applied sciences

MDPI

Article

Signal Reconstruction of Arbitrarily Lack of Frequency Bands from Seismic Wavefields Based on Deep Learning

Xin Li, Fengjiao Zhang * and Liguo Han

College of Geoexploration Science and Technology, Jilin University, Xi Min Zhu Street No. 938, Changchun 130026, China; li_x22@mails.jlu.edu.cn (X.L.); hanliguo@jlu.edu.cn (L.H.)
* Correspondence: zhangfengjiao@jlu.edu.cn

Abstract: Due to the limitations of seismic exploration instruments and the impact of the high frequencies absorption by the earth layers during subsurface propagation of seismic waves, recorded seismic data usually lack high and low frequency information that is needed to accurately image geological structures. Traditional methods face challenges such as limitations of model assumptions and poor adaptability to complex geological conditions. Therefore, this paper proposes a deep learning method that introduces the attention mechanism and Bi-directional gated recurrent unit (BiGRU) into the Transformer neural network. This approach can simultaneously capture both global and local characteristics of time series data, establish mappings between different frequency bands, and achieve information compensation and frequency extension. The results show that the BiGRU-Extended Transformer network is capable of compensating and extending the synthetic seismic data sets with the limited frequency band. It has certain generalization capabilities and stability and can effectively handle various problems in the data reconstruction process, which is better than traditional methods.

Keywords: attention mechanism; signal reconstruction; frequency extension; BiGRU-extended transformer

Citation: Li, X.; Zhang, F.; Han, L. Signal Reconstruction of Arbitrarily Lack of Frequency Bands from Seismic Wavefields Based on Deep Learning. *Appl. Sci.* **2024**, *14*, 4922. https://doi.org/10.3390/app14114922

Academic Editor: Yoshiyuki Kaneda

Received: 29 April 2024
Revised: 3 June 2024
Accepted: 4 June 2024
Published: 6 June 2024

1. Introduction

Ideal broadband seismic data can reveal intricate geological details and provide high-resolution images of subsurface structures. However, due to the excitation characteristics of the explosive source and the absorption of high frequencies by the earth's layers, as well as the suppression of low-frequency noise in data processing, the actual seismic data often partially lack high-frequency and low-frequency signals, resulting in insufficient imaging resolution and limited detection capabilities of deep geological structures [1]. To mitigate these challenges, frequency augmentation processing becomes essential. Frequency extension processing of seismic data plays an important role in data interpretation and inversion. Extending the high-frequency information can help to shorten the wavelet width and improve the spatial resolution in imaging results. At the same time, enhancing the low-frequency information can weaken the energy of the side lobe, dividing the layers more explicitly and precisely [2]. In full waveform inversion, the low-frequency information can improve the inversion's accuracy and estimate the parameters of the subsurface strata closer to the actual situation [3]. Therefore, the research and application of frequency extension methods for seismic data have become one of the key means to improve data resolution.

Traditional methods for extending the frequency range of seismic data typically involve deconvolution, inverse Q filtering, and time-frequency transformation [4]. However, these techniques come with their own set of challenges. Deconvolution may introduce some unstable factors, leading to inaccurate processing results. Inverse Q filtering is effective at enhancing high frequencies but falls short in amplifying low frequencies. The time-frequency transformation method may be affected by noise and other interference, resulting in incomplete and inaccurate frequency recovery. In recent years, some new seismic data

frequency extension methods have begun to emerge with the continuous development of technology. Sun et al. [5] proposed a spectrum extension technique that combined broadband processing and a tuned energy enhancement method, which effectively improved the seismic identification and prediction capabilities of thin-crossing reservoirs. Han et al. [6] solved the sparse inversion problem based on compressed sensing theory and achieved low-frequency compensation of seismic data. Rui et al. [7] proposed a deconvolution low-frequency recovery method based on high- and low-frequency joint measurement, solving the amplitude and phase distortion generated during the signal acquisition process. Ding et al. [8] used compressed sensing technology and a broadband Yu-type low-pass shaping filter to analyze and compensate for seismic low-frequency information. Zhang et al. [9] realized the expansion of lacking low-frequency signals based on compressed sensing theory. Zhang et al. [10] proposed a time-domain seismic data compression method and frequency band extension using nth power operations, which effectively enhanced the low-frequency components of seismic data.

In recent years, the advancement of computer hardware resources has significantly facilitated the application of deep learning in geophysical research [11–13]. Zhang et al. [14] introduced the GAN and combined it with an adaptive bandwidth extension method to use deep learning to improve the resolution of seismic data. Sun et al. [15–17] studied low-frequency bandwidth extension methods through convolutional neural networks (CNNs). Fang et al. [18] proposed a data-driven low-frequency recovery method based on high-frequency signals to improve the reliability of full waveform inversion. Choi et al. [19] enhanced the spectrum by training the U-Net network by considering data characteristics and prior information. Nakayama et al. [20] achieved low-frequency extrapolation through supervised learning. Jin and Hu et al. [21,22] proposed a progressive transfer learning method to supplement low-frequency information in full waveform inversion using extrapolated low-frequency reflection seismic data. Zhang et al. [23] used neural networks to map post-stack data into broadband data, effectively extrapolating low-frequency information. Ni [24] proposed a data-driven high and low-frequency recovery scheme that used the attention mechanism and ResNet structure to achieve frequency band widening of narrowband seismic records. Zhang et al. [25] improved the U-Net network through residual structures and combined the constraints of physical parameters to achieve low-frequency recovery.

In seismic exploration, the choice of wavefield components is essential to meet different exploration needs. Seismic data usually contain multiple wavefield components; each has different propagation characteristics and responses to subsurface structures [26]. Observing and analyzing information such as the amplitude, travel time, and frequency of the reflected waves can help infer the stratum structure and lithology distribution. Particularly, the direct wave of the near-offset has a short propagation path and is less affected by the subsurface medium, offering higher fidelity to the source wavelet. Analyzing near-offset direct waves allows for a more precise determination of the wavelet's shape, frequency, and other properties, thereby facilitating a better understanding of the source characteristics. Effective suppression of multiples can mitigate the crosstalk artifacts frequently observed in stack and offset sections, thus enhancing geological interpretation accuracy [27]. Therefore, when discussing the effects of signal reconstruction, the wavefield components of the original data must be considered, which helps to select appropriate methods and parameters to meet specific exploration needs.

The core issue of this research is to take data-driven bands as the main body, using the effective information and wavefield attributes within the existing seismic frequency bands to effectively realize the information compensation and reconstruction of seismic data in any frequency bands, thereby improving the imaging quality of seismic data. Therefore, we propose a method based on deep learning, incorporating the self-attention mechanism and BiGRU to design a neural network that facilitates the mapping between each frequency band and reconstructs the implicit frequency band characteristics. In order to ensure the objectivity of the experiment, we introduce a set of evaluation metrics to verify the effectiveness and accuracy of the method in this paper through a series

of numerical examples and evaluate the effect of signal reconstruction under different wavefield conditions.

2. Theory and Method

2.1. Theory

The field of application deep learning is constantly expanding, and it is widely used in seismic data processing. The current method of using neural networks to solve the problem of frequency extension is as follows. The data produced by the main frequency are designated as full frequency band data, while the data corresponding to each individual frequency band are referred to as frequency band-slice data. It is then posited that a specific functional relationship exists between the full frequency band data and the frequency band-slice data:

$$Y(x) = G(M(x)) \tag{1}$$

x is the full seismogram represented by a matrix; Y represents the full frequency band data; M represents the frequency band-slice data; G represents the mapping relationship between the two. The mapping relationship is approximated by introducing the neural network BE-Transformer, and the expression is as follows:

$$\overline{Y}(x, \alpha) = \mathrm{Net}(M(x, \alpha)) \tag{2}$$

$\overline{Y}$ represents the predicted complete full band data; Net represents the BiGRU-Extended Transformer network structure; α is a hyperparameter. This mapping is non-linear and can handle complex patterns and laws. The neural network continuously adjusts the learning parameter α to achieve a good fit between the input and output so that:

$$Y(x) \approx \overline{Y}(x, \alpha) \tag{3}$$

During the neural network training process, the implicit relationship between data is established by solving the least squares optimization problem. The objective function expression is as follows:

$$J(M, \alpha, Y) = \mathrm{argmin}\sum_{i} L[\mathrm{Net}(M(x_i, \alpha), Y_i] \tag{4}$$

In the formula, J is the network optimization objective function; L is the loss function. The iterative idea approaches the optimal solution by using the gradient descent method and expresses the network parameters α_{i+1} as:

$$\alpha_{i+1} = \alpha_t + \beta[\nabla\alpha_{i=1,2,\ldots,t} J(M, \alpha_i, Y)] \tag{5}$$

In the formula, α_t is the network parameters after the t-th iteration; β is the update directions; $\nabla\alpha_{i=1,2,\ldots,t}$ is the gradients relative to J. Through optimization algorithms and back-propagation mechanisms, neural networks can gradually approach the objective function and improve the accuracy of predictions.

