CONVOLUTIONAL AUTOENCODER NEURAL NETWORK FOR SEISMIC NOISE REDUCTION

DARAPUREDDY HARITHA and NITTALA SATYAVANI*

CSIR - National Geophysical Research Institute, Hyderabad, India.
satyavani.nittala@gmail.com

(Received June 28, 2021; revised version accepted February 14, 2022)

ABSTRACT

Haritha, D. and Satyavani, N., 2022. Convolutional autoencoder neural network for seismic noise reduction. *Journal of Seismic Exploration*, 31: 267-278.

Seismic noise reduction is one of the main steps in the data processing sequence that aids in proper seismic imaging and interpretation. For the purpose of noise reduction and to recover weaker / masked signals, we propose the scheme of an unsupervised convolutional autoencoder neural network. Cross-entropy is used as the loss function in the network. The adaptive moment estimation plays the role of a backpropagation algorithm that can optimize the loss function. The key parameters of the network, like convolutional layers, filter size, and learning rate have been selected after performing a series of tests with different values for each of the parameters and those results are also presented here. We show that the present network applied to the seismic data shows improvement in the reflections and also allows us to recover some of the weaker / masked signals. The results show that the proposed method is successful in suppressing the noise and enhancing the seismic signals.

KEY WORDS: seismic data, noise reduction, convolutional layers, filter size, convolutional autoencoder.

INTRODUCTION

Noise suppression generally leads to an increase in signal to noise ratio, thereby, improving the quality of the seismic image. The presence of an unwanted signal (noise) in pre-stack data affects the amplitude information, thereby causing problems in seismic processes like amplitude versus offset (AVO) studies (Li and Mallick, 2014), reverse time migration (RTM), full-waveform inversion (FWI) studies (Pratt, 1999) etc. The seismic noise also affects post-stack data, potentially leading to erroneous interpretation. In the last few decades, many conventional methods have been developed in processing techniques for noise reduction. Some of the

0963-0651/22/$5.00

methods were wavelet transform (Wu and Liu, 2008), f - x filtering (Harris and White, 1997), singular vector decomposition (SVD) (Bekara and van der Baan, 2007). These methods make use of spatial coherence of the data and improve SNR, thereby increasing the horizontal resolution, however, in the process, they reduce the tilt and the bending events. On the other hand, the median filters (Bednar, 1983; Duncan and Beresford, 1995) are helpful especially in eliminating peak noise in nonstationary signals. Of these two methods, the median filter is better in denoising and in attaining signal consistency. Both the methods are mostly based on empirical knowledge and therefore the selection of denoising parameter to suppress the noise plays a crucial role. If the parameters are not selected diligently, there might be a loss in signal content and/ or a regain of noise in the data (Wang et al., 2017; Huang et al., 2017a; Zhao et al., 2019). As such, these methods addressing the issue of noise reduction still suffer from problems such as erroneous assumptions and inappropriate parameter settings. One possible solution for the above problem is Deep Learning (DL) because it is data driven and it does not require any prior knowledge.

Machine Learning (ML) can automatically build a model based on training data and can also improve the model based on training. DL methods build the model based on an artificial neural network, which helps to extract the high-level features from the data. DL is a subset of ML. ML is designed in such a way that it learns the features from the data, and it is mostly used in regression, prediction, and classification problems like facial recognition (Rowley et al., 1998), medical diagnosis (Kononenko, 2001), and in seismic exploration and interpolation. Artificial neural networks (ANN) are mostly used in seismic data processing techniques like event picking (Glinsky et al., 1996), tomography (Nath et al., 1999) and in seismic interpretation facies classification (Ross and Cole, 2017) and fault identification (Huang et al., 2017).

DL has gained incredible attention in solving geophysical problems like noise suppression in seismic images using the convolutional neural networks (Jain and Seung, 2009); Multilayer perceptron method to reduce the seismic noise (Burger et al., 2012); trainable nonlinear reaction-diffusion method to remove the Gaussian noise (Chen and Pock, 2017), etc. In addition, for general image noise suppression, denoising convolutional neural networks are proposed by Zhang et al., 2017. The convolutional neural network (CNN) is used to extract the important features from the images/ 2D data using the convolutional layers. CNN is gaining popularity in image processing, segmentation, and classification. The DL methods have achieved good results in terms of reduction in the Gaussian noise. If the noise is other than Gaussian, then the network performance is poor.

Seismic Denoising using DL can be carried out in two ways: 1. by supervised learning, and 2. by unsupervised learning. The supervised learning-based methods need large sets of clean data for labelling, which might not be possible always. On the other hand, the unsupervised learning method does not need clean datasets for labelling and can remove noise from corrupted seismic data. Supervised learning-based noise reduction

(Zhang et al., 2017; Chen et al., 2017) and unsupervised learning-based noise reduction (Gondara et al., 2016; Du et al., 2016; Chen et al., 2019; Song et al., 2020) both provide good results. In a nutshell, it may be said that both, Supervised and Unsupervised DL methods are useful in noise suppression problems, depending on the kind of datasets available to the user.

