CNN-BASED ADAPTIVE SUBTRACTION FOR THE REMOVAL OF SEISMIC MULTIPLES

ZHONGXIAO LI[1, 2], FEI XIE[1, 2], JIAHUI MA[3], ZHEN QI[3] and YIBO WANG[4]

[1] *State Key Laboratory of Shale Oil and Gas Enrichment Mechanisms and Effective Development, Beijing 100083, P.R. China. thulzx@163.com*
[2] *Sinopec Key Laboratory of Seismic Elastic Wave Technology, Beijing 100083, P.R. China.*
[3] *Department of Electronic Engineering, School of Electronic Information, Qingdao University, Qingdao 266071, Shandong Province, P.R. China.*
[4] *Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, P.R. China.*

(Received November 16, 2022; revised version accepted March 10, 2023)

ABSTRACT

Li, Z.X., Xie, F., Ma, J.H., Qi, Z. and Wang, Y., 2023. CNN-based adaptive subtraction for the removal of seismic multiples. *Journal of Seismic Exploration*, 32: 169-184.

In seismic data processing primaries are usually distorted by multiples which need to be removed in advance before seismic imaging. After multiple modelling, adaptive subtraction is essential for removing multiples successfully and can be expressed as a problem of linear regression (LR) with L1 norm minimization constraint on primaries or support vector regression (SVR). Compared to the LR-based method, the SVR-based method achieves better separation of primaries and multiples since it transforms the modelled multiples nonlinearly for a better match with the true multiples in every 2D data window. However, the LR- or SVR-based method may harm primaries or cause residual multiples in complex subsurface media. In this paper a deep convolutional neural network (CNN) is constructed to better express the complicated mismatches between the modelled multiples (input data) and true multiples of the original data (label) than the LR or SVR model. To avoid overfitting to the original data and preserve primaries the L1 norm minimization constraint on primaries and L2 norm minimization constraint on CNN coefficients are used in the optimization problem. During CNN training multiple 2D data windows constructed with one or several gathers are used simultaneously to avoid overfitting. The trained CNN is used in the corresponding training data to remove multiples and then the same flowchart with CNN is used in other gathers. The proposed CNN-based method extracts high-level features of the modelled multiples to remove multiples. It is demonstrated in the synthetic and field data examples that the proposed CNN-based method can better remove multiples and preserve primaries than the LR- or SVR-based method.

KEY WORDS: adaptive subtraction, seismic multiple removal, convolutional neural network.

0963-0651/23/$5.00

INTRODUCTION

Seismic imaging with primaries is an effective tool for investigating the geological structures and the earth's subsurface (Mousa, 2014). Generally, primaries are distorted by multiples, which need to be removed in advance. Adaptive subtraction is very important for the methods with the prediction and subtraction steps to remove multiples in seismic data processing (Berkhout and Verschuur, 1997; Verschuur and Berkhout, 1997; Abma et al., 2005; Guitton, 2005). How to balance primary preservation and multiple removal is crucial for adaptive subtraction.

The commonly-used method removes the mismatches between the true and modelled multiples by using a matching filter (Guitton and Verschuur, 2004). This method can be posed as a problem of linear regression (LR) with L1 norm minimization constraint on primaries (Guitton and Verschuur, 2004; Verschuur, 2006). It can solve the LR problem in the time-space domain, Radon domain, curvelet domain and so on (Li and Li, 2017; Neelamani et al., 2010). Moreover, the method of support vector regression (SVR), which is a state-of-the-art tool of machine learning, has been introduced to express adaptive subtraction as a problem of SVR (Li, 2020). Compared to the LR-based method the SVR-based method can better separate primaries and multiples since it transforms the modelled multiples nonlinearly for a better match with the true multiples. Additionally, there are other adaptive subtraction methods, such as the pattern recognition method (Spitz, 1999; Guitton, 2005), blind source separation method (Kaplan and Innanen, 2008; Donno, 2011) and sparse coding method (Liu et al., 2017).

In this paper we treat the LR-based method with L1 norm minimization constraint on primaries and SVR-based method in the time-space domain as the benchmark for comparison. To deal with the non-stationary characteristics of the seismic data, we conduct the LR- or SVR-based method in overlapping 2D data windows. In every 2D data window an LR or SVR problem is solved, meaning that different LR or SVR problems are solved in different data windows. In this case we can not use the modelled multiples in multiple windows simultaneously for adaptive subtraction. Complicated discrepancies exist between the true and modelled multiples. The LR or SVR model has limited ability to express these discrepancies. We want to further improve the accuracy of adaptive subtraction, especially in complex subsurface media (Donno, 2011; Li, 2020).

The motivation of this paper is to introduce the convolutional neural network (CNN) into adaptive subtraction. As an advanced tool of deep learning (DL), CNN can extract high-level features in a nonlinear way from the training data and shows powerful ability in non-linear regression (LeCun et al., 2015). DL has been applied in many fields, such as the aerospace industry, biological medicine industry and remote sensing (Schmidhuber, 2015; Cheng et al., 2016; Jin et al., 2017). We construct a deep CNN to express the complicated mismatches between the true and modelled multiples. In seismic exploration CNN has been introduced in many aspects, such as seismic inversion, seismic interpretation, data denoising and data interpolation (Yuan et al., 2018; Yu et al., 2019; Gao et al., 2019; Kaur et al., 2019; Geng et al., 2020; Liu et al., 2020). Additionally, CNN and another network U-net have been used in surface related multiple elimination (Siahkoohi et al., 2019; Jiao et al., 2021) with true primaries as labels for network training. Besides, U-net has been used in adaptive subtraction for multiple removal (Bugge et al., 2021; Zhang et al., 2021) and this method also needs true primaries as labels. The lack or imbalance of labels may reduce the accuracy of multiple removal, especially in field data processing. Our proposed CNN-based method does not need true primaries in training CNN. The convolutional autoencoder (CAE) and U-net (Kumar et al., 2021; Li et al., 2021) have been used in adaptive subtraction with the modelled multiples as the input data and the original data as the label and this method is demonstrated to achieve better results than the LR-based method. Our proposed CNN-based method uses different network architecture from CAE and U-net and uses both SVR- and LR-based methods for comparison. The adaptive subtraction method in Li and Gao (2020) uses CNN to extract the features of modelled multiple, which are then subtracted adaptively from the original data with the LR-based method. This method needs to solve different 3D matching filters in every 2D data window of the original data. Therefore, it can not use the original data in multiple 2D data windows simultaneously for adaptive subtraction.

In this paper adaptive subtraction is expressed as a non-linear regression problem with CNN. We treat the modelled multiples and original data as the input data and label, respectively. During the training procedure the original data and modelled multiples, in multiple 2D data windows, are fed into the CNN simultaneously. Different from the LR- or SVR-based method, the proposed CNN-based method estimates one deep CNN for all data windows. After training, we input the modelled multiples into the trained CNN to obtain estimated multiples, which are subtracted directly from the original data to estimate primaries. The deep CNN expresses the complicated mismatches between the true and modelled multiples in a non-linear way. After the computation of convolution and non-linear activation function the deep CNN extracts high-level features of the modelled multiples, which are used to match with the original data. The proposed CNN-based method can better preserve primaries and remove multiples than the LR- or SVR-based method.

