

D²UNet: Dual Decoder U-Net for Seismic Image Super-Resolution Reconstruction

Fan Min, *Member, IEEE*, Linrong Wang, Shulin Pan, Guojie Song

Abstract—Super-resolution reconstruction is an essential task of seismic inversion due to the low resolution and strong noise of field data. Popular deep networks derived from U-Net lack the ability to recover detailed edge features and weak signals. In this paper, we propose a dual decoder U-Net (D²UNet) to explore both detail and edge information of the data. The encoder inputs the low resolution image and the edge image obtained through the Canny algorithm. Edge image can provide rich shape and boundary information, which is helpful to generate more accurate and high-quality data. The dual decoder consists of a main decoder for high-resolution recovery and an edge decoder for edge contour detection. These two decoders interact with a texture warping module (TWM) with deformable convolution. TWM aims to distort realistic edge details to match the fidelity of low resolution inputs, especially the location of edges and weak signals. The loss function is a combination of L_1 loss and multi-scale structural similarity loss (MS-SSIM) to ensure perception quality. Results on synthetic and field seismic images show that D²UNet not only improves the resolution of noisy seismic images, but also maintains the image fidelity.

Index Terms—parallel decoder; seismic images; super-resolution.

I. INTRODUCTION

DU to the unavoidable influence of environmental conditions and instrument performance, field exploration data often have problems such as low resolution and noise interference. We can observe important information from high-quality seismic data [1], such as stratum, reservoir distribution, fault information. Therefore, the quality of seismic data plays a crucial role in the data interpretation [2] of oil and gas exploration and geological exploration. Only on the premise that the acquisition quality is guaranteed, can the processing technology (such as static correction and frequency extension) play its best role. Resolution improving and denoising are two popular technologies to address the above problems.

The resolution of seismic data is the precision description of spatial measurement of underground structures. It includes vertical and horizontal resolution. High resolution processing technology aims to broaden the frequency band and improve

Corresponding author: Fan Min. Tel.: +86 135 4068 5200.

Email: minfan@swpu.edu.cn.

Fan Min is with the School of Computer Science; Lab of Machine Learning; Institute for Artificial Intelligence, Southwest Petroleum University, Chengdu 610500, China (e-mail: minfan@swpu.edu.cn).

Linrong Wang is with the School of Computer Science, Southwest Petroleum University, Chengdu 610500, China (e-mail: 202121000476@swpu.edu.cn).

Shulin Pan is with the School of Earth Science and Technology, Southwest Petroleum University, Chengdu 610500, China (e-mail: shulinpan@swpu.edu.cn).

Guojie Song is with the School of Science, Southwest Petroleum University, Chengdu 610500, China (e-mail: songgj@swpu.edu.cn).

the main frequency. Its essence is to recover the true amplitude of weak signals. Common methods can be divided into three categories: deconvolution, absorption compensation and time-frequency spectrum based frequency extension. The *deconvolution* technology extracts or restores the original signal from the output signal, and expands the frequency spectrum by compressing the seismic wavelet to improve the resolution [3–6]. The *absorption compensation* technology is dominated by inverse Q filtering. Kjartansson attenuation model [7] performs amplitude compensation and phase correction for each frequency component. Inverse Q filtering of continuous wavelet transform [8] restores wider traces. The extended stability factor method of inverse Q filter [9] achieves the required high resolution by introducing parameters and gain restrictions. The *frequency extension* technology mainly uses time-frequency decomposition as the basic means, and processes the high and low frequency components in the time-frequency domain to achieve the effect of compressing wavelet and supporting broadband. Hilbert Huang method (HHT) [10] improves the resolution by calculating the compensation factor in time-frequency domain. Spectral fitting method [11] estimates the seismic wavelet spectrum on the generalized S-transform spectrum to eliminate the influence of the wavelet spectrum. Gabor deconvolution (GD) combined with wavelet scaling method [12] increases the dominant frequency to generate a higher resolution seismic trace.

With the exponential growth of seismic data scale, traditional methods have poor performance in terms of high-resolution recovery, denoising and efficiency. This makes it difficult for them to recover detailed geological information. The support of graphics processor (GPU) enhances the success of deep learning methods in high-resolution and noise reduction tasks [13–16]. In particular, the deep convolution neural network (CNN), as a derivative extension of many methods, can be easily used to accelerate training with GPU. U-Net [17] is a network of encoder-decoder structure, which is widely used in high-resolution [18–20] and denoising [19, 21, 22] tasks. However, most of the methods mentioned above are only for resolution or denoising tasks. SRDUNet [23] based on U-Net can handle both tasks simultaneously. It can significantly enhance the detailed structure and stratigraphic characteristics of seismic images. But the results have artifacts, too smooth edges, and poor recovery effect on weak signals. Multi-task learning (MTL) [24] has been successfully applied in the field of seismic exploration. M-RUSR [18] takes the edge image of the seismic velocity model as an auxiliary task of the high-resolution task of the velocity model. It shows that the edge image learning improves the high-resolution effect of the

velocity model and has better details.

In this paper, we propose a new approach with three aspects:

- Network: we propose a new multi-task network D²UNet with one encoder and two decoders. The encoder takes low resolution and edge image obtained by Canny algorithm as input. The two decoders are used to learn two different tasks in parallel. High resolution restoration of seismic data is the main task, while edge detection is the auxiliary task. The auxiliary task can provide rich shape and boundary information, which is helpful to generate more accurate and high-quality data. They interact through a texture warping module (TWM) [25] with deformable convolution. It can distort real edge details to match the fidelity of low resolution inputs, especially the location of edges and weak signals.

- Training data: in order to deal with the problem of insufficient training data, we refer to the method proposed by [26, 27] to generate clean high-frequency and noisy low-frequency 3D seismic volume pairs, respectively. Then, we use Canny algorithm to extract the corresponding edge images. Finally, noisy low resolution seismic images and edge images are used as input, and clean high resolution seismic images and edge images are used as ground truth.

- Loss function: we introduce the joint loss function. Each task includes L_1 loss and multi-scale structure similarity loss (MS-SSIM). They can improve the perceptual quality of images.

A large number of experimental results are undertaken on synthetic and field seismic data. Results show that the proposed D²UNet network can not only improve the resolution of noisy seismic images, but also maintain the image fidelity.

The rest of the paper is organized as follows. Section II describes the problem definition of super resolution recovery of seismic data. Section III describes the network structure and loss function of the proposed method in detail. Section IV describes the generation method of training data set and the training details of the network. Section V discusses the experimental results of our method on synthetic and field data. Finally, Section VI makes a conclusion.

II. PROBLEM DEFINITION

According to the traditional convolutional model, the seismic record can be modeled as

$$\hat{\mathbf{y}} = \mathbf{W}\hat{r} + \mathbf{n}, \quad (1)$$

where $\hat{\mathbf{y}}$ represents the observation data filtered by the wide-band wavelet $w_h(t)$, \mathbf{W} is the convolution matrix formed by the seismic wavelet $w_h(t)$, \hat{r} is the full-band reflection coefficient or the band-limited reflection coefficient filtered by $w_h(t)$, and \mathbf{n} denotes random noise.

The objective function established by super-resolution inversion method is expressed as

$$J_{\hat{r}} = \min_{\hat{r}} \frac{1}{2} \|\hat{\mathbf{y}} - \mathbf{W}\hat{r}\|_2^2 + \mu Q(\hat{r}), \quad (2)$$

where μ is a regularization coefficient and $Q(\cdot)$ is a constraint function.

In image processing, the super-resolution reconstruction of seismic images is regarded as a low-level visual task, which can be modeled as

$$\mathbf{S}_L = D(\mathbf{S}_H) + \mathbf{n}, \quad (3)$$

where \mathbf{S}_L and \mathbf{S}_H denote low-resolution and high-resolution seismic images, respectively. $D(\cdot)$ denotes a degradation mapping function.

In this work, noisy low-resolution and noise-free high-resolution image are used as input and ground truth, respectively. They are denoted by a pair $\mathbf{S}_{in} = (\mathbf{S}_L, \mathbf{S}_H)$. Our goal is to reconstruct a high-resolution image $\hat{\mathbf{S}}_H$ from a low-resolution image \mathbf{S}_L with noise removed, denoted by

$$\hat{\mathbf{S}}_H = \text{Net}(\mathbf{S}_L, \Theta), \quad (4)$$

where $\text{Net}(\cdot)$ denotes our model, and Θ are the parameters of the network.

III. OUR METHOD

This section presents our method.

A. Architecture

Fig. 1 illustrates the proposed D²UNet network architecture. The input \mathbf{x}^d include two parts, namely the original seismic image and the edge image. Examples are giving by Figs. 2(a) and 2(b). The corresponding high-resolution image and edge image are taken as ground truth. Examples are giving by Figs. 2(c) and 2(d). The Canny edge detection operator [28] is used to obtain edge images from original seismic images. The D²UNet network consists of one encoder and two parallel decoders. The encoder consists of four down-sampling blocks, each of which consists of “max pooling layer + (convolutional layer + batch normalization layer + rectified linear unit (ReLU)) \times 2”. The max pooling layer uses a 2×2 kernel with a step size of two. The convolution layer uses a 3×3 convolution kernel. With four down-sampling blocks, the feature channels are 64, 128, 256, 512 and 1024, respectively.

D²UNet contains two parallel decoders, named edge decoder and main decoder. They all contain four up-sampling blocks, which are the opposite of the down-sampling blocks. The difference is that the up-sampling blocks enlarge the size of the feature map by transposed convolution. Furthermore, the output of the transposed convolution is connected by skip connections with the feature maps from the down-sampling blocks at the same level. Although skip connections can incorporate more fidelity information, these intermediate features are largely affected by the input features, and some high-quality image details are missing. In order to overcome the problem of preserving real details and improving fidelity, the two decoders use a texture warping module with deformable convolution to interact.

