# Deep Learning for Simultaneous Seismic Image Super-Resolution and Denoising

Jintao Li, Xinming Wu, and Zhanxuan Hu

***Abstract*— Seismic interpretation is often limited by low resolution and strong noise data. To deal with this issue, we propose to leverage deep convolutional neural network (CNN) to achieve seismic image super-resolution and denoising simultaneously. To train the CNN, we simulate a lot of synthetic seismic images with different resolutions and noise levels to serve as training data sets. To improve the perception quality, we use a loss function that combines the $\ell_1$ loss and multiscale structural similarity loss. Extensive experimental results on both synthetic and field seismic images demonstrate that the proposed workflow can significantly improve the perception of quality of original data. Compared to conventional methods, the network obtains better performance in enhancing detailed structural and stratigraphic features, such as thin layers and small-scale faults. From the seismic images super-sampled by our CNN method, a fault detection method can compute more accurate fault maps than from the original seismic images.**

***Index Terms*— Deep learning, geophysical image processing, image denoising, super-resolution.**

## I. Introduction

SEISMIC interpretation is sensitive to the quality of seismic data. Due to the limitations of seismic acquisition and processing, the field seismic data often suffer from low resolution and noise corruption, which bring challenges to subsequent seismic interpretation. Two potential technologies to solve these issues are image super-resolution and image denoising.

In the last two decades, many researchers have developed numerous methods to increase the resolution of seismic images. These methods can be roughly grouped into two categories: high-density acquisition [1], [2] and broadband seismic [3]–[5]. The former, as the name suggests, generally increases the horizontal resolution by acquiring a densely sampled data set utilizing a larger number of sources and receivers. While the latter improves the vertical resolution by recording a full range of frequencies, including low- and high-frequency parts. However, both of them are costly in data acquisition and processing. Because they require more sources and receivers to increase the density of samples, better instruments to record a wider range of frequencies during acquisition. Besides, a significantly more computational cost is required for processing the data sets.

For the second potential technology: seismic image denoising, a number of effective methods have been proposed [6]–[11]. These methods enhance the structural and stratigraphic features and attenuate random noise in seismic image by constructing structure-oriented filters to smooth a seismic image along reflections. To construct such structure-oriented filters, researchers can utilize anisotropic diffusion [6], [8], [11], the steered Kuwahara filter [7], plane-wave prediction [9], and steered bilateral filter [10]. Wu and Guo [11] proposed a method to simultaneously enhance reflections, faults, and channels in a seismic image by using the fast explicit diffusion (FED). Although these methods can attenuate the random noise and enhance the structural and stratigraphic features, they also damage some useful details of geological structural features in the seismic image.

In recent years, with the advancement of hardware computing power, especially graphics processor units (GPUs), many deep learning methods have been proposed and achieved success in many computer vision tasks including natural image super-resolution and denoising [12]–[15]. These methods have used deep convolutional neural network (CNN) to achieve remarkable performance. And once the model is well-trained, it takes only a brief amount of time in the application. Inspired by these CNNs, this article deviates from the traditional methods and leverages deep CNN to achieve seismic image super-resolution and denoising simultaneously. Nevertheless, directly using such methods to address the seismic image often encounter two significant issues. The first issue is the lack of training data. Unlike natural images, we cannot obtain a large amount of noise-free field seismic images with high resolution as labels. Some authors [16] select a modern, high-fidelity 3-D seismic survey with well-imaged faults as the donor survey. However, it may be insufficient for training. The second is perception quality. Existing CNN-based methods generally use a mean absolute error ($\ell_1$) loss, which tends to generate blurry and overly smoothed results, especially near the faults, and therefore limits the subsequent seismic interpretation.

To tackle the first issue, we follow the workflow proposed by Wu *et al.* [17] and [18] and generate 800 synthetic seismic volumes. Subsequently, we extract plenty of 2-D inputs from generated 3-D volumes to serve as the training data sets. Besides, to tackle the second issue, we replace the $\ell_1$ loss

Manuscript received June 30, 2020; revised November 6, 2020, January 6, 2021, and January 30, 2021; accepted January 30, 2021. Date of publication February 26, 2021; date of current version December 6, 2021. This work was supported by the National Science Foundation of China under Grant 41974121. *(Corresponding author: Xinming Wu.)*

Jintao Li and Xinming Wu are with the School of Earth and Space Sciences, University of Science and Technology of China, Hefei 230026, China (e-mail: lijintao@mail.ustc.edu.cn; xinmwu@ustc.edu.cn).

Zhanxuan Hu is with the School of Computer Science and OPTIMAL, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: huzhanxuan@mail.nwpu.edu.cn).

Digital Object Identifier 10.1109/TGRS.2021.3057857

with a new objective, a combination of $\ell_1$ loss and multiscale structural similarity (MS-SSIM) [19] loss, which is also used in computer version task [20]. The network used in our method is a variant of U-net [21], which introduces a subpixel layer [22] and several residual blocks. The details can be found in Section IV. Some machine learning algorithms [23] leverages multikernel learning to improve accuracy and stability. To validate the performance of the proposed method, we conduct extensive tests on both synthetic and field seismic data. And the experimental results demonstrate that the network trained on only the synthetic data can significantly improve the perception quality of field seismic data, and the detailed structural and stratigraphic features are enhanced, such as thin layers and small-scale faults. From the seismic images super-sampled by our CNN method, a fault detection method can compute a more accurate fault map than from the original seismic images.

## II. Problem Definition

Image super-resolution and image denoising are both low-level vision tasks and processed similarly. We first analyze the principle of the two problems. For seismic super-resolution, the relationship between high-resolution seismic images and low-resolution seismic images follows the formula:

$$I_L = \text{Down}(I_H, \Theta_1) \tag{1}$$

where $I_L$ and $I_H$ represent as the low-resolution seismic images and high-resolution seismic images, respectively, Down denotes a degradation mapping function and $\Theta_1$ is the parameters of Down. The goal of super-resolution is to reconstruct $I_S$ to approximate the high-resolution seismic images $I_H$ from the low-resolution seismic images $I_L$ through a CNN or other methods.

For seismic denoising, if $I_p$ and $I_N$ denote the clean and noisy seismic images, respectively, then the relationship can be expressed as follows:

$$I_N = I_p + n \tag{2}$$

where $n$ is the noise added to $I_p$. Our purpose is to obtain the noise distribution $n$ and then subtract it from noise seismic images $I_N$ to obtain output images that approximate the clean seismic images $I_p$.

In this work, we tackle those two issues simultaneously. The noise-free seismic images with high resolution and the noise seismic images with low resolution are used as ground truth $I_{gt}$ and input $I_{\text{input}}$, respectively. We recover a seismic approximation $I_{\text{output}}$ of the ground truth $I_{gt}$ from the input, following:

$$I_{\text{output}} = N(I_{\text{input}}, \Theta_2) \tag{3}$$

where $N$ denotes the CNN used, and $\Theta_2$ are the parameters of the network. We aim to leverage deep CNN to achieve seismic image super-resolution and denoising simultaneously. The details of the used CNN model are as follows.

## III. Training Data Sets

Before training a model for super-resolution and denoising together, we need many 2-D high resolution pure seismic images as the ground truth. In practice, however, such data sets are rare. To this end, we follow a workflow provided by Wu *et al.* [17] and [18] to build a realistic structure model, and then extract thousands of 2-D sections for training.

### A. Generate Training Data

We first generate 800 synthetic 3-D seismic cubes with size $256 \times 256 \times 256$, as shown in Fig. 1. In this workflow, we first build an initial reflectivity model with all flat layers. Subsequently, we need to add some structures to simulate field seismic image. The folding and faulting structures are the most important structures of field seismic data. We vertically sheared the initial model to create folding structures and then utilize volumetric vector fields to simulate faulting in the model [Fig. 1(a)]. In this way, we obtain a reflectivity model with realistic folding and faulting structures. We further convolve the generated model with a wavelet to simulate synthetic seismic volumes.

