

Generative Adversarial Network for Desert Seismic Data Denoising

Hongzhou Wang[✉], Yue Li[✉], and Xintong Dong[✉]

Abstract—Seismic exploration is a kind of exploration method for oil and gas resources. However, the disturbance of numerous random noise will decrease the quality and signal-to-noise ratio (SNR) of real seismic records, which brings difficulties to the following works of processing and interpretation. The seismic records of desert region pose a particular problem because of the strong energy noise and the spectrum overlapping between effective signals and random noise. Recent research works demonstrate that a convolutional neural network (CNN) can increase the SNR of seismic records. The optimization of denoising methods based on CNN is principally driven by the loss functions that largely focus on minimizing the mean-squared reconstruction error between denoising records and theoretical pure records. The denoising results estimated by the CNN model are often lacking the perfection of the signal structure. Therefore, when processing seismic records with low SNR, the denoising results often have a lack of effective signal in some traces, which leads to the poor continuity of events. In order to solve this problem, we adopt the strategy of generative adversarial network (GAN) to construct a GAN for denoising. It is divided into two parts: the generator (the denoising network based on CNN) is used to remove noise, while the discriminator is used to guide the generator to restore the structure information of effective signals. The generator and discriminator enhance the performance of each other through adversarial training, and the generator after adversarial training can greatly recover events and suppress random noise in synthetic and real desert seismic data.

Index Terms—Adversarial training, convolutional neural network (CNN), desert seismic data, generative adversarial network (GAN), low signal-to-noise ratios (SNRs).

I. INTRODUCTION

SEISMIC exploration is a kind of important way for oil and gas exploration. It is a geophysical exploration method, which uses the response of artificial seismic waves to infer the lithology and structure of stratum by analyzing wave velocity, travel time, angles of reflection and refraction, and so on. However, this response received by geophones, i.e., seismic data, is often interfered by a large amount of random noise [1]–[4]. Therefore, how to suppress seismic random noise and enhance the signal-to-noise ratio (SNR) of seismic data becomes an important topic, which is of great

Manuscript received January 13, 2020; revised July 26, 2020 and September 14, 2020; accepted October 9, 2020. Date of publication November 2, 2020; date of current version July 22, 2021. This work was supported by the National Natural Science Foundations of China under Grant 41730422. (Hongzhou Wang and Xintong Dong contributed equally to this work.) (Corresponding author: Yue Li.)

The authors are with the Department of Information, College of Communication Engineering, Jilin University, Changchun 130012, China (e-mail: wanghz18@mails.jlu.edu.cn; liyue@jlu.edu.cn; 18186829038@163.com).

Digital Object Identifier 10.1109/TGRS.2020.3030692

significance to analyze stratum structure and explore oil, gas, and mineral resources [1], [2].

As the exploration environment becomes more and more complex, the denoising of seismic data becomes more and more difficult, especially in desert region. Compared with mountain, hilly, grassland, plain, and so on, the desert region poses two obvious problems. First, the energy of desert low-frequency noise is stronger because the desert region is relatively empty, so effective signals are completely submerged by random noise [6], [7], [9]. Second, sand is regarded as the main medium of background noise propagation in the desert region. It can absorb most high-frequency noise due to its selective absorption characteristic [6]. Thus, the dominant frequencies of desert random noise are relatively low [7]. Meanwhile, the effective signals are also mainly in the low-frequency band, so the spectrum overlapping between effective signals and noise is more serious in the desert region [3], [5]–[7]. At the same time, the surface waves also severely disrupt the continuity of effective signals in seismic data. The above noise seriously reduces the quality of seismic data. From what we have mentioned earlier, the denoising of desert seismic data is a huge challenge for us. Recently, experts have proposed various denoising methods to enhance the SNR of seismic data, such as F-X deconvolution [10], [11], wavelet transform [12], curvelet transform [13], shearlet transform [2], empirical mode decomposition (EMD) [14], [15], and time–frequency peak filtering (TFPF) [16], [17]. These denoising methods have good adaptability to conventional regions, but they have little effect on desert seismic data because of its low SNR and serious spectrum overlapping. In order to meet the requirement of desert seismic exploration, we urgently need an effective denoising method for desert seismic data.

In recent years, the methods based on deep learning have attracted much attention due to the rapid development of graphics processing units (GPUs), which increases the computing power and greatly shortens the training time. Among different types of deep neural networks, convolutional neural network (CNN) has been most extensively studied and achieved the state-of-the-art results on various tasks, such as image identification [20], image segmentation [21], and image reconstruction [22]. CNN, as a learning method, can map data into nonlinear space and automatically extract potential features [20]–[26]. These features are more effective and more robust to seismic data containing random noise. Recent research works show that these CNN-based methods play an essential role in suppressing the complex random noise of seismic data [27], [28]. Depending on the complex system and huge parameters, CNN can learn the complex mapping from

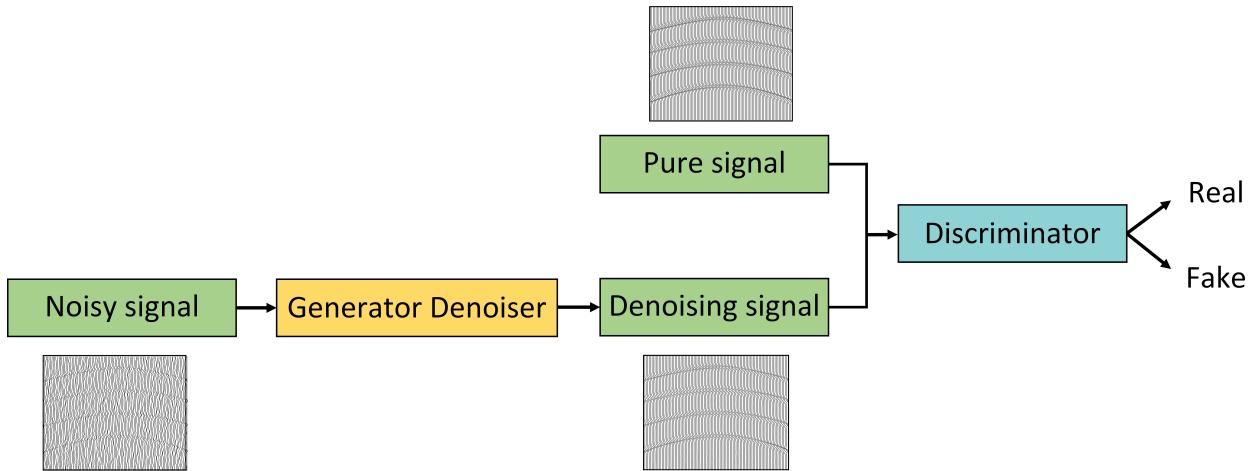


Fig. 1. Adversarial network architecture for denoising.

noisy seismic data to theoretical pure data, so as to recover the effective signals to a great extent. The optimization process of CNN-based methods is usually to minimize the mean square error (mse) between the denoising data and the original pure data [27], [28]. However, the ability of mse to capture perceptually relevant differences, such as structural integrity, is very limited as they are defined based on each sampling point, so this optimizing method is lack of the description of data distribution, especially the structural characteristics. Thus, when processing the desert seismic data with low SNR by based-CNN methods, some effective signals are seriously attenuated in some traces of recovered desert seismic data. This phenomenon greatly decreases the continuity of events and the quality of seismic data.

In order to completely recover the effective signals and suppress desert random noise, we adopt the strategy of generative adversarial network (GAN) and combine CNN with it. The GAN mainly consists of generator and discriminator, and the expected output is obtained by confrontation between the two [29]. GAN has gradually applied to multiple areas, such as image generation [29], super-resolution of image [30], and image translation [31]. In this article, the CNN denoising network is used as the generator of GAN. The CNN can learn the mapping from noisy desert data (the sum of theoretical pure data and desert random noise) to denoising desert data, and the discriminator is trained to distinguish the denoising desert data from theoretical pure data. The result of discriminator (true or false) can guide the CNN denoising network to optimize its parameters, so as to improve the quality of denoising seismic data. Until the discriminant result is true, the discriminator cannot distinguish the denoising desert data from the theoretical pure data, and therefore, such CNN denoising network can completely recover effective signal and suppress desert random noise. In order to train this denoising GAN for desert seismic data, we define a new loss function that consists of reconstruction loss and adversarial loss. The reconstruction loss based on mse is used to achieving particularly high SNR of denoising seismic data. Also, the adversarial loss pushes the denoising estimates to the theoretical seismic signal manifold using a discriminator network that is trained to differentiate

between the denoising seismic data and the theoretical seismic data. Reconstruction loss and adversarial loss are used to ensure both the high SNR and structural integrity of denoising seismic data.

Our contribution lies in redesigning a denoising GAN scheme based on the characteristics of seismic data. By readjusting and modifying the traditional GAN from the three perspectives of training set construction, network structure, and loss function, the adjusted GAN can be more suitable for desert data denoising, especially with better ability to repair events. The trained generator denoiser of GAN is used to process simulated and field seismic records. Compared with the traditional seismic denoising method, the method in this article has an obvious suppression effect on desert noise. At the same time, we also trained an original denoising CNN using mse loss alone. Through comparison, it is found that the denoising CNN under adversarial training has a better ability to recover events while suppressing low-frequency noise. It encourages adversarial learning as a training scheme for more deep-learning-based denoising models. In addition, compared with the original GAN [43]–[45], our scheme also has advantages in recovering events and accelerating model convergence.

The rest of this article is organized as follows. In Section II, we introduce our method. Then, the denoising results of synthetic data and field data are shown in Sections III and IV, respectively. Finally, Section V concludes this article.

