Petroleum Science 21 (2024) 2345–2355

Contents lists available at ScienceDirect

Petroleum Science

journal homepage: www.keaipublishing.com/en/journals/petroleum-science

Original Paper

Application of sparse S transform network with knowledge distillation in seismic attenuation delineation

Nai-Hao Liu [a], Yu-Xin Zhang [b], Yang Yang [a, *], Rong-Chang Liu [c], Jing-Huai Gao [a], Nan Zhang [d]

[a] School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an, 710049, Shaanxi, China
[b] School of Software Engineering, Xi'an Jiaotong University, Xi'an, 710049, Shaanxi, China
[c] PetroChina Research Institute of Petroleum Exploration and Development (RIPED), CNPC, Beijing, 100083, China
[d] Research Institute of Exploration and Development, Yumen Oilfield Company, CNPC, Jiuquan, 735019, Gansu, China

ARTICLE INFO

Article history:
Received 4 May 2023
Received in revised form
6 February 2024
Accepted 5 March 2024
Available online 19 March 2024

Edited by Jie Hao and Meng-Jiao Zhou

Keywords:
S transform
Deep learning
Knowledge distillation
Transfer learning
Seismic attenuation delineation

ABSTRACT

Time-frequency analysis is a successfully used tool for analyzing the local features of seismic data. However, it suffers from several inevitable limitations, such as the restricted time-frequency resolution, the difficulty in selecting parameters, and the low computational efficiency. Inspired by deep learning, we suggest a deep learning-based workflow for seismic time-frequency analysis. The sparse S transform network (SSTNet) is first built to map the relationship between synthetic traces and sparse S transform spectra, which can be easily pre-trained by using synthetic traces and training labels. Next, we introduce knowledge distillation (KD) based transfer learning to re-train SSTNet by using a field data set without training labels, which is named the sparse S transform network with knowledge distillation (KD-SSTNet). In this way, we can effectively calculate the sparse time-frequency spectra of field data and avoid the use of field training labels. To test the availability of the suggested KD-SSTNet, we apply it to field data to estimate seismic attenuation for reservoir characterization and make detailed comparisons with the traditional time-frequency analysis methods.

1. Introduction

Time-frequency analysis (TFA) is widely used for seismic processing and interpretation (Zhao and Song, 2012), which can describe local features of seismic signals. The traditional time-frequency (TF) transforms can be easily divided into linear and non-linear categories. Continuous wavelet transform (CWT), S transform (ST), and short-time Fourier transform (STFT) are representative of the linear TF transform, which has received a lot of attention in recent years. Although STFT and CWT are often used in many seismic tasks and have achieved good performance (Lu and Li, 2013), they still have their unavoidable shortcomings. For instance, STFT usually utilizes a fixed window to analyze non-stationary seismic signals and obtains TF features with fixed TF resolution. Besides, STFT, ST, and CWT are all limited by the Heisenberg uncertainty principle. The non-linear transforms can be divided into several categories, such as quadratic transforms (Wigner, 1997; Alsalmi and Wang, 2021), reassigned transforms (Li et al., 2022a; Wu and Liu, 2009), and sparse transforms (Gholami, 2013; Yuan et al., 2020), which are often limited by the low computational efficiency and the difficulty in selecting parameters.

ST was first proposed by Stockwell et al. (1996), regarded as the succession and development of wavelet transform (WT) and STFT. On the one hand, ST can perform multi-resolution analysis similarly to WT. On the other hand, ST has the capability of single-frequency independent analysis as STFT. Moreover, ST is a tool for both signal analysis and signal synthesis in addition to being a linear reversible time-frequency analysis method. Lately, ST has been applied in many fields, including exploration geophysics (Fang et al., 2021), mechanical engineering (Song et al., 2020), clinical medicine (Chen and Zhang, 2020), etc. Inspired by ST, generalized S transform (GST) was proposed and discussed (Adams et al., 2002). GST successfully breaks the limitations on the form of the window function of ST, which has better frequency aggregation ability. Several excellent versions of GST are proposed and used in seismic exploration,

* Corresponding author.
E-mail address: yang_yang@mail.xjtu.edu.cn (Y. Yang).

https://doi.org/10.1016/j.petsci.2024.03.002

containing seismic time-frequency analysis (Wang et al., 2018), fluvial channel detection (Liu et al., 2019), instantaneous frequency estimation (Lin and Meng, 2011), seismic deconvolution (Zhou et al., 2014), etc. Whereas, it is difficult to choose the optimized parameters for GST. Nowadays, sparse transform has become a new way for improving the resolution of time-frequency analysis (Wang et al., 2016) and many sparse time-frequency analysis methods have been proposed (Orović et al., 2015; Jokanovic and Amin, 2015). Based on the idea of sparsity and ST, sparse S transform (SST) is also suggested for describing seismic time-frequency features. However, since SST often has more parameters than ST, the issue of choosing optimum parameters is inescapably presented in SST, if anything it is worse. Therefore, it usually takes a long time to calculate the corresponding SST spectrum, which is another obvious disadvantage (Yang et al., 2022).

Recently, deep learning (DL) has shown a powerful ability to learn mapping relationships between training data and training labels (Zhong et al., 2022a). For seismic exploration, DL has been successfully used for different scenes, such as fault interpretation (Wu et al., 2019; Li et al., 2022b), lithology prediction (Liu et al., 2020, 2022a), seismic data reconstruction (Kaur et al., 2020; Liu et al., 2022d), noise separation and attenuation (Zhong et al., 2022b; Liu et al., 2022b, 2022c; Dong et al., 2022), horizon picking (Tschannen et al., 2020; Wu et al., 2022), etc. As one of the standard DL methods, supervised deep learning can be applied to solve many complex tasks that need enough training data and corresponding labels. However, it should be noted that the labels for field data are difficult to obtain, meaning that this disadvantage limits its wide applications. To break this bottleneck, transfer learning (TL) becomes an efficient way, which has achieved more and more attention for addressing seismic tasks (Siahkoohi et al., 2019), including seismic fault detection (Cunha et al., 2020), seismic phase picking (Chai et al., 2020), and seismic dip estimation (Ao et al., 2022). The main idea of TL is transferring the pre-trained model from the source domain to the target domain. There are lots of training data and training labels for pre-training a DL model in the source domain. On the contrary, the target domain includes data with a few labels or even without labels. Compared with traditional machine learning, TL emphasizes exacting knowledge learned from the source domain to help the learning process in the target domain. The mentioned knowledge consists of network parameters, data characteristics, and other helpful information. One of the schemes for effectively exacting knowledge is knowledge distillation (KD), proposed for transferring the knowledge from the teacher network to the student network (Hinton et al., 2015). The teacher network is a complex model with superior predictive accuracy, while the student network is often a simple and low-complexity model. The knowledge extracted from the teacher network can be regarded as the flow of the solution procedure (Yim et al., 2017), which is beneficial for training the student network. In other words, knowledge distillation can be regarded as a constraint to help the network learn features and improve accuracy. Sometimes, for unsupervised learning, knowledge distillation can replace labels to help the model achieve good performance and break the limit of labels.

Inspired by KD and SST, we suggest a sparse S transform network with knowledge distillation (KD-SSTNet). We first propose a supervised deep learning network for implementing SST, named the sparse S transform network (SSTNet). To train SSTNet, we build synthetic training data and training labels as the training data set. Here, the synthetic model is generated by well logs and horizons at the study survey, which enhances the similarity between synthetic data and field data. After pre-training SSTNet, we utilize KD to generalize the ability of SSTNet and further propose the sparse S transform network with knowledge distillation (KD-SSTNet). It should be noted that the second step is unsupervised transfer learning. It means that we use a field data set without training labels to train KD-SSTNet, which solves the lack of training labels for field data. To demonstrate the efficiency of KD-SSTNet, we finally apply it to seismic attenuation estimation and make detailed comparisons with the traditional methods. The whole deep learning workflow is shown in Fig. 1.

The main contributions of our work can be summarized as follows.

1. A supervised deep learning network is suggested for implementing SST, which is named the sparse S transform network (SSTNet). It can learn the mapping relationship between synthetic traces and SST spectra.
2. Inspired by KD, we propose the sparse S transform network with knowledge distillation (KD-SSTNet), which enhances the generalized ability of SSTNet.
3. A field data set without labels is utilized to train KD-SSTNet and validate its performance. Moreover, KD-SSTNet is utilized to characterize hydrocarbon reservoirs after estimating seismic attenuation.

The rest of this paper is organized as follows. We first introduce ST and its sparse variant. Then, we describe the structure of SSTNet and the generated training data set. Next, we introduce the details of KD-SSTNet, including the architecture, workflow and knowledge distillation. Afterward, we provide the application of KD-SSTNet in field data, proving its availability based on its accuracy and efficiency in seismic attenuation delineation. Finally, we have some discussions about the limitations and achievements of our work.

