ORIGINAL ARTICLE

Improving signal fidelity for deep learning-based seismic interference noise attenuation

Jing Sun[1,2] ⬤ | Song Hou[3]

[1]Department of Geosciences, University of Oslo, Oslo, Norway

[2]CGG Services (Norway) AS, Oslo, Norway

[3]CGG Services (UK) Ltd, Crawley, West Sussex, UK

Correspondence
Jing Sun, Department of Geosciences, University of Oslo, Sem Sælands vei 1, Oslo No 0371, Norway.
Email: jingsun8803@hotmail.com

Funding information
CGG and NFR (The Research Council of Norway) through an industrial PhD, Grant/Award Number: 314179

Abstract

Deep learning has shown a considerable potential to significantly improve processing efficiency but has not yet been widely deployed to production projects of seismic signal separation such as seismic interference attenuation. The main reasons are: First, the industry has high standards for signal fidelity, which are critical for the success of subsequent seismic imaging, and deep neural network methods have not yet matched the required level; second, the network's interpretability issue has affected many geophysicists and sponsors' trust in the deep learning technique. To develop deep neural network methods towards the end of benefiting real-world production, we first attempt to better understand their performance, especially in how they make use of local and global features of the data. A novel quantitative research of the overall network model behaviour on synthetic data is conducted. We simulate three types of coherent seismic data components in the shot domain, blend them together and then train a network to separate them. In this process, random noise, a component having only learnable local features, is selectively injected into the network's training pairs. Three network models sharing the same architecture are trained individually, and they show distinctive behaviours when applied to the same test data. Step-by-step analysis of each of them reveals that training the network with additional random noise injected into both the input and the output channel of the desired signal can lead to a decent prediction of the coherent noise based on good learning of the global features and, in the meantime, preserve almost all the data information from being lost. We propose this key lesson we learnt as a new method to improve the network's signal fidelity for shot-domain seismic interference attenuation, which is essentially a signal separation task. Its effectiveness is demonstrated on field data from Africa with a comparison to a conventional physics-based seismic interference attenuation method used in production.

KEYWORDS
Attenuation, neural network, noise, seismic data processing, signal processing

INTRODUCTION

The recent development of deep neural networks (DNNs) and powerful graphics processing units (GPUs) has enabled many deep learning (DL) applications in the seismic community (Ovcharenko, 2021; Qu et al., 2021; Song et al., 2021; Wu et al., 2019). The wide adoption of this technology has been well illustrated by the increasing number of publications related to DNNs. However, the interpretability issue of DNN remains to be studied. This is a growing concern for both the DNN users regarding the nature of their black box and the sponsors regarding whether to trust the research and development value of this technique. In addition, recent studies of seismic applications have highlighted the challenges for DNNs to outperform physics-based algorithms in terms of processing quality (Hou & Hoeber, 2020). DNNs can be fooled (Nguyen et al., 2015) and the uncertainty in their performance cannot be easily measured (Ghahramani, 2015). DNNs make use of an enormous set of parameters, which are automatically updated during the data-driven training process but are more out of control compared with those conventional physics-based algorithms, whose parameters can be manually fine-tuned.

The processing task focused on using DL to deal with in this paper is seismic interference (SI) attenuation. SI is a representative type of coherent noise that occurs when unwanted energy from nearby seismic sources not linked to the survey is recorded. It is typically observed as linear or hyperbolic events with high amplitudes appearing at different arrival times in each seismic shot (Jansen et al., 2013). The SI noise within one survey may differ from sail line to sail line depending on the relative position of its source origin to the receivers. As SI noise is generated by powerful dedicated sources for seismic exploration and propagated through the water column, it tends to be well preserved over large distances (Akbulut et al., 1984; Jansen et al., 2013) and may overlap with reflections from sub-surface layers that have significantly lower amplitudes.

Signal fidelity is an essential measure of processing quality for SI attenuation and other signal separation tasks in seismic processing, which refers to how well a given algorithm does at preserving the original seismic signal intact. It is closely related to the accuracy and completeness of the separated noise, as the total information contained in the raw data is determined. Visual inspection is typically used as the primary method of quality control (QC) in actual processing projects for SI attenuation. This is due to the fact that SI noise emerges intermittently in a field survey while the underlying reflection data remain unknown. Calculating the local similarity maps allows for a more in-depth analysis of the signal fidelity after a SI attenuation technique has been applied (Chen & Fomel, 2015; Fomel, 2007). In addition, one useful way for locating signal leakages is to QC the processing results, particularly the SI noise that has been removed, in the CMP-stacked domain.

To improve the signal fidelity of a DNN for application to real production data (Hou & Messud, 2021), we need to understand what our DNN has learnt. Printing feature maps of the hidden layers has been considered as a way for researchers to learn from their designed networks (Sun et al., 2020a, 2020b), but in many cases, these maps are too abstract to analyse. Therefore, in this paper, we designed a synthetic study to investigate the overall DNN model behaviour through quantitative research.

We simulated three types of seismic data components and trained a DNN to separate them from their blend in the shot domain, where they are all coherent. On this basis, we studied the impact of injecting random noise into the training data and validated that this can be a useful way to drive the DNN to form different processing patterns of mapping the given input to the desired output. Sietsma and Dow (1991) suggested adding random noise with proper characteristics as a means of data augmentation. This method was shown to improve the generalized performance of a network (Holmstrom & Koistinen, 1992). Differing from the previous work, our key purpose here is to investigate the DNN's learning focus on local or global features of the data. Random noise, which lacks learnable global features for its randomness, is used as a tool to steer the DNN.

Three processing patterns of the DNN are summarized with an interpretation of different model behaviours. We find that training the DNN with additional random noise injected into both the input and the target channel of the desired signal can steer the DNN to better learn global features of the data while maintaining local features from being discarded. We therefore propose this as a new method for improving the DNN's signal fidelity. We demonstrated this new method on a real data example to attenuate SI noise.

Before closing this section, it is worth noting that the word 'channel' is used as terminology in both fields of seismic processing and DL with different meanings. In this paper, 'channel' is used in the machine learning sense to represent a certain component of an image.

