Reconhecimento automatizado de zonas fraturadas em dados sísmicos abertos com Azimute Completo usando técnicas de Aprendizado Profundo

Lorena da Silva Oliveira Santos[1], Matheus Silva Gonçalves[1], Pablo Machado Barros[2], Pedro Henrique Silvany[2], Carlos Eduardo Borges de Salles Abreu[2], Alexandre Augusto Cardoso da Silva[2], Rogério Espíndola[3], Maria Célia Lopes[3], Lucas Dias[3], Bruno Souza[3]

(1) Kognitus, (2) Petrobras, (3) COPPE-UFRJ

Copyright 2024, SBGf - Sociedade Brasileira de Geofísica

Este texto foi preparado para a apresentação no X Simpósio Brasileiro de Geofísica, Salvador, 08 a 10 de outubro de 2024. Seu conteúdo foi revisado pelo Comitê Técnico do X SimBGf, mas não necessariamente representa a opinião da SBGf ou de seus associados. É proibida a reprodução total ou parcial deste material para propósitos comerciais sem prévia autorização da SBGf.

Resumo

A identificação de fraturas naturais é uma tarefa crítica para a definição do projeto de desenvolvimento da produção e para o gerenciamento de reservatórios portadores de hidrocarbonetos. Trata-se, contudo, de uma atividade complexa por envolver dados adquiridos em múltiplas escalas e suportes de observação. Foi desenvolvida uma metodologia de reconhecimento automatizado de zonas fraturadas em dados sísmicos abertos com Azimute Completo usando técnicas de Aprendizado de Máquina, onde os rótulos para treinamento são obtidos a partir de interpretação de fraturas em perfis de imagem. O reduzido número de dados de treinamento disponíveis impôs o uso de duas estratégias distintas: (i) técnicas de aumento de dados e (ii) realimentação da rede com predições nas imediações dos poços de treinamento. Os resultados iniciais mostraram que a metodologia proposta é promissora, necessitando aprimoramentos futuros.

Introdução

A caracterização de reservatórios é uma etapa fundamental para a previsão da produção de petróleo. A posição das heterogeneidades geológicas (falhas, fraturas, vugs, carstes) influenciam diretamente na dinâmica do reservatório, podendo levar a irrupção precoce de água e gás nos poços produtores. A interpretação destas heterogeneidades, frequentemente encontradas nos reservatórios carbonáticos do Pré-sal, são de suma importância para o desenvolvimento dos campos de petróleo.

A análise e interpretação de dados sísmicos permitem a identificação de estruturas geológicas, no entanto trata-se de uma tarefa desafiadora para as fraturas que estão no limite da resolução dos dados sísmicos. Nesse contexto, a grande riqueza de informações disponibilizadas a partir de dados pré-empilhamento com azimute completo é de fundamental importância, visto que se tem reportado na literatura uma relação entre variação de amplitude vs. afastamento e azimute quando da presença de zonas fraturadas (Fang et al., 2017; Hall and Kendall, 2003). Isso pois essa variação é um indicador de anisotropia (Harwood et. al, 1998), que tem como possível causa a presença de rochas fraturadas na região (Clark et. al, 2009).

Diversas abordagens utilizando aprendizado de máquina vêm sendo reportadas recentemente para aplicações em análise e interpretação de dados sísmicos, particularmente, usando modelos de visão computacional via redes convolucionais (Zheng et. al, 2019; Wang et al., 2020; Cheng et al., 2023; Sheng et. al, 2023). Em especial, muitos avanços têm sido observados no campo de identificação de falhas, produzindo métodos aplicáveis a diferentes contextos (Wu et al., 2019). A tarefa de identificação de fraturas a partir de amplitudes sísmicas também despertou interesse de diversos autores, que frequentemente aplicam técnicas de aprendizado de máquina para tal (Cheng et. al, 2022; Babasafari et. al, 2022; Yuan et. al, 2021). No entanto, o uso de dados pré-empilhamento com riqueza azimutal em conjunto com técnicas de aprendizado profundo ainda pode ser mais bem explorada para essa tarefa.