II. DENOISING GENERATIVE ADVERSARIAL NETWORK

This article proposed a denoising architecture based on the GANs to solve the denoising problem of field seismic records in the desert region. We first give an overview of the general architecture, then provide details on the learning procedure, including the design of loss function and construction of training set, and finally present the principle of denoising.

A. Adversarial Network Architecture

The adversarial network architecture consists of a generator denoiser and a discriminator, as shown in Fig. 1. Among them,

Generator Denoiser

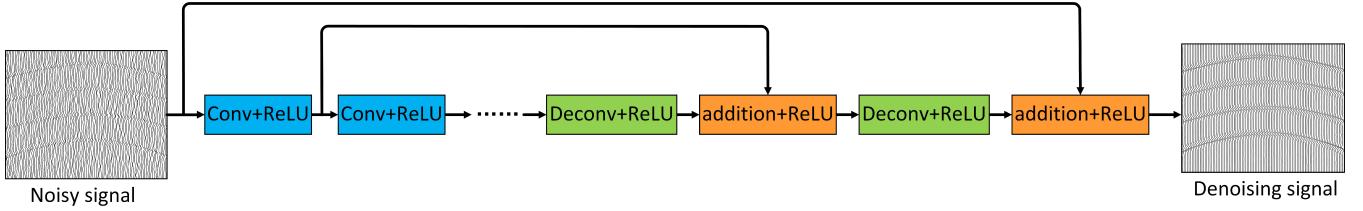


Fig. 2. Network architecture of the generator denoiser.

TABLE I
DEFINITION OF DIFFERENT LAYERS

Layer	Description	Function
Conv	Convolution	High-dimensional features extraction for low-dimensional feature vectors
Deconv	Deconvolution	Reconstruction of low-dimensional vectors from high-dimensional vectors
ReLU	Rectified Linear Unit	$\max(0, \cdot)$
LeakyReLU	Leaky Rectified Linear Unit	$\max(0, \cdot) + leak \times \min(0, \cdot)$
Addition	Addition layer	Add feature maps

the generator denoiser is used to learn the end-to-end mapping from the noisy signal to the theoretical clean signal, and it can generate a denoising signal according to the noisy input. Next, the denoising signal and the theoretical clean signal are used together as the input of discriminator. The discriminator determines whether the input signal is denoised or clean. The generator denoiser continuously improves the ability to suppress noise and recover signals so that the denoised signal is as close as possible to the clean signal, and at the same time, the discriminator continuously improves its discriminating ability. Finally, when the discriminator cannot determine whether the input is a clean signal or a denoised signal, the generator denoiser becomes the appropriate denoising network we need. Next, the architecture of the generator denoiser and the discriminator is introduced separately.

The generator denoiser is derived from the RED-Net architecture [32], as shown in Fig. 2. The definition of each layer is shown in Table I. Conv represents the convolution layer whose function is to extract the abstract features of the input signal while eliminating the noise. Deconv denotes the deconvolution layer whose function is the inverse process of Conv and it can inversely synthesize data from features [35]. ReLU is a commonly used activation function with nonlinearity principles [36], [40]. The generator denoiser consists of multiple convolutional layers and deconvolution layers. The kernel size for convolution and deconvolution is set to 3×3 . Also, we use 64 feature maps for convolutional and deconvolutional layers. The number of convolutional layers and deconvolutional layers is each set to 15. Deriving from the above architecture, in this work, we mainly conduct experiments with the generator denoiser with 30 processing layers. ReLU is added after each convolutional layer and deconvolution layer to achieve a

nonlinear mapping. Skip connections are added symmetrically from a convolutional layer to its corresponding mirrored deconvolutional layer. The convolutional feature maps are passed to and summed with the deconvolutional feature maps elementwise and passed to the next layer after rectification. Skip connections can bring more performance gains to the denoising model. First, the skip connections allow the gradient to be backpropagated to bottom layers directly and thus tackles the problem of gradient vanishing, making training deep networks easier. Second, as the number of processing layers increases, the feature map represents a higher level of features and the details of the signal may be lost. The skip connections can pass signal details from convolutional layers to deconvolutional layers, which is beneficial in recovering the original signal.

The discriminator is shown in Fig. 3, and the definition of each layer is also listed in Table I. The discriminator is mainly composed of multiple convolutional layers. The convolution layer is responsible for feature extraction to capture the abstract information of the input signal. By continuously extracting features from the input signal, the discriminator finally determines the specific category of the input data. The filter number of the discriminator from the first to the fifth unit is 6, 128, 256, 512, and 1024, respectively. The kernel size for these convolutions is set to 3×3 . Finally, a convolution layer with 1×1 convolution kernels is used to dimension reductionality. The filter number of the last convolutional layer is 1.

The generator denoiser receives the noisy input as an encoding to generate a denoised output. At the same time, due to the constraints of the discriminator, the specific category of the signal can be discriminated so that the final output of

Discriminator

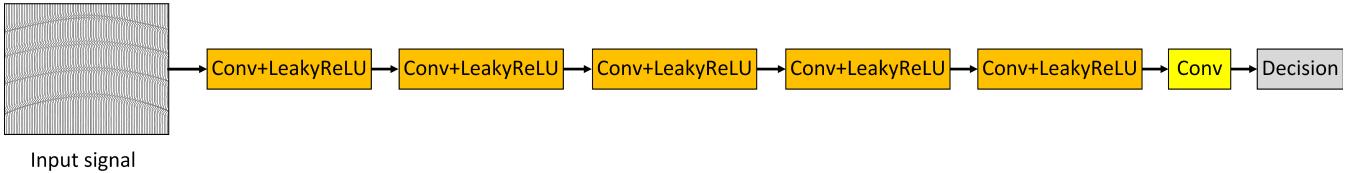


Fig. 3. Network architecture of the discriminator.

the generator denoiser tends to retain a fine signal structure. Finally, the generator denoiser can better restore the structure of the events of effective signal while improving the SNR of the record.

B. Loss Function

Models based on deep learning have a huge amount of parameters. The optimization algorithm can be used to update the parameters of the model to achieve the mapping we need. The loss function is the optimization goal of training neural networks. In general, it can help to judge the performance of model during training and can provide gradient guidance model training in backpropagation.

The definition of the loss function is critical to the performance of the denoising network. The general denoising network is based on mse modeling as shown in (1) in order to make the network's estimate have a high SNR. However, a pixel-size loss function, such as mse, is difficult to guide the model to restore a fine structure of the signal. In general, there are multiple equally plausible ways to reconstruct a signal region that is severely disturbed by noise while also maintaining coherence with the given context. The mse loss is responsible for capturing the overall structure of the noisy region and coherence with regard to its context but tends to average together the multiple modes in predictions. Minimizing the mse makes it difficult for the model to recover the specified signal structure and thus have poor perceptual quality. Also, the adversarial loss can make up for this shortcoming. The adversarial loss tries to make the prediction look real and has the effect of picking a particular mode from the distribution. As shown in Fig. 4, discontinuous seismic events and false seismic events typically occur in the denoising records of mse-based models when processing low SNR records. A model based on mse loss and adversarial loss tends to avoid the occurrence of the situation.

We have improved the loss function. A new loss function, including reconstruction loss based on mse and adversarial loss based on WGAN-GP [34], is proposed to implement a solution with respect to both SNR and perceptually relevant characteristics. Next, the reconstruction loss and the adversarial loss are introduced separately, and finally, the final loss is proposed.

1) Reconstruction Loss: We used the mse as the reconstruction loss function, as shown in the following equation:

$$\mathcal{L}_{\text{rec}} = \frac{1}{2N} \sum_{i=1}^N \|G(y_i; \Theta) - x_i\|_F^2 \quad (1)$$

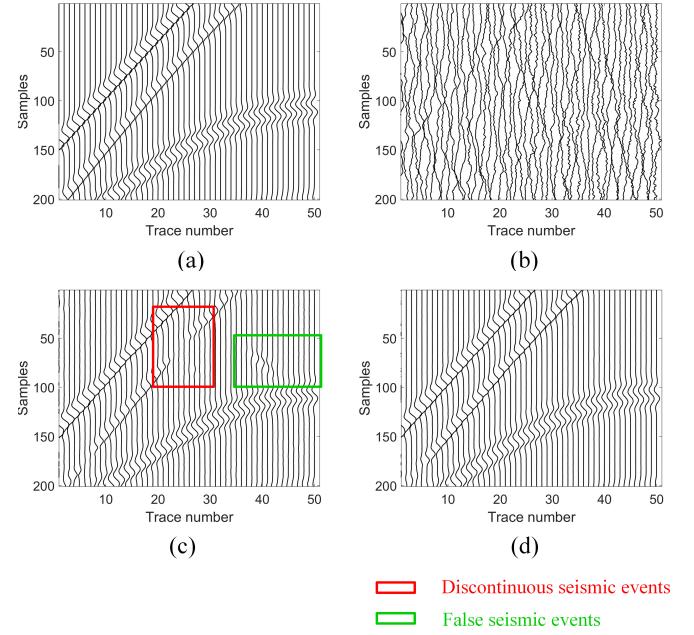


Fig. 4. Local magnification of the denoised signal of models trained with different loss functions. (a) Theoretical seismic data. (b) Noisy seismic data. (c) Denoised signal of the model trained with mse loss. (d) Denoised signal of the model trained with mse loss and adversarial loss.

where G represents the mapping fit by the generator denoiser, $\{x_i, y_i\}_{i=1}^N$ is the training set and construction of it is mentioned in Section II-C. $(Gy_i; \Theta)$ is the predicted output of the network when the input is y_i . x_i is the label of the network (the expectation output). $\|\cdot\|_F$ represents the Frobenious function. Minimizing mse loss during training allows the model to output denoising records with high SNR. However, mse loss tends to direct the model to produce a smooth prediction signal in the noisy signal region rather than a highly accurate texture, with limited the ability to describe the structured features of the signal. We alleviate this problem by increasing the adversarial loss.