As the conventional methods declared in the introduction, a super-resolution seismic image owns a wider frequency band than a native seismic image in general. According to this principle, we convolve the generated reflectivity model [Fig. 1(a)] with a high-frequency wavelet to obtain the 3-D seismic volume [Fig. 1(b)] from which 2-D label images are extracted. To generate the corresponding input training images, we first convolve the same reflectivity model [Fig. 1(a)] with a low-frequency Ricker wavelet to obtain a relatively low-resolution seismic volume [Fig. 1(c)] where random noise is further added to obtain a more realistic seismic volume [Fig. 1(d)]. From the noisy and low-frequency seismic volume, we finally extract the same 2-D sections and downsample them as the input training images. In this way, we obtain training image pairs with the same structures but different resolutions.

Fig. 2(a) and (b) shows the frequency spectrum of 2-D seismic sections extracted from the automatically generated low-resolution [Fig. 1(c)] and high-resolution [Fig. 1(b)] seismic volumes, respectively. We observe that the spectrum [Fig. 2(b)] of a high-resolution seismic section shows a wider frequency band with significant more high-frequency components than the one [Fig. 2(a)] of a low-resolution section. In generating the training data pairs of low-resolution (input) and high-resolution (label or output) seismic images, the peak frequency of the wavelets are randomly chosen, however, we make sure the spectrum band of the high-resolution seismic image is always wider than the one of the corresponding low-resolution image. And the range of peak frequencies is 5–25 Hz. Using randomly varying peak frequencies for different training data pairs is helpful to train a better generalized model for different field data sets which typically show different peak frequencies.

In generating the input seismic data for training, we have added random colored noise into the data as shown in Figs. 1(d) and 3(b), where the added noise looks more realistic than the simple white noise in Fig. 3(a). In order to increase the diversity and generalization of the training data, the signal-to-noise ratio (SNR) for each training sample is randomly defined in the range of [4, 14].

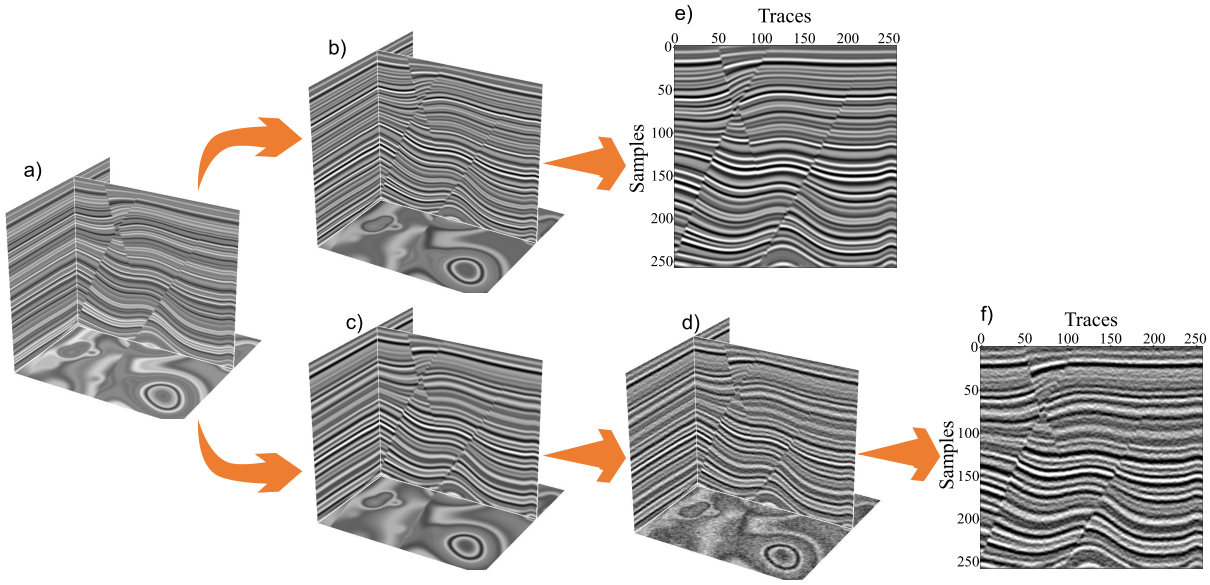


Fig. 1. From a folded and faulted reflectivity model (a), we compute two seismic volumes in (b) and (c) by convolving a high-frequency and low-frequency Ricker wavelet, respectively. From the high-frequency volume, we extract 2-D slices (e) as training labels. In the low-frequency volume, we further add random noise to obtain a noisy seismic volume (d) from which we extract 2-D slices (f) and then downsample them to obtain input training data.

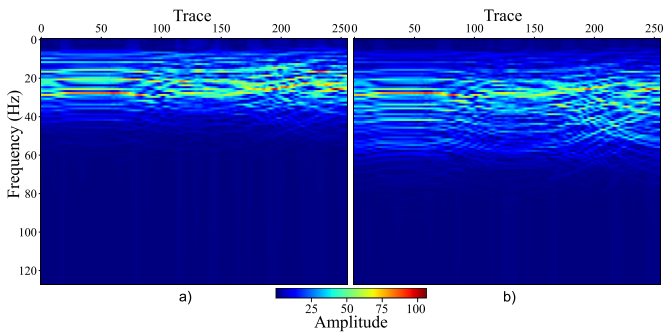


Fig. 2. Amplitude spectrum of 2-D seismic sections extracted from Fig. 1(c) and (b), respectively: (a) spectrum map of the seismic volume convolved with low-frequency wavelets and (b) spectrum map of the seismic volume convolved with high-frequency wavelets.

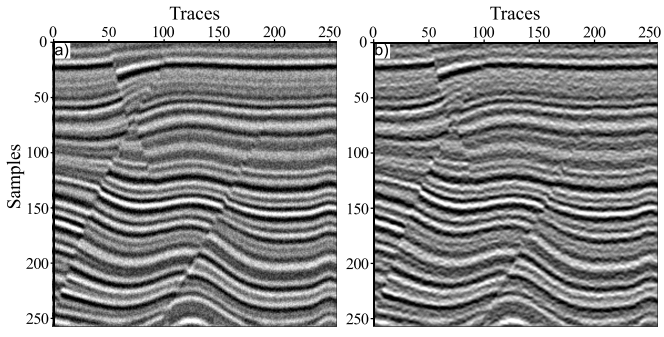


Fig. 3. Comparison of synthetic seismic images with different noise: (a) synthetic image with white noise and (b) synthetic image with colored noise.

To prepare training data sets, we must generate many pairs of 2-D seismic images $I_{gt}$ and $I_{\text{input}}$. Inline or crossline 2-D seismic sections, used as high-resolution images ($I_{gt}$) are extracted from 3-D synthetic seismic volumes with

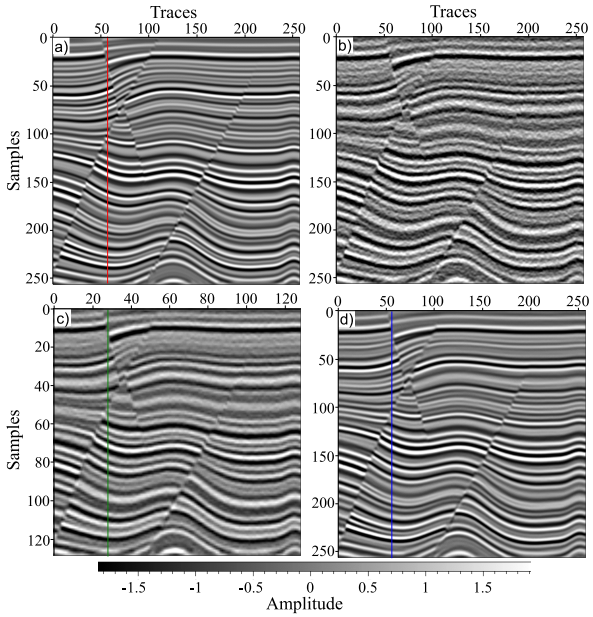


Fig. 4. Experimental results on synthetic seismic data: (a) clean and high resolution seismic section extracted from Fig. 1(b); (b) same section extracted from Fig. 1(d); (c) input seismic section downsampled from (b); and (d) recovered seismic section using our method.