Suppose a noisy low-resolution seismic image and edge image are $\mathbf{x}^d = (\mathbf{x}^s, \mathbf{x}^e) \in \mathbb{R}^{H \times W \times 2}$. The input \mathbf{x}^s is passed through the encoder to get the latent vector $F_{5,3}^E = E(\mathbf{x}^s)$. To keep its true image edges, we decode it back to the image space using the edge decoder D_{edge} , expressed as

$$F_4^e = D_{edge}(F_{5,3}^E), \quad (5)$$

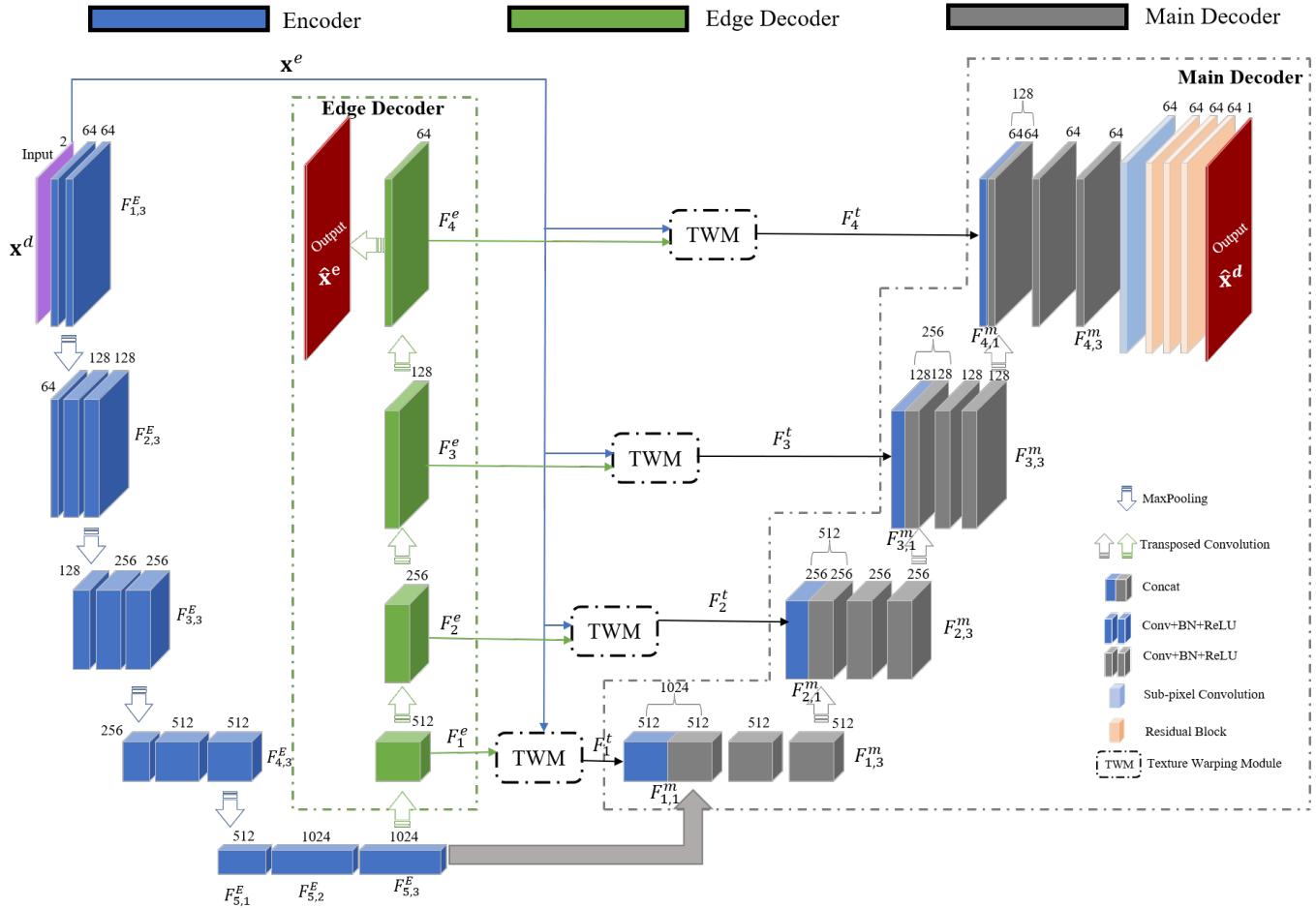


Fig. 1. Our proposed network structure.

where the multi-level features of the edge decoder are represented as F_i^e , $i \in \{1, 2, 3, 4\}$, respectively.

The main decoder D_{main} is designed to generate high-resolution seismic images \mathbf{x}^m , while the edge decoder D_{edge} extracts edge information and details. D_{main} distorts the edge features F_i^e by extracting input features from degenerate inputs at multiple spatial levels. In the texture warping module, we use the maximum spatial resolution input feature \mathbf{x}^e as input, which can maintain the richest fidelity of low resolution image information and edge information respectively.

For the i resolution level, F_i^e and high-quality features \mathbf{x}^e are used as inputs to the texture warping module, expressed as

$$F_i^t = \text{TWM}(F_i^e, \text{Conv}(\mathbf{x}^e)), \quad (6)$$

where TWM is the texture warping module. As shown in Fig. 3, we first concatenate the two input features to generate the offset. Then, the offset is applied to the deformable convolution to distort the edge features to match the fidelity of the input, denoted as

$$\begin{aligned} \text{offset} &= \text{Conv}(\text{Concat}(F_i^e, \text{Conv}(\mathbf{x}^e))), \\ F_i^t &= \text{DeformConv}(F_i^e, \text{offset}), \end{aligned} \quad (7)$$

where DeformConv denotes the deformable convolution.

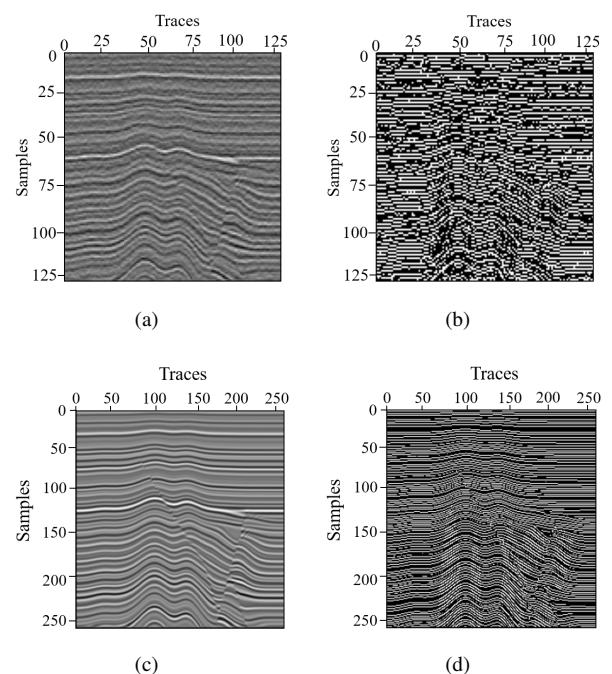


Fig. 2. Inputs: (a) noisy and low resolution seismic image, (b) edge image corresponding to (a), (c) clean and high resolution seismic image, and (d) edge image corresponding to (c).

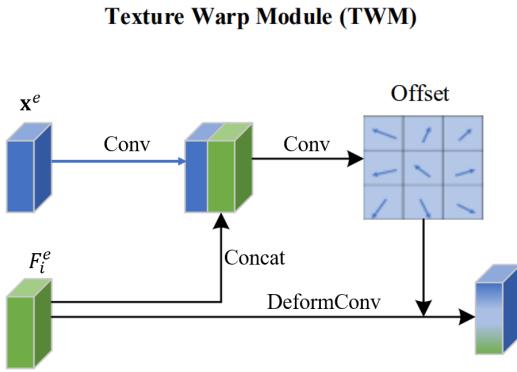


Fig. 3. Texture warping module (TWM).

Then, the features $F_{i,1}^m$ of the main decoder are concatenated by three features, including the distortion feature F_i^t , the feature of the down-sampling block in the encoder $F_{i,3}^E$ and the transposed convolution characteristics in the main decoder. The process can be formulated as

$$\begin{aligned} F_{i,1}^m &= \text{Concat}\left(F_1^t, F_{4,3}^E, \text{Deconv}(F_{5,3}^E)\right), \\ F_{i,1}^m &= \text{Concat}\left(F_i^t, F_{(5-i),3}^E, \text{Deconv}(F_{(i-1),3}^m)\right), \end{aligned} \quad (8)$$

where $i \geq 2$ and Deconv denotes transposed convolution. The features of the edge decoder are used as reference features containing rich edge information. The main decoder jointly learns the main features and “reference features” of seismic data.

In addition, we add a subpixel convolution layer to the four up-sampling blocks of the main decoder, after which connect the three residual blocks. Sub-pixel convolution has achieved important success in both image super-resolution reconstruction accuracy and computational performance [29]. We use a sub-pixel convolutional layer for up-sampling. It can not only reduce the training time and save GPU memory, but also provide more context information to further improve the image quality [30]. Three residual blocks are connected after the sub-pixel convolutional layer, which is able to learn more high-frequency information and details. Each residual block is composed of two “convolutional + batch normalization + ReLU” modules. The skip connection covers two convolutional layers. Finally, we use a 1×1 convolution layer to reduce the number of feature channels to match the ground truth.

