Assisted fault identification and surface extraction by machine learning, a case study from Oman

Fan Jiang, Phill Norlund, Halliburton Landmark; David Dietz, Occidental of Oman*

Summary

Fault interpretation is an important step in the seismic interpretation process and is critical for understanding challenges such as reservoir compartmentalization, fluid migration, and drilling hazards. Recently, assisted fault interpretation workflows leveraging machine learning techniques have become a promising way to automatically identify faults in seismic. Convolutional neural networks (CNNs) are a popular new method to identify fault attributes in seismic data by analyzing image segmentation and feature extraction. In this abstract, we applied a dual-channel CNN architecture to train seismic data and its discontinuity attribute together to increase the fidelity of the fault prediction process. We first trained a model using synthetic data only, we then trained a second model augmenting it with real data from the study area. We then implemented an unsupervised machine learning clustering approach to analyze the fault probability map and extract fault sticks. This automated workflow took less than an hour to complete compared with over a week for an experienced geoscientist to manually pick approximately 200 faults in the same study area. This experience shows that machine learning-based fault imaging and extraction is a valuable tool for fault interpretation. The automated workflow can be used to provide a quick initial fault interpretation or to identify alternative interpretations and better assess fault uncertainty.

Introduction

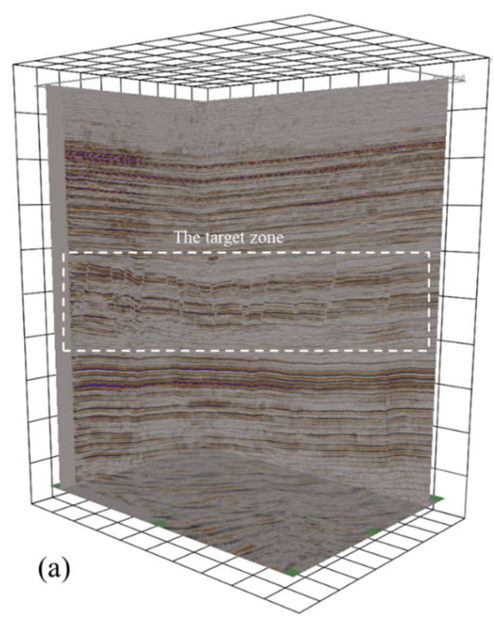
Faults are important for many aspects of hydrocarbon exploration and development. Accurately identifying fault locations is necessary for trap identification and well planning. In recent years, fault interpretation driven by machine-assisted approaches has shown to be a promising way to reduce interpretation cycle time and improve the accuracy of prediction (e.g., Philit et al., 2019, Wu et al., 2019). In this abstract, we apply a machine learning-based fault interpretation workflow to first generate fault probability seismic volumes, and then implement an unsupervised machine learning approach to generate fault sticks from these volumes. The complete workflow took less than an hour compared with over a week to interpret the areas by an experienced geoscientist. These results show that machine learning-based fault interpretation workflow can be considered an important tool for assisting in the interpretation of seismic data.

The study area is a survey of around 460 square kilometers in Oman provided by Occidental of Oman. The 3D seismic is composed of roughly two million 3D seismic traces and is generally of very good quality. In this area, Riedel shear provides the dominant structural setting, creating an en-echelon pattern of normal faults. Several algorithmic automatic fault extraction methods were previously attempted on this dataset with disappointing results.

In this evaluation, we trained a dual-channel convolutional neural network (Jiang and Norlund, 2020) first with synthetic seismic data only, and then with a combination of synthetic data and manually interpreted data from a subset of the study area. The dual-channel network is a combination of seismic data and its discontinuity attribute which improves the accuracy of fault prediction. We then applied an unsupervised machine learning clustering algorithm to generate fault sticks from the fault probability volume.

The study area

Figure 1 shows the seismic data area for the study area in North Oman. The target zone is dominated by en-echelon normal faults created in a trans-tensional, Riedel shear structural setting. Figure 1a illustrates the entire survey area and the faulted subset studied in this abstract. Figure 1b shows a seismic section from the target area where it shows a clear normal fault pattern. Figure 1c is the discontinuity attribute derived from the seismic section.