In the present work, we have used an unsupervised algorithm because of the lack of training samples and also because it can automatically suppress the noise and reconstruct the signals from noisy seismic data. Convolutional autoencoder keeps the spatial information of the input image and extracts the information gently with convolutional layers without any loss of data. The autoencoder uses CNN to reproduce the input in the output layer. The convolutional autoencoder neural network is also used to attenuate the noise from seismic data. The proposed convolutional autoencoder method shows promising results in denoising of the seismic data. The seismic data from the Andaman Offshore region has been used in the present work. The seismic stack data used in this study is obtained from a complex geologic region and conventional processing could not bring out some of the subtle features (Satyavani et al., 2008). In the present study, we demonstrate that the present methodology can be used effectively to improve the seismic reflections and more importantly, retrieve some of the weak/masked signals.

METHODOLOGY AND COMPONENTS OF THE PROPOSED NETWORK

Seismic data is a combination of signal and noise, and the recorded data can be mathematically expressed as below:

$$S = x + N_o \quad , \tag{1}$$

where x is the clean data and N_o is the noise component. We aim to reconstruct the clean data (x) from the contaminated seismic data (S). So, we design the model to map F: $(x + N_o) \rightarrow x$.

Before constructing the model, we design the pre-processing steps for training and testing the datasets. We first normalize the contaminated seismic data using max-min normalization. We normalise the data to bring the different amplitude values into one scale because it helps the network to learn data quickly and converge faster. This normalised dataset adapts to the neural networks via:

$$S^{\star}=(S - min)/(max - min) \quad , \tag{2}$$

where $S^{\star}$ is the normalized data, max and min are maximum and minimum amplitude values of the seismic dataset.

Autoencoder

Autoencoder is a type of unsupervised artificial neural network that contains an encoder and decoder. Basically, in autoencoder, the input data is compressed into a lower dimension to extract the important features and then it is expanded to its original data size using a decoder.

The unlabelled input data will pass into the encoder, and it compresses the data into latent space representation expressed as follows:

$$y = \sigma(W_1 S^{\star} + b_1) \quad , \tag{3}$$

where σ is the non-linear activation function, y is the lower-dimensional data after encoding.

The compressed data (y) acts as input for the decoder which reconstructs the compressed data to the original input data dimension by inverse mapping. The reconstructed data (Z) expressed as follows:

$$Z = \sigma(W_2 y + b_2) \quad . \tag{4}$$

W_1 , W_2 *and* b_1, b_2 are the weights and bias in the network, respectively. W_1, b_1 are the weights and bias between the input and hidden layers, whereas W_2 and b_2 are in between the hidden and out layers. σ is defined as the activation function (Fig. 1).

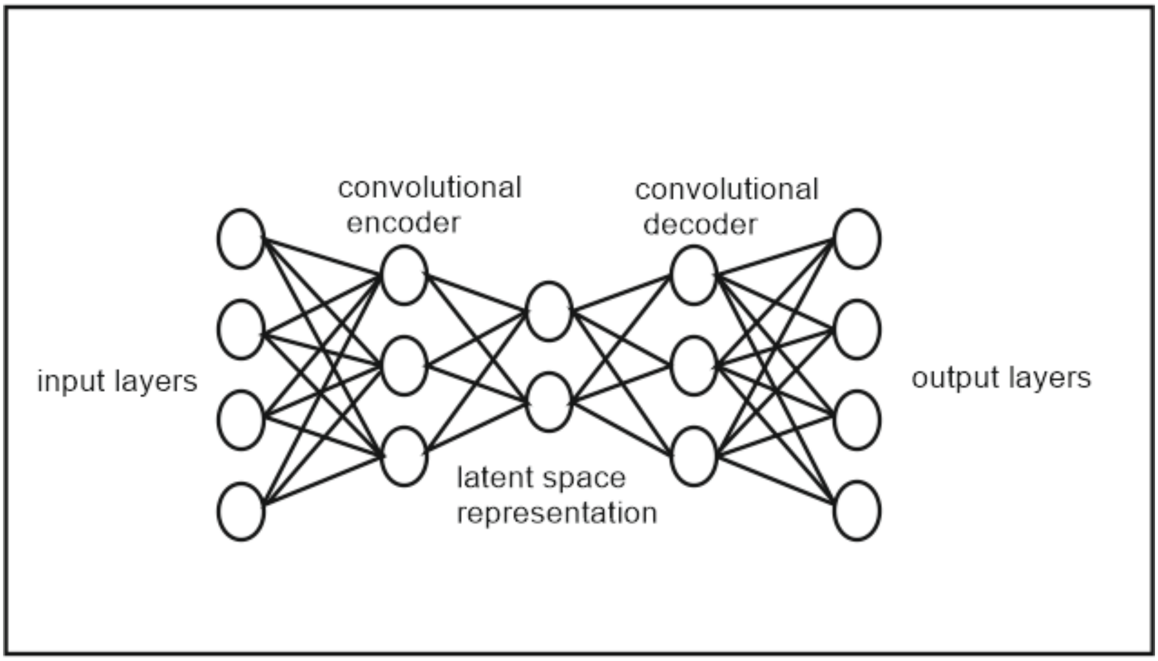

Fig.1. General architecture of convolutional autoencoder neural network.

