Deep learning for simultaneous seismic image super-resolution and denoising

Jintao Li[1], *Xinming Wu*[1] *and Zhanxuan Hu*[2]
[1]*School of Earth and Space Sciences, University of Science and Technology of China,* [2]*Northwestern Polytechnical University.*

SUMMARY

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 design a novel loss function that combines the ℓ_1 loss and multi-scale structural similarity loss. Extensive experimental results on both synthetic and field seismic images demonstrate that the proposed workflow can significantly improve the perception quality of original data. Compared with the conventional method, the network obtains better performance in enhancing detailed structural and stratigraphic features, such as thin layers and small-scale faults.

INTRODUCTION

Seismic interpretation is sensitive to the quality of seismic data. Due to the limitations of seismic acquisition and processing, the field seismic data is often acquired with low resolution and corrupted by noise, which brings challenges to subsequent seismic interpretation. Two potential technologies to solve these issues are image super-resolution and image denoising.

For the seismic image denoising, numerous methods (e.g., Fehmers and Höcker, 2003; Hale, 2009; Liu et al., 2010; Wu and Guo, 2019) have been proposed to attenuate noise and enhance the most dominant structures of reflections and faults. Although these methods can attenuate the noise, blurring the details limits its application in field data. For seismic image super-resolution, high density acquisition (e.g., Zhang et al., 2010; Xiao et al., 2014) and broadband seismic (e.g., Soubaras et al., 2012; Rebert et al., 2012; Wang et al., 2018) are two representative technologies. Nevertheless, both of them are expensive consuming in acquisition and processing.

This paper diverts from the traditional methods and leverages deep convolutional neural network (CNN) to achieve seismic image super-resolution and denoising simultaneously. CNNs have recently achieved success in many computer vision tasks including natural image super-resolution and denoising (e.g., Lim et al., 2017; Zhang et al., 2017; Ulyanov et al., 2018; Dai et al., 2019). And there are numerous methods that have been developed. 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 lot of noise-free field seismic images with high resolution. The second is perception quality. Existing CNN-based methods generally use mean absolute error (ℓ_1) loss, which tends to generate blurry and overly-smoothed results and limits the subsequent seismic interpretation.

To tackle the first issue, we follow the workflow proposed by (Wu et al., 2019, 2020) and generate a lot of synthetic seismic volumes. Subsequently, we extract plenty of 2D inputs from generated 3D volumes to serve as the training data sets. Besides, to tackle the second issue, we replace the ℓ_1 loss by a new objective, a combination of ℓ_1 loss and multi-scale structural similarity (Wang et al., 2003) loss. The network used in our method is a variant of U-net (Ronneberger et al., 2015), which introduces a sub-pixel layer and several residual blocks. The details can be found in the next section. To validate the performance of 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.

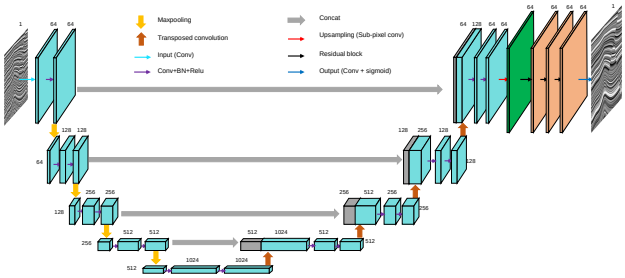

Figure 1: Network architecture used in proposed method.

METHOD

We aim to leverage deep convolutional neural network (CNN) to achieve seismic image super-resolution and denoising simultaneously. The details of used CNN model are as follows.

CNN architecture

The network architecture used in our method is illustrated in Figure 1, which consists of three parts: a stand *U-net*, a *Sub-pixel 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 2×2 and stride 2, two convolution layers with kernel 3×3, and each convolution layer is followed by a batch normalization layer and a rectified linear unit (ReLU). Upsampling block is an opposite design with downsampling block. The goal of introducing *sub-pixel convolution layer* (Shi et al., 2016) is to conduct upsampling. 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 *sub-pixel convolution layer* provides more contextual information through a larger receptive field, which is beneficial for generating more realistic details (Wang et al., 2019). The last part, i.e, the *Residual blocks* learn more high frequency information and details from input to target. In practice, we achieve good performance with just three residual blocks. Finally, we use a convolution layer with kernel 1×1 to reduce the number of feature channels to match the input.