We give the following sections of the remainder of this paper. We illustrate the theory of the proposed CNN-based method, including the architecture of CNN and flowchart of the proposed CNN-based method. Then synthetic and field data examples are given to demonstrate the proposed CNN-based method. At last we give conclusions.

THEORY

Architecture of CNN

For the proposed CNN-based method the following equation is used to express the relationship between the 2D data windows $\mathbf{p}$, $\mathbf{x}$ and $\mathbf{y}$ (Yang and Ma, 2019; Li and Gao, 2020):

$$\mathbf{p}=\mathbf{y}-Net(\mathbf{x};\boldsymbol{\Theta}) \quad , \tag{1}$$

where $\boldsymbol{\Theta}$ represents the network coefficients and $Net(\bullet)$ represents CNN. $\mathbf{p}$, $\mathbf{x}$ and $\mathbf{y}$ represent the estimated primaries, modelled multiples and original data, respectively.

We define the loss function $L(\boldsymbol{\Theta})$ to train the CNN as follows (LeCun et al., 2015; Yu et al., 2019):

$$L(\boldsymbol{\Theta})=\left\|\mathbf{y}-Net(\mathbf{x};\boldsymbol{\Theta})\right\|_1+\mu\left\|\boldsymbol{\Theta}\right\|_2^2 \quad , \tag{2}$$

where μ is the regularization factor. To preserve primaries we use the L1 norm to measure the super-Gaussian distribution of primaries (Guitton and Verschuur, 2004) in eq. (2). Generally, the regularization constraint of the CNN coefficients is used to avoid overfitting during training (Yu et al., 2019; Jiao et al., 2021). L2 or L1 norm can be chosen in the regularization constraint and we choose L2 norm in eq. (2). In this paper the Adam algorithm (Kingma and Ba, 2015) is used to estimate the network coefficients $\boldsymbol{\Theta}$. And it chooses the batch size of S, which means that the 2D data windows with the number of S are fed into CNN simultaneously for training in every epoch.

Fig. 1 illustrates the architecture of CNN. Conv, ReLu and BN constitute the basic computation elements of CNN. They represent the convolution computation, Rectified Linear Units [$\max(0,\cdot)$] and batch normalization, respectively. A number of filters are used in the convolution computation to generate multiple feature windows (Zhang et al., 2017). To ensure the size of the feature window and input window is the same, we pad zeros before convolution in each layer. The BN computation refers to the normalization, the scale and shift computation (He et al., 2016). Generally, we use BN before the ReLu computation. The advantages of using BN computation are low sensitivity to initialization, good training performance and fast training. The ReLu computation is used for nonlinearity to express the complex relationship between the input and output.

In Fig. 1 we use N to denote the depth of CNN. The size of the 2D data window and feature window is chosen as $n\times n$. The layers in CNN can be divided into three types. The input-output relation of the first layer is as follows:

$$\mathbf{o}_k^1=ReLu(\mathbf{a}*\mathbf{f}_k^1+\mathbf{v}_k^1), k=1,2,\cdots,g \quad , \tag{3}$$

where the input $\mathbf{a}$ is a 2D data window, $\mathbf{f}_k^1$ represents the k-th 2D filter with size, $m \times m$. $\mathbf{v}_k^1$ is the corresponding bias, g is the total number of 2D filters and $\mathbf{o}_k^1$ is the k-th feature window. For the layers $2 \sim (N-1)$ the input-output relation is as follows:

$$\mathbf{o}_k^d = ReLu\left(BN(\mathbf{o}_k^{d-1} * \mathbf{f}_k^d + \mathbf{v}_k^d)\right),\ k = 1,2,\cdots,g,\ d = 2,3,\cdots,N-1 \quad , \tag{4}$$

where $\mathbf{f}_k^d$ is the k-th 3D filter with size $m \times m \times g$ in the d-th layer, $\mathbf{v}_k^d$ is the corresponding bias and $\mathbf{o}_k^d$ is the k-th feature window in the d-th layer. Therefore, in the layers $1 \sim (N-1)$ the output is g feature windows. For the last layer the input-output relation is as follows:

$$\mathbf{o}_1^N = \mathbf{o}_1^{N-1} * \mathbf{f}_1^N + \mathbf{v}_1^N \quad , \tag{5}$$

where $\mathbf{o}_1^N$ is the output window with size $n \times n$, $\mathbf{f}_1^N$ is the 3D filter with size $m \times m \times g$ in the N-th layer and $\mathbf{v}_1^N$ is the corresponding bias.

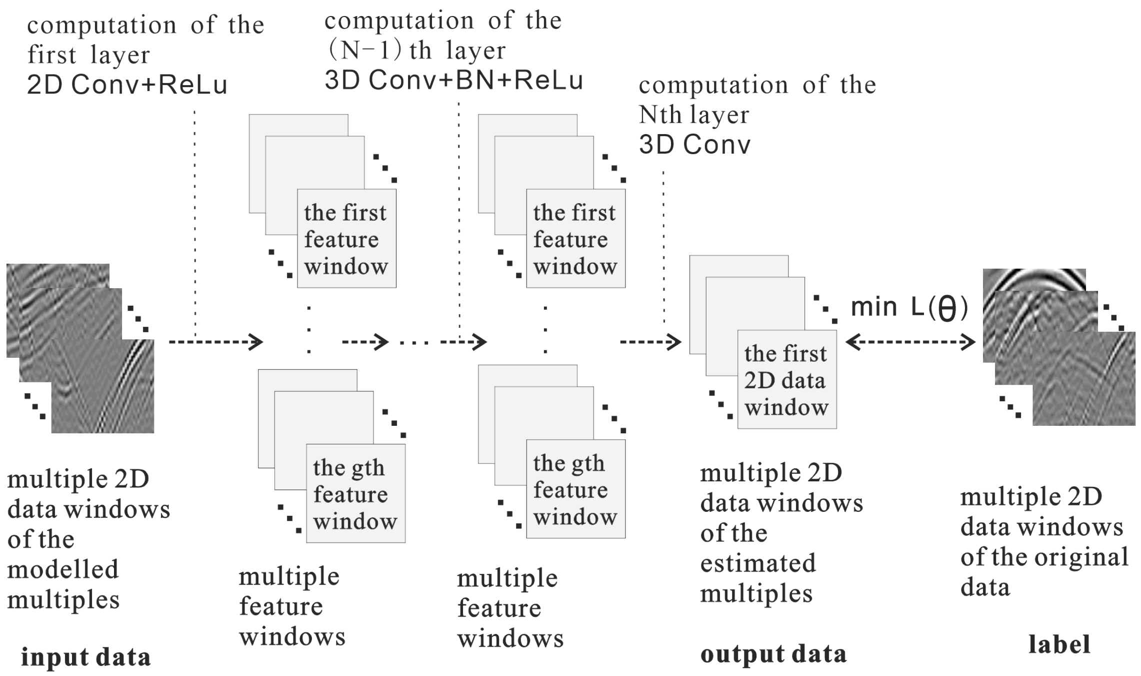

Fig. 1. Architecture of CNN. The input data, output data and label are multiple 2D data windows of the modelled multiples, estimated multiples and original data, respectively. The computation of the first layer is 2D convolution (Conv) followed by Rectified Linear Units (ReLu) to generate multiple feature windows. For layers $2 \sim (N-1)$ the computation is 3D convolution, batch normalization (BN) and Rectified Linear Units to generate multiple feature windows. The computation of the last layer is 3D convolution to give the output data. $L(\mathbf{\Theta})$ is the loss function minimized for CNN training.

