Seismic Data Denoising By Deep Residual Networks

Yuchen Jin, Xuqing Wu, Jiefu Chen, Zhu Han, University of Houston*
Wenyi Hu, Advanced Geophysical Technology, Inc

Summary

Deep learning leverages multi-layer neural networks architecture and demonstrates superb power in many machine learning applications. The deep denoising autoencoder technique extracts better coherent features from the seismic data. The technique allows us to automatically extract low-dimensional features from high dimensional feature space in a non-linear, data-driven, and unsupervised way. A properly trained denoising autoencoder takes a partially corrupted input and recovers the original undistorted input. In this paper, a novel autoencoder built upon the deep residual network is proposed to perform noise attenuation on the seismic data. We evaluate the proposed method with synthetic datasets and the result confirms the effective denoising performance of the proposed approach.

Introduction

Noise attenuation is a very important step for obtaining high-quality data for seismic data post-processing. The sparse representation of seismic data has gained its popularity in recent years. Since natural signals can be compactly expressed, efficient approximations, such as a linear combination of specific atom signals, have been proved to be very useful. Traditionally, most noise suppression methods are applied in a transform domain with fixed-basis functions (e.g. wavelets (Mallat, 2009), curvelets (Starck et al., 2002), and seislet (Fomel et al., 2010), etc. Learning based approaches, on the hand, infer an overcomplete dictionary from a set of examples. The dictionary is typically represented as an explicit matrix and a training procedure is required to adapt the matrix to the examples. It is assumed that the signal can be reconstructed with respect to the dictionary using some sparse linear coefficients.

Dictionary learning can be computationally expensive depending on the learning algorithms. For example, the K-SVD method for designing dictionaries of the sparse representation involves many SVD decompositions for a big dictionary and a large number of training samples. Dictionary learning based noise attenuations have been adopted by the seismic data processing in recent years. Chen et al. (2016) proposed a cascaded approach for learning the dictionary for the seismic noise attenuation. They applied data-driven tight frame construction (Cai et al., 2014) over the seislet transformed noisy data. By performing thresholding on the learned dictionary space, the denoised data could be recovered after the inverse seilet transform. Beckouche et al. evaluated the dictionary learning method in the time-space domain. The data were divided into smaller patches, and a dictionary of patch-size atoms was learned. The method offers a framework to adaptively construct sparse data representation without other transformations. The dictionary can be learned on noisy data or noiseless data. When learning on noisy data, the noise variance needs to be empirically estimated for balancing the data fidelity during denoising. Siahsar et al. (2017) also adopted the data-driven dictionary learning approach. Since meaningful geometric repetitive structures of the seismic data make it intrinsically low-rank in the time-space domain, the sparsity-promoting dictionary learning is reformulated into a non-negative matrix factorization problem. Low-rank property will reduce the noise and the extra L1 norm constraint will archive less correlated atoms in the dictionary matrix.

All aforementioned sparse coding models share a common shallow linear structure. Latest research in deep learning network (LeCun et al., 2015) suggests that non-linear and deep models can achieve superior performance when solving problems in practice. Multi-layer neural networks, such as an autoencoder (Vincent et al., 2010), have been proposed to denoise images. The autoencoder consists of an encoder part and a decoder part. The objective function of the autoencoder can be defined as:

$$\arg\min_{\phi,\psi} \|X - (\psi \circ \phi)X\| ,$$

where $X \in \mathbb{R}^d$, $\phi : X \to Y$, and $\psi : Y \to X$. Given an input x, each hidden layer in the neural network maps it to $z \in \mathbb{R}^k$: $z = \sigma(Wx + b)$, whrere W is a weight matrix, b is the bias vector, and σ is called the activation function such as a sigmoid function or rectified linear unit. While being used as a denoising tool, the autoencoder is trained to recover the original clean data by taking corrupted noisy inputs.

