De-aliased high-resolution Radon transform based on the sparse prior information from the convolutional neural network

Luyu Feng 1,2, Yaru Xue 1,2,*, Chong Chen 1,2, Mengjun Guo 1,2 and Hewei Shen 1,2

1 College of Information Science and Engineering, China University of Petroleum-Beijing, Beijing, 102249, China
2 State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum-Beijing, Beijing, 102249, China

*Corresponding author: Yaru Xue. E-mail: xueyaru@cup.edu.cn

Received 25 November 2021, revised 22 April 2022
Accepted for publication 20 May 2022

Abstract
The resolution of Radon transform is crucial in seismic data interpolation. The high-frequency components usually suffer from serious aliasing problems while the sampling is insufficient. Constraining high-frequency components with unaliased low-frequency components is an effective method for improving the resolution of seismic data. However, it is difficult to obtain high-resolution low-frequency Radon coefficients by traditional analytical methods due to the strong correlation of basis functions. For this problem, a sparse inversion method using the neural network is proposed. First, the convolution model is deduced between the conjugated Radon solution and its ground truth. Then, a convolutional neural network (CNN), with the conjugate Radon solution as input, is designed to realize the deconvolution from the conjugate solution to the sparse and high-resolution Radon solution. Finally, the obtained sparse solution is regarded as prior knowledge of the iteratively reweighted least-squares algorithm. The proposed strategy has a distinct advantage in improving the resolution of low-frequency components, which helps overcome the aliasing. Interpolation experiments on synthetic and field data demonstrate the de-aliased performance of this CNN-based method.

Keywords: de-aliased Radon transform, convolutional neural network, sparse inversion, data reconstruction

1. Introduction
Radon transform (RT) is widely used in seismic data processing, such as data interpolation (Wang et al. 2010; Xue et al. 2014; Shao et al. 2017), multiple attenuation (Wang 2003; Li & Lu 2014; Xue et al. 2016; Ma et al. 2020) and deblending (Xue et al. 2017; Li & Sacchi 2020). A high-resolution RT is very important for accurate seismic processing in exploration geophysics (Cary 1998). However, the limited and discrete acquisition aperture of seismic data results in low resolution and aliasing problems for the traditional RT (Herrmann et al. 2000; Wang 2002). Thus, the sparse solution via inversion becomes a hot topic in seismic data processing.

To achieve a high-resolution RT, it is generally cast as an inverse problem with sparse constraints. Thorson & Claerbout (1985) proposed the basic theory of stochastic inversion of RT, which laid the foundation for the subsequent sparse RT. Hampson (1986) suggested calculating the Radon coefficients in the frequency domain, which greatly improved the computational efficiency. Its inverse matrix is fixed in the calculation process of the least square method. Therefore, its
ability to improve resolution is limited. Based on the Bayes’ theory, Sacchi & Ulrych (1995a) introduced prior knowledge as sparse constraints in the frequency domain, and this high-resolution method has been widely used in seismic signal processing.

Since these stochastic inversion ideas were proposed, many algorithms have been carried out to solve the inverse problem. The iteratively reweighted least-squares (IRLS) method (Scales & Gershenkorn 1988) is the most commonly used optimization method for Radon coefficient estimation, which has a good focusing ability (Sacchi & Ulrych 1995b). During the iteration process, the weighted matrix is updated continuously to improve the resolution by using the results of the previous iteration (Sacchi & Ulrych 1995a).

Considering the spike noise, Ibrahim & Sacchi (2019) introduced a robust RT with the IRLS algorithm to estimate Radon coefficients. The IRLS algorithm was also involved in the work of deblending with a high blending factor and AVO-preservation Radon coefficient estimation (Lin & Sacchi 2020). These studies have been carried out in the frequency domain and the regularization only posed on the curvature parameters, which resulted in the low resolution in the time direction. For this problem, Trad et al. (2003) proposed a mixed time-frequency RT that can effectively suppress multiples, using the IRLS algorithm to estimate the Radon coefficients in the time domain. In the time domain, the dimension of inverse matrix is very large, which leads to high computational costs.