2. S transform and its sparse variant

2.1. S transform

Given a signal $s(t)$, its ST can be defined as

$$\begin{cases} ST(\eta,f) = \int_{-\infty}^{\infty} s(t)w(t)\mathrm{e}^{-2\pi ift}\mathrm{d}t, \\ w(t) = \dfrac{|f|}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(t-\eta)^2 f^2}{2}}, \end{cases} \tag{1}$$

where η is the localization of the Gaussian window w, t represents the time variable, and f is the frequency variable. Eq. (1) can be rewritten as

$$ST(\eta,f) = \int_{-\infty}^{\infty} \bar{s}(\beta+f)\mathrm{e}^{-\frac{2\pi\beta^2}{f^2}}\mathrm{e}^{2\pi i\beta\eta}\mathrm{d}\beta, \tag{2}$$

where $\bar{s}(\beta)$ denotes the Fourier transform of $s(\eta)$, β is the frequency variable.

Eq. (2) describes the ST in the frequency domain. By integrating the time-frequency spectrum over the time axis, we can obtain the correlation between the ST spectrum and the Fourier spectrum of $s(t)$ as

$$\int_{-\infty}^{\infty} ST(\eta,f)\mathrm{d}\eta = \bar{s}(f), \tag{3}$$

where $\bar{s}(f)$ represents the Fourier transform of $s(t)$. In this way, the given signal $s(t)$ can be reconstructed by using ST coefficients as

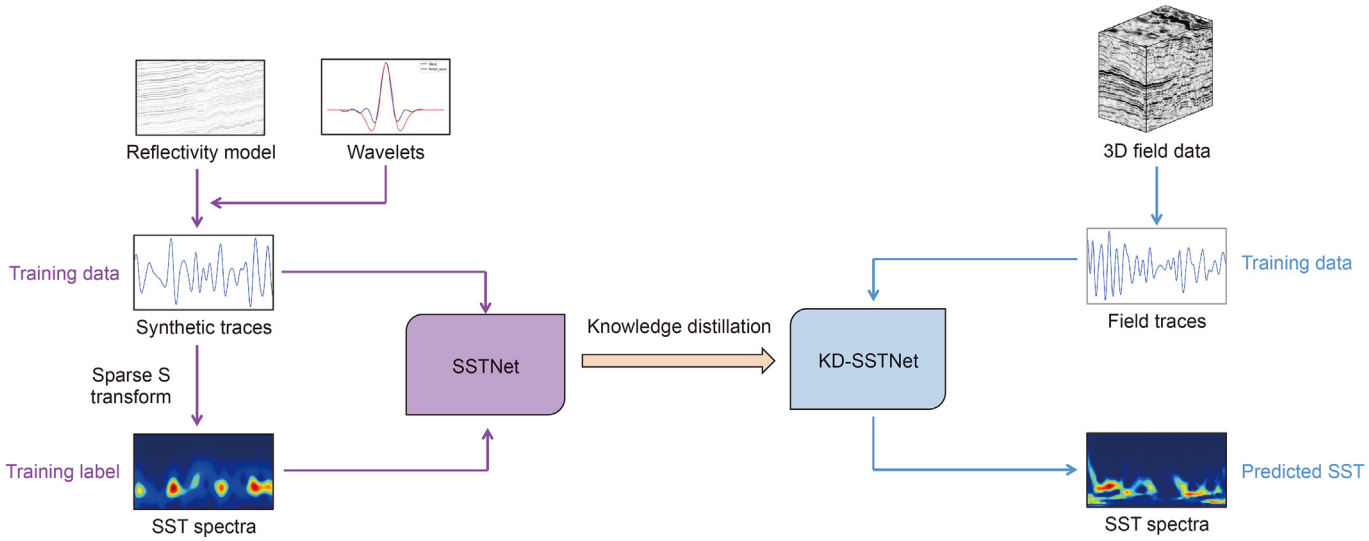

Fig. 1. The simplified workflow of the suggested KD-SSTNet.

$$s(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} ST(\eta, f)\mathrm{d}\eta e^{i2\pi ft}\mathrm{d}f. \tag{4}$$

2.2. Sparse S transform

Based on Eq. (2), the discrete version of the inverse S transform can be written as

$$\bar{s}[m] = \sum_{l=0}^{N-1}\sum_{k=0}^{N-1} ST(l,k)w[m+l-1]e^{-\frac{i2\pi mk}{N}}, \tag{5}$$

where $l = 0, 1, \ldots, N-1$ is the frequency index and k is the time index. w is the window function, and $\bar{s}[m]$ represents the Fourier transform of s. Then, the reconstructed signal can be obtained by inverse Fourier transform as

$$s[k] = \sum_{m=0}^{N-1} \bar{s}[m]e^{\frac{i2\pi mk}{N}}. \tag{6}$$

Due to the linearity of Eq. (5), we regard the time-frequency coefficients of ST as the vector $\mathbf{x}\in\mathbb{C}^{N^2\times 1}$ and set a basis matrix $\mathbf{G}\in\mathbb{C}^{N\times N^2}$. Then, Eq. (5) can be written as

$$\bar{\mathbf{s}} = \mathbf{G}\mathbf{x}, \tag{7}$$

where the matrix $\mathbf{G}$ is an over-complete dictionary, generated by the given window function w. If the signal and window function are given, the ST can be obtained by an inverse problem. Afterward, the objective function is denoted as

$$\min_{\mathbf{x}} \frac{1}{2}\|\bar{\mathbf{s}} - \mathbf{G}\mathbf{x}\|_2^2. \tag{8}$$

Furthermore, a regularization term is added to achieve a unique solution. In this way, Eq. (8) can be defined as

$$\min_{\mathbf{x}} \frac{1}{2}\|\bar{\mathbf{s}} - \mathbf{G}\mathbf{x}\|_2^2 + \alpha\theta(\mathbf{x}), \tag{9}$$

where α is a parameter for adjusting the proportion of the regularization term $\theta(\mathbf{x})$. Usually, we replace $\theta(\mathbf{x})$ with l_1-norm and Eq. (9) can be described as

$$\min_{\mathbf{x}} \frac{1}{2}\|\bar{\mathbf{s}} - \mathbf{G}\mathbf{x}\|_2^2 + \alpha\|\mathbf{x}\|_1, \tag{10}$$

where $\|\mathbf{x}\|_1 = \sum_{p=1}^{N}|x_p|$ and x_p means the pth elements of $\mathbf{x}$. Here, the regularization parameter α is approximately selected as $\sqrt{2\log N_{\text{length}}}\sigma$, where N_{length} is the length of the analyzed signal and σ is the standard deviation of noise (Herrera et al., 2014). Then, we obtain the sparse solution of ST, named sparse S transform (SST).

3. Sparse S transform network

3.1. SSTNet

The structure of the suggested sparse S transform network (SSTNet) is shown in Fig. 2. Here, the widely used Unet model is selected as our baseline. The whole structure can be divided into three parts, including the transformation layer, encoder layer, and decoder layer. The first part is a 1×3 convolution operation to map the input 1D seismic data into the 2D feature image. The following encoder layer includes the repeated 3×3 convolution, 2×2 max pooling operations, batch normalization, and rectified linear unit (ReLU). It should be noticed that each convolution uses a different channel number to extract characteristics in more depth. The framework of the decoder layer is similar to the encoder layer, whose modules contain the repeated 3×3 convolutions, 2×2 up-convolution, batch normalization, and ReLU. At the end of the decoder layer, there is a 1×1 convolution operation to output the final result trained by the whole network. Here, skip connections (copy and crop) are added between the encoder layer and the decoder layer, which can reduce spatial information loss. To be more precise, skip connections help the feature maps calculated by the decoder layer own more low-level semantic information, and make the final results more accurate.