Flowchart of the proposed CNN-based method

To deal with the non-stationary characteristics of the seismic data, we divide the modelled multiples and original data into 2D data windows with overlap in both temporal and spatial directions. For the SVR-based method an SVR problem is solved in every 2D data window with the feature vectors of the modelled multiples and the target values of the original data. Multiples are estimated by inputting the feature vectors of the modelled multiples into the SVR function. We estimate primaries by subtracting the estimated multiples directly from the original data.

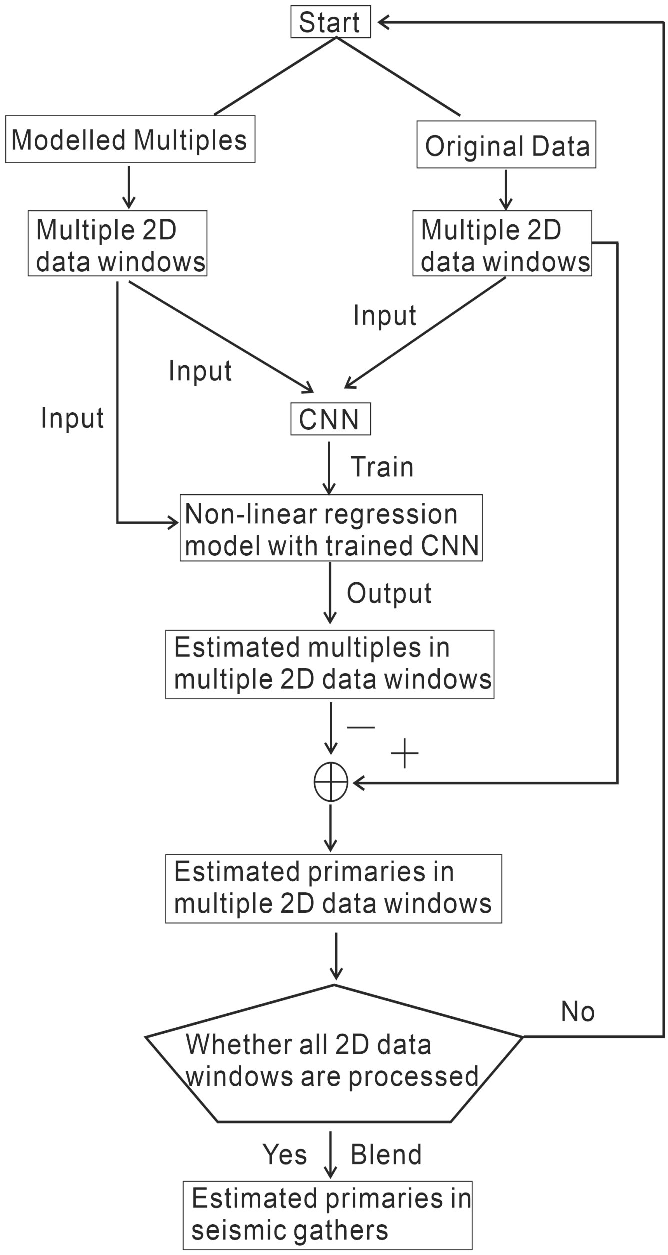

Fig. 2. Flowchart of the proposed CNN-based method.

For the proposed CNN-based method Fig. 2 shows its flowchart. Multiple 2D data windows of the original data and modelled multiples are fed into CNN for training. It is good to apply the CNN-based method on common shot gather, common offset gather or other data gathers. We complete the training by solving the optimization problem with the minimization of the loss function in eq. (2). After that, we input multiple 2D data windows of the modelled multiples into the trained CNN to estimate multiples. These modelled multiples are also used in the training stage. We obtain the estimated primaries by subtracting the estimated multiples directly from the original data. For the proposed CNN-based method we weight and blend the estimated primaries in all 2D data windows to obtain the seismic gathers as described in Li and Li (2018).

The difference between the proposed CNN-based method and the LR- or SVR-based method is summarized as follows:

1) The Mathematical Model: The proposed CNN-based method uses a non-linear regression model defined by CNN and the LR- or SVR-based method uses a LR or SVR model.

2) The Number of 2D Data Windows: The proposed CNN-based method uses multiple 2D data windows simultaneously for training CNN and the LR- or SVR-based method solves different LR or SVR problems in each 2D data window.

3) The Matching Way: After the convolution, BN and ReLu computation, the proposed CNN-based method actually uses the high-level features of the modelled multiple extracted by CNN to match with the original data. The SVR-based method uses the modelled multiples, which are obtained by transforming the modelled multiples with a nonlinear kernel function, to match with the original data. The LR-based method uses the modelled multiples themselves to match with the original data and is apt to cause residual multiples or distorted primaries. The CNN-based method can better separate primaries and multiples than the SVR-based method, which can better separate primaries and multiples than the LR-based method.

EXAMPLES

In this section the synthetic and field data are used. We compare the proposed CNN-based method with the SVR-based method (Li, 2020) and LR-based method (Guitton and Verschuur, 2004).

The synthetic data example

We use the Sigsbee2B dataset (Bishop et al., 2001) in this example. The common offset gather with primaries and multiples from 5050 ms to 7800 ms with 230 traces is shown in Fig. 3a. In this example we use the common offset gather since it can reflect the geologic structures clearly in the seismic profile. This is beneficial to judge the effectiveness of multiple removal. True primaries and multiples are available in this dataset. The corresponding modelled multiples obtained by 2D SRME (Berkhout and Verschuur, 1997; Verschuur and Berkhout, 1997) and true primaries are shown in Figs. 3b and 3c, respectively. Compared with the true multiples, the modelled multiples have temporal, spatial and wavelet difference. In Fig. 3a the black arrow indicates the strong primaries, which are surrounded by weak multiples.

