



Article

Structure-Preserving Random Noise Attenuation Method for Seismic Data Based on a Flexible Attention CNN

Wenda Li ^{1,2} , Tianqi Wu ^{1,*} and Hong Liu ^{1,2}¹ Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100864, China² National Engineering Laboratory for Offshore Oil Exploration, University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: tianqi_wu@cugb.edu.cn; Tel.: +86-010-82998357

Abstract: The noise attenuation of seismic data is an indispensable part of seismic data processing, directly impacting the following inversion and imaging. This paper focuses on two bottlenecks in the AI-based denoising method of seismic data: the destruction of structural information of seismic data and the inferior generalizability. We propose a flexible attention-CNN (FACNN) and realized the denoising work of seismic data. This paper's main work and advantages were concentrated on the following three aspects: (i) We propose attention gates (AGs), which progressively suppressed features in irrelevant background parts and improved the denoising performance. (ii) We added a noise level map M as an additional channel, making a single CNN model expected to inherit the flexibility of handling noise models with different parameters, even spatially variant noises. (iii) We propose a mixed loss function based on MS_SSIM to improve the performance of FACNN further. Adding the noise level map can improve the network's generalization ability, and adding the attention structure with the mixed loss function can better protect the structural information of the seismic data. The numerical tests showed that our method has better generalization and can better protect the details of seismic events.

**Citation:** Li, W.; Wu, T.; Liu, H.Structure-Preserving Random Noise Attenuation Method for Seismic Data Based on a Flexible Attention CNN. *Remote Sens.* **2022**, *14*, 5240.<https://doi.org/10.3390/rs14205240>

Academic Editors: Jingrui Luo, Benfeng Wang and Ru-Shan Wu

Received: 29 August 2022

Accepted: 17 October 2022

Published: 20 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: random noise; attention; noise level; structure-preserving; denoising

1. Introduction

With the deepening of oil exploration, exploration areas are becoming more and more diversified, such as deserts, mountains, plains, oceans, etc. Different exploration areas have different characteristics of seismic data, which raises new challenges for data processing in the oil exploration industry. For example, in desert areas, as influenced by the surface conditions and acquisition environment, the characteristics of seismic records are mainly characterized by the low signal-to-noise ratio (SNR) and severe spectrum aliasing of noise and effective signals [1]. How to enhance the SNR of seismic data is an urgent problem.

Traditional random noise suppression methods can be divided into the following categories. Firstly, filtering techniques based methods, which are based on the difference between the effective signal and the noise in the frequency spectrum, mainly include an f - x deconvolution method (FXDM) [2], the time-frequency peak filtering [3–5], the bandpass filtering [6], and so on. The second kind of algorithm is based on inter-tract correlation, mainly including singular-value decomposition [7], K-L transform [8], etc. At the same time, there is a series of transformation algorithms, such as shearlet transformation [9], wavelet transformation [10], seislet transformation [11], and curvelet transformation [12,13]. The fourth is a reduced-rank class algorithm [14], which takes advantage of the feature that the seismic signal is of low rank. However, the noise signal increases the rank of the seismic signal, mainly including Cadzow filtering [15,16] and principal component analysis.

However, conventional denoising algorithms still have two unbreakable bottlenecks [17], inaccurate assumptions and cumbersome parameter adjustment with manual intervention, which are unfavorable for massive seismic data processing. In recent years, due to the rapid development of computer technology, deep learning has become a popular research

topic. Lecun et al. [18] proved that CNNs with fewer parameters provided superior classification results on the MNIST. Zhang et al. [19] proposed a CNN with 17 layers named DnCNN to realize the image's noise attenuation. Zhang et al. [20] proposed a fast and flexible denoising CNN to improve the denoising ability further. Yu et al. [21] proposed a vision-based crack diagnosis method using a deep convolutional neural network (DCNN) and an enhanced chicken swarm algorithm. Yu et al. [22] proposed a novel method based on deep convolutional neural networks to identify and localize the damages of building structures equipped with smart control devices.

The deep learning methods can extract the secret relationship between noise and clean data and realize the intelligent denoising tasks without unnecessary assumptions and excessive manual intervention. There have been awe-inspiring research advances from previous researchers for AI-based denoising work. Yu and Ma [17] introduced DnCNN in image denoising into seismic data denoising and achieved good results. Based on this, Wang et al. [23] proposed a data augmentation algorithm, which can improve the diversity of the training dataset. Zhang and Liu [24] proposed a novel approach to attenuate seismic random noise based on a deep convolutional neural network (CNN) in an unsupervised learning manner. Dong and Li [25] proposed an adaptive DnCNN based on the determination of high-order statistics and realized the desert seismic data noise attenuation.

Many scholars have further applied the attention mechanism to various computer vision tasks such as noise removal [26] and image segmentation [27]. Since the attention mechanism allows the neural network to focus more on the higher-weighted parts, it can significantly preserve the structural characteristics of the seismic data [28,29].

Although many AI-based denoising methods have obtained favorable results, they still face two intractable problems: the destruction of the detailed seismic data structure and insufficient generalization. Firstly, different AI-based denoising algorithms will destroy the amplitude and phase information of the seismic signal to some extent, specifically the residual part of the useful signal information in the removed noise profile. Secondly, when using the well-trained network model for testing, if the noise characteristics of the testing data are inconsistent with the training, the network cannot reach the best performance.

This paper focuses on two intractable problems: the destruction of the detailed seismic data structure and insufficient generalization. We propose a flexible attention-CNN (FACNN) and realized the denoising work of seismic data. In contrast to the existing AI-based denoisers, FACNN enjoys several desirable properties. We first added a noise level map as an additional channel in the network input data body, making a single CNN model that can handle the noise models with different noise characteristics. Secondly, we integrated the attention gates (AGs) in a standard U-Net model, which progressively suppressed features in irrelevant background parts and improved the preservation of seismic data structure. Thirdly, we used a mixed loss function to improve the performance of FACNN further. Numerical tests demonstrated that the noise level map can improve the network's generalization. The attention structure with the mixed loss function can better protect the details of seismic events.

2. Theory and Method

The theory in this paper focuses on two unavoidable problems in AI-based denoising work: firstly, the destruction of the seismic data structure by the well-trained denoised network model; secondly, the diminished generalization ability of the trained network when the seismic data characteristics change. This section introduces three theories of this paper: the noise level map for solving Problem 1 with reduced generalization ability and the attention layer with the mixed loss function for solving Problem 2 to protect the detailed structure of seismic signals.