B. Loss Function

The loss function of our method is defined as

$$L = \mathbb{W}(L_d, \sigma_d) + \mathbb{W}(L_e, \sigma_e), \quad (9)$$

where weighting operator \mathbb{W} and trainable parameter σ_i balance training between two tasks. L_d and L_e denote the loss of main decoder branch and edge decoder branch, respectively.

Following [31], we define the weight parameter \mathbb{W} as

$$\mathbb{W}(L_i, \sigma_i) = \frac{1}{2\sigma_i} L_i + \log \sigma_i, i \in \{d, e\} \quad (10)$$

where the value of σ_i quantifies the uncertainty associated with the prediction. The logarithmic term is used to prevent σ_i

has increased excessively. By dynamically adjusting the contributions of multiple loss items, the loss can be automatically balanced.

Loss items L_d and L_e include L_1 and MS-SSIM, expressed as

$$\begin{aligned} L_d &= \lambda L_1(\hat{\mathbf{x}}^d, \mathbf{x}^d) + (1 - \lambda) L_{\text{MS-SSIM}}(\hat{\mathbf{x}}^d, \mathbf{x}^d), \\ L_e &= \lambda L_1(\hat{\mathbf{x}}^e, \mathbf{x}^e) + (1 - \lambda) L_{\text{MS-SSIM}}(\hat{\mathbf{x}}^e, \mathbf{x}^e), \end{aligned} \quad (11)$$

where L_1 is widely used for image super resolution. MS-SSIM is a multi-scale image quality assessment method. $\hat{\mathbf{x}}^d$ and \mathbf{x}^d denote recovered high-resolution seismic images and ground truth, respectively. $\hat{\mathbf{x}}^e$ and \mathbf{x}^e denote recovered high-resolution edge images and ground truth, respectively. λ is the weight parameter of L_1 function. According to [23], λ is set to 0.4.

L_1 loss is defined as

$$L_1 = \frac{1}{N} \sum_{i,j} |\hat{\mathbf{x}}(i,j) - \mathbf{x}(i,j)|, \quad (12)$$

where N denotes the total number of pixels. MS-SSIM is a multi-scale image quality assessment method, which is an improved version of SSIM. SSIM is defined as

$$L_{\text{SSIM}}(\hat{\mathbf{x}}, \mathbf{x}) = [l(\hat{\mathbf{x}}, \mathbf{x})]^\alpha \cdot [c(\hat{\mathbf{x}}, \mathbf{x})]^\beta \cdot [s(\hat{\mathbf{x}}, \mathbf{x})]^\gamma \quad (13)$$

where

$$\begin{aligned} l(\hat{\mathbf{x}}, \mathbf{x}) &= \frac{2\mu_{\hat{\mathbf{x}}}\mu_{\mathbf{x}}+c_1}{\mu_{\hat{\mathbf{x}}}^2+\mu_{\mathbf{x}}^2+c_1}, \\ c(\hat{\mathbf{x}}, \mathbf{x}) &= \frac{2\sigma_{\hat{\mathbf{x}}\mathbf{x}}+c_2}{\sigma_{\hat{\mathbf{x}}}^2+\sigma_{\mathbf{x}}^2+c_2}, \\ s(\hat{\mathbf{x}}, \mathbf{x}) &= \frac{\sigma_{\hat{\mathbf{x}}\mathbf{x}}+c_3}{\sigma_{\hat{\mathbf{x}}}\sigma_{\mathbf{x}}+c_3}, \end{aligned} \quad (14)$$

where $\hat{\mathbf{x}}$ and \mathbf{x} denote recovered high resolution images and ground truth, respectively. μ and σ denote the mean and the standard deviation of image, respectively. For example, $\sigma_{\hat{\mathbf{x}}\mathbf{x}}$ denotes the covariance between images $\hat{\mathbf{x}}$ and \mathbf{x} . c_1 , c_2 and c_3 denote three constants to avoid the denominator being too small to reach the best value. $l(\cdot)$, $c(\cdot)$ and $s(\cdot)$ represent a function that calculates brightness or amplitude, contrast, and structure, respectively. α , β and γ represent the corresponding weights of the three functions.

MS-SSIM is defined as

$$L_{\text{MS-SSIM}}(\hat{\mathbf{x}}, \mathbf{x}) = [l_M(\hat{\mathbf{x}}, \mathbf{x})]^{\alpha_M} \cdot \prod_{j=1}^M [c_j(\hat{\mathbf{x}}, \mathbf{x})]^{\beta_j} [s_j(\hat{\mathbf{x}}, \mathbf{x})]^{\gamma_j}, \quad (15)$$

where $M = 5$ denotes that we measure SSIM from five scales. $\alpha = \beta = \gamma = [0.0448, 0.2856, 0.3001, 0.2363, 0.1333]$.

IV. TRAINING DATASETS

The proposed method is supervised and we need a large number of noise-free high-resolution images as ground truth. However, such field data sets are rare in practice. There have been many references [26, 27] confirming the effectiveness and accuracy of the network trained only on synthetic seismic data to deal with seismic-related tasks. Therefore, we extract a large number of high-resolution images by constructing real complex structural models.

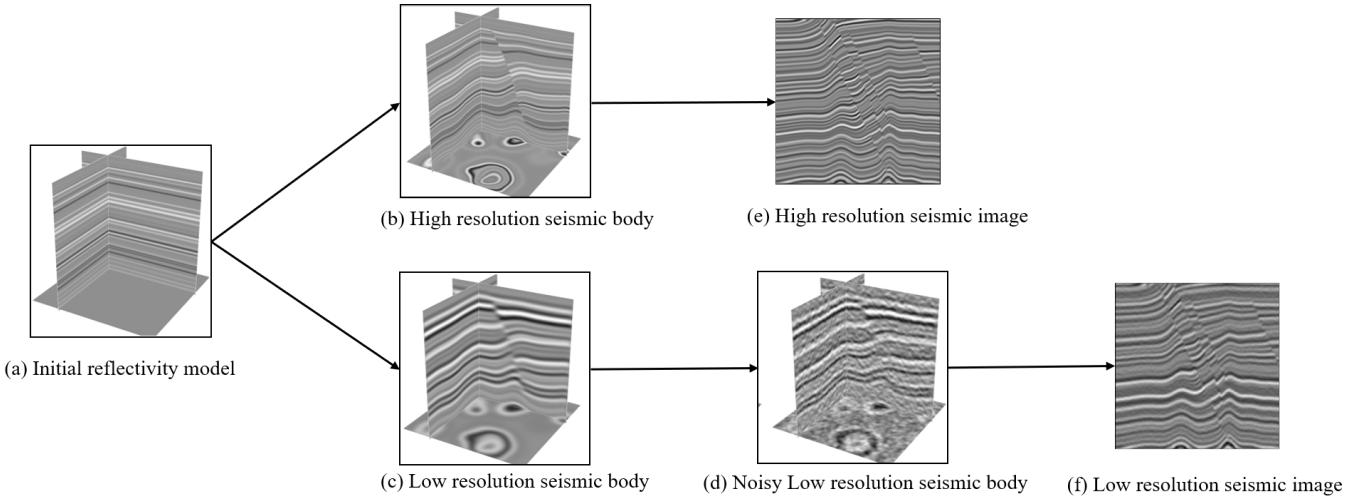


Fig. 4. The workflow of creating synthetic training data sets.

A. Generate training data

Fig. 4 shows our workflow for creating synthetic datasets is similar to [26, 27] (the detailed creation process of 3D data volume is shown in Appendix A), as follows

Step 1: Generate a 1D horizontal reflectance model with random sequences between [-1,1];

Step 2: Using vertical shear model to create folded structure;

Step 3: Fold and fault reflectance models are obtained by adding plane shear;

Step 4: The reflectance model is convolved with high frequency Ricker wavelet to obtain high frequency seismic bodies, and then multiple two-dimensional high frequency seismic images are extracted as ground truth;

Step 5: The reflectance model is convolved with low-frequency Ricker wavelet to obtain low-frequency seismic bodies, and random coloured noise is further added to obtain more realistic seismic bodies, and then multiple two-dimensional noisy low-frequency seismic images are extracted as input.

We follow these steps to create 1600 seismic bodies, including 800 high frequency and their low frequency counterparts. These seismic bodies are randomly divided into three groups, 80% as the training set, 10% as the validation set, and 10% as the test set. From each high-frequency and its corresponding low-frequency seismic body, we have extracted multiple 256×256 and 128×128 2D seismic image pairs. It is worth noting that the random colored noise added to the low-frequency image replaces the common white Gaussian noise. As shown in Fig. 5, the random colored noise is closer to reality than simple white Gaussian noise. In addition, the SNR of the training samples with added noise is randomly defined within the range of [4, 14].

To improve the model performance, we crop the 256×256 2D seismic image into small patches. The size of the patch affects the performance of the model. Too large leads to a large memory footprint, and too small leads to too little information

to recover details. To balance the above issues, we test the model performance with two different patch sizes, including 48×48 and 96×96 . Figs. 6(a) and 6(b) show the PSNR of the validation data set when the patch size is 48 and 96, respectively. We find that the network performance of patch 96 is better than that of patch 48. Figs. 6(c) and 6(d) show the data results of patch 96 and patch 48 respectively. Result shows that the larger patch training network can better recover weak signals (red arrow). In addition, we also perform random horizontal flipping for each sample to increase the diversity of training samples.

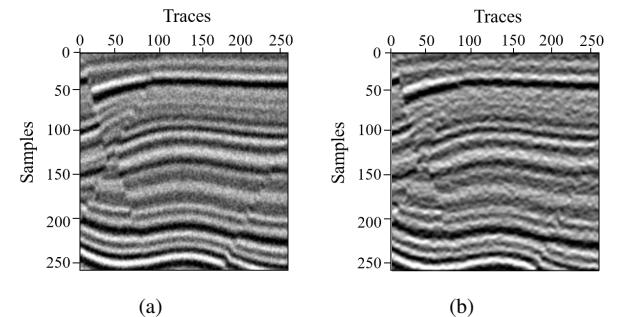


Fig. 5. Comparison of synthetic seismic images with different noise: (a) white gaussian noise, (b) random coloured noise.