2.2. Network Structure

The Transformer model was proposed by the Google team in June 2017 by Ashish Vaswani et al. in the paper "Attention Is All You Need" [28], and the proposal of this model has had a significant impact in the field of Natural Language Processing (NLP). Unlike traditional recurrent neural networks, the Transformer model abandons the sequential structure and, instead, adopts the mechanism called self-attention, which allows the model to be trained in parallel and improves the training speed. In addition, the self-attention mechanism also allows the model to consider global information when processing the input data, further improving the model's ability to understand and generate text [29–31]. The BiGRU is an improved recurrent neural network structure that combines the bidirectional

recurrent neural network (RNN) and the gated recurrent unit (GRU) [32,33]. It aims to solve the gradient vanishing or gradient explosion problems that are prone to occur in traditional RNNs when processing long sequences. It consists of two independent GRU layers, one processing the sequence along the forward direction and the other processing the reverse direction. The bidirectional structure allows BiGRU to capture long-distance dependencies in sequences, thus improving the model's performance. To achieve the acquisition and reconstruction of signals in any arbitrary lack of bands of seismic wavefields, the paper proposes the BiGRU-Extended Transformer (hereafter referred to as BE-Transformer) network.

The study adopts a hybrid architecture of the extended version of Transformer and BiGRU. The introduction of multi-head attention (MHA) allows the model to simultaneously focus on different parts of the input sequence in different heads, thereby enhancing the model's expressiveness and ability to capture complex dependencies. The introduction of Scaled Dot-Product Attention allows the model to learn dependencies in the input sequence based on the similarity between various positions through scaled dot-product operations and weighted sum operations, thereby better capturing semantic information. Position Encoding introduces position information into the embedding vector of sequence data, allowing the model to maintain long sequence dependencies while considering the order information of words or characters in the sequence. BiGRU is introduced to help the model further refine local information and waveform details in seismic images. With such a setting, the network can process local and global features of time series data simultaneously and improve flexibility and prediction performance, thereby more accurately establishing the mapping relationship between frequency band-slice data and full frequency band data.

When processing large-scale data, the inner product operation relied on by the attention mechanism in the Transformer model may cause numerical instability. Introducing the scaling factor allows us to adjust the scale of the inner product result to finely control the inner product operation, which improves the model's performance when dealing with large-scale data, enhances the modeling capability of the model, and reduces the training difficulty. The regularization technique is introduced into the model, and the data are linearly transformed through the Add Norm layer to eliminate the dimensional differences between the eigenvalues, making the model more stable and reliable. The Dropout layer randomly inactivates some neurons to reduce the model's dependence on training data, thereby alleviating the overfitting problem and improving the network's generalization.

In summary, this deep learning model can construct a feature-rich and multi-scale domain, which enhances their capabilities when dealing with complex tasks. The network we used is shown in Figure 1.

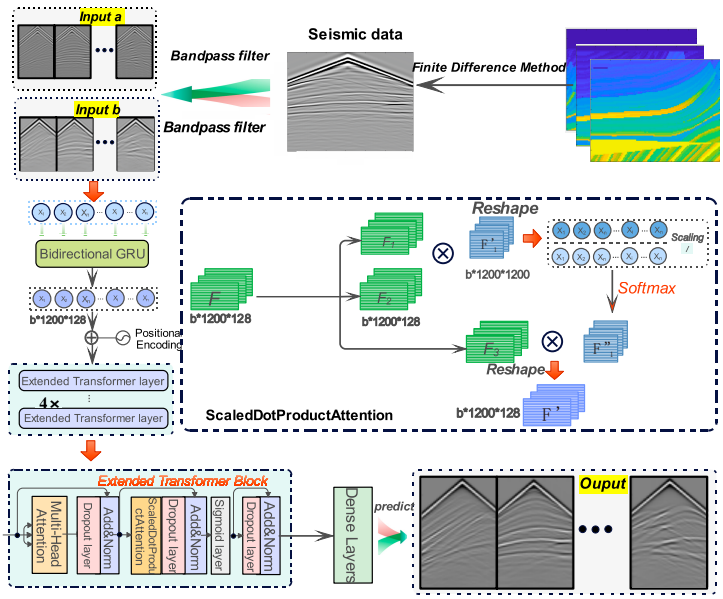

Figure 1. Network processing architecture diagram.

3. Numerical Example

3.1. Data Organization

We choose two different velocity models and generate three different frequency band-slice data sets as the data basis for this experiment to verify the generalization of the network in different situations. Namely, the randomly intercepted velocity sub-model (Figure 2a), the Marmousi II model (Figure 2b), and the four-layer layered model (Figure 2c). Each velocity model was configured with a uniform grid size of 101 × 128, maintaining a grid spacing of 8 m in both horizontal and vertical directions. The finite difference method is used to perform forward modeling on each velocity model, and the Ricker wavelet with a dominant frequency of 20 Hz is selected as the source function. During the forward modeling process, the total recording time is set to 1.2 s, the sampling interval is 0.001 s, and a perfectly matched layer boundary condition of 60 grid point thickness is set. We construct the labeled dataset with full-band seismic records. We generate three different frequency band-slice data as the sample data sets. For the first sample data set, we apply a band-pass filter between 15 Hz and 25 Hz to the full-band data to simulate the band-slice seismic records. Subsequently, we adjust the amplitude in different parts of the first sample data set to generate second and third sample data sets. A total of 1280 seismic records are received, and we randomly select 70% of the entire data set as the training set and the remaining 30% as the test set.

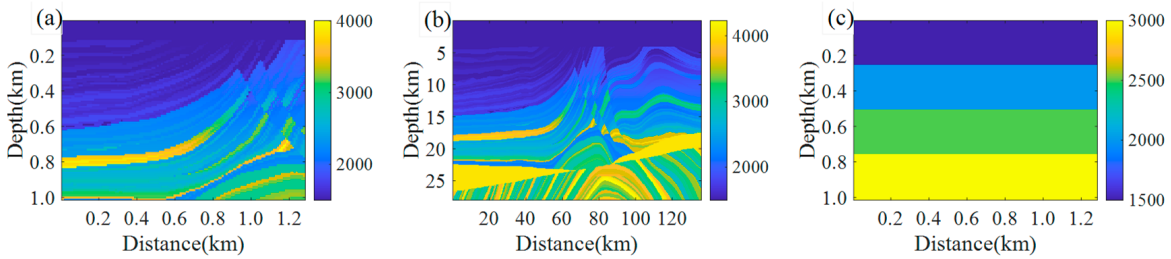

Figure 2. Velocity model (**a**) Marmousi II sub-model; (**b**) Marmousi II model; (**c**) Four-layer velocity model.

Our computing device is NVIDIA GeForce RTX 3080, and our neural network framework is TensorFlow-GPU 2.6.0 with 10 GB of video memory. We built two network models: the BE-Transformer network and the CNN network. During the training process, the Adam algorithm is used to optimize the objective function, and the activation function is Sigmoid. The network models were trained with the same hyperparameters to ensure a fair comparison. The entire training process was carried out for 600 iterations, the learning rate was set to 0.001, the number of training data in each batch was 32, and the inputs and outputs of both networks were 1200 × 128. This setting provides a basis for subsequent model evaluation and comparison.

