International Journal of Sensors and Sensor Networks
2022; 10(2): 33-50
http://www.sciencepublishinggroup.com/j/ijssn
doi: 10.11648/j.ijssn.20221002.13
ISSN: 2329-1796 (Print); ISSN: 2329-1788 (Online)

Research Status and Prospects of Generative Adversarial Networks in Seismic Data Denoising

Yuan Xu[1, 2], Qing Wang[1, 2, *], Shiguang Guo[1, 2]

[1]Key Laboratory of Information and Communication Systems, Ministry of Information Industry, Beijing Information Science and Technology University, Beijing, China

[2]Key Laboratory of the Ministry of Education for Optoelectronic Measurement Technology and Instrument, Beijing Information Science and Technology University, Beijing, China

Email address:

xuyuan0930@bistu.edu.cn (Yuan Xu), qingwang@bistu.edu.cn (Qing Wang)

[*]Corresponding author

To cite this article:

Yuan Xu, Qing Wang, Shiguang Guo. Research Status and Prospects of Generative Adversarial Networks in Seismic Data Denoising. *International Journal of Sensors and Sensor Networks*. Vol. 10, No. 2, 2022, pp. 33-50. doi: 10.11648/j.ijssn.20221002.13

Received: September 13, 2022; **Accepted**: September 29, 2022; **Published**: October 17, 2022

Abstract: Efficient and high-quality processing of seismic data collected by geophone sensors is the core of successful seismic exploration. Seismic denoising is a key step in seismic data processing. Traditional seismic denoising relies on manual empirical parameter selection and comparative analysis, which is time-consuming and limited by subjective errors. Deep learning methods based on convolutional neural networks (CNN) have improved the efficiency of denoising massive amounts of seismic data and reduced the manual errors of traditional methods. However, most CNN methods only consider data loss and are weak in recovering the structure of seismic signals, resulting in severe attenuation of some seismic traces in the recovered effective signals, reducing the continuity of seismic events and the quality of seismic data. Generative adversarial networks (GAN), a popular method for deep learning with unique adversarial ideas and powerful feature extraction capabilities, can overcome the limitations of CNN methods in the field of seismic data denoising. This paper firstly introduces the classification and development process of seismic denoising. Then, starting from the principle of GAN, it introduces the workflow of the original GAN, the objective function in the training process, the existing problems of the original GAN and some mainstream solutions to these problems, and introduces the commonly used model of GAN in the field of earthquake denoising. besides, it summarizes and analyzes the current application and improvement innovation of GAN in the field of seismic denoising, and analyzes the application of GAN in the field of seismic denoising from two aspects of supervised learning and unsupervised learning with examples. Finally, the prospect of GAN for seismic denoising in the future is prospected.

Keywords: Deep Learning, GAN, Seismic Denoising

1. Introduction

Oil and gas are important sources of energy, and products derived from them are present in all aspects of life. Currently the most useful in oil and gas exploration is seismic exploration. Seismic exploration is subdivided into three main parts: acquisition, processing and interpretation. Seismic processing is an intermediate link in seismic exploration. The quality of processing directly determines the success or failure of seismic exploration. In recent years, as the environment of oil and gas exploration targets has become more complex, the requirements for seismic data processing technology have also become more demanding [1, 2]. Seismic denoising is one of the key technologies in seismic data processing, and how to maximize the suppression of seismic data noise and enhance the signal-to-noise ratio of seismic data while protecting the amplitude and frequency band has been the focus of research [3, 4]. However, traditional seismic data denoising methods usually start from a specific perspective, and require artificial settings of processing parameters during data processing, and the processing effect relies on the

experience of handlers. Facing more and more vectors and a large amount of seismic data, there are some subjective factors lead to data errors [5-7].

With the increase in computer computing power, deep learning techniques have been developed. Data-driven deep learning techniques relies less on human involvement and can automatically complete data processing after model training, which is more suitable for the denoising requirements of vector and large amount of seismic data, bringing new opportunities for seismic data denoising [8–11]. Among them, the method based on convolutional neural networks (CNN) is the most effective in terms of noise suppression, achieving a new breakthrough over traditional methods. Zhang first proposed the use of residual convolution for seismic random noise suppression and achieved better processing results than traditional methods [12]; Wang and Jin verified the performance of the CNN denoising algorithm using synthetic data [13, 14]; Han and Li and Yang used residual convolution for seismic data denoising and proved the effectiveness of the model through simulated data testing and application to actual data [15-17], Li also verified the high dependence of the model on high-quality labeled data [16]. Wang proposed a CNN denoising method based on data enhancement to address the problem that CNN methods rely on high-quality labeled data, which can effectively remove noise and protect effective signals compared with traditional methods [18]. Dong and Zhao illustrated the limitations of traditional denoising methods based on the characteristics of desert seismic data which are severely disturbed by low frequency noise, and used CNN for desert seismic data denoising to verify its great potential in the field of desert seismic denoising [19, 20]. In addition to applying CNN methods for noise suppression on 2D seismic data, researchers have proposed a 3D denoising convolutional neural network combined with sample screening based on the non-Gaussian characteristics of 3D seismic random noise, which has higher processing efficiency than traditional methods [21, 22].

Observations of CNN denoising studies have shown that two problems remain with CNN methods for denoising seismic data. First, most CNN methods for seismic denoising are based on supervised learning, which requires the construction of a large number of labeled pairs of data. Although the actual seismic data are large in volume, it is impossible to obtain clean data for labeling due to the interference of complex noise. If the denoising results of other methods are used as labels, the final denoising results may tend to be mapped by traditional methods, losing the meaning of noise suppression using CNN methods [23]. Secondly, the optimization process of CNN-based denoising methods usually involves minimizing the Mean Square Error (MSE) between the denoised data and the theoretically pure data, and MSE have very limited ability to measure structural integrity because they are defined on a per-sample basis, so this approach lacks a description of the data distribution and in particular the structural features [24].

In order to overcome the shortcomings and limitations of CNN denoising, Generative Adversarial Networks (GAN) began to be studied and applied. The GAN proposed by Goodfellow is an unsupervised learning method that learns data distribution indirectly [25]. Unlike traditional generative models that require explicit modelling of data distribution, it is a more efficient method that obtains high quality generative data through competition between generators and discriminators, and guides the integrity of the generated data structure through overall grasp of adversarial training. In the field of seismic data processing, Alwon has conducted a preliminary study on the feasibility of generative adversarial networks for seismic data processing by constructing appropriate generators and discriminators [26]. Picetti designed a generative adversarial network for seismic image processing that can convert low-quality offset images to high-quality images and also convert offset images to the corresponding reflectance images [27]. Subsequently, some papers on the application of GAN to seismic data denoising have been published. Wang and Ma combined attribute coding with GAN for noise suppression of desert seismic data, recovering effective reflections and attenuating noisy signals by introducing new losses, and even recovering weak signals heavily contaminated by seismic noise [28, 29]. Zheng trained a GAN model end-to-end using synthetic seismic data on a single-channel time-series-waveform and verified that the method outperformed conventional methods in terms of performance on real micro seismic data [30]. Li combined an attention module guided by Class Activation Maps (CAM) with GAN to better distinguish between noisy and active signals, resulting in denoising results with less false reflection signals [31]. Yu used DCGAN for random noise suppression of lightning wave signals. By using a model trained on synthetic data to denoise real data, the visual denoising effect and dispersion curves from seismic data proved that the DCGAN model can effectively denoise seismic data and outperform traditional denoising methods [32]. Liu and Li used CycleGAN for seismic data denoising [33, 34]. Liu compared the denoising performance of CycleGAN, GAN and ResNet [35] at different noise levels by quantitative means, and demonstrated that with the enhancement of noise, the model constructed by generative adversarial network, the denoising effect has significant anti-saturation ability. These studies and applications demonstrate the potential and advancement of GAN for seismic data denoising. Figure 1 shows the classification and development of seismic denoising methods, and researchers have been exploring smarter and more efficient methods for denoising seismic data.

Starting from the original GAN, this paper introduces its basic principles, problems and derived models, and generally summarizes the current status of GAN improvements and research applications in seismic denoising. The principles of the different methods and the problems solved are analyzed with examples from both supervised and unsupervised learning, and the prospects for the development of GAN in the denoising process of seismic data are discussed.

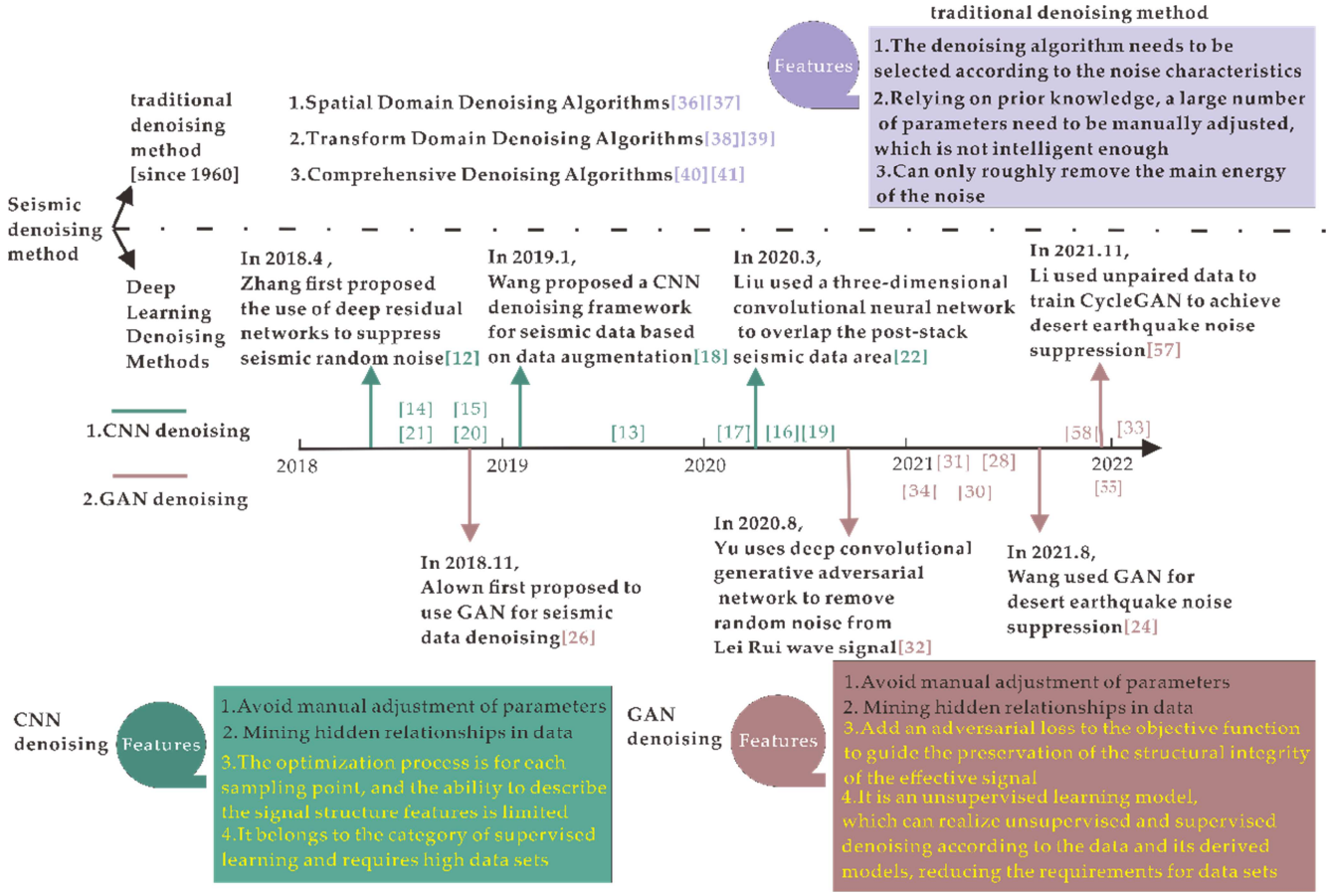


Figure 1. *Classification and Development History of Seismic Denoising Methods (Among them, the traditional denosing method references [36-41]).*

2. The Theory of GAN

2.1. Principle of Operation

The GAN is a network model inspired by the idea of zero-sum games in game theory [42], which consists of a generator and a discriminator, as shown in Figure 2. The generator is used to generate the data and the discriminator is used to determine whether the input data is real or generated by the generator. The generator is used to "deceive" the discriminator, and the discriminator is used to improve the discriminatory ability. The two are constantly optimizing themselves and playing against each other to achieve a dynamic balance.