O objetivo deste trabalho é desenvolver uma metodologia para identificação de fraturas a partir de dados sísmicos pré-empilhamento com riqueza azimutal. Os rótulos de cada exemplo de treinamento se baseiam em informações processadas de perfis de densidade de fratura P32. Esses perfis dizem respeito à densidade de fratura volumétrica, ou seja, indicam a área de fraturas por unidade de volume da rocha (Dershowitz & Herda, 1992) e são produto da interpretação dos perfis de imagem dos poços realizada pela equipe do Ativo de Produção de um campo do pré-sal. Propõe-se a utilização de redes neurais para mapear a presença de fraturas em função da distribuição azimutal de amplitudes sísmicas. A quantidade de dados extraídos a partir desses perfis, no entanto, não permite a criação de redes muito complexas. Para contornar esse fato, propõe-se a utilização de duas estratégias adicionais: aumento de dados e realimentação. O aumento de dados gera novos exemplos para treinamento a partir de transformações dos dados originais utilizando técnicas derivadas de visão computacional, adaptadas para o contexto do presente trabalho de modo a evitar que se gerem dados espúrios. Já a realimentação permite que dados extraídos a partir de predições da própria rede sejam incluídos para o treinamento de novas redes. No entanto, é necessário um processo de validação desses dados para evitar que a rede se degenere, devendo se utilizar

apenas aqueles que se tenha bastante confiança de estarem corretos. Para testar a viabilidade da metodologia utilizaram-se dados provenientes de um campo em produção do pré-sal da bacia de Santos.

Metodologia

A abordagem para inferência de heterogeneidades geológicas, como fraturas e falhas, via redes convolucionais usando dados sísmicos vem apresentando resultados importantes recentemente (e.g. Wu et al., (2019); Yuan et. al, (2021)). Propõe-se neste trabalho seguir essa mesma linha, aplicando redes neurais convolucionais multicanais para a classificação binária de pontos sísmicos a partir de um volume de dados pré-empilhamento com riqueza azimutal em torno deles. Cada canal corresponde a um par azimute-offset de informação. Para fornecer os rótulos necessários para o treinamento, emprega-se um processo de anotação baseado em perfis de poço, detalhado a seguir. Sabe-se que esse processo de anotação gera relativamente poucos dados para treinamento, o que possibilita unicamente o uso de redes com arquiteturas muito simplificadas. Para contornar esse problema, propõe-se implementar duas etapas adicionais: aumento de dados e realimentação da rede, permitindo que predições possam ser incorporadas como dados de treinamento para possibilitar arquiteturas mais complexas.

A Figura 1 ilustra o fluxo de trabalho que se divide em três etapas: rotulagem de dados, aumento de dados e aprendizado supervisionado.

A primeira etapa é denominada rotulagem de dados e engloba todo o processamento dos dados brutos: identificação das fraturas a partir dos perfis P32 de cada poço, compatibilização desses dados com a resolução sísmica e extração das amplitudes sísmicas referentes aos pontos em que foram anotados os dados de poços. Nessa etapa permite-se também configurar o processo de maneira flexível, facilitando análises comparativas. Essas configurações englobam diferentes formas de processamento, como técnicas de *upscaling* e procedimentos de filtragem dos dados de perfil. Destaca-se aqui a possibilidade de selecionar diferentes frequências de corte para o filtro passa-baixa empregado nos dados de perfil. Seleciona-se a menor frequência capaz de cobrir um percentual, definido pelo usuário, do espectro de frequências. O efeito deste percentual será analisado em mais detalhes nas próximas seções.

A segunda etapa é o aumento de dados, que visa ampliar o conjunto de treinamento. Foram selecionadas apenas transformações que não corrompem características geofísicas dos dados sísmicos extraídos: rotações de 90°, 180° e 270° no plano horizontal; inversões horizontais e alterações de brilho pela multiplicação das amplitudes por um fator constante.