In deep learning, loss function is an important concept used to measure the difference between the model predictions and the real data. The predictive accuracy and generalization of the model are improved by minimizing this difference. The loss function is Mean Square Error (MSE), and the formula is:

$$\text{MSE\_Loss} = \frac{1}{n}\sum_{i=1}^{n}(x_i - f(x_i))^2 \tag{6}$$

In the formula, x_i is the real data, $f(x_i)$ is the prediction result of the neural network, and n is the size of the network batch.

The loss function curve during the BE-Transformer network training process is shown in Figure 3. The loss curve decreases rapidly at the beginning of the training and gradually stabilizes as the number of iterations increases.

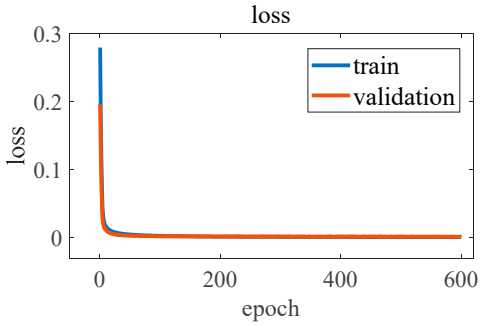

Figure 3. The loss convergence of the network.

We used the Pearson correlation coefficient as the evaluation index when evaluating the accuracy of the prediction model. The Pearson correlation coefficient measures the linear relationship between two variables, with values between −1 and 1. Here, we use it to evaluate the similarity between the predicted output and the expected data. Its expression is:

$$R^2 = \frac{\left(\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})\right)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2} \quad (7)$$

In the formula, x_i and y_i ($i = 1, 2, \ldots, n$) are the predicted output and expected data, respectively, and $\overline{x}$ and $\overline{y}$ are the mean values of the predicted output and expected data, respectively.

Figure 4 shows the Pearson correlation coefficient diagram of the BE-Transformer training process. As the number of iterations increases, the Pearson correlation coefficient gradually increases. After 600 iterations, the Pearson correlation coefficient value reached 99.5%, and various indicators reached stability. It indicates that the network has converged, has strong positive correlations between features and variables, and can capture the similarity between the predicted output and the expected data.

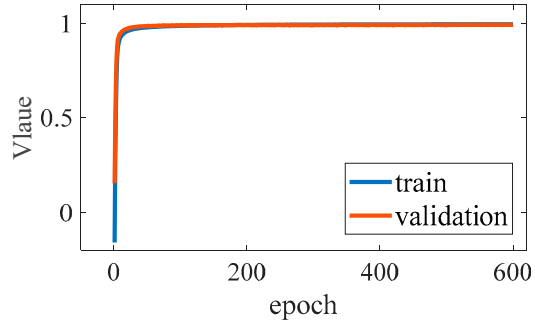

Figure 4. Pearson correlation coefficient plot during training.

3.2. Training Results

The comparison of the prediction results obtained after 600 iterations with the input data and expected data seismic records is shown in Figure 5, where the seismic source is located at x = 500 m. The input data is the first type of band-slice seismic record with only a band-pass filter applied. The label data are the full band seismic record. The comparison shows that the prediction results of the BE-Transformer are consistent with the expected data, and the wave group characteristics are clear. Although the prediction results from CNN also contain the desired features, there are apparent artifacts in the records, and the signal-to-noise ratio is low. To display the prediction effect of the model more intuitively, Figure 6 shows the single-trace waveform comparison between the BE-Transformer and CNN network prediction results and the expected data. The offset of Figure 6a is 56 m, and the offset of Figure 6b is 0 m. The prediction results of the BE-Transformer maintain high consistency and accuracy with the expected data in terms of amplitude and phase. As

for the CNN prediction results, although the main events match the expected data well, the continuous consistency of the phase information is not good, and there is an apparent undesired waveform oscillation phenomenon before the first arrival time, and the overall fitting effect is average. Therefore, from the comparison of macroscopic and local features, the BE-Transformer network has shown excellent reconstruction performance and good generalization ability from the tests on both simple and complex models. The effect is significantly better than that of the traditional CNN network.

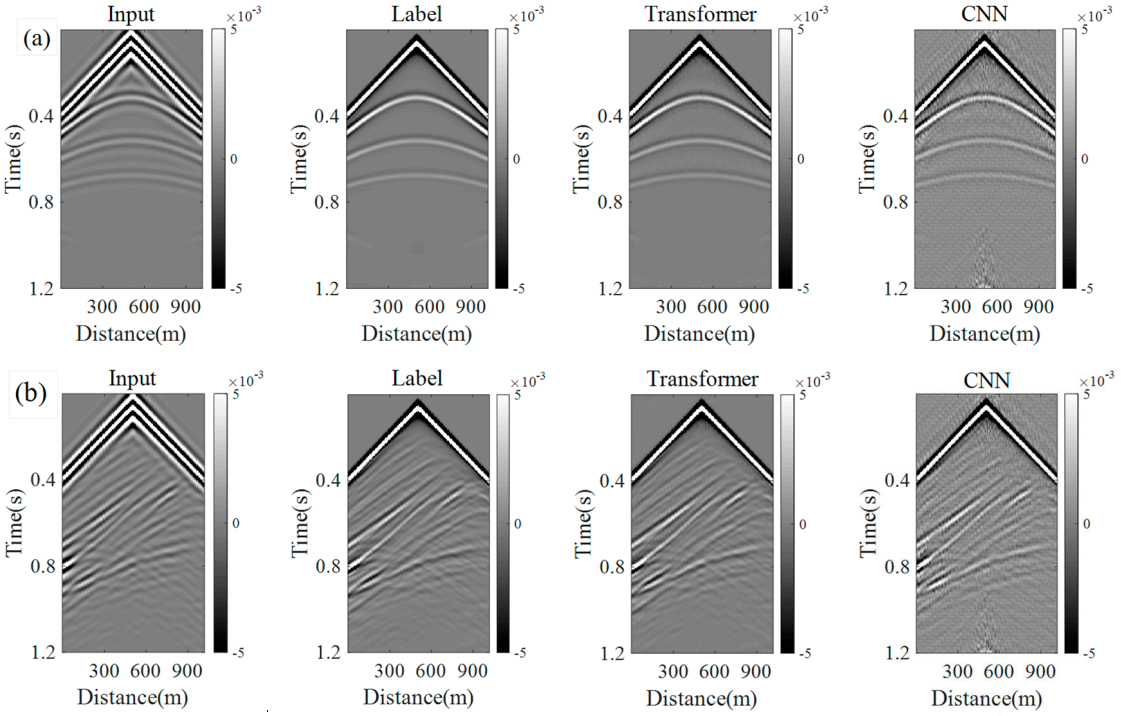

Figure 5. Comparison of BE−Transformer and CNN network prediction results with input and expected data seismic records. (**a**) Layered model; (**b**) Marmousi II velocity submodel.

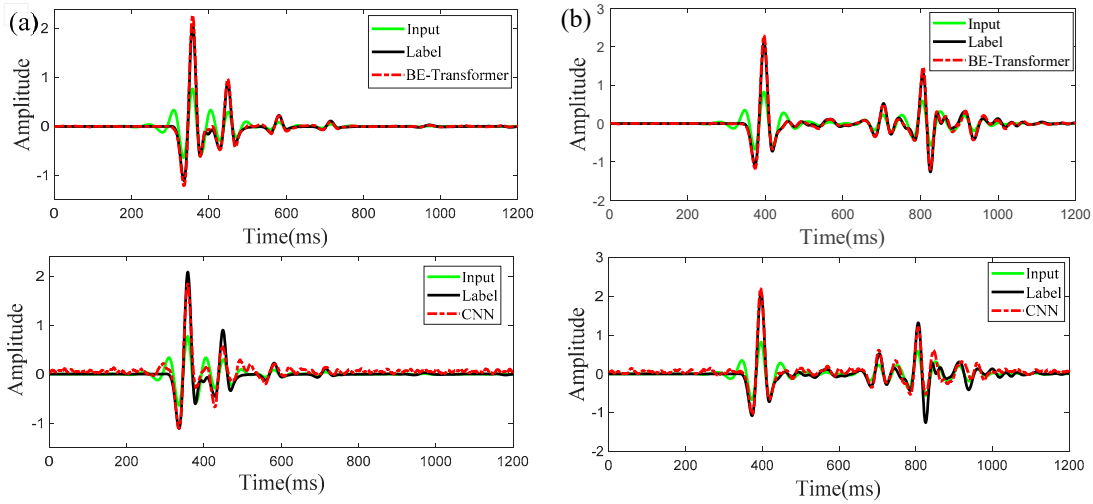

Figure 6. Comparison of single-trace waveforms between BE-Transformer and CNN network prediction results and input and expected data. (**a**) Layered model (offset = 56 m); (**b**) Marmousi II velocity submodel (offset = 0 m).