QUANTITATIVE RESEARCH OF MODEL BEHAVIOURS VIA SYNTHETICS

In this section, we designed a series of synthetic experiments based on a strategy of quantitative research. All the experiments here share the same DNN architecture and test data. We controlled each experiment to have a unique variable in their training data with the next-step experiment, so that each of them could represent a distinctive training of a DNN model. These trained DNN models are observed to perform differently than each other at the inference stage. Our objective is to reason about these distinctive behaviours and to discover the underlying processing patterns of the DNN.
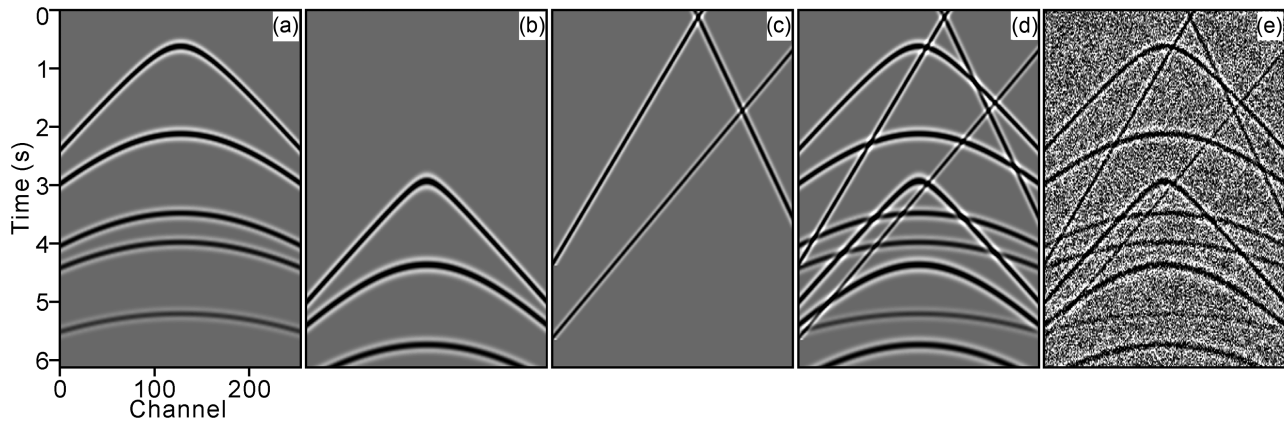
FIGURE 1 Examples of (a) desired signal, (b) blending noise, (c) linear noise, (d) their summation simulated for the training and validation of DNN and (e) with additional random noise injected into (d)

Training data

As previously stated, three types of seismic data components observed coherently in the shot domain are simulated, as shown in Figure 1a–c, respectively. The data component in Figure 1a represents the desired signal in a noisy shot gather. As we can see, the simulated desired signal has events with hyperbolic curvatures as if it were acquired from a source-over-streamer (Vinje et al., 2017) acquisition. The data component in Figure 1b represents a type of hyperbolic coherent noise, which can be regarded as blending noise from a blended acquisition using a short shot point interval or SI noise coming from the side. For convenience, we refer to this data component as 'blending noise'. Under this assumption, 300 blended shots (desired signal plus blending noise) were simulated by using a shooting interval of 3.0 s with a shot-to-shot dithering of maximum ± 500 ms, consisting of one blended batch. In addition, we simulated linear noise as shown in Figure 1c to mimic SI noise coming from ahead or astern; such SI types are most often recorded in real-world acquisition. For each blended shot in such a batch, we simulated different numbers of linear events with different dips and amplitudes. By changing the shooting interval and simulating different numbers of events of the desired signal, 30 batches (9000 shots) were obtained. Twenty-four of them (7200 shots) were used as the training data, and the remaining six (1800 shots) were used for validation. The size of each seismic image is 1536×256. The core task is to separate these three data components from their mixture (Figure 1d) in the shot domain using a DNN.
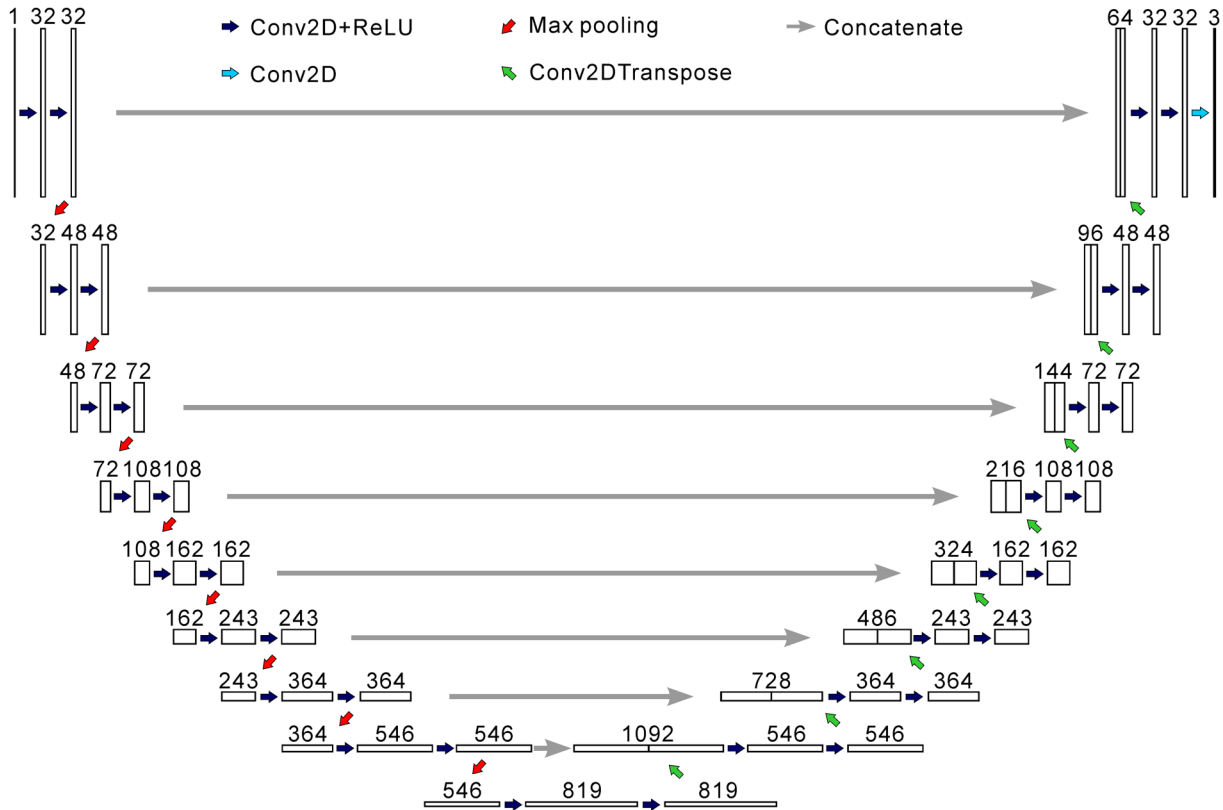
DNN architecture

The DNN we used in this study has a U-Net architecture (first proposed by Ronneberger et al., 2015) as shown in Figure 2. The core idea behind U-Net architectures consists

of three parts: an encoder transforms the input to feature maps that are essentially sparse representations of the input, a decoder reverse-transforms the feature maps back to the target (Hou & Hoeber, 2020), and skip connections that allow more information to be retained from previous layers of the DNN. One key benefit of such an architecture is that it enables the simultaneous usage of global location and context (Alom et al., 2018). In addition, U-Net is fast and computationally effective to run (Sun et al., 2022). Such advantages make U-Net frequently adopted in our studies; a word of experience is that different hyper-parameters are not very sensitive to the DNN model's accuracy compared with the training data sets. Though the scope of this study is limited to U-Net, the proposed idea of analysing DNN's overall model behaviour should be instructive if extended to other DNN architectures.

In our testing of several hyper-parameters, we discovered that they were less sensitive to the DNN model accuracy than the training data sets. In our employed U-Net, the encoder is made up of convolutional layers and max pooling, yielding a multilevel, multi-resolution feature representation. The convolutional layers employ a typical filter size of 3×3 and ReLU activation function. The max pooling employs a stride of 2 and a pool size of 2×2. Correspondingly, the decoder employs transposed convolutional layers with a stride of 2 to up-sample the low-resolution features. Skip connections between the encoding and decoding paths employ concatenation operations to ensure information fusion between high- and low-level information. No activation function is applied at the last layer of the employed U-Net. Adam optimization (Kingma & Ba, 2014) is adopted, and a batch size of 4 is selected.

