Random noise suppression in seismic data: what can deep learning do?

Dawei Liu,Wei Wang, Wenchao Chen, Xiaokai Wang, Yanhui Zhou, Zhensheng Shi，Xi'an Jiaotong University, Xi'an, China*

Summary

In the past few years, deep learning has gained great success in image signal and information processing. What are the challenges of seismic denoising compared to image denoising when using deep learning? First, the clean training seismic data are usually unavailable. Secondly, most of the networks are set for two-dimensional (2D) image denoising. However, seismic data denoising is mainly aiming at three-dimensional (3D) or higher dimensional data. Whether it works when expanding the network from 2D to 3D or higher dimension for seismic denoising? Finally, the most importance thing is to learn what aspects has been enhanced from deep learning compared to conventional methods for seismic denoising. This abstract takes random noise suppression in 3D post-stack seismic data processing as an example to discuss the application of deep learning in seismic denoising. Except for white noise, there are imaging noise and scattered noise due to near surface non-uniformity in seismic data. This means 3D seismic random noise is usually non-Gaussian. According to the features of seismic noise, we propose 3D denoising convolutional neural networks (3D-DnCnn) incorporating sample screening. The seismic data acquired in a survey covering 700 square kilometers are used in our test. A small part of data with high signal-to-noise ratio (SNR) is selected for training. Other data are used for testing. The test results show that the network has similar random noise attenuation performance to conventional methods. Moreover, it can extract features of noise from the overall situation and thus has better suppressing performance of imaging noise. In addition, we adopt residual learning and batch normalization for accelerating training speed. And after network training is satisfactorily completed, its processing efficiency can be significantly faster than conventional denoising methods.

Introduction

Random noise in seismic data, which has no definite frequency and apparent velocity, is usually the irregular interference caused by multiple factors. It leads to the low SNR and resolution of seismic data, and further cause unreliable seismic inversion results. Thus, suppressing random noise is quite essential in seismic processing.

Many advanced signal processing techniques have been used for suppressing random noise during the past few decades. Conventional methods include wavelet transform (Wu and Liu, 2008), $f-x$ prediction filtering (Harris and White, 1997), Karhunen-Loeve (KL) transform (Jones and Levy, 1987) and singular vector decomposition (SVD) decomposition (Bekara and Baan, 2007). However, they have some disadvantages. One common disadvantage of the above techniques is that they make use of the spatial coherence of seismic data and enhance the SNR at the expense of horizontal resolution which might easily attenuate the tilt and bending events. Besides, the median filters (Bednar, 1983; Duncan and Beresford, 1995) which were especially suitable for eliminating peak noise in non-stationary signals were applied to the removal of noise from seismic data and it achieved superior random noise attenuation and signal fidelity than previous processing methods. Many scholars also proposed many kinds of methods to improve the median filter (Liu et al, 2006; Liu et al, 2006; Wang et al, 2012). Wang et al. (2017) proposed a three-step denoising workflow incorporating Cadzow filter (Trickett, 2008) and edge preserving filter (Al-Dossary et al., 2002) for random noise suppression.

In recent years, deep learning has developed rapidly in the field of image denoising due to its favorable denoising performance. Jain and Seung (2009) used convolutional neural networks (CNNs) to suppress image noise. Burger et al. (2012) applied multi-layer perceptron to suppress image noise successfully. Xie et al. (2012) adopted stacked auto-encoders to remove Gaussian noise and achieved similar results compared to that of K-SVD (Elad and Aharon, 2006). Chen and Pock (2017) proposed a trainable nonlinear reaction diffusion model to handle Gaussian noise removal. Zhang et al. (2017) developed denoising convolutional neural networks (DnCNNs) for general image denoising. The deep learning based methods mentioned above can achieve promising performance in the scenario of Gaussian noise. However, the performance will become poor if the noise is non-Gaussian. Deep learning has gained tremendous attention in building hierarchical representation and has also been applied to the field of geophysics. Hami-Eddine et al. (2017) applied neural networks to accelerate seismic interpretation. DeVries et al. (2017) used deep neural network to accelerate viscoelastic calculations with speedups of at least ~55,000%.