This section first introduces the architecture and characteristics of FACNN from two aspects in the following subsections: the attention gates' (AGs) structure and the noise level map, respectively. The whole network architecture is illustrated in Figure 1. U-Net [30] is a fully convolutional neural network for medical image segmentation. U-Net contains a

downsampling layer, upsampling layer, and skip connection operation, which can achieve a very excellent image segmentation function. It is named so because its shape resembles a U-shape and is widely used in deep learning research. Unlike the traditional U-Net, we integrated the noise level map M and AGs into a standard U-Net model. FACNN consists of the encoder and decoder parts, including the convolutional layer ($3 \times 3 \times 3$ Conv), activation function (ReLU), downsampling layer, upsampling layer, skip connection structure, and AGs. The input data shape of FACNN is $(Batchsize, Channels, H, W)$, in which $Channels = 2$ stands for the channel number, H is the time dimension, and W is the surface coordinate. The input data through the encoder structure include $3 \times 3 \times 3$ Conv, ReLU, and downsampling layers. The purpose of ReLU is to increase the deep neural network's nonlinearity. The definition of ReLU is: $ReLU(\cdot) = \max(\cdot, 0)$. In this paper, we used Maxpooling as the downsampling layer, significantly reducing the computational cost and saving GPU memory. The decoder shares an asymmetric structure with the encoder. FACNN also uses the skip connection structure, which can help accelerate the gradient back-propagation and convergence.

2.1. Attention Gates

One of the distinctive features of FACNN is the integration of AGs into a standard U-Net model. Specifically, we merged the AGs [31] into the skip connection operations. Attention gates can help the model assign weights to each input part, extracting critical and vital information and making more accurate predictions without imposing greater computational and storage expenses. This is beneficial to better preserve the structural information of the seismic data during the denoising process of the neural network.

AGs is a tool that allows updating the model parameters of the shallow feature layer according to the spatial region relevant for a target task. Through weight control, AGs can progressively suppress features in irrelevant background parts and make the feature layers more focused on the target part's characteristics. In addition, compared with the previous attention mechanisms, the proposed AGs in this paper is a memory-efficient method, which means it will not add too many extra model parameters. The AGs can be expressed as the following Equation (1):

$$f_{att} = \Psi^T [ReLU(w_l^T l_{i,c} + w_r^T r_{i,c} + b_1)] + b_2 \quad (1)$$

$$coe_{i,c} = \sigma_1(f_{att}(l_{i,c}, r_{i,c}; \theta_{att})) \quad (2)$$

where $l_{i,c}$ and $r_{i,c}$ are the left and right feature map in downsampling and upsampling, respectively. i and c denote the spatial and channel, respectively. We chose ReLU as the activation function. Ψ and w are the convolution operations. b_1 and b_2 are the bias. $\sigma_1(l_{i,c}) = \frac{1}{1+\exp(-l_{i,c})}$ corresponds to the sigmoid function.

After $l_{i,c}$ passes through the AGs, we can obtain a weight coefficient matrix $coe_{i,c}$ focusing on the target, and we multiplied it with $l_{i,c}$ to obtain the output of skip connection $\hat{l}_{i,c}$, as shown in the following Equation (3):

$$\hat{l}_{i,c} = l_{i,c} \cdot coe_{i,c}. \quad (3)$$

Finally, we used the feature concatenation to concat $\hat{l}_{i,c}$ and $r_{i,c}$, which are expressed as Equation (4):

$$\hat{r}_{i,c} = \text{Concat}(\hat{l}_{i,c}, r_{i,c}) \quad (4)$$

The architecture of the AGs is illustrated in Figure 2. The AGs can progressively suppress the irrelevant background features and improve the denoising performance without increasing intolerable extra computation. Adding the attention layer to the traditional neural network can make the network pay more attention to the structural information of the input data so that the network can better protect the structural information of the seismic data.

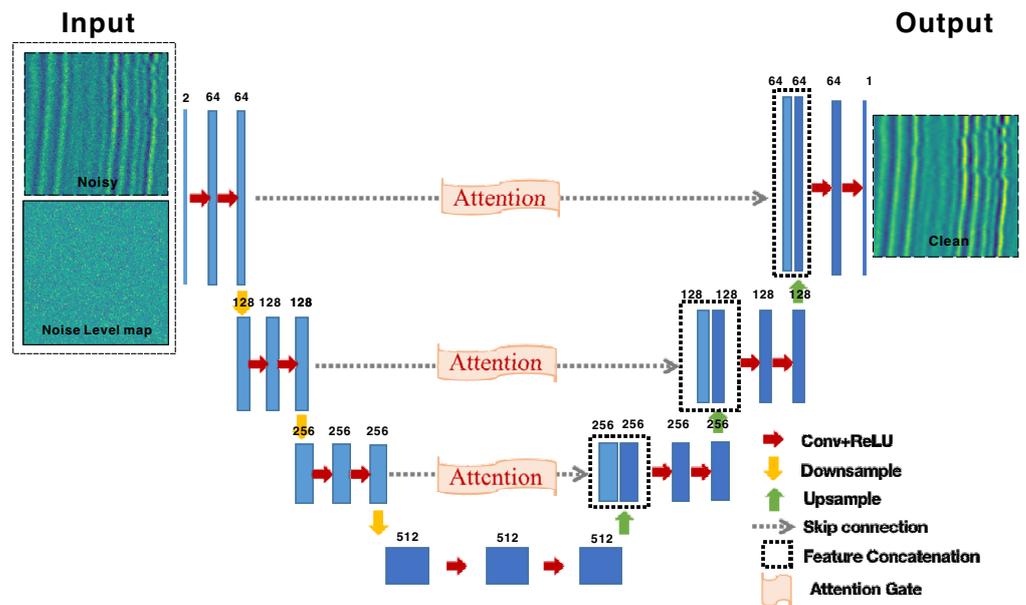


Figure 1. The architecture of FACNN.

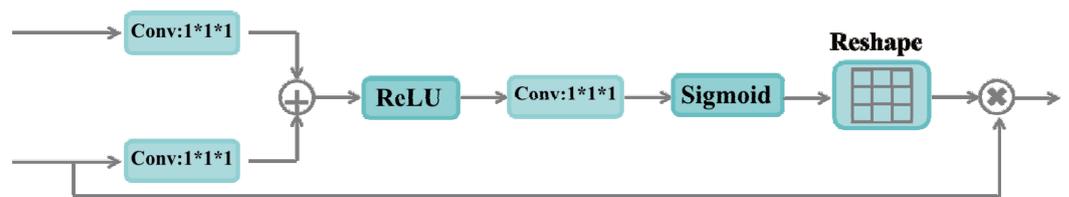


Figure 2. The architecture of the AGs.

2.2. Noise Level Map

The second distinctive feature of FACNN is that we added a noise level map M [20] as an additional channel in the network input to improve the generalization. When using the well-trained network model for testing, if the noise level and the testing data’s characteristics are inconsistent with the training dataset, the network cannot reach the best performance. Then, we introduced the theory of noise level map M .

