Active deep learning for seismic facies classification

Nayguel Costa, Paulo Canas Rodrigues, Luciano Rebouças
Federal University of Bahia, Brazil

Abbreviated abstract: Interpretation of seismic facies is vital for oil and gas exploration; however, it is a costly and labor-intensive task. Although deep learning applications reduce the interpretation time, they depend on sufficient large, annotated training data. Active learning (AL) methods create an optimized labeled training set from unlabeled data. In this ongoing study, we perform seismic facies classification by training a deep neural network on 3-D seismic data. Then, we simulate a random sampling AL, proving that the results can be achieved with a smaller labeled dataset.

Related publications:

– Chai *et al*, IEEE Transactions on Geoscience and Remote Sensing (60), 4507719 (2022)
– Smailagic *et al*, WIREs Data Mining Knowledge Discovery (10), 1353 (2020)

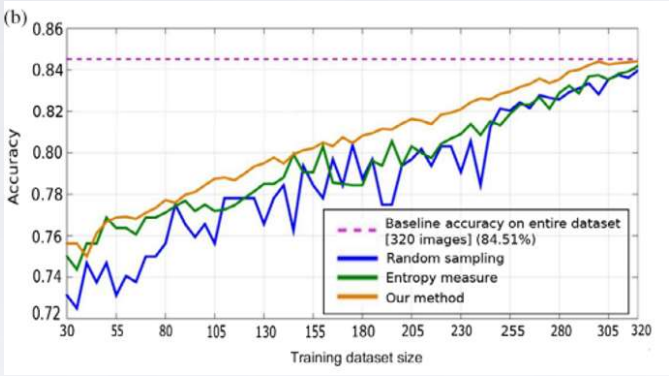
UFBA Universidade Federal da Bahia

nayguele@gmail.com

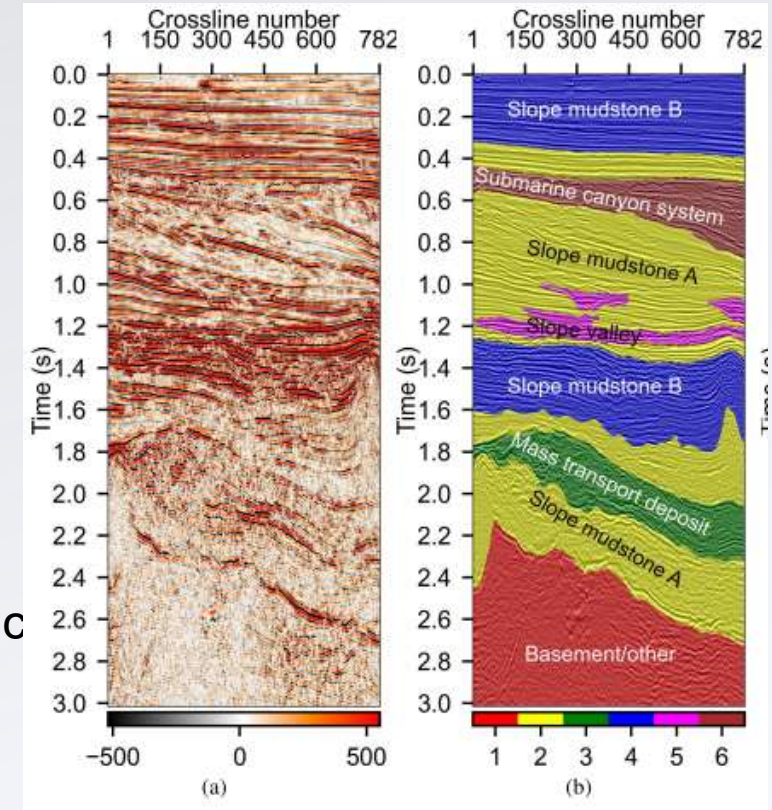
4th Conference on Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022

Problem, Data, Previous Work

- Seismic Facies Classification (SFC) means identifying relevant geological features in the seismic data.
- Deep learning-based algorithms have been successfully applied to SFC (ex: Chai *et al,* 2022).
- Since labeling seismic data is a costly process, sufficient large training data is rare in the real world.
- Active learning methods create an optimized labeled training set from unlabeled data. Active learning has gained popularity in the medical image segmentation domain in recent years (ex: Smailagic *et al, 2020).*



Competing AL methods in a breast cancer dataset (balanced multiclass).