The iterative shrinkage thresholding algorithm is another class of the commonly used method in sparse inversion procedures because of its simplicity of the calculation (Daubechies et al. 2004). Lu (2013) introduced an iterative 2D model shrinkage algorithm to solve the mixed time-frequency RT, which realized the sparse regulation in the time domain and the forward and inverse RT in the frequency domain. This combination not only solves the sparsity constraint in both time and curvature directions, but also reduces computational costs. Later on, Zhang et al. (2013) and Zhang & Lu (2014) extended the 2D algorithm to 3D RT for 3D seismic data interpolation and noise attenuation. With consideration of spike noise, Wang et al. (2019) explored the robust RT based on the iterative threshold algorithm. Gong et al. (2016) achieved prestack data interpolation, using the fast iterative thresholding algorithm to deal with the anisotropic RT.

Decreasing the dimension of inversion parameters is a practical way to improve the computational efficiency. Greedy algorithms or basis pursuit fall into this category (Chen & Donoho 1994; Ng & Perz 2004; Donoho & Elad 2006; Perrone et al. 2020). Wang et al. (2010) and Wang & Ng (2009) introduced a greedy least-squares inversion method to realize the local linear RT. Cao & Ross (2017) implemented 3D high-resolution RT through the matching pursuit algorithm to achieve seismic data interpolation. Xue et al. (2014, 2017) integrated the gradient and curvature parameters of AVO into the RT and adopted the greedy algorithm to overcome the high-dimension inverse parameters in this high-order sparse RT, which could maintain amplitude-preserving information well in interpolation and simultaneous separation experiments.

The RT is an underdetermined inverse problem (Sacchi & Ulrych 1995b), which usually requires numerous time-consuming iterations and hyperparameters setting using this analytical scheme. In addition, sparse offset sampling may cause spatial aliasing, which would pose large challenges for iterative algorithms. An effective way to avoid aliasing is to estimate high-frequency Radon coefficients using unaliased low-frequency components. Herrmann et al. (2000) proposed a non-iterative, de-aliased high-resolution RT, which constrained the current frequency by the Radon coefficients of the previous frequency. This non-iterative low-frequency constraint scheme improves both resolution and computational efficiency. The study by Hugonnet et al. (2018) showed that the de-aliased RT strategy can also be used for 3D RT. This de-aliased RT was also used by Ma et al. (2020) to reduce the effect of inadequate sampling in the crossline direction. At low frequencies, however, the strong correlation in the basis functions further limits the convergence of the analytical method. Thus, it is necessary to explore a new method to avoid these disadvantages while ensuring accuracy.

In fact, neural networks are gradually drawing much attention to high-resolution inversion in various industries, such as medical CT imaging (Gupta et al. 2017; Sun & Guo 2019), remote sensing (Del Carmen Valdes & Inamura 2000; Yuan et al. 2021), computational microscopy (Kamilov et al. 2015; Nehme et al. 2018) and so on. The nonlinear representation ability of neural networks facilitates the solution of the issues mentioned previously. In addition, some iterative inversion algorithms are mapped to the network structure, and the inverse solution is realized through parameter learning (Gregor & LeCun 2010; Yang et al. 2016; Borgerding et al. 2017). In seismic exploration, Zhang et al. (2020) designed a fully convolutional network to achieve seismic full-waveform inversion. Kaur et al. (2020a) adopted the Generated Adversarial Network to characterize inverse Hessian to reduce the influence of the blurring kernel and improve the resolution of the migrated image. They also introduced deep neural networks to implement hyperbolic RT, which achieved a solution equivalent to least squares and reduced the computational cost (Kaur et al. 2020b). These studies have proved that neural networks take advantage of the data-driven nature to improve resolution.

In this paper, a de-aliased high-resolution RT is proposed, by solving the sparse Radon coefficients at low frequencies with a neural network. In section 2, the IRLS algorithm is re-
viewed to solve parabolic RT. Then we establish the inversion model by deducing the relationship between the conjugate solution of RT and the true Radon coefficients. The low resolution of the conjugate solution is caused by the convolution kernel, especially at low frequencies. Therefore, a one-dimensional convolutional neural network (1D CNN) is used to implement this inversion process from the conjugate solution to the sparse solution. Finally, the output of the network is regarded as prior knowledge for the IRLS algorithm to constrain other frequencies. This high-resolution RT inversion scheme based on 1D CNN has a distinct advantage in anti-aliasing. In section 3, we analyze the network mechanism using the low-frequency series inversion. The interpolation experiments show that the proposed method can effectively reconstruct decimated traces and suppress aliasing.