Test data

The test data were produced similarly to the training data, with the major difference being that we created jittered gathers by

FIGURE 2 Architecture of the employed DNN. Each box represents a collection of feature maps from the previous operation, and the numbers above (e.g., 32, 48 and 72) represent the number of feature maps.
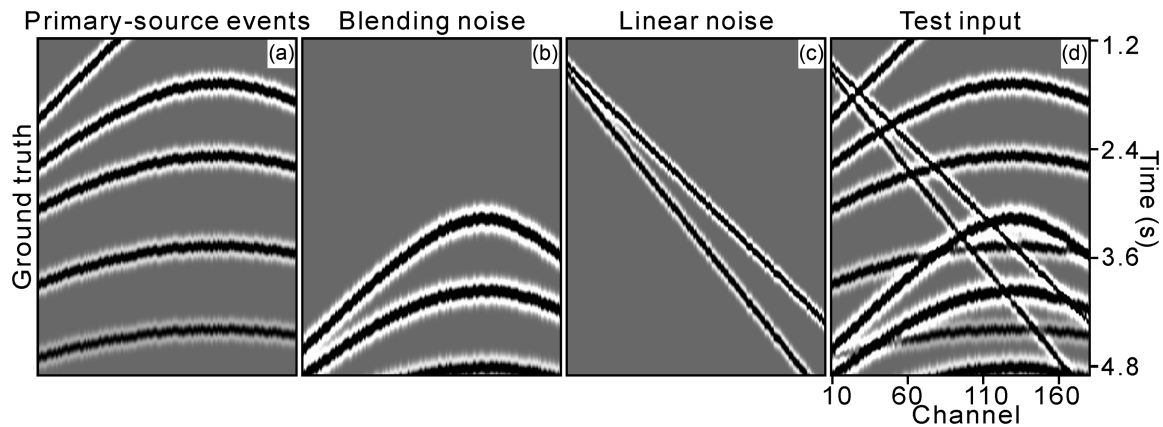
FIGURE 3 Test data of the three synthetic experiments. The figure from left to right shows zoomed sections of the true (a) desired signal, (b) blending noise, (c) linear noise and (d) their summation (i.e., test input).

employing random dither on each trace. Such local jitter can be found on every event of all three coherent components, as shown in Figure 3a–d, which are the zoomed sections of the true desired signal, true blending noise and true linear noise of a test example and their mixture (test input), respectively. The

reason for introducing such unrealistic local jitters is to provide visible local features for seismic events. With the artificially introduced local jitters, we should be able to distinguish the performance of a trained DNN model on the local versus global features of the seismic events. Again, the training

TABLE 1 Summary of the three experiments on synthetic data

Training		Experiment 1	Experiment 2	Experiment 3
Input		Summation	Summation + additional random noise	Summation + additional random noise
Output channels	# 1	Desired signal	Desired signal	Desired signal + additional random noise
	# 2	Blending noise	Blending noise	Blending noise
	# 3	Linear noise	Linear noise	Linear noise

data are always un-jittered (as shown in Figure 1), but the trained models are all tested on the jittered input (as shown in Figure 3d).

Summary of the three experiments

In total, three different DNN models sharing the same architecture (as shown in Figure 2) but trained individually were obtained and applied to the same jittered test data for comparison and analysis. Their distinctive training was based on whether random noise was selectively injected into the training data and, if so, whether it was injected into only the training inputs or both the training inputs and targets. A summary of the three experiments is given in Table 1. It should be noted that we designed our synthetic events to be simple compared with the real-field seismic events, and we also created some unrealistic random jitters on the test input events. This is to emphasize the distinctly visible characteristics among different model behaviours.

Experiment 1

The inputs of the first training data set designed for Experiment 1 consist of only the coherent components, and we train the DNN to separate them into three output channels. Once trained, the model (from now on called Model 1; similarly, DNN models trained in Experiments 2 and 3 in the following subsections will be called Models 2 and 3, respectively) was applied to the test data, of which an example is shown in close-up in Figure 3d. Model 1's predictions for the desired signal, blending noise and linear noise are shown in Figure 4a–c, and we compare them to the corresponding ground truth (Figure 3a–c).

We observe residuals of the local jitters in the difference (Figure 4e) between the true (Figure 3a) and predicted desired signals (Figure 4a). Similar observations can also be found in the comparisons between the ground truth and predicted channels of the blending noise and linear noise (comparing Figures 3b with 4b, and 3c with 4c). However, the summation of three output channels from Model 1 (Figure 4d) is very close to the raw test input (Figure 3d), where almost all the

local jitters on events have been regained. Their difference is nearly invisible (Figure 4f). The performance of Model 1 is similar to the decomposition problem that can occur in conventional signal processing. Therefore, we denominate this processing pattern as a 'decomposition pattern'.

Experiment 2

On the basis of the training data set used in Experiment 1, we further injected random noise into the training inputs (examples comparing training inputs before and after random noise injection are given in Figure 1d and e) and thereby produced the training data set for Experiment 2. This is an attempt to steer the DNN to focus more on learning the global features of the data components, based on the property of the random noise that it lacks learnable global features and only has learnable local features. Random noise of uniform and normal distributions was investigated with no difference found regarding their influence on the DNN. The displayed examples here employ uniform random noise. Note again that in the training process here, we did not ask the DNN to make any prediction for this additional random noise. For this experiment, we trained the DNN to output the three coherent components, free of random noise, into three channels. After being trained, Model 2 was tested on the same data as Experiment 1 (Figure 3d), and its output desired signal, blending noise and linear noise are displayed in Figure 5a–c, respectively. Their summation is shown in Figure 5d.