We can ascribe the denoising problem of seismic data to the following Equation (5):

$$\bar{x} = \underset{x}{\operatorname{argmin}} \frac{1}{2\sigma^2} \|y - x\|^2 + \lambda\Phi(x). \tag{5}$$

where $\frac{1}{2\sigma^2} \|y - x\|^2$ is the data fidelity term with the noise level σ . $\Phi(x)$ is the regularization term, and λ controls the balance between $\frac{1}{2\sigma^2} \|y - x\|^2$ and $\Phi(x)$. This means that we introduced λ to control the trade-off between the noise reduction results and seismic data detail preservation.

Based on some optimization algorithms, the denoising function can be expressed as an implicit function, and λ is absorbed into σ . Then, we can obtain the following Equation (6):

$$\bar{x} = F(y, \sigma; \psi). \tag{6}$$

where σ inherited the characteristics of λ , and we used the seismic data and noise level as the inputs in this paper. However, the seismic data and noise level map M have different dimensions; we cannot directly feed them into our network model. To eliminate the problems of mismatching dimensions, we stretched the noise level σ into a noise level map M composed of σ , which has the same dimension as the input seismic data. Finally, Equation (6) can be expressed as Equation (7):

$$\bar{x} = F(y, M; \psi). \quad (7)$$

We added a noise level map M as an additional channel, controlling the trade-off between noise attenuation results and seismic data detail preservation. We can adjust the noise level map to keep the mean value of the input data body noise constant so that the network always maintains the best performance.

2.3. Mixed Loss Function

The loss function is used to evaluate the difference between the predicted $\hat{x}_i = F(y_i, M_i; \psi)$ and true values x_i of the model, and the choice of the loss function can largely influence the model's performance. M_i and ψ are the noise level maps and collection of all learnable parameters, respectively. Mathematically, the loss function is defined as:

$$\mathcal{L}(\psi) = \frac{1}{2N} \sum_{i=1}^N \|\hat{x}_i - x_i\|^2. \quad (8)$$

where $\mathcal{L}(\psi)$ is the loss function of the training process. N is the total number of pixels. To improve the preservation of the details of the denoised seismic image, we introduced multi-scale structural similarity (MS_SSIM) [32] and formed a new loss function. As an assessment for the image quality method, MS_SSIM is very sensitive to changes in local structure, which can significantly preserve the seismic event details. Mathematically, it can be expressed as:

$$MS_SSIM(a, b) = [l_I(a, b)]^{\alpha_I} \prod_{j=1}^I [c_j(a, b)]^{\beta_j} [s_j(a, b)]^{\gamma_j} \quad (9)$$

where I is the scale. $l(a, b) = \frac{2\mu_a\mu_b+c_1}{\mu_a^2+\mu_b^2+c_1}$, $c(a, b) = \frac{2\sigma_{ab}+c_2}{\sigma_a^2+\sigma_b^2+c_2}$, and $s(a, b) = \frac{\sigma_{ab}+c_3}{\sigma_a\sigma_b+c_3}$ represent three measurements between a and b . μ and σ represent the mean and the standard deviation. σ_{ab} is the covariance between data a and b . c_1, c_2, c_3 are three constants to avoid the denominator being stable. Based on Equation (9), we obtained an MS_SSIM loss as Equation (10):

$$\mathcal{L}^{MS_SSIM} = 1 - MS_SSIM(\hat{x}_i, x_i). \quad (10)$$

Furthermore, the mixed loss function is defined as:

$$\mathcal{L}^{Mix} = \phi \cdot \mathcal{L}^{MS_SSIM} + (1 - \phi) \cdot \mathcal{L}(\psi). \quad (11)$$

Through several experiments, we empirically set the weight ϕ to 0.5. Table 1 shows the SNR values on the same test dataset with different weights ϕ . It shows that FACNN can achieve the best performance when $\phi = 0.5$. Using the loss function with MS_SSIM can better preserve the structural information of the denoised seismic data. The details of the effective events on denoised images can be maintained better through our proposed mixed loss function.

Table 1. The variations of the SNR with respect to the different loss function weights ϕ .

ϕ	0.3	0.4	0.5	0.6	0.7
SNR (dB)	19.86	19.99	20.08	20.02	19.87

3. Numerical Tests

In this section, we mainly tested the denoising performance of the FACNN method proposed in this paper from both synthetic and real seismic data. Some comparison tests were included in the numerical tests to demonstrate the strategy proposed in the theoretical part to protect the detailed structure of the seismic data and to improve the generalization ability.

3.1. Synthetic Data Testing

In this subsection, we first tested FACNN on synthetic data. Firstly, we introduce the construction of synthetic training dataset. We usually constructed the training dataset for synthetic data by adding random noise. Considering the samples' diversity and the well-trained model's generalization, we chose 2D, 3D, pre-stack, and post-stack data with different features to construct the training dataset. Then, we split the dataset into small patches through the Monte Carlo strategy [33], removing the useless patches (almost all-zero data or particularly similar data). Moreover, we performed a data augmentation strategy based on the following operations: rotation transformation, mirror transformation, space-time downsampling, and intensity transformation [34]. Finally, we obtained a total of 22,000 training datasets, of which 18,000 were used for training and 2000 were used for validation, and the rest were used for testing. The dimension of the training input and output is (20, 2, 64, 64) and (20, 1, 64, 64), respectively. The additional channel of the training input is the noise level map M . The learning rate can be adjusted adaptively to accelerate the network's convergence speed, which starts from 10^{-3} and reduces to 10^{-4} . This paper set the batch size, patch size, and epoch to 20, 64×64 , and 100, respectively. The training process involved 100,000 iterations, with 2 h and 5 min total.

Figure 3a,b show the clean sigmoid model and the contaminated image. Then, we compare the denoising capabilities of FACNN, DnCNN [17], industrial RNA [2] (f-x random noise attenuation), and the adaptive prediction filter [35] (APF). The RNA method is the most widely used denoising method for data processing work in the petroleum industry. It achieves good denoising effects based on predicted seismic events and has now become the standard algorithm for denoising seismic data. APF is a state-of-the-art traditional denoising method, an extension of the RNA method, based on an adaptive prediction filter. It can achieve an excellent noise suppression effect. DnCNN is the standard comparison algorithm in neural network denoising for seismic data. We demonstrate the capability of FACNN in seismic data denoising by using two traditional algorithms and one deep learning algorithm as a comparison.