Loss functions

We train our network using a new loss function that combines the ℓ_1 loss and multi-scale structural similarity (MS-SSIM) loss. Due to advantage in improving the performance and convergence over mean squared error (MSE) or ℓ_2 loss (Lim et al., 2017), ℓ_1 loss has been widely used for image super-resolution.

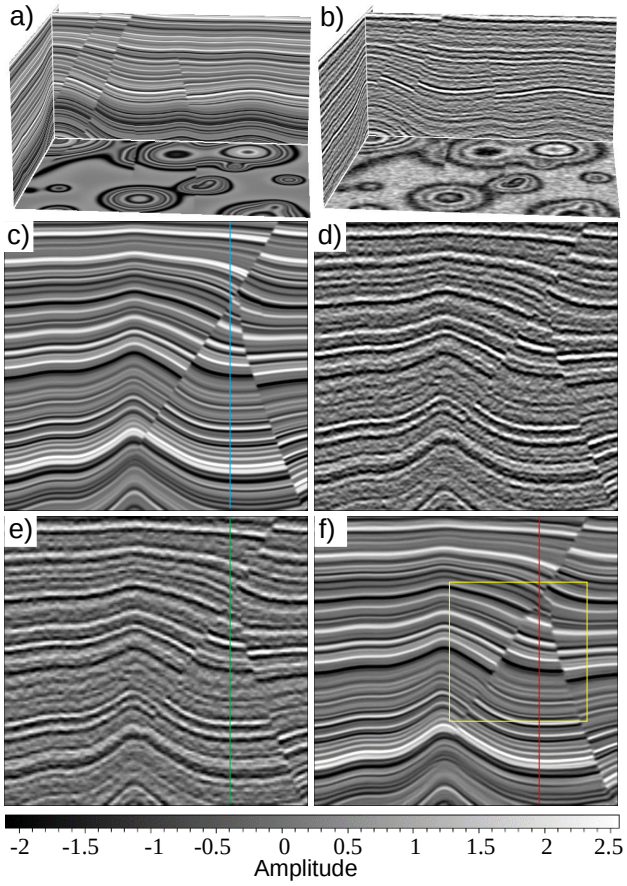

Figure 2: Experimental results on synthetic seismic data. (a) Noiseless 3D synthetic seismic volume; (b) 3D synthetic seismic volume with artificial noise; (c) Noiseless high resolution seismic section extracted from (a); (d) the same section extracted from (b); (e) the input seismic section downsampled from (d); (f) the recovered seismic section using our method.

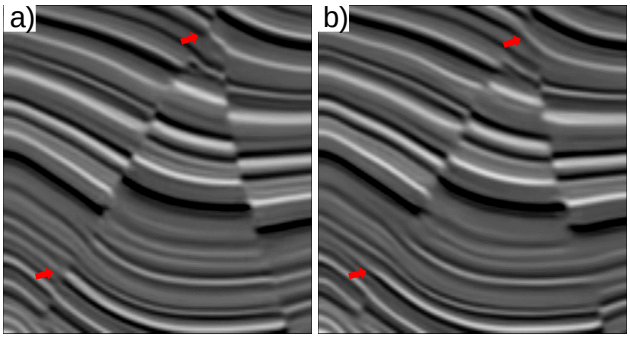

Figure 3: Comparison between the proposed mix loss (a) and traditional ℓ_1 loss (b). The sub-images are extracted from the area of yellow boxes shown in Figure 2(f).

Mathematically, ℓ_1 loss is defined as:

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

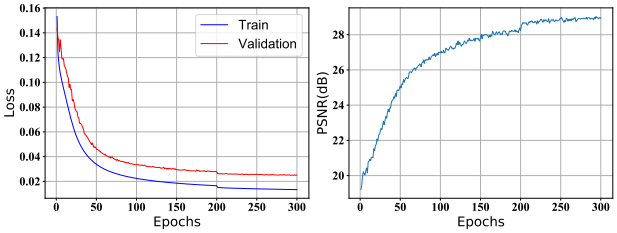

Figure 4: Training record. (a) the loss function values on training and validation data sets; (b) the performance curves of peak signal-to noise ratio (PSNR) on validation data sets.