2. Theory

2.1 The parabolic Radon transform

The adjoint parabolic RT in the time domain is a projection transform, which maps seismic data into the Radon domain along the parabolic trajectory. For one specific frequency component, the RT can be expressed in matrix form (Sacchi & Ulrych 1995b):

\[ D = LM, \]

(1)

where \( D \) is original data, \( M \) is the Radon model to be solved. The adjoint Radon operator \( L \) represents the integral along the path \( \tau = t - qx \)., and it can be represented by

\[
L = \begin{bmatrix}
  e^{-j\omega q_1 x_1^2} & e^{-j\omega q_2 x_2^2} & \cdots & e^{-j\omega q_N x_N^2} \\
  e^{-j\omega q_1 x_1^2} & e^{-j\omega q_2 x_2^2} & \cdots & e^{-j\omega q_N x_N^2} \\
  \vdots & \vdots & \ddots & \vdots \\
  e^{-j\omega q_1 x_{N_x}^2} & e^{-j\omega q_2 x_{N_x}^2} & \cdots & e^{-j\omega q_N x_{N_x}^2}
\end{bmatrix}
\]

(2)

Here, \( q \) is the trajectory curvature, \( N_q \) is the number of \( q \), \( x \) is the offset and \( N_x \) is the number of \( x \). The conjugate solution (Thorson & Claerbout 1985) is

\[ \tilde{M} = L^H D. \]

(3)

Because of the nonorthogonality of the operator, it is not possible to express its inverse by using the conjugate operator \( L^H \) (Trad et al. 2003). Therefore, the result obtained from equation (3) is only a rough approximation (Luo et al. 2008).

For higher resolution, the RT is usually taken as a regularization optimization problem. The objective function is defined as

\[ J(M) = \| D - LM \|_2^2 + \lambda G(M), \]

(4)

where \( G(M) \) is the regular term, \( \lambda \) is the regularization coefficient. Generally, the \( L_2 \) norm regularization achieves a smooth solution, and the \( L_1 \) norm or the Cauchy norm gets a sparse solution. When the \( L_1 \) norm is taken, a solution to equation (4) is

\[
M_{n+1} = \left( L^H L + \frac{1}{|M_n| + \sigma^2} \right)^{-1} L^H D.
\]

(5)

This iterative solution method is called the IRLS method (Scales et al. 1988), whose performance largely depends on the result of the previous iteration \( M_n \). So, high-resolution prior information can help both improve resolution and converge efficiently.

2.2 Inversion model based on CNN

In most cases, this IRLS method can obtain high-resolution Radon coefficients. But at low frequencies, the performance of this method is limited. Next, a novel inversion model is constructed to solve the sparse solution.

Submitting equation (1) to equation (3), the relation between conjugate solution and ground truth is achieved,

\[ \tilde{M} = L^H LM. \]

(6)

Let \( \Gamma = L^H L \), and be characterized by

\[
\Gamma = \begin{bmatrix}
  N_x & \sum \delta(\omega q_1) x_1^2 & \sum \delta(\omega q_2) x_2^2 & \cdots & \sum \delta(\omega q_N) x_N^2 \\
  \sum \delta(\omega q_1) x_1^2 & N_x & \sum \delta(\omega q_2) x_2^2 & \cdots & \sum \delta(\omega q_N) x_N^2 \\
  \vdots & \vdots & \ddots & \vdots & \vdots \\
  \sum \delta(\omega q_1) x_{N_x}^2 & \sum \delta(\omega q_2) x_{N_x}^2 & \cdots & N_x & \sum \delta(\omega q_N) x_{N_x}^2
\end{bmatrix}
\]

(7)

Because \( \Gamma \) is the autocorrelation matrix of the Radon operator \( L \), it forms a conjugate symmetric matrix, namely, the Hermitian matrix, which can be described as a blurring convolution operator (Hu et al. 2013). The convolution operator is constructed by the first column and row of \( \Gamma \). Thus, the matrix form of equation (6) can be reduced to a 1D convolution model:

\[ \tilde{M} = \Lambda \ast M. \]