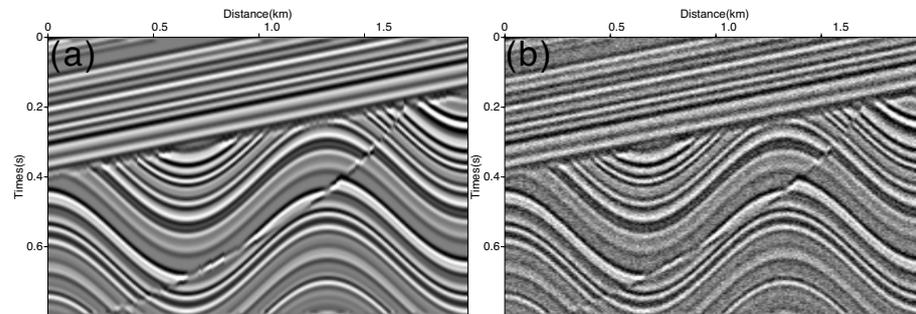
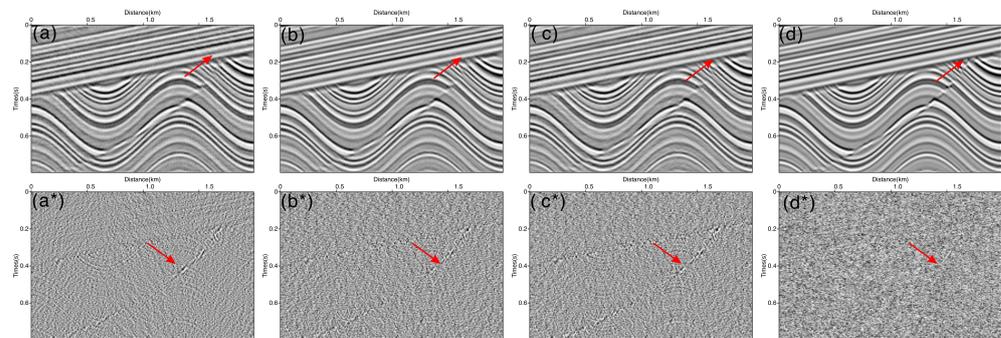
Figure 4a–d are the denoised results of RNA, APF, DnCNN, and FACNN, respectively. As the red arrow points out, we can easily find that the RNA method will leave much residual noise. For the state-of-the-art filtering method APF, the denoising effect was satisfactory, and the residual noise is almost invisible in Figure 4b. The denoising result was better than the RAN method for the widely used DnCNN method, but residual random noise still led to partially blurred data. The FACNN proposed in this paper showed a remarkable denoising performance, and its denoising result was further improved compared with the APF method.

Then, we further compared the ability of the structure preservation of the above four methods. Figure 4a*–d* are the corresponding removed noise profiles of the above four methods. The noise profile of the RNA method contained many effective signals, including geologic folds, faults, and unconformities. The preservation of the effective signals of the APF method was excellent, and only a tiny amount of fault information can be seen. Due to the DnCNN method not adapting to this noise level, much geological information remained in its noise profile. In contrast, the FACNN in this paper had no helpful information residual, and the preservation of the seismic structures was the best among the four methods.

By comparing the removed noise profiles, we can find that the detailed structural information of the seismic data can be better preserved by using our proposed attention structure with the mixed loss function. The variations of the SNR and SSIM for the four methods are shown in Table 2 below. It can be found that the SNR and SSIM of the FACNN method in this paper were the highest among several methods.

Table 2. The comparison of the SNR and SSIM for different methods.

Methods	RNA	APF	DnCNN	FACNN
SNR (dB)	13.98	18.99	18.96	20.08
SSIM	0.781	0.871	0.854	0.918

**Figure 3.** The diagram of the original sigmoid data and noisy data. (a) Original data. (b) Contaminated data.**Figure 4.** The comparison of AI-based denoised results with conventional method denoised results. (a,a*) RNA denoised result and its removed noise. (b,b*) APF denoised result and its removed noise. (c,c*) DnCNN denoised result and its removed noise. (d,d*) FACNN denoised result and its removed noise.

To examine the influence of the denoising methods on the original data amplitude, we extracted the 50th traces of four methods for comparison. The noisy data, RNA, APF, DnCNN, and FACNN methods are from top to bottom. The black line is the original signal. More overlap with the black line means that the method was more successful. As we can observe, FACNN had the best overlap with the original signal, which proves that FACNN is excellent in preserving the amplitude information. This also proves that the method in this paper hardly damages the structural information of the seismic data (Figure 5).

It is worth noting that AI-based methods have advantages over traditional methods. The denoising efficiency of traditional methods is not very high, especially the APF method, which is unsuitable for large-scale data processing because it requires iterations and consumes much time. In addition, we tested different denoising parameters for the two non-AI methods and, finally, chose the best denoising results. The traditional method always needs the selection of many parameters and manual intervention. Both AI-based methods consume only about one second of processing time for the sigmoid model and do not require manual intervention. In this experiment, we only compared the denoising ability of two deep learning algorithms at different noise levels.

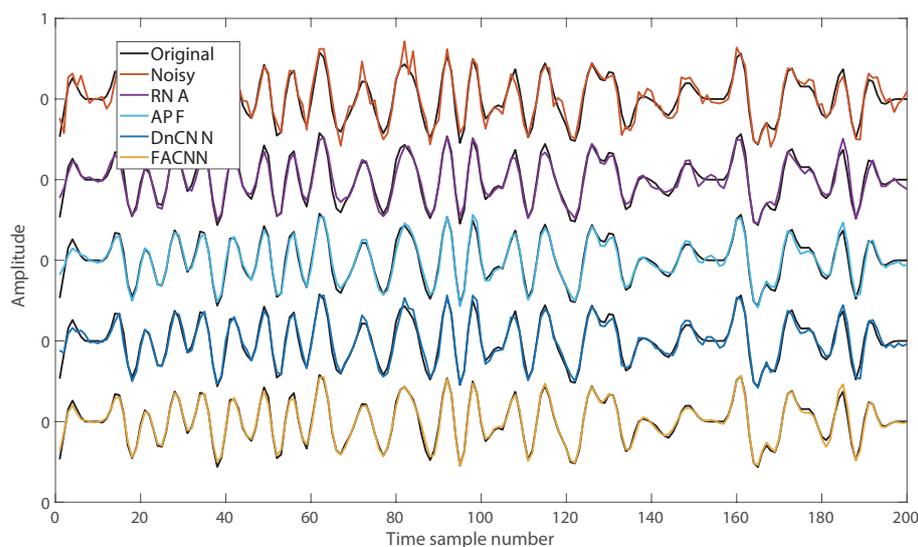


Figure 5. The 50th trace comparison of different methods.

Then, we compared the denoised results of DnCNN with FACNN at different noise levels to examine the superiority of the noise level map M . As we all know, when using a well-trained network model for testing, if the noise levels and the testing data's characteristics are inconsistent with the training dataset, the network cannot reach the best performance. Figure 6 is the comparison of DnCNN and FACNN with respect to the different noise levels, where (a–d) are the DnCNN denoised results and (a*–d*) are the corresponding FACNN denoised. The noise level testing in Figure 6a is the same as the training, and the denoising result of DnCNN was similar to FACNN. Moreover, as the noise level increased, more and more noise remained in the results of DnCNN (b–d). As for FACNN, we can adjust the noise level map so that the network always maintains the best performance. Therefore, its denoising performance only decreased slightly as the noise level increased. At the highest noise level, its denoising result was much better than DnCNN. The variations of the SNR for the two methods are shown in Table 3 below.

