Submission code: RQXKRWLZV5

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

Data preconditioning and parameters selection for deep learning-based first-break picking

Carlos Eduardo Dos Santos Garabito (CPGeo; Brasil), Amir Mardan (WSP Canada)

Data preconditioning and parameters selection for deep learning-based first-break picking

Please, do not insert author names in your submission PDF file

Abstract Summary

First arrival picking is a crucial task in the land seismic data processing workflow, especially for data with rough topography and heterogeneous weathering zones. This interactive process is very tiring and time-consuming, so there are many approaches to automating this process. But even semi-automated and automated processes require quality control, which is also time-consuming. In recent years, various solutions leveraging artificial intelligence techniques have been proposed for first-break picking. These new techniques present other challenges, such as efficient training and calibration of parameters to produce more accurate results. The aim of this work is to investigate strategies for efficient data training to achieve accurate results using a deep learning-based first-break picking algorithm built on a U-Net architecture, which considers the seismic datasets as image and the first-break picking is solved as a segmentation problem. Then, the aim of this work is to study the effects on the accuracy of picks generated using data with different acquisition geometries for network training, data preconditioning and data argumentation. Here we present the preliminary results of this study, in which some shots of the dataset without any preconditioning were used to train the network and the picking accuracy on all the data was satisfactory.

Introduction

In land seismic data processing for oil exploration, the first-break picking process is a critical step, as these travel times represent the first arrivals of the seismic signal reaching the receivers. They are fundamental for determining the near-surface (weathered zone) geological structure and velocity model through refraction interpretation techniques and travel time tomography. This velocity model is typically used for static corrections, which correct for distortions in the reflection data caused by variations of the elevations and weathered zone heterogeneities. This is achieved by applying time shifts to the data traces recorded on the ground surface transforming them to a flat datum (Cox, 1999). Furthermore, an accurate near-surface velocity model can contribute to refining the deeper velocity model, especially when depth imaging is applied directly from the topography (Yilmaz et al. 2022).

Picking errors are common because the process typically involves manual or semi-automatic methods. The manual picking is time-consuming, subjective, and prone to operator variability, while semi-automatic picking still requires time-consuming quality control for accuracy. In recent years, several FB-picking algorithms have been introduced using deep learning (DL) techniques with the aim of making it automatic and thus reducing the excessive time spent on this process and improve the accuracy of the picks. In this study we will focus on the FB-picking algorithm introduced by Mardan et al. (2023), where the reader can find a good review of DL-based FB-picking algorithms. This FB-picking algorithm uses a U-net architecture that employ a convolutional neural network (CNN) specifically designed for image segmentation. This means the seismic shot gather is handled as an image represented by pixels, and the first arrivals correspond to the boundary separating the segmented areas within it.

This work will investigate the impact of preconditioning techniques on picking accuracy when applied to input data to attenuate noise before the first arrivals. We will also study the limitations of using a general training dataset, derived from several seismic survey projects using a similar acquisition geometry, when applied to a specific sedimentary basin. Furthermore, we will investigate how the data augmentation parameters, such as the width of sub-images and the overlap between sub-images of the seismic shot gathers, affect the accuracy of the picks. As previously stated, this study will use the DL-based FB-picking codes develop Mardan et al. (2023). All tests will be carried out using public 2D seismic data from the Barreirinhas Basin, located in north-eastern Brazil. On the other hand, codes will be developed for semi-automatic picking of the first breaks in the dataset to train the neural network and also for quality control of the picks.

Method

The key to have an efficient network that can predict with high accuracy is training with a suitable dataset. A network should be trained with a large amount of data, and the data should represent all possible scenarios that might exist in the real-world problem that is tackled.

To achieve effective results, the theory behind this neural network uses the principles of convolutional neural networks (CNNs) for image segmentation. By transforming seismic shot gathers into images, the model identifies the FBs as boundaries within segmented areas. The U-net architecture is particularly well-suited for this purpose because of its encoder-decoder structure, which allows for high-resolution feature extraction and precise segmentation. The encoder captures essential features from the input data, while the decoder reconstructs the output image with the segmented boundaries, ensuring the picks fit with seismic first-arrival times.

To further enhance the U-Net model's performance, our approach incorporates residual blocks that have been pretrained to solve the ImageNet classification problem. Residual blocks effectively counteract the vanishing gradient problem during training, thereby ensuring deeper layers of the network retain meaningful feature representations. Leveraging pretrained weights from ImageNet allows the model to start from a well-calibrated initialization, improving its ability to generalize to seismic data, and requiring smaller dataset. This transfer learning strategy ensures that the U-Net architecture is not only optimized for segmentation tasks but also benefits from the nuanced feature extraction capabilities developed during the ImageNet problem-solving process.