3.2. Pre-training data set of SSTNet

The training data set is a critical factor for successfully implementing a DL model. To enhance the similarity of synthetic and

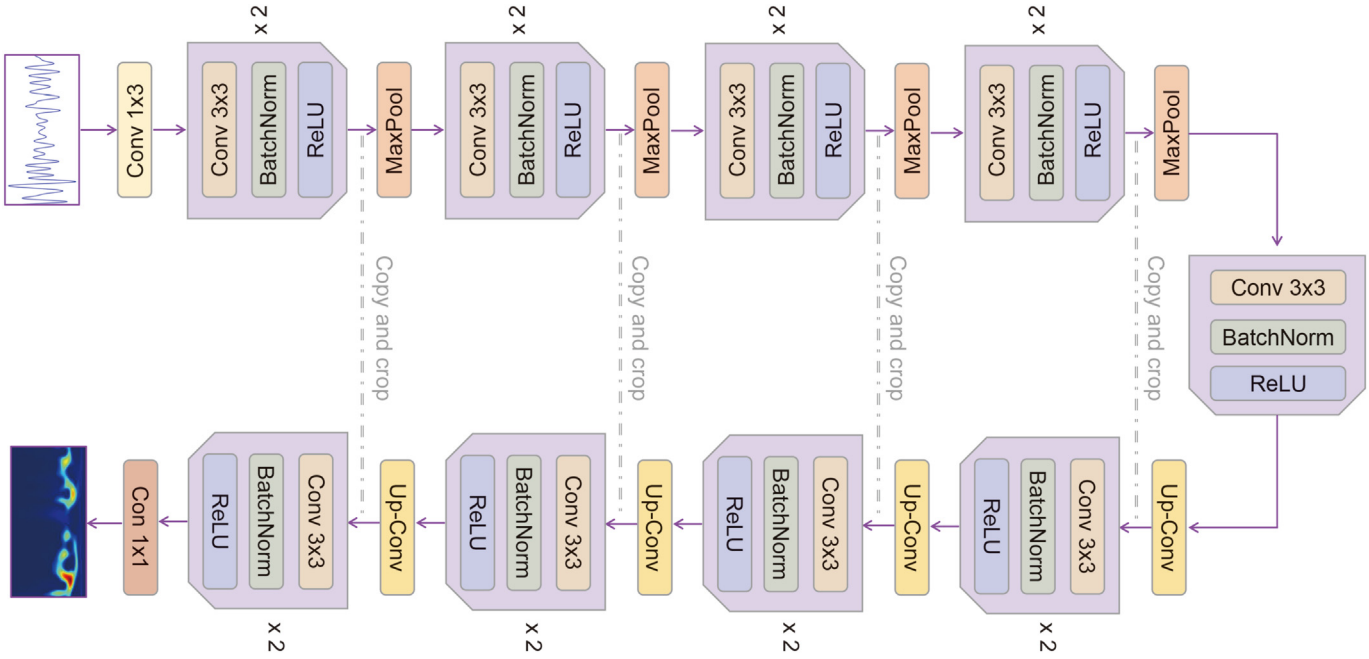

Fig. 2. The simplified architecture of the proposed SSTNet.

field data, we use well logs and horizons at the study area to build a synthetic model and then generate a synthetic data set. The dashed purple rectangle in Fig. 3 indicates the location of the study area, located in the Ordos Basin in northwest China. It is well known that abundant natural gas and oil are in the Ordos Basin. The reservoirs in the Triassic Yanchang Formation of the Ordos Basin have many characteristics, such as high cement content, great heterogeneity, poor sorting, fine particles of rock, and low maturity (Wang et al., 2007). Therefore, the Ordos Basin is one of the representative areas in which low permeability reservoirs developed in China (Yao et al., 2013).

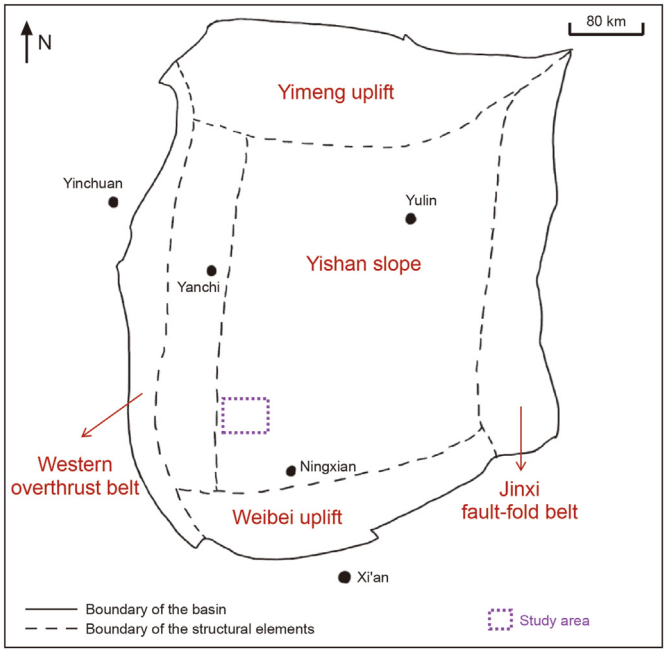

Fig. 3. The location of the study area, indicated by the dashed purple rectangle.

The Triassic Yanchang Formation consists of inland fluvial-delta-lacustrine clastic rock series (Fu et al., 2020), whose reservoir quality in the southern is poorer than the coeval strata of fluvial to deltaic systems in the north and northeast of the Ordos Basin (Xie, 2016). The Triassic can be divided into top, middle, and low Triassic. The top and middle Triassic are both composed of silty mudstone, sandstone interbedded with thin, arenaceous mudstone, and coal seams. The low Triassic includes sandstone with intact cross-bedding or conglomerates, some mud-clast conglomerates, and mudstone (Xie, 2016). The Triassic Yanchang Formation in the southern Ordos Basin is above the middle Triassic and is limited to low resolution and poor reservoir quality. Therefore, it is challenging to define the distribution of reservoirs, which is the problem that our suggested approach is intended to address.

The field data set used in this study mainly includes a 3D post-stack seismic volume, an interpreted horizon, and well logs. For the 3D seismic data volume, there are 801 crosslines and 251 inlines, and the time sampling interval is 1 ms. As mentioned above, the interpolation of well logs is utilized to generate the 2D geological model for building the synthetic training data set. The interpreted horizon T_0 with the structural information is used as a constraint, which can enhance the lateral resolution. Then, the geological reflectivity model is presented in Fig. 4 with 512 traces. The time sample interval is 1 ms and the trace interval is 20 m. We adopt the Ricker wavelet and the statistical wavelet extracted from this study area for implementing the convolution. The statistical wavelet is presented by the blue curve in Fig. 5, while the red curve denotes the simulated Ricker wavelet with the dominant frequency of 35 Hz. This demonstrates that the statistical wavelet in the study area can be accurately fitted by a simulated Ricker wavelet, proving the validity of selecting Ricker wavelets to generate the synthetic data set in this study. The dominant frequencies of Ricker wavelets used in generating synthetic data are from 20 Hz to 50 Hz, with a 5 Hz interval. In this way, we can utilize 7 Ricker wavelets and a statistical wavelet to convolve with the reflectivity sequences in Fig. 4. Afterward, we generate 1401 × 8 synthetic traces as the synthetic training data set.

As mentioned above, the selection of parameter α in Eq. (10) will

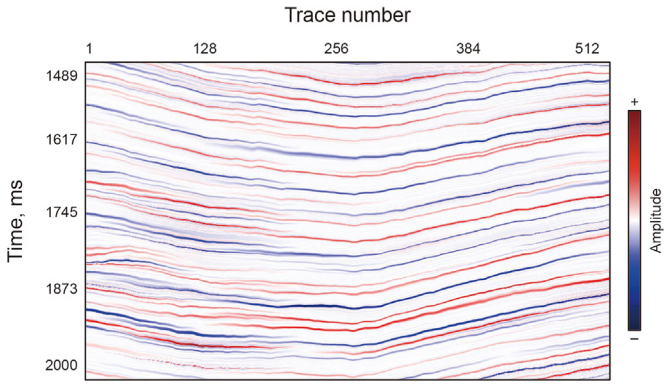

Fig. 4. The synthetic reflectivity model obtained from the geological model, including 512 traces and 512 time samples.

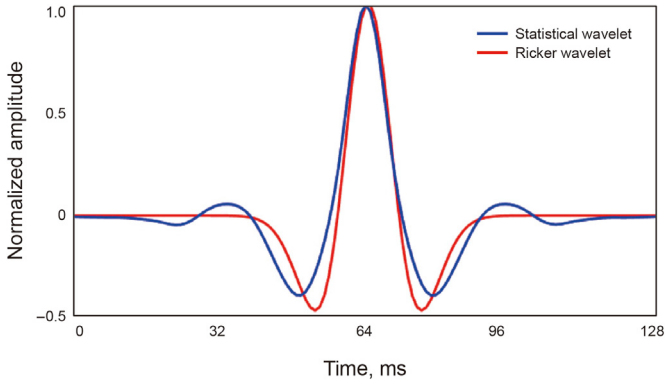

Fig. 5. Statistical wavelet (blue) and Ricker wavelet (red).

greatly influence the final SST spectrum. Fig. 6 depicts a synthetic trace and corresponding SST spectra with different α. When α is small, the calculated SST spectrum will become obscure and vague, as shown in Fig. 6(b). On the contrary, if α is too large in Fig. 6(d), the SST spectrum is too sparse to accurately describe the time-frequency feature. Fig. 6(c) describes the optimal SST spectrum with $\alpha = 0.5$, which was selected by several experiments. Choosing appropriate parameters is more complex for field traces than for synthetic data because of the existence of noise and the impact of seismic exploration factors. We need to implement plenty of manual experiments and take lots of time to find the matching regularization parameters. There is an obvious shortage of sparse-based TF methods, which is the main motivation for carrying out this work.