Deep neural network is capable of learning over-complete feature space by using hidden layers having higher dimensional than the input. Sparse representation can then be learned by imposing L1 regularization. In this work, we proposed a novel denoising approach that utilizes the deep residual learning framework. In the subsequent sections, we will first introduce the residual learning framework. Then we will overview the network structure. Following the implementation details, the performance of the proposed method will be evaluated on a set of syntheic datasets.

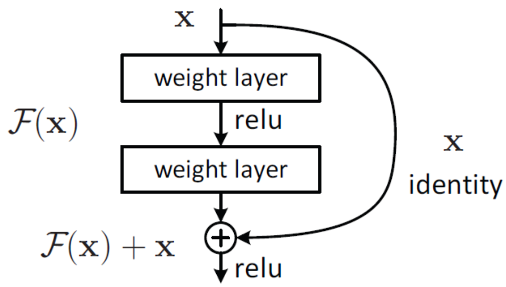


Figure 1. A building block of the DRN (He et al., 2016).

Methodology

The deeper the network is the harder to train it. Problems such as overfitting, vanishing/exploding gradient, and increasing training errors will affect the network to converge. He et al. (2016) addressed these degradation problems and introduced a deep residual network (DRN) structure. Unlike the traditional deep neural network, where a desired underlying mapping is fit by stacked layers, DRN explicitly make those layers fit a residual mapping. For example, as Fig. 1 shows, a new nonlinear mapping $\mathcal{F}(x) := \mathcal{H}(x) - x$ is used during the fitting. DRN hypothesizes that it would be easier to make the residual to zero than fitting an identify mapping by a stack of nonlinear layers. As shown in Fig. 1, the original mapping $\mathcal{H}(x)$ is recast into $\mathcal{F}(x) + x$, which is realized by a feedforward shortcut connection that simply performs identity mapping.

In this paper, we have propose two denoising networks, one with 13 layers and the other with 100 layers. Both of them are residual networks. Fig. 2 and Fig. 3 show the structure of the 13-layer and 100-layer networks. The input size of both networks can be adaptive. Fig. 4 shows the detailed structure for each residual block (rblock). The network is trained with image patches of the size (84x84). The batch size used during the training stage is 128. To train the DRN, we add different levels of noise to the synthetic data. Then the data is randomly cropped into patches and these patches were used as the training sample to learn the parameters of the network. Once the network is welled trained, we can perform denoising on the data regardless its size since the denoising operation is applied to each image patch. Adjacent image patches are overlapping to each other and we find that overlapping will improve denoising performance.

However, recovering original image from overlapping image patches is not a trivial task and it requires extra storage space to index all overlapping areas. In this paper, we design a 2D Kaiser window based masking operator, which has $O(1)$ time complexity. The 2D Kaiser window is an extension of the 1D Kaiser-Bessel window, a one-parameter

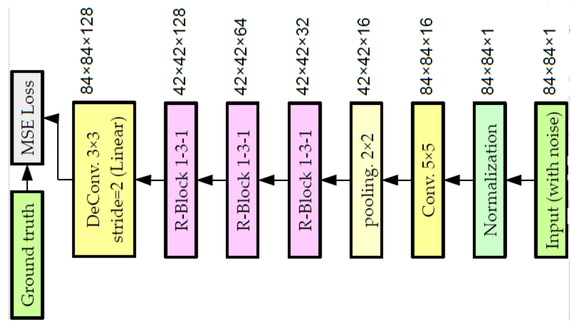


Figure 2. A 13-layer network for seismic data denoising.