As we can see, Model 2 managed to reconstruct the shapes and curvatures of the coherent data components at the correct positions but lost all the local jittering on their events (Figure 5e and f). This is because, in this case, having or lacking learnable global features is the essential difference between the coherent data components that are supposed to be predicted and the random noise that is supposed to be discarded. Therefore, the DNN could focus on making use of global features of all the data components to achieve such a signal separation task, which led to the natural/automatic dropping of the random noise. This experiment demonstrated the possibility of driving the DNN's estimation of data components towards emphasizing their global features. The processing pattern of Model 2 worked similarly
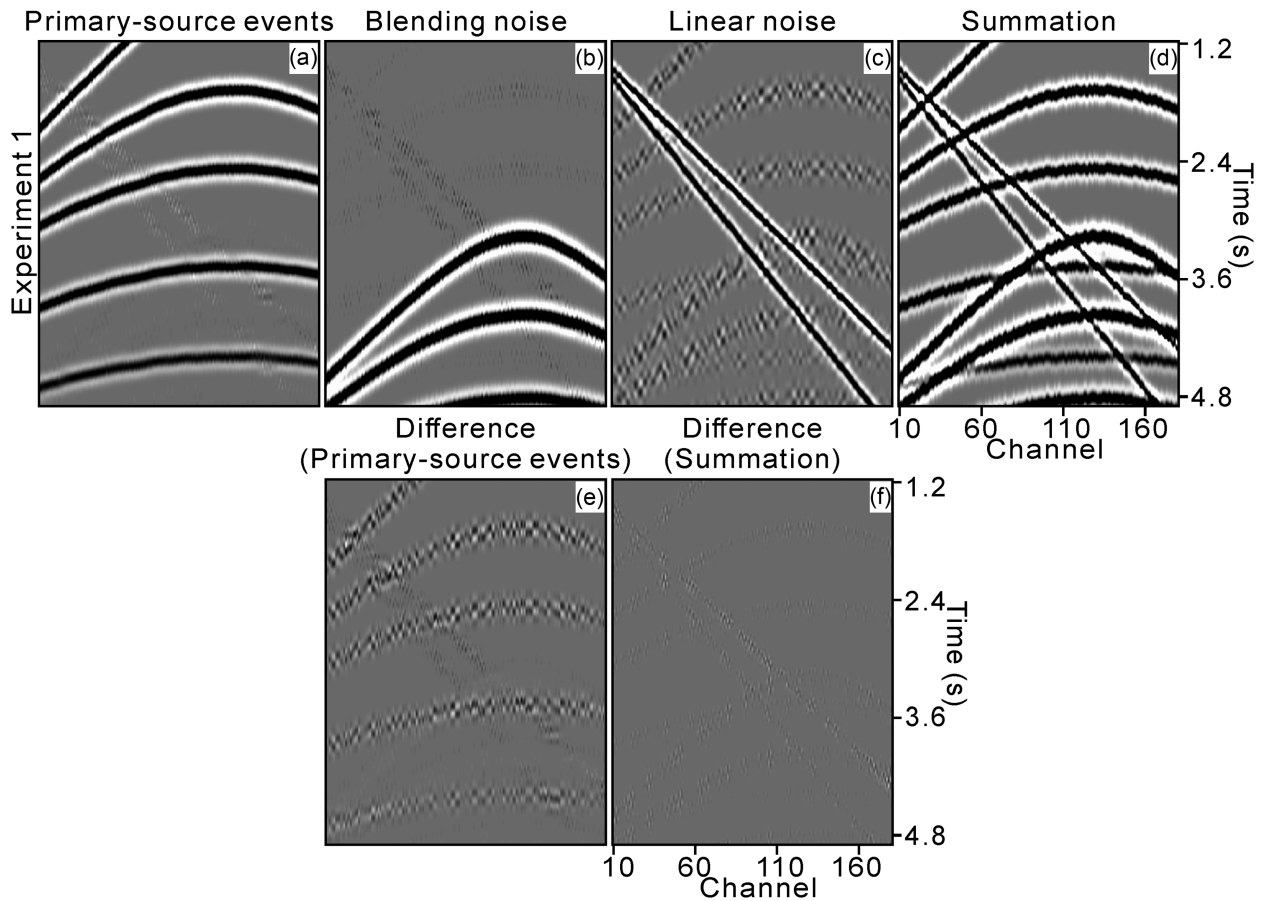
FIGURE 4 Test results of the synthetic Experiment 1. The first row from left to right shows zoomed sections of the predicted (a) desired signal, (b) blending noise, (c) linear noise and (d) the summation of the three predictions. The second row shows (e) the difference between the predicted desired signal and the true desired signal and (f) the difference between the summation of the three predictions and the true test input.

to sparsity-promoting transforms in conventional seismic processing, which reconstruct data from a sparse representation. We therefore denominate this pattern as a 'reconstruction pattern'.
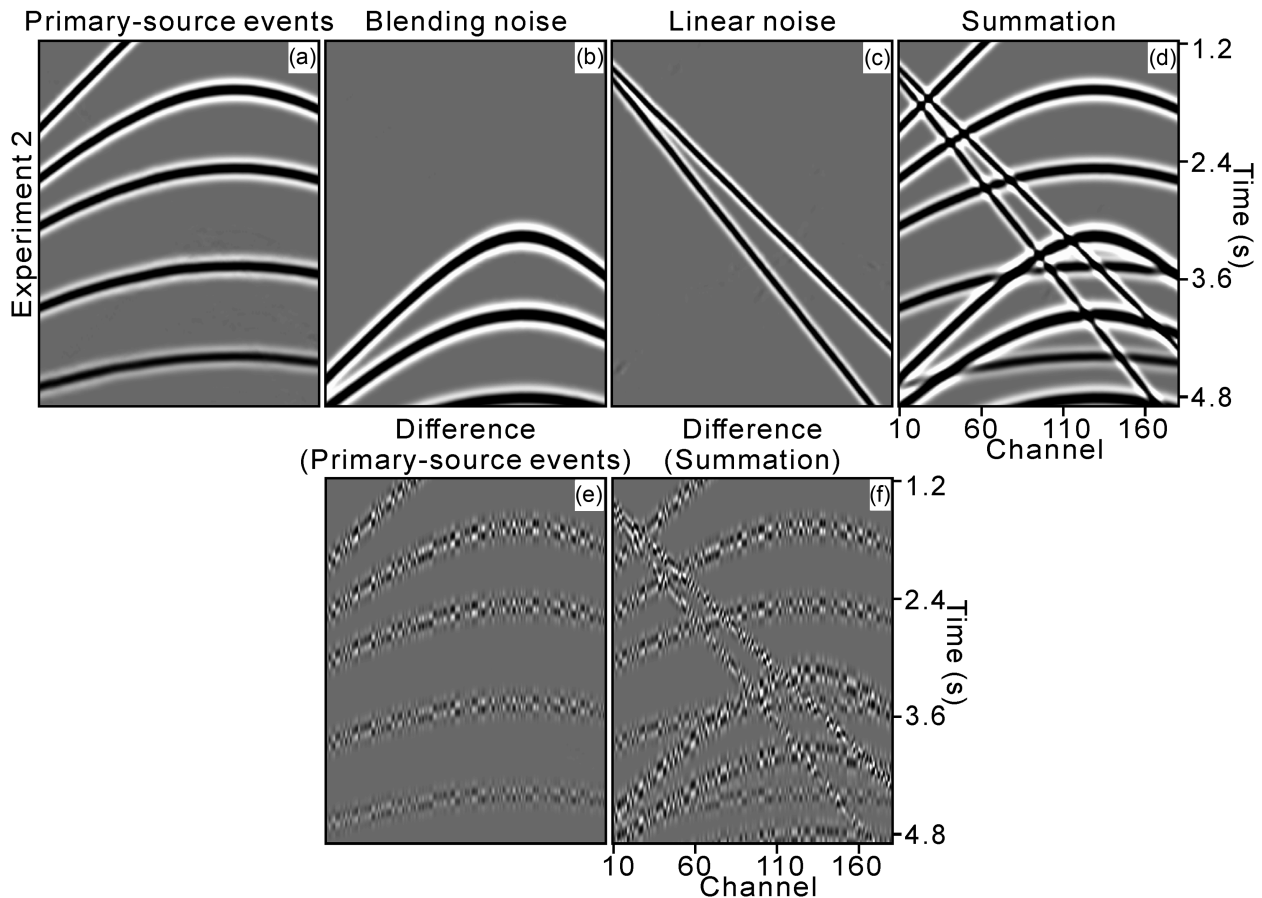
Experiment 3

Based on our observations from Experiment 2, we took the experiment one step further: In addition to injecting random noise into the training inputs, we next trained the DNN to retain the random noise along with one output channel. The intention here is to maintain the good extraction of the global features (as in Experiment 2) while avoiding the loss of data information (e.g., the local jitters on the events which disappeared in Figure 5a).

