Seismic Signal Denoising Method Based on 1-D SE Cycle GAN

Xiang-jie Du

[1]*Student, Chengdu University of Information Technology*

ABSTRACT: Seismic signals contain rich geological information about subsurface rock layers, but this information is often obtained in the form of noisy results due to complex environmental interference factors. The quality of the signal is directly related to the accuracy and reliability of the characteristics of the subsurface medium and the structure of the strata. Traditional methods, due to factors such as their algorithm model and limitations of prior knowledge, cannot achieve ideal results in processing increasingly complex seismic signals. In recent years, with the development of neural network technology, seismic signal denoising methods based on neural network technology have become a research hotspot. CycleGAN does not require paired seismic data and can automatically learn the mapping relationship between noisy and clean signals. In this paper, based on the improvement of CycleGAN, the convolution layer is replaced with a one-dimensional convolution layer suitable for seismic sequences, and an attention mechanism based on a Squeeze-and-Excitation Networks (SENet) is introduced to obtain a one-dimensional CycleGAN based on Squeeze-and-Excitation (1-D SE CycleGAN), which enables the network to focus more on noise features when dealing with one-dimensional seismic data. The simulation experiments and actual data processing experiments show that 1-D SE Cycle GAN can better focus on the useful signals and noise existing in the high-frequency part of the original signal, capture the key features of the seismic signal, and achieve noise reduction effects, further improving the resolution of the seismic signal. Compared with traditional methods, the denoising method of 1-D SE Cycle GAN can remove more noise and obtain signals with higher signal-to-noise ratios in both synthetic and actual seismic signals. It can also retain more details of the original seismic signals, making the stratigraphy and faults clearer.

KEYWORDS: seismic signal, Cycle GAN, denoising.

I. INTRODUCTION

As seismic exploration advances and expands, more and more exploration areas become complex and diverse. The seismic signals collected under these harsh conditions usually contain various types of noise, such as natural noise, artificial noise, instrument noise, and signal processing noise, due to the complexity of the underground structure and limitations of the geophysical exploration equipment. Among them, signal processing noise refers to the noise introduced during the seismic signal processing. For example, interference during data acquisition and transmission, digital filtering, and other processing methods may introduce noise. Noise in seismic signals seriously interferes with the analysis and interpretation of seismic signals, directly affecting subsequent seismic signal processing steps such as high-resolution processing. Therefore, improving the signal-to-noise ratio and accuracy of seismic waveforms is the primary task in seismic signal processing.
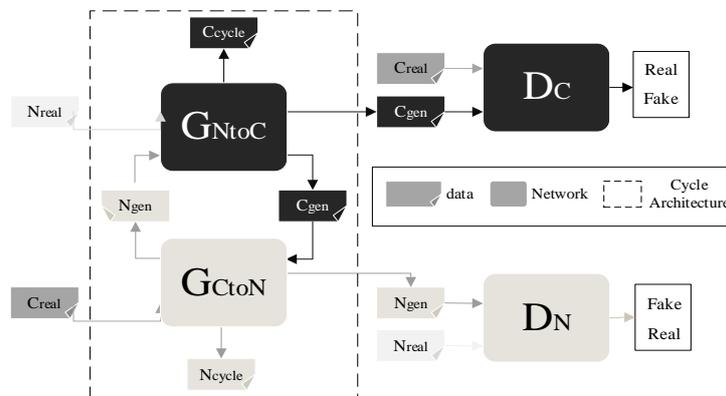
Fig. 1. The working model of Cycle GAN.

Noise in reality is often complex and diverse, and many traditional denoising methods are based on prior information such as the frequency or amplitude of the signal. Different models are established for different types of noise before denoising, including Gaussian white noise models, wavelet models, etc. [1-7]. Only by correct and appropriate modeling can we better grasp the statistical and time-frequency characteristics of noise. However, once the signal becomes too complex to model or the signal is below a high noise level, traditional methods may be limited or defective. Therefore, as seismic signals become increasingly complex and diverse, traditional research on seismic signal denoising methods has become increasingly challenging. In recent years, neural network-based seismic signal denoising methods do not require a large amount of prior knowledge, do not require manual design of complex feature extractors and filters, and can learn and represent the characteristics of complex seismic signals through deep network learning. They have strong non-linear modeling capabilities and can adapt to seismic signal denoising tasks under different acquisition conditions [8-12]. However, many deep learning techniques still lack in network learning ability and ability to retain seismic signal features, and require paired data.

Generative Adversarial Networks (GAN) is an unsupervised neural network model proposed by Ian Goodfellow et al. in 2014 [13]. The GAN model consists of two parts: the generator and the discriminator, which use different objective functions. The model is trained by adversarial learning to generate high-quality sample data. Cycle GAN is a variant of GAN, which is an unsupervised image translation model based on GAN [14]. Cycle GAN has the ability to map between two domains and perform conditional adversarial learning, which allows it to better preserve the features of seismic signals even when paired data is not available. The powerful image transformation capability of Cycle GAN enables the network to learn and recognize the characteristics of noise by training on a training set that contains a wide range of noise types, and then convert seismic signals between the noisy and noise-free domains. Building on Cycle GAN, we propose and investigate a 1-D Cycle GAN network model for one-dimensional seismic signals. We decompose low signal-to-noise ratio (SNR) two-dimensional seismic sections into single-trace seismic signals and process and transform each trace before reassembling them into high SNR two-dimensional seismic sections. As noise in seismic signals is often fine-grained and the signals themselves contain valuable information, we focus on preserving and restoring the seismic signal during the denoising process. Attention mechanisms can improve the filtering effect of the model, allowing the network to more accurately distinguish between noise and signal, thus better preserving the useful information in seismic signals and further improving the denoising accuracy and effectiveness. By adding a channel-wise attention mechanism using a Squeeze-and-Excitation Network (SENet) to the 1-D Cycle GAN [15], we obtain a compressed and excited 1-D Cycle GAN (1-D SE Cycle GAN) that achieves better denoising results than the original 1-D Cycle GAN.