The selection of training samples used in seismic denoising is very crucial. In the absence of clean data, we can use as advanced methods as possible to suppress the noise of the noisy seismic data in order to create desired samples. In addition, there are some approaches to obtain high SNR training samples. For instance, high-density exploration technology is carried out in some local regions and hence high SNR seismic data can be acquired. In addition, surface conditions in some regions are conducive to seismic exploration, and high SNR data are also obtained. In combination with these favorable factors, we carefully select a small region of high SNR from the whole seismic exploration region. Then, the most effective denoising method in the industry today is applied to suppress noise in the small region. Thus, we get the comparatively ideal training sample pairs. Generally, in a large survey region, the underground structures have similarities, which means that the structure of seismic imaging useful signals is similar. Further, the noise in one exploration region is also similar. Deep learning can capture the features of the noise in the complete exploration region from a small high SNR region and use it to suppress noise in the other regions.

In this abstract, a 3D denoising convolutional neural network (3D-DnCnn) used to suppress random noise in seismic data is proposed. First, a small part seismic data from a survey are processed by a three-step denoising workflow (Wang et al., 2017). The filtered data have high SNR and training sample pairs are selected from them. Then, 3D-DnCnn is iteratively trained so that features of random noise can be obtained with the residual learning. Finally, we utilize the well-trained 3D-DnCnn to suppress the random noise contained in test data to demonstrate the fidelity of our method.

Method

In this section, we present a 3D denoising CNN method model different from conventional neural network based algorithms. One aspect is network architecture. Here, we extend the 2D DnCnn to 3D model to make it adapt to seismic data denoising. The other is model learning from training data.

Network architecture

The input of our 3D-DnCnn is the original 3D seismic data volume $\mathbf{y}=\mathbf{x}+\mathbf{v}$, where $\mathbf{x}$ is the useful signal and $\mathbf{v}$ is the noise. We employ the residual learning method to train a residual mapping $\Re(\mathbf{y}) \approx \mathbf{v}$ which is regarded as noise. And then, we can obtain the output $\mathbf{x}=\mathbf{y}-\Re(\mathbf{y})$, which is taken as denoised useful data. The averaged mean squared error between the desired denoised seismic data and estimated ones from original noisy seismic data is set as the loss function. The corresponding formula can be written as:

$$l(\mathbf{\Theta})=\frac{1}{2N}\sum_{i-1}^{N}\left\|\Re(\mathbf{y}_i;\mathbf{\Theta})-(\mathbf{y}_i-\mathbf{x}_i)\right\|_F^2 \qquad (1)$$

A locally-optimal of the trainable parameters $\mathbf{\Theta}$ can be found by the back-propagation method (Rumelhart et al., 1986) after iterative training. $\{(\mathbf{y}_i,\mathbf{x}_i)\}_{i=1}^{N}$ represents N noisy-clean training sample pairs.

According to the empirical parameter setting in VGG-net (Simonyan and Zisserman, 2014), we choose the size of convolutional filters to be $3\times3\times3$ and remove all pooling layers to make sure the network input and output have the same size. Therefore, we can obtain the result of the receptive field of 3D-DnCnn with depth of d by $(2d+1)\times(2d+1)\times(2d+1)$. The receptive field size which capture information for denoising has strong relationship with the effective patch size. By comparing different denoising method, we set the receptive field size to $35\times35\times35$ with network depth of 17 while patch size is cube with length 40~60 so that more context information for denoising can be captured. Fig.1 shows the specific architecture of the proposed 3D-DnCNN. Latent clean data can be removed in the hidden layers and only noise remains.