B. Training details

After data augmentation, we generate 3,600 low-high resolution image pairs. Each sample is normalized to [0,1].

Adam optimizer [32] is used for training, and set $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$, momentum = 0.9. The initial learning rate is initialized to $1e-4$, and the number of training iterations is set to 150. The experiments are carried out in Pytorch GPU framework, and D²UNet is trained and tested on NVIDIA GeForce GTX 1650 GPU. The training process takes about 12 h. Fig. 7(a) illustrates the loss curves for training and validation, and Fig. 7(b) illustrates the weights of respective

losses. A larger weight for a certain loss term indicates higher confidence associated with the task (lower uncertainty).

Fig. 8 shows the TWM training process (only the first 12 channels are shown), which corresponds to the Edge Decoder process. The first column represents the result F_i^e of Edge Decoder after each transposed convolution. We find that the feature map has the characteristics of seismic waveform until $i = 3$. The second column represents the offset calculated in TWM. We find that when $i = 2$, there is an obvious waveform shape in the offset. The third column represents the feature map after the offset is applied. The seismic waveform of the feature map added with offsets is clearer and more accurate. This example is enough to show that TWM has the ability to extract important features faster and more accurately.

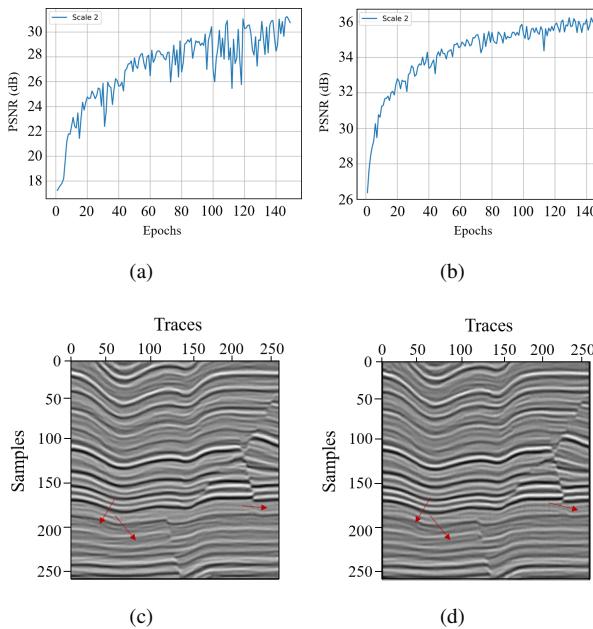


Fig. 6. Patch size: 48×48 , (a) performance curves, (c) validation set experimental result. Patch size: 96×96 , (b) performance curves, (d) validation set experimental result.

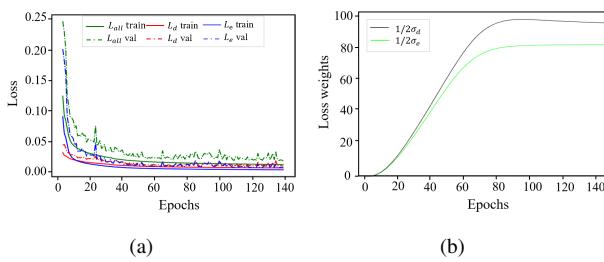


Fig. 7. Training record: (a) training and validation loss curves; (b) weights of respective losses. A larger weight for a certain loss term indicates higher confidence associated with the task (lower uncertainty).

V. EXPERIMENTS

In this section, we demonstrate the performance of the proposed method on synthetic and field data, and compare them with SRDUNet [23].

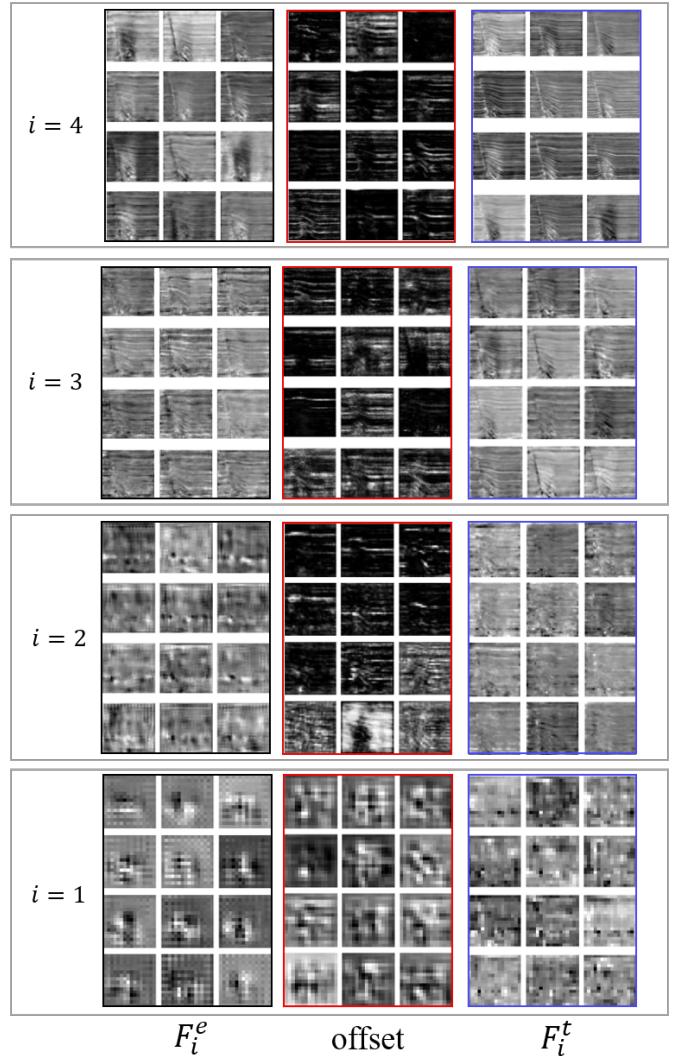


Fig. 8. Example of TWM training process. It corresponds to the Edge Decoder process in Fig. 1.

A. Evaluation on synthetic data

We first evaluate the performance of the D²UNet model on synthetic data. The data does not involve training and validation datasets. Fig. 9 shows the test results of synthetic data, where Figs. 9(a) and 9(b) are original noisy low resolution seismic image and ground truth, respectively. Figs. 9(c) and 9(d) are high-resolution seismic images recovered from SRDUNet and D²UNet respectively. Compared with noisy low resolution seismic images, the high resolution seismic images restored by SRDUNet and our method have enhanced structural features and clearer edge features, such as faults. In particular, blurry traces in low resolution images can also be recovered well (shown by the yellow arrow). In addition, two methods can achieve resolution improving and denoising simultaneously. Compared with the SRDUNet method, our method is more robust to weak signals, as shown by the red arrow in the figure.

In order to compare the recovery effects more clearly, we compare the amplitude characteristics of a randomly selected seismic trace in the recovery results and report the results in

Fig. 10. These three records are from ground truth (blue), SRDUNet restored seismic profiles (green) and D²UNet restored seismic profiles (orange), which correspond to the colored vertical lines in Fig. 10 respectively. The waveform shapes of the three curves are generally consistent, but the trace amplitude recovered by our method is closer to the ground truth. From the test results of synthetic data, our method can transform noisy low resolution images into clean high-resolution seismic images with enhanced fault, thin layer and clear edge features, which is conducive to subsequent seismic interpretation. In addition, our method is effective, that is, it takes 25.85 seconds to process 150 128×128 seismic images.

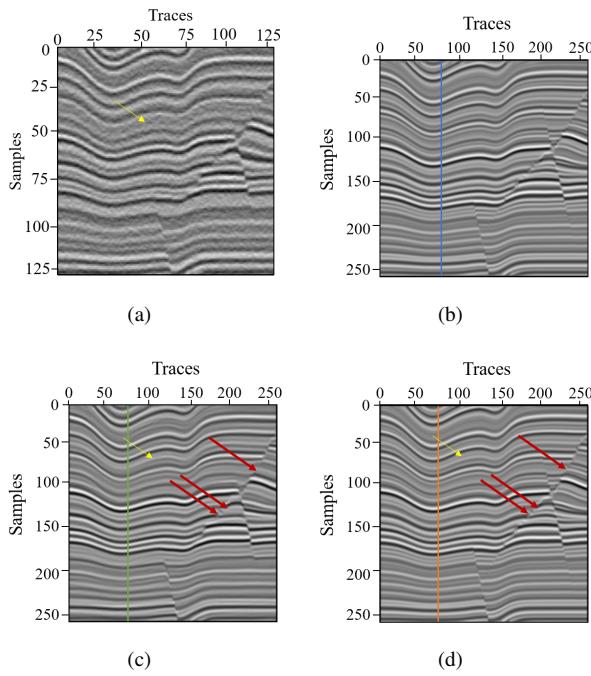


Fig. 9. Synthetic data test results: (a) noisy and low resolution seismic image, (b) clean and high resolution seismic image, (c) recovered seismic image using SRDUNet and (d) recovered seismic image using D²UNet.

B. Evaluation on field data

In the previous section, we explain the effectiveness of the D²UNet method on synthetic data. In this section, we test the trained D²UNet with multiple field seismic data.