Note that real-field seismic data do not always have dramatic local features like these artificially introduced local jitters, either in the desired signal or the coherent noise, but it is still important to preserve the information from the raw data as completely as possible during the processing process.

In the specific operation of Experiment 3, we selected the channel of the desired signal as the one to retain the random

noise. This is based on the needs and experience of real-world processing projects. If our hypothesis was correct, most of the local features would be retained in the selected channel by this newly trained model, and for our processing task of SI attenuation, retaining the local features mostly in the channel of the desired signal is more in line with our consideration of signal fidelity.

Predictions from the new trained Model 3 are shown in Figure 6a–c. As we can see, Model 3 performed about as well as Model 2 in the prediction of blending noise (Figures 6b and 5b) and linear noise (Figures 6c and 5c), which should have followed the same principle of the reconstruction pattern, but performed dramatically differently in the prediction of the desired signal (Figures 6a and 5a) due to the presumed presence or absence of random noise in that channel. This is because the adoption of the reconstruction pattern requires the model to have well learned and extracted the global features of the to-be-reconstructed data. Such a requirement was always met when producing all the output channels in the case of Model 2, but not for Model 3. The lack of global features in the random noise made Model 3 struggle to reconstruct the output channel of the desired signal where the random noise

FIGURE 5 Test results of the synthetic Experiment 2. The first row from left to right shows zoomed sections of the predicted (a) desired signal, (b) blending noise, (c) linear noise and (d) the summation of the three predictions. The second row shows (e) the difference between the predicted desired signal and the true desired signal and (f) the difference between the summation of the three predictions and the true test input.

was to be retained (Figure 6a). Consequently, the DNN here deduced this output channel of the desired signal by subtracting the other two reconstructed random-noise-free channels (Figure 6b and c) from the given input (Figure 3d). Note that the DNN implicitly learned this process during the training phase; we did not code the DNN to perform any subtraction operations. We denominate this process of Model 3 for producing the channel of the desired signal as a 'subtraction pattern'.

This can explain why the local jitters on the events of the desired signal are well preserved in Figure 6a, and why the local jitters belonging to the blending noise and linear noise were also left there. A clearer observation of the latter can be found in Figure 6e, which is the difference between the true desired signal (Figure 3a) and Model 3's prediction (Figure 6a). After adding up all the output channels, we regained the local jitters on the events of all three coherent data components in Figure 6d. It hardly differs from the test input (Figure 3d), as shown in Figure 6f.

Summary of the learnt lessons

So far, we have displayed DNN's distinctive model behaviours in Experiments 1 to 3 and analysed the different processing logics behind them. The three DNN processing patterns we learnt from the above synthetic experiments can be summarized as:

1. In Experiment 1, the DNN simply decomposes the random noise-free input seismic to the different output channels (our so-called decomposition pattern). Therefore, loss of data features can be found in each of them, but their summation is approximately equal to the raw input.

2. In Experiment 2, random noise is injected into the training inputs, but we never train the DNN to predict such additional random noise into any output channel. In this case, the DNN learns to use our so-called reconstruction pattern to reconstruct all the output channels free of random noise. During this process, DNN is more focused on learning the
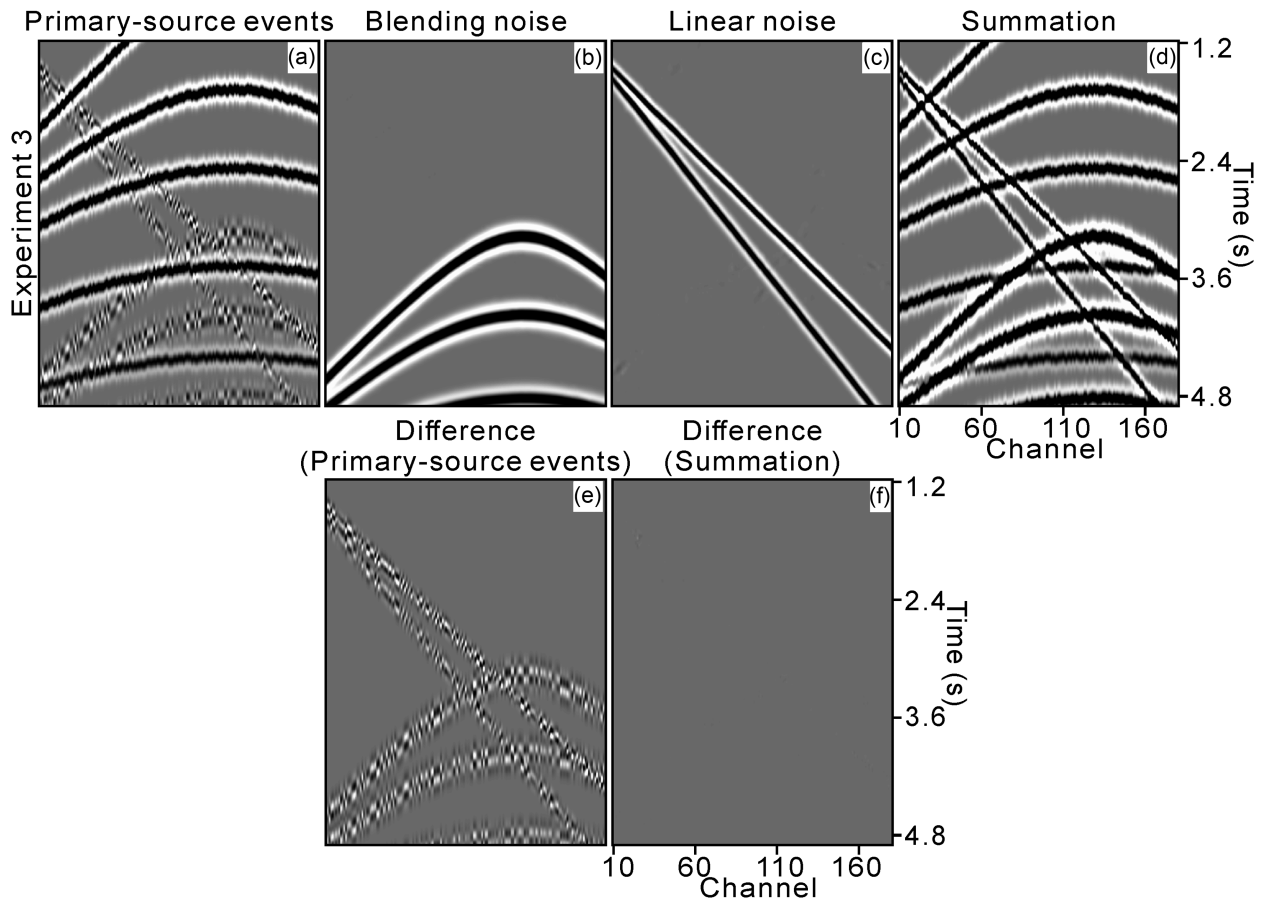
FIGURE 6 Test results of the synthetic Experiment 3. The first row from left to right shows zoomed sections of the predicted (a) desired signal, (b) blending noise, (c) linear noise and (d) the summation of the three predictions. The second row shows (e) the difference between the predicted desired signal and the true desired signal and (f) the difference between the summation of the three predictions and the true test input.

global features of data, such as the moveout of each seismic event. As a trade-off, some local features such as the local variance of amplitude and time shift would be overlooked.