Reverse-time migration (RTM) was performed on the three data sets: original data (input data), expected data (label data), and BE-Transformer predicted results. We use

true velocity models to perform the RTM. The RTM imaging results and depth curves are shown in Figures 7 and 8, respectively. For the layered model, the comparison between the original and expected data shows that artifacts will appear in the imaging results, and the signal-to-noise ratio is low when the data lack low-frequency and high-frequency components. After signal reconstruction, the signal-to-noise ratio of the data is significantly improved, the velocity structure of the subsurface medium is correctly modeled, and the stratal continuity is good. The comparison between the input data and the prediction results shows that the artifacts in the input data have been significantly improved and eliminated, the details of the image are richer, and the imaging quality has been significantly improved. For the Marmousi II model, the main strong reflection interface of the expected data is clear, and the layer information will be accurate. However, due to insufficient frequency band information, the original data have side lobes and artifacts, and there is also certain layer distortion in thin layer areas and lower resolution. The data reconstructed by the BE-Transformer network have better resolution capabilities in the thin layer of shallow and the reservoir of deep, and the imaging accuracy is significantly improved.

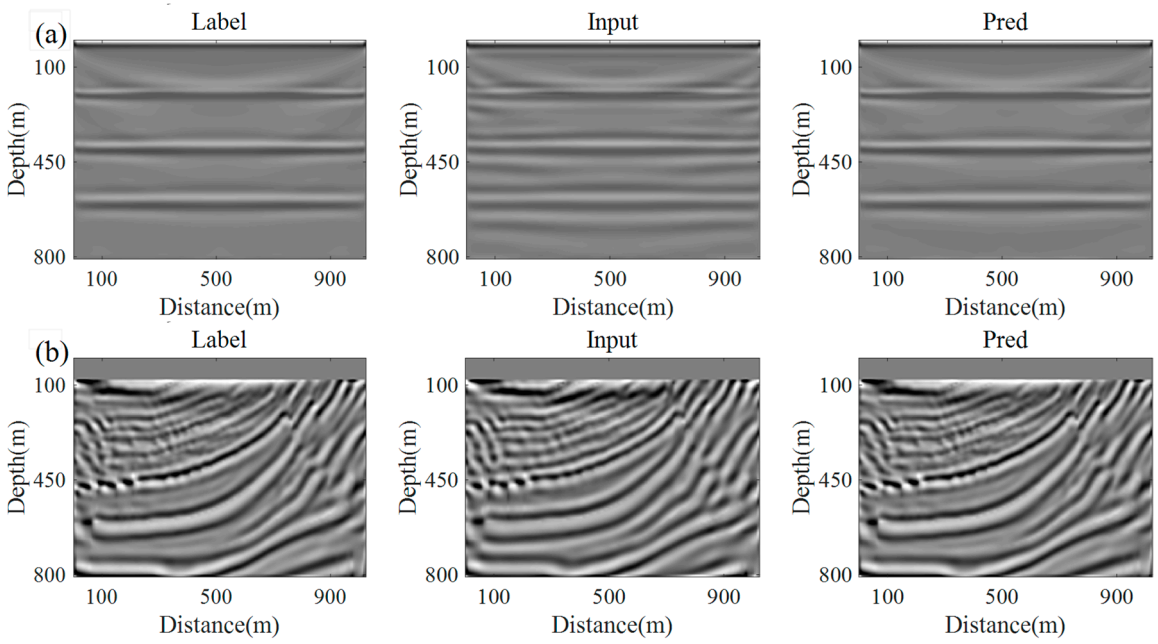

Figure 7. Comparison of BE-Transformer prediction results with original data and expected data RTM imaging results. (**a**) Layered model; (**b**) Marmousi II velocity submodel.

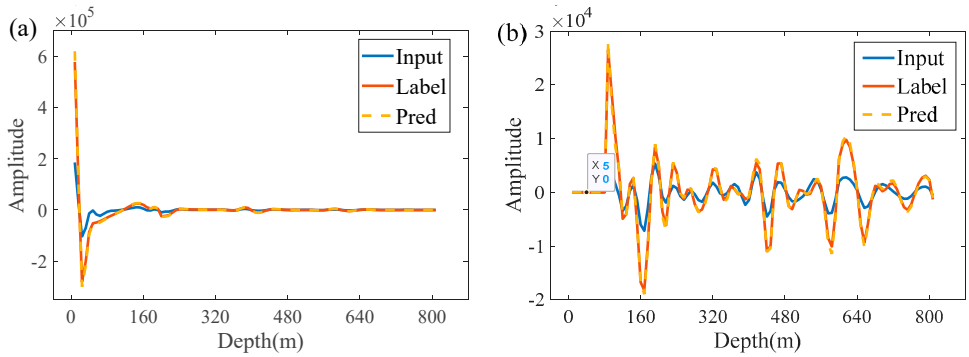

Figure 8. Comparison of BE-Transformer prediction results and original data and expected data reflectance curves. (**a**) Layered model; (**b**) Marmousi II velocity submodel.

We use the second and third sample data sets to test the signal reconstruction effect and verify the BE-Transformer network's generalization ability. In addition to band-pass

filtering 15−25 Hz, we also attenuate the amplitudes within different frequency bands of the filtered data set to simulate the condition of damage across various frequency ranges. Compared with the first data set, the second data set lacks some middle frequency. The third data set lacks high and low frequency and the frequency amplitude is very narrow. After 600 iterations, the BE-Transformer network prediction results of the layered model and the Marmousi II velocity model in the two cases are shown in Figures 9 and 10, along with the seismic record comparisons of input data and expected data. The predicted results have good waveform matching with the expected data, and the differences in phase and relative amplitude are very small, thus achieving amplitude-preserving prediction of seismic data.