high-frequency wavelets [Fig. 4(a)]. And the low-resolution seismic images ($I_{\text{input}}$) [Fig. 4(c)], used as input, are obtained by downsampling the same 2-D sections by a factor of 2 [Fig. 4(b)] which are generated from the volumes with random noise. We expect our network to upscale the input downsampled image and broaden its frequency band like the conventional methods which achieve the super-resolution by using dense receivers and broad frequencies during the data acquisition. In our experiment, we divided 600 of the total 800 3-D volumes to contribute training set, 75 volumes for

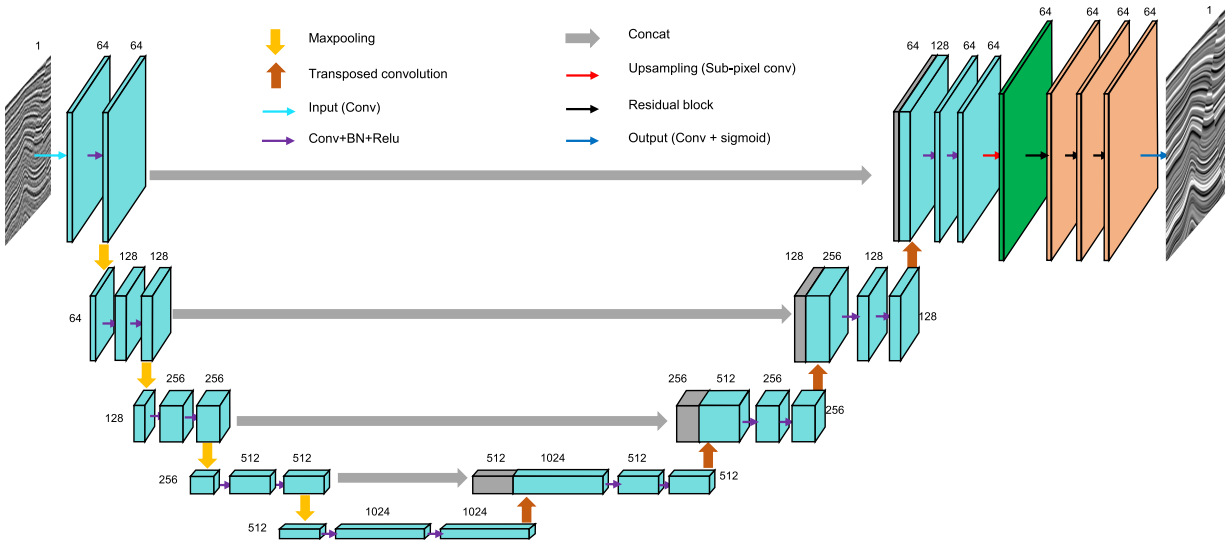


Fig. 5. Network architecture used in our proposed method.

validation, 75 volumes for the test set. The rest 50 volumes are retained to contribute the slices for fault detection. From each 3-D volume, we extract two 2-D sections for generating one of training/validation/testing bins. Our work aims at reconstructing the high-resolution images $I_{\text{output}}$ from low-resolution images with noise $I_{\text{input}}$. In particular, $I_{\text{output}}$ is expected to close to the original high-resolution images $I_{gt}$.

### B. Data Augmentation

Data augmentation is one of the most useful methods for improving the performance of deep models. There are some successful instances in geophysics [24]. In order to avoid using a large memory footprint for training, we first crop the pairs of 2-D seismic images into some small patches at random. However, the information contained in a small patch is often insufficient to recovery the details between $I_{\text{input}}$ and $I_{gt}$ [25]. Thus, we choose the size of the input seismic patches to be $96 \times 96$ to balance the above problems, and the corresponding size of the high-resolution seismic patches is $192 \times 192$. In addition, we add some simple geometric manipulation with randomly horizontal flipping the pairs of patches to increase the diversity of training data sets.

## IV. Architecture and Training Details

The network of this article is a variant of U-net. To obtain more realistic results, we use a MS-SSIM loss function [20] to avoid the overly smoothed structural edges. Finally, we estimate our method on synthetic data sets.

### A. CNN Architecture

The network architecture used in our method is illustrated in Fig. 5, which consists of three parts: a stand *U-net*, a *subpixel layer* and several *residual blocks*. The *U-net* is an encoder-decoder network and includes four downsampling blocks and corresponding upsampling blocks. Each downsampling blocks consists of a max-pooling layer with kernel $2 \times 2$ and stride 2, two convolution layers with kernel $3 \times 3$, and each convolution layer is followed by a batch normalization layer and a rectified linear unit (ReLU). The number of feature channels are 64, 128, 256, 512, and 1024, respectively, through four downsampling blocks. Upsampling block is an opposite design with the downsampling block, and it enlarges the feature size by a transposed convolution layer. Then, we concatenate the output of the transposed convolution layer with the feature maps from the downsampling block at the same level. After that, the output is fed to two convolution layers and the layers yield the same design as downsampling except for the feature number.

The goal of introducing *subpixel convolution layer* [22] is to conduct upsampling. As another way to increase the resolution by a dense acquisition without spatial aliasing, we want to simulate the methods to achieve high-resolution by the subpixel convolution layer. Besides, the image size is small in the main part of our network, i.e., the U-net. Thus, we can reduce training time and save GPU memory. In practice, if we remove the subpixel convolution layer and the input size is $256 \times 256$, the training time will be two to three times longer. Here, we first increase the feature channels by convolution and then reshaping them to enlarge the resolution of inputs. Unlike transposed convolution used in the first part, the *subpixel convolution layer* provides more contextual information through a larger receptive field, which is beneficial for generating more realistic details [26].

The last part, i.e., the *residual blocks* learn more high-frequency information and details from input to target. Each residual block contains two convolution layers, and each convolution layer is followed by a batch normalization layer and a ReLU. Then, a skip connection covers the two convolution layers. In practice, we achieve good performance with just three residual blocks. When we removed these three residual blocks, the peak signal-to-noise ratio (PSNR) on the test set dropped from 29.068 to 28.803. It is more obvious to observe the performance improvement by using the residual blocks. Fig. 6 shows a field example using the

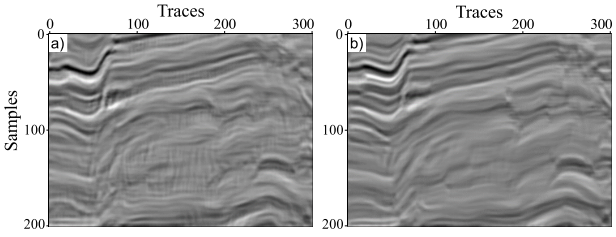


Fig. 6. Comparison test in a field example: (a) result without the residual blocks and (b) result with the residual blocks.

trained U-Net with and without three residual blocks. The result with residual blocks shows fewer artifacts and more recovered details. Finally, we use a convolution layer with kernel $1 \times 1$ to reduce the number of feature channels to match the ground truth.

### *B. Loss Functions*

We train our network using a new loss function that combines the $\ell_1$ loss and MS-SSIM loss. Due to the advantage in improving the performance and convergence over mean squared error (MSE) or $\ell_2$ loss [12], $\ell_1$ loss has been widely used for image super-resolution. Mathematically, $\ell_1$ loss is defined as

$$\mathcal{L}^{\ell_1} = \frac{1}{N}\sum_{i,j} |I_{\mathrm{SR}}(i,j) - I_{\mathrm{HR}}(i,j)| \tag{4}$$

where $N$ is the total number of pixels. In practice, however, the network trained using only the $\ell_1$ loss will generate unsatisfying high-resolution images with smooth textures. The reason is that $\ell_1$ loss minimizes the only pixel-wise distance between output and target and ignore the texture structures. To tackle this issue, we introduce a more sophisticated loss term derived from MS-SSIM, which is also used in computer version tasks [20].