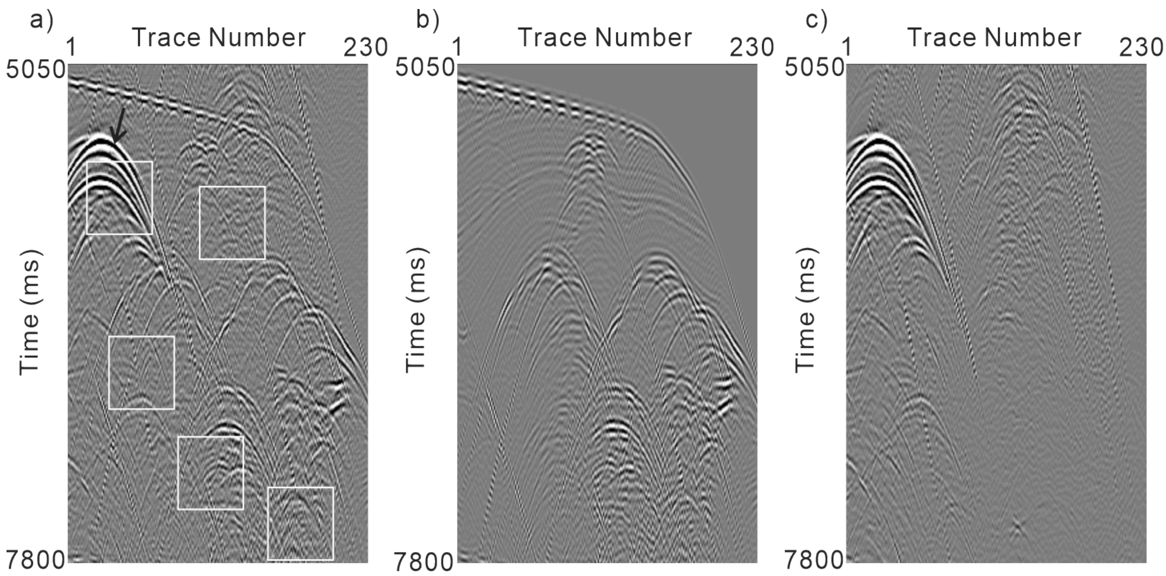

Fig. 3. The synthetic data example: (a) The original data, (b) The modelled multiples, and (c) The true primaries.

For the proposed CNN-based method we choose the CNN depth $N = 8$, the batch size $S = 5$, the epoch number 50, the 2D data window of size 50×50 and 64 filters of size 3×3 in each layer by trial and error. The total number of 2D data windows for training CNN is 1131. The Adam algorithm uses 5 data windows out of these 1131 data windows to estimate the CNN parameters at one time. In Fig. 4a and 4b we give five 2D data windows obtained from the original data and modelled multiples, respectively. The five 2D data windows in Fig. 4a correspond to the five white rectangles in Fig. 3a. The proposed CNN-based method can utilize the rich multiple information in five 2D data windows simultaneously to remove multiples. We choose the window sizes P = 52, Q = 52 and the local patch size $K = 5$, $R = 3$ by trial and error for the SVR-based method. The filter length of the LR-based method has the same meaning as the local patch size of the SVR-based method (Li, 2020). We choose the same window size and filter length of the LR-based method as the window size and local patch size of the SVR-based method to compare the performance of SVR and LR models.

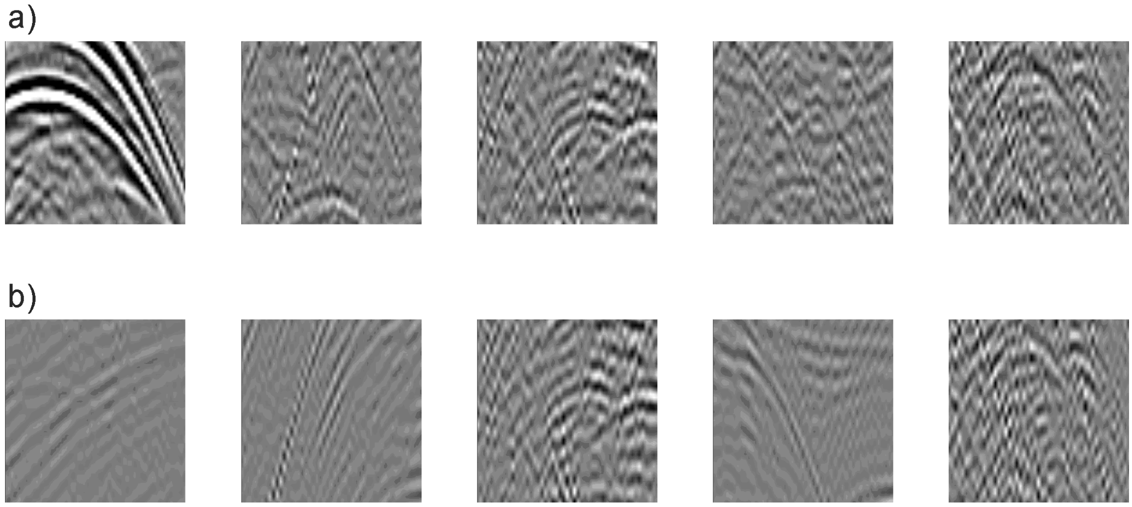

Fig. 4. The synthetic data example: (a) Five 2D data windows of the original data, and (b) Five 2D data windows of the modelled multiples.

To give a quantitative evaluation on the performance of different methods, we define the signal-to-noise ratio as

$$S / N = 10\log10\left(\|\mathbf{p}_0\|_2^2 \big/ \|\mathbf{p} - \mathbf{p}_0\|_2^2\right) \quad ,$$

where $\mathbf{p}$ denotes the estimated primaries and $\mathbf{p}_0$ denotes the true primaries.

For the proposed CNN-based method Figs. 5a and 5b shows its estimated primaries and removed multiples. For the SVR-based method Figs. 6a and 6b show its estimated primaries and removed multiples. For the LR-based method Figs. 7a and 7b show its estimated primaries and removed multiples. For the estimated primaries in Figs. 5a, 6a and 7a the S/Ns are 17.0, 13.5 and 12.7, respectively. The proposed CNN-based method obtains higher S/N than the SVR-based method, which obtains higher S/N than the LR-based method.

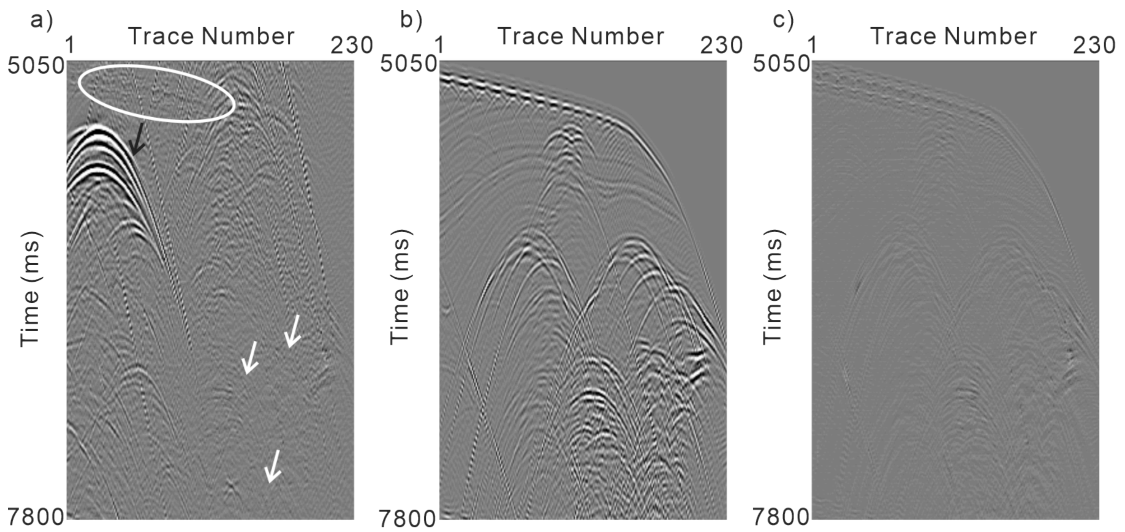

Fig. 5. The synthetic data example: (a) The estimated primaries of the proposed CNN-based method. (b) The removed multiples of the proposed CNN-based method, and (c) The difference gather by subtracting Fig. 3c from Fig. 5.