A usable GAN is generally trained several times iteratively, with the general training process being: 1) Initialize the generator and discriminator parameters. 2) Perform the following in each round of training: a) Fix the generator parameters, train the discriminator parameters, and input the real samples and generated sample data into the discriminator respectively, so that the real data is judged as real, and the data generated by the generator is judged as fake. b) Fix the discriminator parameters and train the generator parameters. A set of random noise is fed into the generator to generate "fake" data. The generated "fake" data is passed into the discriminator, which then discriminates the generated data. The motivation for the generator optimization is to generate data that more closely matches the distribution of the real data, so that the discriminator cannot accurately determine whether the input data is generated data or real data.

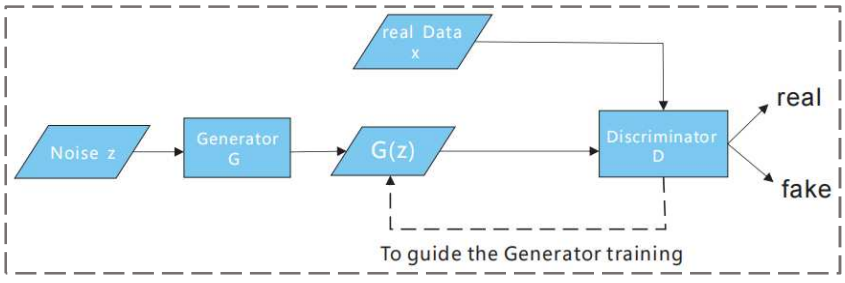


Figure 2. *Generating adversarial network's structure and training process.*

2.2. Objective Function

The main goal of GAN is to minimize the distance between the true distribution and the generated distribution. The original GAN model uses Jensen-Shannon (JS divergence) [43] to measure the difference between the two distributions. According to this standard objective function is designed as

$$\min_{G}\max_{D} V(D,G) = E_{x\sim P_r}[\log D(x)] + E_{z\sim P_g}[\log(1-D(G(z)))] \quad (1)$$

where D denotes the discriminator, G denotes the generator, x denotes the true data, P_r denotes the true data probability distribution, z denotes the generated data, and P_g denotes the generated data probability distribution. For the discriminator, its objective function is to maximize V(D,G) , that is, try to measure the distance between the generated distribution and the true distribution (measure the loss). For the generator, the goal is to minimize this distance (minimize the loss), ultimately allowing the generator to produce data that resembles the distribution of the real data.

2.3. Problems with the Original GAN

The original GAN was immature and suffered from many problems, among which gradient disappearance and training instability and pattern collapse severely limited the development of GAN. The reasons for the above problems are analyzed below:

1) Gradient disappearance

Starting with the objective function, the discriminator measures the optimal distance and the generator can generate fake data disguised as real data. The objective function of the generator at this point is

$$C(G) = 2JS[P_r \,\|\, P_g] - 2\log 2 \quad (2)$$

It can be seen that the goal of the generator optimization is to reduce the JS divergence of the two distributions by continuous training. However, in GAN training, there is a high probability of no or negligible overlap between the generated data distribution and the real data distribution. At this time, no matter what the true distance between the two distributions is, the value of the JS divergence is always log2. This causes the gradient of the generator to disappear and prevents further optimization of the generator through training.

2) Training Instability and Pattern Collapse Severely

The essence of pattern collapse is that the generator no longer seeks diversity in order to avoid greater penalties by producing large amounts of similar data. From (1) and (2) and divergence [44], the deformed generator objective function is

$$L_G = KL(P_g \,\|\, P_r) - 2JS(P_r \,\|\, P_g) \quad (3)$$

$KL[P_g \,\|\, P_r]$ expands according to the KL divergence definition as

$$\mathrm{KL}[\mathrm{P_g} \,\|\, \mathrm{P_r}] = \mathrm{P_g(x)}\log\frac{\mathrm{P_g(x)}}{\mathrm{P_r(x)}} \quad (4)$$

From formula (3) (4), we can see two problems in GAN training: a) The essence of training generators is minimization L_G . From (1)-(3) it can be seen that minimization means both minimizing the divergence of KL between the generated data and the real data and maximizing the divergence of JS between the two distributions. This leads to gradient oscillations, causing the GAN training to become unstable. b) Formula (4) punishes different errors differently. From equations (4), it can be seen that when P_g tends to 0 and P_r tends to 1, the KL divergence is close to 0 and the generator obtains a small gradient penalty; when P_g tends to 1 and P_r tends to 0, the KL divergence tends to infinity and the generator obtains a huge gradient penalty. Among them, the first error indicates the lack of diversity of the generator, and the second indicates the lack of accuracy of the generator. Since the second penalty is much larger than the first, the generator is more willing to generate some duplicate data, resulting in the phenomenon of mode collapse.

2.4. Improvements to the GAN

An analysis of the original GAN problem shows that the cause of a number of problems is that its objective function uses JS divergence to measure the distance between the generated distribution and the true distribution, and in order to solve the problems posed by the original GAN, the researchers have chosen to put aside the JS divergence to seek a breakthrough.

2.4.1. WGAN and WGAN-GP

In response to the problems of the original GAN, ARJOVSKY proposed Wasserstein GAN (WGAN) [45]. It uses the EM distance to measure the distance between two distributions, and the resulting objective function of the WGAN is derived as

$$V(G,D) = \max_{D\in 1-Lipschitz}\{E_{x\sim P_{data}}[D(x)] - E_{x\sim p_g}[D(x)]\} \quad (5)$$

The advantage of EM distance is that the distance between two distributions can still be measured even if there is no overlap between them or if the overlap is negligible. This avoids problems such as gradient disappearance, training instability and pattern crashes. However, in practical experiments WGAN often suffers from difficulties in training and slow convergence. This is due to the fact that the parameters are directly trimmed to the range [-c, c] using weight trimming in order to satisfy the Lipschitz restriction, forcing the discriminator's objective function to be smooth. At this point in the optimal strategy, the parameters tend to go to extremes, either taking maximum or minimum values, and the fit of the model becomes poor, while the parameter c is difficult to determine and can cause the gradient to disappear or explode if it is not chosen properly.

Gulrajani proposed an improved version of WGAN, WGAN-GP [46], which uses a gradient penalty to satisfy the Lipschitz continuity condition. This is done by adding a gradient penalty (extra loss) to the discriminator objective function, linking the parameters to the constraints to achieve the Lipschitz restriction. The optimized objective function is

$$V(G,D) \approx \max_D \{E_{x\sim P_{data}}[D(x)] - E_{x\sim p_g}[D(x)] - \lambda E_{x\sim penalty}[\max(0, \| \nabla_x D(x) \| - 1)]\} \quad (6)$$

The penalty term in this objective function is used to penalize behavior greater than 1 on gradient update, thus solving the problem of parameter centralization and the difficulty of determining the extent of cropping due to weight cropping in WGAN.

In addition to improving GAN from the objective function, researchers have also combined GAN with other ideas to improve GAN models, allowing the adversarial learning ideas of GAN to permeate applications in a wider range of fields.

2.4.2. DCGAN

Convolutional neural networks (CNN) have shown great advantages in feature extraction making them widely used in the field of supervised learning, but their application in unsupervised learning has been less than satisfactory. Both the generator (G) and discriminator (D) of the original GAN use fully connected neural networks, which gives it limited ability to extract features. The convolutional generative adversarial network (DCGAN) proposed by Radford [47] combines a deep convolutional neural network (DNCNN) [48] with a GAN, replacing the network structure of both the generator and discriminator of the GAN with a deep convolutional neural network, using the powerful feature extraction capability of the DNCNN to enhance the generative capability of the generator and the discriminatory capability of the discriminator. In addition to using CNN as the main structure of the generator and discriminator, DCGAN also uses a series of training techniques, such as using Batch Normalization (BN) technique to solve the problem of deeper training of the network structure; eliminating pooling and using filters with step size to complete the convolution operation, effectively retaining the feature information; using ReLu activation function for all layers except the last layer of the generator, which uses Tanh activation function to reduce the risk of gradient disappearance; using Leaky ReLu as the activation function for all layers of the discriminator. DCGAN is suitable for most image generation scenarios.

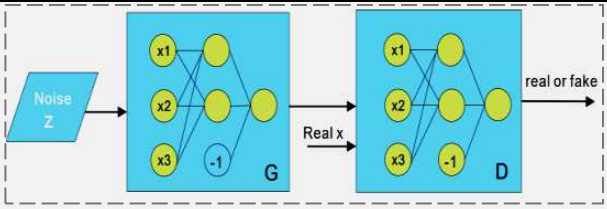
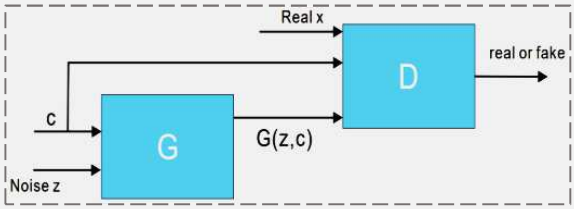
2.4.3. CGAN

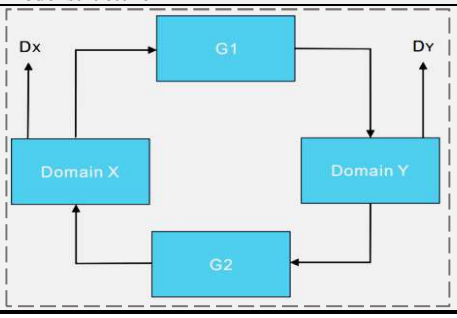
The generator input to the original GAN is a random noise vector, which cannot be controlled to generate the specified data. Mirza proposed Conditional Generative Adversarial Networks (CGAN)[49], which adds extra information c to both the generator and discriminator, turning unsupervised GAN into supervised learning, an improvement that enables the network to generate data in a prescribed direction. GAN applies to the scenario for which the object is specified to be generated.