MS-SSIM, an assessment for image quality, is sensitive to local structure variations and more appropriate for the human visual system (HVS). It is an improved version of SSIM [27] and can be mathematically defined as

$$\mathrm{SSIM}(\mathbf{x},\mathbf{y}) = [l(\mathbf{x},\mathbf{y})]^{\alpha} \cdot [c(\mathbf{x},\mathbf{y})]^{\beta} \cdot [s(\mathbf{x},\mathbf{y})]^{\gamma} \tag{5}$$

where

$$\begin{aligned} l(\mathbf{x},\mathbf{y}) &= \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \\ c(\mathbf{x},\mathbf{y}) &= \frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \\ s(\mathbf{x},\mathbf{y}) &= \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}. \end{aligned} \tag{6}$$

Here, $\mathbf{x}$ and $\mathbf{y}$ are two images. $\mu_i$, $\sigma_i$ represent the mean and the standard deviation of image $i$, $\sigma_{xy}$ is the covariance between images $\mathbf{x}$ and $\mathbf{y}$. And $c_1, c_2, c_3$ are three constants to avoid the situations where the denominator is too small to be stable. $l(\mathbf{x},\mathbf{y})$, $c(\mathbf{x},\mathbf{y})$, $s(\mathbf{x},\mathbf{y})$ represent three measurements between $\mathbf{x}$ and $\mathbf{y}$: luminance or amplitude in seismic image, contrast and structure, respectively. $\alpha, \beta, \gamma$ are the corresponding weights of three measurements and have to be positive.

TABLE I
PSNR ON THE SAME TEST DATA SET BUT WITH DIFFERENT LOSS FUNCTION WEIGHTS $\alpha$

| $\alpha$ | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 |
|---|---|---|---|---|---|
| PSNR | 28.880 | 28.958 | 29.007 | 29.068 | 28.904 |

In general, we set $c_1 = 1 \times 10^{-4}, c_2 = c_3/2 = 9\ c_1$. And MS-SSIM is defined as

$$\mathrm{MS\text{-}SSIM}(\mathbf{x},\mathbf{y}) = \left[l_M(\mathbf{x},\mathbf{y})\right]^{\alpha_M} \cdot \prod_{j=1}^{M} \left[c_j(\mathbf{x},\mathbf{y})\right]^{\beta_j} \left[s_j(\mathbf{x},\mathbf{y})\right]^{\gamma_j} \tag{7}$$

where $M$ is the scale. In generally, $M$ is set as 5, $\alpha_5 = 0.1333$, $\alpha = \beta = \gamma = [0.0448, 0.2856, 0.3001, 0.2363, 0.1333]$. The *multiscale* means that we measure the SSIM in different scales, i.e., we first zoomed-out view the pairs by a factor $2^{j-1}$, then calculate each term of SSIM and finally multiply them together with the corresponding weights $\alpha_j, \beta_j, \gamma_j$. We zoomed-out view the image pairs by simply using an average pool layer. It must be noted that the range of MS-SSIM values is not 0 to 1. Because the covariance $\sigma_{xy}$ can be a negative value which may lead $s(\mathbf{x},\mathbf{y})$ to be a negative number. To modify it as a loss function, we need to normalize it to 0 to 1, and we can simply perform MS-SSIM = (MS-SSIM + 1)/2 to achieve the goal. A ReLU layer can also be used for the same purpose. The higher the value of A, the closer the two pictures are.

To improve the perception quality of recovered images, we combine the $\ell_1$ loss and MS-SSIM loss and obtain a new loss function defined as

$$\mathcal{L}^{\mathrm{Mix}} = \alpha \cdot \mathcal{L}^{\mathrm{MS\text{-}SSIM}} + (1-\alpha) \cdot \mathcal{L}^{\ell_1} \tag{8}$$

where

$$\mathcal{L}^{\mathrm{MS\text{-}SSIM}} = 1 - \mathrm{MS\text{-}SSIM}(I_{\mathrm{SR}}, I_{\mathrm{HR}}) \tag{9}$$

and $\alpha$ is the weight of loss function and we empirically set $\alpha = 0.6$. Table I shows the PSNR values on the same test data set with different weights $\alpha$. It shows that our network achieves the best performance when $\alpha = 0.6$, A field example in Fig. 7 also visually demonstrates that $\alpha = 0.6$ is the best choice. When $\alpha = 0.6$, the recovered seismic image yields fewer artifacts and looks more realistic.

The comparison between $\ell_1$ loss and mix loss is illustrated in Fig. 8. The three seismic images are extracted from the upper left region shown in Fig. 4 using different loss functions and their corresponding ground truth. It is modestly that the output of the mix loss [Fig. 8(b)] shows sharper discontinuities near faults than the $\ell_1$ loss [Fig. 8(c)] in faults. The mix loss leads to realistic and perception results. This phenomenon is suggestive where the areas are denoted by the red arrows. For example, near the area highlighted by the left arrow in the $\ell_1$ loss result [Fig. 8(c)], the fault is smoothed out and the reflections are continuous across the fault, which is untrue compared to the ground truth [Fig. 8(a)]. This will mislead the following seismic interpretation tasks such as fault detection. By using the mix loss, we are able to better preserve the fault discontinuities as denoted by the red arrows in Fig. 8(b).

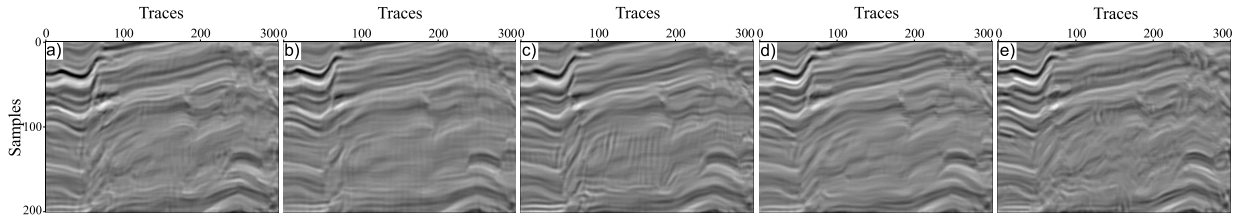


Fig. 7. Results (a)–(e) by using a loss function weight ($\alpha$) of 0.3, 0.4, 0.5, 0.6, 0.7, respectively.

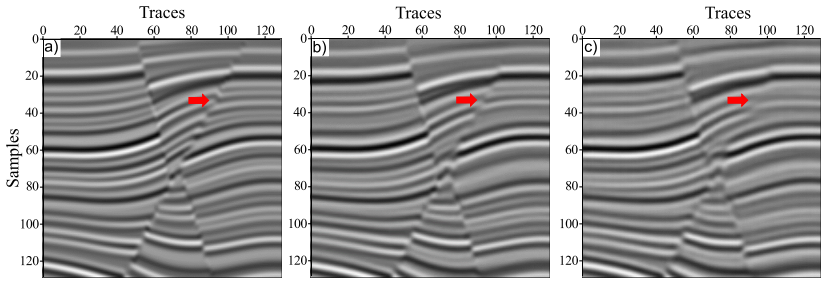


Fig. 8. Comparison between the ground truth (a), the result of the proposed mix loss (b), and traditional $\ell_1$ loss (c). These subimages are extracted from the upper left region shown in Fig. 4.

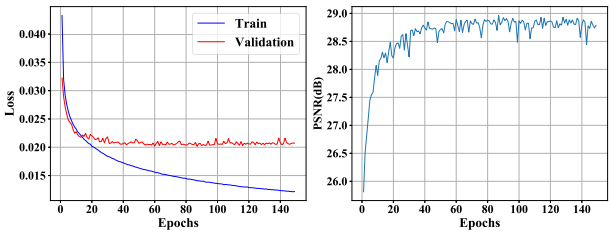


Fig. 9. Training record: (a) loss function values on training and validation data sets and (b) performance curves of PSNR on validation data sets.