Before model training, we first normalize synthetic traces as

$$\mathbf{u_i}' = \frac{\mathbf{u_i}}{\max\limits_{1\le i\le N_{\text{trace}}, 1\le j\le N_t} u_{i,j}}, \tag{11}$$

where $\mathbf{u_i}$ and $\mathbf{u_i}'$ represent synthetic traces before and after standardization. N_{trace} indicates the trace number, and N_t is the time length of a synthetic trace. $\max\limits_{1\le i\le N_{\text{trace}}, 1\le j\le N_t} u_{i,j}$ is the maximum value of the whole synthetic data set. Afterward, we randomly select 3000 traces with their SST spectra to pre-train SSTNet. Here, we divide them into training data and validation data with a ratio of 8:2.

3.3. Pre-training details of SSTNet

In the process of model pre-training, to evaluate the performance of SSTNet, we utilize the mean squared error (MSE) to calculate the discrepancies between the predicted results and labels as

$$loss_{\text{MSE}} = \frac{1}{sum}\sum_{z=1}^{sum}(Label_z - Result_z)^2, \tag{12}$$

where *sum* is the total sample number of the training data. $Label_z$ denotes the labels of input data and $Result_z$ represents the corresponding results by the network. Besides, the Adam optimization

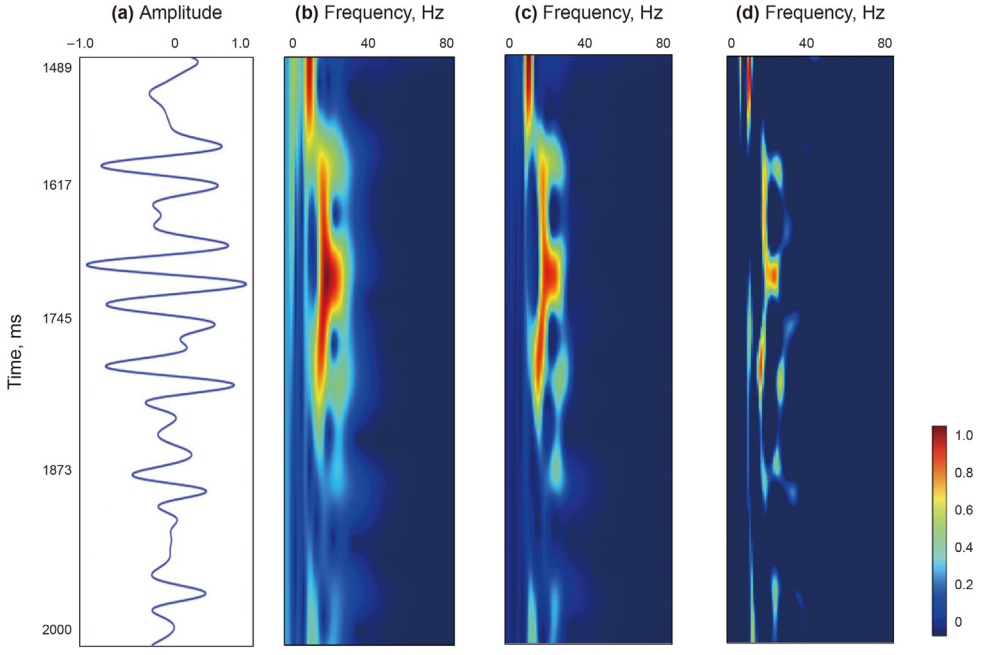

Fig. 6. **(a)** A random synthetic trace and the corresponding SST spectra with **(b)** $\alpha = 0.005$, **(c)** $\alpha = 0.5$, and **(d)** $\alpha = 50$.

algorithm is selected as the optimizer and the learning rate is set to 0.0005. It should be noted that all models run in a runtime environment with the PyTorch deep learning library on Python 3.7. The server provided for all computations has a 36-core processor, 128 GB RAM, and an NVIDIA GTX 3090 (24 GB GPU memory). The whole training time is 2936 s, i.e., approximately 49 min. Fig. 7 describes the training (blue) and validation (red) loss curves. It is obvious that the loss curves keep going down and finally converge. It is easy to conclude that SSTNet is well-trained and has high accuracy. To verify this, we select the extra 100 synthetic traces as the testing data set and show their results obtained using the proposed model in Fig. 8. Fig. 8(a) is a synthetic trace randomly chosen from the testing data set. Fig. 8(b)–(d) are its corresponding ST spectrum, SST label, and predicted SST spectrum using SSTNet. It can be found that the SST spectrum is more sparse than the ST spectrum, which can characterize time-frequency features of seismic data more clearly. Comparing the SST label with the predicted result using SSTNet, there are almost no discrepancies, meaning that the proposed model has accurately learned the mapping relationship between synthetic traces and SST spectra.

4. Sparse S transform network with knowledge distillation

4.1. KD-SSTNet

Based on the pre-trained SSTNet, we add a measure as the knowledge to distill, named the sparse S transform network with knowledge distillation (KD-SSTNet), whose structure is presented in Fig. 9.

We do not modify the modules of the whole structure, including the number of layers, the number of convolution kernels, and skip connections. The grey parts symbolize the freezing parameters in Fig. 9, which means using the fine-tuned parameters of SSTNet. On the contrary, the colorful parts in Fig. 9 need to be retrained in the target domain. The added measure can be written as

$$KD_{\text{target}} = \frac{\sum |F_t|^2}{\left|\left|\sum |F_t|^2\right|\right|_2} \tag{13}$$

where F_t represents the feature map in the target domain. $|\cdot|$ and $||\cdot||$ mean the corresponding absolute value and l_2 norm. Similarly, KD_{source} in Fig. 9 is the same calculation, whose feature map comes from the source domain. In other words, KD_{source} is the knowledge we distilled from the pre-trained SSTNet to help train KD-SSTNet in the target domain. More detailed information about knowledge distillation is described in the next subsection. It should be noted that we add this measure to the first and last convolution layers of the decoder layer. We have done comparative experiments to judge the quantity and location of this measure. The result shows that we added this measure in all four layers of the decoder layer, which is no different from the result we added only in the first and last layers. It means that putting this measure in all four layers is redundant.

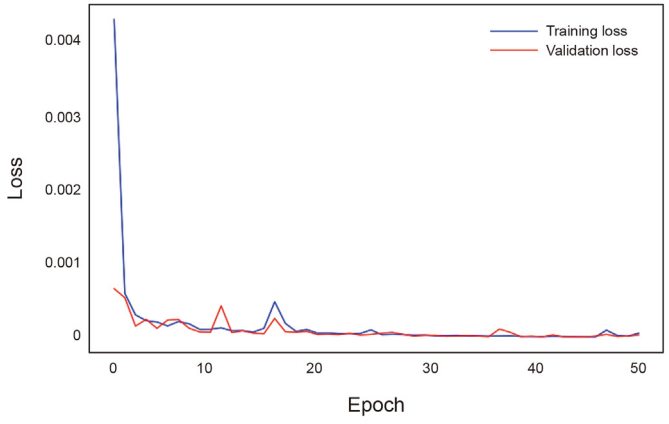

Fig. 7. Training (blue) and validation (red) loss curves.

4.2. Knowledge distillation

To obtain valuable knowledge, we adopt SSTNet with fine-tuned parameters to predict 50 synthetic traces. In this process, we record the measures in the first and last layers of the decoder layer of pre-trained SSTNet, which are indicated as KD_{source} in Fig. 9. The aim is to optimize the performance of KD-SSTNet in the target domain by replacing labels with prior knowledge in the source domain.

After acquiring the knowledge we distilled, we modify the loss function utilized in KD-SSTNet as

$$\begin{aligned} loss_{\text{KD}} &= \sum_{m=1}^{2} ||KD_{\text{target}} - KD_{\text{source}}||_2 \\ &= \sum_{m=1}^{2} \left|\left| \frac{\sum |F_t|^2}{\left|\left|\sum |F_t|^2\right|\right|_2} - \frac{\sum |F_s|^2}{\left|\left|\sum |F_s|^2\right|\right|_2} \right|\right|_2 \end{aligned} \tag{14}$$

where F_s represents the feature map in the corresponding convolution layer of SSTNet in the source domain. Here, we set the maximum of m as 2 because we only emphasize two convolution layers, as we described above.