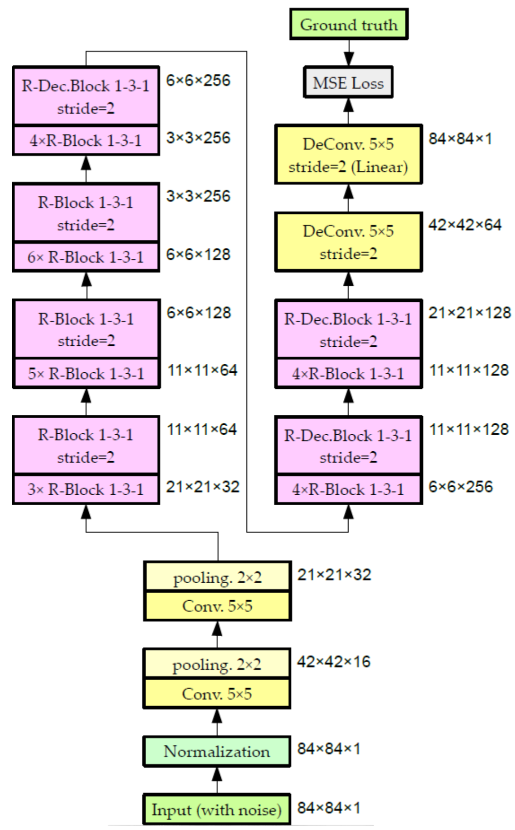


Figure 3. A 100-layer network for seismic data denoising.

family of window functions (Fig. 5). We use this window as a filter to scan the original image along both axises and extract patches. In the meantime, we generate a single-value image the same size as the original image. Then the same filtering process is applied to the single-value image too and the filtered image is used as a "mask". To recover the original image from image patches, we first place patches back to their locations, sum values in overlapped areas, and

10.1190/segam2018-2998619.1

then use the mask as the "divisor" and perform a pixel-wise division. The workflow is shown in Fig. 5.

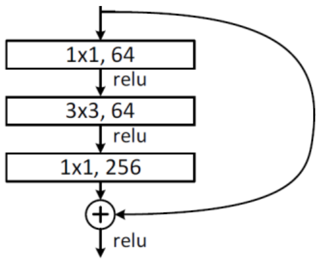


Figure 4. A residual building block (rblock) as in Fig. 2.

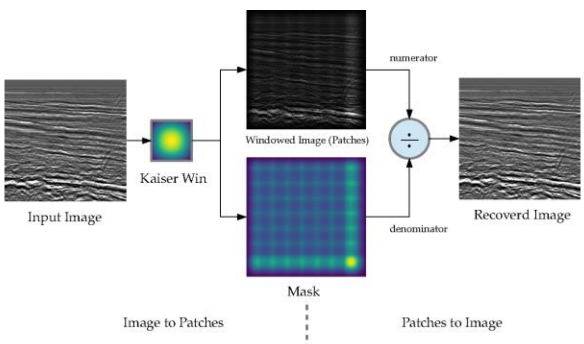


Figure 5. Recover the image from patches by using the Kaiser window based mask.

Experiments

In this section, we use the DRN to attenuate the noise in seismic data. We conduct a comprehensive studies with regard to the selection of the network structure and parameters. We use synthetic examples to evaluate the denoising performance of the proposed deep residual framework. Gaussian white noise is added to simulate noisy data by assuming that the seismic noise is caused by a diversity of spatially distributed, uncorrelated, different but low-frequency sources.

Compared to the 40,000 steps required for training the 100-layer network, the 13-layer network will converge in 20,000 steps. However, the 100-layer network has smaller squared loss among the two. Fig. 6 shows the comparison of training losses of these two networks. Since we perform patch-wise denoising, different stride sizes will affect the size of the overlapped areas among adjacent patches. If $stride-size/batch-size$ equals to 1, then there would be no overlappig. The less the ratio is, the larger the overlapping area will be. In order to evaluate the denoising performance, we use signal-to-noise ratio (SNR) as a quantitative evaluation metric to justify the reconstruction result. The SNR is calculated as:

$$SNR = 10\log_{10}\left(\frac{\|S_{clean}\|_F^2}{\|S_{clean} - S_{denoised}\|_F^2}\right).$$