### C. Training Details

As discussed in the section of generating training data, we generated 1200 pairs of the 2-D image for training. And every input seismic image and label seismic image is normalized to [0, 1] by the following formula:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{10}$$

where $x^*$ is the normalized seismic image, $x_{\max}$ and $x_{\min}$ are maximum value and minimum value of each input seismic image, respectively. We then preprocess all the seismic images by data augmentation discussed before.

We train our model with ADAM optimizer [28] and set the parameter $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$. The learning rate is initialized to $1e-4$. We set batch size as 16 and totally extract $16 \times 1000$ patch pairs from training data sets. We train our network over 150 epochs. We provide the training details in Fig. 9, where Fig. 9(a) reports the loss function values on training and validation data sets; and Fig. 9(b) reports the performance curves of PSNR on validation data sets. Although the loss function and PSNR curves do not converge until nearly 100 epochs, it only takes us about 4 h to finish a training task that works in NVIDIA Tesla V100.

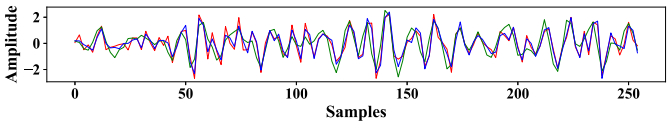


Fig. 10. Three traces extracted from the same position where the vertical lines shown in Fig. 4. The green, blue, and red curves, respectively, represent the traces extracted from the input, output, and ground truth seismic images.

We first evaluate the performance of our CNN model on synthetic seismic images, i.e., the test sets that are not involved in training and validation. The experimental results are shown in Fig. 4, where Fig. 4(a) is the pure seismic images with high resolution used as ground truth, Fig. 4(c) and (d) are the input noise seismic image with low resolution and the output seismic image recovered by our CNN model respectively. Compared with the low-resolution noise seismic image, the recovered seismic image provides enhancing structural features and sharper geologic edges, especially in faults and seismic horizons. And the thin layers are recovered well even if there are only some blurred traces almost invisible to human eyes on the low-resolution image. In addition, the result also offers an effect of denoising. It is obvious that the seismic section between two seismic horizons of recovered images is more smoothed and cleaner compared with the input seismic images.

Furthermore, we compare the amplitude characteristics of three traces and report the results in Fig. 10. These three traces are extracted from the output seismic section (red) and the corresponding pair of the low-resolution seismic image (green) and the ground truth (blue) where the corresponding color vertical lines. The waveforms of three curves are in approximate agreement keeping the shape. But the red one yields more details than the green curve. The ground truth curve maintains similar characteristics with the output trace. This characteristic is well manifested in the range of samples

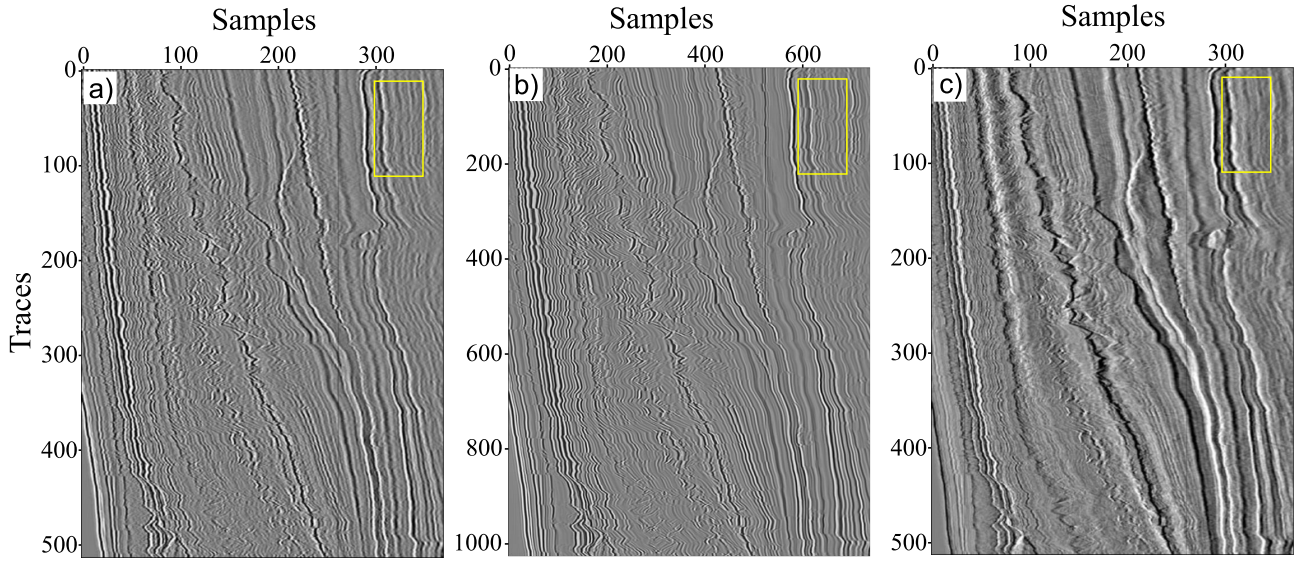


Fig. 11. Comparison between our method and BroadSeis: (a) 2-D native field seismic section; (b) result recovered by our method; and (c) result by BroadSeis.

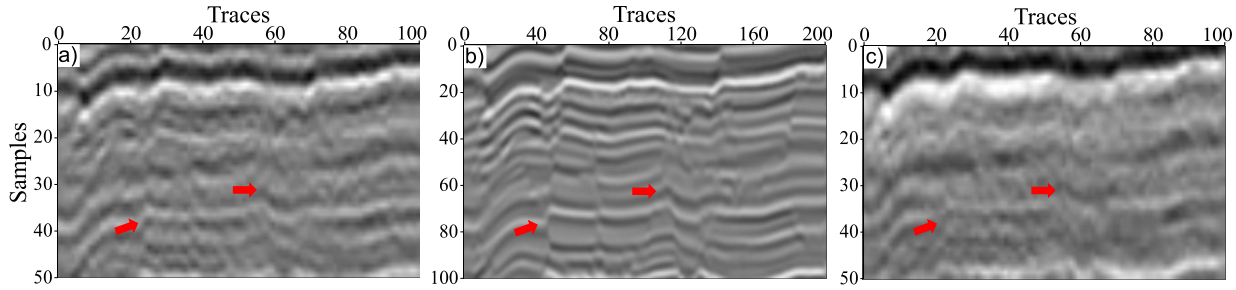


Fig. 12. Comparison of Fig. 11 in detail: (a), (b), and (c) are three patches (the yellow boxes) extracted from Fig. 11.

0–50 of the curve, where many thin layers are covered by random noise in Fig. 4(c). These details may be the faults or thin layers that appear blurry or have a small change of amplitude compared to surroundings in low-resolution seismic images. In a words, the clean and high-resolution seismic sections with enhanced faults and thin layers are generated by applying the input seismic images to the CNN model, and can facilitate subsequent seismic interpretation. Our method of simultaneous super-resolution and denoising is highly efficient. It takes only several seconds to process all 150 images, each with a size of $128 \times 128$.

## V. Applications

We feed several field seismic images to the well-trained model and make a comparison with a conventional method. We also deploy the results of our CNN model into fault detection to confirm the ability of our well-trained model.

### A. Several Real Examples

Our CNN model achieves good effectiveness and generalization on both synthetic and field seismic data sets even if it is trained with synthetic seismic data only. To verify the capability of the model, we feed some 2-D field seismic images acquired at different 3-D surveys to the well-trained CNN model. And the sampling interval for each example is 4 ms. Before applying the field seismic images, each image is normalized as same as the synthetic data sets to make it consistent with training. Besides, the dimension size of the input seismic image is not fixed and is only required to be dividable by $2^t$, or we need to resize the input seismic image so that its dimension can be dividable by $2^t$, where $t$ is the downsampling times of the architecture. We use $t = 4$ in our experiment.