II. METHOD

A. 1-D SE Cycle GAN denoising method

Fig. 1 shows the general framework of the proposed model. The generator GNtoC learns a mapping from N domain {Nreal} seismic images to C domains {Creal}:

$$C_{gen} = G_{NtoC}(N_{real}) \qquad (1)$$

Where GNtoC can be optimized by minimizing (1) in the iterative training of the network:

$$L(G_{NtoC}) = E[\log(D_C(G_{NtoC}(N \quad (2)$$

The term "E[*]" represents the expected value, and D_C is the discriminator that matches G_{CtoN}, whose goal is to distinguish whether the input image comes from the real cleandata domain$\{C_{real}\}$ or is generated by G_{NtoC}. $D(G_{NtoC}(N_{real}))$ represents the probability that the discriminator D_C judges the input $G_{NtoC}(N_{real})$ as coming from the real domain, where a value closer to 1 indicates that the discriminator is more likely to believe that $G_{NtoC}(N_{real})$ comes from the real domain. The optimization method of the discriminator is to minimize (3). The closer the result is to 1, the stronger the discrimination ability of D_C, and the more accurate the judgment result.

$$L(D_C) = E[\log(D_C(C_{real})) + \text{k} \quad (3)$$

In Cycle GAN, GCtoN is the mirror version of GNtoC, which aims to learn the mapping from clean data domain {Creal} to noisy data domain {Nreal}, generating the same image Ngen as the domain {Nreal}

$$N_{gen} = G_{CtoN}(C_{real}) \quad (4)$$

Similarly,D_N is the discriminator paired with G_{CtoN}, whose goal is to distinguish as much as possible whether the input image comes from the real noisy data domain $\{N_{real}\}$ or is generated by G_{CtoN}.

The first stage of Cycle GAN is to train two sets of generators and discriminators that are each other's mirror images. In theory, adversarial training enables GCtoN and GNtoC to learn the transformation of each other's domain. However, when the sample size is large enough, the two generators can map the same input image set to any random arrangement of images in the target domain, and any learned mapping can induce an output distribution that matches the target distribution. Therefore, to compress the target space of the mapping function, Cycle GAN adds a cyclic loss objective function in the second stage to ensure cycle consistency:

$$L(G_{cycle}) = E[\|G_{HtoL}(G_{LtoH}(LR_{real})) -$$

$$+ E[\|G_{HtoL}(G_{LtoH}(LR_{real})) - LR_{real}\|_1] \quad (5)$$

Here, $\|*\|_1$ represents the L1 norm. A smaller result from (5) indicates that the images generated by the

two generators are closer to the original images.

Similarity (SSIM) is a widely used metric for image quality evaluation that considers the impact of image brightness, contrast, and structure information, based on the sensitivity of human eyes to structural information [16]. In high-resolution processing of seismic signals, we use SSIM as an evaluation metric to assess the similarity and signal-to-noise ratio improvement of seismic signals before and after processing, thereby determining the effectiveness and superiority of the algorithm. The simplified module coefficient SSIM algorithm is:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C)(2\sigma_x\sigma}{(\mu_x^2 + \mu_y^2 + C)(\sigma_x^2 + \text{(}} \quad (6)$$

Here, μ represents the brightness of the image, which indicates the average amplitude of the signal in a single seismic trace; σ represents the contrast of the image, which indicates the smoothness of the signal in a single seismic trace. To ensure the stability of the formula, a constant C is added. The SSIM value ranges from 0 to 1, where 1 indicates that two images are identical. When the SSIM value is close to 1, it means that the quality of the two images is very similar, and it is difficult for the human eye to distinguish their differences; while when the SSIM value is close to 0, it means that there are significant differences in the quality of the two images. The SSIM loss does not have the generative and adversarial characteristics, therefore, for G_{NtoC}, the generated signal used for comparison with the real signal C_{real} is not $G_{NtoC}(N_{real})$, but rather $G_{NtoC}(C_{real})$:

$$S_{NtoC} = SSIM(G_{NtoC}(C_{real}), \quad (7)$$

As a result, the corresponding SSIM objective functions for the two generators are:

$$L(G_{SSIM-NtoC}) = E(1 - S_{NtoC}) \quad (8)$$

and:

$$L(G_{SSIM-CtoN}) = E(1 - S_{CtoN}) \quad (9)$$

Therefore, we can represent the final objective function of Cycle GAN as follows:

$$L(G, D) = L(G_{NtoC}) + L(G_{CtoN}) + L(D_C)$$

$$+ \beta_1(L(G_{SSIM-NtoC}) + \beta_2 L(G_{SSIM-CtoN} \quad (10)$$

whereαand β_1, β_2 are adjustable coefficients. After experimental verification and balancing with other objective functions, we obtained the optimal parameters of 5 and 10, 10 respectively.