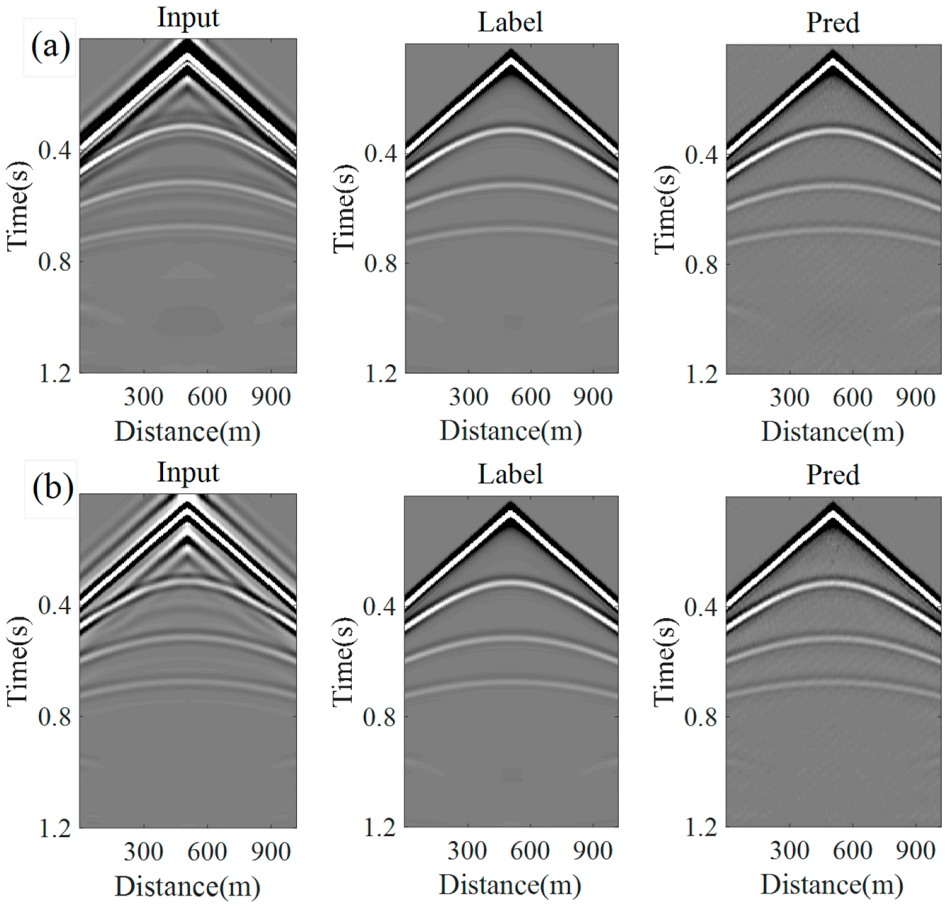

Figure 9. Comparison chart between BE-Transformer prediction results, input data, and expected data seismic records (Layered model). (**a**) is the simulated case lacking some middle frequency; (**b**) is the simulated case lacking high and low frequency and the frequency amplitude is very narrow.

To show the reconstruction effect of seismic data in each frequency band more intuitively, spectral analysis was performed on the above dataset, and the results are shown in Figure 11. Input(a–c) refers to the three different types of band-slice seismic data sets, which are the input of the network. The Label is the network label full band seismic data set. Pred(a–c) are the prediction results of the three different types. Whether a simple or complex model, the network can learn and capture the intrinsic correlation between frequency band-slice data set and full frequency band data set. The frequency bands are balanced and linearly gained, which achieves the information extension and compensation of arbitrary lacking frequency bands of the seismic data from local to whole and from unknown to known. This further proves the generalization ability of the BE-Transformer model to cope with various complex seismic data processing scenarios.

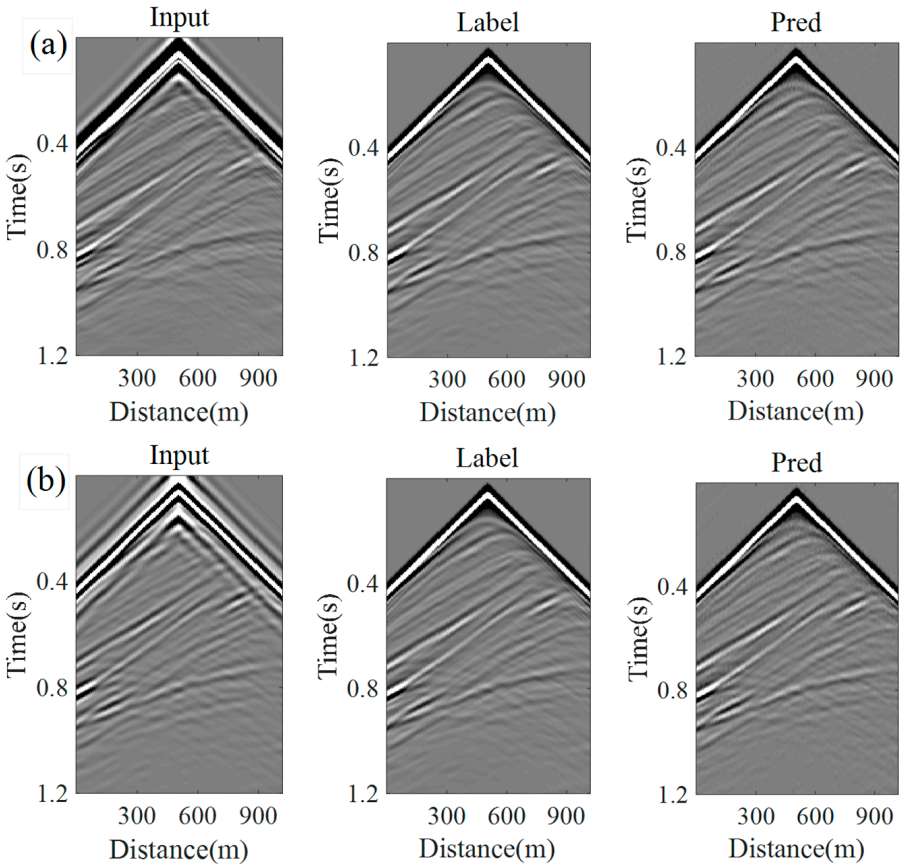

Figure 10. Comparison chart between BE-Transformer prediction results, input data, and expected data seismic records (Marmousi II velocity submodel). (**a**) is the simulated case lacking some middle frequency; (**b**) is the simulated case lacking high and low frequency and the frequency amplitude is very narrow.

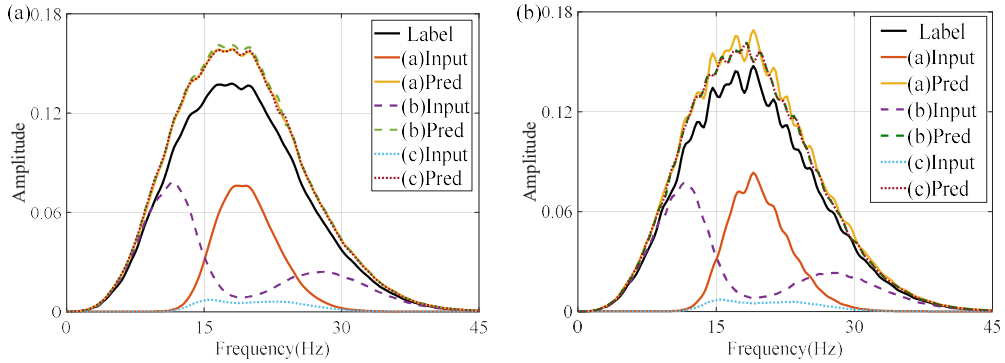

Figure 11. Comparison of BE-Transformer prediction results, input data, and expected data amplitude spectrum. (**a**) Layered model; (**b**) Marmousi II velocity submodel.

Considering that noise may impact the signal compensation reconstruction, random noise with a signal-to-noise ratio of 20 was added to the frequency band-slice data (only applied band-pass filter with 15−25 Hz) and entered into the model for anti-noise experiments. The comparison of seismic records before and after noise addition is shown in Figure 12. After testing, the noise has been effectively suppressed, and the effective signal has been restored and enhanced. The BE-Transformer network also has a good processing effect on data containing noise and has good anti-noise performance.

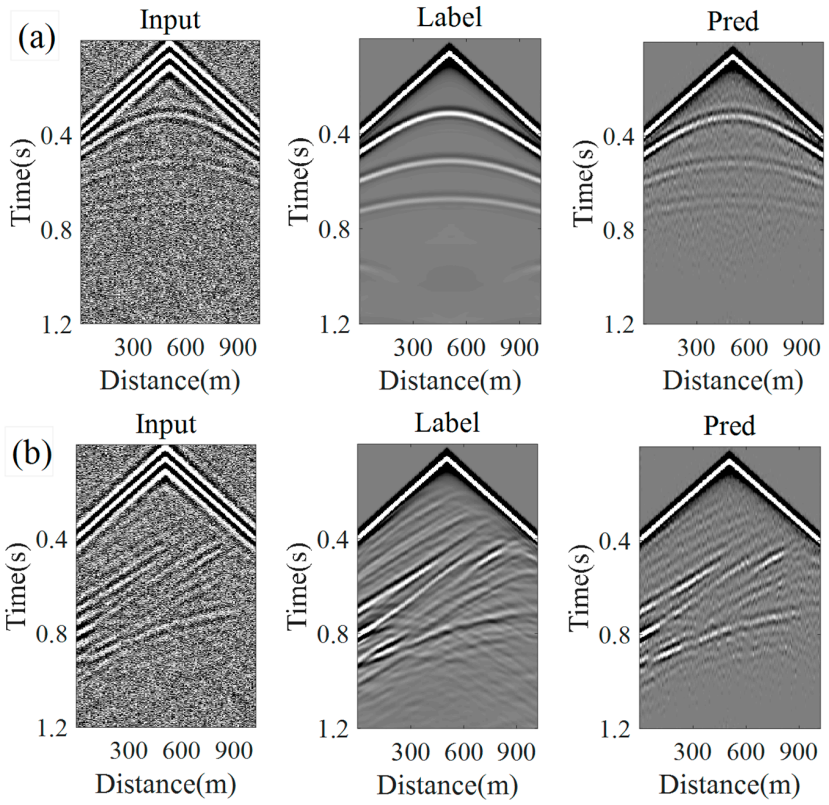

Figure 12. Anti-noise experiments. (**a**) Layered model; (**b**) Marmousi II velocity submodel.