3. Furthermore, from Experiment 3 we see that if one of the output channels (here, the desired signal) is expected to retain not only one coherent data component but also the additional random noise, the DNN model will tend to reconstruct the other random noise-free channels first, and then produce the selected channel injected with random noise by subtracting the reconstructed channels from the raw input. Note again that this 'subtraction pattern' behaviour of the DNN model is implicitly learnt during the training phase, but not by manual coding. In this case, not only can the coherent noise be well predicted based on the more focused learning of the global features as in Experiment 2, but also all the local features of the raw input can be preserved through the joint use of the subtraction pattern. A notable expense here is that the local jitters from the other data components are all left on the selected channel (as shown in Figure 6a) and make it look rather noisy. Despite this, extending what we learnt from Experiment 3 to real processing should not be a problem

in general, because real-field seismic data normally do not contain such dramatic local features like the local jitters we designed in these synthetic experiments.

The above three processing patterns could be selectively used for different purposes based on their own merits. For SI attenuation in the shot domain, the goal is to extract the SI noise as completely and accurately as possible while keeping the signal barely damaged. With this concern, we consider cloning the DNN's overall processing behaviour in Experiment 3 and propose it as a new method for improving the DNN's signal fidelity. A demonstration of this proposed method is given in the section to follow on shot-domain SI noise attenuation of field seismic data from Africa.

FIELD DATA DEMONSTRATION

To verify the effectiveness of the proposed method, we applied it to real-field SI-contaminated data acquired from a marine survey in Africa. As a benchmark method, we apply a conventional physics-based workflow designed for

commercial production. We also introduce a so-called reference case for our network with no additional random noise injected into the training process.

Conventional physics-based SI attenuation

To train the DNN, we manually constructed SI-contaminated data by randomly blending SI-free shots with records containing almost pure SI noise produced from a conventional method used in real production. In the conventional mean, SI noise is often attenuated based on a strategy of data resorting to obtain a more incoherent distribution of the SI noise in the common receiver or common offset domain, or the use of data transforms, such as the τ–p transform, to discriminate the noise via differences in dips/curvature from the underlying signal. The conventional physics-based method employed in this study involved preconditioning of common shot gathers (CSGs) by separation of dips of primary signal and SI noise where possible, and rank-reduction (Trickett et al., 2012) based denoising to improve the signal-to-noise ratio (S/N) of coherent events (both wanted signal and unwanted SI noise). Next, SI noise was modelled by a sparse τ–p inversion (Zhang & Wang, 2015) scheme and adapted to the input CSGs with a complex wavelet adaptation (Ventosa et al., 2012). The training and validation data sets consisted of 7000 and 1000 images, respectively. The same DNN architecture as displayed in Figure 2 was adopted again in this section of field data demonstration.

Comparison of results on signal fidelity

We first defined a reference case for our DNN, in which the above manually blended data were fed into the DNN as training (and validation) inputs, and the DNN's learning targets were made up of two channels: one was the SI-free shots, and the other was the SI noise.

Accordingly, the proposed method indicates that a DNN of the same architecture was trained from scratch with identical simulated random noise injected into the training (and validation) inputs and the target channel of SI-free shots. After being independently trained, the two DNN models were applied to the same field data contaminated by SI noise during acquisition. Their performances in removing SI noise are shown in Figure 7, followed by a comparison to the above-mentioned conventional method.

Figure 7a shows an example contaminated with two types of SI noise originating from different directions. To remove different types of SI, different sets of parameters need to be manually selected when using the conventional method. In contrast, when applying the DNN-based approaches, they can be removed simultaneously. Figure 7b and c in the second row

shows respectively the SI-attenuated shot and the extracted SI of the reference case; the third and last rows show the corresponding results of the proposed method and the conventional method, respectively. Shot gathers of large two-way traveltime (TWT) (around 6.6–8.0 s) in which the SI noise mainly exists are zoomed-in displayed in red boxes for a clearer comparison of the three adopted methods. As a QC metric, the local similarity maps (Chen & Fomel, 2015; Fomel, 2007) between the SI removal result and the SI noise that was extracted using each method are calculated and shown in blue boxes.

Comparing Figure 7d (red box v), i.e., the SI-attenuated results of the proposed method to Figure 7b (red box ii), i.e., the SI-attenuated results of the reference case, significantly less signal leakage can be observed, as indicated by the green arrows. A more obvious comparison can be found in their predicted SI, as Figure 7e to c (red boxes vii to iv). The blue boxes vi and iii show the local similarity maps of the predicted SI noise to the SI-attenuated result from the DNN with and without using the proposed method. As we can see, when introducing the proposed method, the DNN predicts SI noise that has considerably lower similarity with the SI-attenuated section than in the reference case.

The SI-attenuated results of the proposed method and the conventional method in Figure 7d and f (red boxes v and viii) look very similar at first glance, but when moving to their extracted SI, we can find that a noticeably more accurate prediction of SI was achieved by the proposed method, as indicated by the yellow arrows in red boxes vii and x of Figure 7e and g. Overall, the proposed method has shown a great improvement in DNN's signal fidelity compared with the reference case, and its results have reached a comparable level to the conventional physics-based method. This is also supported by their local similarity maps, as shown in blue boxes vi and ix.

In addition to the shot-domain example in Figure 7, the SI attenuation performances of the three approaches are also compared in the stacked-CMP domain and shown in Figure 8. As can be seen in Figure 7, SI attenuation is a more challenging issue in the sections of large TWT, such that the CMP stacks are zoomed from 5.0 s to 8.0 s in Figure 8. Figure 8a shows a CMP stack of the field SI-contaminated data from the survey in Africa. The corresponding SI removal results from the reference case, the proposed method and the conventional method are shown in Figure 8b, d and f, respectively. As observed, the difference among the three SI removal results is not very visible. A better comparison can be made by comparing the extracted SI noise from the three methods, as shown in Figure 8c, e and g.