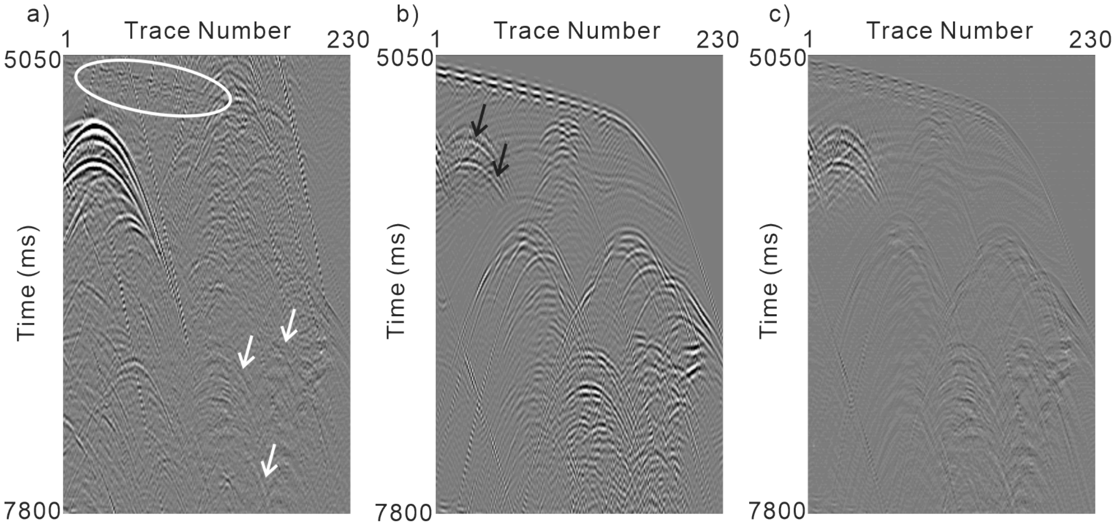

Fig. 6. The synthetic data example: (a) The estimated primaries of the SVR-based method, (b) The removed multiples of the SVR-based method, and (c) The difference gather by subtracting Fig. 3c from Fig. 6a.

By comparing the areas indicated by the white arrows in Figs. 5a, 6a and 7a we can see that the LR- or SVR-based method causes obvious residual multiples. Moreover, the LR- or SVR-based method causes some overfitting to primaries. The black arrows in Fig. 6b and 7b indicate the distorted primaries. In addition, the black arrow in Fig. 7a indicates that the LR-based method generates the false events around the strong events of primaries. Fig. 5c shows the difference gather of the proposed CNN-based method by subtracting Fig. 3c from Fig. 5a. Fig. 6c shows the difference gather of the SVR-based method by subtracting Fig. 3c from Fig. 6a. Fig. 7c shows the difference gather of the LR-based method by subtracting Fig. 3c from Fig. 7a. The difference gather, which equals to the primary error, can measure whether the multiples are effectively removed or the primaries are well preserved. It can be observed that the primary error energy of the LR-based method is larger than that of the SVR-based method, which is larger than that of the proposed CNN-based method.

The example demonstrates that using the deep CNN to express the complicated discrepancies between the true and modelled multiples with plenty of training data, the proposed CNN-based method can extract the high-level features of the modelled multiples to remove the residual multiples effectively. We use the L1 norm minimization constraint on primaries and the regularization constraint of the network coefficients in eq.(2). Therefore, the proposed CNN-based method can avoid overfitting to primaries. The strong primaries indicated by the black arrow in Fig. 5a are preserved effectively by the proposed CNN-based method. For the data window on the far left of Fig. 4, Fig. 8 shows 16 feature windows (out of 64) after the ReLu computation in the 2nd, 5th and 8th layer, respectively. The proposed CNN-based method uses these high-level features of the modelled multiples rather than the modelled multiples themselves to match with the original data. Compared to the LR- or SVR-based method, the proposed

CNN-based method preserves primaries effectively while removing multiples.

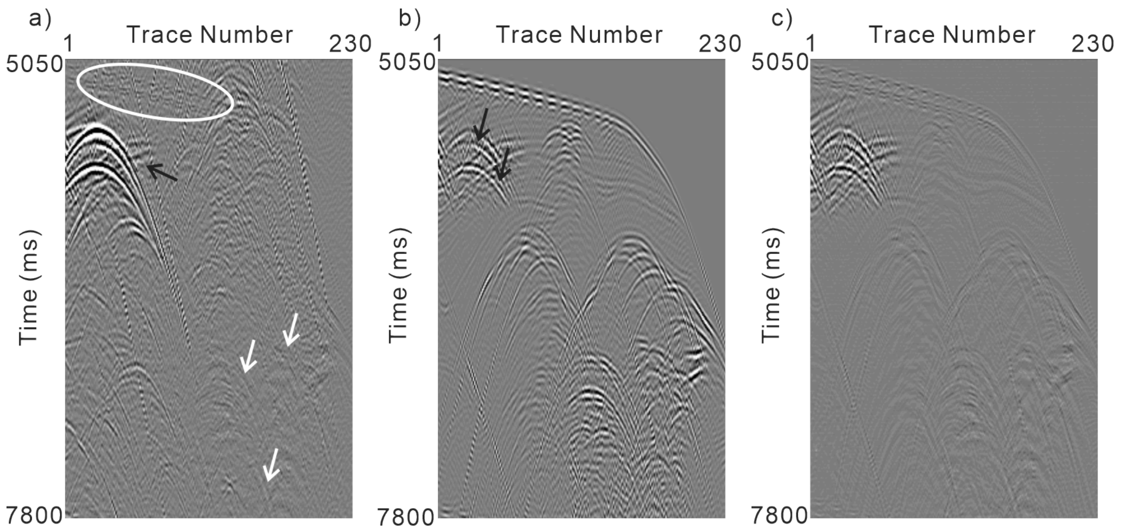

Fig. 7. The synthetic data example: (a) The estimated primaries of the LR-based method. (b) The removed multiples of the LR-based method, and (c) The difference gather by subtracting Fig. 3c from Fig. 7a.