Convolutional autoencoder neural network

Conventional autoencoder (AE) ignores the structural image information because the data in the network is in vector form. This can introduce parameter redundancy and forces the network to learn the global structure. So, it becomes difficult for AE to extract the hidden features of data (Zhang et al., 2019). Therefore, we have used a convolutional neural network, i.e., CNN with autoencoder (Fig. 1) which helps to extract the hidden features and reduce the noise in seismic data.

In general, the latent representation of the n-th feature map with convolutional layers can be represented as

$$y_n = \begin{cases} \sigma(W_n \otimes S^{\star}{}_n + b_n), & n = 1 \\ \sigma(W_n \otimes y_{n-1} + b_n), & n \in (2, \dots m) \end{cases} \tag{5}$$

where $\otimes$ denotes convolutional operator, m is the number of convolutional layers.

In general, the activation function is used to help the network to learn the complicated patterns in data and also it can increase the non-linearity into the network. Different types of activation functions are used in DL. We have considered a rectified linear unit (ReLU), $\sigma(x) = \max(0, x)$ as the activation function in hidden layers because it can solve vanishing gradient problem and also allows the model to learn faster (Krizhevsky et al., 2017). The Sigmoid function, $\sigma(x) = 1/(1 + e^{-x})$ acts as the activation functions in the output layer in our work.

Max-pooling layers are introduced in CNN to reduce the dimensionality of the data. The main advantage of using max-pooling is to speed up the calculation and prevent overfitting. In the present work, max-pooling layers are added after each convolutional layer in the encoder part. The output generated from the max-pooling, producing the smallest dimension of the whole network in latent space, acts as input to the decoder part in the convolutional autoencoder. Using upsampling (inverse of max-pooling) helps to reconstruct the compressed data into its original input size.

The final aim in neural networks is to decrease cost function/loss function/error function. The loss function is used to measure the performance of the network. During the optimization process, the model error is calculated by choosing a suitable loss function for the network. We have used cross-entropy (Kline et al., 2005) as a loss function in our network. To reduce the loss function, we have used the adaptive moment estimation (adam) (Kingma and Ba, 2014) as a backpropagation algorithm in our work. After obtaining the denoised image, we restore the amplitude back to the original by performing denormalization.

SEISMIC DATA AND THE NETWORK ARCHITECTURE

The marine seismic stack data from the Andaman offshore region (Fig.2) is used in our study and the dataset consists of CDP 380-1080 within the time of 1808-3400 ms, i.e, 700 traces and 796 time samples. Some of the events are blurred because of the noise in seismic data that could not be reduced by conventional processing. In the present study, we make an attempt to apply the CNN autoencoder network to this data to reduce the noise. Before training the dataset, we have normalized the seismic data using max-min normalization. Generally, the choice of CNN parameters dictates the quality of the seismic image and therefore, selection of the parameters is critical to CNN methods. In our present study, selection of the parameters was done using a trial-and-error approach and we notice that the hyperparameters effecting the performance of our network are the number of convolutional layers and associated neurons, filter size and learning rate. For better noise suppression results, we have tested the each of these parameters sequentially using different sets of values and the results are summarised as follows:

