Submission code: 79NZG0YA95

See this and other abstracts on our website: https://home.sbgf.org.br/Pages/resumos.php

A Real Benchmark Swell Noise Dataset for Performing Seismic Data Denoising via Deep Learning

Pablo Machado Barros (Cenpes-Petrobras), André Bulcão (Petrobras), Giovanny Alberto Meneses Arboleda (COPPE; Federal University of Rio de Janeiro), Lessandro De Souza Sadala Valente (COPPE; Federal University of Rio de Janeiro), Roosevelt Sardinha (COPPE; Federal University of Rio de Janeiro), Isabelle Rodrigues Vaz de Melo (COPPE; Federal University of Rio de Janeiro), Alexandre Esukov (COPPE / Federal University of Rio de Janeiro), ALBINO AVELEDA (COPPE; Federal University of Rio de Janeiro), SERGIO NETTO (COPPE; Federal University of Rio de Janeiro)

A Real Benchmark Swell Noise Dataset for Performing Seismic Data Denoising via Deep Learning

Abstract.

Deep learning (DL) gained great importance in geophysics exploration due to its great generalization capacity [1], which allows the prediction of complex system states. This ability can be leveraged in various applications in the oil and gas industry, such as seismic denoising. However, the difficulty of applying and exploring the different methods often lies in the lack of open datasets that can be used as a benchmark, given their scarcity, which makes testing and reproducibility of results difficult. This work presents a benchmark dataset of synthetic seismic data with real extracted swell noise and introduces a new evaluation metric that can capture small variations in model results. Additionally, a comparison between two widely known DL-based denoising models is performed on the proposed dataset, which constitutes a benchmark for accelerating the development of new solutions for seismic data denoising. The results show the effectiveness of the models for the seismic data-denoising task. However, the preservation of signal integrity remains a challenging problem.

The proposed benchmark dataset is composed by the addition of real swell noise to synthetic seismic data. The synthetic data is the simulation of the seismic responses of Marmousi [2] and SEAM [3], two open-source structural models, and was used to generate four synthetic datasets. The noise data were obtained from a filtering process of raw seismic data from two different acquisitions. Since the noise data were on a different scale, a rescaling process was needed before being added to the synthetic clean data. The noise rescaling was calculated to obtain the desired signal-to-noise ratio (SNR) from the noisy data (clean synthetic data + real noise). The two real noise files were combined with four synthetic seismic datasets at four different noise levels. The noise files were rescaled to a desired SNRe of the noisy data L = 1, 2, 5, and 10, resulting in 16 noisy data files. Two supervised image denoising DL methods were used to test the proposed benchmark dataset: the Fully Convolutional Network (FCNN) [4] and the generative module of SRGAN [5], [6]. To quantify the results of the models, a new metric called signal-to-noise relative ratio (SNR2) is proposed. The SNR2 is a relative measure that is calculated as one minus the ratio of the original data's SNR and the model prediction's SNR. This metric allows small variations in the model results to be recorded. To test the robustness of the models and the difficulty of the database, 32 different experiments were performed in a cross-validation fashion, taking into account the noise level and the seismic file. These experiments can be used as benchmarks for future works, and the noise files can be combined with other values of SNR or with different seismic data to generate new data. The results show that the proposed metric is highly sensitive to the mistakenly removed signal in the denoising process. To obtain higher SNR2 values, the model must mitigate the noise and preserve the signal. The proposed benchmark dataset can be used to develop future DL models for seismic denoising since it presents a challenging open problem. The models tested reached similar values across all experiments when evaluated by the SNR2 metric, and were able to mitigate the noise component even in the files with high noise levels.

[1] S. Yu and J. Ma, "Deep Learning for Geophysics: Current and Future Trends," *Rev. Geophys.*, vol. 59, no. 3, p. e2021RG000742, 2021, doi: 10.1029/2021RG000742.
[2] A. Brougois, M. Bourget, P. Lailly, M. Poulet, P. Ricarte, and R. Versteeg, "Marmousi, model and data," presented at the EAEG Workshop - Practical Aspects of Seismic Data Inversion,

European Association of Geoscientists & Engineers, May 1990, p. cp. doi: 10.3997/2214-4609.201411190.

[3] C. Regone, J. Stefani, P. Wang, C. Gerea, G. Gonzalez, and M. Oristaglio, “Geologic model building in SEAM Phase II — Land seismic challenges,” *Lead. Edge*, vol. 36, no. 9, pp. 738–749, Sep. 2017, doi: 10.1190/tle36090738.1.

[4] C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, and C.-W. Lin, “Deep learning on image denoising: An overview,” *Neural Netw.*, vol. 131, pp. 251–275, Nov. 2020, doi: 10.1016/j.neunet.2020.07.025.

[5] C. Ledig *et al.*, “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 105–114. doi: 10.1109/CVPR.2017.19.

[6] R. Abma and J. Claerbout, “Lateral prediction for noise attenuation by t-x and f-x techniques,” *GEOPHYSICS*, vol. 60, no. 6, pp. 1887–1896, Nov. 1995, doi: 10.1190/1.1443920.