Fig. 11 shows three field examples captured directly from the paper [23]. Figs. 11(a), 11(d) and 11(g) are three field seismic data acquired in different exploration. The second and

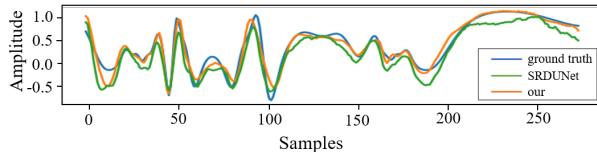


Fig. 10. Three traces extracted from the same position where the vertical lines shown in Fig. 9. The blue, green, and orange curves, respectively, represent the traces extracted from the ground truth (Fig. 9(b)), SRDUNet output (Fig. 9(c)), and D²UNet output (Fig. 9(d)).

third columns of Fig. 11 are the corresponding results obtained by simultaneous high-resolution and denoising methods based on SRDUNet and D²UNet, respectively. The bottom of each big figure is the local enlarged view corresponding to the yellow box. Compared with the original data, the recovery results of SRDUNet and D²UNet show more clear and detailed geological structure and stratigraphic characteristics (such as small-scale faults and thin beds), and remove noise. But from the details, our method is better than SRDUNet. According to the local enlarged view, the D²UNet method has a stronger recovery ability for weak signals and a higher resolution for thin layers (yellow arrow). From the local enlarged view in the middle of SRDUNet, we find that the restoration effect of SRDUNet for fuzzy geological structures and features is poor, and there is too smooth phenomenon. In contrast, the stratum recovered by D²UNet is more real and rich (yellow arrow). It can be seen from the recovery results of three field examples that, although there are significant differences between training data and field data, well trained models work well in synthetic data and field data testing. This further shows that the D²UNet method trained only with synthetic data has certain generalization and robustness.

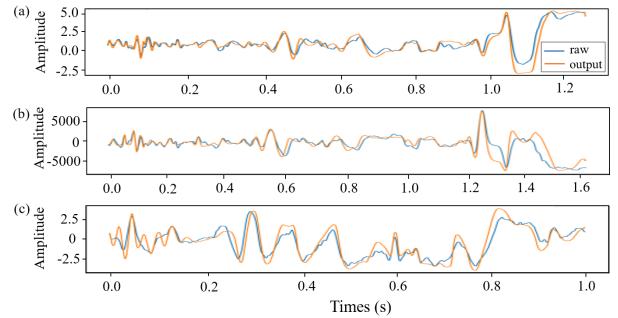


Fig. 12. Traces analysis of the field seismic images (blue) and the results (orange) of our D²UNet method shown in Fig. 11.

In addition, we also compared our high-resolution results with the original traces. Fig. 12 shows some seismic traces that are extracted from the original images and our high-resolution images in Fig. 11. Fig. 12(a) shows the randomly extracted seismic trace from Figs. 11(a) and 11(c), Fig. 12(b) is one trace extracted from the second data (Figs. 11(d) and 11(f)), and Fig. 12(c) is one trace extracted from the third data (Figs. 11(g) and 11(i)). Among them, blue represents the original trace, and orange represents the trace of the output image. Compared with the original trace, our results have similar characteristics and waveform trends, but have more details.

In addition, we also conduct spectral analysis on three seismic data, as shown in Fig. 13. The amplitude of each frequency in the spectrum is obtained by averaging all traces in the seismic profile. The blue and orange curves represent the amplitude spectra of the original data and the output data, respectively. Results show that the frequency band of the seismic profile obtained by D²UNet is wider than that of the original data, especially in the high-frequency part.

The second field example is selected from the logging data of Luzhou 203. The test data includes 879 traces and

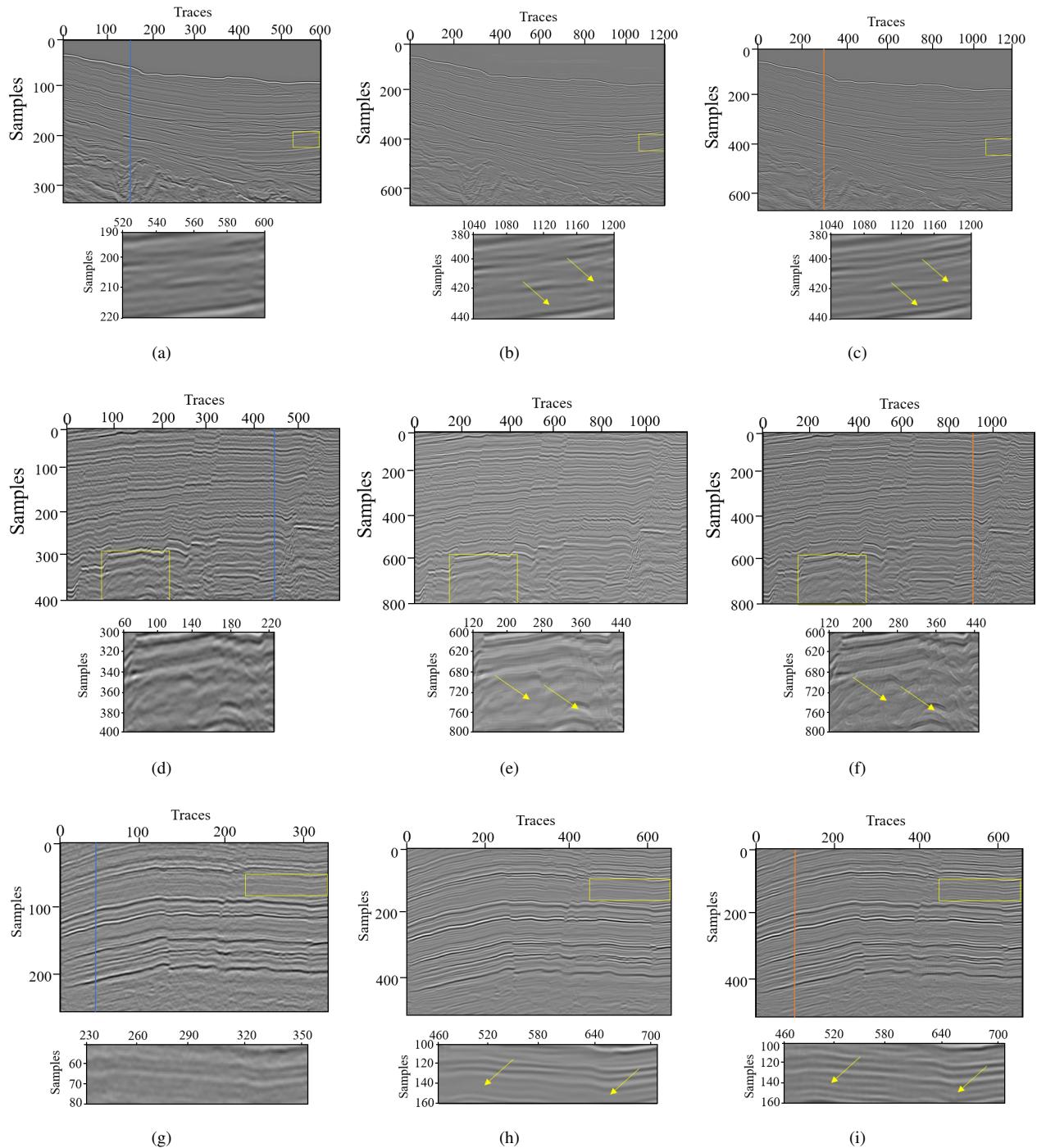


Fig. 11. Experimental results of proposed method on three field seismic data. First field data: (a) low resolution seismic image, recovery results of using (b) SRDUNet and (c) D²UNet methods; Second field data: (d) low resolution seismic image, recovery results of using (e) SRDUNet and (f) D²UNet methods; Third field data: (g) low resolution seismic image, recovery results of using (h) SRDUNet and (i) D²UNet methods. The bottom of each subfigure shows a partial enlargement of the yellow box.

	Synthetic data			Field data				Logging data			
	Avg PSNR	Avg SSIM	All Time	Avg SSIM	Avg DS	Avg PSNR	All Time	SSIM	DS	PSNR	Time
SRDUNet	30.256	0.976	14.126	0.917	27	9.405	3.719	0.524	8	3.629	0.567
D ² UNet	36.024↑	0.989↑	21.812	0.956↑	34↑	12.732↑	4.361	0.826↑	18↑	7.117↑	0.765

TABLE I

COMPARISON OF EXPERIMENTAL RESULTS BETWEEN D²UNET AND SRDUNET. AVG PSNR: AVERAGE PSNR (DB); AVG SSIM: AVERAGE PSNR SSIM; AVG DS: AVERAGE DOMINANT SPECTRUM (HZ). THE SYNTHETIC DATA INCLUDES 150 SEISMIC IMAGES. THE FIELD DATA CORRESPONDS TO THE THREE DATA IN FIG. 11. LOGGING DATA CORRESPONDS TO FIG. 14(A).

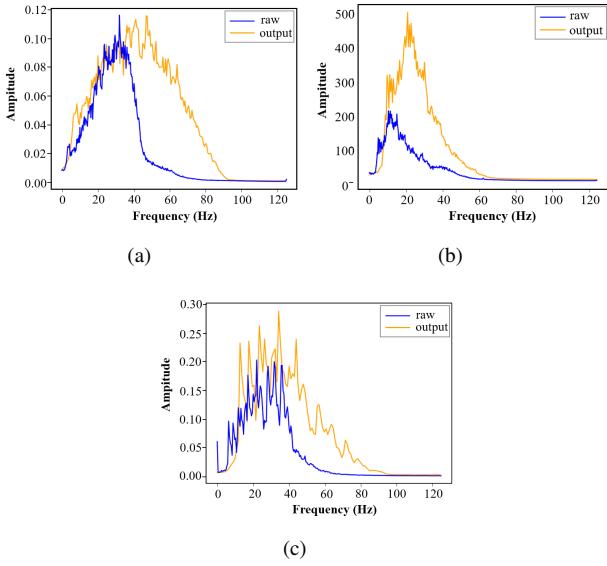


Fig. 13. Spectrum analysis of three field seismic images: (a) the first field data, (b) the second field data, and (c) the third field data.