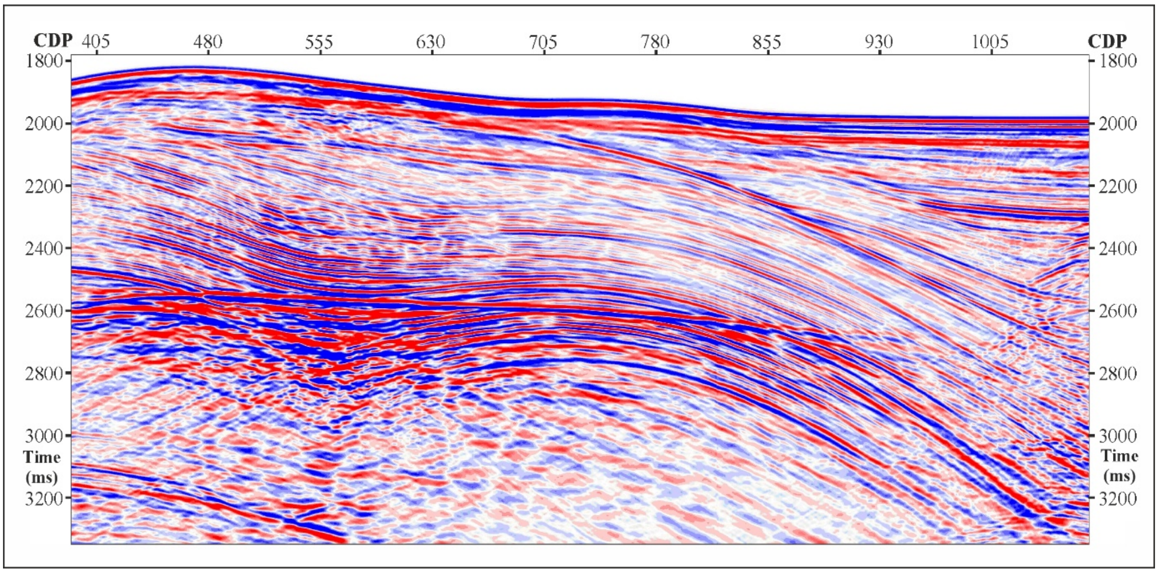

Fig. 2. Seismic stack image obtained from Andaman Offshore from conventional seismic data processing (modified after Satyavani et al., 2008).

Convolutional layers

We first tested the number of convolutional layers that can provide improved images. Initially, we considered five convolutional layers with 8,16,16,8,1 neurons in the respective layers for the network. This configuration leads to data loss (indicated with an arrow) as shown in Fig.3(b). We then increased the number of layers from five to seven with 8,16,16,16,16,8,1 neurons and we observe that the noise is reduced (indicated as ovals) as shown in Fig. 3(c). We further changed the number of convolutional layers to nine with 8,16,16,16,16,16,16,8,1 neurons and we observe that the data is attenuated, as shown by ovals in Fig. 3(d). Based on the above test results, we selected seven convolutional layers as the most promising for our network. Keeping the number of convolutional layers fixed at 7, we have sequentially changed the number of neurons within the convolutional layers by a trial-and-error method, from 8,16,16,16,16,8,1 neurons to 64,16,16,16,16,64,1.

