Introduction

The high-resolution method is a vital technique for broadening the effective bandwidth of the observed stacked seismic data, to promote any thin layers identification. There are two commonly used model-driven methods to improve the resolution of seismic data: least-squares inverse filtering and sparse-spike inversion. The method within a least-squares framework rely on the assumption that the reflectivity of the underground medium has a spectrum of white noise (A.J.Berkhout, 1977; Cooke and Schneider, 1983; Bickel and Martinez, 1983). However, these methods often have difficulty yielding stable and unique results because the wavelet is band-limited (Du et al., 2018; Yuan and Wang, 2013). To avoid meaningless results, many sparse-spike inversion (SSI) methods have been proposed (Sacchi et al., 1994; Debye and Van Riel, 1990). This type of methods assume that the reflectivity sequence is sparse. By constraining the reconstructed reflection coefficients with sparseness regularization, these methods can obtain a sparse reflectivity estimation with stable increased bandwidth. To sum up, the traditional methods mentioned above invert for a high-resolution seismic data by relying on mathematical models with prior information or certain assumptions.

In recent years, data-driven methods have been introduced to improve seismic resolution (Kim and Nakata, 2018; Zhang et al., 2019). It implements the feature learning of training set with multiple levels of nonlinear mapping through multiple processing layers of the network. However, the feasibility of these methods depends on the quality and quantity of the labeled data. In this work, we use the CycleGAN-based framework of image-to-image translation to improve the resolution of seismic data. The key step of the method is to build a training set, which includes the realistic synthetic seismic traces generated from the well-log data (labeled data), and the original seismic data itself (unlabeled data). Then, by feeding the training set to train network, a complex nonlinear mapping from low-resolution to high-resolution seismic data can be effectively obtained. We conduct numerical experiments to demonstrate our method using both synthetic data and field data, the synthetic data results show that the proposed method has a better performance than the traditional method. In addition, field data applications further demonstrate its applicability.

Theory

In this work, we adopt the CycleGAN to improve seismic data resolution. Specifically, CycleGAN is composed of two pairs of GANs. Each GAN is used to approximate the mapping from one domain to another (Zhu et al., 2017). Therefore, we consider the frequency bandwidth reconstruction and low-pass filtering process in seismic data as a special forward and backward form of domain learning. This process can be approximated by using two pairs of generative and discriminative sub-models.

In addition, the CycleGAN adds an additional cycle-consistent loop through the architecture of two pairs of sub-models. The cycle-consistent loop not only use labeled synthetic data but also unlabeled raw seismic data to adjust parameters of the network (Benaim and Wolf, 2017). Figure.1 shows the CycleGAN workflow used to improve the resolution of seismic data. In Figure.1a, two generators take data from the one domain as input and outputs new data for the another, and two discriminators classify whether the generated data belongs to the corresponding domain, respectively. In this case, the low-resolution seismic data from domain Y are converted into corresponding high-resolution results, and the discriminator is used to distinguish whether the output data distribution belongs to domain X. In addition, as shown in Figure.1b, a wide bandwidth seismic traces output by the Generator G could be used as input to the Generator F and its result should match the original narrow bandwidth seismic traces. Similarly, the corresponding reverse process should be consistent (Figure.1c).

To mitigate the dependence on labeled synthetic data and utilize the knowledge contained in unlabeled seismic data, we input unlabeled raw seismic data into the CycleGAN and make the results consistent with the input data by cycle-consistent loop (i.e., the input seismic data remains unchanged after being mapped through Generator G and Generator F).

As we all know, training set can affect the performance of many machine learning applications. Here, we
Figure 1: Training a CycleGan to improve seismic data resolution. In a), the network makes the translation between the low-resolution and high-resolution seismic traces. b) Generator G converts the low-resolution trace (domain Y) to high-resolution trace (domain X), and then Generator F converts result back to the original trace, c) represents the reverse process similar to b).

generate different labeled data according to different processing regions, so that the network extracts the characteristics of the region for improving the resolution of seismic data. According to the convolution model, a synthetic seismic trace can be obtained by convolving the reflection coefficient and wavelet to build labeled data. Therefore, we first use the reflection coefficient calculated from the well-log data and the wavelet to generate synthetic seismic trace (domain X) that can tie up with near-well trace. Next, we convolute the same well-log data with a wide frequency bandwidth wavelet, and then regard the generated seismic trace (domain Y) as the broadband version of the seismic trace. To learn the appropriate mapping function from the training set, a patch-based processing is required. The patch samples used for training CycleGAN (Figure.1) are obtained through a sliding time window with a fixed step along the synthetic seismic trace, and we divide the seismic data through the same patch-based processing. When processing synthetic example and field data example, the time window and step size we choose are 64×1 and 1, respectively. In addition, we perform normalization ($x' = x / \max(|x|)$) on the raw seismic data, and normalize the training data using the statistics computed from the corresponding seismic data.