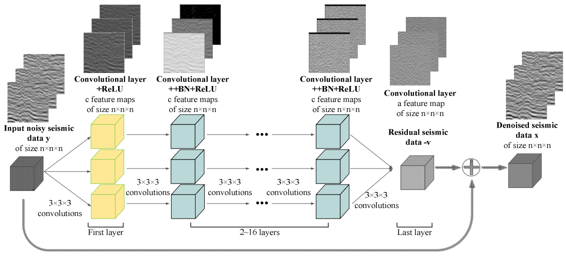

Fig.1. Architecture of the proposed 3D-DnCnn network for seismic data denoising

It is claimed that the size of output data should keep the same as the input one in seismic data processing, but this may lead to boundary artifacts. Padding strategy may reduce boundary artifacts and is also an important part of network architecture. We overcome the above problems by directly padding zeros before convolution. The field data examples show there does not exist any boundary artifacts by simple zero padding strategy.

Model learning

A small part of data is selected for training. If directly fed by the whole small region, the network will learn bad features from poorly denoised structure. This reduces the denoising effect of the network. In order to provide reliable samples, we calculate the similarity based on gradient structure tensor (Bakker, 2002) and choose the samples with high similarity.

Residual learning formulation and batch normalization are adopted for accelerating training convergence and improving denoising performance in our method. Using the original mapping $F(\mathbf{y})$ to predict $\mathbf{x}$ and using the residual mapping $\Re(\mathbf{y})$ to predict $\mathbf{v}$ are two forms to asymptotically approximate the desired functions. However, since learning perturbations with reference is much easier

10.1190/segam2018-2998114.1

than learning the function as a new one, residual learning are easier to be optimized when the difference between the input data and output data is small (He et al., 2016). The amplitude of random noise in seismic data is much smaller than useful data, so we choose residual learning formulation to learn the features of random noise.

It has been proved that mini-batch stochastic gradient descent (SGD) is an effective way to train deep networks. With mini-batch SGD, the parameters

$$\mathbf{\Theta}=\arg\min_{\mathbf{\Theta}}\frac{1}{N}\sum_{i=1}^{N}l\left(\mathbf{y}_i,\mathbf{\Theta}\right) \tag{2}$$

can be optimized in steps, at each step training a mini-batch $\mathrm{x}_{1\ldots m}$ of size m, where $\mathrm{x}_{1\ldots N}$ is the training data set. The parameters $\mathbf{\Theta}$ can be updated according to gradient descent:

$$\mathbf{\Theta}^{'}=\mathbf{\Theta}-\frac{\alpha}{m}\sum_{i=1}^{m}\frac{\partial l\left(\mathbf{y}_i,\mathbf{\Theta}\right)}{\partial\mathbf{\Theta}} \tag{3}$$

Although mini-batch SGD is simple and effective, its training is complicated on account of internal covariate shift, i.e., the changes in the distributions of internal inputs to each layer during training. The internal covariate shift can be reducing by batch normalization which inserts a normalization step

$$\hat{\mathbf{y}}^{(k)}=\frac{\mathbf{y}^{(k)}-E[\mathbf{y}^{(k)}]}{\sqrt{Var\left[\mathbf{y}^{(k)}\right]+\varepsilon}}, \tag{4}$$

and a scale and shift step

$$\mathbf{Y}^{(k)}=\gamma\hat{\mathbf{y}}^{(k)}+\beta \tag{5}$$

before ReLU with m-diminsional input $\mathbf{y}=\left(\mathbf{y}^{(1)}\ldots\mathbf{y}^{(m)}\right)$. In the above formula, the only two extra parameters $\gamma=\sqrt{\mathrm{Var}\left[\mathbf{y}^{(k)}\right]}$ and $\beta=\mathrm{E}\left[\mathbf{y}^{(k)}\right]$ per activation added by batch normalization can be updated with back-propagation. Batch normalization can accelerate training speed, get better performance and make networks less sensitive to initialization. Thus, we integrate residual learning and batch normalization to get better denoising performance.