(8)
Figure 1. Convolution kernels of different frequencies.

where * denotes convolution and

\[
\Lambda = \left[ \sum_{x_i} e^{i\omega(N_q-1)\Delta q x_i^{\perp}} \sum_{x_i} e^{i\omega(N_q-2)\Delta q x_i^{\perp}} \ldots N_x \right] \ldots \sum_{x_i} e^{i\omega(2-N_q)\Delta q x_i^{\perp}} \sum_{x_i} e^{i\omega(1-N_q)\Delta q x_i^{\perp}} \right]. \quad (9)
\]

Now high-resolution Radon coefficient estimation is transformed from the conjugate solution to the sparse solution by a 1D convolution instead of from the data domain to the Radon domain.

The convolution kernel is a mixed-phase wavelet. Parameters such as frequency and spatial sampling interval of seismic data will affect the convolution kernel. Figure 1 depicts the real parts of the convolution kernel with different frequencies. The lower the frequency, the more seriously blurred happen. This brings much difficulty for the low-frequency inversion.

To be sure, most deconvolution is realized by using matrix inversion like equation (5). We instead regard it as 1D convolution in equation (8) to facilitate the processing and analysis in the following CNN inversion. Since CNN was proposed, it has attracted extensive attention from researchers (Lecun et al. 1998; Szegedy et al. 2015; Krizhevsky et al. 2017). As a deep learning model, CNN has the ability of hierarchical learning features (Zeiler & Fergus 2014; Abadi et al. 2016) and gets the sparse representation of signals (Dong et al. 2016; Zhu et al. 2017).

To achieve high resolution, 1D CNN is designed to complete the inversion process of the RT as shown in figure 2. It includes nine convolutional layers and a fully connected layer. Each convolution layer has 16 convolution kernels with the size of 3 × 1, and a rectified linear unit function following it. The zero-padding approach is taken to maintain the input size and output size consistent after each convolution. Convolution operation and nonlinear activation function ensure feature extraction of input data. The fully connected layer integrates the highly abstract features after multiple convolutions.

To implement this inversion process, the real part and the imaginary part of complex numbers are connected in series and fed to the network for training. The outputs of the network are then combined into complex numbers. 1D CNN is trained by supervised learning. The loss function is defined as

\[
f = \| \mathbf{M} - \hat{\mathbf{M}} \|_2^2, \quad (10)
\]

which ensures the convergence of the network. Here, \( \hat{\mathbf{M}} \) is the output of the network. With the deepening of the neural network layer, the blurred Radon coefficients become more and more sparse, and the neural network learns to achieve high-resolution inversion. After training the CNN, the conjugate

Figure 2. Block diagram of the 1D CNN.
Figure 4. Synthetic data. (a) Theoretical synthetic profile. (b) Theoretical Radon model. (c) f-k spectrum of (a). (d) Synthetic gather with 66% traces regularly decimated. (e) f-k spectrum of (d).

Figure 5. Series inversion: (a), (b) and (c) represent the three lots of test data input to the network; (d), (e) and (f) represent the corresponding network output.
solution of low resolution can be directly fed into the neural network and obtain the sparse Radon coefficients.

2.3. The de-aliased Radon inversion based on CNN constrain

Next, to achieve the de-aliased Radon inversion, the frequency-constrained approach based on the sparse solution is used. The frequency constraint method is effectively applied in RT. The non-iterative high-resolution RT (NIHRT) method adopts the high-quality dominant frequency result as prior to constrain the other frequency data inversion (Chen & Lu 2011). In de-aliased, high-resolution RT method, the constraints are bootstrapped from the low-frequency inversion result to high frequency (Herrmann et al. 2000). Here, for anti-aliasing, the similar strategy is used and only one unaliased low-frequency component needs to be trained. Then, the output of the network is embedded
Figure 8. Radon models of regularly decimated data estimated by (a) the IRLS method, (b) the NIHRT method and (c) the proposed method.