Seismic inline 1, (Parihaka dataset) and accompanying label section.

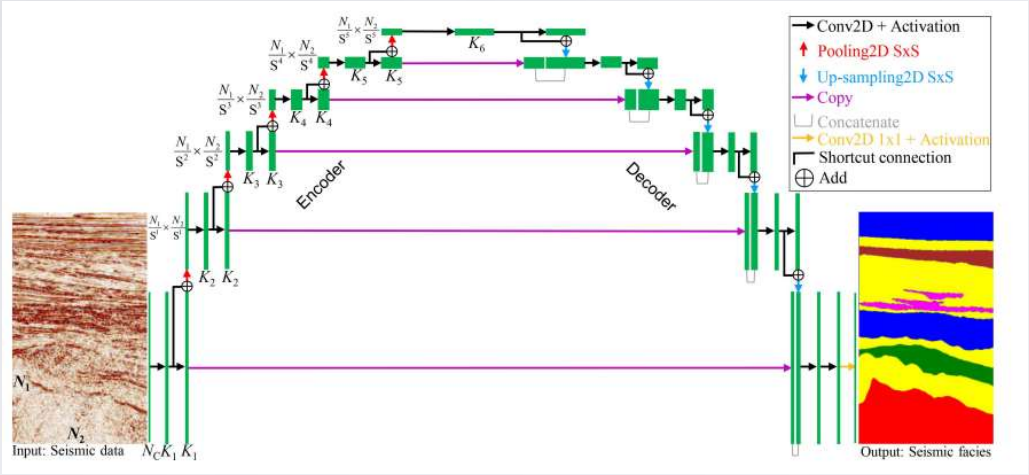
UFBA Universidade Federal da Bahia

nayguele@gmail.com

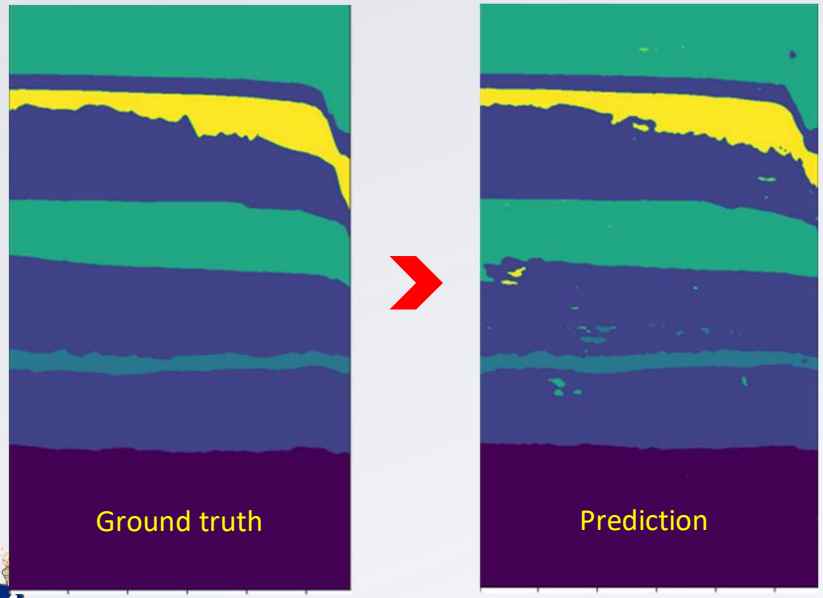
4th Conference on Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022