We found that, when the characteristics of the test data changed, the network performance of the DnCNN method decreased significantly. In contrast, the FACNN method proposed in this paper did not deteriorate the performance because we can adjust the noise level map so that the noise mean value of the input data remains constant. This also proves again that adding the noise level map can improve the generalization ability of the well-trained neural network.

Table 3. The variations of SNR with respect to the noise level.

Noise Level	40	50	60	70
DnCNN (dB)	18.96	16.81	15.68	13.67
FACNN (dB)	20.08	20.01	19.46	19.01

Through the previous tests, we demonstrated that the proposed FACNN in this paper can effectively protect the structural information of seismic signals. We used the control variable method for comparison to further demonstrate the contribution of the proposed attention structure and the mixed loss function. In this experiment, we compared the FACNN method of this paper, the method using the attention structure with the traditional loss function, and the method using the traditional U-Net with the mixed loss function proposed. We verified the effect of the attention and the mixed loss function proposed in this paper with the remaining variables held constant.

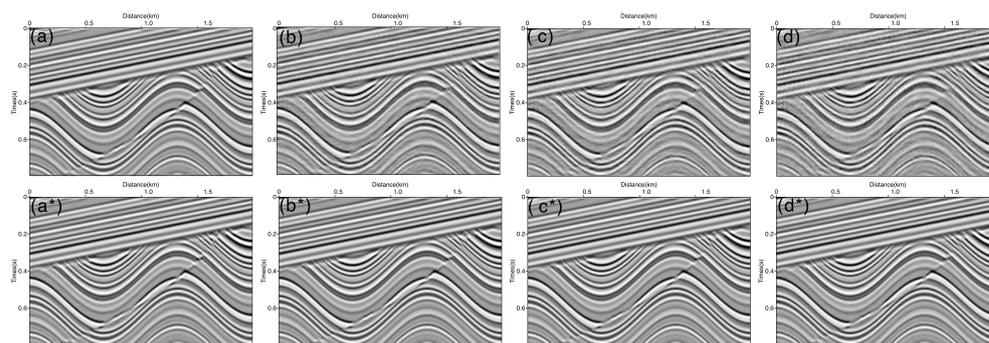


Figure 6. The comparison of DnCNN and FACNN with respect to the different noise levels. (a–d) are the DnCNN denoised results. (a*–d*) are the corresponding FACNN denoised.

Figure 7a,a* show the denoising results and the removed noise of FACNN, respectively. (b) and (b*) are the denoising results and noise profiles of the attention-CNN method based on the traditional loss function. (c) and (c*) are the denoising results and noise profiles of the U-Net method based on the proposed mixed loss function. We can find some effective signals in (b*) and (c*), and the residual information in (c*) is more than (b*). By comparing (a*) with (b*), we can demonstrate that the mixed loss function can protect the seismic signal structure. Furthermore, by comparing (a*) with (c*), it can be proven that the AGs can also play a role in protecting the seismic structure. The effective residual signal in (b*) is less than that in (c*), which proves that the AGs can protect the seismic signal structure better than the mixed loss function. This demonstrates that both the attention structure and the mixed loss function can contribute to the protection of the seismic data's detailed structure. The variations of the SNR and SSIM for the three methods are shown in Table 4 below. The SSIM tests also demonstrated that the attention structure and the mixed loss function proposed in this paper can protect the details of the seismic signal structure.

Table 4. The comparison of the SNR and SSIM for different methods.

Methods	FACNN	Attention-CNN with Tradition Loss	U-Net with Mixed Loss
SNR (dB)	20.13	19.45	19.36
SSIM	0.921	0.906	0.899

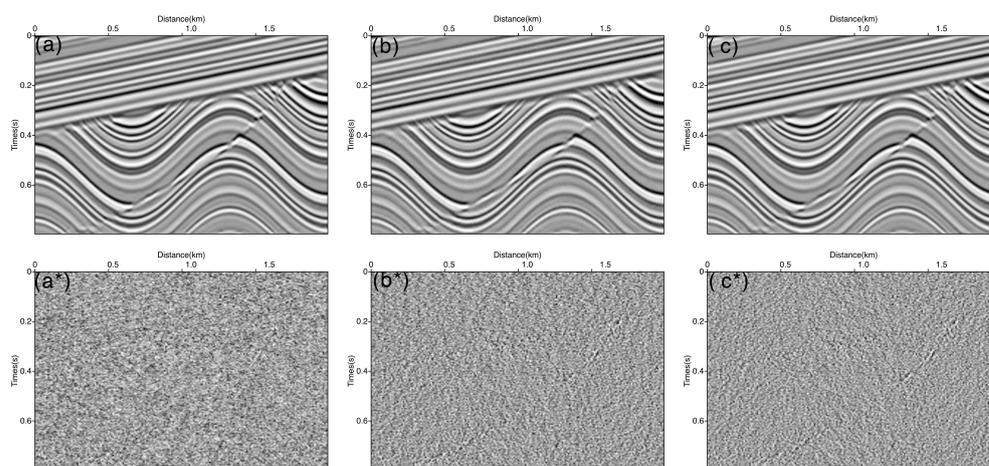


Figure 7. (a,a*) FACNN denoised result and its removed noise. (b,b*) Attention-CNN denoised result and removed noise, with tradition loss function. (c,c*) U-Net denoised result and removed noise, with proposed loss function.

3.2. Real Data Testing

To denoise the seismic field data, we added real data to the synthetic dataset to build a comprehensive training dataset, where the labels of the field data were denoised by the state-of-the-art APF method. The real seismic data were obtained from the field 3D seismic data of some oil fields in China, including land seismic data, desert seismic data, and marine seismic data. Through suitable processing and cropping, a total of 5000 training samples of real data were constructed to build a comprehensive training set. In addition, we used the same data augmentation strategy for the real data to form a comprehensive dataset with diversity [4,34]. Then, we used the transfer learning method [17], which means the well-trained synthetic data model was used as the pre-trained model, and finally, we obtained the denoised model adapted to the field seismic data. It is worth noting that during the training process, we built the noise level map by using the noise removed by the APF method so that the well-trained model can adjust to the data with different characteristics by adjusting the noise level map M .

In this section, we tested FACNN's effects in the field desert seismic data and the land post-stack seismic data. By comparing the processing results of different field seismic data, we can demonstrate the generalization ability and the ability to protect the structural features of seismic data.

Figure 8a is a part of the field desert post-stack seismic data, on which some noise residues can be observed. In this experiment, we compared the performance of the FACNN method in this paper with the DNCNN method and the traditional APF method in real seismic data denoising.