4. Discussion of the Impact of Different Wavefield Components

During the signal reconstruction process, testing the effect of data training under different wavefield conditions is essential to improve the accuracy and reliability of seismic data processing. Therefore, by establishing labeled datasets under different wavefield conditions, the impact of different wavefield conditions on the signal reconstruction effect can be more accurately simulated, thereby optimizing the data processing algorithm. Here, we consider the wavefield in four cases. The absorbing boundary and free surface were used as the top boundary conditions. At the same time, the wavefield data with and without multiples were simulated under each surface boundary condition. Taking the four labeled data obtained by the layered velocity model shown in Figure 2c as an example, the single shot records with the source located at x = 500 m are shown in Figure 13a–d, respectively. The top boundary conditions of Figure 13a,b are absorbing boundaries, and Figure 13c,d are free surfaces. Figure 13a is the without free surface-related multiples wavefield but including direct wave, Figure 13b is the primary scattered field without direct wave, Figure 13c is the full wavefield, and Figure 13d is the full wavefield without direct wave.

We compared the seismic records and single-trace waveforms of input data, expected data, BE-Transformer network, and CNN network prediction results under four wave field conditions. Figure 14 shows the comparison of the source's seismic records at x = 500 m. The prediction results of the BE-Transformer are highly consistent with the expected data events, and the fit of the waveform features is very high. Although the CNN network prediction results have a high degree of waveform matching, there are apparent artifacts. The prediction results in wavefields other than the no-free surface multiple wave field are very poor, the phase axis is discontinuous, the signal-noise ratio is low, and the information is obviously distorted. Figure 15 shows a comparison of single-trace waveforms randomly selected under four wavefield conditions. The offset of Figure 15a is 56 m, the offset of Figure 15b is 8 m, the offset of Figure 15c is 16 m, and the offset of Figure 15d is 16 m. The prediction results of the BE-Transformer network have high waveform consistency with the

expected data in terms of amplitude and phase under arbitrary wave field conditions. However, the CNN network prediction results show a large number of waveform mismatches. The waveforms are discontinuous and there were many noise points and mutation points, and the fitting effect was not ideal. In summary, the BE-Transformer network performs well under different wave field conditions and can effectively deal with various problems in the reconstruction process of seismic data. In contrast, traditional CNN networks have apparent limitations in handling complex wavefield conditions.

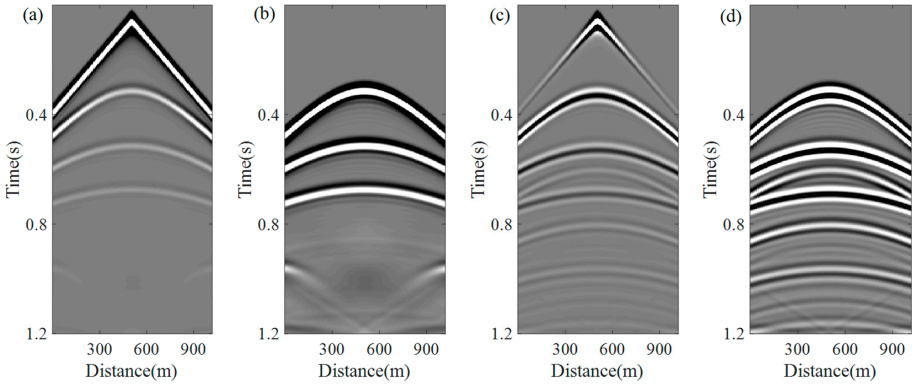

Figure 13. Labeled datasets under different wavefield conditions. (**a**) without free surface-related multiples wavefield; (**b**) primary scattered field; (**c**) full wavefield; (**d**) the full wavefield without direct wave.

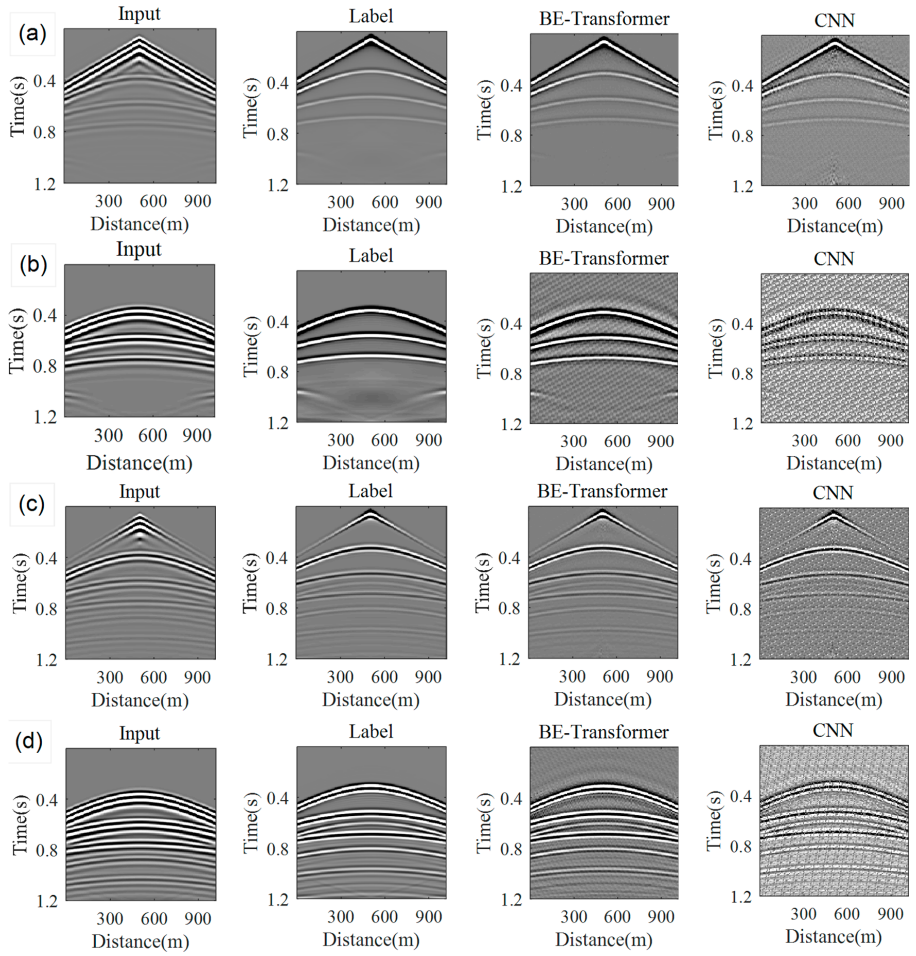

Figure 14. Comparison of BE-Transformer and CNN prediction results with input data and expected data seismic records under four wave field conditions. (**a**) without free surface-related multiples wavefield; (**b**) primary scattered field; (**c**) full wavefield; (**d**) the full wavefield without direct wave.