B.   Network Architecture
        Inspired by Goodfellow, Zhu, et al, our 1-D Cycle GAN belongs to the overall category of deep learning networks. After experimental testing, the optimal structure was obtained and is shown in Fig. 2.
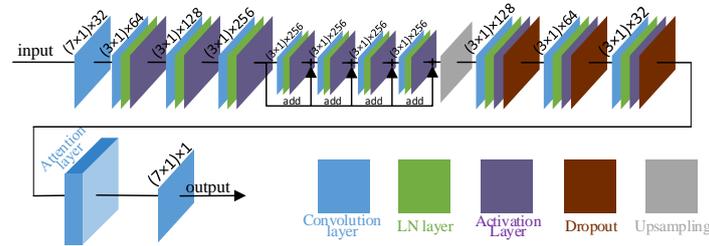
Fig. 2. The main network of the generator
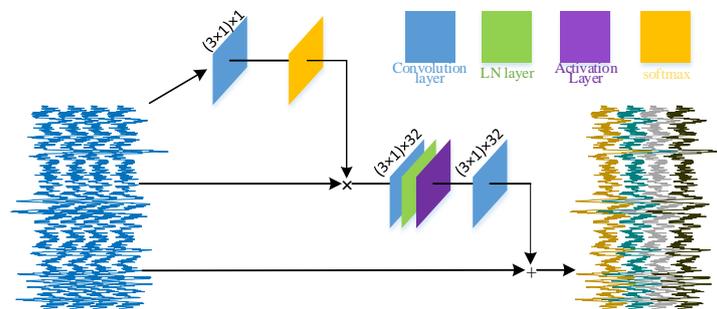
Fig. 3. The structure of the attention layer.

        Fig. 2 shows the main network of the generator, including the encoder, the residual network, and the decoder. The encoder consists of the first four layers, which include a convolutional layer and three convolutional+activation+LN layers. In the convolutional layer, "$(7\times1)\times32$" represents 32 one-dimensional convolution kernels of size 7×1 with a stride of 1. They perform convolutions along the time/depth dimension of a seismic signal, producing 32 one-dimensional feature vectors. Leaky ReLU is used instead of ReLU for the activation layer, which has better training speed and effectiveness [17]. The generator uses LN for normalization, mainly because in seismic sequence data, the time dimension is usually a dimension of the sample data, and the differences between samples are often larger on this dimension than on other dimensions, resulting in traditional normalization being less effective on time-series data than on other types of data, such as images. During the training process, LN can keep the average output of the activation function around 1, preventing gradient vanishing or exploding and improving the stability and generalization performance of the model [18]. The residual network consists of four residual blocks and passes the input directly to the output through skip connections, which can effectively alleviate the problem of gradient vanishing during the training of deep neural networks and improve the training speed and accuracy of the model [19]. The residual network is connected to the decoder via an upsampling layer. The upsampling layer uses linear interpolation to calculate the interpolation by averaging the distance between two known data points to the interpolation point. The upsampling layer doubles the length of the output feature vector to keep the generator output the same length as the original input.

The first three convolutional layers of the encoder have a stride of 1, while the fourth layer has a stride of 2, which reduces the length of the output feature vector to half the length of the input vector. This reduces the amount of data input to the residual network and improves the learning speed of the network. The first three layers of the decoder are the same as those of the encoder, except for the number of convolutional filters, and a Dropout layer is added after each layer. Adding Dropout layers in the encoder may destroy some important features of the input data, leading to a decrease in model performance. Meanwhile, in the decoder, some features used to generate samples may rely too much on certain features of the encoder, which may not correspond to key information in the real dataset. Therefore, during decoder training, Dropout randomly drops some neurons, preventing the decoder from relying too much on certain features of the encoder and improving the model's generalization ability [20].

The output of the 12th layer of the generator is 32 feature vectors, which are fed into the attention layer. The feature vectors are first compressed into a single weight vector through a 3×1×1 convolution layer and a softmax layer. The weight vector is then multiplied with the 32 feature vectors from the input to perform attention weighting, resulting in a weighted feature vector group. The feature vector group then goes through two layers of 3×1×32 convolution to obtain 32 feature vectors with assigned weights, which are added back to the original input to recalibrate the features of each channel. The structure of the attention layer is shown in Fig. 3.

The main network of the discriminator in Fig. 4 consists of five layers, similar to the encoder of the generator. Each layer has a convolutional stride of 2, and the last layer has a stride of 1 with a 1x1 convolutional kernel. The output of the final layer is a feature vector containing 512 neurons, which is then connected to a sigmoid function to calculate a probability value.

II.   EXPERIMENT

A.   Training details

To train a neural network with denoising capability, it is necessary to create datasets with both noisy and noise-free data. The noisy dataset needs to include multiple types of noise to ensure that the trained neural network has sufficient generalization ability and robustness. Our goal is to obtain high signal-to-noise ratio and visually clean seismic signals from noisy seismic signals with different noise levels and types. Fig. 4 is an example of creating a seismic dataset.
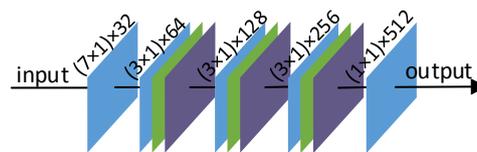
Fig. 4. Structure of discriminator
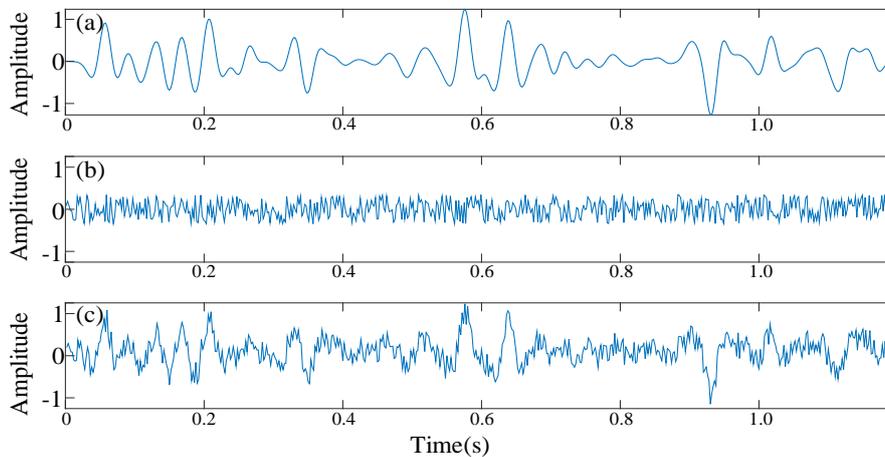
Fig. 5. Example of synthetic seismic signals with and without noise. (a. seismic signal generated by convolving synthetic reflection coefficients with a Ricker wavelet. b. Random Gaussian noise c. Seismic signal with random Gaussian noise added.)