Figure 9. Reconstruction results of regularly decimated data. (a) Data reconstructed by the IRLS method. (b) Data reconstructed by the NIHRT method. (c) Data reconstructed by the proposed method. (d) Error profile of (a). (e) Error profile of (b). (f) Error profile of (c).
in the IRLS algorithm (Gersztenkorn et al. 1985; Scales & Gersztenkorn 1988), as the prior knowledge to constrain the inversion matrix of other frequencies:

$$M^*_{ij} = \left( L^H_{ij}L_{ij} + \lambda \frac{1}{|F(M^*_j, \Phi)|} + \epsilon^2 \right)^{-1} L^H_{ij}D_j,$$

(11)

where $F(M, \Phi)$ is the output of the network, $\Phi$ represents trained network parameters, $f_0$ represents an unaliasing low frequency and $\epsilon^2$ is the stability factor to avoid the denominator being 0. This frequency constraint method ensures that the strong part energy is stronger, and the weak part energy is further weakened. As a result, the low-frequency high-resolution Radon coefficients constraint scheme can promote the coefficients sparsity of other frequencies, which can deal with the aliasing problem.

The entire workflow of the proposed RT inversion scheme is given in figure 3. The ground truth $M_{i_0}$ and the conjugate solution $\tilde{M}_{i_0}$ of decimated data are synthesized and used as data pairs to train the network. Through training, the error between network output $\tilde{M}_{i_0}$ and $M_{i_0}$ is reduced. The mapping relationship between $M_{i_0}$ and $\tilde{M}_{i_0}$ is embedded in the network parameter $\Phi$. Once the network training is completed, the conjugate Radon solution $\tilde{M}_{i_0}$ of the test data obtained by equation (3) can be fed into the network. The output of the 1D CNN $F(\tilde{M}_{i_0}, \Phi)$ is then embedded into the IRLS algorithm to estimate the final Radon profile according to equation (11). The main computational burden of this RT method lies in the training of the low-frequency components.

3. Examples

In this part, the performance of the CNN-based inversion scheme for RT is shown. We demonstrate the effectiveness of the network using the low-frequency series inversion and analyze the network mechanism. The de-aliased interpolation experiments are carried out on synthetic and field data using the CNN-based method.

3.1. Low-frequency high-resolution inversion using CNN

An example of synthetic data with seven events is given first to demonstrate that CNN can achieve high-resolution inversion for low-frequency data, as shown in figure 4a. The theoretical Radon model is displayed in figure 4b, with the dominant frequency of 29 Hz. The number of time samples is 401 with $t \in [0, 1.6]$ s, and the number of curvature samples is 120 with $q \in [-0.3, 0.8]$ s. The $f-k$ spectrum of the original synthetic profile is shown in figure 4c. After 66% of the traces are regularly decimated (figure 4d), there is obvious spatial aliasing in its $f-k$ spectrum, as depicted in figure 4e. To suppress aliasing, we select the Radon coefficients of 5 Hz to provide prior information. Usually, the inversion at low frequency is more difficult to converge than at high frequency. Here the CNN helps us to improve this challenging task.

To train the network, a dataset is generated according to the parameters of the theoretical Radon model. We generate 33,000 Radon profiles with Ricker wavelets of 29 Hz. The amplitude of wavelets is between 0.5–1. There are seven events in each Radon profile. The 5 Hz coefficients $M_{i_0}$ of these models are used as ground truth. The corresponding synthetic profiles are decimated 66% traces, and then their 5 Hz conjugate solutions are fed into the network: 30,000 data pairs constitute the training data and the other 3000 data pairs serve as the test data. We set the batch sizes to 32 and the epochs to 300. The Adam
Figure 11. \( f-k \) spectrum of figure 9: (a)–(c) are obtained by the IRLS, NIHRT and proposed method. (d), (e) and (f) are residuals of (a), (b) and (c), respectively (residuals are magnified by five times).

The algorithm is used to optimize networks, and the learning rate is 0.001.

This trained network can achieve a sparse solution, which is demonstrated with three samples selected randomly from the test set. Figure 5 parts a–c show their conjugate results, which is obvious that the resolution is very low and cannot distinguish the different events. Since the data are in the frequency domain, the vertical coordinates indicate their modulus and the horizontal coordinates indicate the curvature. We feed these samples into the trained network and get the output, as shown in figure 5 parts d–f with the red line. They show good sparsity, and the estimated result in the red line is almost the same as the ground truth in the blue line. Even adjacent curvature events can be clearly separated, as shown in figure 5d.