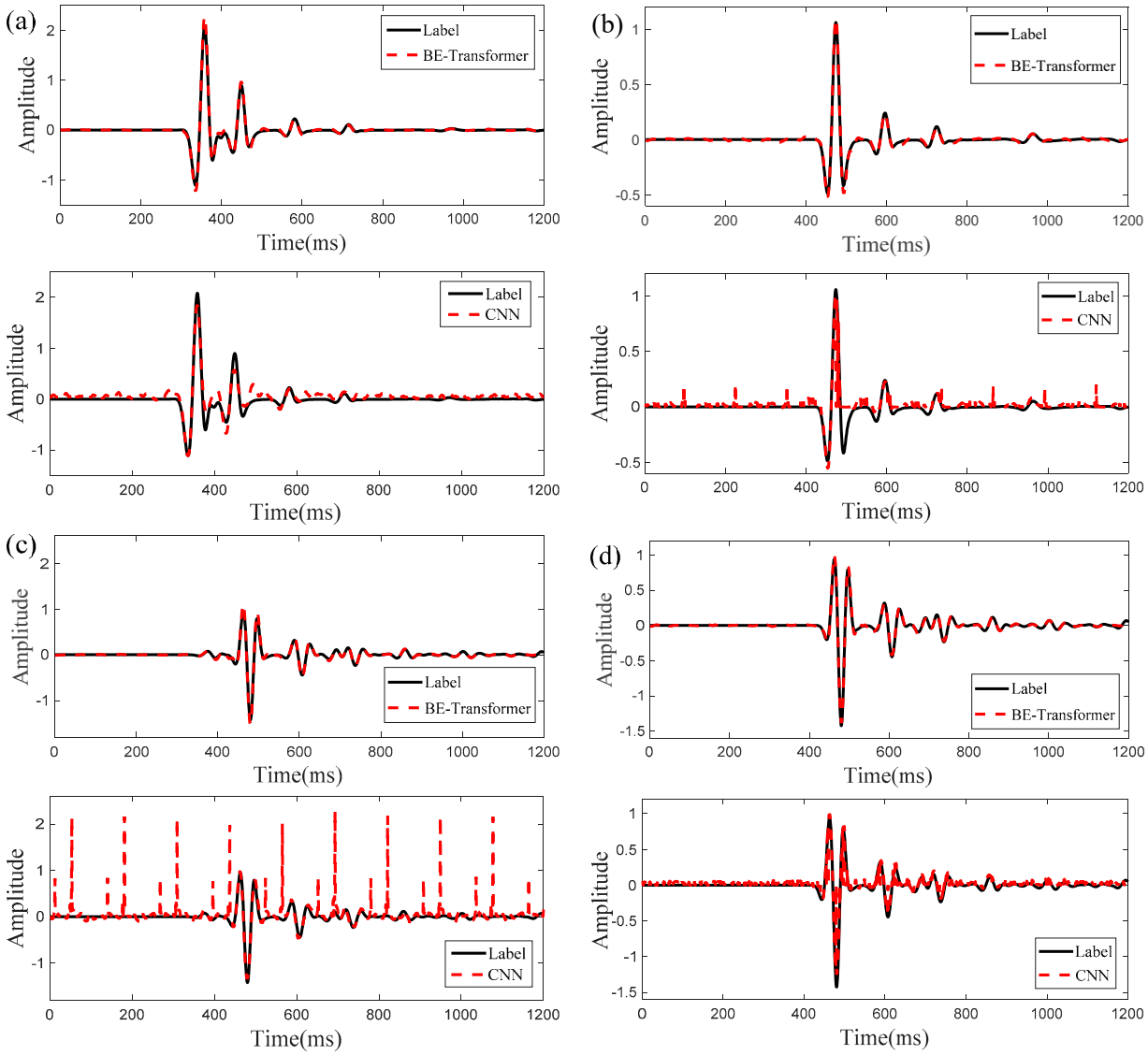

Figure 15. Comparison of BE-Transformer, CNN network prediction results, and expected data single-trace waveforms under four wave field conditions. (**a**) without free surface-related multiples wavefield; (**b**) primary scattered field; (**c**) full wavefield; (**d**) the full wavefield without direct wave.

To more intuitively demonstrate the signal compensation and reconstruction effect of the BE-Transformer network model, Figure 16 shows the spectrum analysis results of seismic records under four wavefield conditions. While the energy signal reconstructs the specific frequency band information in the original observation record, it also has a certain degree of richness in high and low-frequency bands. The results of the experiments under different wavefield conditions show that the prediction results are most accurate in the wavefields without free surface-related multiples. Training with data from a free-surface multiple-free wavefield containing direct waves and the entire wavefield yields better results than using wavefields without direct waves. The presence of multiples has little impact on the experiment, highlighting the importance of retaining direct wave components before signal reconstruction. This provides new insights for the data preprocessing of our reconstruction method.

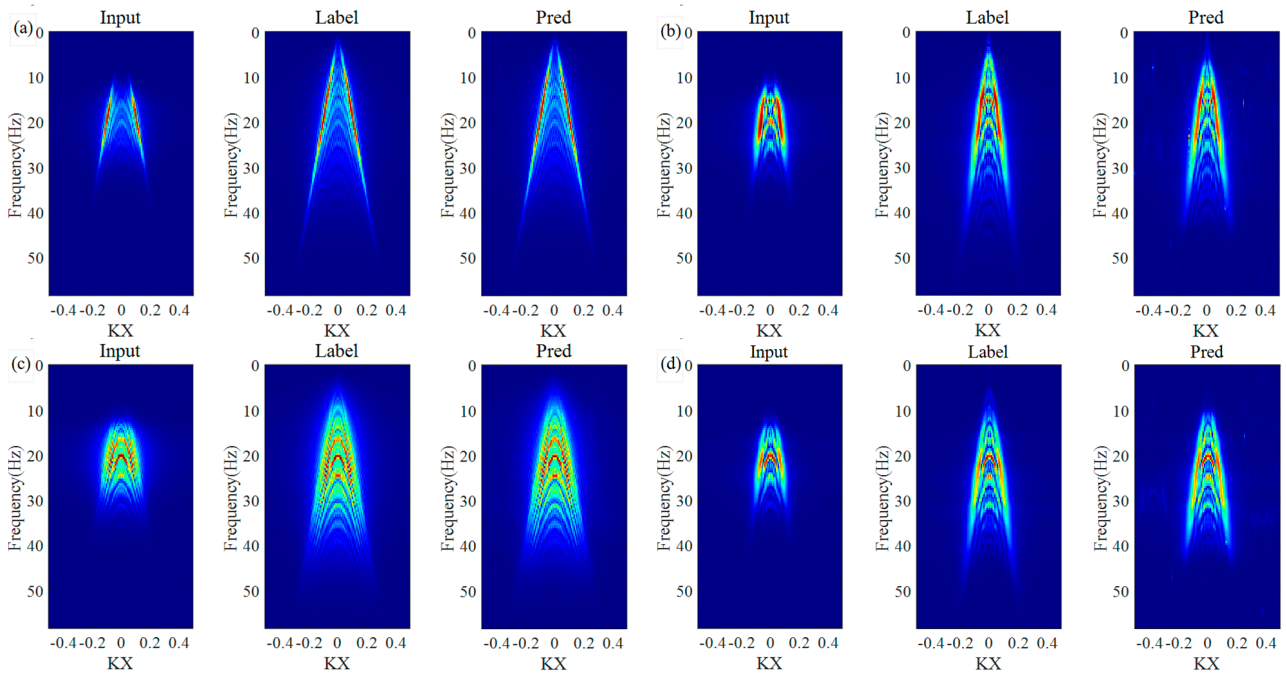

Figure 16. Comparison chart of FK spectrum of input data, expected data, and BE-Transformer prediction results under four wavefield conditions. (**a**) without free surface-related multiples wavefield; (**b**) primary scattered field; (**c**) full wavefield; (**d**) the full wavefield without direct wave.

5. Conclusions

This study introduces a novel deep learning approach designed to reconstruct signals missing in certain frequency bands within seismic wavefields. Numerical examples show that the BiGRU-Extended Transformer network model can process local and global features of time series data at the same time, flexibly utilize existing seismic frequency band information, and obtain and reconstruct the signals with some frequency band missing in the full frequency band. This process effectively compensates for missing information, facilitating signal reconstruction. By comparing the impact of different wavefield conditions, we find that the importance of retaining direct waves before signal reconstruction. This provides new insights for the data preprocessing of our reconstruction method. The experiments in this paper were conducted under idealized conditions and have not yet touched on the application level of transfer learning. We plan to introduce real seismic data in future research to further validate the model's effectiveness and generalization ability in practical applications.