Fig. 5(a) shows an example of synthetic seismic signals with and without noise. Adding noise to a clean signal, it can be seen that the areas of the original signal with smaller amplitudes have been severely distorted by noise and lost their original waveform. We synthesized 30,000 reflection coefficient sequences, each with 600 sampling points. Then, we randomly extracted 30,000 Ricker wavelets with frequencies between 10-80 Hz from seismic signals, and convolved each of them with every reflection coefficient sequence to obtain 30,000 one-dimensional seismic signals without noise, which were used as the noise-free training set. Next, we divided these 30,000 noise-free signals into three groups, each with 10,000 signals, and added Gaussian white noise randomly selected from three different intervals of 0-5dB, 5-10dB, and 10-15dB to them, resulting in 30,000

noisy signals with different noise levels, which were used as the noisy training set. These signals were

paired and fed into the network for training.
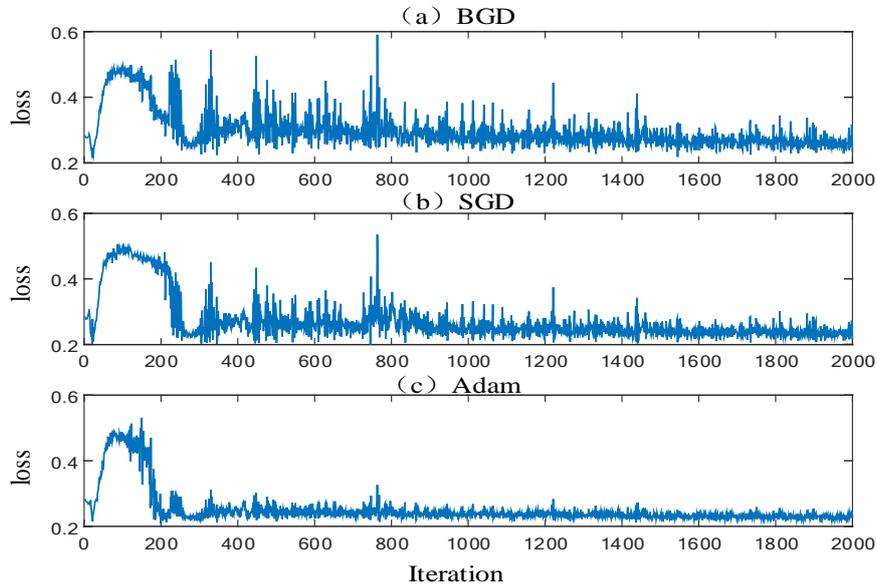
Fig. 6. Convergence effects of different gradient descent methods. (a. BGD method. b. SGD method c. Adam method.)

After each mini-batch training, we randomly flip some discriminator losses with a flipping factor of 0.3, which means the complements of three out of ten discriminator results are set to 1. The training time and effectiveness of BGD, SGD [21], and Adam [22] backpropagation algorithms were tested separately. Fig. 6 shows the convergence effect of the generator G_{NtoC} during the training process from 0 to 2000 for different gradient descent methods.

It can be seen that when using the Adam method, the loss function of BGD and SGD methods did not converge between 180 and 800 iterations, and the loss function tended to increase.

This means that the network's performance was unstable during this period, and stopping the training would result in very poor seismic signals generated by the G_{NtoC} generator. After 800 iterations, all three methods tended to converge, but it can be seen from the comparison that the loss function of the Adam method is more stable and tends to a lower value. Table 1 shows the time required for training 2000 iterations under the three methods. The SGD method is the fastest, and its time is set to 1, while the time of the other two methods is normalized. A comprehensive analysis shows that Adam achieves the best training effect in a relatively short time.

TABLE I. TRAINING TIME (NORMALIZED)

Method	BGD	SGD	Adam
Time cost	2.726	1	1.422

B.   Synthetic data experiment and analysis
We randomly synthesized a reflection sequence and convolved it with a 40Hz Ricker wavelet to obtain a synthesized seismic signal. Then, we extended it to a 2D seismic section by 50 traces. Adding 5dB