Then, the trained network is used to process the data in figure 4d. The conjugate solution of 5 Hz is calculated by equation (3) and fed into the trained network to get the sparse solution. To analyze the mechanism of the convolutional network more clearly, the feature maps of the second, fifth and ninth convolutional layers are extracted and depicted in figure 6 parts a–c. With the deepening of the convolution layers, the features gradually become concentrated. This variation indicates that CNN can capture the sparse characteristics hidden in the data. Finally, the fully connected layer is used to select the features for linear characterization, as shown by the black line in figure 6d. The results of the 11th iteration of IRLS are also plotted in this figure. Figure 1 has shown that the low-frequency convolution kernel brings serious blur in the conjugate solution. Even so, our network still shows good performance at the low frequency, which outweighs the analytical IRLS method.

Figure 7a displays the estimated 5 Hz data by performing the forward RT on Radon coefficients, and the errors are presented in figure 7b. The fidelity of the estimated data is higher than that of the IRLS method due to the higher accuracy of...
Figure 12. Noisy, decimated data. (a) Synthetic example with conflicting events and (b) its f-k spectrum. (c) Profile with noise and 66% traces regularly decimated and (d) its f-k spectrum.

the Radon coefficients obtained by the neural network. Owing to the non-uniqueness of the underdetermined problem, IRLS can estimate the data within a reasonable error. However, due to the strong correlation of the low-frequency basis function, it is hard to obtain sparse solutions, as shown in figure 6d. As a result, it is not possible to provide appropriate prior for data where spatial aliasing occurs. This conclusion is further confirmed in the following interpolation results and their f-k spectra.

3.2. Reconstruction of regularly decimated traces

The interpolation is then carried out using the neural-network-based method. All other frequency components are obtained directly using the network output as prior according to equation (11). Figure 8 parts a–c show the Radon panels obtained by IRLS, NIHRT and the proposed method.

To overcome aliasing, in the iterative process of the IRLS method, the constraints are bootstrapped from the low-frequency inversion result to high-frequency components. In the NIHRT method, data at 5 Hz are iteratively constrained...
Figure 14. Reconstruction results of noisy decimated data. (a) Data reconstructed by the IRLS method. (b) Data reconstructed by the NIHRT method. (c) Data reconstructed by the proposed method.

Table 1. SNR of noisy, decimated data interpolation

<table>
<thead>
<tr>
<th>Different algorithm</th>
<th>IRLS</th>
<th>NIHRT</th>
<th>Proposed method</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNR (dB) of t-x domain</td>
<td>6.4061</td>
<td>5.3007</td>
<td>10.1358</td>
</tr>
</tbody>
</table>

to all other frequency components. Obviously, our method can obtain a much sparser Radon model. The corresponding interpolation results and residual profiles are shown in figure 9 parts a–f. Some residuals are still visible in the IRLS and NIHRT results, while few residuals are in the proposed method. We extract the 70th trace to compare the waveforms of interpolation results, which is presented in figure 10a. The red line represents the waveform of the original trace. The blue, green and black lines represent the waveforms of the interpolation results obtained by the IRLS algorithm, the NIHRT algorithm and the proposed method. The errors in figure 10b illustrate that the CNN-based method can reconstruct the decimated data more accurately.

The f-k spectra are shown in figure 11 parts a–c to further analyze the performance of different methods. For a clear comparison, the residuals of the f-k spectra are magnified by five times, as shown in figure 11 parts d–f. The IRLS method and NIHRT method lose more energy in the reconstruction process than the proposed method, which corresponds to the residual event information in figure 9d and e. The proposed method retains the energy of the original data while suppressing the artifacts, and the reconstruction effect is obviously better than the other two methods. Unaliased low-frequency sparse Radon coefficients facilitate accurate interpolation.