Examples

We convolute a 55 Hz Ricker wavelet with Marmousi reflectivity model, and then add 10% random noise to generate the synthetic data (Figure.2(a)). This data consists of 737 traces and 749 sampling points with an interval of 2 ms. Next, we use a 20-180 Hz band-pass wavelet to construct the referenced ideal high-resolution result from the Marmousi reflectivity model, and display it in the Figure.2(b). We assume that the seismic trace at CDP 200 is a well-logging trace, which can be synthesized by convoluting a wavelet with the reflectivity from logging data. Therefore, we extract the 200th trace as labeled data from the synthetic data and the high-resolution result respectively. The remaining 736 seismic traces in the synthetic data are treated as unlabeled data and used to detect the results of broadening the effective bandwidth. We also use the SSI method to retrieve the sparse reflectivity, and then apply the 20-180 Hz band-pass wavelet to the estimated reflectivity to obtain the high-resolution result, as shown in Figure.2(c). Figure.2(d) shows the result after applying the trained CycleGAN to the synthetic data. From Figure.2, we can observe that the high-resolution data using the proposed method (Figure.2(d)) is more continuous and has a higher signal-to-noise ratio compared to SSI, and is closer to the referenced ideal result in Figure.2(b). In order to precisely examine the frequency bands of different data, we calculate the normalized amplitude spectrum (AMP) of the synthetic example, the SSI method, the proposed method, and the referenced ideal result, and show it in Figure.2(e). From the 2(e), we can find that both methods broaden the frequency band of the synthetic example. The green curve draws the AMP for SSI, which achieves a good performance when the frequency band is 0-80 Hz. However, SSI result
Figure 2: (a) The synthetic data, (b) the high-resolution data (referenced ideal result). (c) High-resolution data by the SSI method; (d) high-resolution data by the proposed method. (e) Normalized AMP of synthetic example, the high-resolution data by the SSI method, and that by the proposed method, and normalized AMP of the referenced ideal result.

(green curve) and the referenced ideal result (black curve) has poor fitting quality in the frequency band of 80-180 Hz. In addition, the value of SSI is close to zero when the frequency band is 150-180 Hz, so that the frequency content of this part is not recovered. In comparison, the proposed method result (red curve) is distinctly close to the reference ideal result (black curve) within the entire frequency band (i.e., 0-180 Hz).

To further test the feasibility of our proposed method, we improve the seismic resolution on a field data (Figure.3(a)). The data has 108 seismic traces and 611 sampling points with an interval of 2 ms. We first extract the well-logging reflectivity series at CDP 81, and then convolve wavelets of different frequency bandwidth with this reflectivity series as the labeled data. In order to generate realistic labeled data, we use the narrow bandwidth wavelet that is consistent with the frequency band of the field data. Specifically, we use 5-10-70-80 Hz and 5-10-25-45 Hz frequency bandwidth wavelet as wide bandwidth wavelet and narrow bandwidth wavelet to build the labeled data. The estimated results using the proposed method are displayed in Figure.3(b). The result provided by the proposed method have more thin

Figure 3: (a) The field data, (b) the high-resolution data by the proposed method. (c) Normalized AMP of the field data, and that by the proposed method.
layer information, particularly the continuous reflection events in the deep part (about 1s). These events are indistinguishable in the field data, but events become more distinguishable in the result using the proposed method. We also show the AMP of the field data (yellow curve), and the proposed method result (red curve) in the Figure.3(c). From the Figure.3(c), we can see that the proposed method retrieve the high-frequency constituents of the field data. Additionally, the low-frequency components of the proposed method result are consistent with the original data (about 5-38 Hz), so it is pointed out that the method has higher fidelity. In summary, the proposed method has better performance in broadening the effective bandwidth and preserving the frequency components of the original seismic data.

Conclusions

In this article, we introduce a new framework of data-driven for improving seismic resolution based on a machine learning approach. We utilize the CycleGAN to adaptively learn the bidirectional functional mappings between wide bandwidth and narrow bandwidth seismic traces, such that the trained network can be used to broaden the frequency band of the seismic data. Because of the nature of our CycleGAN-based approach, the proposed framework is only rely on the data itself (observed seismic data and well-log data) and thus can be much more flexible in practice. The experimental results on the synthetic data show that the proposed method provides a precise high-resolution estimation compared with SSI. More importantly, the application on the field data further demonstrates that CycleGAN-based approach broaden the bandwidth obviously and stably in the seismic data while preserving the low-frequency constituents of the original seismic data.

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (42074141, 41874141) and the Strategic Cooperation Technology Projects of CNPC and CUP (ZLZX2020-03).

References