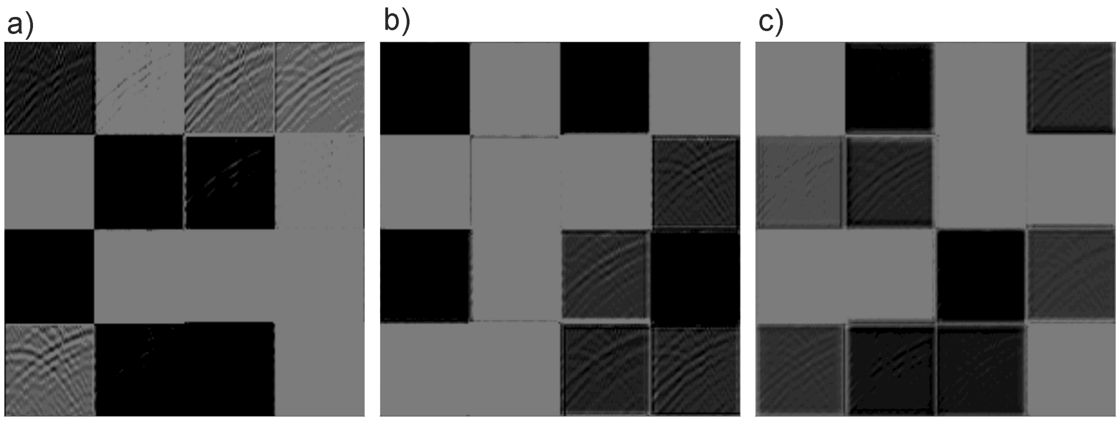

Fig. 8. The synthetic data example: (a) 16 feature windows (out of 64) after the ReLu computation in the 2nd layer, (b) 16 feature windows (out of 64) after the ReLu computation in the 5th layer, and (c) 16 feature windows (out of 64) after the ReLu computation in the 8th layer.

The field data example

In this field data example we use a common-offset gather from 3200 ms to 7998 ms with 500 traces for adaptive subtraction. The time interval is 2ms. Fig. 9a and 9b shows the original data and modelled multiples, respectively. In Fig. 9a there are weak primaries, which are covered by strong multiples.

For the proposed CNN-based method we choose the CNN depth $N = 16$, the batch size $S = 16$, the epoch number 50, the 2D data window of size

50×50 and 64 filters of size 3×3 in each layer by trial and error. The total number of 2D data windows for training CNN is 966. Fig. 10a and 10b shows the estimated primaries and removed multiples by the proposed CNN-based method. We choose the window size P = 30, Q = 30 and the local patch size K = 5, R = 5 by trial and error for the SVR-based method. Figs. 10c and 10d show the estimated primaries and removed multiples of the SVR-based method. For the LR-based method we choose the same window size and filter length as the window size and local patch size of the SVR-based method. Fig. 10e and 10f show the estimated primaries and removed multiples of the LR-based method.

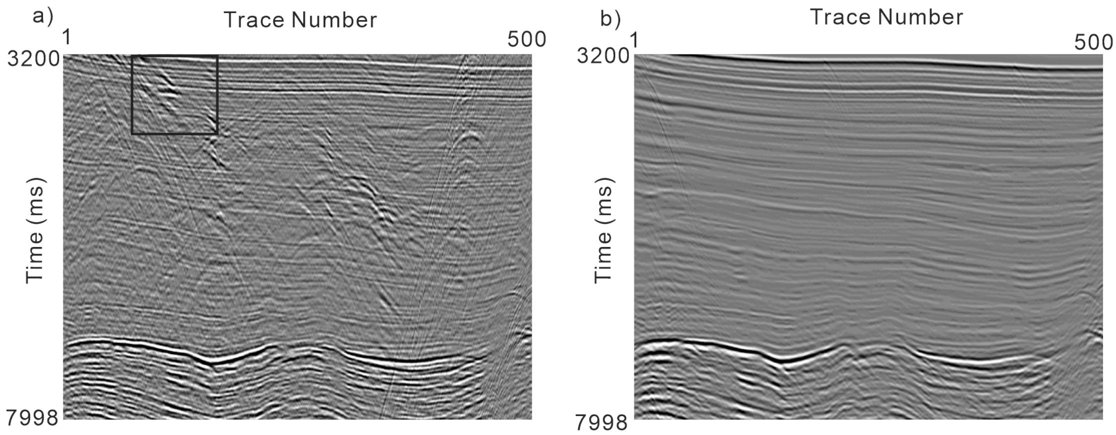

Fig. 9. The field data example: (a) The original data, (b) The modelled multiples.

In Figs. 10a, 10c and 10e the white arrows indicate that the CNN-based method removes more residual multiples than the SVR- or LR-based method. Fig. 11a shows the magnified result corresponding to the black rectangle in Fig. 9a. Fig. 11b, 11c, 11d, 11e, 11f, 11g and 11h show the magnified results corresponding to Fig. 9b, 10a, 10b, 10c, 10d, 10e and 10f, respectively. By comparing the areas indicated by the black arrows in Figs. 11c, 11e and 11g we can see that the LR- or SVR-based method removes weak primaries mistakenly. The white ellipses in Fig. 11h indicate that the LR-based method causes more distorted primaries than the SVR- or CNN-based method. In areas where strong multiples exist, the proposed CNN-based method preserves weak primaries effectively. Therefore, the proposed CNN-based method can better balance primary preservation and multiple removal than the LR- or SVR-based method.

CONCLUSION

In this paper we introduce a CNN into adaptive multiple subtraction. We construct a deep CNN to express the complicated mismatches between the true and modelled multiples. The original data and modelled multiples in multiple 2D data windows are fed into a CNN simultaneously for training. The proposed CNN-based method makes the high-level features of the modelled multiples match with the original data. In this way it expresses adaptive multiple subtraction as a problem of non-linear regression and can better balance primary preservation and multiple removal than the LR- or SVR-based method. Synthetic and field data examples are used to test the proposed CNN-based method and demonstrate its effectiveness.