Figure 15. f-k spectrum of figure 14. (a)–(c) are obtained by the IRLS, NIHRT and proposed method, respectively.
3.3. Reconstruction of noisy regularly decimated traces

The proposed method also shows good anti-noise performance. Figure 12a is a synthetic example with conflicting events, and its f-k spectrum is shown in figure 12b. Random band-limited noise is added to the synthetic data, and 66% of traces are regularly decimated, as shown in figure 12c. Both aliasing and noise exist in its f-k spectrum, which is presented in figure 12d.

The trained network is not suitable for figure 12c because its data characteristics are different from the previous dataset. To avoid retraining the network, we adopt the transfer learning strategy. First, a small sample dataset is generated according to the method mentioned before. Unlike the previous dataset, there are four events in each Radon profile. The 5 Hz coefficients $M_{j_0}$ of these models are used as ground truth. Then, the profiles are synthesized according to these Radon models. Random band-limited noise is added to these profiles, 66% of traces are regularly decimated, and then their 5-Hz conjugate solutions are fed into the network. The training set includes 500 pairs of data, and the test set includes 100 pairs of data. The previous trained network is...
regarded as a pre-trained model. These small sample data are used to fine-tune the full connection layer of the pre-trained model, and other training parameters remain unchanged. The output from the network is shown in figure 13. The accurate output of the network can be achieved only by adjusting the full connection layer of the network with small samples, which greatly reduces computational costs.

Data reconstruction results are shown in figure 14. Compared with figure 14a and b, the events in figure 14c are more complete and clearer. We calculate the recovered signal to noise ratio (SNR) of profiles to conduct quantitative comparisons according to

\[
    \text{SNR} = 20 \log_{10} \frac{||x_{\text{true}}||_2}{||x_{\text{true}} - x_{\text{est}}||_2},
\]

where \(x_{\text{true}}\) is the ground truth, and \(x_{\text{est}}\) is the estimated data. The SNR of noisy, regularly decimated profile is 1.8427 dB. After reconstruction, as shown in Table 1, the proposed method has a higher SNR than the IRLS algorithm and the NIHRT method. Figure 15 shows the f-k spectra corresponding to figure 14, as marked by ellipses, manifesting that the CNN-based method is valid in suppressing spatial aliasing and noise.

3.4. Reconstruction of irregularly decimated traces

The field seismic data are often irregularly sampled because of acquisition conditions, which can also cause aliasing problems. The field data (figure 16a) derived from the Gulf of

![Figure 17. Data augmentation. (a) Radon model estimated by the least square method: (b) and (c) are different Gauss operators, and (d) and (e) are the new Radon models synthesized by (b) and (c), respectively.](image_url)

![Figure 18. Estimated 5 Hz data from irregularly decimated data. (a) Estimated data by IRLS method and 1D CNN. (b) Error.](image_url)
Mexico (http://seismiclab.physics.ualberta.ca/) contains 92 traces. Figure 16b exhibits its f-k spectrum. We demonstrate the performance of the algorithm with seismic data that randomly decimated 32 traces (figure 16c) with SNR being 4.0840 dB. There are fewer aliasing artifacts at low frequencies in its f-k spectrum compared to higher frequencies, as depicted in figure 16d. Therefore, the 5 Hz data are still used to be sparse prior in interpolation experiments.

Due to lack of field datasets, data augmentation is required. First, the Radon model of decimated data is obtained by applying the least square method, shown in figure 17a. From the estimated Radon model, it is found that the main energy distribution is near the regions with curvatures of 0 and 0.5 s, while the other parts have relatively low energy distribution. Using these two curvature points as centers, the Gaussian function is generated, and they are connected in series:

\[
R_{m,n} = \begin{bmatrix}
e^{-\frac{(q-c_1)^2}{2\sigma_1^2}} & e^{-\frac{(q-c_2)^2}{2\sigma_2^2}}
\end{bmatrix},
\]

where center point \(c_1 = 3.5\), \(c_2 = 78\), ensuring that the two peaks of the Gaussian operator are at 0 and 0.5 s, respectively. Operators of different shapes can be constructed by randomly setting standard deviation \(\sigma\), such as figure 17b and c. The standard deviations of the two Gaussian functions in figure 17b are 6.8159 and 4.1038, respectively, and 7.23 and 7.51 in figure 17c.