Figure 8b–d are the denoised results of the APF, DnCNN, and FACNN, respectively. We can find that the denoising results of the APF method (b) and DnCNN (c) were very similar, which demonstrated that the transfer learning method made the deep learning model outperform the traditional methods. Moreover, the denoising result of FACNN was further improved compared with DnCNN, demonstrating the effectiveness of AGs and the mixed loss function proposed in this paper.

Figure 8b*–d* are the removed noise profiles by APF, DnCNN, and FACNN, respectively. We can observe that DnCNN had many residual effective signals due to not adapting to the characteristics of the real seismic data in this region. However, in Profile (d*), we can hardly find the residual geological structure information. This demonstrated that the FACNN in this paper can also better preserve the structural information of seismic signals by the noise level map, AGs, and mixed loss function.

Figure 9 is the comparison of the three methods of the FK spectrum in real seismic data, where (a–d) represent the FK spectrum of the field seismic data, the denoising results of the APF method, the denoising results of DnCNN, and the denoising results of FACNN, respectively. All three methods can suppress the noise in Figure 9a. There is still a slight residual noise in the high and low wave numbers in (b,c). In contrast, the FACNN method in this paper can suppress the random noise better. In addition, in Figure 9d, we can observe that the effective signal of the low wavenumber part is better retained. This proved that FACNN can better protect the characteristics of the effective signal while maintaining good denoising ability.

Next, we further verified the generalization ability of FACNN with field land seismic data. The generalization ability of the method in this paper can be demonstrated by testing the field seismic data from different blocks. In this experiment, we compared the denoising performance of a standard U-Net network without using the noise level map and FACNN with the attention structure and noise level map in this paper.

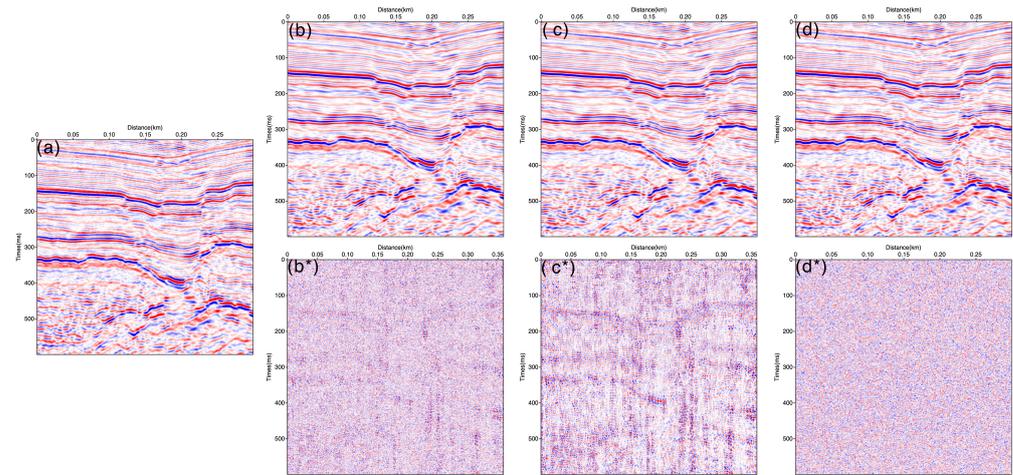


Figure 8. The comparison of three methods’ denoised results in real seismic data. (a) The field desert post-stack seismic data. (b,b*) Denoised result and removed noise profile using APF method. (c,c*) Denoised result and removed noise profile using DnCNN. (d,d*) Denoised result and removed noise profile using FACNN.

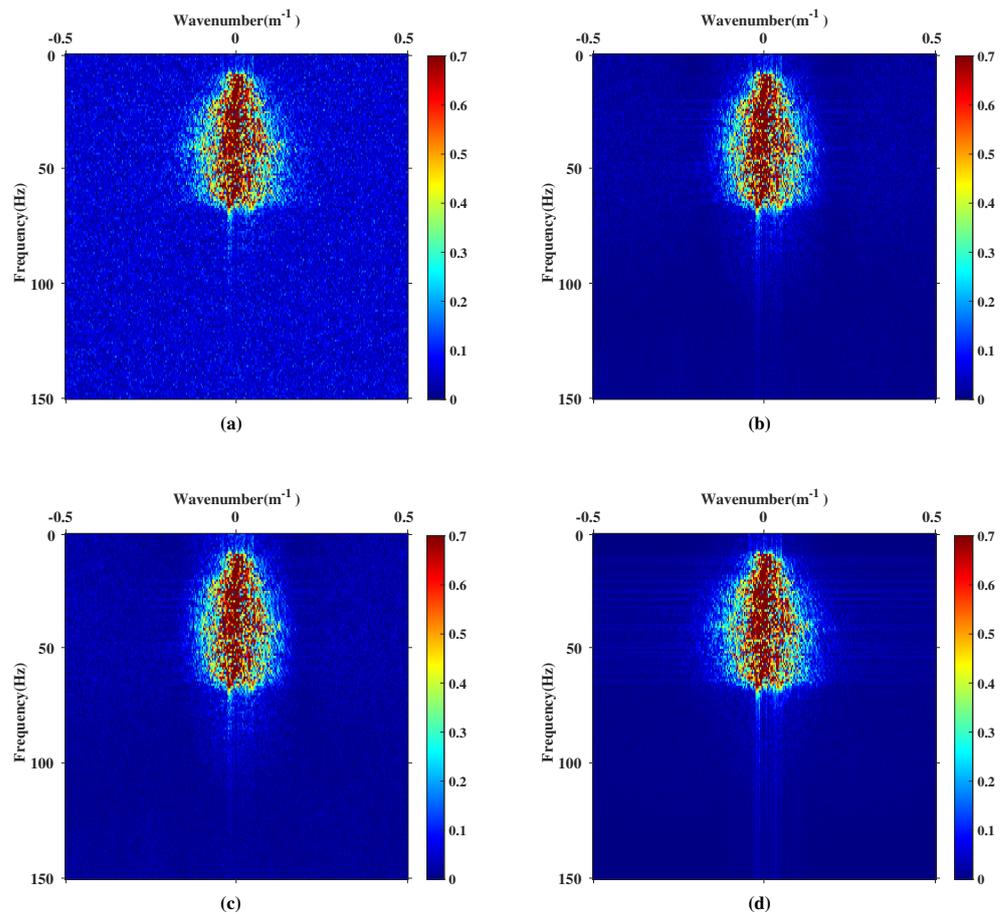


Figure 9. The comparison of three methods of the FK spectrum in real seismic data. (a) The FK spectrum of field desert post-stack seismic data. (b) The FK spectrum of the denoised result using the APF method. (c) The FK spectrum of the denoised result using DnCNN. (d) The FK spectrum of the denoised result using FACNN.