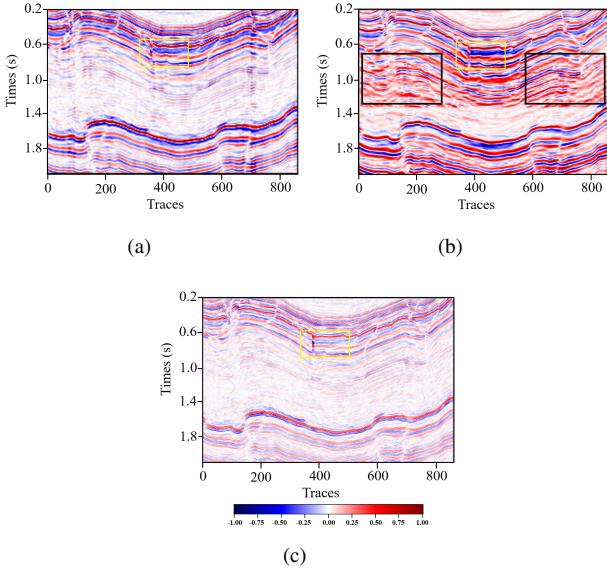


Fig. 14. Comparison of logging data effects: (a) original seismic profile, (c) seismic profile restored by SRDUNet method, (e) seismic profile restored by D²UNet method; (b), (d) and (f) are the enlarged versions of the yellow boxes in the corresponding images.

the sampling interval is 0.02 s. Fig. 14 shows the recovery effects of the two methods on the original logging data. From the overall effect, the recovery result of SRDUNet method is not completely correct, as shown in the black box in Fig. 14(b). In contrast, our method is more accurate in terms of visual quality. Fig. 15 shows the local enlarged view corresponding to the yellow box in Fig. 14. One data covered in the image is logging data, which contains 0.3 s. We find that logging data in SRDUNet and D²UNet are in good agreement. In particular, the result of D²UNet restoration has more detailed stratigraphic structure and edge characteristics, as shown by the black arrow in the figure. In addition, we also compared various Fourier frequency spectrums to see

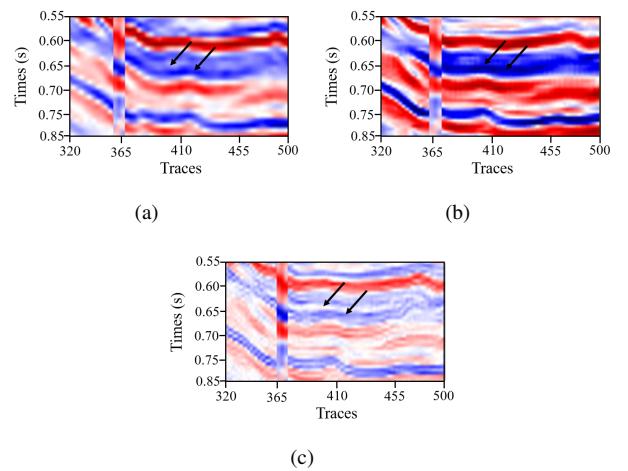


Fig. 15. Comparison of logging data effects: (a) original seismic profile, (c) seismic profile restored by SRDUNet method, (e) seismic profile restored by D²UNet method; (b), (d) and (f) are the enlarged versions of the yellow boxes in the corresponding images.

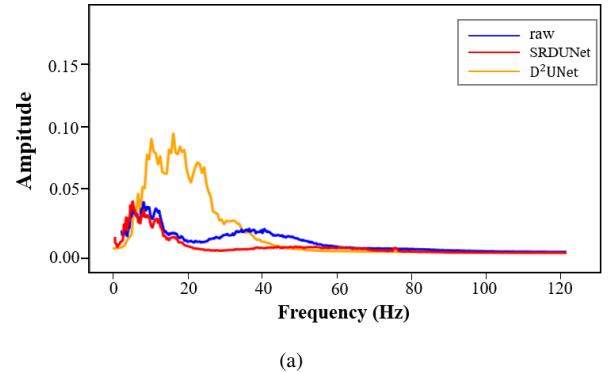


Fig. 16. Spectrum analysis of the logging data.

how the dominant frequency is highlighted in the characteristic frequency band (5-40 Hz) of seismic records. The results show that the frequency band obtained by our method is wider and the information gain provided is clearer (Fig. 16), while the frequency band obtained by SRDUNet method is not much different from the original data.

Finally, we summarize the experimental results of the two methods on all data in Table I. In the table, except for time, they all represent average indicators. For synthetic data, we test 150 128×128 seismic images. The proposed method is improved in both PSNR and SSIM, and the test time is only increased by a few seconds. For field data and logging data, SSIM, dominant spectrum and PSNR constitute a joint indicator. The indices of the proposed method are better than SRDUNet, especially the logging data.

VI. CONCLUSION

In this paper, we designed a network called D²UNet, which had an encoder and two parallel decoders. By comparing the synthetic data with the field data and the high-resolution results of the test algorithm, our results could more clearly reflect the detailed structure and stratigraphic characteristics, especially the edge and weak signals. Although the network only used

synthetic data for training, it still had good generalization ability. If the training data set could be enriched, it was expected that the performance of field data would be further improved.

VII. ACKNOWLEDGEMENTS

This work is supported by Central Government Funds of Guiding Local Scientific and Technological Development (No. 2021ZYD0003) and the National Natural Science Foundation of China (No. 41674141), and Nanchong Municipal Government-Universities Scientific Cooperation Project (Nos. SXHZ045, 19XHZ0030).

APPENDIX A 3D SYNTHETIC SEISMIC CONSTRUCTION

The workflow of creating synthetic data is first proposed by [33].

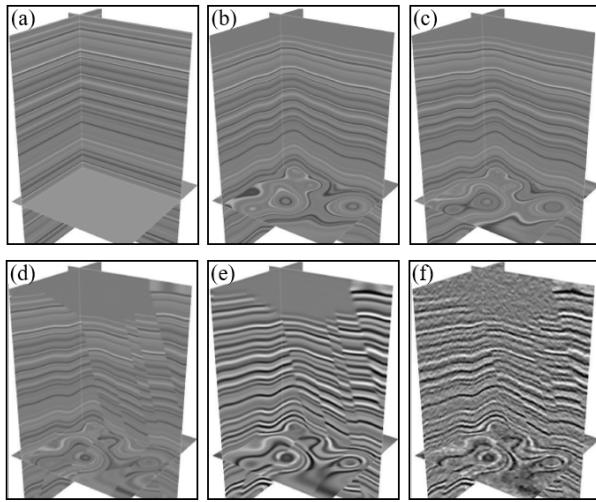


Fig. 17. Workflow for creating a 3D synthetic training dataset: (a) horizontal reflectivity model, (b) adding folding structure, (c) adding plane shear, (d) folding fault reflectivity model, (e) synthetic seismic image obtained by convolution of reflectivity model and Ricker wavelet, and (f) adding noise.

First, we generate a 1D horizontal reflectivity model $r(x, y, z)$ with a random sequence between [-1,1] (Fig. 17(a)). Second, we obtain fold and fault reflection models through vertical shear. The folding structure is defined as

$$s_1(x, y, z) = a_0 + \frac{1.5z}{z_{\max}} \sum_{k=1}^{k=N} b_k e^{\frac{(x-c_k)^2+(y-d_k)^2}{2\sigma_k^2}}, \quad (16)$$

where $\frac{1.5z}{z_{\max}}$ represents a linear scaling function designed to attenuate vertical folding from bottom to top. $e^{\frac{(x-c_k)^2+(y-d_k)^2}{2\sigma_k^2}}$ represents a 2D Gaussian function, which produces a folding structure that changes laterally. Different combinations of a_0, b_k, c_k, d_k and σ_k produce different folding structures with different spatial changes. We obtain the folded model $r(x, y, z + s_1(x, y, z))$ (Fig. 17(b)), which is obtained by sinc interpolation of the original reflectivity model $r(x, y, z)$. Third, we also add some planar shearing defined as follows

$$s_2(x, y, z) = e_0 + fx + gy, \quad (17)$$

where e_0, f and g are randomly selected from a predefined range. Plane cutting only controls the horizontal direction, and the vertical direction remains unchanged. We obtain a new reflectivity model $r(x, y, z + s_1 + s_2)$ (Fig. 17(c)), which is obtained by applying the plane displacement $s_2(x, y, z)$ to $r(x, y, z + s_1)$. Fourth, we add plane faults with different directions (dip and strike) and displacements. As shown in Fig. 17(d), we add six fault planes that are not too close apart. There are two definitions of fault displacement distribution: linear function and Gaussian function. Under the linear function, the fault displacement increases or decreases linearly along the dip direction. Under the Gaussian function, the fault displacement decreases in all directions along the fault plane from the fault center. We set the maximum fault position of each fault between 0 and 40 samples. Fifthly, we convolve the model obtained in the previous step with Ricker wavelet with different peak frequencies to obtain 3D seismic images (Fig. 17(e)). Finally, we add colored noise to it to further improve the authenticity of the image (Fig. 17(f)).

APPENDIX B ABLATION STUDY

D²UNet takes seismic data and edge images as input. It has a dual decoder structure and aims to gradually integrate seismic data features and edge features. The two decoders are connected via TWM, which is designed to distort real stratigraphic details to match the fidelity of low resolution inputs, particularly fault. In this section, we use Fig. 19(a) as the test data to conduct qualitative ablation studies of these key designs.