random Gaussian noise to the section resulted in a noisy signal. We then denoised the signal using two methods and showed the results in Fig. 7.
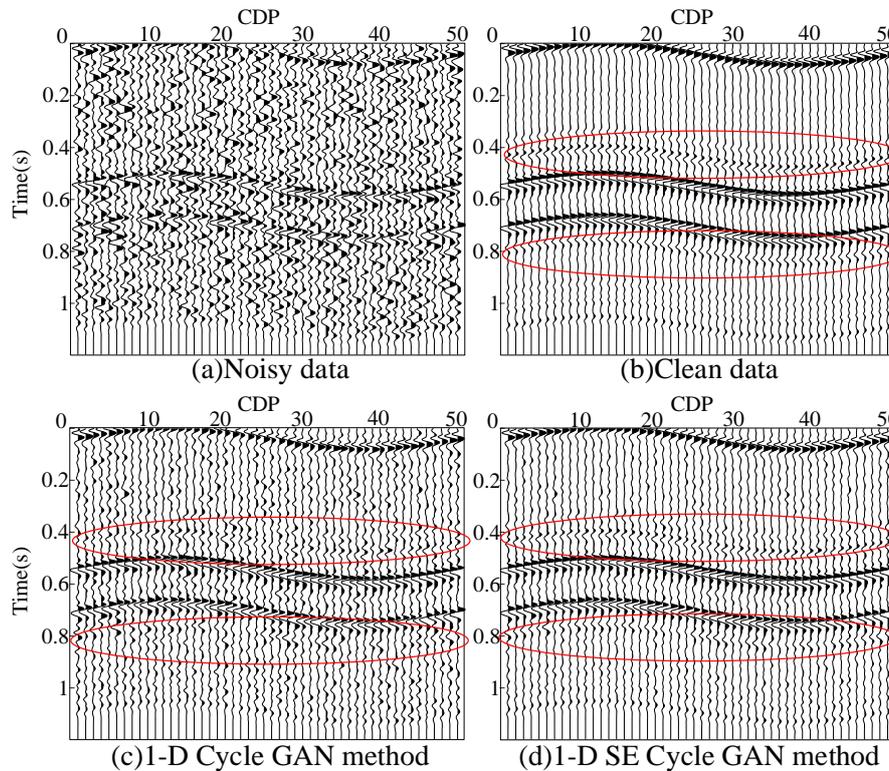
Fig. 7. The denoising performance of 1-D Cycle GAN and 1-D SE Cycle GAN at 5db noise level. (a. Noisy data
b. Clean data c. Denoising performance of 1-D Cycle GAN d. Denoising performance of 1-D SE Cycle GAN)

Fig. 7(a) shows the synthetic seismic signal with 5 dB noise added, while Fig. 7(b) shows the noise-free synthetic seismic signal. Fig. 7(c) and(d) display the denoising results of the conventional 1-D Cycle GAN method and the 1-D SE Cycle GAN-based method, respectively. It can be seen that the 1-D SE Cycle GAN-based method achieves better denoising performance. In addition, there is useful signal present in the red boxed area of the noise-free synthetic seismic signal, but the conventional 1-D Cycle GAN method fails to recover this useful part and leaves it contaminated by noise, leading to data loss. However, the 1-D SE Cycle GAN-based method preserves the information of this part and effectively removes the

noise. The color residual sections of the denoising results are also shown in Fig. 8. It shows the residual sections after denoising with the above two methods. The useful information in the original signal can be seen to be removed by the regular 1-D Cycle GAN method in the area indicated by the red arrow, which may lead to significant loss of useful geological information and greatly affect the denoising effect. However, in Fig. 8(c), the residual of the 1-D SE Cycle GAN-based method is very close to that of the noise-free section in Fig. 8(a) and the residual of the noisy section, which means that the signal after denoising with the 1-D SE Cycle GAN-based method is very close to the original noise-free signal.
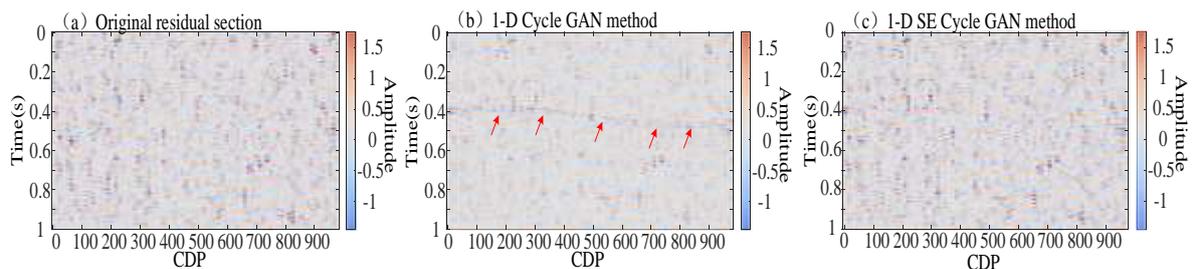
Fig. 8. Residual sections after denoising with two methods. (a. original residual section b. Residual section of 1-D Cycle GAN method c. Residual section of 1-D SE Cycle GAN method.)