4.3. Fine-tuning details of KD-SSTNet

We select field traces at Inline 2400–2403 and Inline 2647–2650 (total 6400 traces) as the fine-tuned data set. Next, we randomly select 5000 field traces without SST labels as fine-tuned training data. The other field traces, except for these 6400 traces, are built as the blind testing data set. The learning rate is 0.0001, and the number of iterations is 100. The whole training time is approximately 6 h. Fig. 10 presents the fine-tuned loss curve, defined in Eq. (14), which has a convergence tendency after 100 epochs.

To verify the efficiency of KD-SSTNet, we apply it to the blind testing data set. Fig. 11(a) depicts a field trace randomly selected from the blind testing data set. Fig. 11(b)–(e) represent their ST spectrum, SST label, and SST spectrum predicted using SSTNet and KD-SSTNet. The predicted SST spectrum is similar to the SST label and sparser than the ST spectrum, which can describe time-frequency features more accurately. Meanwhile, comparing the Fig. 11(d) and (e), we can find that the result predicted by the KD-SSTNet is sparser and more accurate than the spectrum calculated by the SSTNet. In other words, knowledge distillation has a positive impact on the result prediction. This also demonstrates the necessity and validity of introducing knowledge distillation in this study. Although KD-SSTNet takes a lot of computing time, it is the best performance we can achieve under the circumstance of no field labels.

To further demonstrate the efficiency of KD-SSTNet, we compare the calculation times of different methods for predicting 100 traces, as shown in Table 1. Note that we implement ST and SST by using an Inter(R) i9-10980XE CPU. Besides, we also utilize the Renyi entropy (RE) to evaluate the sparsity of TF spectra, written as

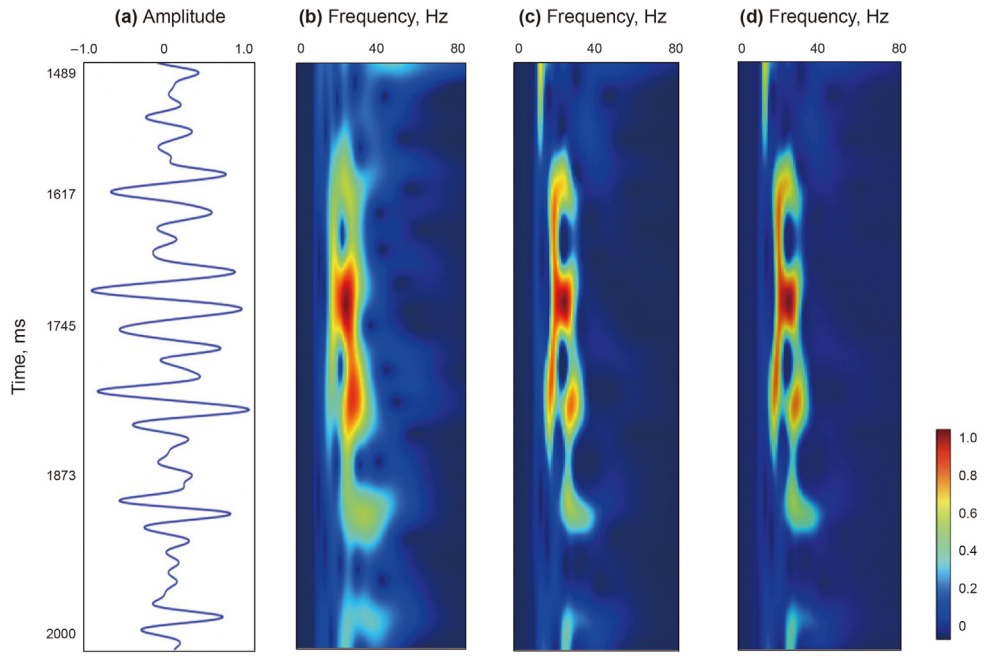

Fig. 8. **(a)** A synthetic trace randomly selected from the blind testing data set, **(b)** ST spectrum, **(c)** SST label, **(d)** SST spectrum predicted using SSTNet.

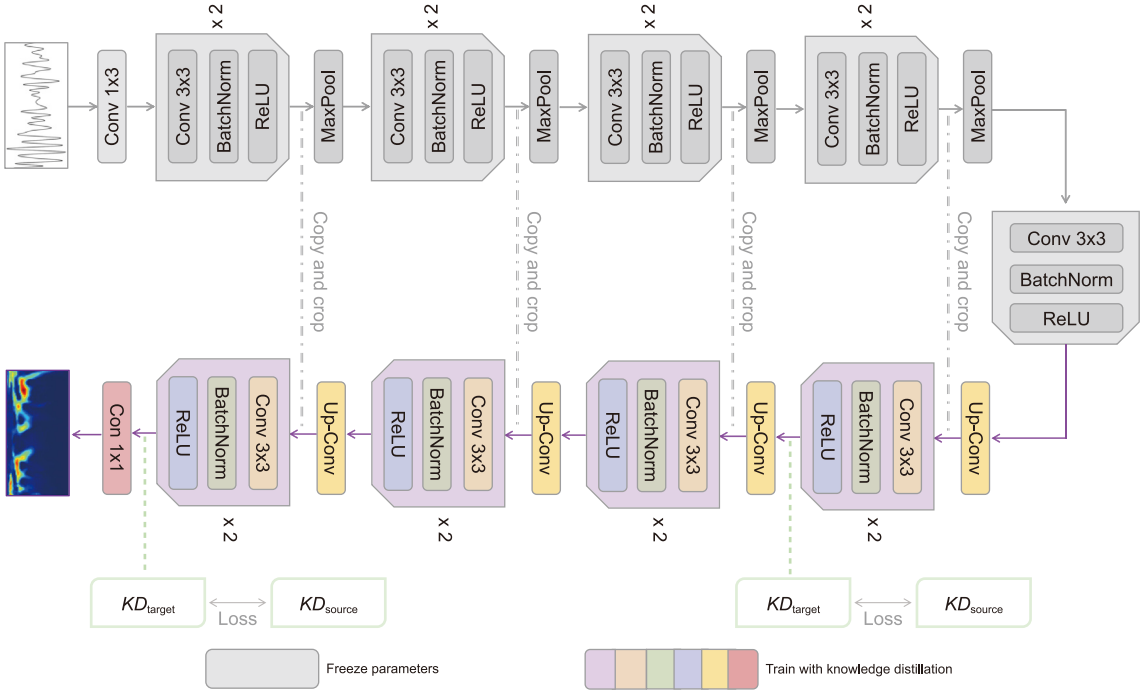

Fig. 9. The simplified architecture of the suggested KD-SSTNet.

$$RE = -\frac{1}{2}\log_2\left(\sum_{a=1}^{A}\sum_{b=1}^{B}\left[\frac{TF[a,b]}{\sum_{a=1}^{A}\sum_{b=1}^{B}|TF[a,b]|}\right]^{\delta}\right); \quad \delta > 2, \tag{15}$$

where $TF[\cdot]$ denotes the time-frequency spectrum, such as the ST spectrum and SST spectrum. $|TF[\cdot]|$ indicates its absolute norm. δ is selected as 3 in this example. The calculation time of KD-SSTNet is shorter than that of SSTNet and the RE of KD-SSTNet is smaller than that of SSTNet. Besides, the RE value of KD-SSTNet is a little smaller than that of SST, indicating that KD-SSTNet has excellent calculation efficiency and prediction ability.

5. Field data applications

Before validating the efficiency of KD-SSTNet, we first extract the amplitude attribute and several commonly used seismic attributes. Fig. 12 shows the horizontal amplitude surface and three

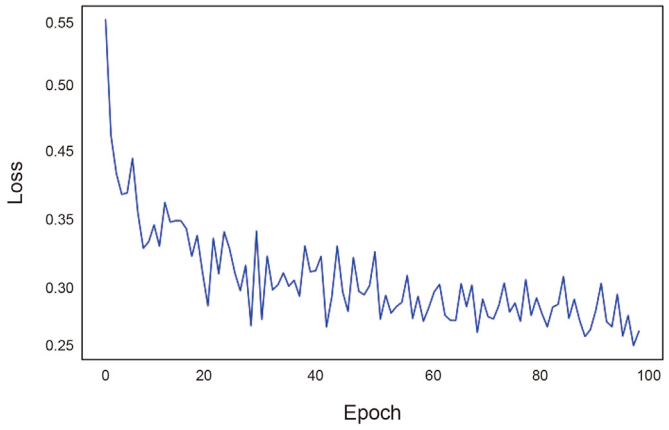

Fig. 10. The loss curve of fine-tuning KD-SSTNet.

attribute surfaces, i.e., (b) main frequency, (c) instantaneous frequency, and (d) weighted frequency. Here, the white dots represent the dry well boreholes at T_0 and the black dots denote the productive well boreholes with hydrocarbon reservoirs. It should be noted that we only show the attribute surface between Inline 2404 and Inline 2646 due to the selection of the fine-tuning data set (Inline 2400–2403 and Inline 2647–2650). By comparing these four images in Fig. 12, we cannot accurately distinguish the dry and productive wells, meaning that these attributes cannot be effectively utilized for reservoir characterization. Moreover, these attributes suffer from heavy noise contained in this 3D field volume.