Por fim, a terceira etapa é a fase de aprendizado supervisionado e compreende o pré-processamento, com regularização da distribuição de azimutes/ângulos de afastamento e normalização; a definição de arquitetura de rede e demais configurações de treinamento (número de épocas, função de perda, etc); o treinamento da rede propriamente dito; o cálculo de métricas de desempenho, como acurácia (ACC) e coeficiente de correlação de Matthews (MCC); e predição de resultados, gerando um cubo sísmico com as fraturas identificadas em cada nó da sísmica. A possibilidade de realimentação é indicada ao final do processo, sendo realizada para as predições que forem aprovadas. Nota-se que o processo pode seguir múltiplas vezes, aumentando a complexidade da rede.

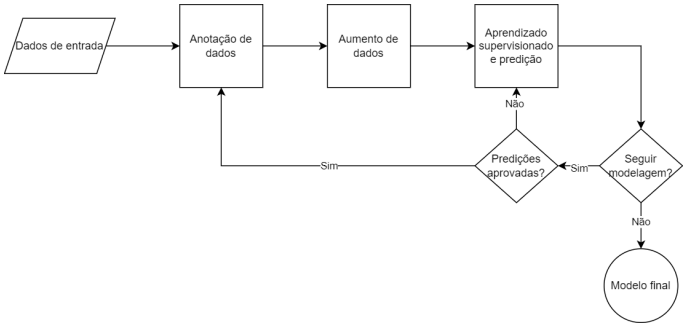

Figura 1 - Fluxograma da metodologia adotada.

Resultados

A metodologia permite diferentes configurações para abordar o problema, que foram avaliadas pela definição de diferentes cenários de treinamento, com o uso das diferentes técnicas de aumento de dados, diferentes frequências de filtragem dos perfis P32, diferentes cenários de conjuntos de testes e diferentes arquiteturas de redes. Esses cenários referem-se à combinação de cada um dos casos descritos a seguir. Ressalta-se que arquiteturas de rede e dimensões de cubos menores são necessárias, para não gerar sobreajuste, devido à baixa quantidade de dados anotados disponíveis. Dessa forma, inicialmente não se aplicam camadas de convolução as redes, visto estas camadas possuírem muitos parâmetros livres para essa aplicação. Apenas duas arquiteturas, referentes a redes totalmente conectadas, são analisadas inicialmente: 4x4x2 e 8x4x2, em que cada valor se refere ao número de neurônios na camada em questão. A utilização de camadas de convolução será verificada quando dados suficientes estiverem disponíveis, via etapas de realimentação/aumento de dados.

- Cenários de aumento de dados:
  - I: Dados originais
  - II: Dados originais e rotacionados
  - III: Dados originais e invertidos
  - IV: Dados originais e modificados pela multiplicação da amplitude por uma constante
- Percentuais de frequência de corte:
  - percf06: abrange 60% do espectro de frequência

  - percf08: abrange 80% do espectro
  - percf1: sem filtragem
- Formas de validação:
  - A: 3 poços para treinamento e 10 para teste
  - B: 8 para treinamento e 5 para teste
  - C: 10 para treinamento e 3 para teste
  - D: 80% das amostras para treino e 20% para teste
- Duas arquiteturas de rede totalmente conectadas:
  - 4x4x2: camadas com 4, 4 e 2 neurônios, respectivamente
  - 8x4x2: camadas com 8, 4 e 2 neurônios, respectivamente

Os resultados dos diferentes cenários de conjuntos de testes (Figura 2) apresentam medianas de MCC próximas a zero e amplitudes que atingem valores negativos, inclusive. Os resultados indicam que essas primeiras redes não são capazes de realizar predições confiáveis para o volume sísmico como um todo (casos A, B e C). No entanto, as predições realizadas nas imediações dos dados de treinamento podem ser úteis (caso D). Esse resultado é interessante principalmente do ponto de vista da realimentação.

