):1745–1756

Lu Y, Mak M (2020) Improving speech emotion recognition with adversarial data augmentation network. IEEE transaction on neural networks and learning systems. Early access.

Mallat SG, Zhang Z (1993) Matching pursuits with time-frequency dictionaries. IEEE Trans Signal Process 41(12):3397–3415

Matheron G (1975) Random sets and integral geometry. John Wiley & Sons

Meng F, Fan Q, Li Y (2022) Self-supervised learning for seismic data reconstruction and denoising. IEEE Geosci Remote Sens Lett 19:1–5

Moreno-Barea FJ, Jerez JM, Franco L (2020) Improving classification accuracy using data augmentation on small data sets. Expert Syst Appl 161(15):113696–113709

Mousavi SM, Langston CA (2016) Hybrid seismic denoising using higher-order statistics and improved wavelet block thresholding. Bull Seismol Soc Am 106(4):1380–1393

Naghizadeh M, Sacchi MD (2018) Ground-roll attenuation using curvelet downscaling. Geophysics 83(3):V185–V195

Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381(13):607–609

Oropeza V, Sacchi MD (2011) Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis. Geophysics 76(3):V25–V32

Paszke A, Gross S, Massa F, et al (2019) PyTorch: An imperative style, high-performance deep learning library. Adv Neural Inf Process Syst 32.

Pochet A, Diniz PHB, Lopes H, Gattass M (2019) Seismic fault detection using convolutional neural networks trained on synthetic poststacked amplitude maps. IEEE Geosci Remote Sens Lett 16(3):352–356

Remez T, Litany O, Giryes R, Bronstein AM (2018) Class-aware fully convolutional gaussian and poisson denoising. IEEE Trans Image Process 27(11):5707–5722

Romano Y, Elad M (2015) Boosting of image denoising algorithms. SIAM J Imag Sci 8(2):1187–1219

Saad OM, Chen Y (2020) Deep denoising autoencoder for seismic random noise attenuation. Geophysics 85(4):V367–V376

Saad OM, Chen Y (2021) A fully unsupervised and highly generalized deep learning approach for random noise suppression. Geophys Prospect 69(4):709–726

Sacchi M, Porsani M (1999) Fast high resolution parabolic radon transform. 89th Annual international meeting, SEG, Expanded abstracts 1477–1480.

Sahoo SK, Makur A (2013) Dictionary training for sparse representation as generalization of k-means clustering. IEEE Signal Process Lett 20(6):587–590

Schmarje L, Santarossa M, Schroder S, Koch R (2021) A survey on semi-, self- and unsupervised learning for image classification. IEEE Access 9:82146–82168

Schultz P, Claerbout J (1978) Velocity estimation and downward continuation by wavefront synthesis. Geophysics 43:691–714

Serra J (2011) Historical overview of image analysis and mathematical morphology. Pattern Recognit Image Anal 21(2):167

Trickett S (2008) F-Xy Cadzow noise suppression: CSPG CSEG CWLS convention. Abstracts 27:303–306

Tukey, (1974) Nonlinear (nonsuperposable) methods for smoothing data. Proc Congr Rec EASCON 74:673–681

Van Der Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9:2579–2625

Vautard R, Ghil M (1989) Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series: Physica D. Nonlinear Phenomena 5:395–424

Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol P (2010) Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J Mach Learn Res 11:3371–3408

Wang X, Wen B, Ma J (2019) Denoising with weak signal preservation by group-sparsity transform learning. Geophysics 84(6):V351–V368

Wang C, Zhu Z, Gu H (2020a) Low-rank seismic denoising with optimal rank selection for Hankel matrices. Geophys Prospect 68(3):892–909

Wang YH, Liu XW, Gao FX, Rao Y (2020b) Robust vector median filtering with a structure-adaptive implementation. Geophysics 85(5):V407–V414

Wright J, Peng Y, Ma Y, Ganesh A, Rao S (2009) Robust principal component analysis: exact recovery of corrupted low-rank matrices by convex optimization. Advances in neural information processing systems 22 - proceedings of the 2009 conference 2080–2088.

Xue YJ, Cao JX, Wang XJ, Li YX, Du J (2019) Recent developments in local wave decomposition methods for understanding seismic data: application to seismic interpretation. Surv Geophys 40(5):1185–1210

Yang Q, Yan P, Zhang Y, Yu H, Shi Y, Mou X, Kalra M, Zhang Y, Sun L, Wang G (2018) Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss. IEEE Trans Med Imaging 37(6):1348–1357

Yang L, Chen W, Wang H, Chen Y (2021) Deep learning seismic random noise attenuation via improved residual convolutional neural network. IEEE Trans Geosci Remote Sens 59(9):7968–7981

Yu S, Ma J, Wang W (2019) Deep learning for denoising. Geophysics 84(6):V333–V350

Yu F, Koltun V (2016) Multi-scale context aggregation by dilated convolutions. 4th international conference on learning representations, ICLR 2016 - conference track proceedings.

Zhang Z, Alkhalifah T (2019) Regularized elastic full-waveform inversion using deep learning. Geophysics 84(5):R741–R751

Zhang C, van der Baan M (2018) Multicomponent microseismic data denoising by 3D shearlet transform. Geophysics 83(3):A45–A51

Zhang K, Zuo W, Chen Y, Meng D, Zhang L (2017) Beyond a gaussian denoiser: residual learning of deep CNN for image denoising. IEEE Trans Image Process 26(7):3142–3155

Zhang Z, Wang X, Jung C (2019) DCSR: dilated convolutions for single image super-resolution. IEEE Trans Image Process 28(4):1625–1635

Zhao Y, Niu FL, Zhang ZH, Li X, Chen JH, Yang JD (2020) Signal detection and enhancement for seismic crosscorrelation using the wavelet-domain Kalman filter. Surv Geophys 42(1):43–67

Zhong T, Li Y, Wu N, Nie PF, Yang BJ (2015) A study on the stationarity and Gaussianity of the background noise in land-seismic prospecting. Geophysics 80(4):V67–V82

Zhong T, Cheng M, Dong XT, Wu N (2022) Seismic random noise attenuation by applying multi-scale denoising convolutional neural network. IEEE Trans Geosci Remote Sens 60:1–13

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Xintong Dong[1] · Jun Lin[1] · Shaoping Lu[2,3] · Xingguo Huang[1] · Hongzhou Wang[1] · Yue Li[4]

✉ Xingguo Huang
   xingguohuang@jlu.edu.cn

[1] College of Instrumentation and Electrical Engineering, Jilin University, Changchun 130026, Jilin, China

[2] School of Earth Sciences and Engineering, Sun Yat-Sen University, Guangzhou 510275, Guangdong, China

[3] Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519000, Guangdong, China

[4] College of Communication Engineering, Jilin University, Changchun 130012, Jilin, China