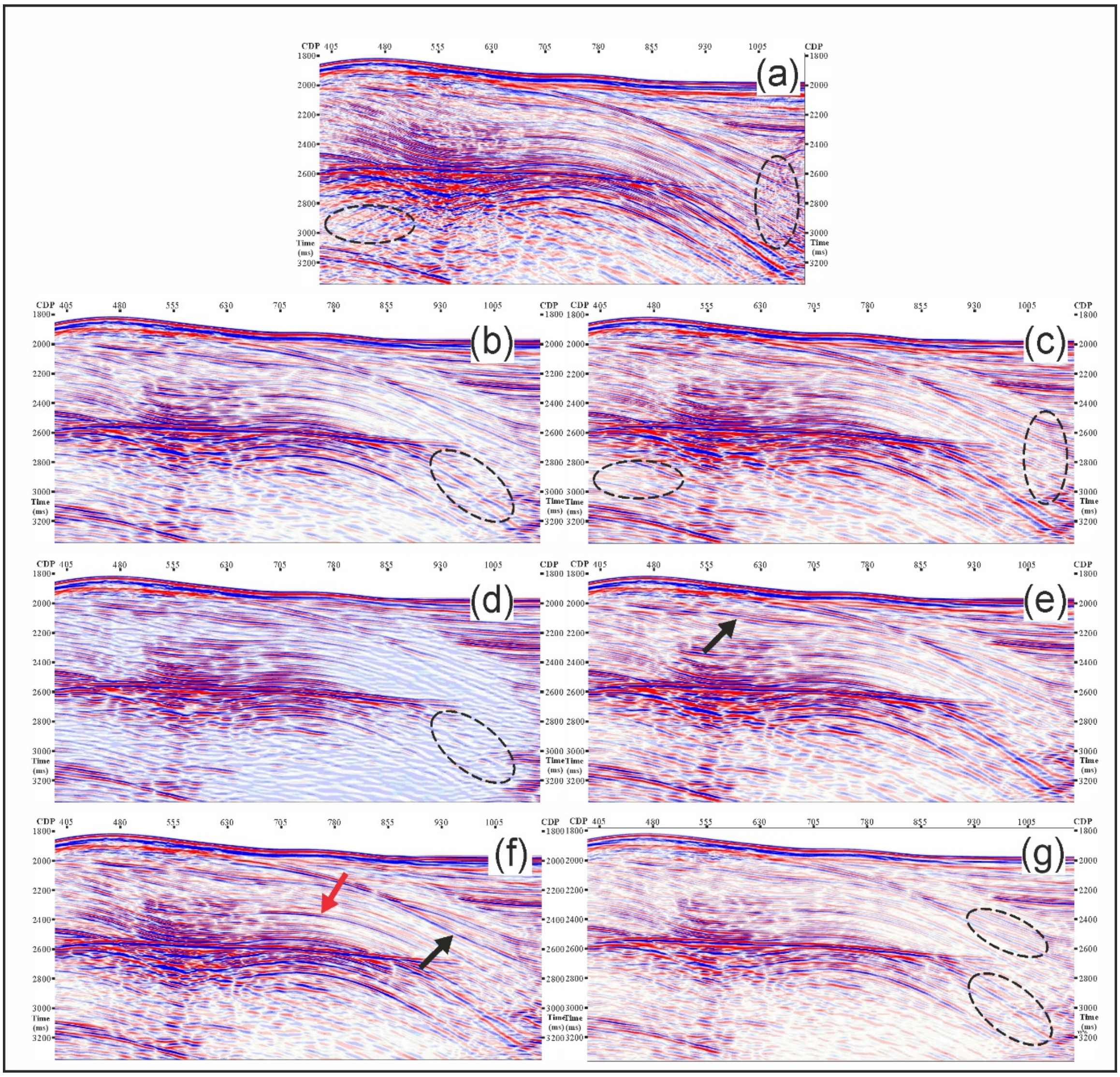

Fig. 3. (a) Original Seismic image. (b) The seismic image obtained after using five convolutional layers with 8,16,16,16,16,8,1 neurons in the respective layers. The dotted ovals indicate that the signal loss in output data. (c) The seismic image obtained after increasing the number of convolutional layers to seven with 8,16,16,16,16,8,1 neurons in respective layers. The dotted ovals indicate that the data is denoised to some extent. (d) The seismic image obtained after increasing the layers to nine with 8,16,16,16,16,16,16,8,1 neurons. The dotted ovals indicate the signal attenuation in denoised data.(e) The seismic image obtained after using seven convolutional layers with 32,16,16,16,16,32,1 neurons in the respective layers. The arrow indicates the continuity of the event is improved in output data. (f) The seismic image obtained after changing the configuration with 48,16,16,16,16,48,1 neurons in the respective layers. Weak signals recovery is indicated with red arrow and increase in the continuity of the events is marked with a black arrow. (g) The seismic image obtained after using seven convolutional layers with 64,16,16,16,16,64,1 neurons in the respective layers. The dotted ovals indicate that the signal loss in output.

We notice that when the neuron configuration was kept at 32,16,16,16,16,32,1 the continuity of event is improved as indicated with an arrow in Fig. 3(e) compared to the original data [Fig. 3(a)]. We have also observed the signal [Fig. 3(f)] which was earlier masked in noise and also an increase in the continuity of the events when the neuron configuration is kept at 48,16,16,16,16,48,1 in the convolutional layers. When the number of neurons was 64,16,16,16,16,16,1 we observed that there is a data loss, indicated with dotted ovals in predicted data as seen in Fig. 3(g).

Therefore, 48,16,16,16,16,48,1 neurons in convolutional layers are selected for our network as this configuration gives the best noise suppression result. We can see clear events in the denoised section shown in Fig. 3(f). After fixing the convolutional layer values, we tested the filter size for our network.

Filter size

The size of the convolutional filter also influences the denoising network. We have tested different filter sizes like 2*2, 3*3, 4*4 in our network for noise suppression. We have fixed the number of convolutional layers to seven with 48,16,16,16,16,48,1 neurons and then we have tested the filter sizes to achieve the best-denoised result. By employing the 2*2 filter, most of the signals are attenuated along with the noise data and the resolution also is reduced (Fig. 4b). The 3*3 filter size achieves significant noise attenuation and brings out the subtle features in the seismic data [Fig.4(c)]. When we used a 4*4 filter size, the noise in raw data [Fig. 4(a)] is not suppressed [Fig. 4(d)] and hence it is discarded for our network. Therefore, we have selected a 3*3 filter as the filter size best suited for our network.