2) Adversarial Loss: In addition to the reconstruction loss described earlier, we also add the adversarial loss to the loss function. It encourages our generator denoiser to achieve better denoising performance by trying to fool the discriminator. In the original GANs, the adversarial loss is based on cross entropy [29]. Therefore, one choice of adversarial loss can be expressed as follows:

$$\begin{aligned} \mathcal{L}_{\text{Original_adv}}(G, D) = & \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(x)] \\ & + \mathbb{E}_{y \sim p_{\text{data}}(y)} [\log(1 - D(G(y)))] \end{aligned} \quad (2)$$

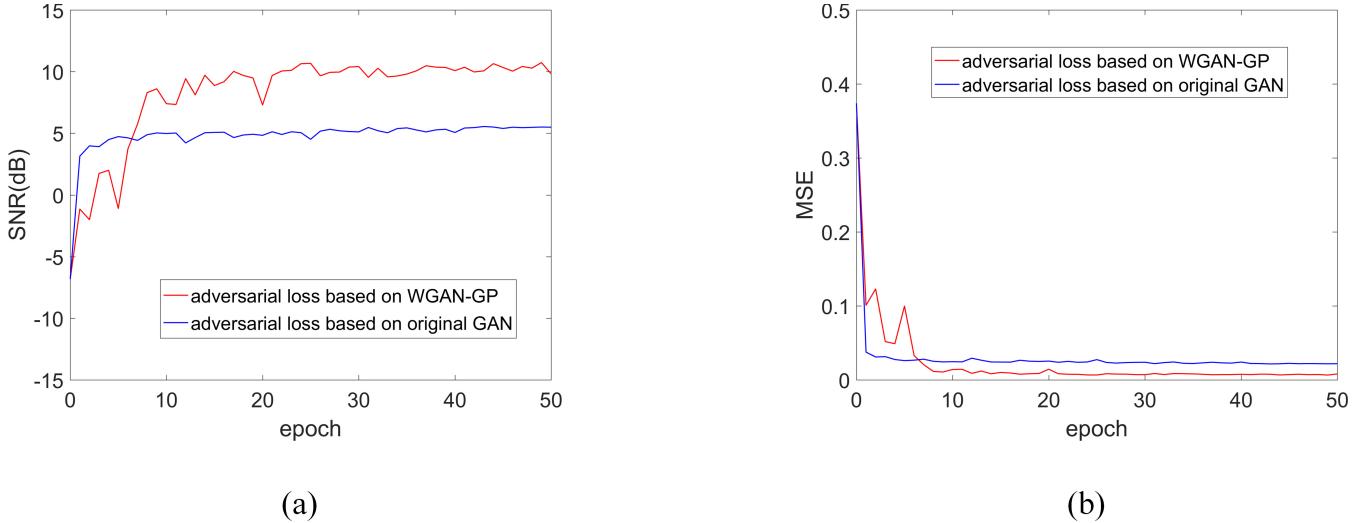


Fig. 5. Denoising performance of the generator denoiser model with two adversarial losses during adversarial training. The model performance is evaluated every epoch. (a) SNR of synthetic seismic data after being processed by generator denoiser with two adversarial losses. (b) MSE of synthetic seismic data after being processed by generator denoiser with two adversarial losses.

where $\mathbb{E}_{\sim \mathbb{P}}[f(.)]$ represents the mathematical expectation of function $f(.)$ for \cdot under distribution p , G and D represent the mappings generated by the generator denoiser and discriminator, respectively, x represents the clean signal, and y represents the noisy signal. In this case, the optimization target can be described as

$$G^* = \arg \min_G \max_D \mathcal{L}_{\text{Original_adv}}(G, D). \quad (3)$$

Among them, the generator denoiser tries to maximize the loss function, and the discriminator wants to minimize the loss function.

However, in practical applications, the drawbacks of adversarial loss proposed from the original GAN are also obvious. Adversarial training between generator denoiser and discriminator cannot be effectively controlled. If the discriminator is trained too well, the generator denoiser will not be able to learn, resulting in poor denoising performance of the model. In contrast, if the discriminator is trained too weakly, it will not play a significant role in indicating the generator denoiser. Therefore, we need to choose a new adversarial loss that reflects the training situation of GAN.

In various improved versions of GAN, WGAN [33] used the Wasserstein distance to describe the degree of difference between the two distributions. The discriminator can approximate the Wasserstein distance. With the help of the discriminator, the optimized generator denoiser can reduce the Wasserstein distance and effectively makes the denoised distribution closer to the clean distribution. WGAN provides a reliable indicator of the training progress. Moreover, the indicator is highly correlated with the quality of the generated sample so that the degree of model training can be monitored by this loss.

WGAN-GP is an improved version of WGAN [34]. WGAN-GP adds gradient penalty to the loss mentioned in WGAN, making model training more stable. The loss function

of WGAN-GP is as follows:

$$\begin{aligned} \mathcal{L}_{\text{adv}} = & \mathbb{E}_{\tilde{x} \sim \mathbb{P}_g}[D(\tilde{x})] - \mathbb{E}_{x \sim \mathbb{P}_r}[D(x)] \\ & + \lambda \mathbb{E}_{\hat{x} \sim \mathbb{P}_{\hat{x}}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2] \end{aligned} \quad (4)$$

where $\mathbb{E}_{\sim \mathbb{P}}[f(.)]$ represents the mathematical expectation of function $f(.)$ for \cdot under distribution \mathbb{P} , \mathbb{P}_g is the generator denoiser distribution, \mathbb{P}_r is the distribution over the pure signal, $\mathbb{P}_{\hat{x}}$ is defined as a distribution sampling uniformly along straight lines between pairs of points sampled from \mathbb{P}_g and \mathbb{P}_r , $\|\cdot\|_2$ represents the L_2 -norm, ∇ is the gradient symbol, and λ is the penalty factor. Also, we make $\lambda = 10$. The loss function includes the Wasserstein distance between the distribution of the denoised signal and the clean signal and increases the gradient penalty to improve training stability. By minimizing the adversarial loss during training, the output of the model will be pushed to the manifold of the clean signal.

We evaluated the performance of (2) and (4) as adversarial loss in practice. Fig. 5 shows the denoising performance of the generator denoiser with two adversarial losses during adversarial training. By comparison, it can be found that the loss of WGAN-GP has better performance on improving SNR of the seismic record. Moreover, it is no longer necessary to balance the training of the generator denoiser and the discriminator, that is, the better the generator denoiser is trained, the more advantageous it is to upgrade the discriminator. Therefore, we finally choose the loss based on WGAN-GP as the solution of adversarial loss.

3) *Final Loss Function:* The final loss function is expressed as

$$\mathcal{L} = \lambda_{\text{rec}} \mathcal{L}_{\text{rec}} + \lambda_{\text{adv}} \mathcal{L}_{\text{adv}}. \quad (5)$$

Among them \mathcal{L}_{rec} is the reconstruction loss, \mathcal{L}_{adv} is the adversarial loss, λ_{rec} and λ_{adv} are the corresponding weights for them, respectively. We set λ_{rec} and $\lambda_{\text{adv}} = 0.01$. The ultimate training goal for the entire network is to minimize this loss function.

TABLE II
PARAMETER OF FORWARD MODELING

Parameter	Specification
Seismic wavelet	Ricker wavelet, Zero-phase wavelet, Mixed-phase wavelet
Dominant frequency(Hz)	15-25
Interval(m)	20
Sampling frequency(Hz)	500
Apparent velocity(m/s)	500-4000
Waveform parameter	1-8

C. Construction of the Training Set

However, even more important than the choice of an appropriate learning algorithm is the available data in our training data set. No algorithm will be able to make good predictions without informative and discriminatory features. Obviously, the model proposed in this article is data-based. During network training, the noisy signal is required as the network input and the corresponding clean signal is used as the label (the expectation output) for semisupervised training network. However, in actual seismic exploration, clean seismic records are basically not available due to complex noise interference. If the denoised signal by other methods is used as a label, the network will fit the denoising mapping of the traditional method, which is meaningless. Therefore, we try to use seismic wavelets with variable parameters for forward modeling to obtain the theoretical clean seismic signal.

The Ricker wavelet, zero-phase wavelet, and mixed-phase wavelet are used to simulate effective signals in desert seismic data. Through the forward modeling based on the above acoustic wave equation, multiple noiseless synthetic gathers with different reflections. We synthesized 100 theoretically clean signals whose sizes are 1000×200 (sampling point number \times trace number). The specific settings of parameters are shown in Table II. The equations of these three seismic wavelets are listed as follows.

Ricker Wavelet:

$$f(t) = A[1 - 2 \times (\pi f_0(t - t_0))^2] \times e^{-(\pi f_0(t - t_0))^2}. \quad (6)$$

Zero-Phase Wavelet:

$$f(t) = A \cos[2\pi f_0(t - t_0)] \times e^{-\left(\frac{2\pi f_0(t - t_0)}{r_1}\right)^2}. \quad (7)$$

Mixed-Phase Wavelet:

$$f(t) = A \sin[2\pi f_0(t - t_0)] \times e^{-\left(\frac{2\pi f_0(t - t_0)}{r_2}\right)^2} \quad (8)$$

where A , t_0 , and f_0 represent the amplitude, beginning time, and main frequency, respectively. r_1 and r_2 are the waveform parameters of zero-phase wavelet and mixed-phase wavelet, respectively. We set the patch size as 50×50 and stride as 10 and crop 36×100 patches to train the model. Finally, the training set of clean data $X = \langle x_1, x_2, \dots, x_e \rangle$ can be obtained by amplitude normalization.