After applying KD-SSTNet to the 3D field data, we extract the low- and high-frequency components. Then, we estimate seismic attenuation by comparing the difference between the low- and high-frequency components, denoted in Fig. 13(b). More details about the qualitative attenuation estimation can be easily found in Yang et al. (2021). Furthermore, we calculate the corresponding result using ST in Fig. 13(a) for a comparison. Note that we do not show the result calculated using SST due to its low computational

Table 1
The calculation time and Rayleigh entropy (RE) of different methods for predicting 100 field traces.

Methods	CPU/GPU time, s	RE ($\delta = 3$)
ST	33.789/–	13.732
SST	50.212/–	12.554
SSTNet	33.579/6.622	13.226
KD-SSTNet	33.124/6.454	12.495

efficiency and difficulty in parameter selection. As described above, the black dots (W4, W5, W6) in Fig. 13 represent the productive well boreholes with hydrocarbon reservoirs and the white dots (W1, W2, W3) denote the dry well boreholes at T_0. Comparing Fig. 13(a) and (b), we have several main observations. First, KD-SSTNet provides a smoother attenuation surface than ST, benefiting from its better anti-noise performance. This benefits seismic attenuation delineation and further reservoir characterization. Second, the interpretation results based on well logs seem to be unmatched by the attenuation surface computed using ST. For example, W3 shows the sandy-brown color (strong attenuation) in Fig. 13(a), indicating that there is an apparent hydrocarbon reservoir. This is inconsistent with the interpretation results based on well logs. Whereas, KD-SSTNet provides an attenuation surface, matched with all well boreholes at T_0, proving the effectiveness of the suggested model for delineating seismic attenuated reservoirs. Based on these images in Fig. 13 and the above discussions, it can easily conclude that KD-SSTNet can availably delineate seismic attenuation and then accurately describe hydrocarbon reservoirs, which demonstrates its availability and practicability in seismic exploration.

6. Discussions

We suggest a KD-SSTNet for implementing sparse time-frequency analysis of 3D field data at the Ordos Basin, which has been applied for seismic attenuation estimation and then reservoir characterization. There are two critical factors when implementing

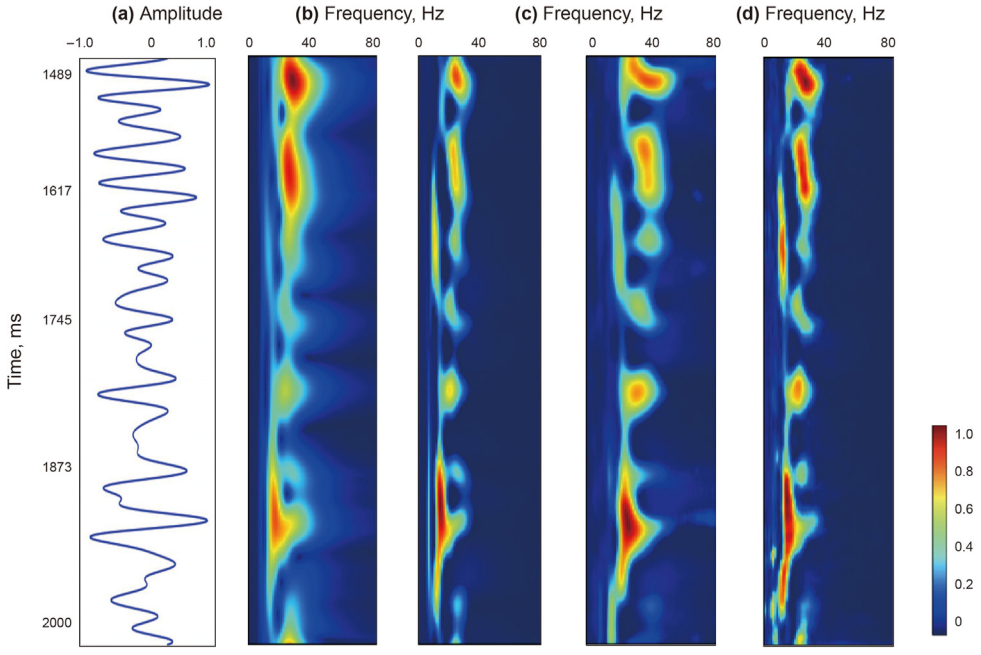

Fig. 11. **(a)** A field trace randomly selected from the blind testing data set, **(b)** ST spectrum, **(c)** SST label, TF spectra calculated using **(d)** SSTNet and **(e)** KD-SSTNet.

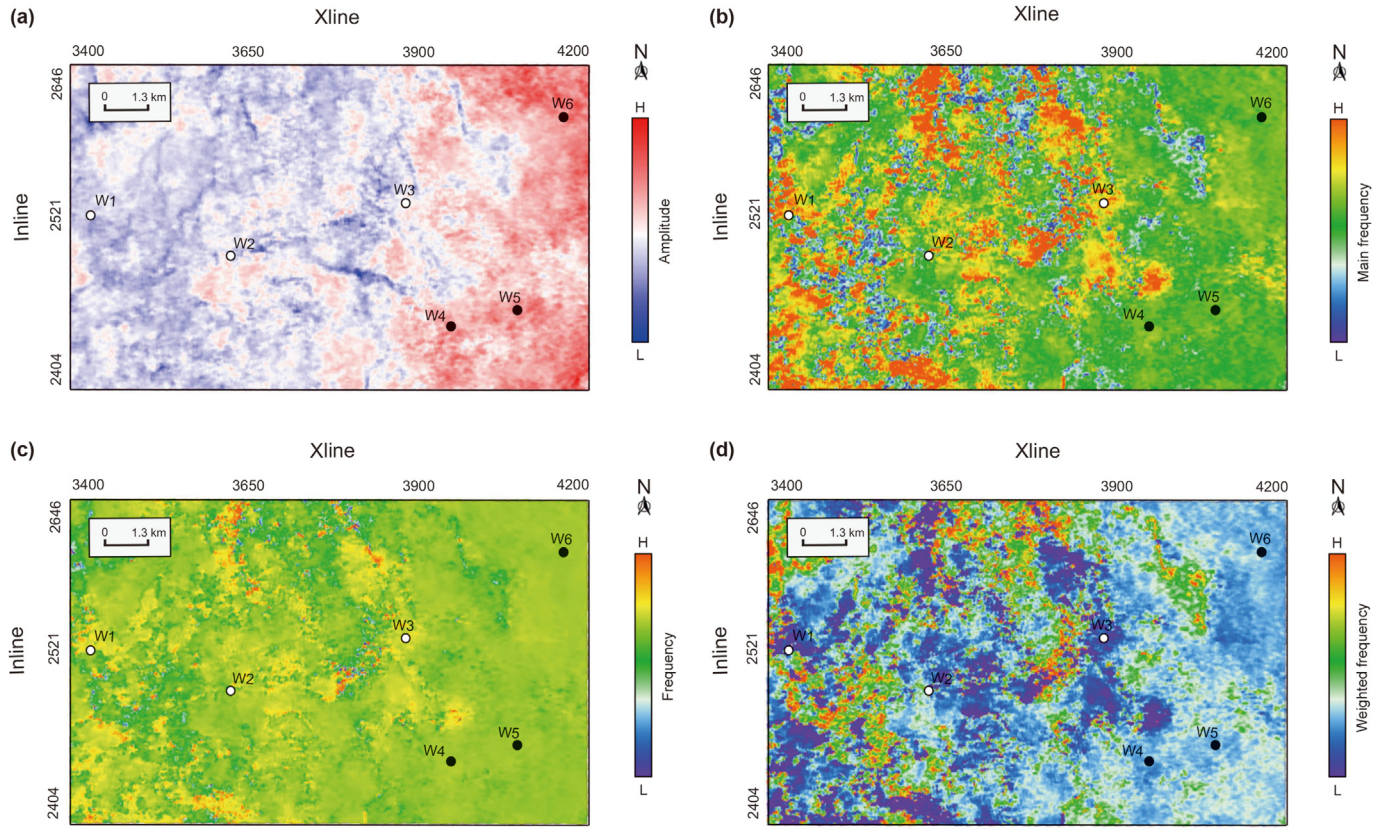

Fig. 12. **(a)** The horizontal amplitude surface and several commonly used attribute surfaces, **(b)** main frequency, **(c)** instantaneous frequency, and **(d)** weighted frequency. The white dots represent dry well boreholes and the black dots denote the productive well boreholes with hydrocarbon reservoirs.