To further validate the advantages of the 1-D SE Cycle GAN-based method, we synthesized noisy signals at different levels and compared the results of the traditional deconvolution method and the wavelet soft-thresholding denoising method with the 1-D SE Cycle GAN-based method. The results are shown in the following images.
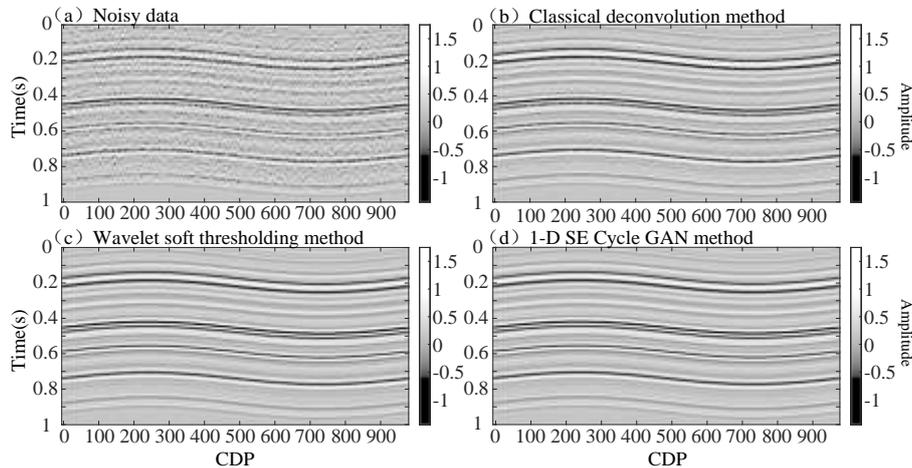
Fig. 9. The result of denoising experiment with noise level of 10dB (a. Noisy signal b. Result of denoising using traditional deconvolution method c. Result of denoising using wavelet soft thresholding method d. Result of denoising using 1-D SE Cycle GAN method)
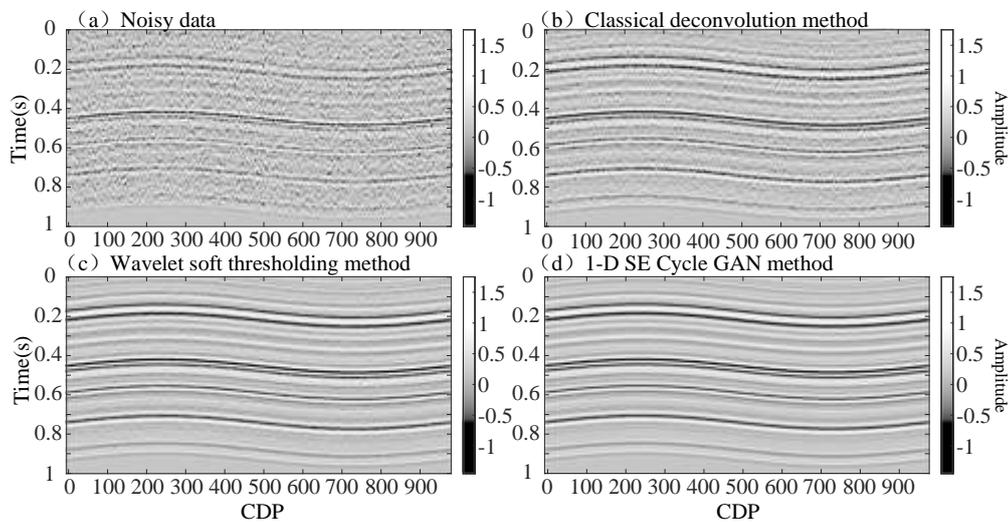
Fig. 10. The result of denoising experiment with noise level of 5dB (a. Noisy signal b. Result of denoising using traditional deconvolution method c. Result of denoising using wavelet soft thresholding method d. Result of denoising using 1-D SE Cycle GAN method)
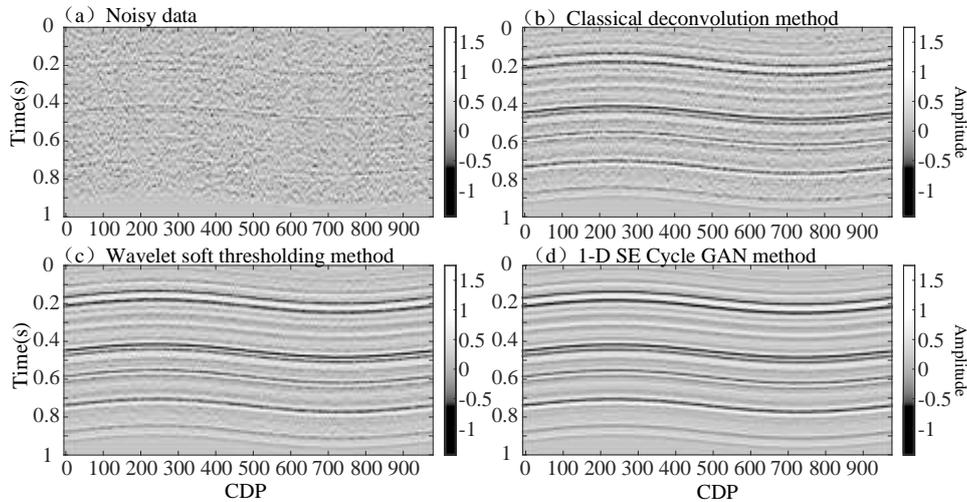
Fig. 11. The result of denoising experiment with noise level of 1dB (a. Noisy signal b. Result of denoising using traditional deconvolution method c. Result of denoising using wavelet soft thresholding method d. Result of denoising using 1-D SE Cycle GAN method)

In Fig. 9, at a noise level of 10 dB, the original section was less contaminated by noise, and the denoising effects of traditional deconvolution method, wavelet soft-thresholding denoising method, and the method based on 1-D SE Cycle GAN were all good. However, the section processed by the traditional deconvolution method still contained subtle noise, which was difficult to separate and filter out due to the spectrum being very close to that of the original signal. Comparing the results, it can be seen that the denoising results of the wavelet soft-thresholding method and the method based on 1-D SE Cycle GAN are both better than that of the traditional deconvolution method. In Fig. 10, at a noise level of 5dB, the result of the traditional deconvolution method still had obvious noise pollution, and the section after denoising by the wavelet soft thresholding method also had slight noise. This is because when the noise level is high, the energy of the noise signal is closer to the energy of the signal, making it difficult to distinguish between noise and signal coefficients in certain cases in the wavelet domain, making it difficult to accurately determine which coefficients are noise and which are signal coefficients. However, the 1-D SE Cycle GAN-based method still maintained its advantage. In Fig. 11, at a noise level of 1dB, the original section was severely contaminated by high-level noise, and most of the geological information was no longer visible. The results of the traditional deconvolution method and the wavelet soft thresholding method had obvious noise, while the denoising effect of the 1-D SE Cycle GAN-based method was obvious, and there was almost no noise affecting the visual effect of the section. We calculated the Signal-to-Noise Ratio (SNR) of the above denoising experiment results and presented them in Table 2.