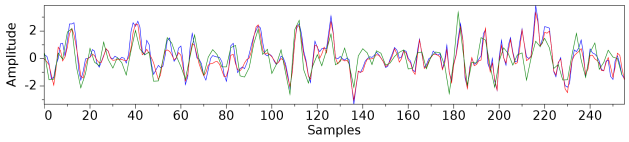

Figure 5: Three traces extracted from the same position where the vertical lines shown in Figure 2. Output section trace (red), corresponding synthetic input section trace (green) and ground truth trace (blue).

where N is total number of pixels. In practice, however, the network trained using only the ℓ_1 loss will generate unsatisfying high-resolution images with smooth textures. The reason is that ℓ_1 loss minimizes only pixel-wise distance between output and target and ignore the texture structures. To tackle this issue, we introduce a new loss term derived from multi-scale structural similarity (MS-SSIM).

MS-SSIM, an assessment for image quality, is sensitive to change in local structure and more appropriate for human visual system (HVS). Mathematically, the loss derived from MS-SSIM is defined as:

$$\text{MS-SSIM}(\mathbf{x},\mathbf{y}) = [l_M(\mathbf{x},\mathbf{y})]^{\alpha_M} \cdot \prod_{j=1}^{M} [c_j(\mathbf{x},\mathbf{y})]^{\beta_j} [s_j(\mathbf{x},\mathbf{y})]^{\gamma_j}, \tag{2}$$

where M is the scale and $l(\mathbf{x},\mathbf{y})$, $c(\mathbf{x},\mathbf{y})$, $s(\mathbf{x},\mathbf{y})$ represent three measurements between $\mathbf{x}$ and $\mathbf{y}$: luminance or amplitude in seismic image, contrast and structure. To improve the perception quality of recovered images, we combine the ℓ_1 loss and MS-SSIM loss and obtain a new loss function defined as:

$$\mathscr{L}^{\text{Mix}} = \alpha \cdot \mathscr{L}^{\text{MS-SSIM}} + (1-\alpha)\cdot \mathscr{L}^{\ell_1} \tag{3}$$

where

$$\mathscr{L}^{\text{MS-SSIM}} = 1 - \text{MS-SSIM}(I_{SR}, I_{HR}), \tag{4}$$

and α is the weight of loss function and we empirically set $\alpha = 0.5$.

The comparison between ℓ_1 loss and mix loss is illustrated in Figure 3. It is obvious that the output of ℓ_1 loss (Figure 3b) is more smooth than the mix loss (Figure 3a) in faults. The mix loss leads to realistic and perception results.

Training data sets

Before training model for super-resolution and denoising together, we need a lot of 2D 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., 2019, 2020) and generate a lot of synthetic 3D seismic data with size $256\times256\times256$. In this workflow, we first obtain a high resolution and noise-free 3D seismic volume with

well-imaged faults and layers (Figure 2a). And then, we add some random noise to simulate realistic and generalized field seismic data (Figure 2b).

To prepare training datasets, we must generate many pairs of 2D seismic images I_{HR} and I_{LR}. Inline or crossline 2D seismic sections are extracted from 3D synthetic noise-free seismic volumes as high resolution images (I_{HR}) (Figure 2c). And the low resolution seismic images (I_{LR}) (Figure 2e), used as input, are obtained by downsampling the same 2D sections by a factor of 2 (Figure 2d) which are generated from the volumes with random noise. We apply 80% of the datasets for training, 10% for validation and rest for test. Our work aims at reconstructing the high resolution images I_{SR} from low resolution images with noise I_{LR}. In particular, I_{SR} is expected to close to the original high resolution images I_{HR}.

Data augmentation

Data augmentation is one of the most useful methods for improving the performance deep models. In order to avoid using a large memory footprint for training, we first crop the pairs of 2D seismic images into a lot of small patches at random. However, the information contained in a small patch is often insufficient to recovery the details between I_{LR} and I_{HR} (Kim et al., 2016). Thus, we choose the size of the input seismic patches to be 96×96 to balance the above problems, and the corresponding size of the high resolution seismic patches is 192×192. In addition, we randomly horizontal flip and vertical flip the pairs of patches to increase the diversity of training datasets.