2.4.4. CycleGAN

The CGAN mentioned above is adding extra information, and its extra information is the label information, which is equivalent to the fact that CGAN training must have labeled pairs of data, which limits its application scenarios. In order to address this problem and enable the transformation between different domains, researchers have proposed a cyclic consistent generative adversarial network (CycleGAN) [50], trained using unpaired data, in conjunction with pairwise learning [51]. The core idea of CycleGAN is cyclic consistency. The data in the x-domain is transformed to the y-domain by generator G1, which can then be transformed to the data in the x-domain by generator G2 (as in the CycleGAN model structure in Table 1). CycleGAN is suitable for scenarios where styles are shifted and data are not paired. Table 1 lists a comparison of the features of each derived model.

Table 1. *Comparison table of derived models of GAN.*

Name	Model structure	Advantages	Disadvantages	Application Scenarios
DCGAN		Generate rich images	Poor quality of generated images and unstable models	Most image generation tasks
CGAN		Generate specified data based on the given information	Need labelled dataset, transformed to supervised learning, unbalanced training, poor quality of generated data	Scenarios with additional given information

Name	Model structure	Advantages	Disadvantages	Application Scenarios
CycleGAN		Low data requirements, no need for pairs of images, just images of different domains	Generate images with low resolution	Most domain transformation scenarios

3. Application of GAN in Seismic Data Denoising Processing

As GAN has made breakthroughs in other fields, researchers have thought of introducing GAN to the field of seismic denoising. To get a clearer picture of the current research status of GAN in the field of seismic denoising, Table 2 shows the applications of GAN in the field of seismic denoising in recent years. As can be seen from Table 2, there are three main directions of innovation in applying GAN to the field of seismic data denoising: first, application field innovation, applying GAN methods to uncovered seismic data, verifying the feasibility of GAN denoising methods and comparing them with traditional denoising methods; second, improving the structure of the original GAN model by replacing the fully connected network structure of the original GAN model with a network structure that is more capable of extracting features, making the model more capable of denoising; third, improving the loss function by introducing other ideas to improve the denoising capability while enhancing the protection of the effective signal.

Table 2. *Comparison of GAN in seismic denoising.*

Methods	Innovations	Learning styles	Comparison methods	Evaluation indicators
GAN [30]	Innovation in application: GAN for micro seismic data denoising	Supervised learning	EEMD, DWT	Single-channel time-sequence waveform plot
DCGAN [32]	Innovation in application: GAN for random noise removal from Ray Ripple signals	Supervised learning	F-X deconvolution, wavelet transform	Seismic data maps, dispersion curves
CGAN [52] [53, 54]	Innovation in application: for ground roll noise removal [53] Structural improvements: generators use U-Net structures	Supervised learning	S transform, AGRA, and f-k filtering	Seismic data maps, amplitude spectra, SNR spectra, PSNR, SSIM
DNGAN [24] CADN [55]	Loss function improvement: introducing data loss	Supervised learning	band-pass filter, RED-Net, original GAN, DNGAN, EMD, RPCA	Seismic data maps, SNR, MSE
CycleGAN [33, 34, 56]	[56] Loss function improvement: introducing signal estimation to obtain training data [33, 34] Structural improvement: residual-based learning	Supervised learning	LTF, RNR FXDM, DNCNN Original GAN, ResNet	Seismic data maps, spectral curves, PSNR, MSE
CycleGAN [57, 58]	Structural improvements: generators use deep residual networks	Unsupervised learning	Band-pass filter, wavelet threshold method, damping reduction rank method	Seismic data maps, SNR, MSE (RMSE), amplitude spectra
Att-DCDN [28] RAGAN [29]	Loss function improvement: introducing attribute information	Unsupervised learning	VMD, DNCNN	Seismic data maps, SNR, MSE, amplitude spectra, power spectra
U-GAT-IT [30]	Loss function improvement: introducing an attention mechanism	Unsupervised learning	Wavelet transform, F-X deconvolution, DNCNN	Seismic data maps, SNR, MSE

When performing denoising tasks using deep learning methods, CNN methods are widely used in seismic denoising because they can map the data to a non-linear space and can automatically perform the extraction of potential features, achieving better results than traditional methods [59]. However, the CNN method still suffers from weak ability to restore the structural integrity of the signal and strong reliance on labelled data. Generative adversarial networks are highly effective in guiding data to recover structural integrity due to their unique adversarial properties. Currently, deep learning algorithms can be divided into two categories, supervised learning and unsupervised learning, depending on whether the dataset is labeled or not [60]. In particular, supervised learning requires the construction of a training set based on pairs of data containing labels, which is demanding on the dataset; unsupervised learning can solve the problem of the training model relying on the labels of the training set and reduce the requirements of the dataset. In the following, the application of GAN in seismic denoising is analyzed with examples from both supervised and unsupervised learning.

3.1. GAN Denoising Based on Supervised Learning

The initial function of a GAN is to generate a large amount of data with probability distributions close to the probability distribution of real data, an unsupervised learning process [61-64]. In seismic denoising, Dong and Wang replaced the generator in the GAN with a denoiser having a CNN structure, using the constructed paired dataset and guiding the CNN denoising network through a discriminator to optimize the parameters [24, 55]. As shown in Figure 3 is the flowchart of its implementation of seismic data denoising using GAN, which mainly includes two parts: denoiser and discriminator. The denoising process is as follows: the synthetic seismic data (a superposition of theoretically pure seismic data and actual noise) is fed into the denoiser, which denoises the data by constructing a loss between the theoretically pure seismic data and the output data of the denoiser and continuously minimizing this loss to optimize the denoising result, thus obtaining denoised data. The discriminator improves the ability to discriminate between denoising data and pure data, and the denoiser improves its denoising ability to generate purer data to "deceive" the discriminator. The final discriminator guides the denoiser to generate denoising results with better continuity of seismic events.

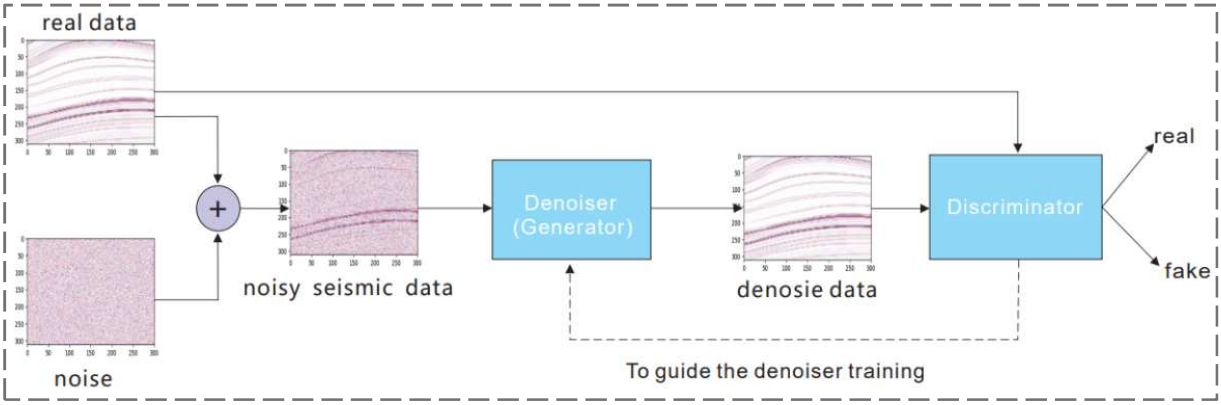


Figure 3. *Denoising Flowchart.*

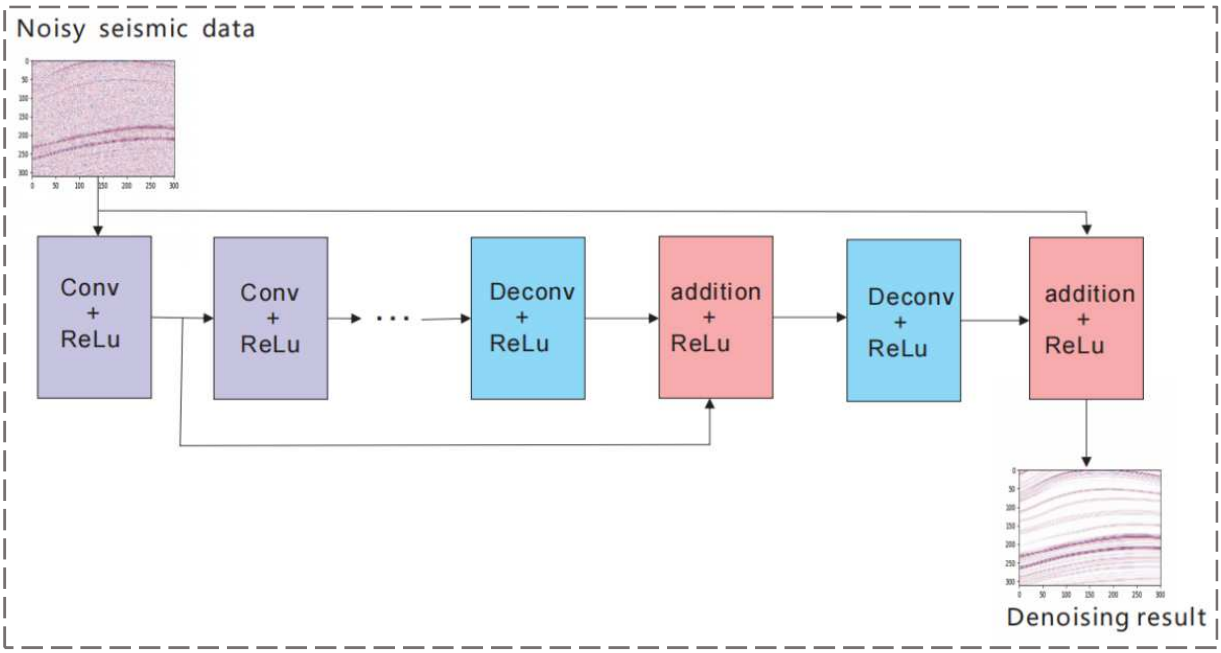


Figure 4. *Denoiser construction.*

The denoiser constructed in Figure 3 is a convolutional neural network based on RED-Net [65], whose structure is shown in Figure 4, with Conv denoting the convolutional layer, whose basic function is to extract potential features of the input data and obtain a feature map. Deconv denotes deconvolution, which is the inverse process of convolution. Deconvolution allows the synthesis of low-dimensional data based on the reconstruction of high-dimensional features extracted by convolution [66]. ReLu is a commonly used nonlinear activation function that can increase the speed of gradient descent and reduce operating costs [67]. The basic functions of ReLu are

$$f(r) = \max(0, r) \tag{7}$$

Where r denotes the input to the ReLu activation function and Addition denotes the feature map obtained by convolution and the feature map obtained by inverse fold convolution, using a skip connection, to achieve feature map fusion. Feature map fusion can lead to additional performance gains for denoising models. Firstly, the back propagation of the gradient to the bottom layer is achieved, which solves the problem of gradient disappearance, accelerates the convergence of the network and reduces the difficulty of network training [68]. Secondly, as the depth of the convolution layer increases, low-dimensional features of the signal are transformed into high-dimensional features resulting in loss of detail information. The skip connection allows the signal to be passed directly from the convolutional layer to the deconvolutional layer for feature map fusion, facilitating the recovery of the original signal and improving the signal-to-noise ratio.