Figure 10a shows a part of the land real seismic data with a low SNR. Figure 10b,c are the denoised results of U-Net without the noise level map and attention-CNN with

the noise level map, respectively. The improvement of (b,c) lies in the AGs and mixed loss functions proposed in this paper. The rest of the network structure, training set, and parameters were the same. We can find that the denoising results of FACNN (c) were significantly better than those without the noise level map (b). In comparing the noise difference profiles (d,e), we found that using the noise level map led to better generalization and better event protection of the seismic data. Furthermore, we can find that, by using the structure-preserving method proposed in this paper, we can protect the structural information of the seismic signal.

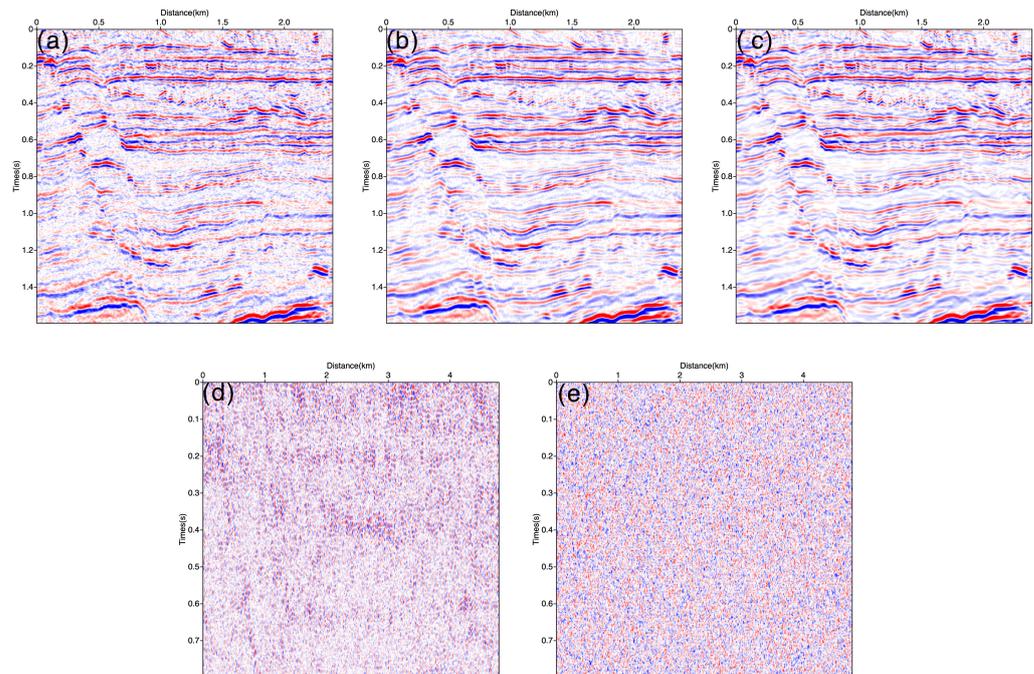


Figure 10. The comparison of the denoised results in real seismic data. (a) The contaminated real seismic data. (b,d) Denoised result and removed noise using U-Net without the noise level map. (c,e) Denoised result and removed noise using attention-CNN with the noise level map.

Figure 11 compares the FK spectrum in real seismic data. (a–c) are the FK spectrum of contaminated real seismic data, the denoised result using U-Net without the noise level map, and the denoised result using attention-CNN with the noise level map. It can be found that the FK spectrum of Figure 11a contains a large number of random noise residues in the whole wavenumber range. The random noise in Figure 11b,c is well suppressed. We can find that the noise in the whole wavenumber part of (c) is significantly better suppressed. It is worth noting that the morphology and characteristics of the effective signal are destroyed in the U-Net-based suppression results in (b). In contrast, in the Attention-CNN-based suppression results in (c), the information of the effective signal is preserved completely. This reaffirmed that the FACNN proposed in this paper has good generalization and structure-preserving ability.

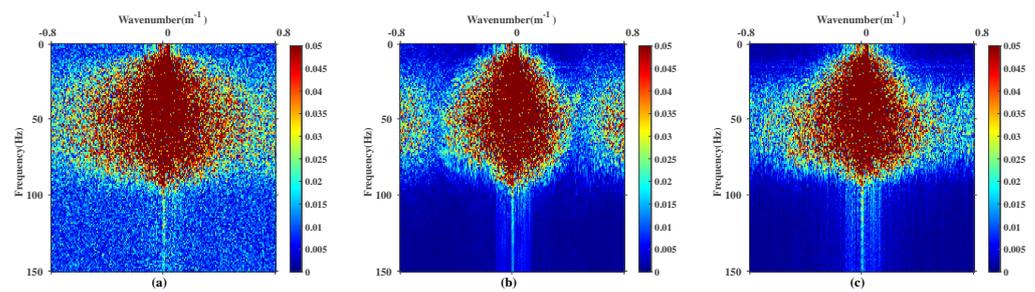


Figure 11. The comparison of the FK spectrum in real seismic data. (a) The FK spectrum of contaminated real seismic data. (b) The FK spectrum of the denoised result using U-Net without the noise level map. (c) The FK spectrum of the denoised result using attention-CNN with the noise level map.

4. Discussion

This paper proposed solutions for two problems in AI-based denoising of seismic data. Firstly, for the problem that the trained neural network lacks generalization, this paper added a noise level map to improve the network's generalization ability. Secondly, for the problem that the denoising method will damage the structural information of seismic signals, this paper added the attention layer and mixed loss function to improve the ability of the network to protect the structural information of seismic signals.

The research ideas in this paper can be extended from the following aspects. Firstly, the noise level map proposed in this paper can be combined with the self-supervised learning algorithm for the random noise removal aspect. Secondly, the structure-preserving algorithm proposed in this paper can be applied to other types of noise removal work in seismic data, such as surface waves and linear noise.

5. Conclusions

In this paper, we proposed a flexible attention-CNN (FACNN) and realized the intelligent denoising work of seismic data. We integrated the attention gates (AGs) with the mixed loss function in a standard U-Net model. Then, we added a noise level map as an additional channel in the network input data body, making a single CNN model that can handle the noise models with different noise characteristics, even spatially variant noises. Adding the noise level map can improve the network's generalization ability, and adding the attention structure with the mixed loss function can better protect the structural information of the seismic data. The testing results on synthetic and real data demonstrated the superiority of our proposed method. Extensive experimental results showed that our method has better generalization and can better protect the details of seismic events.

Author Contributions: Conceptualization, W.L.; methodology, W.L.; software, W.L. and T.W.; validation, W.L.; formal analysis, W.L.; investigation, W.L.; resources, T.W. and H.L.; writing—original draft preparation, W.L.; writing—review and editing, H.L.; visualization, T.W.; supervision, H.L.; project administration, H.L.; funding acquisition, H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the National Engineering Laboratory for Offshore Oil Exploration Opening Funds.

Data Availability Statement: Data associated with this research are available and can be obtained by contacting the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. 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 Sens.* **2020**, *58*, 4680–4690. [[CrossRef](#)]
2. Canales, L.L. Random noise reduction. In *SEG Technical Program Expanded Abstracts*; Society of Exploration Geophysicists: Houston, TX, USA, 1984; pp. 525–527.