Training details

We totally extract 1500 pairs of inline or crossline 2D seismic sections from 200 synthetic 3D seismic volumes. Every input seismic image and target seismic image are normalized to $[0,1]$ by substracting and dividing the minimum value and the value's range of each seismic image. We then pre-process all the seismic images by data augmentation discussed before.

We train our model with ADAM optimizer (Kingma and Ba, 2014) and set the parameter $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 10^{-8}$. The learning rate is initialized to $1e-4$ and decayed by 0.5 once the number of epoch reaches 200. We set batch size as 16 and totally extract 16×1000 patches from training datasets. We train our network over 300 epochs. We provide the training details in Figure 4, where (a) reports the loss function values on training and validation data sets; and (b) reports the performance curves of peak signal-to noise ratio (PSNR) on validation data sets. Although the loss function and PSNR curves do not converge until nearly 300 epochs, it only takes us about 7 hours to finish a training task that works in NVIDIA Tesla V100 SXM2.

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 Figure 2, where Figure 2c is the pure seismic images with high resolution used as ground truth, Figures 2e and 2f 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 horizon. 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 horizon of recovered images is more smoothed compared with the inputs.

Furthermore, we compare the amplitude characteristics of three traces and report the results in Figure 5. 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 approximately agreement keeping the shape. But the red one yields more details than the green curve. The ground truth curve maintain the similar characteristics with the output trace. This characteristics is well manifested in the range of samples 200 to 220 of the curve, where many thin layers are covered by random noise in Figure 2e. 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 CNN model and can facilitate to 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 128×128.

Applications

Our CNN model achieves good effectiveness and generalization on both synthetic and field seismic datasets even if it is trained with synthetic seismic data only. To verify the capability of the model, we feed some 2D field seismic images acquired at different 3D surveys to the trained CNN model. Before applying the field seismic images, each image is normalized as same as the synthetic datasets to make it be 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.

Figure 6 shows a real example where the native seismic image (Figure 6a) is directly captured from the paper of BroadSeis (Soubaras et al., 2012). Feeding this native image into our trained CNN model, we obtain an improved image (Figure 6b) where the noise are effectively removed and the resolution of detailed features (e.g., thin layers and small-scale faults) are significantly improved. Our result (Figure 6b) shows even more details than the improved image (Figure 6c) 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 miniseconds to compute the result shown in Figure 6b. Figures 8a, 8b, and 8c show a zoomed in view of the yellow boxes in the native image, our result, and BroadSeis image, respectively. From these sub-images, we can more clearly observe that our CNN-based method (Figure 8b) significantly enhances the detailed structures of thin layers and the faults with small throws (as denoted by the red arrows).

Figure 7 shows another two field examples, where Figures 7a and 7b are the two native field seismic images which are acquired at different surveys. Figures 7c and 7d are the corresponding results computed by using our CNN-based method of simultaneous super-resolution and denoising. 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 datasets, is well generalized for various field datasets. Compared to the native images (Figures 7a and 7b), our re-

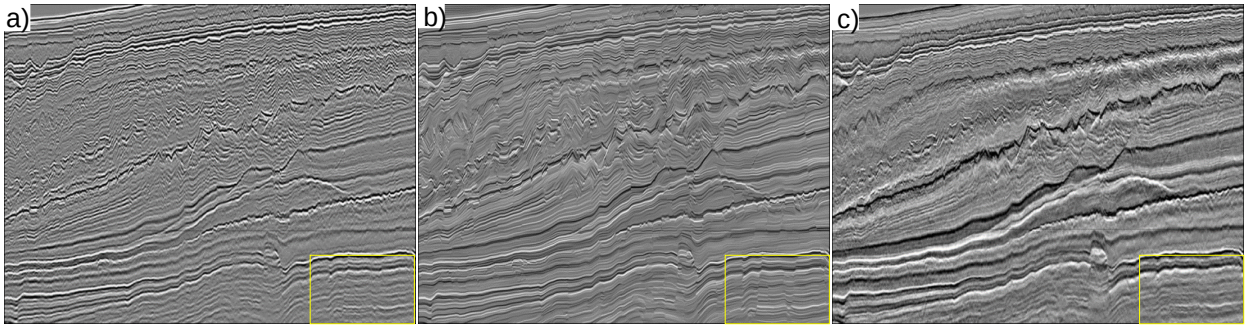

Figure 6: Comparison betweent our method and traditional method. (a) The 2D native field seismic section, (b) the result recovered by our method, (c) the result recovered by traditional method.