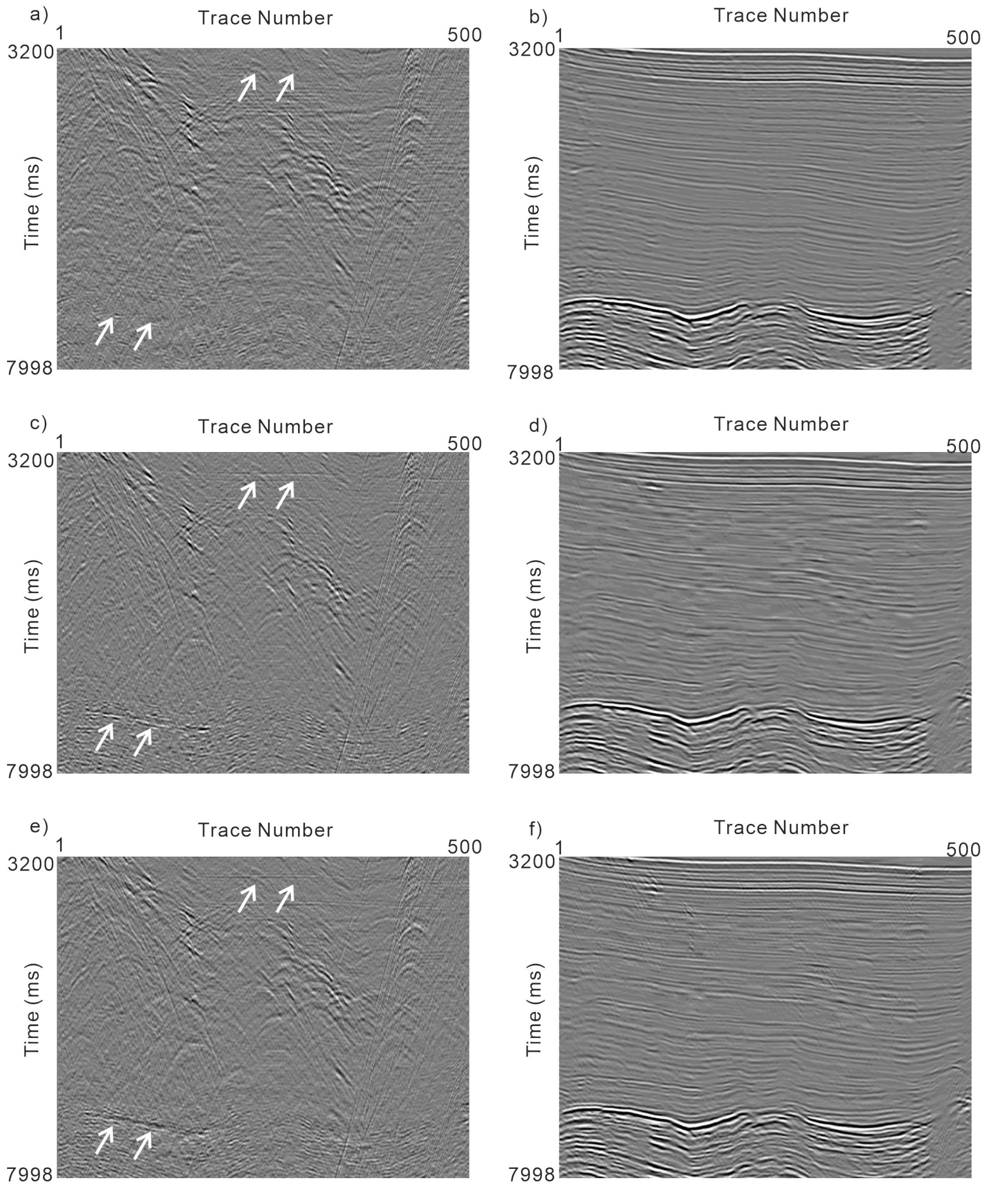

Fig. 10. The field data example: (a) The estimated primaries of the proposed CNN-based method, (b) The removed multiples of the proposed CNN-based method, (c) The estimated primaries of the SVR-based method, (d) The removed multiples of the SVR-based method, (e) The estimated primaries of the LR-based method, and (f) The removed multiples of the LR-based method.

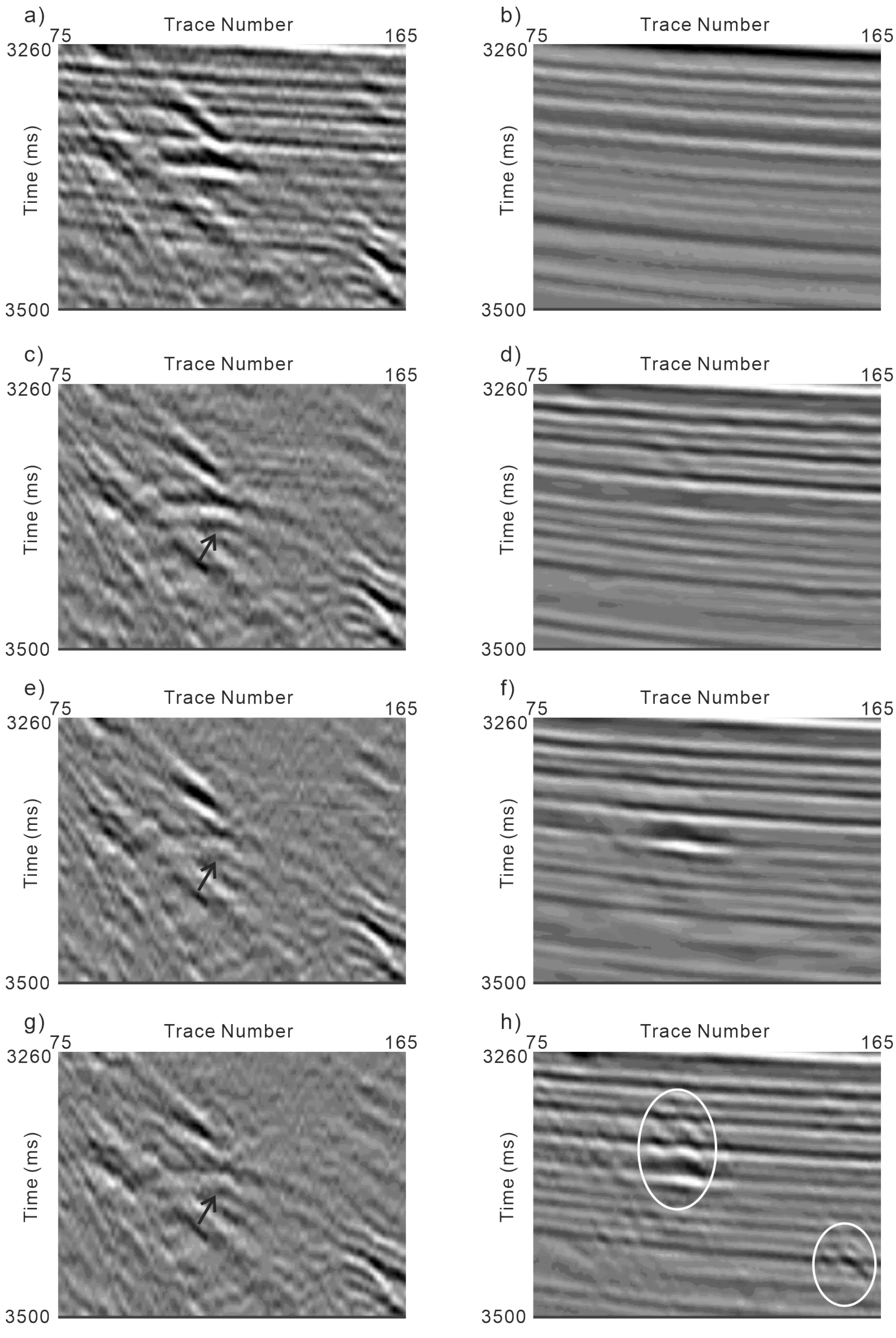

Fig. 11. The field data example: (a) The magnified result of the original data corresponding to the black rectangle in Fig. 9(a). (b) The magnified result of the modelled multiples, (c) The magnified result of the estimated primaries of the proposed CNN-based method, (d) The magnified result of the removed multiples of the proposed CNN-based method, (e) The magnified result of the estimated primaries of the SVR-based method, (f) The magnified result of the removed multiples of the SVR-based method, (g) The magnified result of the estimated primaries of the LR-based method, and (h) The magnified result of the removed multiples of the LR-based method.