Normally noisy data is defined as

$$y = e + n \tag{8}$$

where y represents noisy data, e represents theoretically pure data (labels), and n represents noise. The end-to-end mapping relationship from noisy data to pure data established by the denoiser is

$$\tilde{e} = G(y;\theta) \tag{9}$$

where $\tilde{e}$ denotes the denoised data, G denotes the mapping relationship established by the denoiser and θ denotes the network parameters. In order to optimize the mapping relationship and continuously reduce the gap between the pure and denoised data, it is necessary to construct a loss function. The optimal mapping is obtained by minimizing the losses to obtain denoised data with high signal-to-noise ratio. The MSE is used in the paper as a reconstruction loss function to optimize the denoising process. It is calculated as

$$L_{MSE} = \frac{1}{2M}\sum_{i=1}^{M} \| G(y_i;\theta) - e_i \|_F^2 \tag{10}$$

where M denotes M pairs of data, $\| \bullet \|_F$ denotes the F norm and i denotes the i-th pair of data. It can be seen by (10) that MSE tends to generate smooth predicted signals and lacks high precision textures, resulting in its limited ability to characterize the structure of the signal. This requires the discriminator to be involved in forming a confrontation with the denoiser, mitigating this problem through confrontation.

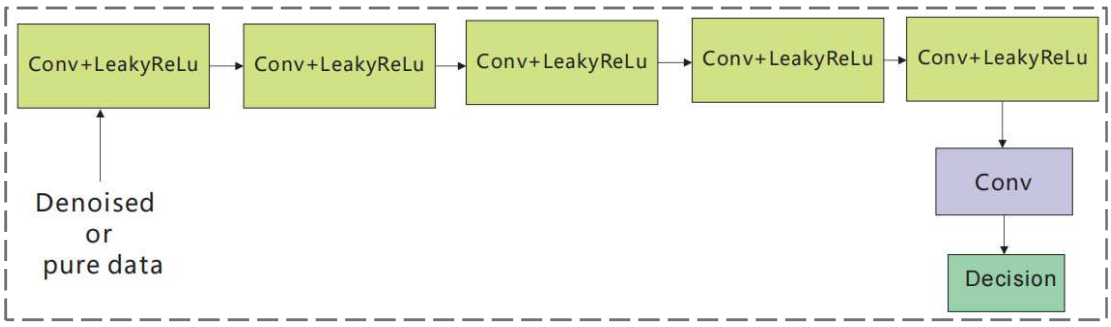


Figure 5. *Discriminator structure.*

Figure 5 shows the detailed structure of the discriminator in Figure 3. It is composed of multiple convolutional layers, and the discriminator, unlike the denoiser, is used to compress the classification, so the use of LeakyReLu as the activation function is more conducive to reducing information loss. The discriminator continuously extracts the data features of denoise data and theoretically pure data through multiple convolutional layers, discriminates the data classes, constrains the denoiser to maintain a good signal structure, and finally obtains a denoised signal with a complete structure and high signal-to-noise ratio. This adversarial process can be achieved by optimizing the adversarial loss, which is defined as

$$L_{\mathrm{ad}}(D,G) = E_{e\sim p_{data}}[\log D(e)] + E_{\tilde{e}\sim p_d}\{\log[1 - D(\tilde{e})]\} \tag{11}$$

Where $L_{ad}(D,G)$ denotes the adversarial loss, E is the mathematical expectation, e denotes the theoretically pure data, p_{data} denotes the theoretically pure data probability distribution, $\tilde{e}$ denotes the denoised data, p_d denotes the probability distribution of the denoised data and D is the mapping relationship established by the discriminator. From equations (10) and (11), it can be seen that in order to obtain better denoising performance, the denoiser tries to minimize the MSE and the discriminator needs to improve its discriminatory power. At this point it is necessary to maximize $L_{ad}(D,G)$, so the adversarial objective function of the denoiser and discriminator is defined as

$$\mathrm{G}^*, \mathrm{D}^* = \arg\min_{\mathrm{G}} \max_{\mathrm{D}} \mathrm{L_{ad}}(\mathrm{D},\mathrm{G}) \tag{12}$$

Where $\min_G$ denotes minimizing the denoiser loss, $\max_D$ denotes maximizing the adversarial loss, and arg denotes the value of the variables in D and G when the denoiser loss is minimized and the adversarial loss is maximized. Based on the optimization objective, a new loss function is constructed as

$$L = L_{MSE} + \eta L_{ad}(D, G) \quad (13)$$

Where η is a balancing factor to balance the proportion of MSE and counteracting losses. The value of this hyper-parameter η is determined by assigning different values to η and training the network model several times.

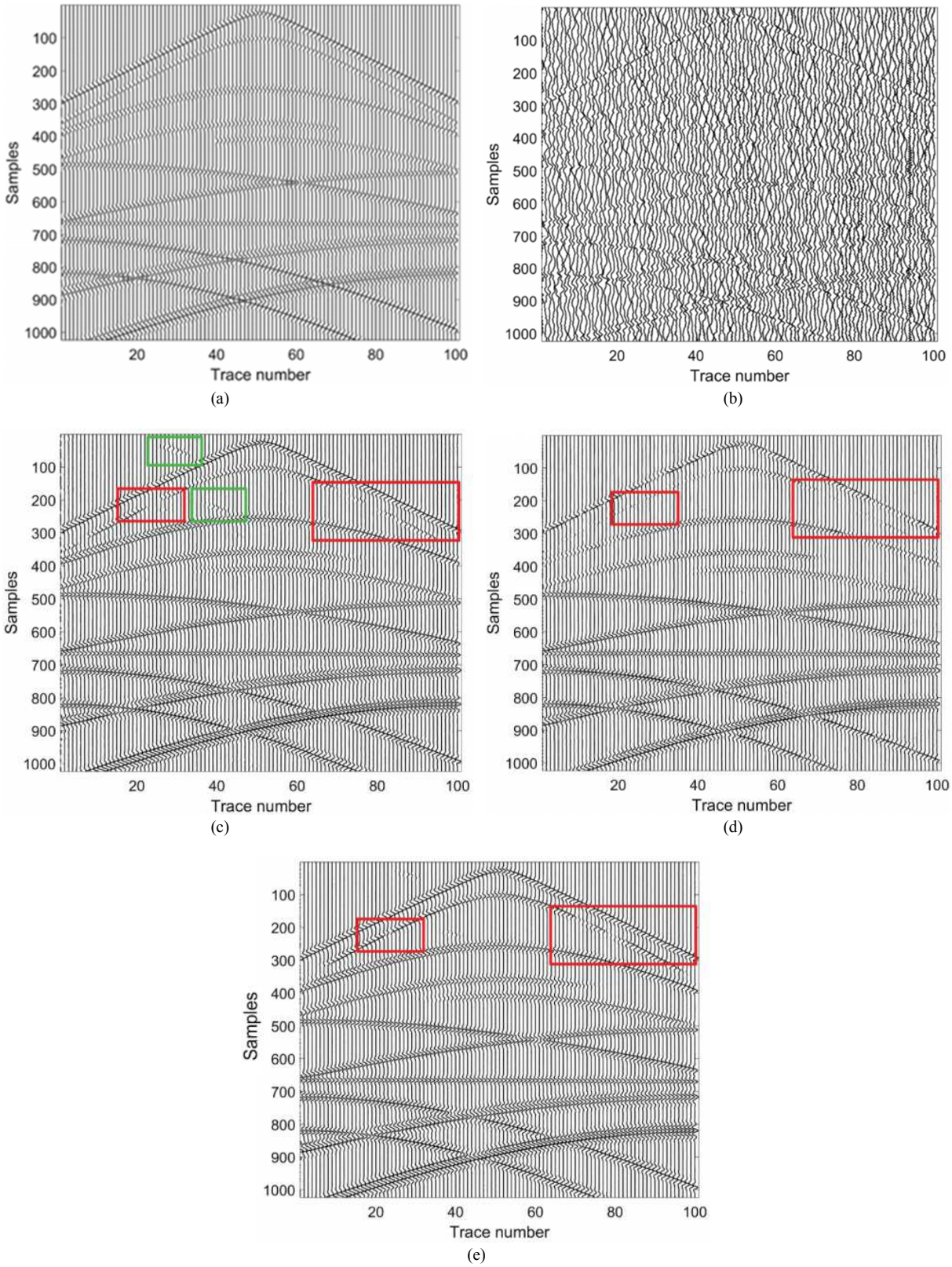


Figure 6. *Analysis of denoising results of forward synthesis data. (a) Theoretical desert seismic data;(b) Synthetic noisy seismic data; (c) Denoising result of RED-Net; (d) Denoising result of original GAN; (e) Denoising result of DnGAN. (Resource: [24]).*