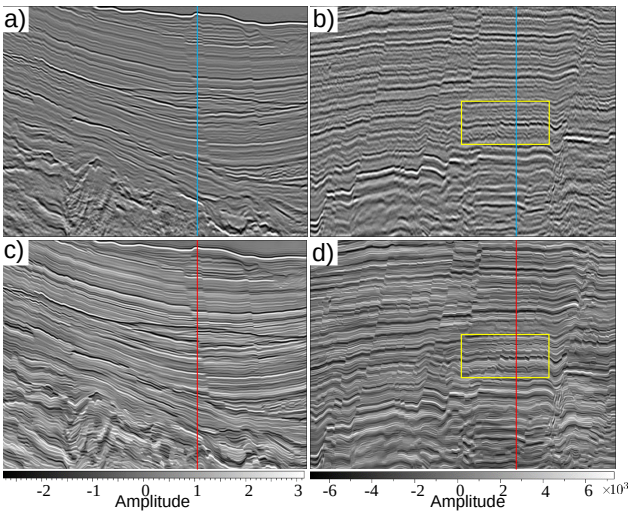

Figure 7: Experimental results of proposed method on field seismic data. (a) and (b) are two field seismic sections, (c) and (d) are corresponding recovered results. In particular, (c) possesses a large amplitude differed with the synthetic seismic sections used for training.

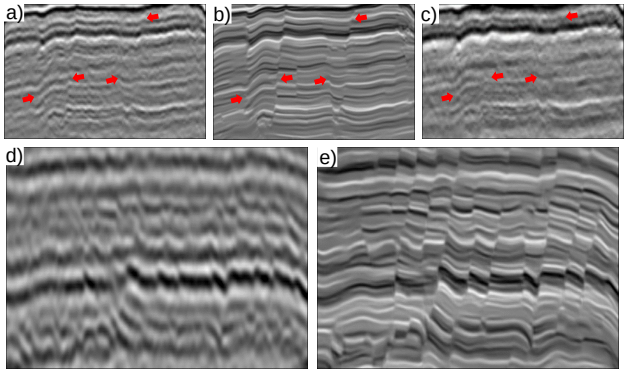

Figure 8: Comparison in details. (a), (b) and (c) are three patches (the yellow boxes) extracted from Figure 6. (d) and (e) are extracted from Figures 7b and 7d (the yellow boxes).

sults 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. Two sub-

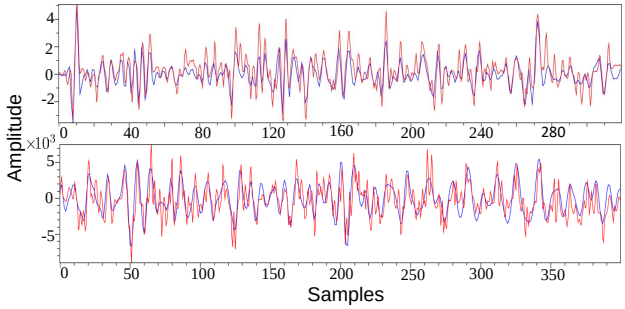

Figure 9: Traces analysis of the field seismic seismic images (blue) and the results (red) shown in Figure 7.

images (Figures 8d and 8e) are extracted from the yellow box area in Figures 7b and 7d to provide a detailed comparison.

Figure 9 shows some seismic traces that are extracted from the native images and our improved images in Figure 7. The top subfigure in Figure 9 shows the traces of first field data shown in Figures 7a and 7c, and the traces in the bottom figure are extracted from the second data (Figures 7b and 7d). 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. 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 Figure 7a, where the results are more likely to contain artifacts.

CONCLUSIONS

In this paper, we developed a novel method for achieving seismic image super-resolution and denoising simultaneously. Our proposed method is based on CNN and 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 in the input seismic images. For future work we want to leverage transfer learning to reduce the gap between synthetic and field seismic data. Such that the performance on field seismic data can be further improved.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation of China under grant no. 41974121.

REFERENCES

Dai, T., J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, 2019, Second-order attention network for single image super-resolution: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 11065–11074.