There are apparent signal leakages that can be found in the reference case of the employed DNN, as marked in the magenta and grey ellipses in Figure 8c, and they have been significantly reduced when applying the proposed method as shown in Figure 8e. Compared with the conventional method
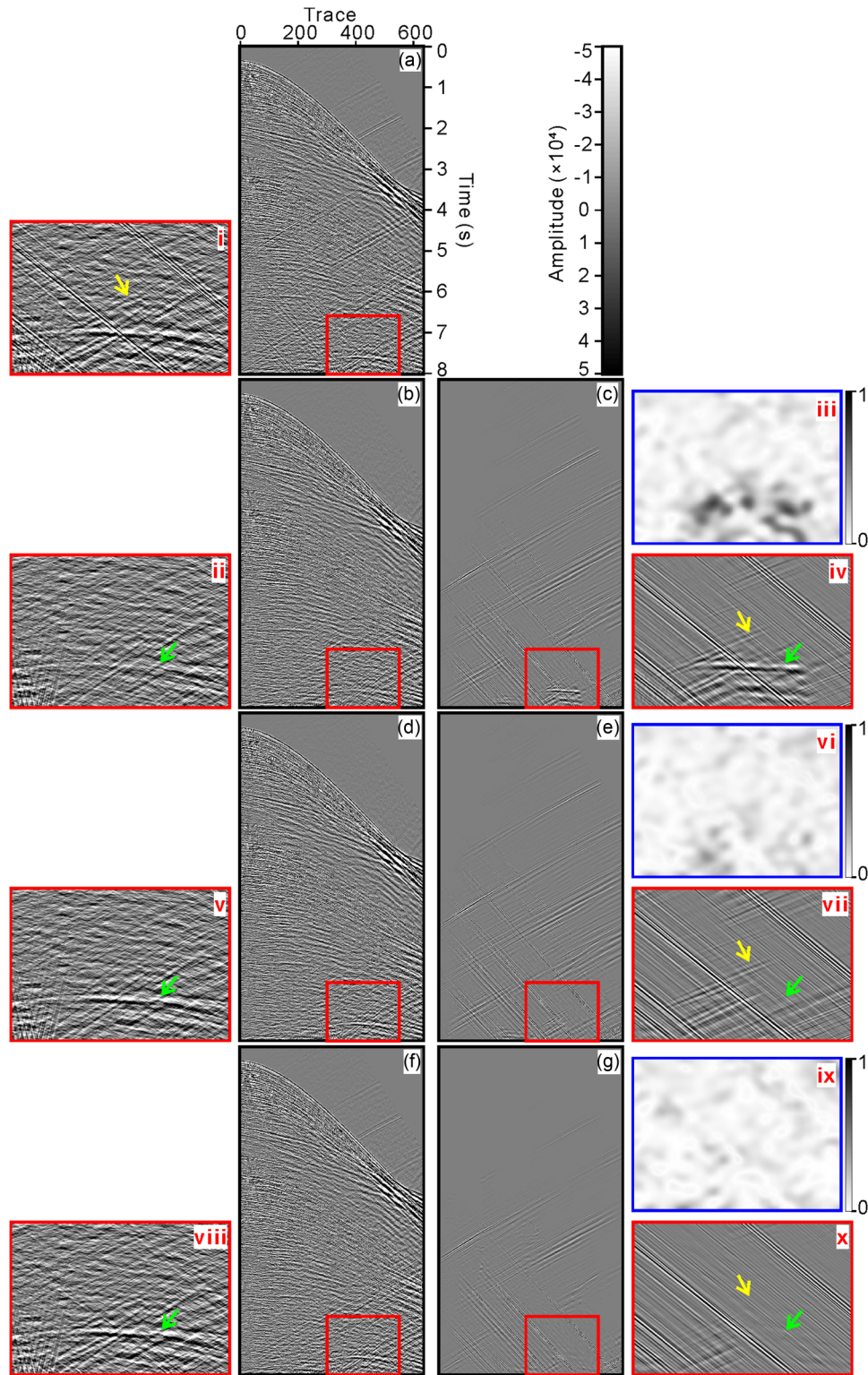
FIGURE 7 A field data example of using three different approaches to attenuate SI noise coming from two directions during acquisition: (a) is the SI-contaminated shot gather; (b) and (c) are the SI-attenuated result and the extracted SI of the reference case, respectively; correspondingly, (d) and (e) represent such outputs of the proposed method, followed by (f) and (g) of the conventional method. The red and blue boxes respectively show zoomed sections of interest and their local similarity maps: (iii) is calculated between (ii) and (iv); (vi) is calculated between (v) and (vii); (ix) is calculated between (viii) and (x).
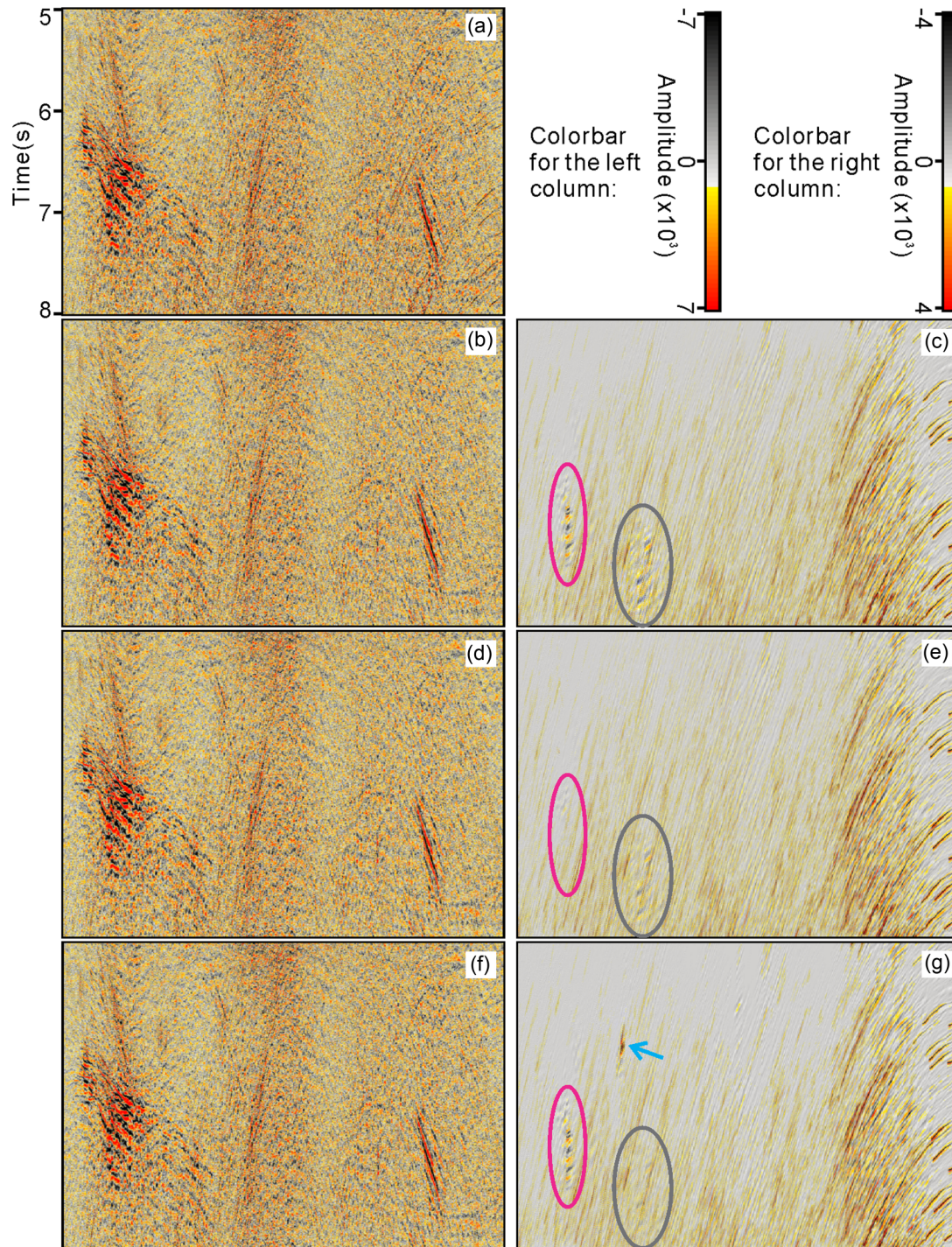
FIGURE 8 Example CMP stacks of (a) the field data contaminated by SI during acquisition; (b) and (c) are the SI-attenuated result and the extracted SI of the reference case, respectively; (d) and (e) represent such outputs of the proposed method, followed by (f) and (g) of the conventional method.