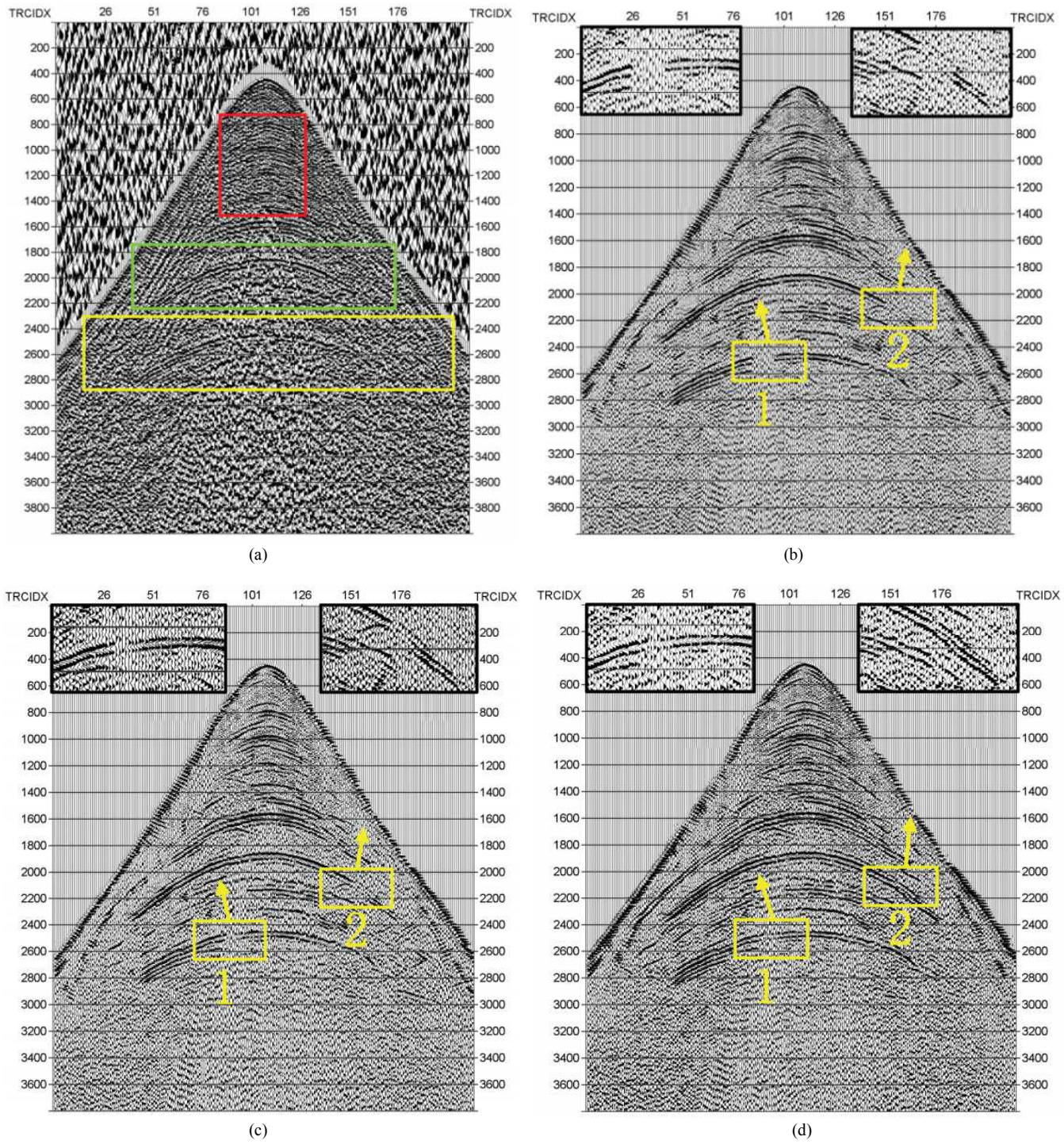


Figure 7. *Analysis of the denoising results of really noisy seismic data. (a) Field data in the desert region; (b) Denoising results of the field data by RED-Net; (c) Denoising results of the field data by original GAN; (d) Denoising results of the field data by DnGAN. (Resource: [24]).*

Figure 6 shows the results of denoising tests on different deep learning models using seismic model forward data. Among them, (a)-(e) denote theoretically pure data, synthetic noise-bearing data, RED-Net denoising results, original GAN denoising results, and DnGAN (the model in Figure 3 is named DnGAN in [24]) denoising results, respectively. From (c) in Figure 6, it can be seen that some spurious seismic events (green boxes) appear in the denoising results of RED-Net (a denoising model using only CNN structure), which is caused by using only MSE as loss. In addition to this, a comparison of (a) and (e) in Figure 6 reveals that in the case of (a), in addition to the presence of spurious seismic events, the continuity of seismic events is not as good as in (c) (red box). As can be seen from (d) and (c) in Figure 6, the denoising results using DnGAN [24] have better seismic event continuity, and the original GAN, although showing weak event repair, is not as effective as DnGAN (red box). This would suggest that a model structure combining data

loss and adversarial loss is more effective in restoring seismic event continuity and suppressing spurious seismic events. This advantage has also been demonstrated in the processing of actual seismic data. Figure 7 shows the results of each model for real seismic data. (a)-(d) in Figure 7 represent the real noisy seismic data, RED-Net denoising result, original GAN denoising result, and DnGAN denoising result, respectively. In this case, it can be seen from (a) in Figure 7 that in addition to low frequency desert seismic noise (red boxes), there is also strong surface wave interference (green and yellow boxes) in the real noise-bearing seismic data. However, as can be seen from (b)-(d) in Figure 7, the deep learning denoising method not only suppresses the desert noise at low frequencies, but also does a good job of suppressing the surface waves. And as can be seen from the enlarged portion (yellow box) labelled in (b)-(d) in Figure 7, DnGAN outperforms RED-Net and the original GAN in terms of event repair.

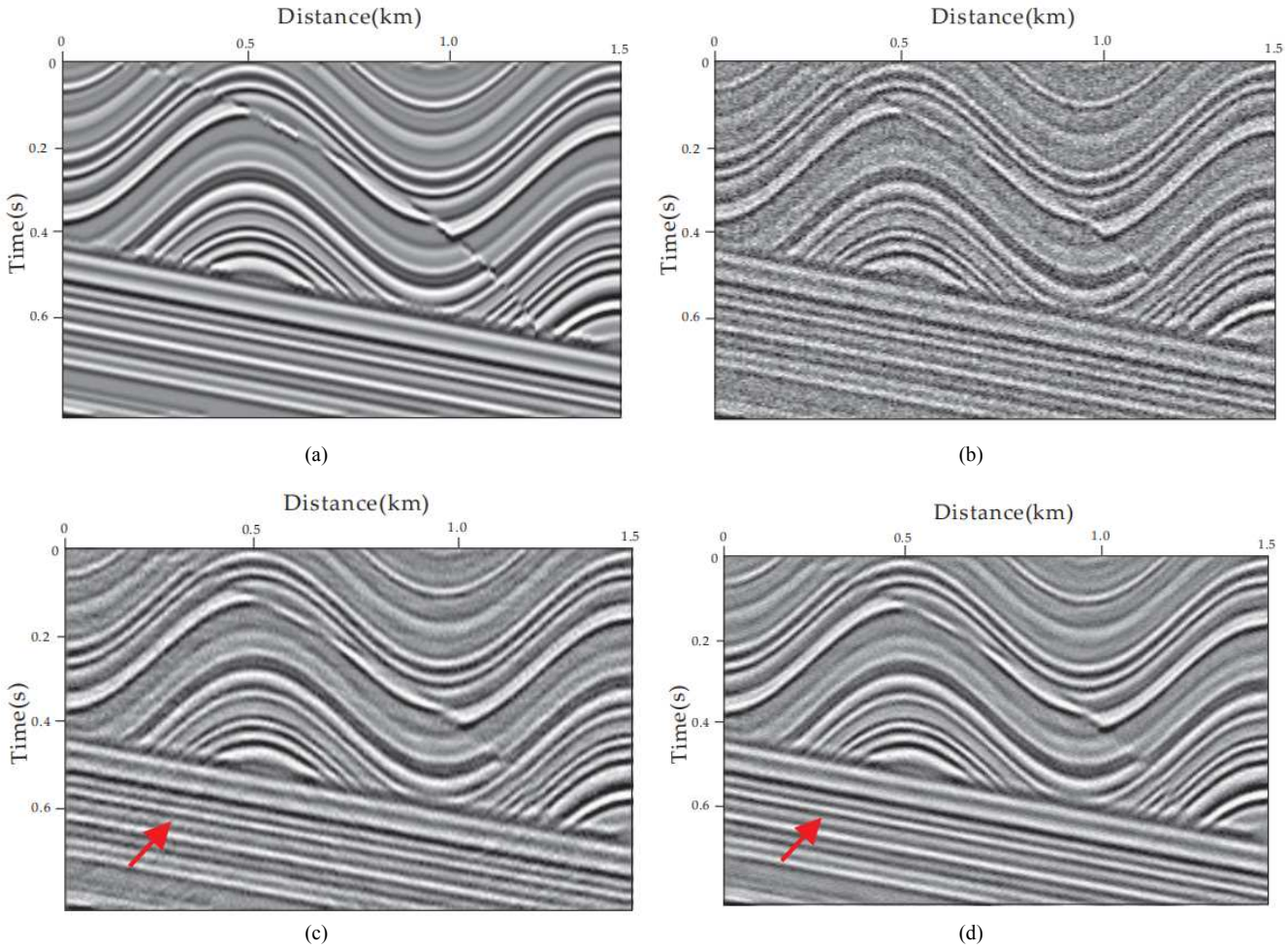


Figure 8. *Analysis of denoising results of forward synthesis data. (a) Theoretical seismic data; (b) Synthetic noisy data; (c) Denoising results of DNCNN; (d) Denoising results of RCGAN. (Resource: [34]).*

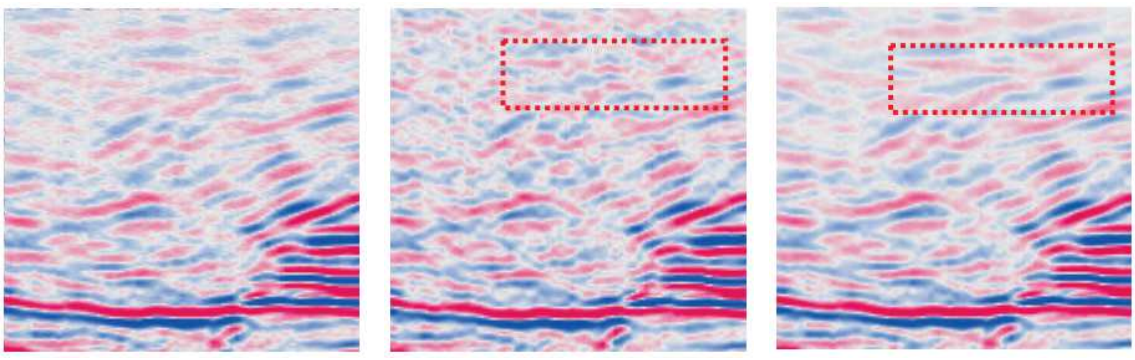


Figure 9. *A partial zoom of the denoising results of the real seismic data. (a) Field data; (b) Denoising results of the field data by DNCNN; (c) Denoising results of the field data by RCGAN. (Resource: [34]).*

In besides, Li applied CycleGAN (the denoising process of CycleGAN will be described in detail later) to seismic data denoising, naming the CycleGAN with the addition of residual networks as RCGAN [34], and trained the RCGAN model using the constructed paired dataset, which also reflects the role of adversarial in restoring continuity of seismic events. Figure 8 shows the results of denoising the synthetic seismic data using this model. Where (c) is the denoising result using DNCNN and (d) is the denoising result using RCGAN. The part indicated in the red arrow shows that although the DNCNN can also suppress the noise signal better, the poor continuity caused by a slight gap in the reflected seismic events. As shown in the red dashed box in Figure 9 for the real data, the DNCNN results in poor continuity of seismic events, while the RCGAN results not only provide effective suppression of seismic noise, but also recover better continuity of seismic events than the DNCNN.

3.2. GAN Denoising Based on Unsupervised Learning

When a GAN based on supervised learning is used for seismic data denoising, the training model must rely on pairs of seismic data. However, actual seismic data is difficult to find matching clean data to use as a target for model learning. CycleGAN, a variant of GAN, was originally designed to achieve mapping of images from source to target domains without using additional input information, rather than a one-to-one mapping relationship between specific data in two domains. In order to reduce the requirement for training samples and improve the generalization ability of the model, Li and Wu used unpaired data to train the CycleGAN model to achieve the mapping from noise-containing domain seismic data to noise-free domain seismic data for the purpose of denoising [57, 58].