The use of GPUs (Graphics Processing Units) or MPS (Metal Performance Shaders) with PyTorch accelerates computations, enabling the network to handle extensive datasets efficiently while maintaining precision. A detailed description of the theory about the DL-based first-break picking algorithm used in this work can be found in Mardan et al. (2023) and the code is publicly available to download from https://github.com/geostack/first_break_picking.

In this study, we will incorporate advanced preconditioning techniques to minimize noise distortion before feeding the data into the network. Preconditioning steps such as filtering and normalizing the shot gathers enhance the clarity of arrivals, allowing the model to discern boundaries more accurately. Parameters such as sub-image width and overlap are fine-tuned to optimize segmentation performance and account for variable acquisition geometries.

Results

The carry out the FB picking initial tests with the DL-based FB picking codes, we used a 2D dataset from the Barreirinhas Basin consisting of 225 shot gathers acquired with split-spread geometry. At this stage of the study, no pre-processing to the dataset. The first arrivals for training the network were manually picked using the BotoSeis software (Garabito 2019). We used 10% (i.e. 22 shot gathers) of the dataset for training the first-break picking algorithm. Figure 1 shows that the training of the model is progressing effectively, since both the loss decreases and accuracy increases. The validation loss closely follows the training loss, suggesting that the model is generalizing well to unseen data and not significantly overfitting. Both loss and accuracy stabilize around epochs 7-8, so there is no further training beyond this point. Accuracy is being measured using the cross-entropy loss function. The neural network processes a sub-image of the seismic shot and, for every single pixel, it calculates a probability that it belongs to each of the two classes. Thus, the cross-entropy loss function quantifies the difference between the network's predicted probabilities and the true labels for each pixel.

After training, we ran the prediction algorithm on an unknown dataset, i.e. data that the model had never seem before. Figure 2 and 3 show the results of the automatic first arrivals picked values (blue lines) for the shot gathers. As can be seen, the prediction is correct when there is no noise in the data before the first arrivals (Figure 2). On the other hand, when there is noise before the first arrivals, the prediction result is inaccurate and picking of the first arrivals fails.

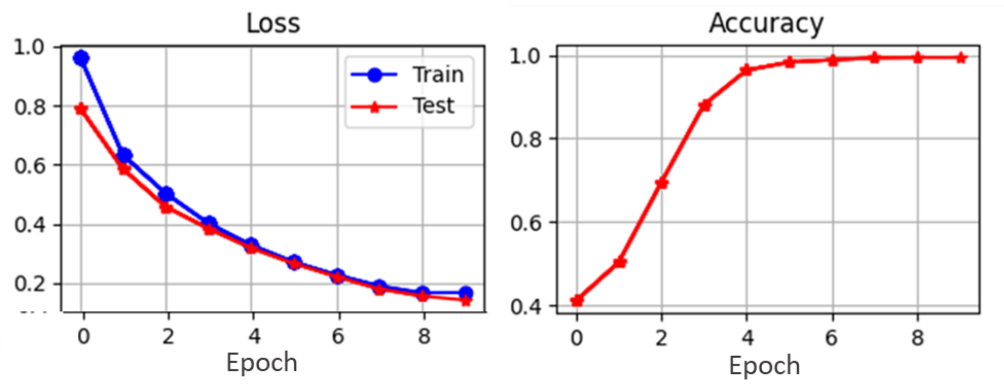

Figure 1: Training results for first-break picking using 22 shot gathers of the Barreirinhas Basin dataset. a) the model learning evolution and b) the model's predictions.