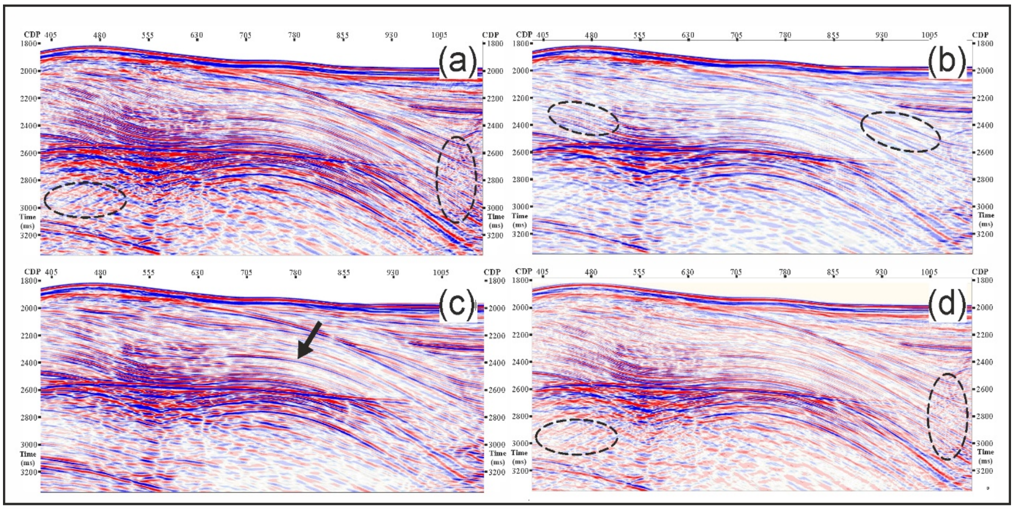

Fig. 4. (a) Original Seismic image. (b) The seismic image obtained when using 2*2 filter size to our network. The dotted ovals indicate the events attenuation. (c) The seismic image obtained after using a 3*3 filter size. The arrow indicates the subtle features in the seismic data. (d) The seismic image obtained after increasing the filter size to 4*4. The dotted ovals indicate noisy events are not suppressed in output data and some of the signals are attenuated.

Learning rate

Learning rate (lr) is the hyperparameter that influences the network performance. The network cannot perform better if we select the learning rate poorly. We tested the performance of our network with different learning rates and observed the noise reduction patterns. When we tested our data with lr = 0.0001 and epoch 50, then the processing time for noise suppression is nearly 190sec and we notice some data loss. We changed the lr value to 0.009 with epoch 50, then the data loss is reduced. Further, we tested data with lr = 0.006 and epoch 50 then the network took 165 sec for processing and gave a slightly good result compared to the above result. We then changed the number of epochs from 50 to 55 with lr = 0.006, and we achieved a good result. After increasing the epochs from 55 to 60, we have not found any changes in data, and the output looks similar to the one achieved with 55 epochs. Therefore, we fixed epoch values for the network as 55. The performance of the network at higher speed and model stability is achieved at lr of 0.006. Based on the performance of denoising, we have selected the lr as 0.006 for our network.

Role of parameters

We have selected the specific parameters to our data based on the trial-and-error method. In our analysis, we have found out that the most important parameters which influence noise reduction are convolutional layers and filter size. When we decrease the number of convolutional layers and filter size, the signal is attenuated in the dataset. When we increase the convolutional layers and filter size, noise suppression doesn’t occur. However, the same is not to with the learning rate. We have observed that the learning rate does not bring much change in the noise reduction pattern compared to other parameters in the network. The parameters discussed in the above sections are applicable to the present dataset. However, we wish to mention that the parameters guiding the neural network depend upon the size and complexity of the dataset. Therefore, the parameters discussed in this section are unique to this particular dataset.

RESULTS

We tested our method on seismic data from the Andaman offshore region (Satyavani et al., 2008) within the range of CDP380 to CDP1080 and time from 1808 to 3400 ms with different convolutional layers, filter sizes and learning rates. Based on the above test results, we have chosen a configuration of seven convolutional layers with a neuron distribution of 48,16,16,16,16,48,1 in each layer. Max-pooling and Upsampling have been used after every convolutional layer to reduce and enlarge the dimensions of the data in the encoder and decoder part, respectively. The filter size of 3*3 is selected for the present network as it shows improvement in noise reduction. The learning rate of our network is 0.006. After 55 epochs, the learned convolutional autoencoder neural network was able to regain the test dataset dimensions successfully. Finally, noise attenuated seismic data (output from the network) is restored to its original dimension by de-normalization. The output from the network shows significant noise reduction and recovery of some of the events. The initial seismic data is compared to the improved seismic data as shown in Fig. 5a and Fig. 5b respectively. In the denoised seismic data, the weak amplitude signals are recovered (indicated by red arrow) which were not seen in initial stack data because of the presence of noise (indicated by ovals in Fig. 5b). The seismic reflections are also showing increased continuity (indicated by black arrows) and amplitude enhancement compared to the original seismic data. We have also noticed that the background noise is significantly reduced in Fig. 5(b). Our results show that the proposed method achieves a good result in noise suppression, and we infer that this methodology with appropriate parameters can be applied for noise reduction of seismic data.