The CycleGAN model used to implement the denoising of seismic data consists of two mirror-symmetric GANs.

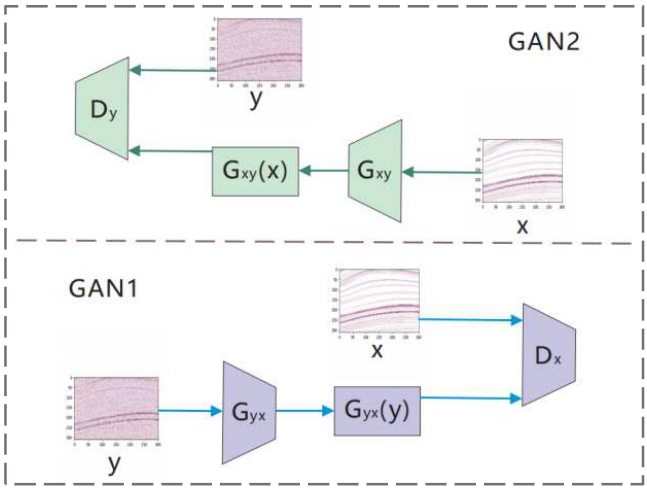


Figure 10. *Denoising model based on CycleGAN.*

Figure 10 shows a seismic denoising model based on the CycleGAN structure, which contains two generators and two discriminators. The process of GAN1 is to map the y-domain (noise-containing domain seismic data) data to the x-domain (clean domain seismic data) by means of the generator G_{yx}, which optimizes the mapping relationship after confrontation with the discriminator D_x. The GAN2 process is the same, taking the data in the x-domain, mapping it to the y-domain after the generator G_{xy}, and using the discriminator D_y to guide the generation of data that approximates the target domain. In this cyclic adversarial training process, the final generator that achieves the mapping of the data from the y-domain to the x-domain is the desired denoiser. In order to optimize the training process and get better domain mapping results, the loss function is constrained by fighting against the loss, which is defined as

$$loss_{GAN1}(G_{yx}, D_x, y, x) = E_{x\sim P_{data}(x)}[\log D_x(x)] + E_{y\sim P_{data}(y)}\{\log[1 - D_x(G_{yx}(y))]\} \tag{14}$$

$$loss_{GAN2}(G_{xy}, D_y, x, y) = E_{y\sim P_{data}(y)}[\log D_y(y)] + E_{x\sim P_{data}(x)}\{\log[1 - D_y(G_{xy}(x))]\} \tag{15}$$

Where $P_{data}(x)$ denotes the clean domain data probability distribution, $P_{data}(y)$ denotes the noisy domain data probability distribution, y denotes noisy data, x denotes theoretically pure data, G_{xy} denotes the generator that implements the mapping from clean domain seismic data to noisy domain seismic data, G_{yx} denotes the generator that implements the mapping from noisy domain seismic data to clean domain seismic data, D_x denotes the clean data discriminator, and D_y denotes the noisy data discriminator.

When training GAN1 and GAN2 using adversarial loss, CycleGAN proposes the use of cyclic consistent loss to constrain this conversion process in order to diversify the data and to avoid the generator converting all the data in the domain to the same data in another domain. This constraint process is shown in Figure 11: after the y-domain data is mapped to the x-domain ($G_{yx}(y)$ via the generator G_{yx}, this data can be mapped back to the y-domain (y-) via G_{xy} (similarly, the x-domain data can be mapped back to the x-domain (x-) via G_{yx} after being mapped to the y-domain ($G_{xy}(x)$) via the generator G_{xy}) [69]. The constraint is to make the target domain data generated by the generator as close as possible to the input data by continuously reducing the cyclic consistency loss, thus maximizing the retention of information about the seismic event. At this point the cyclic consistency loss (reconstruction loss) is defined as

$$loss_{cycle}(G_{yx}, G_{xy}) = E_{x\sim P_{data}(x)}\{\| G_{yx}(G_{xy}(x)) - x \|_1\} + E_{y\sim P_{data}(y)}\{\| G_{xy}(G_{yx}(y)) - y \|_1\} \tag{16}$$

where $\|\bullet\|_1$ denotes the L_1 paradigm. Cyclic consistency loss keeps consistency between inputs x and y and outputs x' and y', guiding the generator to learn purposefully and preventing degradation of the network model [70].

The final loss function is made up of three components

$$loss(G_{yx}, G_{xy}, D_x, D_y) = loss_{GAN1}(G_{yx}, D_y, y, x) + loss_{GAN2}(G_{xy}, D_x, x, y) + \lambda loss(G_{yx}, G_{xy}) \tag{17}$$

where λ is a control factor to control the share of adversarial and cyclic consistent losses in the final loss. Therefore the final objective function is

$$\min_{G_{yx}, G_{xy}} \max_{D_x, D_y} loss(G_{yx}, G_{xy}, D_x, D_y) = loss_{GAN1}(G_{yx}, D_y, y, x) + loss_{GAN2}(G_{xy}, D_x, x, y) + \lambda loss(G_{yx}, G_{xy}) \tag{18}$$

Model training is guided by an optimized objective function, resulting in a generator for seismic data denoising, which is represented as

$$x' = G_{yx}(y) \tag{19}$$

In the same way as GAN denoising with supervised learning, both the generator and discriminator of CycleGAN use CNNs to extract data features. In this case, a skip-connected structure is also used in the generator to avoid network degradation as the number of layers increases, which is used to retain the original data information; the discriminator uses a full convolutional network to continuously extract features and obtain judgement results to guide the generator training.

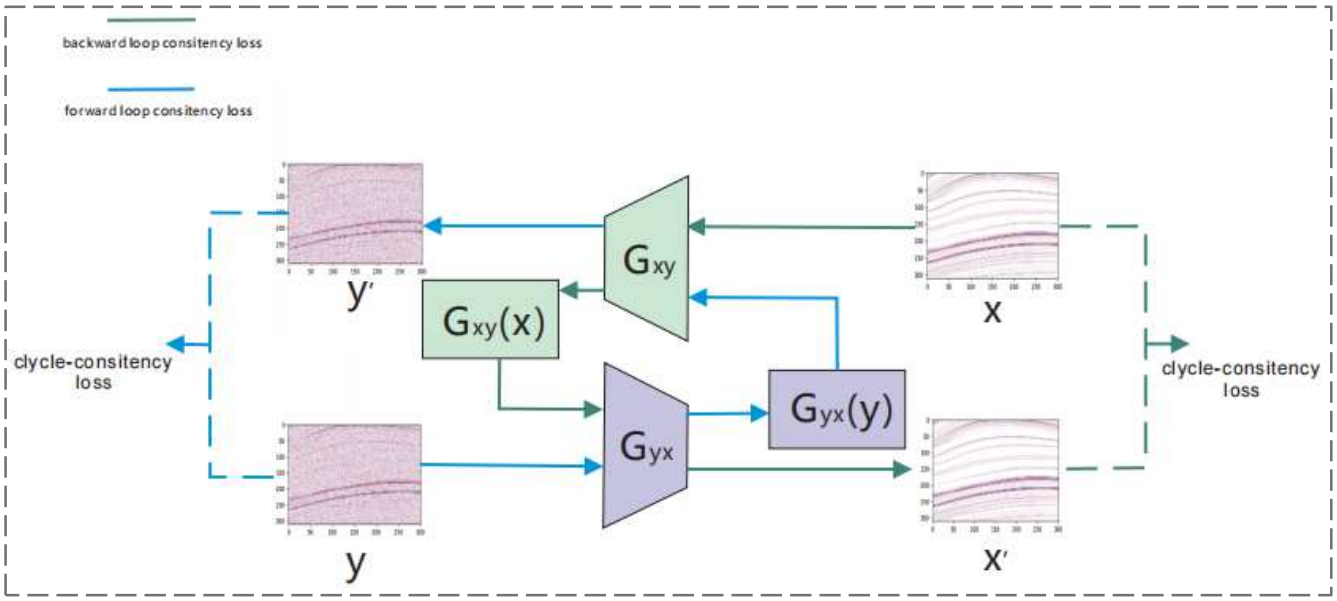


Figure 11. *Cycle Consistent Loss Process.*

Wu used unpaired data to train the CycleGAN model and verified the effectiveness of the CycleGAN model by conducting denoising experiments on synthetic and real seismic data. Figure 12 shows the results of different methods for denoising synthetic seismic data [58]. (a)-(d) in Figure 12 represent theoretically pure data, synthetic noisy data, denoising results using wavelet thresholding, and denoising results using CycleGAN, respectively. As can be seen from (c) and (d) of Figure 12, in the part of (b) that is more severely affected by random noise (red arrow), the situation is improved after denoising by both methods, but CycleGAN has better results. (e) in Figure 12 shows the results of the spectral analysis of the 240th channel of data extracted from the red box in (c) and (d). From (e), it can be seen that in the range of 90Hz to 180Hz, the CycleGAN denoising method almost completely removes the random noise. The effectiveness of CycleGAN trained using unpaired data was similarly validated on real seismic data. (a)-(c) in Figure 13 represent the real noisy data, wavelet threshold denoising result, and CycleGAN denoising result, respectively. As can be seen from the red boxes in (b) and (c) in Figure 13, the CycleGAN model trained using unpaired data not only outperforms conventional methods in terms of noise suppression, but also in terms of its ability to restore signal continuity.

In addition to enabling the denoising of seismic data in the unsupervised learning case through domain mapping, the researchers also introduced attribute information into the GAN model to achieve noise suppression of seismic data in the unsupervised learning case through attribute annotation, reducing the requirements on the dataset. Wang and Ma achieved suppression of seismic noise by using attribute

annotation on the data and constructing new loss constraints in the GAN model. Their experimental results show that using an attribute-based unsupervised learning GAN can achieve better results than traditional methods [28, 29].