Fehmers, G. C., and C. F. Hocker, 2003, Fast structural interpretation with structure-oriented filtering: Geophysics, **68**, 1286–1293, doi: https://doi.org/10.1190/1.1598121.

Hale, D., 2009, Structure-oriented smoothing and semblance: CWP Report, 635.

Kim, J., J. Kwon Lee, and K. Mu Lee, 2016, Accurate image super-resolution using very deep convolutional networks: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1646–1654, doi: https://doi.org/10.1109/CVPR.2016.182.

Kingma, D. P., and J. Ba, 2014, Adam: A method for stochastic optimization: arXiv preprint arXiv:1412.6980.

Lim, B., S. Son, H. Kim, S. Nah, and K. Mu Lee, 2017, Enhanced deep residual networks for single image super-resolution: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 136–144.

Liu, Y., S. Fomel, and G. Liu, 2010, Nonlinear structure-enhancing filtering using plane-wave prediction: Geophysical Prospecting, **58**, 415–427, doi: https://doi.org/10.1111/j.1365-2478.2009.00840.x.

Rebert, T., R. Sablon, N. Vidal, P. Charrier, and R. Soubaras, 2012, Improving pre-salt imaging with variable-depth streamer data: 82nd Annual International Meeting, SEG, Expanded Abstracts, 1–5, doi: https://doi.org/10.1190/segam2012-1067.1.

Ronneberger, O., P. Fischer, and T. Brox, 2015, U-net: Convolutional networks for biomedical image segmentation: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 234–241, doi: https://doi.org/10.1007/978-3-319-24574-4_28.

Shi, W., J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, 2016, Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1874–1883.

Soubaras, R., R. Dowle, and R. Sablon, 2012, BroadSeis: enhancing interpretation and inversion with broadband marine seismic: CSEG Recorder, **37**, 40–46.

Ulyanov, D., A. Vedaldi, and V. Lempitsky, 2018, Deep image prior: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 9446–9454.

Wang, Y., J. Wang, X. Wang, W. Sun, and J. Zhang, 2018, Broadband processing key technology research and application on slant streamer: International Geophysical Conference, SEG Global Meeting Abstracts, 135–138, doi: https://doi.org/10.1190/IGC2018-034.

Wang, Z., J. Chen, and S. C. Hoi, 2019, Deep learning for image super-resolution: A survey: arXiv preprint arXiv:1902.06068.

Wang, Z., E. P. Simoncelli, and A. C. Bovik, 2003, Multiscale structural similarity for image quality assessment: 37th Asilomar Conference on Signals, Systems and Computers, 1398–1402, doi: https://doi.org/10.1109/ACSSC.2003.1292216.

Wu, X., Z. Geng, Y. Shi, N. Pham, S. Fomel, and G. Caumon, 2020, Building realistic structure models to train convolutional neural networks for seismic structural interpretation: Geophysics, **85**, no. 4, WA27–WA39, doi: https://doi.org/10.1190/geo2019-0375.1.

Wu, X., and Z. Guo, 2019, Detecting faults and channels while enhancing seismic structural and stratigraphic features: Interpretation, **7**, no. 1, T155–T166, doi: https://doi.org/10.1190/INT-2017-0174.1.

Wu, X., L. Liang, Y. Shi, and S. Fomel, 2019, FaultSeg3D: using synthetic datasets to train an end-to-end convolutional neural network for 3D seismic fault segmentation: Geophysics, **84**, no. 3, IM35–IM45, doi: https://doi.org/10.1190/geo2018-0646.1.

Xiao, F., J. Yang, B. Liang, M. Zhang, R. Li, F. Li, H. Xiao, X. Lei, Q. Liu, and T. Heesom, 2014, High-density 3D point receiver seismic acquisition and processing – a case study from the Sichuan Basin, China: First Break, **32**, 81–90.

Zhang, K., W. Zuo, Y. Chen, D. Meng, and L. Zhang, 2017, Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising: IEEE Transactions on Image Processing, **26**, 3142–3155, doi: https://doi.org/10.1109/TIP.2017.2662206.

Zhang, Y.-G., Y. Wang, and J.-J. Yin, 2010, Single point high density seismic data processing analysis and initial evaluation: Shiyou Diqiu Wuli Kantan (Oil Geophysical Prospecting), **45**, 201–207.