Examples

Fig.2 shows one time slice from a post-stack noisy oilfield seismic data. As is shown by a red rectangle in Fig.2, the noisy seismic data $\mathbf{y}$ with the size of 66 square kilometers is selected for training. $\mathbf{x}$ is filtered by a three-step denoising workflow (Wang et al., 2017) from $\mathbf{y}$. The training sample $\left\{(\mathbf{y},\mathbf{x})\right\}$ is fed to 3D-DnCnn. The training parameters are shown as follows. The stride is chosen as 10 and thus the number of training sample pairs is about 100,000 after screening. The mini-batch size is set as 6. The 15 epochs are setting. The original field seismic data from the rest about 640 square kilometers regions are used in test. After well training, the whole exploration region can be denoised by 2 M4000 GPU within 2 hours, which is 10 times faster than the conventional method we used (Wang et al., 2017).

Fig.3 shows the zooming noise profiles from crossline and inline respectively. The yellow lines drawn in Fig.2 indicate the locations of the crossline and inline. It can be seen from Fig.3 that both the conventional method and our proposed method have suppressed random noise obviously. In other words, the network successfully learns the conventional denoising method with the given samples. In addition, our method get better suppression result on imaging noise as shown by the black rectangle boxes. From the loss function in equation (1), we can see that the denoising method based on deep learning adopts a global optimization strategy, which makes the network learn the noise structure from the whole 3D seismic data. On the contrary, conventional methods are generally based on local data. This is the reason why the deep learning method can significantly suppress imaging noise. Based on this, deep learning based methods are expected to have superior suppression performance for surface scattered noise compared to conventional methods.

Conclusions

In this abstract, we propose a random noise removal strategy in seismic data based on deep learning. Random noise in seismic data is usually non-Gaussian and possesses distinct spatial structural characteristics in 3D. We apply an excellent performing conventional method combined sample screening to get the comparatively ideal training sample pairs. The features of noise can be learned from overall situation by 3D-DnCnn with the given samples. The denoising results of field seismic data demonstrate that our method can remove most of noise with reserving the desired signals and suppressing considerable imaging noise which conventional methods are hard to remove. Moreover, similar to image denoising, the denoising performance of our network will also be improved with better sampling strategy. We adopt batch normalization and residual learning to speed up the convergence of mini-batch SGD and boost the denoising performance of our model.

Acknowledgments

The research was funded by National Natural Science Foundation of China (41774135, 41504092, 41504093), China Postdoctoral Science Foundation (2016T90925, 2015M572567), National Key Research and Development Program of China (2017YFB0202902), and the Fundamental Research Funds for the Central Universities. We would like to thank Daqing oilfield for providing the dataset.

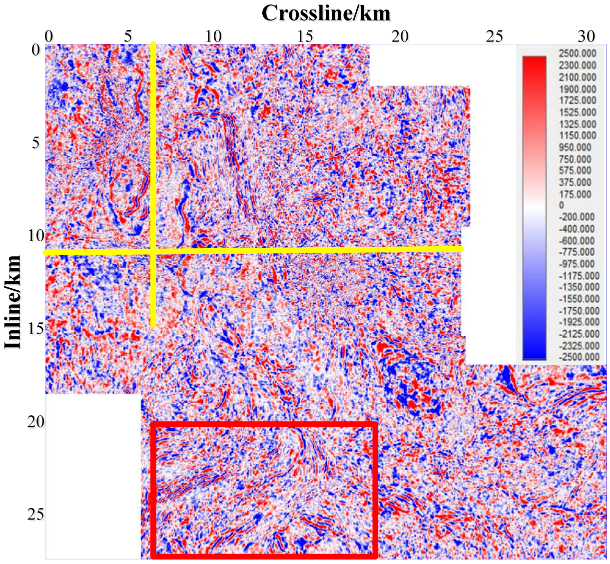

Fig.2. A time slice from a post-stack noisy oilfield seismic data.