We have set three variants of the network, as shown in Table II:

- (1) Variant 1 (Fig. 18(a)): input as seismic image; single decoder without TWM;
- (2) Variant 2 (Fig. 18(b)): input as seismic image and edge image; single decoder without TWM;
- (3) Variant 3 (Fig. 18(c)): input as seismic image and edge image; dual decoder without TWM;
- (4) D²UNet (Fig. 18(d)): input as seismic image and edge image; dual decoder with TWM.

Variant 1 directly uses the coding features of seismic data, and does not use edge features. As shown in Fig. 19(b), it is difficult to reconstruct high-resolution seismic images using only seismic data, resulting in a large number of artifacts, and the stratum is very fuzzy. Variant 2 uses both seismic and edge images as input. From the visualization of Fig. 19(c), we can observe that the stratum and boundary are gradually clear. These two results show that edge image loss can provide rich shape and boundary information, which is helpful to generate higher quality data. However, there are still a lot of artifacts, and the stratum definition is not enough. We speculate that the network may not fully extract the features of edge images.

Variant 3 uses a parallel decoder structure, which decouples high-resolution seismic reconstruction into two sub tasks, namely, generating high-resolution seismic images and edge images simultaneously. During decoding, the features of the

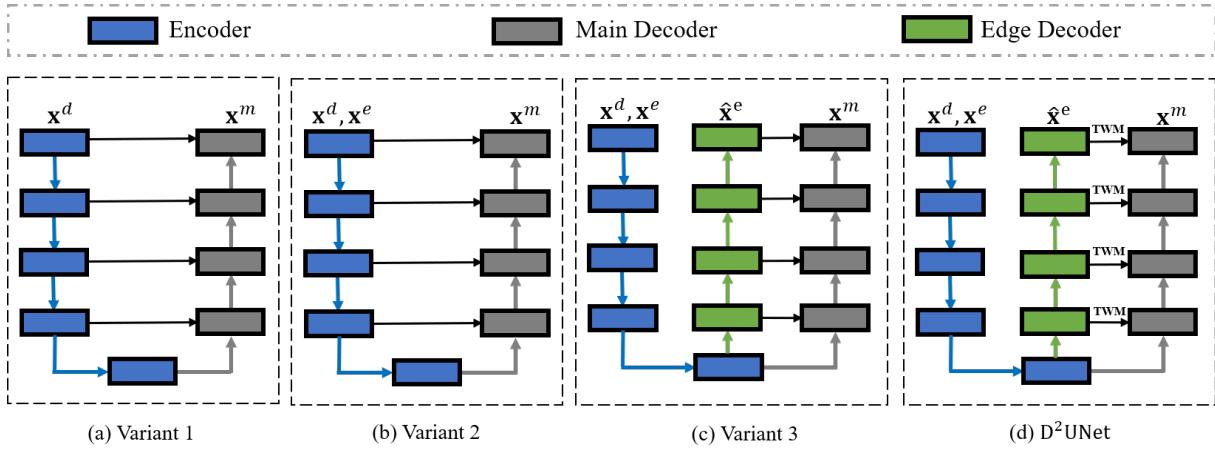


Fig. 18. Illustration of the architecture variants. (a) Variant 1, (b) Variant 2, (c) Variant 3, and (d) D²UNet.

edge decoder are directly spliced with the features of the same level of the main decoder. As shown in the Fig. 19(d), there is almost no artifact, and the resolution gradually increases. However, as shown in the yellow box in the figure, the details of recovery are not real enough. It is speculated that the network recover the noise in the original data as part of the stratum.

D²UNet adds TWM to connect two parallel decoders, which aims to distort the real stratum details to match the fidelity of low resolution input, especially at locations with strong noise and low resolution. Fig. 19(e) shows that the seismic image restored after adding TWM has the highest resolution and more realistic details. From the PSNR recorded in Table II, we can also see that the reconstruction result of D²UNet is optimal.

TABLE II
ABLATION RESULTS. *Inp Feat.*: EDGE IMAGE; *Dual Dec.*: DUAL DECODER;
TWM: TEXTURE WARPING MODULE

Models	Configurations			Test
	Inp Feat.	Dual Dec.	TWM	
Variant 1				24.885
Variant 2	✓			24.910
Variant 3	✓	✓		25.056
D ² UNet	✓	✓	✓	26.564

APPENDIX C NOISE ROBUSTNESS ASSESSMENT

The possible noise of seismic data should be considered when designing a new resolution enhancement method. Therefore, more experiments are done to evaluate the performance of the proposed method when using seismic data with different intensity noise. Specifically, we evaluate the denoising performance of D²UNet trained with colored noise synthetic data on Gaussian noise.

Synthetic data are still generated using the method in Appendix A. The noisy data is generated by adding the Gaussian

noise of $\sigma = 0.2, 0.4, 0.6$ and 0.8 to the clean synthetic data. The SNR of the noisy data is 13.934, 7.943, 4.454 and 0.875 respectively. Table III shows the quantitative comparison of D²UNet's performance on noisy data. We can see that with the increase of σ , the seismic quality decrease significantly (the SNR value is very low). However, the reconstruction results of D²UNet can still guarantee a high SNR and PSNR, with MSE controlled below 0.1 and SSIM above 0.6.

TABLE III
QUANTITATIVE COMPARISON OF D²UNET RESTORATION RESULTS UNDER DIFFERENT LEVELS OF GAUSSIAN NOISY DATA.

Gaussian noise	σ	0.2	0.4	0.6	0.8
Nosiy seismic data	SNR	13.934	7.943	4.454	0.875
Result	SNR	16.924	13.773	11.875	9.405
	PSNR	17.938	14.889	13.092	10.817
	MSE	0.016	0.033	0.049	0.084
	SSIM	0.843	0.764	0.705	0.619

The top row of Fig. 20 corresponds to seismic data with different noise intensity. When $\sigma = 0.8$, noise almost covers up some details of stratum and fault. It is worth noting that the D²UNet network uses colored noise data for training. As shown in the bottom line of the Fig. 20, D²UNet can still produce a pleasant visual effect for suppressing Gaussian noise. Although the evaluation index of the results decreases slightly with the increase of noise level, it is acceptable. Therefore, D²UNet shows the ability to deal with different kinds of noise (Gaussian noise and colored noise) and different intensities of noise.

ACKNOWLEDGEMENTS

This work is supported by Central Government Funds of Guiding Local Scientific and Technological Development (No. 2021ZYD0003), the National Natural Science Foundation of China (No. 41674141), and Nanchong Municipal Government-Universities Scientific Cooperation Project (Nos. SXHZ045, 19XHZ0030).

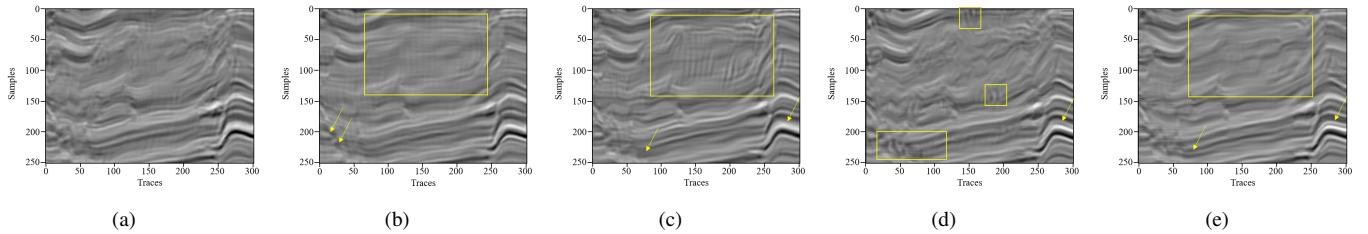


Fig. 19. (a) Origin data, (b) Variant 1, (c) Variant 2, (d) Variant 3, and (e) D²UNet.

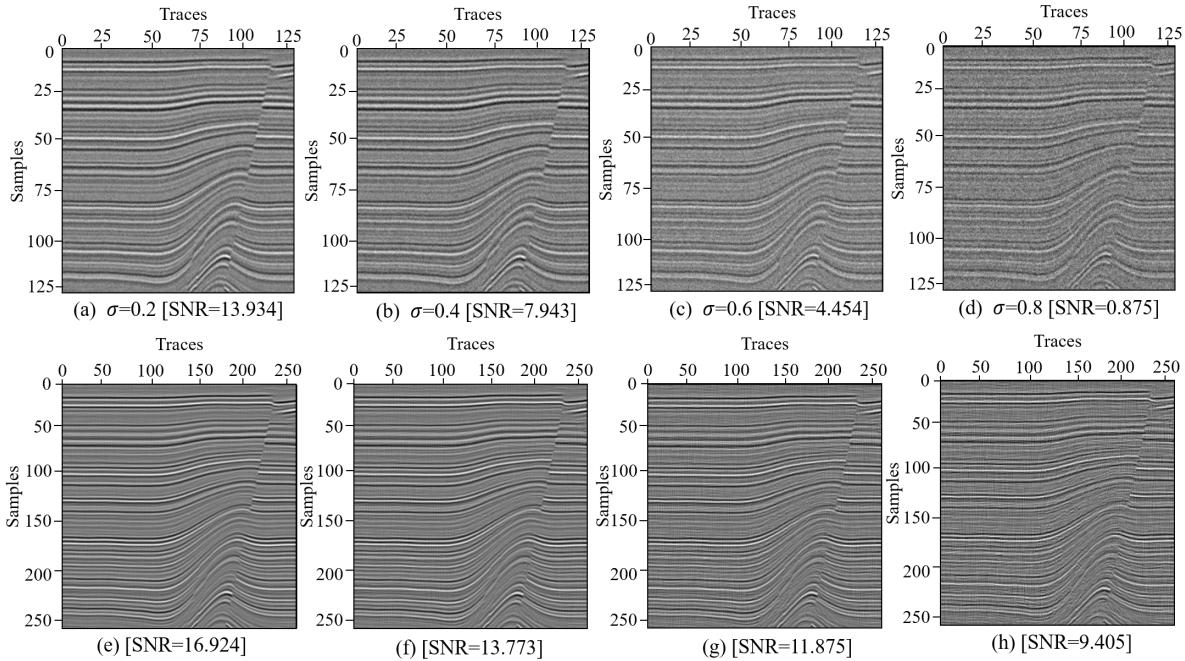


Fig. 20. Evaluation of D²UNet on noisy seismic data. The top line represents noisy data. The bottom line represents the high-resolution results obtained by D²Unet.