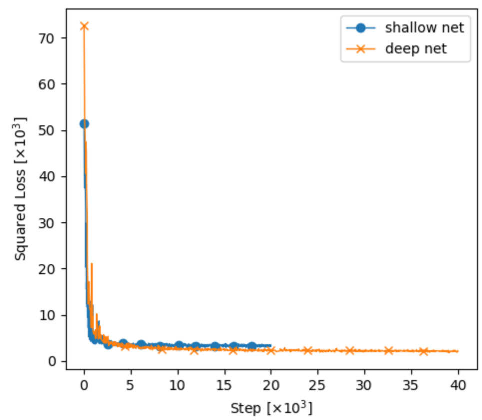


Figure 6. Training loss: 13-layer vs. 100-layer network

Fig. 7 and Fig. 8 show the denoising performance by using different network depths, stride and batrch-size ratios, and Kaiser window shapes. The shape of the Kaiser window is controlled by the parameter α. The main-lobe width of the Kaiser window becomes larger when α increases.

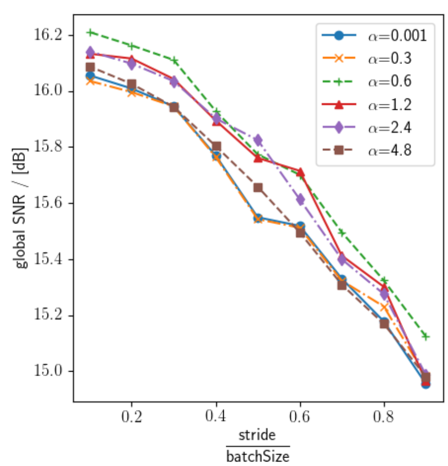


Figure 6. SNR value for the 100-layer network

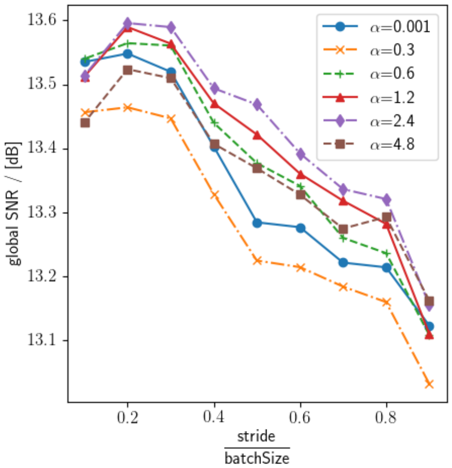


Figure 7. SNR value for the 13-layer network

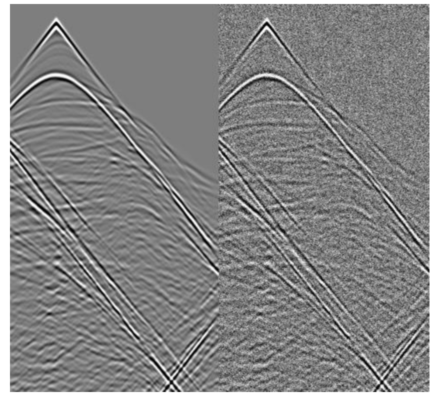
It is obviously that large overlapping is preferred. The shape of the Kaiser window does play a role in tuning the denoising performance for the network. Our test also shows that the 100-layer network delivers better denoising results than the 13-layer network. As shown in Fig. 8, we add Gaussian white noise to the clean image. Then we apply both 13-layer and 100-layer deep residual network for denoising. Examine the denoised results in Fig. 8(b) and Fig. 8(c), we find that the deep residual network is an ideal denoising tool for removing random noises while keeping coherent details.