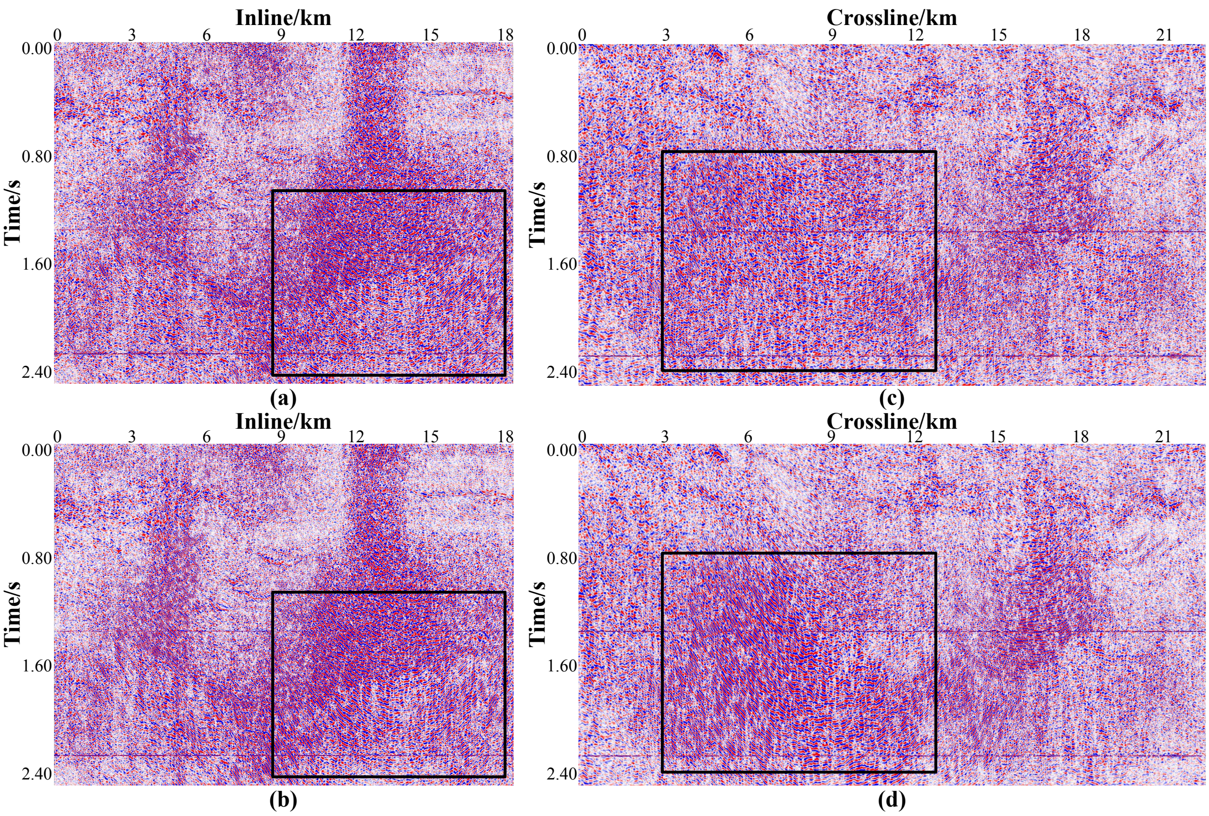

Fig.3. The profile of removed noise. (a). Removed noise by conventional methods at crossline=6.5km. (b). Removed noise by our method at crossline=6.5km. (c). Removed noise by conventional methods at inline=11km. (d). Removed noise by our method at inline=11km.

REFERENCES

Al-Dossary, S., K. J. Marfurt, and Y. Luo, 2002, 3D edge preserving smoothing for seismic edge detection: 72nd Annual International Meeting, SEG, Expanded Abstracts, 2478–2481, https://doi.org/10.1190/1.1817300.

Bakker, P., 2002, Image structure analysis for seismic interpretation: Delft University of Technology.

Bednar, J. B., 1983, Applications of median filtering to deconvolution, pulse estimation, and statistical editing of seismic data: Geophysics, **48**, 1598–1610, https://doi.org/10.1190/1.1441442.

Bekara, M., and M. Van der Baan, 2007, Local singular value decomposition for signal enhancement of seismic data: Geophysics, **72**, no. 2, V59–V65, https://doi.org/10.1190/1.2435967.

Burger, H. C., C. J. Schuler, and S. Harmeling, 2012, Image denoising. Can plain neural networks compete with BM3D?: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2392–2399.