C. Processing of actual seismic data.

We conducted experiments on actual seismic data from the Yuanba area to verify the performance of the 1-D SE Cycle GAN-based method in high-resolution processing of seismic signals. The seismic signals in this area are affected by various types of noise, and there are also multiple types of noise in the signals themselves, which seriously affect the quality and reliability of seismic signals and increase the difficulty of seismic data processing. We randomly selected 370 continuous seismic signals from the data in the Yuanba area to form a two-dimensional seismic section, with each signal having a sampling time of 300ms. We used traditional deconvolution, wavelet soft-threshold denoising, and 1-D SE

TABLE II. SNR AFTER DENOISING WITH THREE METHODS AT DIFFERENT NOISE LEVELS

Method		SNR1(dB)	SNR2(dB)	SNR3(dB)
noise level		1	5	10
deconvolution		9.75	11.46	12.04
Wavelet soft-thresholding		11.85	12.61	14.31

1-D SE Cycle GAN	13.19	14.28	14.65

Cycle GAN-based methods to process these 370 signals, and the results are shown in Fig. 12. The following 740 continuous seismic signals were randomly selected from different locations in the same region to form a two-dimensional seismic section, with each signal having a sampling time of 600ms. The traditional deconvolution method, the wavelet soft threshold denoising method, and the 1-D SE Cycle GAN-based method were used to process these 740 signals, and the results are shown in Fig. 13.
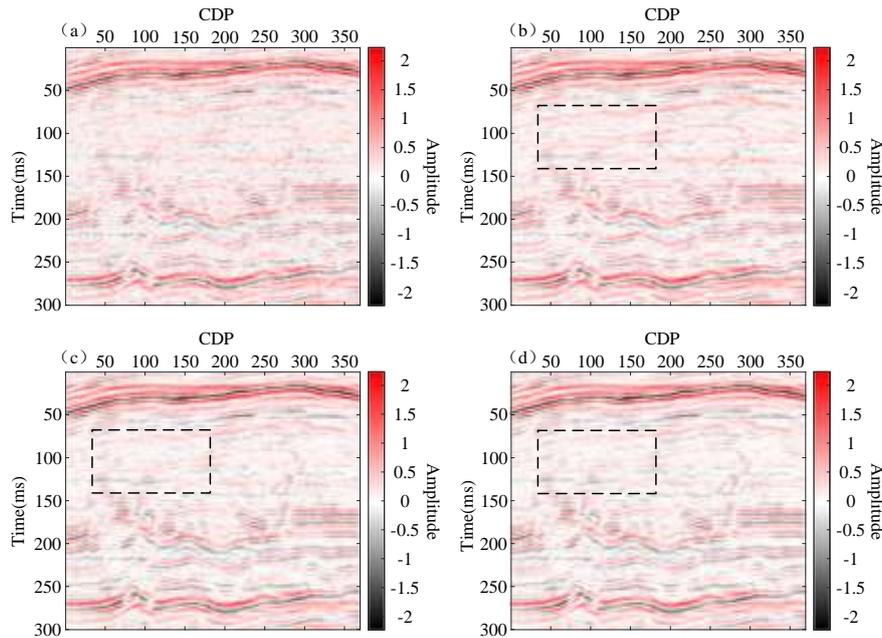
Fig. 12. Different denoising results of various methods on real seismic section (a. Original section b. Traditional deconvolution method c. Wavelet soft thresholding method d. 1-D SE Cycle GAN-based method)
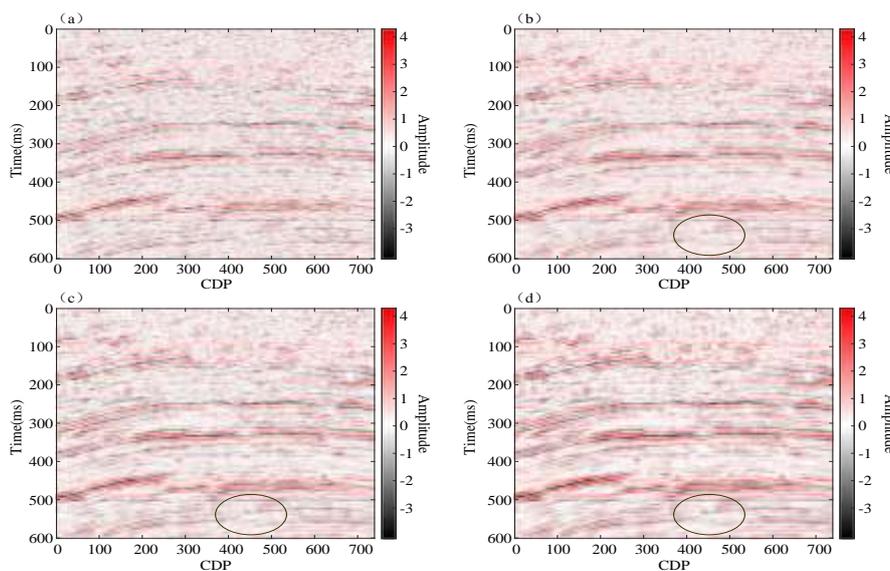
Fig. 13. Different denoising results of various methods on real seismic section (a. Original section b. Traditional deconvolution method c. Wavelet soft thresholding method d. 1-D SE Cycle GAN-based method)