Fig. 11 shows a real example where the native seismic image [Fig. 11(a)] is directly captured from the paper of BroadSeis [3]. Feeding this native image into our trained CNN model, we obtain an improved image [Fig. 11(b)] where the noise is effectively removed and the resolution of detailed features (e.g., thin layers and small-scale faults) are significantly improved. Our result [Fig. 11(b)] shows even more details than the improved image [Fig. 11(c)] by the BroadSeis technique. The BroadSeis requires more expensive acquisition costs and computational costs for processing while our CNN-based method requires no extra cost and takes only milliseconds to compute the result shown in Fig. 11(b). Fig. 12(a)–(c) show a zoomed-in view of the yellow boxes in the native image, our result, and the BroadSeis image, respectively. From these

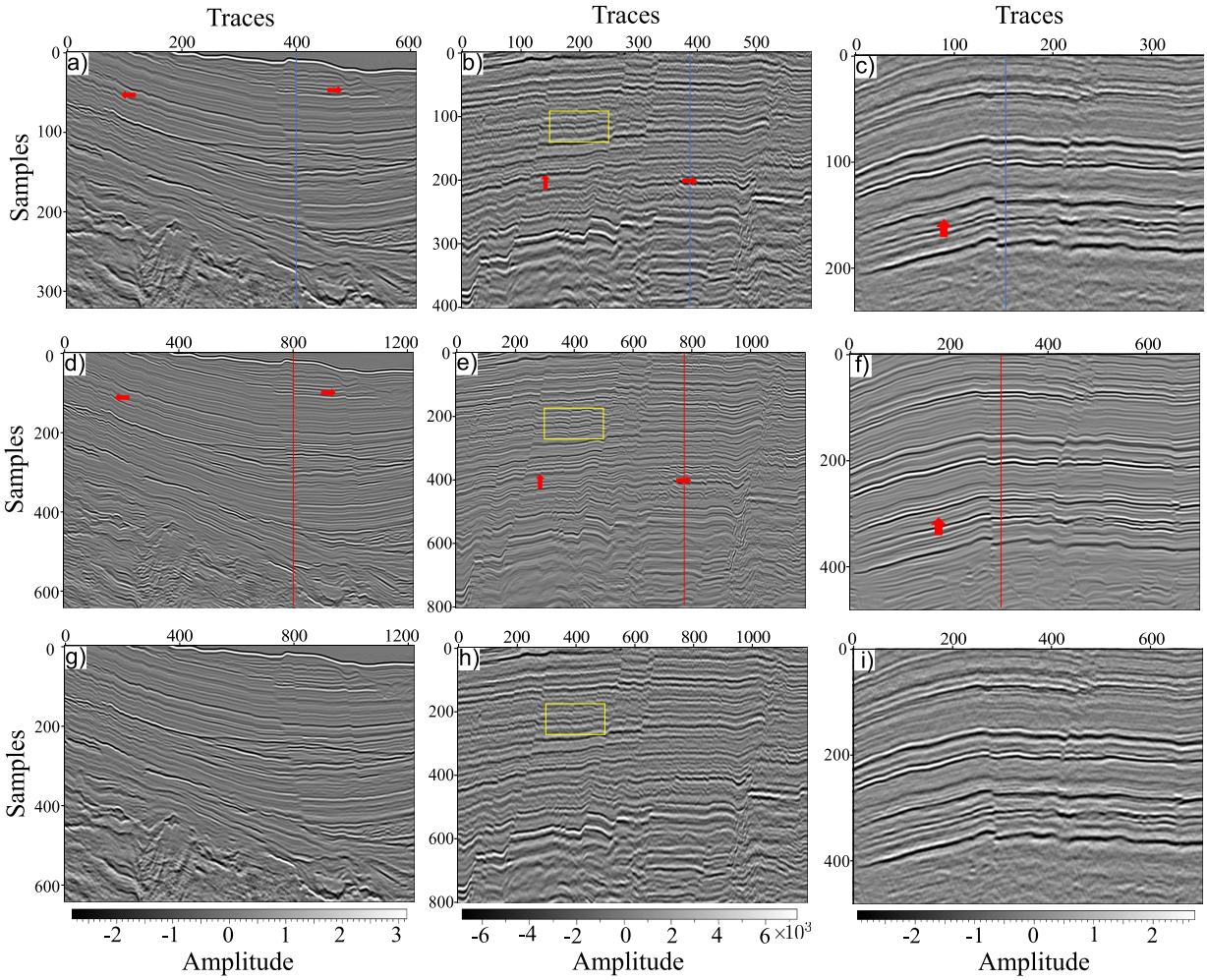


Fig. 13. Experimental results of proposed method on field seismic data: (a), (b), and (c) are three field seismic sections; (d), (e), and (f) are corresponding recovered results; (g), (h), and (i) are corresponding interpolation results (using bicubic). In particular, (b) shows large amplitude values (see the color bar at the bottom) which are much different from the synthetic seismic sections used for training.

subimages, we can more clearly observe that our CNN-based method [Fig. 12(b)] significantly enhances the detailed structures of thin layers and the faults with small throws (as denoted by the red arrows). Those faults are relatively small-scale ones but they certainly exist. They can be seen roughly in the native seismic image but appear very blurry, while the image after processing by our method provides a better view of those faults.

Fig. 13 shows another three field examples, where Fig. 13(a)–(c) are the three native field seismic images that are acquired at different surveys. Fig. 13(d)–(f) are the corresponding results computed by our CNN-based method of simultaneous super-resolution and denoising. Fig. 13(g)–(i) are the corresponding seismic images obtained by upsampling the native image through bicubic interpolation with a factor of 2. The size of the image obtained by bicubic is consistent with the output image of CNN. The interpolated images look very similar to the native images, which indicates that the conventional interpolation method is not helpful to improve the resolution of details of the original images. Compared to the native images (the top row of Fig. 13) and the corresponding interpolated images (the bottom row of Fig. 13), our results show much more clear structures with noise removed and higher resolution of detailed structural and stratigraphic features such as small-scale faults and thin layers. Three subimages (Fig. 14) are extracted from the yellow box area of the second field data in Fig. 13 to provide a detailed comparison. Although the structural features and the amplitude values in these field images are significantly different from our synthetic training data, our trained CNN model still works well as in the synthetic tests. This indicated that our CNN model, trained with only synthetic data sets, is well generalized for various field data sets.

Fig. 15 shows some seismic traces that are extracted from the native images and our improved images in Fig. 13. Due to the bottom row of Fig. 13 is just an interpolated result and very similar to the native images, we do not provide the traces of the three bottom images. The top subfigure in Fig. 15 shows the traces of first field data shown in Fig. 13(a) and (c), and the traces in the middle figure are extracted from the second data

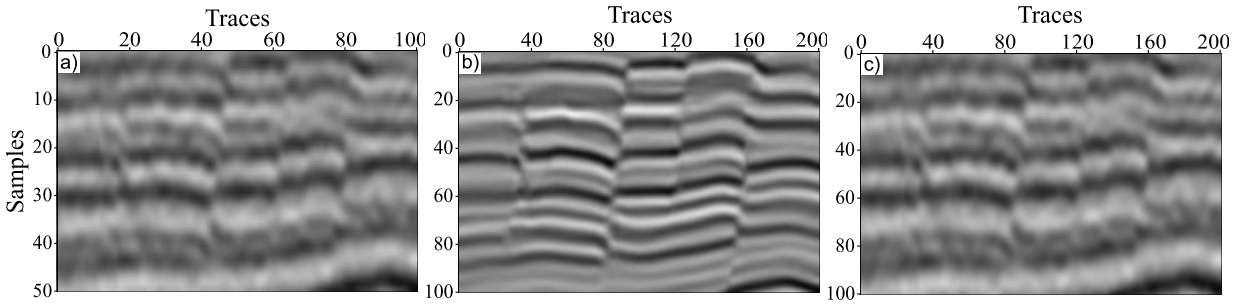


Fig. 14. Comparison in detail: (a), (b), and (c) are extracted from Fig. 13(b), (e), and (h) (the yellow boxes).