Synthesizing noisy signals requires a lot of noise data. In order to make the added noise as close as possible to the noise in the real seismic record, this article uses actual desert noise data collected from the desert region of the Tarim Basin in China to synthesize noisy seismic data. The noise data we used have 2000 traces, each with 30 000 sampling points, and

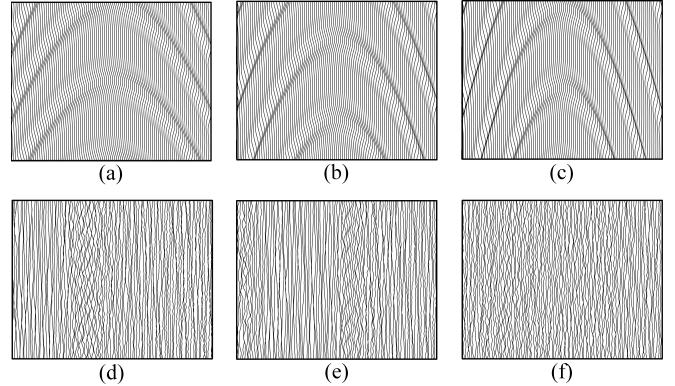


Fig. 6. Patches of clean signal and noise in the training set. (a)–(c) Patches of clean signal. (d)–(f) Patches of noise.

the sampling interval is 2 ms. We set the patch size as 50×50 and stride as 50 and crop 600×40 noise patches. Then, the noise set $N = \langle n_1, n_2, \dots, n_e \rangle$ can be obtained by amplitude normalization. During the model training, the distribution of noise will be implicitly estimated by the GAN.

To learn a single model for the denoising tasks with different SNRs, the following operations are performed. One set of signal data x_i and noise data n_i is randomly extracted from the clean signal set $X = \langle x_1, x_2, \dots, x_e \rangle$ and the noise set $N = \langle n_1, n_2, \dots, n_e \rangle$, respectively. Some patches of clean signal and noise are shown in Fig. 6. The noise data n_i are randomly scaled by 1–5 times and then added to the clean signal x_i to generate a noisy signal y_i . We loop through this process to get the training data set $\{x_i, y_i\}_{i=1}^N$. Among them, x_i is the label when the network input is y_i .

D. Principle of Denoising

Seismic data collected in the field usually contain noise interference and can be expressed as

$$y = x + n \quad (9)$$

where x refers to the theoretical seismic signal and n refers to the noise interference. In our proposed method, the generator denoiser is designed to learn the end-to-end mapping function between the noisy signal y and the theoretical clean signal x . The predicted clean signal can be expressed as

$$\tilde{x} = G(y; \Theta) \quad (10)$$

where G is the generator denoiser and $\Theta = \{W, b\}$ is the network parameter of the generator denoiser that needs to be optimized, including the weight W and the offset b .

TABLE III
NETWORK TRAINING PARAMETERS

Hyper-parameters	Specification
Optimizer	Adam
Patch size	64×64
Batch size	128
epoch	50
Learning rate range	$[10^{-3}, 10^{-5}]$

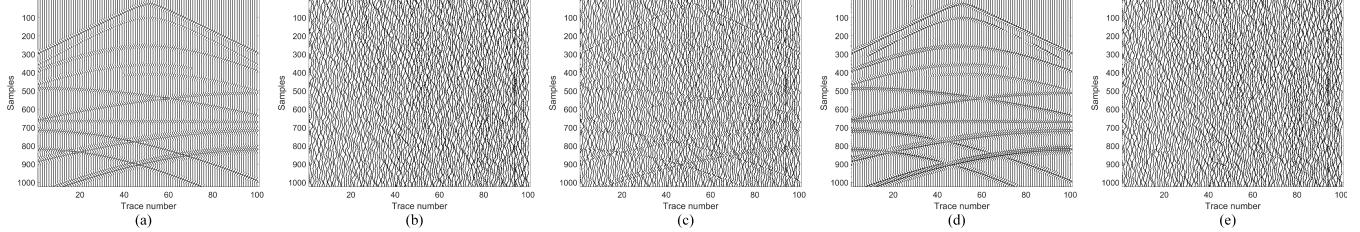


Fig. 7. Time-domain analysis of the denoising result. (a) Theoretical desert seismic data. (b) Desert low-frequency noise. (c) Synthetic noisy seismic data ($\text{SNR} = -8.4201\text{dB}$ and $\text{mse} = 0.5383$). (d) Denoising results of DnGAN ($\text{SNR} = 10.2249\text{ dB}$ and $\text{mse} = 0.0074$). (e) Difference map of DnGAN.

In order to optimize the generator denoiser G in (7), we used the discriminator D to semisupervise the training of the generator denoiser. The loss function as shown in (5) combining the reconstruction loss and the adversarial loss is used to guide the training process of the model. We used the Adam optimizer [37] to optimize the network. By using the optimizer to minimize the loss function, the weight parameters in the model are inversely adjusted layer by layer according to the gradient provided by the loss function [38]. The denoising ability of the network is gradually improved by frequent iterative training. More details of network training can be found in Table III.

After the training of the network is completed, the generator denoiser G is used as a denoising network to process the noisy seismic records. For the noisy seismic record Y , the predicted clean signal can be expressed as

$$\tilde{x} = G(y). \quad (11)$$

III. PROCESSING AND ANALYSIS OF SYNTHETIC DATA

A. Comparison and Analysis of Denoising Results

In this section, we use the trained generator denoiser to process the synthetic noisy data. The synthetic data used for testing are independent of the synthetic data used for training. For the convenience of description, the method proposed in this article is named DnGAN. The seismic events can reflect stratum information. In general, due to the diversity of underground layers, seismic events contain three types: curved event, fracture event, and crossed event, and the first one is the most common. In order to further verify the effectiveness of the method proposed in this article, we directly use the abovementioned three wavelets to construct different types of events. We can obtain the synthetic noise-free seismic data [see Fig. 7(a)], which is simulated by an acoustic equation with a source function given by three wavelets with a central

frequency of 15 Hz. In addition, some of the actual desert noise as shown in Fig. 7(b) (not in the noise training set) is selected and added to the clean signal to obtain a synthetic noisy seismic record (-8.4201 dB) as shown in Fig. 7(c) for denoising network testing. We assume that these data are received by 200 receivers, each receiving 1024 samples at a sampling rate of 500 Hz. Fig. 7(d) shows the denoising result of DnGAN, and Fig. 7(e) shows the corresponding difference map (the difference between the noisy seismic data and the denoised seismic data). Fig. 7(d) shows that DnGAN can effectively suppress low-frequency noise and there is basically no residual noise in the denoising result. Fig. 7(e) shows that there is almost no effective signal in the difference map, which indicates that the method hardly damages the effective signal when suppressing noise. In addition, we also analyze the f-k spectra of each signal. Fig. 8(a)–(e) shows the f-k spectra of the clean signal, the actual noise, the noisy seismic data, the denoising result, and the difference map, respectively. It can be seen from Fig. 8(d) and (e) that the signal overlapping with the noise is well recovered, which proves that the proposed method can suppress the noise overlapping with the effective signal. The different profiles show that DnGAN has excellent amplitude retention performance.

B. Comparison and Analysis of Denoising Results

In order to verify the superiority of the proposed method, we compared the DnGAN method with several traditional denoising methods, including bandpass filter and wavelet transform filter. At the same time, a denoising convolutional encoder-decoder network (RED-Net) training by the mse loss is also used for comparison to highlight the performance gains that the adversarial training brings to the denoising network. In order to show the superiority of our DnGAN scheme, we also added the result of original GAN for comparison. The clean signal is shown in Fig. 7(a). The denoising

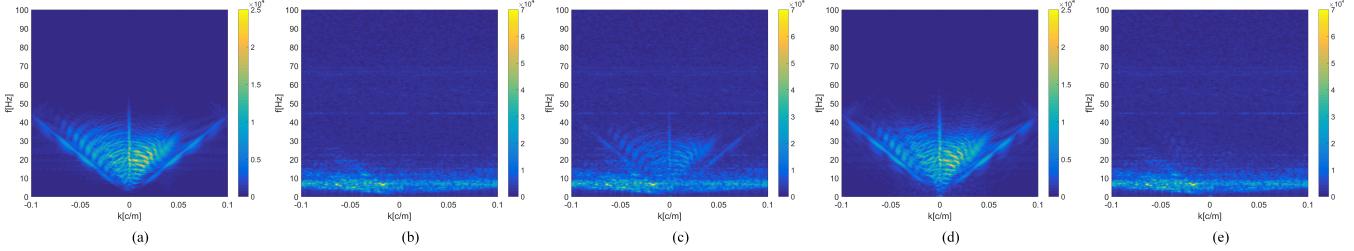


Fig. 8. F-K domain analysis of the denoising result. (a) F-K spectrum of theoretical desert seismic data. (b) F-K spectrum of desert low-frequency noise. (c) F-K spectrum of synthetic noisy seismic data. (d) F-K spectrum of the denoising results of DnGAN. (e) F-K spectrum of the difference map of DnGAN.

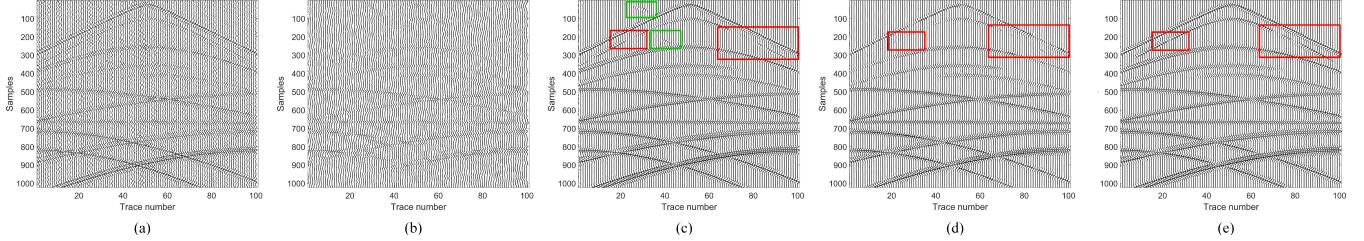


Fig. 9. Comparison for denoising results. (a) Denoising result of bandpass filter (SNR = 3.9981 dB and mse = 0.0308). (b) Denoising result of wavelet transform filter with soft threshold (SNR = -2.8502 dB and mse = 0.1493). (c) Denoising result of RED-Net (SNR = 10.5841 dB and mse = 0.0068). (d) Denoising result of original GAN (SNR = 8.8552 dB and mse = 0.0101). (e) Denoising result of DnGAN (SNR = 10.2249 dB and mse = 0.0074).