It can be seen that the original seismic section is contaminated to a certain extent by noise, with most of the noise components concentrated in the middle and high amplitude parts. The traditional deconvolution method, wavelet soft thresholding method, and 1-D SE Cycle GAN-based method all have a certain degree of denoising effect. However, compared with the wavelet soft thresholding method and the 1-D SE Cycle GAN-based method, the control effect of the traditional deconvolution method on small noise is not ideal, and some details in the original section are still covered by noise. Comparing Fig. 12(c), 12(d), 13(c), and 13(d), it can be seen that the useful information contained in the original noisy section is shown more clearly and completely in Fig. 12(d) and 13(d), especially the part marked by the black dashed box. The 1-D SE Cycle GAN-based method maximally preserves the useful signals of the original section while removing most of the noise. Therefore, the 1-D SE Cycle GAN-based method also has significant advantages in denoising processing of actual seismic signals, preserving most of the real signal details in the actual signal while removing most of the noise, and providing better guidance for the determination of strata and faults.

III.   CONCLUSION

The noise contained in seismic signals is often random noise, distributed in both the low-frequency and high-frequency parts of seismic signals. By adding a self-attention mechanism based on the compression and excitation network to the 1-D Cycle GAN, the 1-D SE Cycle GAN is obtained, which has the property of weighting and combining the input signals, thus more effectively removing noise and retaining useful information of the signal. According to the results of the residual section experiments, the 1-D Cycle GAN removes useful information of the signal as noise during the denoising process, while the method based on 1-D SE Cycle GAN removes more noise and retains useful information of the signal. Then, by comparing with traditional deconvolution methods and wavelet soft threshold denoising methods, conclusions can be drawn through synthetic experiments: under low noise levels, the method based on 1-D SE Cycle GAN and the wavelet soft threshold denoising method have similar effects and can effectively improve the signal-to-noise ratio. However, as the noise level increases, the wavelet soft threshold denoising method becomes less effective due to the difficulty of distinguishing high-frequency amplitude noise. The method based on 1-D SE Cycle GAN can learn the characteristics of

this noise and effectively remove it, maintaining a higher signal-to-noise ratio even under high noise levels. When comparing the actual data, the denoising method based on 1-D SE Cycle GAN removes more noise, retains more details of the original seismic signals, and makes the geological layers and faults clearer.

REFERENCES

[1] Gabor, D. Theory of communication. Part 1: The analysis of information. Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering, 1946, 93(26), 429-441.

[2] Liner, C. L., & Loupas, M. L. Seismic data processing with the wiener filter. Geophysics, 1977, 42(5), 1048-1053.

[3] Huang, N. E., Shen, Z., Long, S. R., et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 1998, 454(1971), 903-995.

[4] Wu, Y., Xia, H., & Zhang, P.Research on the denoising of seismic signal based on EEMD and thresholding method. Measurement, 2021, 179, 109412.

[5] Xu, X., Zhang, X., & Gao, J.Iterative time-frequency peak filtering for random noise attenuation in seismic data. Journal of Applied Geophysics, 2021, 186, 104348.

[6] Zhang, J., Yang, S., Chen, Q., & Wu, R. S.An adaptive matching pursuit algorithm for seismic data denoising. Journal of Applied Geophysics, 2019, 170, 103803.

[7] Liu, Z., Chen, G., & Chen, Q.A robust and effective sparse decomposition method for seismic data denoising. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(7), 6111-6121.

[8] Wu, X., Xu, Z., Cai, H., & Cai, Y. Seismic data denoising using deep convolutional neural networks. Journal of Applied Geophysics, 2017, 135, 38-49.

[9] Wang, D., Wang, H., Jiang, L., et al. H. A GAN-based seismic data denoising method. IEEE Geoscience and Remote Sensing Letters, 2018, 15(4), 562-566.

[10] Zhu, H., Xie, R., Chen, Y., & Wang, J.Seismic signal denoising via deep

convolutional neural network with 2D transfer learning. Journal of Geophysics and Engineering, 2020, 17(3), 560-572.

[11] Wang, D., Wang, H., Jiang, L., et al. H. A GAN-based seismic data denoising method. IEEE Geoscience and Remote Sensing Letters, 2018, 15(4), 562-566.

[12] Zhang, Y., Chen, J., Chen, W., & Luo, Y. Multi-level residual network for seismic data denoising. Journal of Applied Geophysics, 2020, 174, 103977.

[13] Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. Generative adversarial nets. In Advances in neural information processing systems, 2014. (pp. 2672-2680).

[14] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision, 2017, (pp. 2223-2232).

[15] Vaswani, A., Shazeer, N., Parmar, N. Attention is all you need. In Advances in neural information processing systems, 2017, (pp. 5998-6008).

[16] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 2004, 13(4), 600-612.

[17] Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2015). Rectifier nonlinearities improve neural network acoustic models. In Proceedings of the 30th international conference on machine learning (ICML-13) (pp. 3-11).

[18] Ba, J. L., Kiros, J. R., & Hinton, G. E. Layer normalization. arXiv preprint 2016. arXiv:1607.06450.

[19] He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, (pp. 770-778).

[20] Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint, 2012, arXiv:1207.0580.

[21] Robbins, H., & Monro, S. A stochastic approximation method. Annals of Mathematical Statistics, 1951, 22(3), 400-407.

[22] Kingma, D. P., & Ba, J. Adam: A method for stochastic optimization. arXiv preprint, 2014, arXiv:1412.6980.