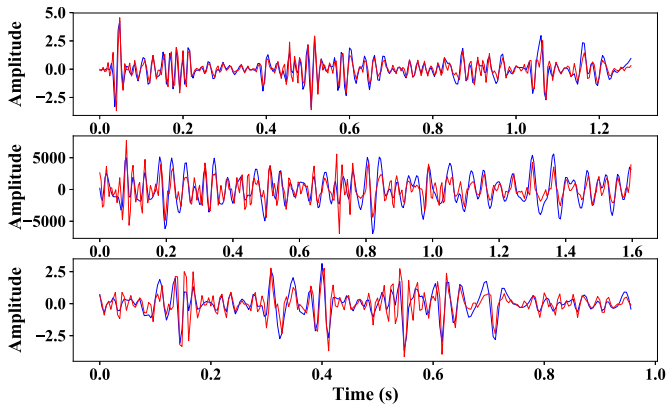


Fig. 15. Traces analysis of the field seismic images (blue) and the results (red) of our CNN method shown in Fig. 13.

[Fig. 13(b) and (d)]. The traces in the bottom subfigure are extracted from the third field data [Fig. 13(c) and (f)]. The blue curves represent the native seismic image traces and the red ones are the traces of our output images. Compared to the blue curves (native images), the red curves (our results) show similar waveform trends and characteristics but much more details.

Spectrum analysis of the three field seismic data is illustrated in Fig. 16 where each frequency amplitude is averaged over all the traces in a 2-D section. The blue and red curves represent the amplitude spectrum of our output seismic sections and the input sections, respectively. As we expect, the frequency bands of our output seismic sections are wider than the bands of native input seismic sections, especially in high-frequency part. It must be noted that some of the recovered detailed features are not necessarily true in the results of our CNN, especially in the areas with quite low data quality, such as the lower left area of Fig. 13(a), where the results are more likely to contain artifacts.

### *B. Fault Detection*

Fault detection is one of the most important tasks in seismic interpretation as faults often indicate the locations of petroleum reservoirs. The experiment on synthetic and field data shows fault detection can be significantly benefited from our CNN method.

We first apply fault detection on 100 synthetic 2-D sections where we know the ground truth of the faults for comparison. Fig. 17(a) shows true faults overlaid on the high-resolution seismic label image. Fig. 17(b)–(d) show the fault maps computed from the seismic label image, the output seismic image of our network, and the input low-resolution and noisy seismic image, respectively, by using the same fault likelihood scanning method [29], [30]. We observe that the fault maps computed from our output seismic image [Fig. 17(c)] and the seismic label image [Fig. 17(b)] show almost the same fault features. Compared to the fault map computed from the input noisy and low-resolution seismic image [Fig. 17(d)], the fault map computed from our output seismic image shows much cleaner and higher resolution fault features. To further quantitatively evaluate the fault detection results, we compute the curves of the pixel-wise accuracy and mean intersection over union (MIoU) based on the ground truth of the faults [Fig. 18(a)]. The MIoU is defined [31] as

$$\text{MIoU} = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \tag{11}$$

where the total number of classes is $k+1$, and $p_{ij}$ is the number of pixels of class $i$ inferred to belong to class $j$. To calculate these two assessments, we assume the calculated faults are accurate when the likelihood value is greater than a threshold. These curves also show that fault detection in our output seismic image (green curves) is significantly more accurate than in the corresponding input seismic image (blue curves). The accuracy of input increases as the threshold increases, while the other two curves change very little. Because the fault detected result of the input seismic image is greatly affected by the noise. This indicates that our network is helpful for the next step of seismic structure interpretation by simultaneously improving the seismic resolution and removing the noise in the seismic image.

We also illustrated a real example. Fig. 19(a) and (b), respectively, show the fault detection results computed from the native seismic image [Fig. 13(b)] and the corresponding super-sampled and denoised seismic image [Fig. 13(e)]. Almost all the individual faults detected on our CNN results are sharper than those computed on the native seismic image. In addition, the fault features are less noise and more continuously tracked. This means that the positions of the faults are predicted more accurately after processing the seismic data by using our CNN method.

In summary, we conclude that our method is indeed beneficial to fault detection, and the images processed by our CNN

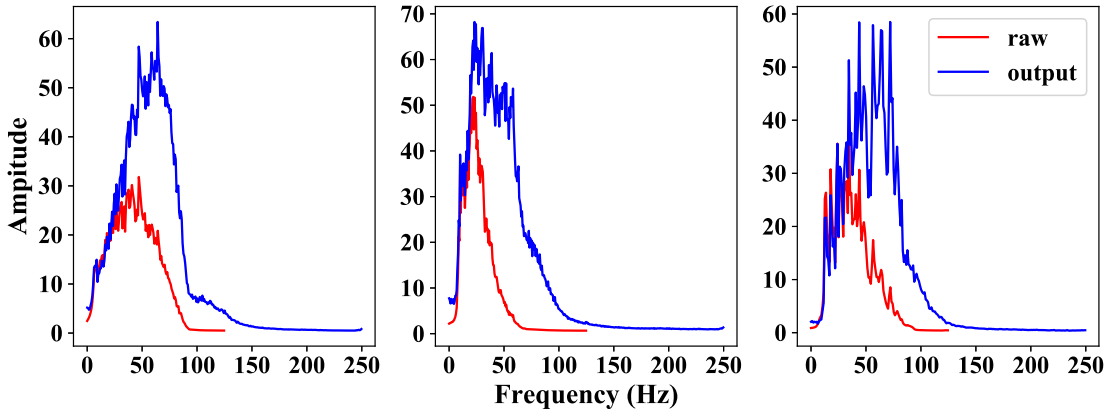


Fig. 16. Spectrum analysis of three field seismic images.

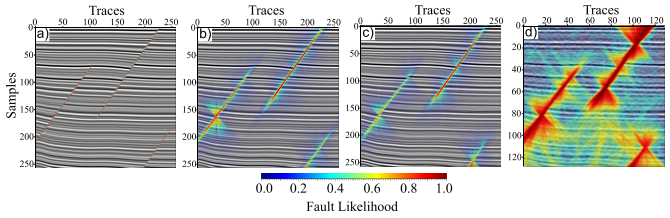


Fig. 17. Comparison of fault detection: (a) true fault map overlaid on the seismic label image; (b) fault likelihood map scanned from the high-resolution seismic label; (c) fault likelihood map scanned from the output seismic; and (d) fault likelihood map scanned from the input low-resolution and noisy seismic.

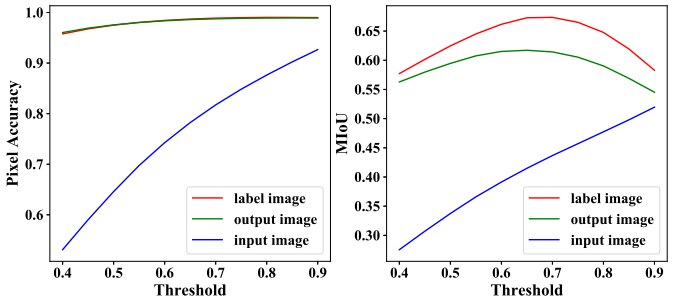


Fig. 18. Quantitative evaluation of the fault maps computed from the high-resolution seismic label image (red), recovered seismic image (green) by our network, and low-resolution and noisy seismic image input to the network: (a) pixel accuracy and (b) MIoU.