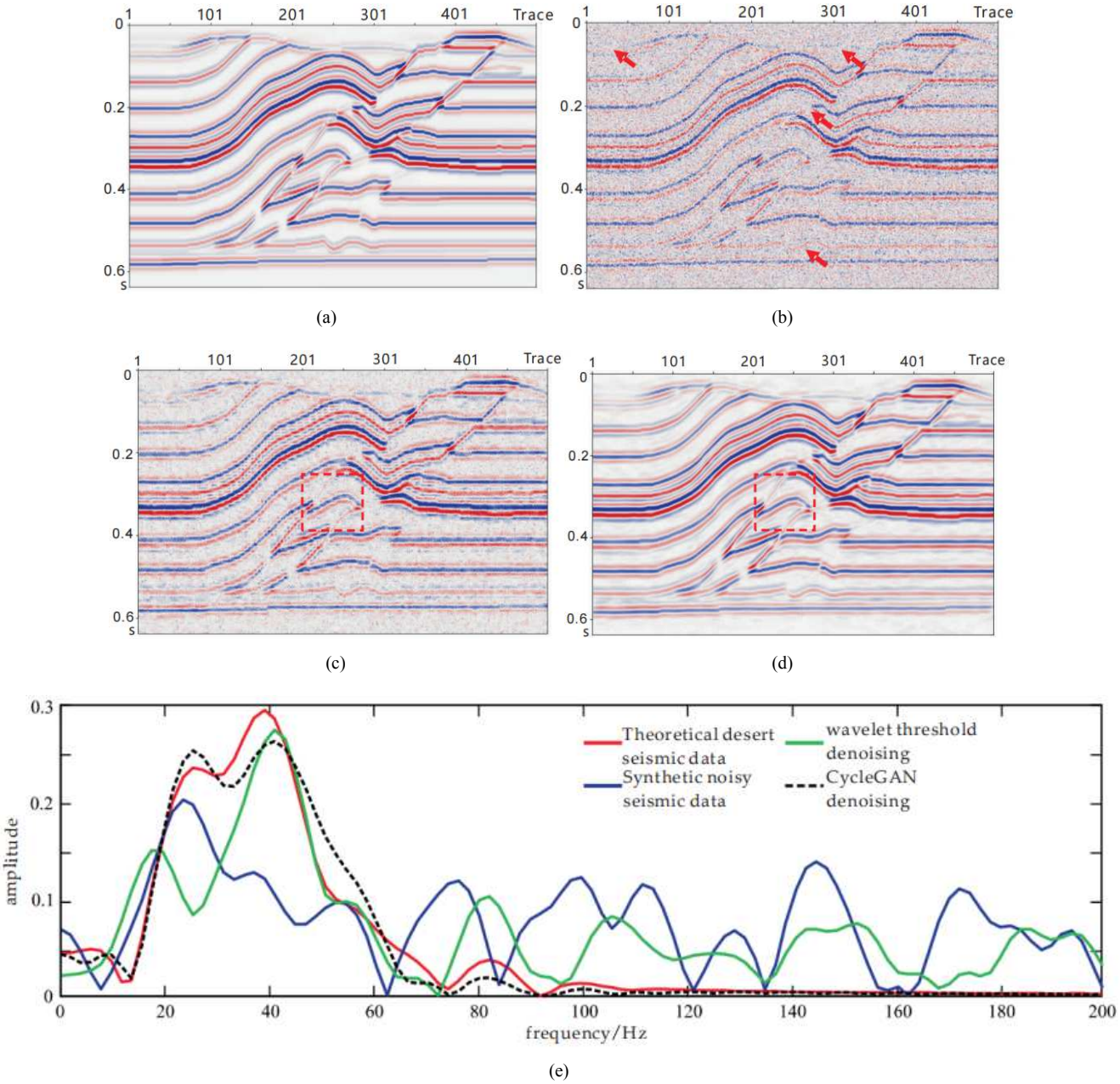


Figure 12. *Analysis of synthetic data denoising results. (a) Theoretical seismic data; (b) Synthetic noisy seismic data; (c) Denoising result of Wavelet Threshold; (d) Denoising result of CycleGAN; (e) comparison of spectrum before and after denoising of trace 240 of complex model seismic profile. (Resource: [58]).*

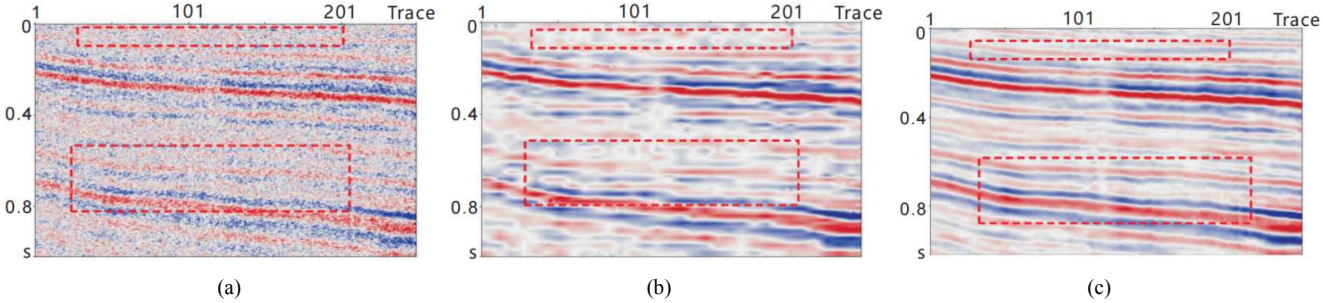


Figure 13. *Real seismic data denoising results. (a) Post-stack seismic profile;(b) Denoising result of Wavelet Threshold;(c) Denoising result of CycleGAN. (Resource: [58]).*

4. Conclusion

4.1. Summary

Traditional domain change filtering based seismic signal denoising techniques have yielded better results in suppressing conventional rules noise interference. However, more intelligent signal denoising techniques are needed to better protect the broadband data of seismic signals and reduce subjective human errors. This paper presents a detailed introduction to the theory of generative adversarial networks and their derived models, discusses the application of GAN in the field of seismic denoising with examples, analyses the feasibility and effectiveness of GAN for seismic denoising in terms of both supervised and unsupervised learning, and achieves the following understanding:

1) Most of the commonly used deep learning models for denoising tasks are CNN models based on supervised learning, which use a single data loss to construct a loss function between denoised data and pure labeled data, and reduce this loss through continuous optimization to achieve the purpose of denoising. This CNN model, which relies only on data loss, is not friendly to recovering seismic event continuity when used for seismic data denoising. Generating an adversarial network is able to guide the recovery of seismic events as a whole due to its adversarial nature. A model combining CNN and GAN not only inherits the denoising advantages of CNN but also the unique adversarial nature of GAN makes the denoising results with high signal to noise ratio and good event continuity.
2) Supervised learning-based GAN models have made some breakthroughs in restoring continuity of seismic events, but supervised learning relies on constructed paired datasets, which also limits its application scenarios. The unsupervised learning GAN model breaks through the limitation that model training must rely on paired data and opens up new ideas for seismic data denoising research. CycleGAN, a typical unsupervised learning model, has the core idea of mapping two different domains of data instead of one-to-one correspondence of data, reducing the requirement for data sets. In addition, the researchers also introduced attribute information into GAN training, reducing the requirement for paired data by setting attribute information to the data.

4.2. Prospect

As the seismic data acquisition environment and seismic data processing interpretation changes, there is an increasing demand for deep reflection seismic data, marine multicomponent seismic data, shallow desert strong attenuation seismic data and other complex tectonic weak signal seismic data processing. How to better suppress complex noise, protect the effective signal amplitude and signal band, and obtain high resolution data for seismic interpretation is an urgent need in the field of seismic data processing now and in the future. The evolving GAN technology applied to the field of seismic data denoising can improve the denoising quality in the following ways:

1) In terms of dataset construction, for data-driven deep learning techniques, it is extremely important to build complete datasets, and the quality of the training set directly affects the model training effect, which can be improved by introducing advanced exploration data modelling techniques. The current method of extending the seismic data for small samples is mostly scale transformation, increasing the training samples through scaling, inversion, deflection, etc. The generative adversarial network is originally a method for generating new data by mining the potential features of the data, so it can be considered to build a training dataset by generating data from the generative adversarial network to verify its feasibility;
2) In terms of algorithms, model performance and feature extraction can be improved by introducing cutting-edge techniques such as hyperparameter optimization, multi-scale feature extraction and migration learning into the GAN. In addition, advanced techniques in the field of image denoising can also be combined, and advanced methods for natural image denoising can be combined with GAN to design algorithms suitable for seismic profile processing;
3) In terms of model training, as the original GAN model suffers from gradient disappearance, training instability and pattern collapse, a stable GAN model suitable for denoising seismic data can be constructed by combining GAN optimization methods to stabilize the training of the GAN.

Author Contributions

Conceptualization, Y. X. and Q. W.; methodology, Y. X.; formal analysis, Y. X.; investigation, Y. X.; visualization, Z. L.; supervision, S. G.; writing—original draft preparation, Y. X. and Q. W.; writing—review and editing, Y. X. and Q. W.; All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by National Natural Science Foundation of China grant number: 41904117; General project of the scientific research plan of the Beijing Municipal Education Commission, number: KM202111232012.

Data Availability Statement

Not applicable.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Zhao, B.; Yong, X. Progress and development direction of Petro China intelligent seismic processing and interpretation technology. China Petroleum Exploration. 2021, 26, 12-23.

[2] Yang, W.; Wei, X. Development plan for intelligent geophysical prospecting technology of applied geophysical+AI. Oil Forum. 2019, 38, 40-47.

[3] Zhu, L.; Liu, E.; McClellan, J. H. Seismic Data Denoising through Multiscale and Sparsity-Promoting Dictionary Learning. GEOPHYSICS 2015, 80 (6), WD45–WD57.

[4] Zhang Junhua, Lv Ning, Tian Lianyu, et al. A comprehensive review of seismic data denoising methods and techniques. Advances in Geophysics, 2006, 2.

[5] Han, G.; Jie, Z. Simultaneous Denoising and Interpolation of Seismic Data via the Deep Learning Method. 33 (1), 15.

[6] Yang, L.; Chen, W.; Wang, H.; Chen, Y. Deep Learning Seismic Random Noise Attenuation via Improved Residual Convolutional Neural Network. IEEE Trans. Geosci. Remote Sensing 2021, 59 (9), 7968–7981.

[7] Chen Tian, Yi Yuanyuan. Random Noise Suppression for Seismic Data Based on Deep Convolutional Neural Networks. Chinese Journal of Seismology, 2021, 43 (4): 474-482.

[8] Yu, S.; Ma, J.; Wang, W. Deep Learning for Denoising. GEOPHYSICS 2019, 84 (6), V333–V350.

[9] Zhang Chaoming, Wen Xiaotao, Zhang Xiaoqi, Lan Yunlin, He Yilong. Seismic data denoising method based on DNCNN and constrained convolution. Advances in Geophysics, 2022, 1-18.

[10] Richardson, A.; Feller, C. Seismic Data Denoising and Deblending Using Deep Learning. arXiv July 2, 2019.

[11] Wang, E.; Nealon, J. Applying Machine Learning to 3D Seismic Image Denoising and Enhancement. Interpretation 2019, 7 (3), SE131–SE139.

[12] Zhang, F.; Liu, D.; Wang, X.; Chen, W.; Wang, W. Random Noise Attenuation Method for Seismic Data Based on Deep Residual Networks. In International Geophysical Conference, Beijing, China, 24-27 April 2018; Society of Exploration Geophysicists and Chinese Petroleum Society: Beijing, China, 2018; pp 1774–1777.

[13] Wang, F.; Chen, S. Residual Learning of Deep Convolutional Neural Network for Seismic Random Noise Attenuation. IEEE Geosci. Remote Sensing Lett. 2019, 16 (8), 1314–1318.

[14] Jin, Y.; Wu, X.; Chen, J.; Han, Z.; Hu, W. Seismic Data Denoising by Deep-Residual Networks. In SEG Technical Program Expanded Abstracts 2018; Society of Exploration Geophysicists: Anaheim, California, 2018; pp 4593–4597.

[15] Han, W.; Zhou, Y. Deep learning convolutional neural networks for random noise attenuation in seismic data. Geophysical Prospecting for Petroleum. 2018, 57, 862-869.