Conclusions

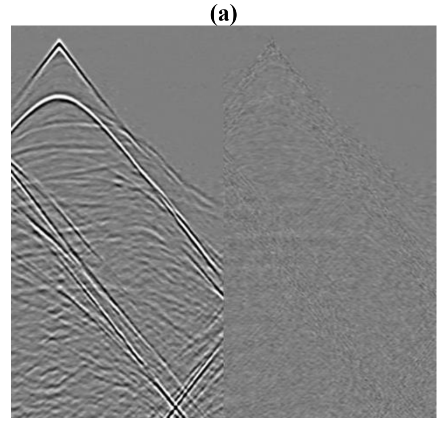
In this paper, we have presented a high performance denoising technology for seismic data sets based on the deep residual networks. One of the advantages of using this framework for denoising is that it does not require any manual parameter tuning once the model is sufficiently trained in prior. Our experiments indicate that the denoising results can be improved by selecting proper network structures. Learning via patches, the proposed Kaiser window filtering approach could significantly improve the denoising performance with regard to the SNR value and computational cost.

Acknowledgment

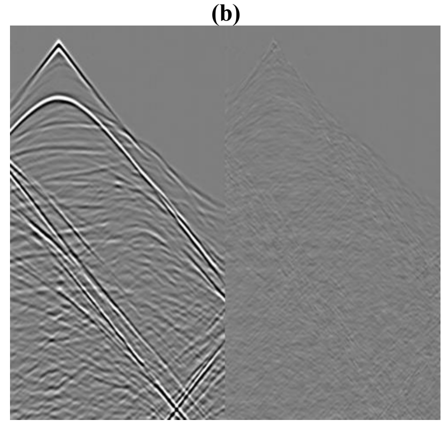
This material is based upon work supported by National Science Foundation under Grant No. 1746824.



(a)



(b)



(c)

Figure 8. Denoising using 13-layer and 100-layer residual networks. (a) left panel: a clean image without noise, right panel: the image with added white noise; (b) left panel: denoised image by the 13-layer network, right panel: differences between the clean image and the denoised image on the left. (c) left panel: denoised image by the 100-layer network, right panel: differences between the clean image and the denoised image on the left.

10.1190/segam2018-2998619.1

REFERENCES

Beckouche, S., and J. Ma, 2014, Simultaneous dictionary learning and denoising for seismic data: Geophysics, **79**, no. 3, A27–A31, https://doi.org/10.1190/geo2013-0382.1.

Cai, J.-F., H. Ji, Z. Shen, and G.-B. Ye, 2014, Data-driven tight frame construction and image denoising: Applied and Computational Harmonic Analysis, **37**, 89–105, https://doi.org/10.1016/j.acha.2013.10.001.

Chen, Y., J. Ma, and S. Fomel, 2016, Double-sparsity dictionary for seismic noise attenuation: Geophysics, **81**, no. 2, V103–V116, https://doi.org/10.1190/geo2014-0525.1.

Fomel, S., and Y. Liu, 2010, Seislet transform and seislet frame: Geophysics, **75**, no. 3, V25–V38, https://doi.org/10.1190/1.3380591.

He, K., X. Zhang, S. Ren, and J. Sun, 2016, Deep residual learning for image recognition: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770–778, https://doi.org/10.1109/CVPR.2016.90.

LeCun, Y., Y. Bengio, and G. Hinton, 2015, Deep learning: Nature, **521**, 436, https://doi.org/10.1038/nature14539.

Mallat, S., 1999, A wavelet tour of signal processing: Academic Press.

Siahsar, M. A. N., S. Gholtashi, V. Abolghasemi, and Y. Chen, 2017, Simultaneous denoising and interpolation of 2D seismic data using data-driven non-negative dictionary learning: Signal Processing, **141**, 309–321, https://doi.org/10.1016/j.sigpro.2017.06.017.

Starck, J.-L., E. J. Candes, and D. L. Donoho, 2002, The curvelet transform for image denoising: IEEE Transactions on Image Processing, **11**, 670–684, https://doi.org/10.1109/TIP.2002.1014998.

Vincent, P., H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, 2010, Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion: Journal of Machine Learning Research, **11**, 3371–3408.