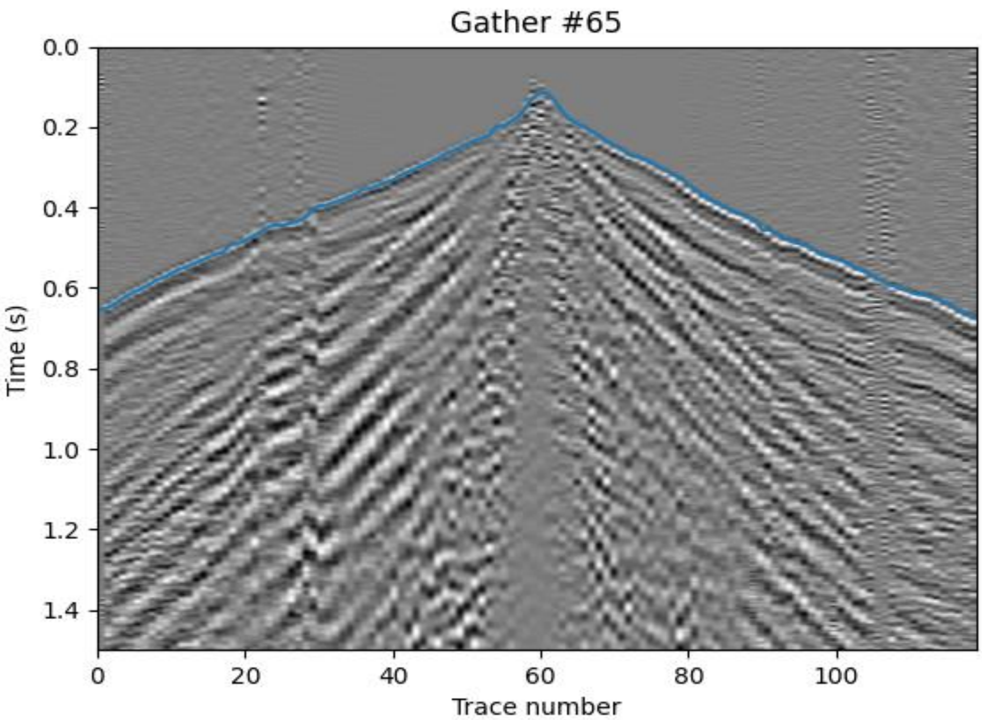

Figure 2: Shot gather 65 of the seismic dataset of the Barreirinhas Basin. The blue line corresponds to the values picked for first-break.

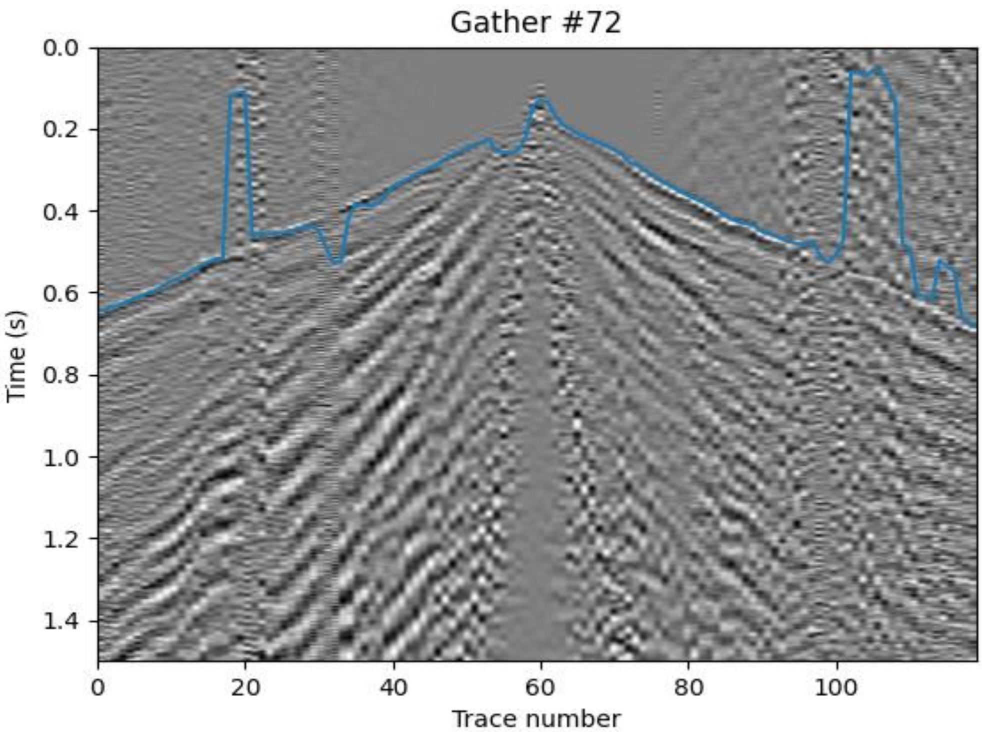

Figure 3: Shot gather 72 of the seismic dataset of the Barreirinhas Basin. The blue line corresponds to the values picked for first-break.

Conclusions

We present the results of the first tests of the DL-based first-break picking algorithm applied on seismic data of the Barreirinhas Brasilian in Brazil. These results demonstrate that challenges must be overcome for this automated algorithm to work properly. Our ongoing research will focus on preconditioning the data and calibrating the parameters to achieve more accurate results.

References

Cox, M. (1999) Static Correction for Seismic Reflection Surveys. Society of Exploration Geophysicists, Tulsa, Oklahoma, US.

Garabito, G., 2019, Tips for seismic data processing with BotoSeis & Seismic Unix, https://germangarabito.wordpress.com

Mardan, A., Blouin, M., Fabien-Ouellet, G., Giroux, B., Vergniault, C. & Gendreau, J. (2024) A fine-tuning workflow for automatic first-break picking with deep learning. Near Surface Geophysics, 22, 539–552. https://doi.org/10.1002/nsg.12316

Yilmaz, O., Gao, K., Delic, M., Xia, J., Huang, L., Jodeiri, H. & Pugin, A., (2022) A reality check on full-wave inversion applied to land seismic data for near-surface modeling. The Leading Edge, 41, 40–46.