Methods

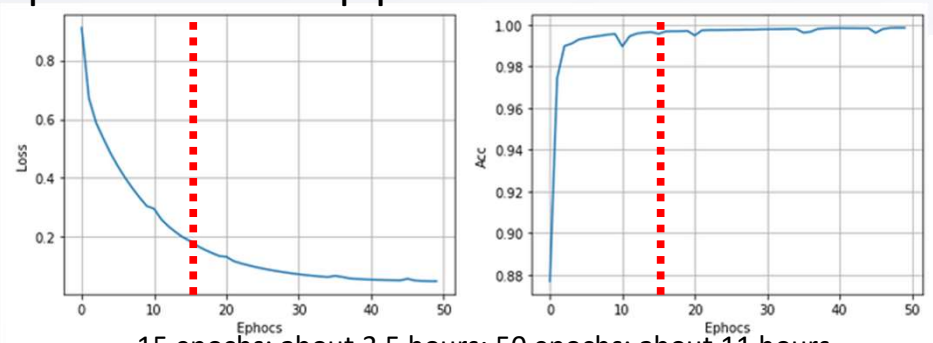
- We perform Seismic facies classification by training an end-to-end encoding-decoding deep network revised from U-Net.
- The metrics and continuity of predicted facies along the test sections indicate that the results are consistent with specialist's interpretation.



Chai et al (2022)



- In order to implement AL experiments, the model must be trained several times. Due to the size of the dataset this is very time consuming.
- To deal with this problem, we choose to interrupt the training to 15 epochs in the AL pipeline.



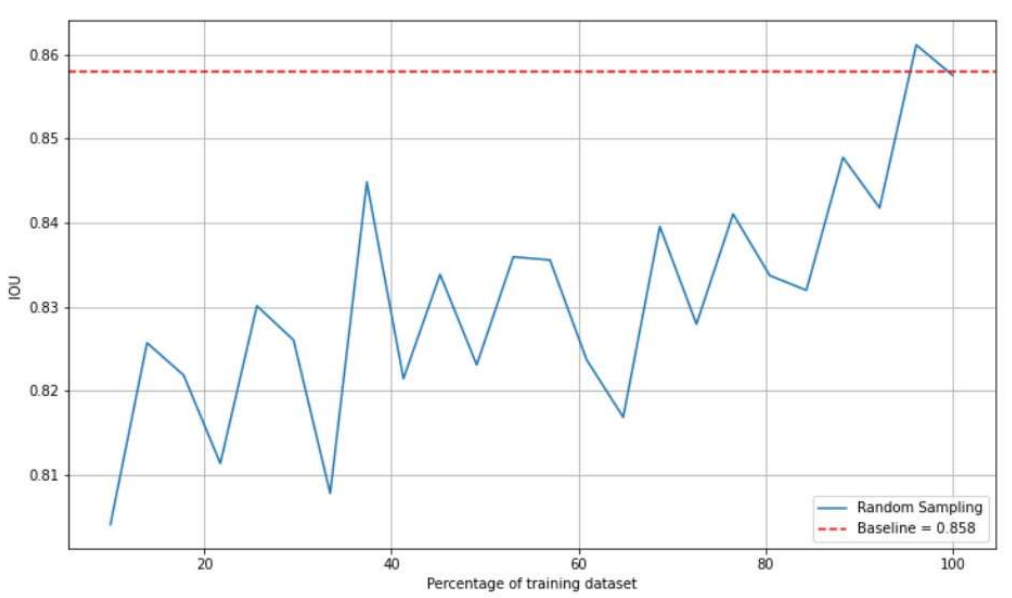
15 epochs: about 3.5 hours; 50 epochs: about 11 hours

UFBA Universidade Federal da Bahia

nayguele@gmail.com

Results and Conclusions

Metric	50 epochs	15 epochs
Accuracy	0.945	0.919
F1-score	0.896	0.835
F1-score (weighted)	0.945	0.921
IOU	0.823	0.745
IOU (weighted)	0.904	0.858
MCC	0.923	0.886



Metrics with 15 epochs are smaller, but still good. The random sampling AL curve shows that good results can be achieved with less than half of the labeled training dataset. The shape of the curve reflects characteristics of the 3D seismic data (ex: high correlation between adjacent slices).
The implementation of a more robust AL method is underway, aiming to achieve better results with fewer labeled examples.

UFBA Universidade Federal da Bahia

nayguele@gmail.com

4th Conference on Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022