Chen, Y., and T. Pock, 2017, Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration: IEEE Transactions on Pattern Analysis and Machine Intelligence, **39**, 1256–1272, https://doi.org/10.1109/TPAMI.2016.2596743.

DeVries, P. M., T. B. Thompson, and B. J. Meade, 2017, Enabling large-scale viscoelastic calculations via neural network acceleration: Geophysical Research Letters, **44**, 2662–2669, https://doi.org/10.1002/2017GL072716.

Duncan, G., and G. Beresford, 1995, Some analyses of 2-D median fk filters: Geophysics, **60**, 1157–1168, https://doi.org/10.1190/1.1443844.

Elad, M., and M. Aharon, 2006, Image denoising via sparse and redundant representations over learned dictionaries: IEEE Transactions on Image Processing, **15**, 3736–3745, https://doi.org/10.1109/TIP.2006.881969.

Hami-Eddine, K., B. de Ribet, P. Durand, and P. Gascue, 2017, A growing machine learning approach to optimize use of prestack and poststack seismic data: 87th Annual International Meeting, SEG, Expanded Abstracts, 2149–2152, https://doi.org/10.1190/segam2017-17767532.1.

Harris, P., and R. White, 1997, Improving the performance of f-x prediction filtering at low signal-to-noise ratios: Geophysical Prospecting, **45**, 269–302, https://doi.org/10.1046/j.1365-2478.1997.00347.x.

He, K., X. Zhang, S. Ren, and J. Sun, 2016, Deep residual learning for image recognition: IEEE Conference on Computer Vision and Pattern Recognition, 770–778.

Jain, V., and S. Seung, 2009, Natural image denoising with convolutional networks: Advances in Neural Information Processing Systems, 769–776.

Jones, I., and S. Levy, 1987, Signal-to-noise ratio enhancement in multichannel seismic data via the Karhunen-Loéve transform: Geophysical Prospecting, **35**, 12–32, https://doi.org/10.1111/j.1365-2478.1987.tb00800.x.

Liu, C., Y. Liu, B. Yang, D. Wang, and J. Sun, 2006, A 2D multistage median filter to reduce random seismic noise: Geophysics, **71**, no. 5, V105–V110, https://doi.org/10.1190/1.2236003.

Liu, Y., C. Liu, and D. Wang, 2008, A 1D time-varying median filter for seismic random, spike-like noise elimination: Geophysics, **74**, no. 1, V17–V24, https://doi.org/10.1190/1.3043446.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams, 1986, Learning representations by back-propagating errors: Nature, **323**, 533–536, https://doi.org/10.1038/323533a0.

Simonyan, K., and A. Zisserman, 2014, Very deep convolutional networks for large-scale image recognition: arXiv preprint arXiv:1409.1556.

Trickett, S., 2008, F-xy Cadzow noise suppression: 78th Annual International Meeting, SEG, Expanded Abstracts, 2586–2590, https://doi.org/10.1190/1.3063880.

Wang, W., J. Gao, C. Chen, and Z. Zhu, 2012, Random seismic noise suppression via structure-adaptive median filter: Chinese Journal of Geophysics, **55**, 1732–1741.

Wang, W., X. Wang, H. Zeng, and Q. Liang, 2017, Preconditioning point-source/point-receiver high-density 3D seismic data for lacustrine shale characterization in a loess mountain area: Interpretation, **5**, no. 2, SF177–SF188, https://doi.org/10.1190/INT-2016-0107.1.

Wu, Z., and T. Liu, 2008, Wavelet transform methods in seismic data noise attenuation: Progress in Geophysics, **23**, 493–499.

Xie, J., L. Xu, and E. Chen, 2012, Image denoising and inpainting with deep neural networks: Advances in Neural Information Processing Systems, 341–349.

Zhang, K., W. Zuo, Y. Chen, D. Meng, and L. Zhang, 2017, Beyond a gaussian denoiser. Residual learning of deep CNN for image denoising: IEEE Transactions on Image Processing, **26**, 3142–3155, https://doi.org/10.1109/tip.2017.2662206.

10.1190/segam2018-2998114.1