(a)

10.1190/segam2021-3581804.1

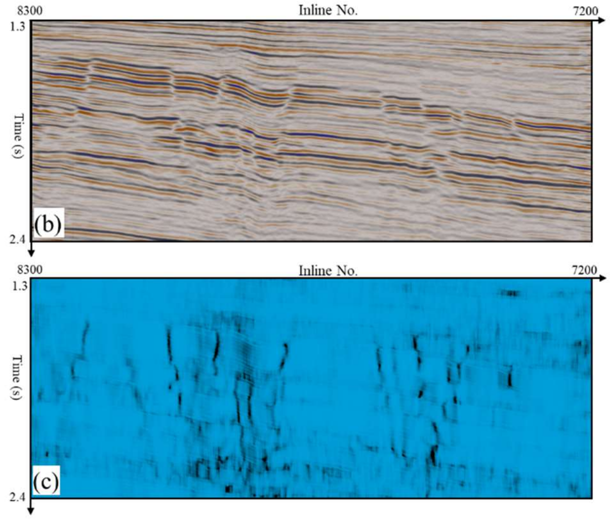


Figure 1: The study area in Oman. (a) The entire survey area and the target zone; (b) the seismic section with normal fault pattern; (c) The discontinuity attribute derived from (b).

Fault prediction by a synthetic deep learning model

We started with a model trained with synthetic data only to predict an initial fault probability volume. A dual-channel convolutional neural network was implemented to train the model. The training dataset was generated using a synthetic data generator, then the synthetic seismic data was used to calculate a discontinuity attribute. The final training data was composed of seismic and its discontinuity attribute together. A normalization scheme was also applied to the dual-channel data to constrain the amplitude range of input data in a similar magnitude. This helps to avoid overfitting when predicting the fault probability volume. Once we have trained the synthetic model, we then calculate the discontinuity attribute for the testing data and apply the same normalization scheme. We used manually interpreted faults as the benchmark to validate our algorithm. Figure 2(a) shows a section view with faults predicted by the synthetic trained model. The manually interpreted faults are then overlain on the predicted fault probability volume. The result shows that the deep learning model trained by synthetic data can successfully identify most of the fault segments from the raw seismic data.

This fault probability volume could be used by geoscientists during the initial interpretation phase to guide their work. It additionally could be used post-interpretation to help identify any missed or alternate interpretations. This further analysis can both improve the quality of the interpretation and bring insight into any uncertainty around the interpretations. How to best train machine learning models

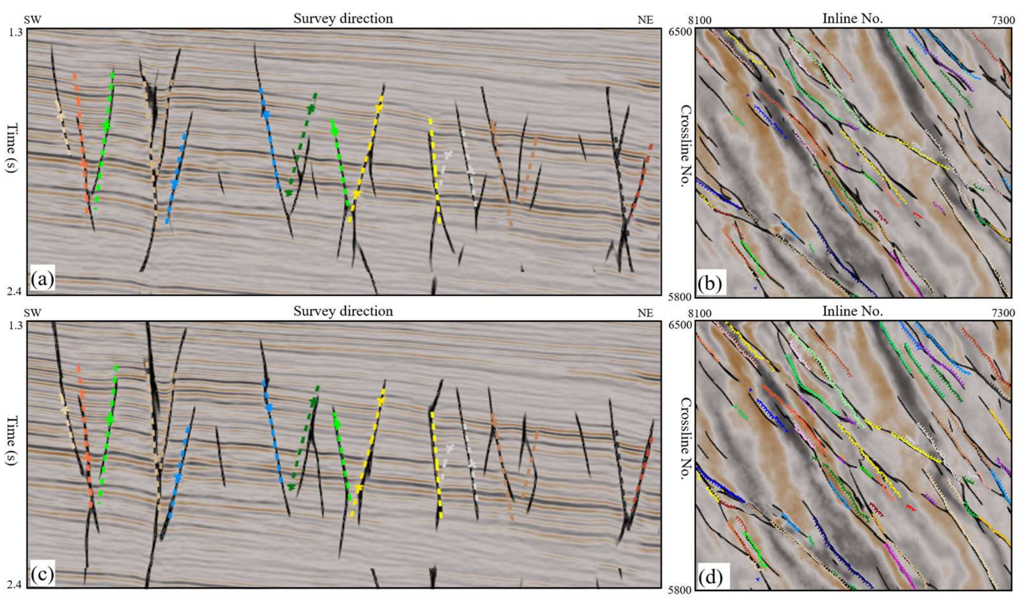


Figure 2: Fault prediction by our machine learning models (black lines) overlain on top of the original seismic amplitude along with manually interpreted fault sticks (dashed colored lines). (a) & (b) section and time slice comparing the model generated with synthetic data only. (c) & (d) section and time slice comparing the model augmented with real data. Overall, the machine learning results match very well with the manual interpretation with the augmented model generating cleaner, more continuous faults. However, there are some areas where the synthetic-data-only model performs better.

to further understand the geology is still a topic of research, however, it is still necessary to require an experienced interpreter to make the final decisions in the fault identification process.