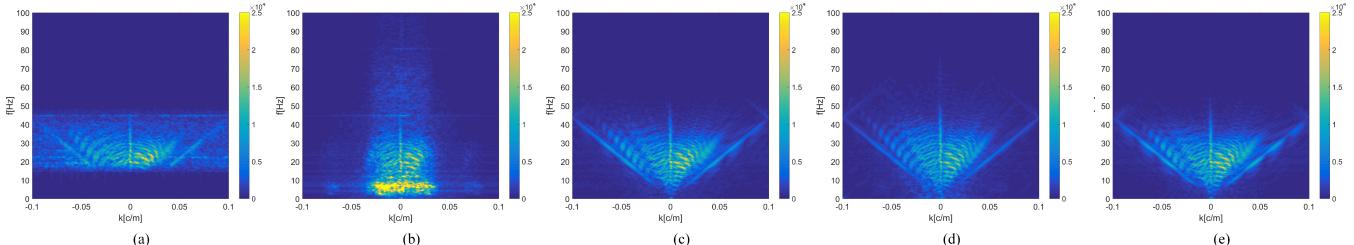


Fig. 10. Comparison and analysis of F-K spectrum. (a)–(e) F-K spectrum of denoising results after applying bandpass filter, wavelet transform with soft threshold, RED-Net, original GAN, and DnGAN in Fig. 9(a)–(e), respectively.

results of bandpass filter, wavelet transform filter, RED-Net, the original GAN, and DnGAN are shown in Fig. 9(a)–(e), respectively. By observing the denoising results of the five methods in Fig. 9, we can see that the denoising result of DnGAN in Fig. 9(e) is closest to the theoretical desert seismic data in Fig. 7(a). We can deduce that DnGAN suppresses the low-frequency noise more thoroughly and restores the effective signal more clearly. In contrast, Fig. 9(a) and (b) shows that a large amount of low-frequency noise remains in the denoising results of two traditional methods. Although the effective signal has been restored to some extent, its continuity is still poor. For example, in Fig. 9(a), noise at the same frequency as the effective signal is difficult to be suppressed by the bandpass filter. In Fig. 9(b), due to the serious frequency-band overlap between the noise and the effective signal, the soft threshold function cannot accurately select the wavelet coefficients, resulting in a large amount of low-frequency noise residual in the denoising result. Moreover, the recovery of the effective signal in the denoising result is not ideal. In Fig. 9(c), we can see that although RED-Net suppresses most of the low-frequency noise, the effective signal severely disturbed by the strong energy noise in the area marked by the red box is not recovered. In the area marked by the green box, because of the

noise with strong energy, there are false predicted events in the denoising result of RED-Net. It is because RED-Net trained with mse loss tends to output smooth results. In contrast, in Fig. 9(e), there is no false event in the denoising result of DnGAN. Also, signals in the red box are recovered more continuously. The original GAN also has the ability to repair the shaft slightly when suppressing the noise. However, it can be seen from the red box in Fig. 9(d) that the ability of original GAN to repair the signal is weaker than DnGAN. It shows that DnGAN has a better performance than the original GAN. We attribute this improvement to the introduction of the new loss function, especially the adversarial loss based on WGAN-GP.

Next, we analyze the denoising results of the clean seismic data and the five methods in the frequency domain. Fig. 10 shows the F-K spectrum of the denoising results of the five methods. From Fig. 10(a)–(e), we can find that the F-K spectrum of the denoising result of DnGAN as shown in Fig. 10(e) is the closest to the F-K spectrum of the clean seismic data, as shown in Fig. 8(a). Therefore, DnGAN can suppress most of the low-frequency noise and recover the effective signal completely. In contrast, from Fig. 10(a) and (b), we can see that the suppression effect of traditional methods on low-frequency

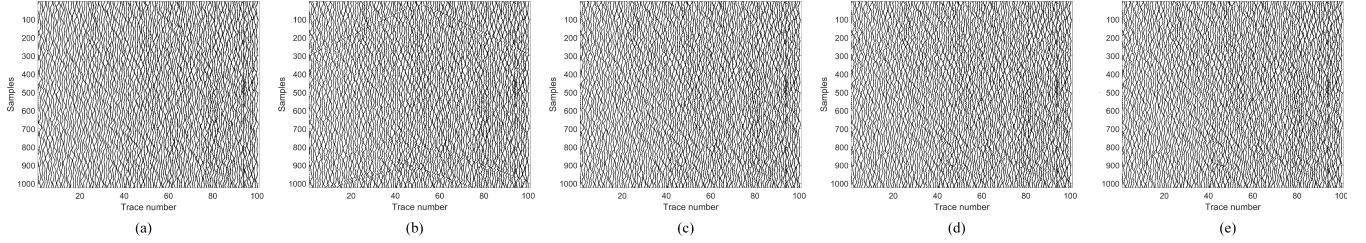


Fig. 11. Difference map of different denoising methods. (a)–(e) Difference map of bandpass filter, wavelet transform filter with soft threshold, RED-Net, original GAN, and DnGAN, respectively.

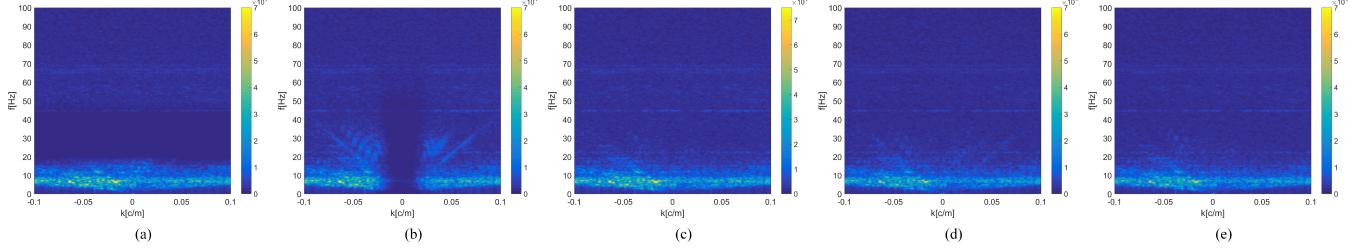


Fig. 12. F-K spectrum of difference map of different methods. (a)–(e) F-K spectrum of difference map of bandpass filter, wavelet transform filter with soft threshold, RED-Net, original GAN, and DnGAN in Fig. 11(a)–(e), respectively.

noise is not ideal. For example, in the F-K spectrum of the denoising result of the bandpass filter [see Fig. 10(a)], the low-frequency signal and noise are removed together. At the same time, the traditional method does not have a strong recovery effect on the effective signal. Especially in the F-K spectrum of the denoising result of the wavelet transform filter with soft threshold as shown in Fig. 10(b), a large number of effective signal components are lost. In Fig. 10(c) and (e), the F-K spectrum of the denoising results of DnGAN and RED-Net is basically the same, which means that DnGAN maintains the basic denoising function of RED-Net.

C. Comparison and Analysis of Difference Map

In the noise suppression of seismic data, the protection of effective signals is an important factor to measure the performance of the algorithm. In this article, we compare the algorithm's ability to protect the effective signal by observing whether there are effective signals in the difference map. The difference map is the difference between the noisy record and the denoising result. Obviously, the closer the difference map is to the added low-frequency noise, the better the algorithm protects the effective signal. In Fig. 11(a)–(e), the difference map of bandpass filter, wavelet transform filter with soft threshold, RED-Net, original GAN, and DnGAN are sequentially presented, respectively. By observation, it can be found that the difference record of DnGAN [see Fig. 11(e)] is closest to the added low-frequency noise [see Fig. 7(b)]. In the difference map of the two traditional methods [see Fig. 11(a) and (b)], the residual of the effective signal is very significant. In the difference result of RED-Net and original GAN, although there is no effective signal residual, the difference map is not as close to the original noise as DnGAN.

Next, we analyze the pure noise and difference map of different methods in the frequency domain. In Fig. 12(a)–(e),

we give the F-K spectrum of the difference map of five methods. By observing, we can get a conclusion similar to the previous one. The F-K spectrum of the difference map corresponding to DnGAN is closest to the F-K spectrum of pure noise as shown in Fig. 8(b), which shows that DnGAN can effectively protect the effective signal. In contrast, there are a large number of effective signal components in the F-K spectrum of the difference map corresponding to the two traditional methods. We can infer that the traditional methods also cause the energy loss of the effective signal while suppressing the low-frequency noise. Compared with RED-Net and original GAN, the noise component in the F-K spectrum of the difference map corresponding to DnGAN is more comprehensive, indicating that DnGAN has a better ability to suppress noise.

D. Comparison of Multiple Test Results

In order to measure the denoising performance of various methods for records with different SNRs, we selected six low-frequency noise data for different energy intensities. The noise training set in this article does not include these five low-frequency noises. The selected five kinds of noise data are added to the synthetic theoretical seismic data of Fig. 7(a) to construct five kinds of noisy seismic data with different SNRs. Then, bandpass filter, wavelet transform filter with soft threshold, RED-Net, the original GAN, and DnGAN are used to process the constructed noisy seismic data, and these denoising results are measured by calculating SNR and mse.