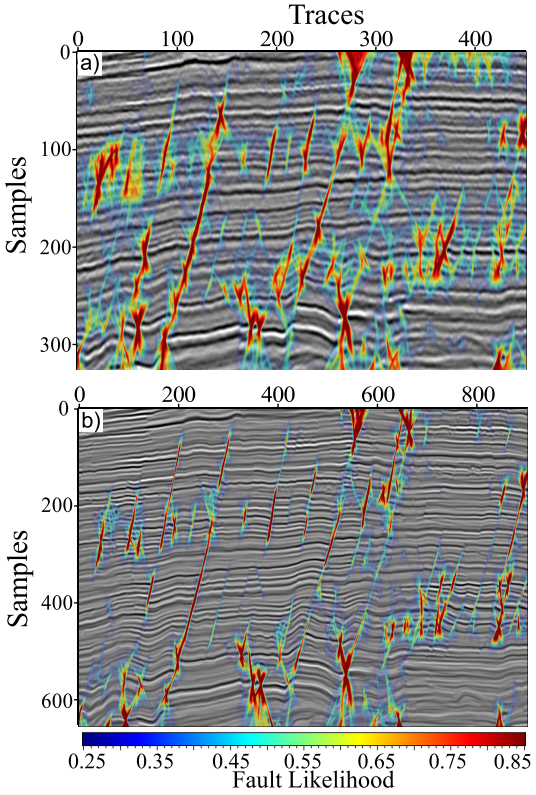


Fig. 19. Comparison of fault detection: (a) fault detection in the native seismic image and (b) fault detection in the seismic image enhanced by our CNN method.

model does provide a better view of fault detection with clearer and sharper fault features, and more accurate fault locations.

## VI. Conclusion

In this article, we developed a novel CNN-based method for achieving seismic image super-resolution and denoising simultaneously. Due to lacking seismic labels in the real world, we generate plenty of synthetic seismic sections to train our CNN. And we use a loss function that combines the $\ell_1$ loss and MS-SSIM loss. That loss function can improve the perception of quality and alleviate overly smoothed geological edges. Our proposed method performs well on both synthetic and field seismic data. Multiple examples demonstrate that our method is able to significantly enhance the detailed structural and stratigraphic features (e.g., thin layers and small-scale faults) in the input seismic images. Besides, a result of fault detection confirms the effectiveness of our method.

However, our proposed method still has some limitations. Some recovered detailed features need to be further verified as some of them (especially those in the areas with low data quality) may be artifacts. And the loss gap between the training and validation may indicate slight overfitting. For future work, we will investigate more details of the tradeoffs between overfitting, training data set size, and architecture.

Besides we want to leverage transfer learning to reduce the gap between synthetic and field seismic data. By doing this, the performance of field seismic data is expected to be further improved.

## REFERENCES


[1] Y.-G. Zhang, Y. Wang, and J.-J. Yin, "Single point high density seismic data processing analysis and initial evaluation," *Shiyou Diqiu Wuli Kantan(Oil Geophys. Prospecting)*, vol. 45, no. 2, pp. 201–207, 2010.

[2] F. Xiao *et al.*, "High-density 3D point receiver seismic acquisition and processing–a case study from the Sichuan Basin, China," *First Break*, vol. 32, no. 1, 2014.

[3] R. Soubaras, R. Dowle, and R. Sablon, "Broadseis: Enhancing interpretation and inversion with broadband marine seismic," *CSEG Recorder*, vol. 37, no. 7, pp. 40–46, 2012.

[4] T. Rebert, R. Sablon, N. Vidal, P. Charrier, and R. Soubaras, "Improving pre-salt imaging with variable-depth streamer data," in *Proc. SEG Tech. Program Expanded Abstr.*, Sep. 2012, pp. 1–5.

[5] Y. Wang, J. Wang, X. Wang, W. Sun, and J. Zhang, "Broadband processing key technology research and application on slant streamer," in *Proc. Int. Geophys. Conf.*, Beijing, China, Apr. 2018, pp. 135–138.

[6] G. C. Fehmers and C. F. Höcker, "Fast structural interpretation with structure-oriented filtering," *Geophysics*, vol. 68, no. 4, pp. 1286–1293, 2003.

[7] N. M. AlBinHassan, Y. Luo, and M. N. Al-Faraj, "3D edge-preserving smoothing and applications," *Geophysics*, vol. 71, no. 4, pp. P5–P11, Jul. 2006.

[8] D. Hale, "Structure-oriented smoothing and semblance," Colorado School Mines, Golden, CO, USA, CWP Rep. 635, 2009.

[9] Y. Liu, S. Fomel, and G. Liu, "Nonlinear structure-enhancing filtering using plane-wave prediction," *Geophys. Prospecting*, vol. 58, no. 3, pp. 415–427, May 2010.

[10] D. Hale, "Structure-oriented bilateral filtering," Colorado School Mines, Golden, CO, USA, CWP Rep. 695, 2011, pp. 239–248.

[11] X. Wu and Z. Guo, "Detecting faults and channels while enhancing seismic structural and stratigraphic features," *Interpretation*, vol. 7, no. 1, pp. T155–T166, Feb. 2019.

[12] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 136–144.

[13] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[14] V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep image prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9446–9454.

[15] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11065–11074.

[16] P. Lu, M. Morris, S. Brazell, C. Comiskey, and Y. Xiao, "Using generative adversarial networks to improve deep-learning fault interpretation networks," *Lead. Edge*, vol. 37, no. 8, pp. 578–583, Aug. 2018.

[17] X. Wu, L. Liang, Y. Shi, and S. Fomel, "FaultSeg3D: 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, 2019.

[18] X. Wu, Z. Geng, Y. 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, Jul. 2020.

[19] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, vol. 2, 2003, pp. 1398–1402.

[20] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 47–57, Mar. 2017.

[21] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2015, pp. 234–241.

[22] W. Shi *et al.*, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.

[23] X. Liu, X. Chen, J. Li, X. Zhou, and Y. Chen, "Facies identification based on multikernel relevance vector machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7269–7282, Oct. 2020.

[24] F. Li, H. Zhou, Z. Wang, and X. Wu, "ADDCNN: An attention-based deep dilated convolutional neural network for seismic facies analysis with interpretable spatial–spectral maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1733–1744, Feb. 2021.

[25] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.

[26] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super-resolution: A survey," 2019, *arXiv:1902.06068*. [Online]. Available: http://arxiv.org/abs/1902.06068

[27] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[28] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[29] D. Hale, "Methods to compute fault images, extract fault surfaces, and estimate fault throws from 3D seismic images," *Geophysics*, vol. 78, no. 2, pp. O33–O43, Mar. 2013.

[30] X. Wu and D. Hale, "3D seismic image processing for faults," *Geophysics*, vol. 81, no. 2, pp. IM1–IM11, Mar. 2016.

[31] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," 2017, *arXiv:1704.06857*. [Online]. Available: https://arxiv.org/abs/1704.06857


**Jintao Li** received the B.S. degree in geophysics from the University of Science and Technology of China (USTC), Hefei, China, in 2020, where he is pursuing the M.S. degree with Computational Interpretation Group (CIG).

His research interests include deep-learning applications on geophysics including seismic super-resolution, denoising, and seismic facies analysis.

**Xinming Wu** received the Ph.D. degree in geophysics from the Colorado School of Mines, Golden, CO, USA, in 2016.

He was a Post-Doctoral Fellow with the Bureau of Economic Geology, University of Texas at Austin, Austin, TX, USA. He is a Professor with the School of Earth and Space Sciences, University of Science and Technology of China (USTC), Hefei, China. His research interests include image processing, machine learning, 3-D seismic interpretation, subsurface modeling, and geophysical inversion.

Dr. Wu received the Society of Exploration Geophysicists (SEG's) awards for J. Clarence Karcher Award, 2020 and Honorary Lecturer, South and East Asia, 2020. He was also a recipient of the Best Paper Award in Geophysics in 2016, the Best Student Poster Paper Award at the 2017 SEG Annual Convention, and the Honorable Mention Award for Best Paper at the 2018 SEG annual convention.

**Zhanxuan Hu** is pursuing the Ph.D. degree with Northwestern Polytechnical University, Xi'an, China.

His research interests include topics in machine learning and its application in geophysics.