Fault prediction by an augmented deep learning model including manually interpreted faults

To improve the synthetic data trained model, we picked a small subset of the seismic cube (and its manually interpreted faults) and used them as additional training data and labels to re-train the synthetic model. The selected seismic used to re-train the model represents ~1.7% of the total survey area and ~3% of the total training data. Therefore, the model is still dominated by synthetic data but augmented by a small portion of real data. Figure 3a shows the selected seismic used to re-train the model, several selected fault planes used as additional training labels in Figure 3b. Figure 3c shows the converted fault mask from Figure 3b.

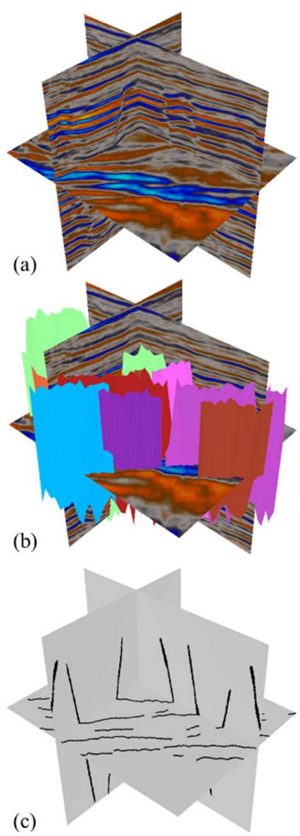


Figure 3: The additional data used to re-train the synthetic model. (a) The selected 3D seismic data; (b) Fault sticks picked by an interpreter; (c) The converted 3D fault mask.

Figures 2 (c) and (d) show the prediction result from the deep learning model re-trained with a subset of real data from the survey of interest. We observed that most faults are correctly predicted compared with the manual interpretation and the original synthetic model. Additionally, the augmented model shows cleaner, more continuous faults. Like the synthetic-only model it still missed identifying a few fault locations where manual faults were picked. This could be due to either the weak discontinuity along the seismic section or limitations of the deep learning algorithm. There are also several places where a predicted fault segment from the deep learning model does not make geological sense compared with the manually interpreted faults. This should be improved by including more valid data points from both the synthetic and the augmented data inputs.

We noticed that in some areas the synthetic-only model shows better predictions than the model including the manual interpretations. This highlights some of the limitations of including non-synthetic data. Firstly, even in the small subset of the volume we selected not every fault in the seismic has been interpreted. These un-interpreted faults get included in the model, reducing the model's ability to identify all faults, Secondly, interpretation can be a highly subjective process. Even experienced geoscientists can interpret the same seismic section in many different ways (Alcalde et al., 2017). By including just one geoscientists interpretation into the model you are adding that person's biases (both good & bad) into the model and thus limiting the ability of the model to generate genuine alternate & valid predictions.

We may find that as the interpreter trains the machine learning tool, the machine learning tool can also help train the interpreter. If the machine learning results are not as expected, the interpreter can update the user-provided training. This interactive feedback loop can help an interpreter refine their fault picking skills to get an optimally trained CNN. Also critical in this process is to have the synthetic training dataset match the geologic environment. Trying to predict a trans-tensional fault pattern using a synthetic dataset from compressional models would be futile. The closer the structure and stratigraphy of the synthetic model to the real geology, the better.

Fault extraction by unsupervised machine learning approach

As the last step, we implemented an unsupervised machine learning clustering approach to generate fault sticks from the fault probability maps. A machine learning clustering approach is used to find the K nearest neighbors of a specific point or location, then go around to find all neighbors within a certain radius (Rabbani et al., 2006). Once all neighbor points are found, we consider them as seed points, growing from the point that has the minimum curvature value until there are no unlabeled points in the cloud. Figure 4 shows the fault planes extracted in the study area. To interpret faults from the seismic data, an experienced geoscientist spent over a week manually picking approximately 200 correlated faults whilst the machine learning approach only took less than an hour to identify the fault locations as well as generate the fault sticks. Having editable fault sticks provides the interpreter with the ability to keep the machine learning results where it was successful and make edits to the fault sticks where a manual interpretation could improve the results. This leverages the power of the machine learning capabilities and the geologic knowledge of the interpreter.