as shown in Figure 8g, the DNN using the proposed method has reached a similar level in SI attenuation accuracy and signal fidelity. Although the proposed method still does worse in the very deep section indicated by the grey ellipse, it has performed better than the conventional method in the areas marked by the magenta ellipse and the blue arrow.

CONCLUSIONS

Understanding model behaviours is important for developing DNN-based approaches for seismic processing, and this was the main motivation for this research. A DNN consists of an enormous set of parameters, which are updated automatically

during the training process. This is an advantage over the conventional methods, which usually require a laborious manual adjustment of parameters, but it also makes the DNN-based approach a 'black box'. Therefore, in our application, we proposed a novel quantitative research of the overall DNN behaviour based on synthetic data. This work enabled us to further propose a new method for improving the signal fidelity of a DNN in the separation of coherent signals, specifically, through the use of added random noise. Its effectiveness has been demonstrated on field data examples of a SI noise attenuation task. The comparison with the reference case shows that the proposed method can significantly improve the DNN's signal fidelity with an overall level of results comparable to the conventional physics-based method used in production. This verifies that the proposed method is valuable for real processing. We also hope the perspective presented in this paper can inspire new ideas for more research in the seismic community and contribute to demystifying DL.

DATA AVAILABILITY STATEMENT
Seismic data used in this paper are the property of CGG.

ORCID
Jing Sun https://orcid.org/0000-0002-4660-1106

REFERENCES
Alom, M.Z., Hasan, M., Yakopcic, C., Taha, T.M. & Asari, V.K. (2018) Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation. *arXiv*, preprint arXiv: 1802.06955.

Akbulut, K., Saeland, O., Farmer, P. & Curtis, T. (1984) Suppression of seismic interference noise on Gulf of Mexico data. In: *54th Annual International Meeting*, Atlanta, USA: Society of Exploration Geophysicists (SEG), pp. 527–529, Expanded Abstracts.

Chen, Y. & Fomel, S. (2015) Random noise attenuation using local signal-and-noise orthogonalization. *Geophysics*, 80(6), WD1–WD9.

Fomel, S. (2007) Local seismic attributes. *Geophysics*, 72(3), A29–A33.

Ghahramani, Z. (2015) Probabilistic machine learning and artificial intelligence. *Nature*, 521, 7553.

Holmstrom, L. & Koistinen, P. (1992) Using additive noise in backpropagation training. *IEEE Transactions on Neural Networks*, 3(1), 24–38.

Hou, S. & Hoeber, H. (2020) Seismic processing with deep convolutional neural networks: opportunities and challenges. In: *84th Annual International Conference and Exhibition*. European Association of Geoscientists and Engineers (EAGE), pp. 1–5, Extended Abstracts.

Hou, S. & Messud, J. (2021) Machine learning for seismic processing: The path to fulfilling promises. In *First International Meeting for Applied Geoscience and Energy*. Denver, USA: Society of Exploration Geophysicists (SEG), pp. 3204–3208, Expanded Abstracts.

Jansen, S., Elboth, T. & Sanchis, C. (2013) Two seismic interference attenuation methods based on automatic detection of seismic interference moveout. In: *75th Annual International Conference and Exhibition*. London, UK: European Association of Geoscientists and Engineers (EAGE), Extended Abstracts.

Kingma, D.P. & Ba, J. (2014) Adam: a method for stochastic optimization. *arXiv*, 1412.6980.

Nguyen, A., Yosinski, J. & Clune, J. (2015) Deep neural networks are easily fooled: high confidence predictions for unrecognizable images. In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA: Institute of Electrical and Electronics Engineers (IEEE), pp. 427–436.

Ovcharenko, O. (2021) Data-driven methods for the initialization of full-waveform inversion. PhD thesis. King Abdullah University of Science and Technology.

Qu, S., Verschuur, E., Zhang, D. & Chen, Y. (2021) Training deep networks with only synthetic data: DL-based near-offset reconstruction for closed-loop surface-related multiple estimation on shallow-water field data. *Geophysics*, 86, A39–A43.

Ronneberger, O., Fischer, P. & Brox, T. (2015) U-Net: convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Lecture Notes in Computer Science, 9351. Cham: Springer, pp. 234–241.

Sietsma, J. & Dow, R.J.F. (1991) Creating artificial neural networks that generalize. *Neural Networks*, 4(1), 67–79.

Song, C., Alkhalifah, T. & Waheed, U.B. (2021) Solving the frequency-domain acoustic VTI wave equation using physics-informed neural networks. *Geophysical Journal International*, 225, 846–859.

Sun, J., Hou, S., Vinje, V., Gordon, P. & Gelius, L.J. (2022) Deep learning-based shot-domain seismic deblending. *Geophysics*, 87(3), V215–V226.

Sun, J., Slang, S., Elboth, T., Larsen Greiner, T., McDonald, S. & Gelius, L.J. (2020a) A convolutional neural network approach to deblending seismic data. *Geophysics*, 85(4), WA13–WA26.

Sun, J., Slang, S., Elboth, T., Larsen Greiner, T., McDonald, S. & Gelius, L.J. (2020b) Attenuation of marine seismic interference noise employing a customized U-Net. *Geophysical Prospecting*, 68(3), 845–871.

Trickett, S., Burroughs, L. & Milton, A. (2012) Robust rank-reduction filtering for erratic noise. In *Technical Program*. Society of Exploration Geophysicists (SEG), pp. 1–5.

Ventosa, S., Le Roy, S., Huard, I., Pica, A. & Duval, L. (2012) Unary adaptive subtraction of joint multiple models with complex wavelet frames. In *Technical Program*, Society of Exploration Geophysicists (SEG), pp. 1–5.

Vinje, V., Lie, J.E., Danielsen, V., Dhelie, P.E., Siliqi, R., Nilsen, C.I. et al. (2017) Shooting over the streamer spread: a novel approach in seismic marine acquisition and imaging. In: *79th Annual International Conference and Exhibition*. Paris, France: European Association of Geoscientists and Engineers (EAGE), pp. 1–5.

Wu, X., Liang, L., Shi, Y. & Fomel, S. (2019) FaultSeg3D: using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. *Geophysics*, 84, IM35–IM45.

Zhang, Z.G. & Wang, P. (2015) Seismic interference noise attenuation based on sparse inversion. In *Technical Program*, Society of Exploration Geophysicists (SEG), pp. 4662–4666.