3. Zhang, J.; Li, Y.; Wu, N. Noise attenuation for seismic data by hyperbolic-trace time-frequency peak filtering. *IEEE Geosci. Remote Sens. Lett.* **2014**, *12*, 601–605. [[CrossRef](#)]
4. Zhao, X.; Li, Y.; Zhuang, G.; Zhang, C.; Han, X. 2D TFPF based on Contourlet transform for seismic random noise attenuation. *J. Appl. Geophys.* **2016**, *129*, 158–166. [[CrossRef](#)]
5. Xiong, M.; Li, Y.; Wu, N. Random-noise attenuation for seismic data by local parallel radial-trace TFPF. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 4025–4031. [[CrossRef](#)]
6. Ma, H.; Qian, Z.; Li, Y.; Lin, H.; Shao, D.; Yang, B. Noise reduction for desert seismic data using spectral kurtosis adaptive bandpass filter. *Acta Geophys.* **2019**, *67*, 123–131. [[CrossRef](#)]
7. Freire, S.L.M.; Ulrych, T.J. Application of singular value decomposition to vertical seismic profiling. *Geophysics* **1988**, *53*, 778–785. [[CrossRef](#)]
8. Al-Yahya, K.M. Application of the partial karhunen-loève transform to suppress random noise in seismic sections 1. *Geophys. Prospect.* **1991**, *39*, 77–93. [[CrossRef](#)]
9. Liang, X.; Li, Y.; Zhang, C. Noise suppression for microseismic data by non-subsampled shearlet transform based on singular value decomposition. *Geophys. Prospect.* **2018**, *66*, 894–903. [[CrossRef](#)]
10. Zhang, R.; Ulrych, T.J. Physical wavelet frame denoising. *Geophysics* **2003**, *68*, 225–231. [[CrossRef](#)]
11. Fomel, S.; Liu, Y. Seislet transform and seislet frame. *Geophysics* **2010**, *75*, V25–V38. [[CrossRef](#)]
12. Cao, J.; Zhao, J.; Hu, Z. 3D seismic denoising based on a low-redundancy curvelet transform. *J. Geophys. Eng.* **2015**, *12*, 566–576. [[CrossRef](#)]
13. Hennenfent, G.; Herrmann, F.J. Seismic denoising with nonuniformly sampled curvelets. *Comput. Sci. Eng.* **2006**, *8*, 16–25. [[CrossRef](#)]
14. Chen, Y.; Zhang, D.; Jin, Z.; Chen, X.; Zu, S.; Huang, W.; Gan, S. Simultaneous denoising and reconstruction of 5-D seismic data via damped rank-reduction method. *Geophys. J. Int.* **2016**, *206*, 1695–1717. [[CrossRef](#)]
15. Huang, W.; Feng, D.; Chen, Y. De-aliased and de-noise cadzow filtering for seismic data reconstruction. *Geophys. Prospect.* **2020**, *68*, 553–571. [[CrossRef](#)]
16. Naghizadeh, M.; Sacchi, M. Multidimensional de-aliased cadzow reconstruction of seismic records. *Geophysics* **2013**, *78*, A1–A5. [[CrossRef](#)]
17. Yu, S.; Ma, J.; Wang, W. Deep learning for denoising. *Geophysics* **2019**, *84*, V333–V350. [[CrossRef](#)]
18. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
19. Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Trans. Image Process.* **2017**, *26*, 3142–3155. [[CrossRef](#)]
20. Zhang, K.; Zuo, W.; Zhang, L. FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Trans. Image Process.* **2018**, *27*, 4608–4622. [[CrossRef](#)]
21. Yu, Y.; Rashidi, M.; Samali, B.; Mohammadi, M.; Nguyen, T.N.; Zhou, X. Crack detection of concrete structures using deep convolutional neural networks optimized by enhanced chicken swarm algorithm. *Struct. Health Monit.* **2022**, *21*, 14759217211053546. [[CrossRef](#)]
22. Yu, Y.; Wang, C.; Gu, X.; Li, J. A novel deep learning-based method for damage identification of smart building structures. *Struct. Health Monit.* **2019**, *18*, 143–163. [[CrossRef](#)]
23. Yuqing, W.A.N.G.; WenKai, L.; JinLin, L.; Meng, Z.; YongKang, M. Random seismic noise attenuation based on data augmentation and CNN. *Chin. J. Geophys.* **2019**, *62*, 421–433.
24. Zhang, M.; Liu, Y.; Chen, Y. Unsupervised Seismic Random Noise Attenuation Based on Deep Convolutional Neural Network. *IEEE Access* **2019**, *7*, 179810–179822.
25. Dong, X.T.; Li, Y.; Yang, B.J. Desert low-frequency noise suppression by using adaptive DnCNNs based on the determination of high-order statistic. *Geophys. J. Int.* **2019**, *219*, 1281–1299. [[CrossRef](#)]
26. Tian, C.; Xu, Y.; Li, Z.; Zuo, W.; Fei, L.; Liu, H. Attention-guided CNN for image denoising. *Neural Netw.* **2020**, *124*, 117–129. [[CrossRef](#)] [[PubMed](#)]
27. Dong, X.; Lei, Y.; Tian, S.; Wang, T.; Patel, P.; Curran, W.J.; Jani, A.B.; Liu, T.; Yang, X. Synthetic MRI-aided multi-organ segmentation on male pelvic CT using cycle consistent deep attention network. *Radiother. Oncol.* **2019**, *141*, 192–199. [[CrossRef](#)] [[PubMed](#)]
28. Yu, J.; Wu, B. Attention and hybrid loss guided deep learning for consecutively missing seismic data reconstruction. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–8. [[CrossRef](#)]
29. Saad, O.M.; Oboué, Y.A.S.I.; Bai, M.; Samy, L.; Yang, L.; Chen, Y. Self-attention deep image prior network for unsupervised 3D seismic data enhancement. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–14.
30. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference On Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2015; pp. 234–241.
31. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention u-net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999.
32. Li, J.; Wu, X.; Hu, Z. Deep learning for simultaneous seismic image super-resolution and denoising. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–11.

33. Yu, S.; Ma, J.; Osher, S. Monte Carlo data-driven tight frame for seismic data recovery. *Geophysics* **2016**, *81*, V327–V340. [[CrossRef](#)]
34. Li, W.; Wang, J. Residual learning of cycle-GAN for seismic data denoising. *IEEE Access* **2021**, *9*, 11585–11597. [[CrossRef](#)]
35. Liu, Y.; Liu, N.; Liu, C. Adaptive prediction filtering in txy domain for random noise attenuation using regularized nonstationary autoregression. *Geophysics* **2015**, *80*, V13–V21. [[CrossRef](#)]