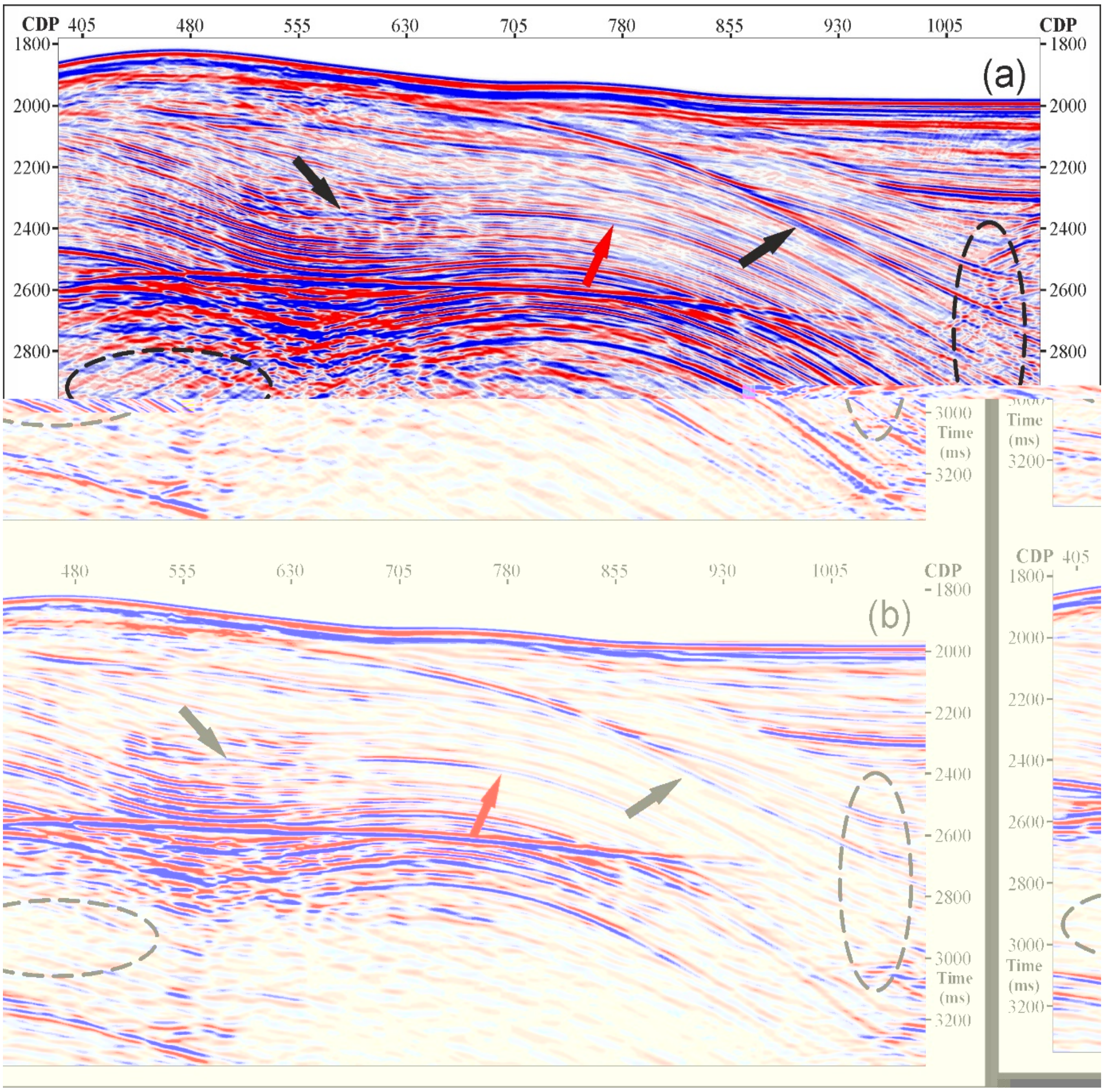

Fig. 5. (a) Original Seismic image. (b) Seismic image obtained from the CNN network. The reduction in noise is indicated by dashed ovals. The weak amplitude signals that are recovered are indicated by the red colour arrow. The reflectors indicated by the black arrows show more continuity and amplitudes are enhanced than the original seismic image.

CONCLUSION

In the present work, we have proposed a convolution autoencoder neural network and also applied it to the seismic data from the Andaman Offshore region to reduce the seismic noise and improve the seismic image. The noise is attenuated using convolutional layers, max-pooling, and up-sampling in the network. The hyperparameters like convolutional layers, filter size, and learning rate affects the noise reduction performance in our work. The obtained result demonstrates that the network improves seismic data quality considerably. The denoised result shows that our proposed method not only attenuates the noise in the seismic data but also recovers the blurred events from the contaminated data. The hyperparameters discussed in the present work are suitable for our dataset, however, they generally vary from dataset to dataset.

ACKNOWLEDGEMENT

The authors wish to thank Director, CSIR-National Geophysical Research Institute for his kind consent to publish this work (Ref No: NGRI/Lib/2021/pub-80). The first author is grateful to the Department of Science and Technology, Government of India, for providing Inspire Fellowship (IF18030). The author thanks Mr. Animireddy Ramesh for the help in preparing the figures.

REFERENCES

Bednar, J.B., 1983. Applications of median filtering to deconvolution, pulse estimation, and statistical editing of seismic data. Geophysics, 48: 1598-1610.

Bekara, M. and van der Baan, M., 2007. Local singular value decomposition for signal enhancement of seismic data. Geophysics, 72(2): V59-V65.

Burger, H.C., Schuler, C.J. and Harmeling, S., 2012. Image denoising. Can plain neural networks compete with BM3D? IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2392-2399.

Chen, H., Zhang, Y., Kalra, M.K., Lin, F., Chen, Y., Liao, P., Zhou, J. and Wang, G., 2017. Low-dose ct with a residual encoder-decoder convolutional neural network. IEEE Transact. Medic. Imag., 36: 2524-2535.