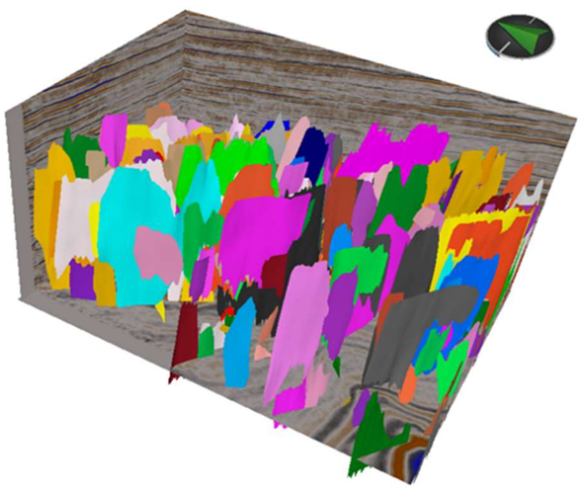


Figure 4: Fault planes extracted by unsupervised machine learning approach.

Discussions and Conclusions

In this abstract, we implemented a dual-channel convolutional neural network to successfully generate fault probability volumes from a dataset in Oman. Our first volume was created using only synthetic data and this machine learning-based approach correctly predicted faults in much of the seismic data as verified by comparing to the manual fault interpretation. The volume also identified some missed faults as well as valid alternative interpretations.

Our second volume used both synthetic data and manual interpreted faults whilst in most areas it performed superior to the synthetic only volume (correctly identifying faults missed by the synthetic only model and generating cleaner, more continuous fault probabilities) in some areas it performed less well highlighting the limitations of using subjective data such as manual fault interpretations to train machine learning models.

After creating the fault probability volume, we then successfully ran an unsupervised machine learning clustering algorithm to extract fault sticks from the fault probability volume. The resulting fault planes compared very favorably with the manual interpretations in most areas. However, we also observed that some machine learning faults did not make geological sense, which could due to missing the necessary fault types to train a model, e.g. strike-slip fault, reverse fault, or due to weak signal-to-noise ratio. The manual fault interpretation used in this project required an experienced geoscientist who spent over a week manually picking segments whilst the machine learning approach took less than an hour to complete the workflow.

Machine learning fault extraction is a valuable new addition to the discontinuity and curvature tools currently used by exploration and development geophysicists. An exploration geophysicist can use the machine learning fault extraction to help with initial recon mapping of an area. A development geophysicist can use the machine learning fault extraction to provide fault locations for well planning. In all cases, the machine learning results will still require detailed review by an interpreter to assess the validity of the automated fault prediction.

Acknowledgments

We thank Occidental Petroleum Corporation for providing the seismic and fault data and their permission to use it in this research.

REFERENCES

Alcalde, J., C. E. Bond, G. Johnson, R. W. H. Butler, M. Cooper, and J. Ellis, 2017, The importance of structural model availability on seismic interpretation: Journal of Structural Geology, **97**, 161–171, doi: https://doi.org/10.1016/j.jsg.2017.03.003.

Jiang, F., and P. Norlund, 2020, Seismic attribute-guided automatic fault prediction by deep learning: 82nd Annual International Conference and Exhibition, EAGE, Extended Abstract, 1–5, doi: https://doi.org/10.3997/2214-4609.202010542.

Philit, S., S. Lacaze, and F. Pauget, 2019, Innovative automatic fault detection using a volume 3D scanning method: 89th Annual International Meeting, SEG, Extended Abstract, 1923–1927, doi: https://doi.org/10.1190/segam2019-3214754.1.

Rabbani, T., F. V. D. Heuvel, and G. Vosselmann, 2006, Segmentation of point clouds using smoothness constraint: Proceedings of the ISPRS Commission V Symposium, Vol. **35**, 248–253.

Wu, X., L. Liang, Y. Shi, and S. Fomel, 2019, FaultSeg3d: Using synthetic data sets 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.