[16] Li, S.; Chen, W. Pre-stack Random Noise Deep Residual Network Suppression Method. Oil Geophysical Prospecting. 2020, 55, 493-503.

[17] Yang, L.; Chen, W.; Liu, W.; Zha, B.; Zhu, L. Random Noise Attenuation Based on Residual Convolutional Neural Network in Seismic Datasets. IEEE Access 2020, 8, 30271–30286.

[18] Wang, Y.; Lu, W. Seismic random noise suppression based on data augmentation and CNN [J]. CHINESE J GEOPHYS-CH, 2019, 62 (01): 421-433.

[19] Dong, X.; Zhong, T.; Li, Y. New Suppression Technology for Low-Frequency Noise in Desert Region: The Improved Robust Principal Component Analysis Based on Prediction of Neural Network. IEEE Trans. Geosci. Remote Sensing 2020, 58 (7), 4680–4690.

[20] Zhao, Y.; Li, Y.; Dong, X.; Yang, B. Low-Frequency Noise Suppression Method Based on Improved DNCNN in Desert Seismic Data. IEEE Geosci. Remote Sensing Lett. 2019, 16 (5), 811–815.

[21] Liu, D.; Wang, W.; Chen, W.; Wang, X.; Zhou, Y.; Shi, Z. Random-Noise Suppression in Seismic Data: What Can Deep Learning Do? In SEG Technical Program Expanded Abstracts 2018; Society of Exploration Geophysicists: Anaheim, California, 2018; pp 2016–2020.

[22] Liu, D.; Wang, W.; Wang, X.; Wang, C.; Pei, J.; Chen, W. Poststack Seismic Data Denoising Based on 3-D Convolutional Neural Network. IEEE Trans. Geosci. Remote Sensing 2020, 58 (3), 1598–1629.

[23] Zhang, M.; Liu, Y.; Chen, Y. Unsupervised Seismic Random Noise Attenuation Based on Deep Convolutional Neural Network. IEEE Access 2019, 7, 179810–179822.

[24] Wang, H.; Li, Y.; Dong, X. Generative Adversarial Network for Desert Seismic Data Denoising. IEEE Trans. Geosci. Remote Sensing 2021, 59 (8), 7062–7075.

[25] Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. arXiv June 10, 2014.

[26] Alwon, S. Generative Adversarial Networks in Seismic Data Processing. In SEG Technical Program Expanded Abstracts 2018; Society of Exploration Geophysicists: Anaheim, California, 2018; pp 1991–1995.

[27] Picetti, F.; Lipari, V.; Bestagini, P.; Tubaro, S. A Generative-Adversarial Network For Seismic-Imaging Applications. In SEG Technical Program Expanded Abstracts 2018; Society of Exploration Geophysicists: Anaheim, California, 2018; pp 2231–2235.

[28] Wang, S.; Li, Y.; Wu, N.; Zhao, Y.; Yao, H. Attribute-Based Double Constraint Denoising Network for Seismic Data. IEEE Trans. Geosci. Remote Sensing 2021, 59 (6), 5304–5316.

[29] Ma, H.; Sun, Y.; Wu, N.; Li, Y. Relative Attributes-Based Generative Adversarial Network for Desert Seismic Noise Suppression. IEEE Geosci. Remote Sensing Lett. 2022, 19, 1–5.

[30] Zheng, J.; Wu, Z. Generative adversarial network-based denoising method for micro seismic data. Research & Explorationin Laboratory, 2021, 40 (05): 18-21.

[31] Li, Y.; Luo, X.; Wu, N.; Dong, X. The Application of Semisupervised Attentional Generative Adversarial Networks in Desert Seismic Data Denoising. IEEE Geosci. Remote Sensing Lett. 2022, 19, 1–5.

[32] Yu, R.; Zhang, Y. Random noise removal from Rayleigh wave signals based on deep convolutional generative adversarial networks. Prog Geophys, 2020, 35 (06): 2276-2283.

[33] Liu, Y.; Wei, H. Random noise removal from seismic data based on convolutional recurrent generative adversarial networks. Journal of Jilin University (Information Science Edition), 2022, 1-9.

[34] Li, W.; Wang, J. Residual Learning of Cycle-GAN for Seismic Data Denoising. 2021, 9, 13.

[35] Ma, W.; Pan, Z.; Guo, J.; Lei, B. Achieving Super-Resolution Remote Sensing Images via the Wavelet Transform Combined With the Recursive Res-Net. IEEE Trans. Geosci. Remote Sensing 2019, 57 (6), 3512–3527.

[36] Rudin, L. I.; Osher, S.; Fatemi, E. Nonlinear Total Variation Based Noise Removal Algorithms. Physica D: Nonlinear Phenomena 1992, 60 (1–4), 259–268.

[37] Bonar, D.; Sacchi, M. Denoising Seismic Data Using the Nonlocal Means Algorithm. GEOPHYSICS 2012, 77 (1), A5–A8.

[38] Gulunay, N. FXDECON and Complex Wiener Prediction Filter. In SEG Technical Program Expanded Abstracts 1986; Society of Exploration Geophysicists, 1986; pp 279–281.

[39] Hennenfent, G.; Herrmann, F. J. Seismic Denoising with Nonuniformly Sampled Curvelets. Comput. Sci. Eng. 2006, 8 (3), 16–25.

[40] Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering. IEEE Trans. on Image Process. 2007, 16 (8), 2080–2095.

[41] Dong, W.; Shi, G.; Li, X. Nonlocal Image Restoration With Bilateral Variance Estimation: A Low-Rank Approach. IEEE Trans. on Image Process. 2013, 22 (2), 700–711.

[42] Mertens, J.-F.; Zamir, S. The Value of Two-Person Zero-Sum Repeated Games with Lack of Information on Both Sides. Int J Game Theory 1971, 1 (1), 39–64.

[43] Fuglede, B.; Topsoe, F. Jensen-Shannon Divergence and Hilbert Space Embedding. In International Symposium onInformation Theory, 2004. ISIT 2004. Proceedings.; IEEE: Chicago, Illinois, USA, 2004; pp 30–30.

[44] Raiber, F.; Kurland, O. Kullback-Leibler Divergence Revisited. In Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval; ACM: Amsterdam The Netherlands, 2017; pp 117–124.

[45] Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein GAN. arXiv December 6, 2017.

[46] Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A. Improved Training of Wasserstein GANs. arXiv December 25, 2017.

[47] Radford, A.; Metz, L.; Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. arXiv January 7, 2016.

[48] Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. IEEE Trans. on Image Process. 2017, 26 (7), 3142–3155.

[49] Mirza, M.; Osindero, S. Conditional Generative Adversarial Nets. arXiv November 6, 2014.

[50] Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A. A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In 2017 IEEE International Conference on Computer Vision (ICCV); IEEE: Venice, 2017; pp 2242–2251.

[51] Xia, Y.; He, D.; Qin, T.; Wang, L.; Yu, N.; Liu, T.-Y.; Ma, W.-Y. Dual Learning for Machine Translation. arXiv November 1, 2016.

[52] Si, X.; Yuan, Y.; Ping, F.; Zheng, Y.; Feng, L. Ground Roll Attenuation Based on Conditional and Cycle Generative Adversarial Networks. In SEG 2019 Workshop: Mathematical Geophysics: Traditional vs Learning, Beijing, China, 5–7 November 2019; Society of Exploration Geophysicists: Beijing, China, 2020; pp 95–98.

[53] Yuan, Y.; Si, X.; Zheng, Y. Ground-Roll Attenuation Using Generative Adversarial Networks. GEOPHYSICS 2020, 85 (4), WA255–WA267.

[54] Si, X. Ground Roll Attenuation with Conditional Generative Adversarial Networks. In SEG Technical Program Expanded Abstracts 2020; Society of Exploration Geophysicists: Virtual, 2020; pp 1511–1515.

[55] Dong, X.; Li, Y. Denoising the Optical Fiber Seismic Data by Using Convolutional Adversarial Network Based on Loss Balance. IEEE Trans. Geosci. Remote Sensing 2021, 59 (12), 10544–10554.

[56] Kaur, H.; Fomel, S.; Pham, N. Seismic Ground-roll Noise Attenuation Using Deep Learning. Geophysical Prospecting 2020, 68 (7), 2064–2077.

[57] Li, Y.; Wang, H.; Dong, X. The Denoising of Desert Seismic Data Based on Cycle-GAN With Unpaired Data Training. IEEE Geosci. Remote Sensing Lett. 2021, 18 (11), 2016–2020.

[58] Wu, X.; Zhang, H. Random noise suppression method for seismic data based on cyclic consistent Generative adversarial network. Oil Geophysical Prospecting. 2021, 56, 958-968.

[59] Zhu, W.; Mousavi, S. M.; Beroza, G. C. Seismic Signal Denoising and Decomposition Using Deep Neural Networks. IEEE Trans. Geosci. Remote Sensing 2019, 57 (11), 9476–9488.

[60] Pan, Z.; Yu, W.; Yi, X.; Khan, A.; Yuan, F.; Zheng, Y. Recent Progress on Generative Adversarial Networks (GANs): A Survey. IEEE Access 2019, 7, 36322–36333.

[61] Alnujaim, I.; Kim, Y. Augmentation of Doppler Radar Data Using Generative Adversarial Network for Human Motion Analysis. Healthc Inform Res 2019, 25 (4), 344.

[62] Leinonen, J.; Guillaume, A.; Yuan, T. Reconstruction of Cloud Vertical Structure With a Generative Adversarial Network. Geophysical Research Letters 2019, 46 (12), 7035–7044.

[63] Wolterink, J. M.; Leiner, T.; Viergever, M. A.; Isgum, I. Generative Adversarial Networks for Noise Reduction in Low-Dose CT. IEEE Trans. Med. Imaging 2017, 36 (12), 2536–2545.

[64] Sandfort, V.; Yan, K.; Pickhardt, P. J.; Summers, R. M. Data Augmentation Using Generative Adversarial Networks (CycleGAN) to Improve Generalizability in CT Segmentation Tasks. Sci Rep 2019, 9 (1), 16884.

[65] Mao, X.; Shen, C.; Yang, Y.-B. Image Restoration Using Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections. 9.

[66] Dumoulin, V.; Visin, F. A Guide to Convolution Arithmetic for Deep Learning. arXiv January 11, 2018.

[67] Eckle, K.; Schmidt-Hieber, J. A Comparison of Deep Networks with ReLU Activation Function and Linear Spline-Type Methods. Neural Networks 2019, 110, 232–242.

[68] Ma, H.; Yao, H.; Li, Y.; Wang, H. Deep Residual Encoder–Decoder Networks for Desert Seismic Noise Suppression. IEEE Geosci. Remote Sensing Lett. 2020, 17 (3), 529–533.

[69] Liu, H.; Zhang, J. A study on the impact of introducing feature loss on CycleGAN. Computer Engineering and Applications, 2020, 56 (22): 217-223.

[70] Li, C.; Liu, H.; Chen, C.; Pu, Y.; Chen, L.; Henao, R.; Carin, L. ALICE: Towards Understanding Adversarial Learning for Joint Distribution Matching. 9.