The equations of SNR and mse are shown as follows:

$$\text{SNR} = 10 \log \frac{\sum_i \sum_t |x(i, t)|^2}{\sum_i \sum_t |\hat{x}(i, t) - x(i, t)|^2} \quad (12)$$

$$\text{MSE} = \frac{1}{MN} \sum_i \sum_t |\hat{x}(i, t) - x(i, t)|^2 \quad (13)$$

TABLE IV
SNR AND MSE OF MULTIPLE TEST RESULTS [SNR(dB)/MSE]

Original Record	band-pass filter	wavelet transform filter	RED-Net	Original GAN	DnGAN
1.1223/0.0598	7.4367/0.0140	2.8141/0.0405	13.1050/0.0038	11.8849/0.0050	12.3598/0.0044
-2.3995/0.1346	6.6874/0.0166	1.2954/0.05575	12.6247/0.0042	11.1713/0.0059	12.0044/0.0049
-4.8983/0.2392	5.8162/0.0203	-0.2028/0.0812	12.0196/0.0042	10.4423/0.0070	11.5314/0.0054
-6.8365/0.3738	4.9035/0.0250	-1.5890/0.1117	11.3195/0.0057	9.6608/0.0084	10.8979/0.0063
-8.4201/0.5383	3.9981/0.0308	-2.8502/0.1493	10.5841/0.0068	8.8552/0.0101	10.2249/0.0074
-9.7591/0.7327	3.1258/0.0377	-4.0037/0.1947	9.0142/0.0097	7.5179/0.0137	8.7731/0.0103

where $\hat{x}(i, t)$ is the denoising data, $x(i, t)$ is the clean data, $i = 1, \dots, N$ represents trace number, and $t = 1, \dots, M$ represents the number of sampling points. Both larger SNR and smaller mse indicate a better denoising effect. The SNR and mse of the denoising results of the five methods are calculated separately, as shown in Table IV. Among them, we can see that the denoising result of RED-Net has the smallest mse and the largest SNR because of supervised learning with mse loss. The results of DnGAN under semisupervised learning have a slight decrease in SNR but retain a better signal distribution. At the same time, the results of DnGAN are better than those of the original GAN. It shows that DnGAN is more advantageous and more stable than the original GAN in improving the SNR. At the same time, the DnGAN can suppress the noise in records with different SNRs, indicating that the model has a very good generalization performance for desert seismic records.

IV. FIELD DATA PROCESSING

When applying GAN to process field data, we did not retrain the network but applied the same network that has been processed synthetic data. In order to verify the practical application value of DnGAN, we use bandpass filter, wavelet transform filter with soft threshold, RED-Net, the original GAN, and DnGAN to process the actual common-shot record shown in Fig. 13(a). The denoising results and difference maps for each method are shown in Fig. 13. Since the noise is always present in the actual exploration, the clean signal cannot be obtained and the SNR and the mse of the denoising record cannot be calculated. Therefore, for the real common-shot record, the quality of the denoising result is usually judged by observing the continuity of the events of effective signal.

The seismic data in Fig. 13(a) comes from the desert region of Tarim Basin, China. This common-shot record with random noise interference is received by 200 detectors, each acquiring 2000 samples with a sampling frequency of 500 Hz. We can see that there are a lot of low-frequency noises in this actual common-shot record, including low-frequency random noise, surface wave, and prestack noise. The low-frequency noise seriously destroys the continuity of the events. For example, in the red box of Fig. 13(a), the low-frequency noise seriously interferes with the events. In the green box and the yellow box of Fig. 13(a), the events are completely cut off due to the interference of the surface wave with strong energy. The following can be found by observing Fig. 13(b)–(k). The bandpass filter can suppress some of the low-frequency noise

(random noise and surface wave) but also damages the effective signal in the same frequency band as the noise, as shown in Fig. 13(c). Wavelet transform filter with soft threshold suppresses random noise and surface waves, but there are still a lot of low-frequency random noises in the denoised record. Moreover, the effective signal in the denoising result has poor amplitude retention, resulting in poor continuity of the events. In Fig. 13(j), we can find that the proposed method effectively suppresses most of the low-frequency random noise, surface wave, and prestack noise. In the area corresponding to the red box in Fig. 13(a), the low-frequency random noise is suppressed more thoroughly, and the events of the effective signal can be clearly observed. In the area corresponding to the green box and the yellow box in Fig. 13(a), the surface wave with strong energy is also suppressed, and the continuity of the events becomes better. Also, from the difference map as shown in Fig. 13(k), we can find that DnGAN has no damage to the effective signal. At the same time, the denoising results of RED-Net and DnGAN are compared to highlight the practical effects of adversarial training. In particular, in Fig. 13(j), the signal in the yellow disturbed by the surface wave and the effective signal in the yellow interfered by the low-frequency noise are also recovered. However, it is not restored by RED-Net in Fig. 13(f). These facts show that DnGAN performs better than RED-Net in restoring effective signals in actual seismic records. We believe that it is the performance improvement that the adversarial training brings to the model denoising. At the same time, from the recovery of events in Fig. 13(j) and (h), especially from the local magnification marked by the yellow box, it can be seen that DnGAN has better repair capabilities for seismic events than the original GAN. In summary, compared with the traditional method, DnGAN proposed in this article can suppress low-frequency noise more effectively and restore the events and has a better recovery effect than RED-Net under the guidance of mse loss and original GAN.

It is worth noting that although we did not intentionally require the model to remove the surface wave during training, the model still suppressed the surface wave in the denoising of the actual record. Surface waves are generated by seismic waves traveling along the ground, so most of the seismic background noise received without source motivation is random noise. According to the principle of deep learning, the DnGAN model trained by desert random noise in this article should only suppress the random noise, not the mixed noise (random noise and surface waves). Although surface waves and effective signals are homologous, the difference of propagation

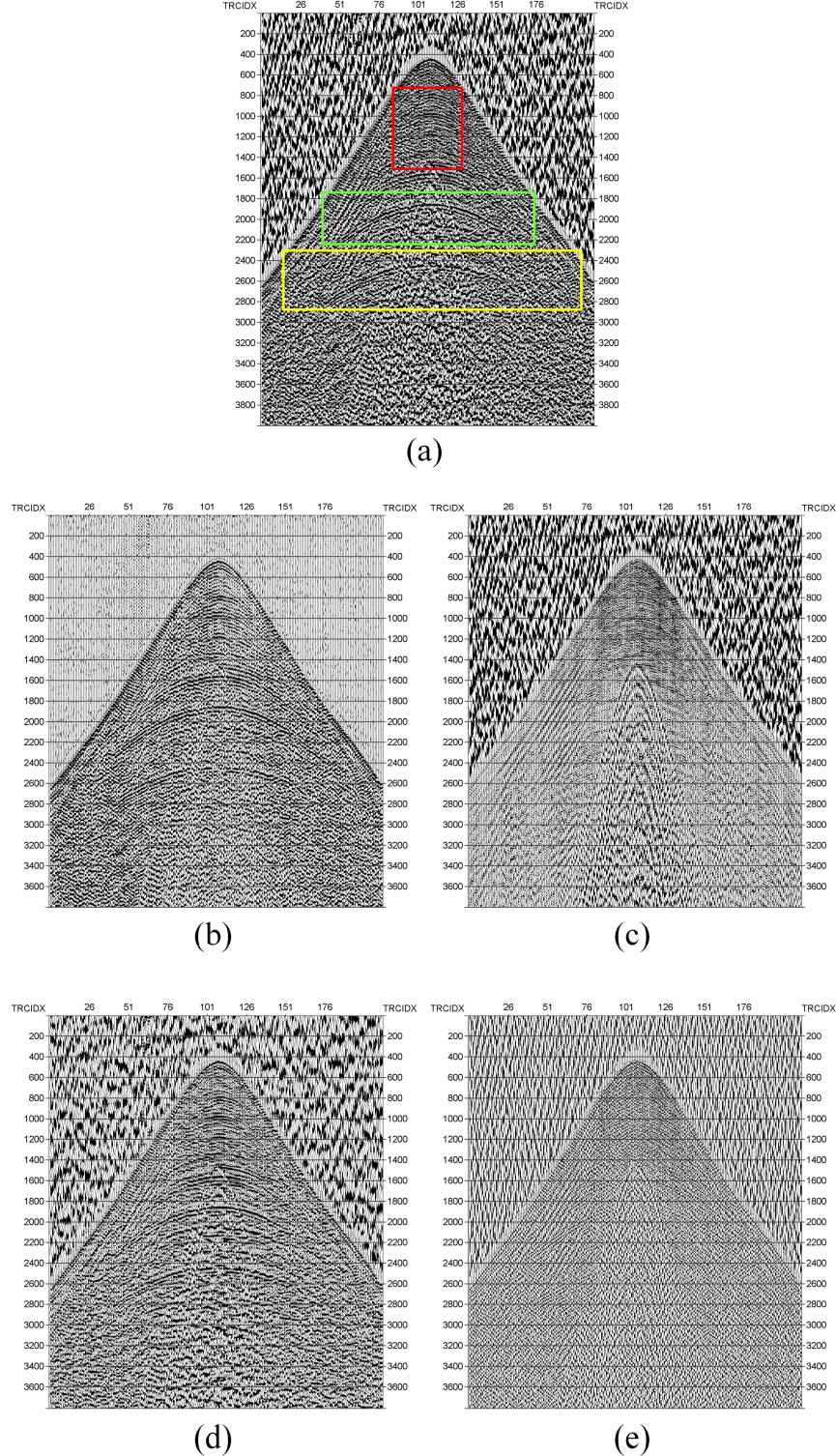


Fig. 13. Denoising results of field seismic data in the desert region. (a) Field data in the desert region. (b) Denoising results of the field data by bandpass filter. (c) Difference map of the field data by bandpass filter. (d) Denoising results of the field data by wavelet transform with soft threshold. (e) Difference map of the field data by wavelet transform with soft threshold.

medium leads to the great difference of wave velocity. In desert areas, the propagation velocity of the surface wave is about 300–500 m/s, whereas the propagation velocity of the effective signal is usually higher than 500 m/s. This difference leads

to that the trained DnGAN model considers surface waves as noise component and remove it.