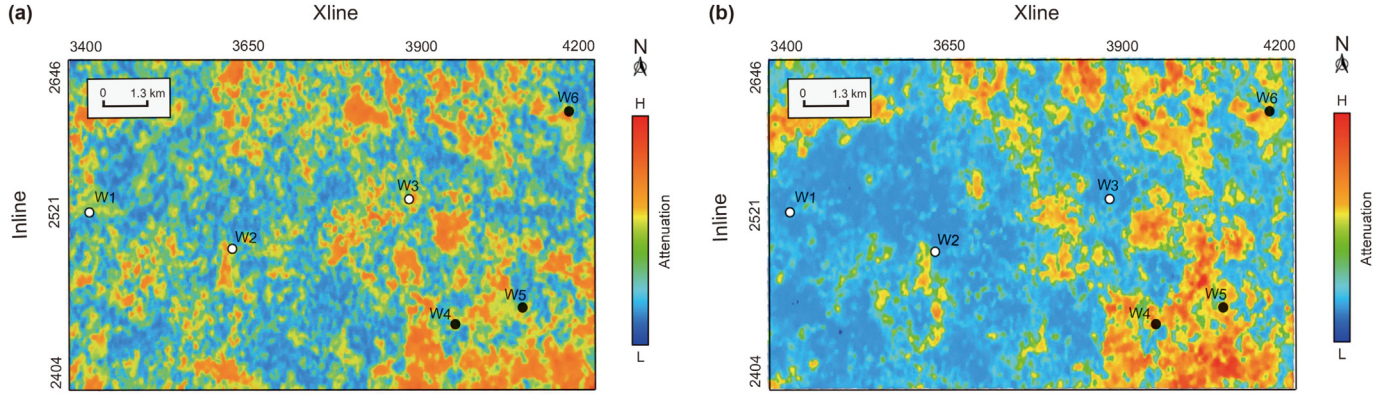

Fig. 13. Seismic attenuation surfaces computed using **(a)** ST and **(b)** KD-SSTNet. The white dots represent dry well boreholes and the black dots denote the productive well boreholes with hydrocarbon reservoirs.

KD-SSTNet, discussed as follows.

Basic time-frequency transform and baseline of SSTNet: SST is a well-liked and simple-to-implement time-frequency analysis tool for seismic interpretation. Therefore, we adopt it as the basic transformation in this study. Note that SSTNet is suggested to map a seismic trace to a time-frequency spectrum. Hence, other excellent time-frequency transforms can also be easily selected as the basic transform in this study, such as continuous wavelet transform and generalized S transform. For mapping synthetic traces to SST spectra, we select the commonly used UNet as the baseline in this study. Note that we did not pay much attention to the selection of the baseline. As we all know, there are several excellent DL models proposed for addressing different geological issues, such as residual networks, transformer networks, wavelet-based networks, etc. Exploring different performances when selecting different DL models must be interesting work in the future.

Pre-training and fine-tuned data sets: Except for the baseline of the suggested model, the pre-training data set is also a key factor for the success of the proposed workflow. Note that it needs the corresponding training labels when implementing model pre-training. Therefore, we propose to utilize a synthetic data set for model pre-training. Moreover, to reduce the variance between synthetic data and field data, we generate synthetic traces by using well logs and horizons at the study survey. When applying KD-

SSTNet to another seismic survey, we suggest generating the related synthetic pre-training data set, which is the main limitation of this study. After model pre-training, we utilize a few field traces without training labels for model fine-tuning. Here, knowledge distillation based transfer learning is proposed to implement model fine-tuning at the target domain. The advantage of KD is to relax the need for fine-tuned labels at the target domain. Certainly, its computational efficiency should be further enhanced.

7. Conclusions

We suggest a KD-SSTNet for seismic time-frequency analysis. First, we build SSTNet and utilize the synthetic data set, generated from well logs and interpreted horizons, to pre-train the proposed SSTNet. Next, we combine it with knowledge distillation to design KD-SSTNet. Afterward, we extract field traces from a 3D post-stack seismic volume without SST spectra labels to train KD-SSTNet. In this way, we can break the dependency of field data labels when training a DL model to solve geological issues. Finally, we apply KD-SSTNet to estimate seismic attenuation and compare it with traditional ST and commonly used seismic attributes. The attenuation slice calculated using KD-SSTNet can describe hydrocarbon reservoirs more accurately than the contrastive slices, which demonstrates the availability and practicability of the proposed KD-SSTNet in seismic exploration.

Declaration of competing interest

There is no other professional or personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.

Data availability statements

The data underlying this article will be shared on reasonable request to the corresponding author.

CRediT authorship contribution statement

Nai-Hao Liu: Writing – original draft, Methodology. **Yu-Xin Zhang:** Software, Methodology. **Yang Yang:** Writing – review & editing, Writing – original draft. **Rong-Chang Liu:** Data curation, Conceptualization. **Jing-Huai Gao:** Supervision. **Nan Zhang:** Validation, Conceptualization.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (42274144, 42304122, and 41974155), the Key Research and Development Program of Shaanxi (2023-YBGY-076), the National Key R&D Program of China (2020YFA0713404), and the China Uranium Industry and East China University of Technology Joint Innovation Fund (NRE202107). The authors deeply appreciate the valuable comments from editors and five anonymous reviewers.

References

Adams, M., Kossentini, F., Ward, R., 2002. Generalized s transform. IEEE Trans. Signal Process. 50, 2831–2842. https://doi.org/10.1109/TSP.2002.804085.

Alsalmi, H., Wang, Y., 2021. Mask filtering to the wigner-ville distribution. Geophysics 86, V489–V496. https://doi.org/10.1190/GEO2021-0193.1.

Ao, Y., Lu, W., Xu, P., Jiang, B., 2022. Seismic dip estimation with a domain knowledge constrained transfer learning approach. IEEE Trans. Geosci. Rem. Sens. 60, 1–16. https://doi.org/10.1109/TGRS.2021.3061438.

Chai, C., Maceira, M., Santos-Villalobos, H.J., Venkatakrishnan, S.V., Schoenball, M., Zhu, W., Beroza, G.C., Thurber, C., Team, E.C., 2020. Using a deep neural network and transfer learning to bridge scales for seismic phase picking. Geophys. Res. Lett. 47. https://doi.org/10.1029/2020GL088651 e2020GL088651.

Chen, P., Zhang, Q., 2020. Classification of heart sounds using discrete time-frequency energy feature based on s transform and the wavelet threshold denoising. Biomed. Signal Process Control 57, 101684. https://doi.org/10.1016/j.bspc.2019.101684.

Cunha, A., Pochet, A., Lopes, H., Gattass, M., 2020. Seismic fault detection in real data using transfer learning from a convolutional neural network pre-trained with synthetic seismic data. Comput. Geosci. 135, 104344. https://doi.org/10.1016/j.cageo.2019.104344.

Dong, X., Lin, J., Lu, S., Huang, X., Wang, H., Li, Y., 2022. 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. 43, 1363–1394. https://doi.org/10.1007/s10712-022-09702-7.

Fang, Y., Chen, H., Hu, Y., Li, R., Li, J., 2021. Application of adaptive parameterized s-transform to delta sandstone reservoir identification. Geophys. Prospect. 69, 1689–1699. https://doi.org/10.1111/1365-2478.13129.

Fu, J., Li, S., Niu, X., Deng, X., Zhou, X., 2020. Geological characteristics and exploration of shale oil in chang 7 member of triassic yanchang formation, ordos basin, nw China. Petrol. Explor. Dev. 47, 931–945. https://doi.org/10.1016/S1876-3804(20)60107-0.

Gholami, A., 2013. Sparse time–frequency decomposition and some applications. IEEE Trans. Geosci. Rem. Sens. 51, 3598–3604. https://doi.org/10.1109/TGRS.2012.2220144.

Herrera, R.H., Han, J., van der Baan, M., 2014. Applications of the synchrosqueezing transform in seismic time-frequency analysis. Geophysics 79, V55–V64. https://doi.org/10.1190/GEO2013-0204.1.

Hinton, G., Vinyals, O., Dean, J., et al., 2015. Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531 2.

Jokanovic, B., Amin, M., 2015. Reduced interference sparse time-frequency distributions for compressed observations. IEEE Trans. Signal Process. 63, 6698–6709. https://doi.org/10.1109/TSP.2015.2477056.