ACKNOWLEDGMENTS

The authors appreciate the support of Open Fund of Sinopec Key Laboratory of Seismic Elastic Wave Technology (No. 33550000-22-ZC0613-0286), which focuses on the seismic multiple separation method based on the interpretable closed-loop optimization algorithm driven neural network. Moreover, the authors appreciate the National Natural Science Foundation of China (No. 41804110), China Postdoctoral Science Foundation (No. 2022M723127) and Open Research Foundation of SINOPEC Key Laboratory of Geophysics (No. 33550006-22-FW0399-0020).

REFERENCES

Abma, R., Kabir, N., Matson, K.H., Michell, S., Shaw, S.A. and McLain, B., 2005. Comparisons of adaptive subtraction methods for multiple attenuation. The Leading Edge, 24: 277-280.

Berkhout, A.J. and Verschuur, D.J., 1997. Estimation of multiple scattering by iterative inversion, Part I: Theoretical considerations. Geophysics, 62: 1586-1595.

Bishop, K., Keliher, J., Paffenholz, J., Stoughton, D., Michell, S., Ergas, R. and Hadidi, M., 2001. Investigation of vendor demultiple technology for complex subsalt geology. Expanded Abstr, 71 st Ann. Internat. SEG Mtg., San Antonio: 1273-1276.

Bugge, A.J., Evensen, A.K., Lie, J.E. and Nilsen, E.H., 2021. Demonstrating multiple attenuation with model-driven processing using neural networks. The Leading Edge, 40: 831-836.

Cheng, G., Zhou, P. and Han, J., 2016. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. IEEE Transact. Geosci. Remote Sens., 54: 7405-7415.

Donno, D., 2011. Improving multiple removal using least-squares dip filters and independent component analysis. Geophysics, 76(5): V91-V104.

Gao, Z., Pan, Z., Zuo, C., Gao, J. and Xu, Z., 2019. An optimized deep network representation of multimutation differential evolution and its application in seismic inversion. IEEE Transact. Geosci. Remote Sens., 57: 4720-4734.

Geng, Z., Wu, X., Shi, Y. and Fomel, S., 2020. Deep learning for relative geologic time and seismic horizons. Geophysics, 85: 1-47.

Guitton, A. and Verschuur, D.J., 2004. Adaptive subtraction of multiples using the L1-norm. Geophys.Prosp., 52: 27-38.

Guitton, A., 2005. Multiple attenuation in complex geology with a pattern-based approach. Geophysics, 70**(Issue No. ???):**: V97-V107.

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. IEEE Conf. Comput. Vis. Pattern Recognit., Las Vegas: 770-778.

Jiao, Q., Ma, J., Chi, L., Liao, Z. and Li, C., 2021. Surface-related multiple attenuation based on deep learning. 82nd EAGE Conf., Amsterdam: 1-5.

Jin, K.H., McCann, M.T., Froustey, E. and Unser, M., 2017. Deep convolutional neural network for inverse problems in imaging. IEEE Transact. Image Process., 26: 4509-4522.

Kaplan, S.T. and Innanen, K.A., 2008. Adaptive separation of free-surface multiples through independent component analysis. Geophysics, 73(4): V29-V36.

Kaur, H., Pham, N. and Fomel, S., 2019. Seismic data interpolation using CycleGAN. Expanded Abstr., 89 th Ann. Internat. SEG Mtg., San Antonio: 2202-2206.

Kingma, D.P. and Ba, J.L., 2015. Adam: A Method for stochastic optimization. Internat. Conf. Learning Representat., San Diego: 1-15.

Kumar, A., Hampson, G. and Rayment, T., 2021. Adaptive subtraction using a convolutional neural network. First Break, 39: 35-45.

LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. Nature, 521: 436-444.

Liu, D., Wang, W., Wang, X., Wang, C., Pei, J., and Chen, W., 2020. Poststack seismic data denoising based on 3-D convolutional neural network. IEEE Transact. Geosci. Remote Sens., 58: 1598-1629.

Liu, J., Lu, W., and Zhang, Y., 2017. Adaptive multiple subtraction based on sparse coding. IEEE Transactions on Geoscience and Remote Sensing, 55: 1318-1324.

Li, Z.-X., 2020. Adaptive multiple subtraction based on support vector regression. Geophysics, 85(3): V57-V69.

Li, Z. and Gao, H., 2020. Feature extraction based on the convolutional neural network for adaptive multiple subtraction. Marine Geophys. Res., 41: 10.

Li, Z.-X. and Li, Z.-C., 2018. Accelerated 3D blind separation of convolved mixtures based on the fast iterative shrinkage thresholding algorithm for adaptive multiple subtraction. Geophysics, 83(1): V99-V113.

Li, Z.-X. and Li, Z.-C., 2017. Accelerated parabolic Radon domain 2D adaptive multiple subtraction with fast iterative shrinkage thresholding algorithm and its application in parabolic Radon domain hybrid demultiple method. J. Appl. Geophys., 143: 86-102.

Li, Z., Sun, N., Gao, H., Qin, N. and Li, Z., 2021. Adaptive subtraction based on U-net for removing seismic multiples. IEEE Transact. Geosci. Remote Sens., 59: 9796-9812.

Mousa, W.A., 2014. Imaging of the SEG/EAGE salt model seismic data using sparse f-x finite-impulse-response wavefield extrapolation filters. IEEE Transact. Geosci. Remote Sens., 52: 2700-2714.

Neelamani, R.N., Baumstein, A. and Ross, W.S., 2010. Adaptive subtraction using complex-valued curvelet transforms. Geophysics, 75(2): V51-V60.

Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural Netw., 61: 85-117.

Siahkoohi, A., Verschuur, D.J. and Herrmann, F.J., 2019. Surface-related multiple elimination with deep learning. Expanded Abstr., 89th Ann. Internat. SEG Mtg., San Antonio: 4629-4634.

Spitz, S., 1999. Pattern recognition, spatial predictability, and subtraction of multiple events. Geophysics, 18: 55-58.

Verschuur, D.J. and Berkhout, A.J., 1997. Estimation of multiple scattering by iterative inversion; Part II, Practical aspects and examples. Geophysics, 62: 1596-1611.

Verschuur, D.J., 2006. Seismic Multiple Removal Techniques: Past, Present and Future. EAGE, Houten.

Yang, F. and Ma, J., 2019. Deep-learning inversion: A next-generation seismic velocity model building method. Geophysics, 84(4): R585-R584.

Yuan, S., Liu, J., Wang, S., Wang, T. and Shi, P., 2018. Seismic waveform classification and first-break picking using convolution neural networks. IEEE Geosci. Remote Sens. Lett., 15: 272-276.

Yu, S., Ma, J. and Wang, W., 2019. Deep learning for denoising. Geophysics, 84(6): V333-V350.

Zhang, D., Leeuw, M. and Verschuur, E., 2021. Deep learning-based seismic surface-related multiple adaptive subtraction with synthetic primary labels. Expanded Abstr., First Internat. Mtg. Appl. Geosci. Energy, Denver: 2844-2848.

Zhang, K., Zuo, W., Chen, Y., Meng, D. and Zhang, L., 2017. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Transact. Image Process., 26: 3142-3155.