The generalization ability is a crucial evaluation criterion for a denoising model in practice applications. The main purpose

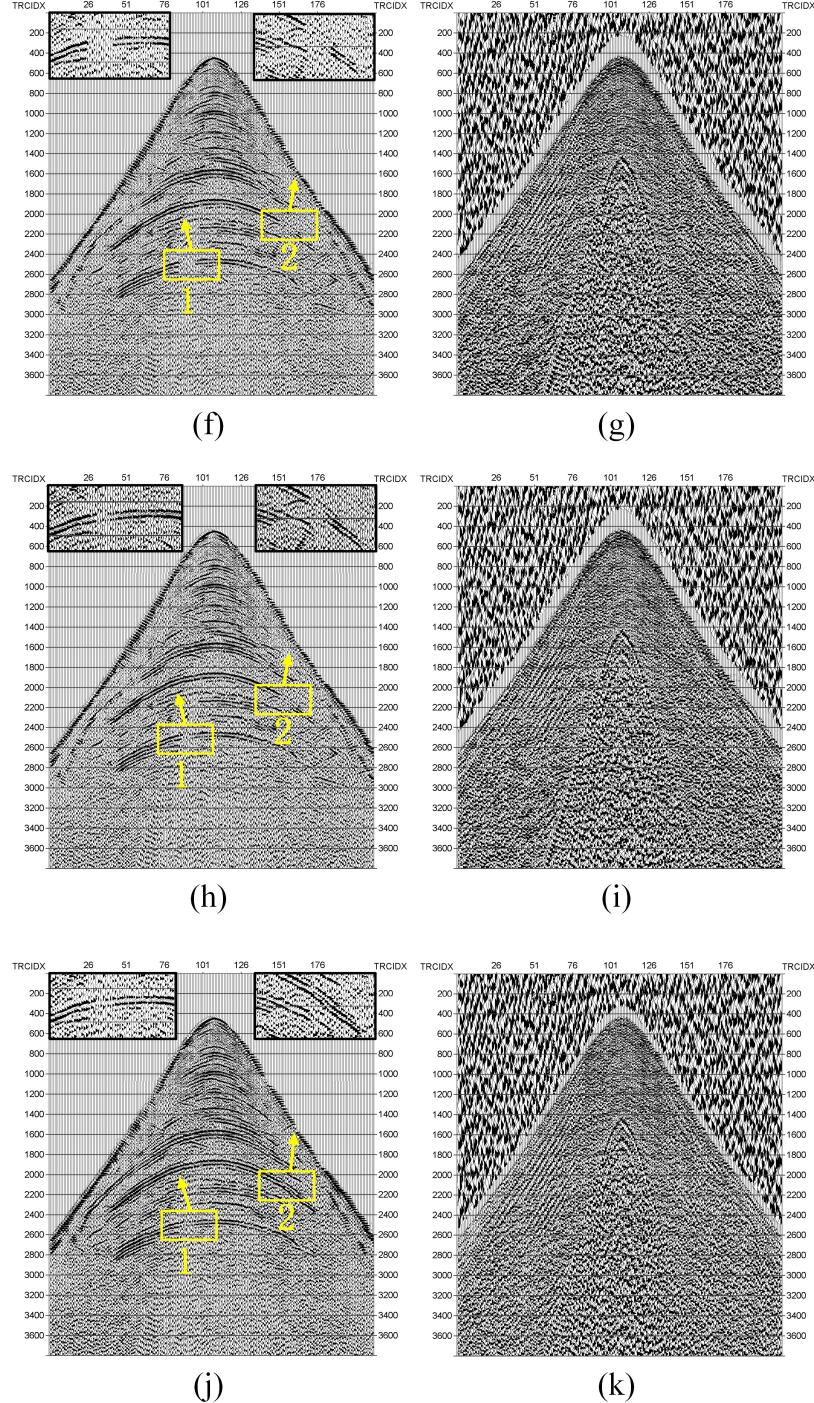


Fig. 13. (Continued.) Denoising results of field seismic data in the desert region. (f) Denoising results of the field data by RED-Net. (g) Difference map of the field data by RED-Net. (h) Denoising results of the field data by original GAN. (i) Difference map of the field data by original GAN. (j) Denoising results of the field data by DnGAN. (k) Difference map of the field data by DnGAN.

of the model is to obtain capture features and learns the laws from the training data during training. Therefore, the model should have the generalization ability that can make the trained model perform well when processing different data with a similar pattern.

To verify the generalization ability of the DnGAN, we applied the trained model on seismic shot gathers from the same survey area but receiving by different receiver lines. The result is shown in Fig. 14. It can be seen that the effective

signal has been completely restored, indicating that our model has a good generalization ability for seismic records in desert areas.

Moreover, to better present the generalization ability of our model, we applied it to process the data from different regions (such as grass and mountains area). The results are shown in Fig. 15. The results show that our model can also denoise these seismic records from different regions, but some noise is still preserved. This is because our model is trained based

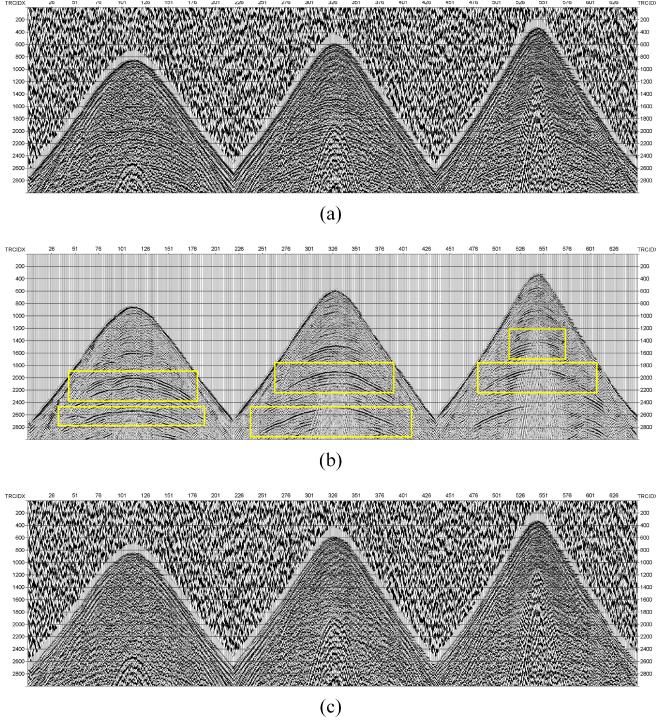


Fig. 14. Denoising results of the real desert shot gathers from the same survey area but received by different receiver lines. (a) Real desert shot gathers. (b) Denoised shot-gathers of (a). (c) Difference map.

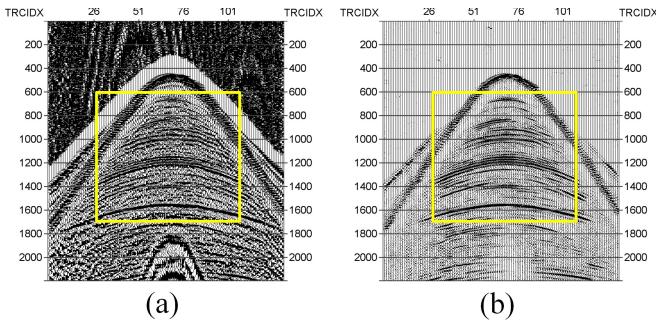


Fig. 15. Denoising results of real seismic record acquired from grass and mountain areas. (a) Real seismic record from the grass area. (b) Denoising record of (a). (c) Real seismic record from the mountain area. (d) Denoising record of (c).

on the training data from the desert region, and the denoising performance is limited due to a lack of training data from the corresponding regions.

The generalization ability of the model is attributed to the denoising algorithm of DnGAN. For the denoising problem, DnGAN extracts the signal features and discriminating the

signal features from the background noise features. This means that the purpose of denoising model is to recognize the signal that should be preserved, and anything different from the signal is removed by the model. This is the reason that our model can denoise the seismic records from different regions where the noise features vary. Also, it is a potential guarantee for the generalization performance of the model.

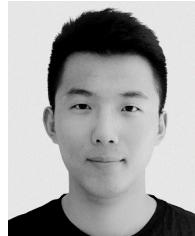
V. CONCLUSION

Seismic records obtained during seismic exploration often include complex random noise interference, which seriously affects the subsequent processing of seismic records. For the records obtained in the desert region, the intensity of the desert noise is large, which makes the desert seismic records have extremely low SNR. In addition, due to the severe spectral aliasing of desert seismic data, traditional denoising methods cannot effectively separate effective signals and low-frequency noise. Although the CNN trained with the guidance of mse loss can effectively improve the SNR of the record when processing desert records, some effective signals are seriously attenuated in some traces of recovered desert seismic data. Aiming at the suppression of desert low-frequency noise and the better restoration of effective signals, this article proposes an adversarial network architecture for denoising. The generator denoiser is used to suppress the noise and improve the SNR of records, and the discriminator is used to guide the generator denoiser to recover the signal structure. The generator denoiser and the discriminator improve each other's performance through games. Adversarial learning makes the discriminator unable to distinguish the data in the denoising domain and the theoretical domain, indirectly making them belong to a distribution, so as to guide the generator denoiser update parameters to achieve better results as a regularization method. The trained generator denoiser can greatly suppress the noise and improve the SNR while greatly improving the ability to recover the event.