REFERENCES

- [1] I. Bruner and E. Landa, "Fault interpretation from high-resolution seismic data in the northern Negev, Israel," *Geophysics*, vol. 56, no. 7, pp. 1064–1070, 07 1991.
- [2] K. C. Miller, S. H. Harder, D. C. Adams, and T. O'Donnell, "Integrating high-resolution refraction data into near-surface seismic reflection data processing and interpretation," *Geophysics*, vol. 63, no. 4, pp. 1339–1347, 08 1998.
- [3] J. A. Cadzow and X. K. Li, "Blind deconvolution," *Digital Signal Processing*, vol. 5, no. 1, pp. 3–20, 01 1995.
- [4] K. Zhang, Y.-D. Li, G.-Z. Liu, and Y.-L. Jia, "Multi-resolution seismic signal deconvolution," *Chinese Journal of Geophysics*, vol. 42, no. 4, pp. 529–535, 04 1999.
- [5] S. D. Billings, B. R. Minty, and G. N. Newsam, "Deconvolution and spatial resolution of airborne gamma-ray surveys," *Geophysics*, vol. 68, no. 4, pp. 1257–1266, 01 2003.
- [6] C.-W. Liu, M.-Y. Sun, N.-X. Dai, W. Wu, Y.-W. Wei, M.-J. Guo, and H.-H. Fu, "Deep learning-based point-spread function deconvolution for migration image deblurring," *Geophysics*, vol. 87, no. 4, pp. S249–S265, 06 2022.
- [7] K. Einar, "Constant Q-wave propagation and attenuation," *Journal of Geophysical Research*, vol. 84, no. B9, p. 4737, 08 1979.
- [8] I. L. S. Braga and F. S. Moraes, "High-resolution gathers by inverse Q filtering in the wavelet domain," *Geophysics*, vol. 78, no. 2, pp. V53–V61, 02 2013.
- [9] Y. Li, G.-L. Zhang, J. Duan, C.-J. He, H. Du, F. Luo, Y.-Z. Zhan, and J. Wang, "Extended stable factor method for the inverse Q-filter," *Geophysics*, vol. 85, no. 3, pp. T155–T163, 04 2020.
- [10] X.-J. Wang, B.-S. Chen, and S.-Y. Cao, "Research and application of HHT amplitude frequency recovery processing technology," *China Offshore Oil and Gas*, vol. 1, no. 1, pp. 19–22, 1 2009.
- [11] X.-M. Shang, R. Diao, Y.-P. Feng, and C.-X. Zhao, "The application of spectral modeling method to high resolution processing of seismic data," *Geophysical and Geochemical Exploration*, vol. 63, no. 4, pp. 75–80, 1 2014.
- [12] A. Niloofer, R. K. Amin, K. R. Abolghasem, and S. Mehrdad, "Seismic resolution enhancement using

- scale transform in the time-frequency domain,” *Geophysics*, vol. 83, no. 6, pp. V305–V314, 10 2018.
- [13] K. Zhang, W.-M. Zuo, Y.-J. Chen, D.-Y. Meng, and L. Zhang, “Beyond a Gaussian denoiser: Residual learning of deep cnn for image denoising,” *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 02 2017.
- [14] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, “Enhanced deep residual networks for single image super-resolution,” 08 2017, pp. 1132–1140.
- [15] V. Lempitsky, A. Vedaldi, and D. Ulyanov, “Deep image prior,” 12 2018, pp. 9446–9454.
- [16] T. Dai, J.-R. Cai, Y.-B. Zhang, S.-T. Xia, and L. Zhang, “Second-order attention network for single image super-resolution,” 08 2019, pp. 11 057–11 066.
- [17] R. Olaf, F. Philipp, and B. Thomas, “U-Net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention*, 11 2015, pp. 234–241.
- [18] Y.-S. Li, J.-Y. Song, W.-K. Lu, P. Monkam, and Y. Ao, “Multitask learning for super-resolution of seismic velocity model,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 9, pp. 8022–8033, 11 2021.
- [19] H.-L. Chen, J.-H. Gao, Z.-Q. Gao, D.-Y. Chen, and T. Yang, “A sequential iterative deep learning seismic blind high-resolution inversion,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, no. 1, pp. 7817–7829, 08 2021.
- [20] Y.-S. Li, J.-Y. Song, W.-K. Lu, P. Monkam, and Y. Ao, “Super-resolution of seismic velocity model guided by seismic data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, no. 1, pp. 1–12, 08 2022.
- [21] T.-T. Bai, H.-X. Zhao, and Z.-Q. Wang, “A u-net based deep learning approach for seismic random noise suppression,” vol. 59, no. 9, 09 2022, pp. 6165–6168.
- [22] O. M. Saad, Y. A. S. I. Obou, M. Bai, L. Samy, L.-Q. Yang, and Y.-K. Chen, “Self-attention deep image prior network for unsupervised 3-D seismic data enhancement,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, no. 2, pp. 1–14, 09 2022.
- [23] J.-T. Li, X.-M. Wu, and Z.-X. Hu, “Deep learning for simultaneous seismic image super-resolution and denoising,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, no. 1, pp. 1–11, 01 2022.
- [24] C. Rich, *Multitask Learning*, 05 1998, vol. 60, no. 1, pp. 95–133.
- [25] Y.-C. Gu, X.-T. Wang, L.-B. Xie, C. Dong, G. Li, Y. Shan, and M.-M. Cheng, “VQFR: Blind face restoration with vector-quantized dictionary and parallel decoder,” *European Conference on Computer Vision*, vol. 43, no. 10, pp. 3365–3387, 05 2022.
- [26] X.-M. Wu, L.-M. Liang, Y.-Z. Shi, and S. Fomel, “Fault-Seg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation,” *Geophysics*, vol. 84, no. 3, pp. IM35–IM45, 04 2019.
- [27] X.-M. Wu, Z.-C. Geng, Y.-Z. Shi, N. Pham, S. Fomel, and G. Caumon, “Building realistic structure models to train convolutional neural networks for seismic structural interpretation,” *Geophysics*, vol. 85, no. 4, pp. WA27–WA39, 01 2020.
- [28] C. John, “A computational approach to edge detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, 11 1986.
- [29] W.-Z. Shi, J. Caballero, F. Huszr, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z.-H. Wang, “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” vol. 6, no. 1, 12 2016, pp. 1874–1883.
- [30] Z.-H. Wang, J. Chen, and S. C. H. Hoi, “Deep learning for image super-resolution: A survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 10, pp. 3365–3387, 03 2021.
- [31] C. Roberto, G. Yarin, and K. Alex, “Multi-task learning using uncertainty to weigh losses for scene geometry and semantics,” *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7482–7491, 06 2018.
- [32] P. K. Diederik and B. Jimmy, “Adam: A method for stochastic optimization,” *arXiv e-prints*, p. 1412.6980, 12 2014.
- [33] X.-M. Wu and D. Hale, “3D seismic image processing for faults,” *Geophysics*, vol. 81, no. 2, pp. IM1–IM11, 02 2016.



Fan Min (M09) received the M.S. and Ph.D. degrees from the School of Computer Science and Engineering, University of Electronics Science and Technology of China, Chengdu, China, in 2000 and 2003, respectively. He visited the University of Vermont, Burlington, Vermont, from 2008 to 2009.

He is currently a professor with Southwest Petroleum University, Chengdu. He has published more than 100 refereed papers in various journals and conferences, including IEEE Transactions on SMC: Systems, Pattern Recognition, Knowledge-Based Systems, Computers and Geosciences, and Information Sciences.



Linrong Wang was born in Pengzhou, Sichuan, China in 1999. She is a graduate student with Southwest Petroleum University, Chengdu, China. She has published a paper in Expert Systems With Applications. Her research interest is seismic inversion.



Shulin Pan was born in DeZhou, Shandong, China in 1979. He received the M.S. and Ph.D. degrees in Geophysical Prospecting and Information Technology from Chengdu University of Technology in 2005 and 2008. Since 2011, he is a Professor with Southwest Petroleum University, sichuan, china.

His research interest includes automatic first-arrival time picking, Seismic data Interpolation and AVAZ (amplitude versus incident and azimuthal angle) inversion. His research has been supported by Science and Technology Cooperation Project of the CNPC-SWPU Innovation Alliance (2020CX020000), Open Projects Fund of the State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation (grant no. PLN201733), and National Natural Science Foundation of China (grant nos. 41204101, 41704134).



Guojie Song received the M.S. and Ph.D. degrees from the Department of mathematical science, Tsinghua University, Beijing, China, in 2008 and 2011, respectively.

He is currently a professor with Southwest Petroleum University, Chengdu. He has published more than 50 refereed papers in various journals and conferences, including the Geophysics, BSSA, et. al. His current research interests include forward modeling and inversion of seismic wave, deep learning.