Then data are augmented by the Gaussian operator:

\[
S_{m,n,w} = w(R_{m,n} \circ S_o) + (1-w)S_o,
\]

where \(S_o\) denotes the least-squares Radon model, \(w\) is the weight and its value can be 0–1. The diversity of the generated Radon models can be ensured by adjusting the weights \(w\) and standard deviations in \(R_{m,n}\). Figure 17d and e are

Figure 19. Interpolation of irregularly decimated data. Interpolation result using the (a) IRLS, (b) NIHRT and (c) CNN-based method. (d) Error of (a). (e) Error of (b). (f) Error of (c).
the new Radon models synthesized by two operators in figure 17b and c, respectively. Generated by Gaussian operators, the Radon models are sparse.

Seismic profiles are generated from these Radon models. These profiles are then decimated in the following method:

\[ T_\Pi(X) = \begin{cases} 
0, & (m, n) \in \Pi \\
X_{m,n}, & \text{otherwise} 
\end{cases} \]  

(15)

where \( \Pi \) represents the index subset corresponding to the decimated traces.

Then we fine-tune the pre-trained model using the target-domain data. 1000 data pairs constitute the training data, and 300 data pairs serve as the test data. The weights and other parameters of the convolution layer remain unchanged, and only the parameters of the full connection layer are updated.

Figure 18a shows the modulus of the estimated 5 Hz data. The ground truth is represented by the red line, the data estimated by the network are indicated by the black line and the data estimated by the IRLS method are displayed by the blue line. Figure 18b displays the residuals, with less error in the network estimation compared to the IRLS algorithm.

Figure 20. The f-k spectrum of figure 19: (a)–(c) are obtained by the IRLS, NIHRT and CNN-based method, and (d), (e) and (f) are errors of (a), (b) and (c), respectively (errors are magnified twice).

<table>
<thead>
<tr>
<th>Different algorithm</th>
<th>IRLS</th>
<th>NIHRT</th>
<th>Proposed method</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNR (dB) of t-x domain</td>
<td>10.3409</td>
<td>11.4384</td>
<td>12.8543</td>
</tr>
</tbody>
</table>

Table 2. SNR of irregularly decimated data interpolation
Figure 19a–c are the profiles reconstructed by IRLS, NIHRT and the CNN-based method, and their errors are depicted in figure 19d–f. When the interval of the missing trace is too large, as marked by ellipses, the IRLS and NIHRT method cannot reconstruct the decimated data. The reconstruction effect of the CNN-based method is improved obviously. Table 2 lists the SNR of three methods, indicating that the interpolation effect of the proposed method is superior to the other two methods.

Figure 20 parts a–c show the $f-k$ spectra of profiles estimated by the three algorithms, and their errors are shown in figure 20d–f. The profile estimated by the proposed method has less aliasing, especially for the elliptically marked parts. We conclude that this is because the conjugate solution of low-frequency data is blurred seriously, as illustrated in equation (6), so it is difficult to realize inversion by analytical methods. The network can estimate low-frequency coefficients with a high resolution due to the nonlinear characterization ability. This can provide appropriate prior information for other frequency components, which ensures the accuracy of interpolation.

Figure 21 shows the SNR at different frequencies of the reconstructed profile, where the black line represents the CNN-based method. Compared with the other two algorithms, the SNR of the CNN-based method is significantly improved between 17 and 50 Hz, because the aliasing of this profile occurs mainly in this frequency range (see figure 16d).

4. Conclusions

Anti-aliasing is the key to successful application of the RT in decimated data reconstruction. We propose a high-resolution RT inversion scheme based on a convolution neural network. Driven by data, the trained neural network can extract sparse features. The experiments demonstrate that the proposed CNN method can provide a more high-resolution RT than the traditional IRLS method. This method has special advantages on low-frequency data sparse inversion, which has good application potential in seismic data dealiasing. The experiments prove that the CNN-based method achieves considerable performance on seismic data interpolation.

Acknowledgements

This work is supported by PetroChina Innovation Foundation (grant no. 2020D-5007-0301).

Conflict of interest statement. None declared.

References


Daubechies, I., Defrise, M. & De Mol, C., 2004. An iterative thresholding algorithm for linear inverse problems with a sparsity con-