REFERENCES

- [1] N. Wu, Y. Li, and B. Yang, "Applications of the trace transform in surface wave attenuation on seismic records," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4997–5007, Dec. 2011.
- [2] X. Dong, H. Jiang, S. Zheng, Y. Li, and B. Yang, "Signal-to-noise ratio enhancement for 3C downhole microseismic data based on the 3D shearlet transform and improved back-propagation neural networks," *Geophysics*, vol. 84, no. 4, pp. V245–V254, Jul. 2019.
- [3] H. Ma, J. Yan, Y. Li, C. Zhang, and H. Lin, "Desert seismic random noise reduction based on LDA effective signal detection," *Acta Geophysica*, vol. 67, no. 1, pp. 109–121, Feb. 2019.
- [4] A. Aghayan, P. Jaiswal, and H. R. Siahkoohi, "Seismic denoising using the redundant lifting scheme," *Geophysics*, vol. 81, no. 3, pp. V249–V260, May 2016.
- [5] X. Dong, Y. Li, N. Wu, Y. Tian, and P. Yu, "The S-STK/LTK algorithm for arrival time picking of microseismic signals," *J. Geophys. Eng.*, vol. 15, no. 4, pp. 1484–1491, May 2018.
- [6] T. Zhong, S. Zhang, Y. Li, and B. Yang, "Simulation of seismic prospecting random noise in the desert by a Brownian-motion-based parametric modeling algorithm," *Comp. Rendus Geosci.*, vol. 351, no. 1, pp. 10–16, Jan. 2019.
- [7] G. Li, Y. Li, and B. Yang, "Seismic exploration random noise on land: Modeling and application to noise suppression," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4668–4681, Aug. 2017.

- [8] T. Zhong, Y. Li, N. Wu, P. Nie, and B. Yang, "A study on the stationarity and Gaussianity of the background noise in land-seismic prospecting," *Geophysics*, vol. 80, no. 4, pp. V67–V82, Jul. 2015.
- [9] H. Ma, Z. Qian, Y. Li, H. Lin, D. Shao, and B. Yang, "Noise reduction for desert seismic data using spectral kurtosis adaptive bandpass filter," *Acta Geophys.*, vol. 67, no. 1, pp. 123–131, Feb. 2019.
- [10] P. E. Harris and R. E. White, "Improving the performance of f-x prediction filtering at low signal to noise ratios," *Geophys. Prospecting*, vol. 45, no. 2, pp. 269–302, Mar. 1997.
- [11] K. Chen and M. D. Sacchi, "Robust f-x projection filtering for simultaneous random and erratic seismic noise attenuation," *Geophys. Prospecting*, vol. 65, no. 3, pp. 650–668, May 2017.
- [12] S. Cao and X. Chen, "The second-generation wavelet transform and its application in denoising of seismic data," *Appl. Geophys.*, vol. 2, no. 2, pp. 70–74, Jun. 2005.
- [13] S. Lianyu *et al.*, "Curvelet transform and its application in seismic data denoising," in *Proc. Int. Conf. Inf. Technol. Comput. Sci.*, 2009, pp. 396–399, doi: [10.1109/ITCS.2009.86](https://doi.org/10.1109/ITCS.2009.86).
- [14] J. L. Gómez and D. R. Velis, "A simple method inspired by empirical mode decomposition for denoising seismic data," *Geophysics*, vol. 81, no. 6, pp. V403–V413, Nov. 2016.
- [15] J. Han and M. van der Baan, "Microseismic and seismic denoising via ensemble empirical mode decomposition and adaptive thresholding," *Geophysics*, vol. 80, no. 6, pp. KS69–KS80, Nov. 2015.
- [16] M. Xiong, Y. Li, and N. Wu, "Random-noise attenuation for seismic data by local parallel radial-trace TFPF," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4025–4031, Jul. 2014.
- [17] X. Deng, H. Ma, Y. Li, Q. Zeng, and G. Zhuang, "Spatiotemporal adaptive Time-Frequency peak filtering for seismic random noise attenuation," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 10, pp. 2105–2109, Oct. 2015.
- [18] X. Liang, Y. Li, and C. Zhang, "Noise suppression for microseismic data by non-subsampled shearlet transform based on singular value decomposition: Suppressing noise for microseismic data," *Geophys. Prospecting*, vol. 66, no. 5, pp. 894–903, Jun. 2018.
- [19] M. Schuld, I. Sinayskiy, and F. Petruccione, "An introduction to quantum machine learning," *Contemp. Phys.*, vol. 56, no. 2, pp. 172–185, 2015.
- [20] K. He *et al.*, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778, doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [21] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
- [22] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
- [23] P. Patro, K. Kumar, and G. S. Kumar, "Applications of three layer CNN in image processing," *J. Adv. Res. Dyn. Control Syst.*, vol. 2018, no. 1, pp. 510–512, 2018.
- [24] T. Tirer and R. Giryes, "Super-resolution via image-adapted denoising CNNs: Incorporating external and internal learning," *IEEE Signal Process. Lett.*, vol. 26, no. 7, pp. 1080–1084, Jul. 2019.
- [25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [26] C. Szegedy, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–6.
- [27] X. T. Dong, Y. Li, and B. J. Yang, "Desert low-frequency noise suppression by using adaptive DnCNNs based on the determination of high-order statistic," *Geophys. J. Int.*, vol. 219, no. 2, pp. 1281–1299, Nov. 2019.
- [28] H. Ma *et al.*, "Deep residual encoder-decoder networks for desert seismic noise suppression," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 13, pp. 529–533, Mar. 2019.
- [29] I. J. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Process. Syst.*, 2014, pp. 2672–2680.
- [30] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4681–4690, doi: [10.1109/CVPR.2017.19](https://doi.org/10.1109/CVPR.2017.19).
- [31] P. Isola *et al.*, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1125–1134, doi: [10.1109/CVPR.2017.632](https://doi.org/10.1109/CVPR.2017.632).
- [32] X. Mao, C. Shen, and Y. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 2802–2810.
- [33] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, *arXiv:1701.07875*. [Online]. Available: <https://arxiv.org/abs/1701.07875>
- [34] I. Gulrajani *et al.*, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5767–5777.
- [35] D. Vincent and V. Francesco, "A guide to convolution arithmetic for deep learning," 2016, *arXiv:1603.07285*. [Online]. Available: <https://arxiv.org/abs/1603.07285>
- [36] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. ICML*, 2010, pp. 1–5.
- [37] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent.*, San Diego, CA, USA, 2015.
- [38] Y. A. LeCun *et al.*, "Efficient backprop," in *Proc. Conf. Neural Netw.*, 2012, p. 48, doi: [10.1007/978-3-642-35289-8_3](https://doi.org/10.1007/978-3-642-35289-8_3).
- [39] E. K. Kumar, P. V. V. Kishore, M. T. K. Kumar, D. A. Kumar, and A. S. Sastry, "Three-dimensional sign language recognition with angular velocity maps and connived feature ResNet," *IEEE Signal Process. Lett.*, vol. 25, no. 12, pp. 1860–1864, Dec. 2018.
- [40] K. Eckle and J. Schmidt-Hieber, "A comparison of deep networks with ReLU activation function and linear spline-type methods," *Neural Netw.*, vol. 110, pp. 232–242, Feb. 2019.
- [41] W. Yang *et al.*, "Deep edge guided recurrent residual learning for image super-resolution," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5895–5907, Dec. 2017.
- [42] W. Liu and J. Lee, "An efficient residual learning neural network for hyperspectral image superresolution," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 4, pp. 1240–1253, Apr. 2019.
- [43] Y. Yuan, X. Si, and Y. Zheng, "Ground-roll attenuation using generative adversarial networks," *Geophysics*, vol. 85, no. 4, pp. WA255–WA267, Jul. 2020.
- [44] X. Si, Y. Yuan, F. Ping, Y. Zheng, and L. Feng, "Ground roll attenuation based on conditional and cycle generative adversarial networks," in *Proc. Workshop, Math. Geophys.*, 2020, pp. 95–98, doi: [10.1190/iwmg2019_23.1](https://doi.org/10.1190/iwmg2019_23.1).
- [45] P. Xie, J. Boelle, and H. Puntous, "Generative-adversarial network-based fast-noise removal on land-seismic data," in *Proc. SEG Tech. Program Expanded Abstr.*, 2018, pp. 2171–2175, doi: [10.1190/segam2018-2995310.1](https://doi.org/10.1190/segam2018-2995310.1).



Hongzhou Wang received the B.S. degree from the Department of Communication Engineering, Jilin University, Changchun, China, in 2012, where he is pursuing the M.S. degree.

His main research interests include deep learning, image processing, seismic event identification, weak signal detection, and seismic record denoising and reconstruction.



Yue Li received the B.S. degree in automatic control from the University of Technology, Changchun, China, in 1982, and the Ph.D. degree in earthscope from Jilin University, Changchun, in 2001. She is a Professor of communication and information system with the Department of Communication Engineering, Jilin University. She has authored more than 150 articles on weak signal detection, real-time signal processing, nonlinear dynamic systems, random signal processing, Radon–Wigner transform, particle filter, time–frequency analysis, weak seismic information extraction, and medical signal processing. She has coauthored two books, *Methodology of Periodic Oscillator Detection* and *Periodic Oscillator System and Detection*.



Xintong Dong received the B.S. degree from the Department of Communication Engineering, Jilin University, Changchun, Jilin, China, in 2012, where he is pursuing the Ph.D. degree.

His main research interests include seismic event identification, weak signal detection, and seismic record denoising based on deep learning.