Author Contributions: Conceptualization, X.L., F.Z. and L.H.; Methodology, X.L., F.Z. and L.H.; Software, X.L.; Validation, X.L., F.Z. and L.H.; Formal analysis, X.L.; Investigation, X.L.; Writing—original draft, X.L., F.Z. and L.H.; Writing—review & editing, X.L., F.Z. and L.H.; Visualization, X.L.; Supervision, F.Z.; Project administration, F.Z. and L.H.; Funding acquisition, F.Z. and L.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the National Natural Science Foundation of China (No. 42130805, No. 42074154, No. 41974138).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest. Zhang, P. The Study on Full Waveform Inversion Based on Low-Frequency Seismic Wavefield Reconstruction. Ph.D. Thesis, Jilin University, Changchun, China, 2018.

References

1. Zhang, P. The Study on Full Waveform Inversion Based on Low-Frequency Seismic Wavefield Reconstruction. Ph.D. Thesis, Jilin University, Changchun, China, 2018.
2. Fang, S. Bandwidth Extension Method of Seismic Data Based on Full-band Extension Filter. Master's Thesis, China University of Petroleum, Beijing, China, 2022.
3. Shang, X.; Zhang, P.; Han, L.; Yang, Y.; Zhou, Y. Joint FWI of Active Source Data and Passive Virtual Source Data Recon-structed Using an Improved Multidimensional Deconvolution. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5905614. [CrossRef]
4. Chen, Z. Seismic Data Frequency Extension Based on Deep Learning. Master's Thesis, China University of Petroleum, Beijing, China, 2022.
5. Sun, X.; Zhang, Y.; Zhang, Y.; Zhou, C. Application of seismic frequency extension technology in dynamic analysis of thin layer reservoir development. *Oil Geophys. Prospect.* **2010**, *45*, 695–699.
6. Han, L.; Zhang, Y.; Han, L.; Yu, Q. Compressed Sensing and Sparse Inversion Based Low-Frequency Information Compen-sation of Seismic Data. *J. Jilin Univ. (Earth Sci. Ed.)* **2012**, *42*, 259–264.
7. Rui, Y.; Zou, Z.; Wang, S.; Shang, X.; Zhang, Y.; Ma, R. A method for simultaneous recovery of low-frequency amplitude and phase from conventional detector data based on high- and low-frequency joint measurement. *Oil Geophys. Pro-Specting* **2017**, *52*, 631–643.
8. Ding, Y.; Du, Q.; Liu, L.; Zhang, Q. Feature analysis and compensation of seismic low-frequency based on compressed sensing and broad-band YU-type low-passing shaping filter. *Chin. J. Geophys.* **2019**, *62*, 2267–2275.
9. Zhang, B.; Zhang, J.; Wu, Y. Research on protection and extension for seismic low frequencies. *Prog. Geophys.* **2019**, *34*, 1139–1144.
10. Zhang, P.; Han, L.; Gao, R.; Zhang, F.; Xing, Z. Frequency extension and robust full-waveform inversion based on nth power operation. *Acta Geophys.* **2020**, *68*, 1317–1333. [CrossRef]
11. Dong, X.; Li, Y. Denoising the Optical Fiber Seismic Data by Using Convolutional Adversarial Network Based on Loss Balance. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 10544–10554. [CrossRef]
12. Dong, X.; Lin, J.; Lu, S.; Huang, X.; Wang, H.; Li, Y. Seismic Shot Gather Denoising by Using a Supervised-Deep-Learning Method with Weak Dependence on Real Noise Data: A Solution to the Lack of Real Noise Data. *Surv. Geophys.* **2022**, *43*, 1363–1394. [CrossRef]
13. Wang, H.; Lin, J.; Dong, X.; Lu, S.; Li, Y.; Yang, B. Seismic velocity inversion transformer. *Geophysics* **2023**, *88*, R513–R533. [CrossRef]
14. Zhang, H.; Wang, W.; Wang, X.; Chen, W.; Zhou, Y.; Wang, C.; Zhao, Z. An implementation of the seismic resolution en-hancing network based on GAN. In Proceedings of the Society of Exploration Geophysicists International Exposition and Annual Meeting, San Antonio, TX, USA, 15–20 September 2020.
15. Sun, H.; Demanet, L. Extrapolated full-waveform inversion with deep learning. *Geophysics* **2020**, *85*, R275–R288. [CrossRef]
16. Sun, H.; Demanet, L. Extrapolated full waveform inversion with convolutional neural networks. In Proceedings of the Society of Exploration Geophysicists International Exposition and Annual Meeting, San Antonio, TX, USA, 15–20 September 2020.
17. Sun, H.; Demanet, L. Low frequency extrapolation with deep learning. In Proceedings of the 2018 SEG International Exposition and Annual Meeting, Anaheim, CA, USA, 14–19 October 2018.
18. Fang, J.; Zhou, H.; Li, Y.E.; Zhang, Q.; Wang, L.; Sun, P.; Zhang, J. Data-driven low-frequency signal recovery using deep-learning predictions in full-waveform inversion. *Geophysics* **2020**, *85*, A37–A43. [CrossRef]
19. Choi, Y.; Jo, Y.; Seol, S.J.; Byun, J.; Kim, Y. Deep learning spectral enhancement considering features of seismic field data. *Geophysics* **2021**, *86*, V389–V408. [CrossRef]
20. Nakayama, S.; Blacquière, G. Machine-learning-based data recovery and its contribution to seismic acquisition: Simulta-neous application of deblending, trace reconstruction, and low-frequency extrapolation. *Geophysics* **2021**, *86*, P13–P24. [CrossRef]
21. Hu, W.; Jin, Y.; Wu, X.; Chen, J. Progressive transfer learning for low-frequency data prediction in full-waveform inver-sion. *Geophysics* **2021**, *86*, R369–R382. [CrossRef]
22. Jin, Y.; Hu, W.; Wang, S.; Zi, Y.; Wu, X.; Chen, J. Efficient Progressive Transfer Learning for Full-Waveform Inversion with Extrapolated Low-Frequency Reflection Seismic Data. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5908810. [CrossRef]
23. Zhang, H.; Yang, P.; Liu, Y.; Luo, Y.; Xu, J. Deep Learning-Based Low-Frequency Extrapolation and Impedance Inversion of Seismic Data. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 7505905. [CrossRef]
24. Ni, J. Prediction of High and Low Frequency Earthquakes Based on Neural Network. Master's Thesis, Northeast Petroleum Uni-versity, Daqing, China, 2023.
25. Zhang, Y.; Zhou, Y.; Song, L.; Dong, H. Low frequency continuation of seismic data based on physically constrained U-Net network. *Oil Geophys. Prospect.* **2023**, *58*, 31–45.
26. Yin, Y. The Study of Joint Processing Method of Seismic Data from Multiple Sources Based on Convolutional Neural Net-Works. Ph.D. Thesis, Jilin University, Changchun, China, 2023.

27. Wang, T. Study on the Elimination and Migration of Surface Related Multiple by Sparse-Constrain Inversion. Ph.D. Thesis, Jilin University, Changchun, China, 2017.
28. Vaswani, A.; Shazeer, N.; Parmar, N. Attention Is All You Need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 6000–6010.
29. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2021**, arXiv:2010.11929.
30. Jiang, K.; Han, Q.; Du, X.; Ni, P. Structural dynamic response reconstruction and virtual sensing using a sequence to se-quence modeling with attention mechanism. *Autom. Constr.* **2021**, *131*, 103895. [CrossRef]
31. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer us-ing Shifted Windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.
32. Schuster, M.; Paliwal, K. Bidirectional Recurrent Neural Networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681. [CrossRef]
33. Cho, K.; Van Merrienboer, B.; Gulcehre, C.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *arXiv* **2014**, arXiv:1406.1078.