Chen, Y. and Pock, T., 2017. Trainable nonlinear reaction-diffusion: A flexible framework for fast and effective image restoration. IEEE Transact. Pattern Analys. Mach. Intellig., 39: 1256-1272.

Chen, Y., Zhang, M., Bai, M. and Chen, W., 2019. Improving the signal-to-noise ratio of seismological datasets by unsupervised machine learning. Seismol. Res. Lett., 90: 1552-1564.

Du, B., Xiong, W., Wu, J., Zhang, L., Zhang, L. and Tao, D., 2016. Stacked convolutional denoising auto-encoders for feature representation. IEEE Transact. Cybernet., 47: 1017-1027.

Duncan, G. and Beresford, G., 1995. Some analyses of 2-D median f-k filters. Geophysics, 60: 1157-1168.

Glinsky, M.E., Clark, G.A. and Cheng, P.K.Z., 1996. Automatic event picking in prestack migrated gathers using a probabilistic neural network. Geophysics, 66: 1488-1496.

Gondara, L., 2016. Medical image denoising using convolutional denoising autoencoders. 16th Internat. IEEE Conf. Data Mining Workshops (ICDMW): 241-246.

Harris, P. and White, R., 1997. Improving the performance of f-x prediction filtering at low signal-to-noise ratios. Geophys. Prosp., 45: 269-302.

Huang, W., Wang, R., Chen, X. and Chen, Y., 2017a. Double least-squares projections method for signal estimation. IEEE Transact. Geosci. Remote Sens. Lett., 55: 4111-4129.

Huang, L., Dong, X. and Clee, T.E., 2017. A scalable deep learning platform for identifying geologic features from seismic attributes. The Leading Edge, 36(3): 249-256.

Jain, V. and Seung, S., 2009. Natural image denoising with convolutional networks. Advan. Neur. Informat. Process. Syst., **(Vol. No. ???)** 769-776.

Kingma, D.P. and Ba, J., 2014. Adam: a method for stochastic optimization. CoRR., Vol. abs/ 1412.6980, 1-15.

Kline, D.M. and Berardi, V.L., 2005. Revisiting squared-error and cross-entropy functions for training neural network classifiers. Neural Comput. Applicat., 14: 310-318.

Kononenko, I., 2001. Machine learning for medical diagnosis: history, state of the art and perspective. Artific. Intellig. Medic., 23: 89-109.

Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. Communicat. ACM, 60(6): 84-90.

Li, T. and Mallick, S., 2014. Multicomponent, multi-azimuth pre-stack seismic waveform inversion for azimuthally anisotropic media using a parallel and computationally efficient non-dominated sorting genetic algorithm. Geophys. J. Internat., 200: 1136-1154.

Nath, S.K., Chakraborty, S., Singh, S.K. and Ganguly, N., 1999. Velocity inversion in cross-hole seismic tomography by a counter-propagation neural network, genetic algorithm, and evolutionary programming techniques. Geophys. J. Internat., 138: 108-124.

Pratt, R.G., 1999. Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model. Geophysics, 64: 888-901.

Ross, C.P. and Cole, D.M., 2017. A comparison of popular neural network facies-classification schemes. The Leading Edge, 36(4): 340-349.

Rowley, H.A., Baluja, S. and Kanade, T., 1998. Neural network-based face detection. IEEE Transact. Patt. Analys. Mach. Intellig., 20: 23-38.

Satyavani, N., Sain, K., Lall, M. and Kumar, B.J.P., 2008. Seismic attribute study for gas hydrates in the Andaman Offshore India. Marine Geophys. Res., 29: 167-175.

Song, H., Yang, G., Chen, W., Xue, Y., Zhang, H. and Zhang, X., 2020. Seismic random noise suppression using deep convolutional autoencoder neural network. J. Appl. Geophys., 178: 104071.

Wang, Y., Zhou, H., Zu, S., Mao, W. and Chen, Y., 2017. Three-operator proximal splitting scheme for 3D seismic data reconstruction. IEEE Geosci. Remote Sens. Lett., 14: 1830-1834.

Wu, Z. and Liu, T., 2008. Wavelet transform methods in seismic data noise attenuation. Progr. Geophys., 23: 493- 499.

Zhang, K., Zuo, W., Chen, Y., Meng, D. and Zhang, L., 2017. Beyond a Gaussian denoiser. Residual learning of deep CNN for image denoising. IEEE Transact. Image Process., 26: 3142-3155.

Zhang, M., Liu, Y., Bai, M. and Chen, Y., 2019. Seismic noise attenuation using unsupervised sparse feature learning. IEEE Transact. Geosci. Remote Sens. Lett., 57: 970-972.

Zhao, Q., Du, Q., Sun, W. and Chen, Y., 2019. Iterative double Laplacian-scaled low-rank optimization for under-sampled and noisy signal recovery. Expanded Abstr., 89th Ann. Internat. SEG Mtg., San Antonio: 4675-4679.