Kaur, H., Fomel, S., Pham, N., 2020. Seismic ground-roll noise attenuation using deep learning. Geophys. Prospect. 68, 2064–2077. https://doi.org/10.1111/1365-2478.12985.

Li, R., Chen, H., Fang, Y., Hu, Y., Chen, X., Li, J., 2022a. Synchrosqueezing polynomial chirplet transform and its application in tight sandstone gas reservoir identification. Geosci. Rem. Sens. Lett. IEEE 19, 1–5. https://doi.org/10.1109/LGRS.2021.3071318.

Li, S., Liu, N., Li, F., Gao, J., Ding, J., 2022b. Automatic fault delineation in 3-d seismic images with deep learning: data augmentation or ensemble learning? IEEE Trans. Geosci. Rem. Sens. 60, 1–14. https://doi.org/10.1109/TGRS.2022.3150353.

Lin, W., Meng, X., 2011. An adaptive generalized s-transform for instantaneous frequency estimation. Signal Process 91, 1876–1886. https://doi.org/10.1016/j.sigpro.2011.02.010.

Liu, N., Gao, J., Zhang, B., Wang, Q., Jiang, X., 2019. Self-adaptive generalized s-transform and its application in seismic time–frequency analysis. IEEE Trans. Geosci. Rem. Sens. 57, 7849–7859. https://doi.org/10.1109/TGRS.2019.2916792.

Liu, N., Huang, T., Gao, J., Xu, Z., Wang, D., Li, F., 2022a. Quantum-enhanced deep learning-based lithology interpretation from well logs. IEEE Trans. Geosci. Rem. Sens. 60, 1–13. https://doi.org/10.1109/TGRS.2021.3085340.

Liu, N., Wang, J., Gao, J., Chang, S., Lou, Y., 2022b. Similarity-informed self-learning and its application on seismic image denoising. IEEE Trans. Geosci. Rem. Sens. 60, 1–13. https://doi.org/10.1109/TGRS.2022.3210217.

Liu, N., Wang, J., Gao, J., Yu, K., Lou, Y., Pu, Y., Chang, S., 2022c. NS2NS: self-learning for seismic image denoising. IEEE Trans. Geosci. Rem. Sens. 60, 1–11. https://doi.org/10.1109/TGRS.2022.3217289.

Liu, N., Wu, L., Wang, J., Wu, H., Gao, J., Wang, D., 2022d. Seismic data reconstruction via wavelet-based residual deep learning. IEEE Trans. Geosci. Rem. Sens. 60, 1–13. https://doi.org/10.1109/TGRS.2022.3152984.

Liu, X.-Y., Zhou, L., Chen, X.-H., Li, J.-Y., 2020. Lithofacies identification using support vector machine based on local deep multi-kernel learning. Petrol. Sci. 17, 954–966. https://doi.org/10.1007/s12182-020-00474-6.

Lu, W., Li, F., 2013. Seismic spectral decomposition using deconvolutive short-time fourier transform spectrogram. Geophysics 78, V43–V51. https://doi.org/10.1190/GEO2012-0125.1.

Orović, I., Draganić, A., Stanković, S., 2015. Sparse time–frequency representation for signals with fast varying instantaneous frequency. IET Radar. Sonar & Navigation 9, 1260–1267. https://doi.org/10.1049/iet-rsn.2015.0116.

Siahkoohi, A., Louboutin, M., Herrmann, F.J., 2019. The importance of transfer learning in seismic modeling and imaging. Geophysics 84, A47–A52. https://doi.org/10.1190/GEO2019-0056.1.

Song, X., Zhao, J., Song, J., Dong, F., Xu, L., Zhao, J., 2020. Local demagnetization fault recognition of permanent magnet synchronous linear motor based on s-transform and pso–lssvm. IEEE Trans. Power Electron. 35, 7816–7825. https://doi.org/10.1109/TPEL.2020.2967053.

Stockwell, R., Mansinha, L., Lowe, R., 1996. Localization of the complex spectrum: the s transform. IEEE Trans. Signal Process. 44, 998–1001. https://doi.org/10.1109/78.492555.

Tschannen, V., Delescluse, M., Ettrich, N., Keuper, J., 2020. Extracting horizon surfaces from 3d seismic data using deep learning. Geophysics 85, N17–N26. https://doi.org/10.1190/GEO2019-0569.1.

Wang, D., Fu, J., Lei, Q., Luo, A., 2007. Exploration technology and prospect of low

permeability oil-gas field in ordos basin. Lithologic Reservoirs 19, 126–130.
Wang, Q., Gao, J., Liu, N., Jiang, X., 2018. High-resolution seismic time–frequency analysis using the synchrosqueezing generalized s-transform. Geosci. Rem. Sens. Lett. IEEE 15, 374–378. https://doi.org/10.1109/LGRS.2017.2789190.
Wang, Y., Peng, Z., He, Y., 2016. Time-frequency representation for seismic data using sparse s transform. In: 2016 2nd IEEE International Conference on Computer and Communications (ICCC), pp. 1923–1926. https://doi.org/10.1109/CompComm.2016.7925036.
Wigner, E.P., 1997. On the quantum correction for thermodynamic equilibrium: Part I: Physical Chemistry. Part II. Solid State Phys 110–120. https://doi.org/10.1103/PhysRev.40.749.
Wu, H., Li, Z., Liu, N., 2022. Variable seismic waveforms representation: weak-supervised learning based seismic horizon picking. J. Petrol. Sci. Eng. 214, 110412. https://doi.org/10.1016/j.petrol.2022.110412.
Wu, X., Liang, L., Shi, Y., Fomel, S., 2019. Faultseg3d: using synthetic data sets to train an end-to-end convolutional neural network for 3d seismic fault segmentation. Geophysics 84, IM35–IM45. https://doi.org/10.1190/GEO2018-0646.1.
Wu, X., Liu, T., 2009. Spectral decomposition of seismic data with reassigned smoothed pseudo wigner–ville distribution. J. Appl. Geophys. 68, 386–393. https://doi.org/10.1016/j.jappgeo.2009.03.004.
Xie, X., 2016. Provenance and sediment dispersal of the triassic yanchang formation, southwest ordos basin, China, and its implications. Sediment. Geol. 335, 1–16. https://doi.org/10.1016/j.sedgeo.2015.12.016.
Yang, Y., Gao, J., Wang, Z., Li, Z., 2021. Seismic absorption qualitative indicator via sparse group-lasso-based time–frequency representation. Geosci. Rem. Sens. Lett. IEEE 18, 1680–1684. https://doi.org/10.1109/LGRS.2020.3006340.
Yang, Y., Gao, J., Wang, Z., Liu, N., 2022. Data-driven time-frequency method and its application in detection of free gas beneath a gas hydrate deposit. IEEE Trans. Geosci. Rem. Sens. 60, 1–13. https://doi.org/10.1109/TGRS.2021.3138540.
Yao, J., Deng, X., Zhao, Y., Han, T., Chu, M., Pang, J., 2013. Characteristics of tight oil in triassic yanchang formation, ordos basin. Petrol. Explor. Dev. 40, 161–169. https://doi.org/10.1016/S1876-3804(13)60019-1.
Yim, J., Joo, D., Bae, J., Kim, J., 2017. A gift from knowledge distillation: fast optimization, network minimization and transfer learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4133–4141. https://doi.org/10.1109/CVPR.2017.754.
Yuan, S.-Y., Yang, S., Wang, T.-Y., Qi, J., Wang, S.-X., 2020. Inverse spectral decomposition using an lp-norm constraint for the detection of close geological anomalies. Petrol. Sci. 17, 1463–1477. https://doi.org/10.1007/s12182-020-00490-6.
Zhao, T., Song, W., 2012. An application of matching pursuit time-frequency decomposition method using multi-wavelet dictionaries. Petrol. Sci. 9, 310–316. https://doi.org/10.1007/s12182-012-0214-9.
Zhong, T., Cheng, M., Dong, X., Li, Y., Wu, N., 2022a. Seismic random noise suppression by using deep residual u-net. J. Petrol. Sci. Eng. 209, 109901. https://doi.org/10.1016/j.petrol.2021.109901.
Zhong, T., Wang, W., Lu, S., Dong, X., Yang, B., 2022b. Rmchn: a residual modular cascaded heterogeneous network for noise suppression in das-vsp records. Geosci. Rem. Sens. Lett. IEEE 20, 1–5. https://doi.org/10.1109/LGRS.2022.3229556.
Zhou, H., Tian, Y., Ye, Y., 2014. Dynamic deconvolution of seismic data based on generalized s-transform. J. Appl. Geophys. 108, 1–11. https://doi.org/10.1016/j.jappgeo.2014.06.004.