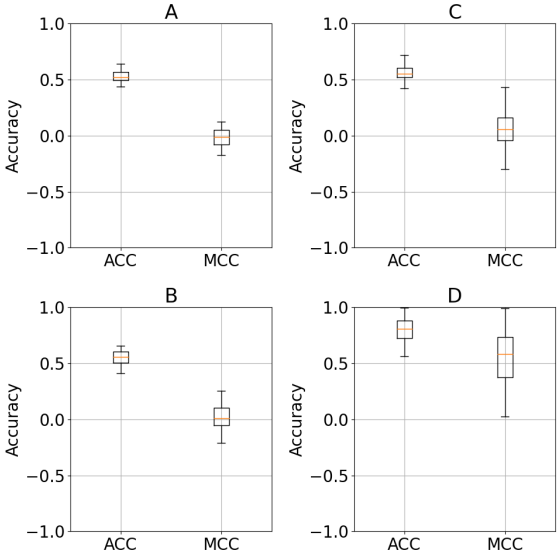

Figura 2 - Resultados obtidos para cada um dos cenários de formas de validação, onde ACC é a acurácia e MCC o coeficiente de correlação de Matthews.

A Figura 3 ilustra os resultados obtidos em função das técnicas de aumento de dados empregadas. Os cenários testados mostram que elas trazem pouca informação útil às previsões além da informação contida nos dados originais (caso I). O caso de aumento de dados por brilho (caso IV) apresenta resultados que parecem melhores, mas correspondem mais a um artefato, ligado ao fato de usar dados modificados nos conjuntos de treinamento e teste, que a um real efeito positivo do aumento de dados.

Na Figura 4 nota-se que o percentual para seleção da frequência de filtragem nos resultados do aprendizado possui um impacto grande, principalmente ao se comparar o caso sem filtragem (percf1) com os demais (percf06 e percf08). Nessa comparação, percebe-se que o limite inferior do MCC dos dois últimos casos supera o limite superior do caso sem filtragem.

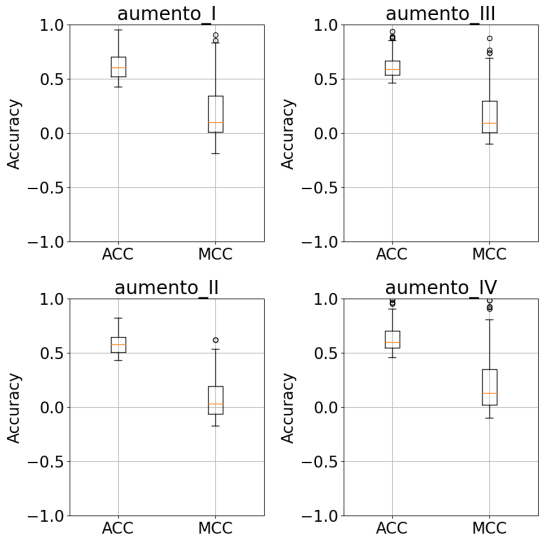

Figura 3 - Resultados obtidos para cada um dos cenários de aumento de dados.

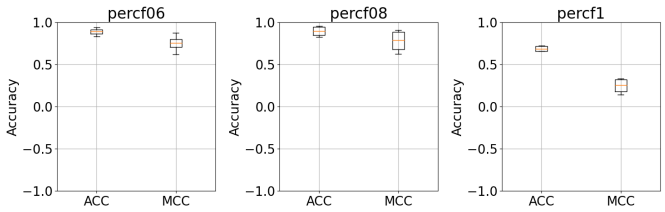

Figura 4 - Resultados dos cenários de percentual de frequência de filtragem.

A Figura 5 apresenta as predições resultantes para cenários abrangendo 3 diferentes percentuais de frequência, de modo a ilustrar o tipo de resposta da metodologia. Nota-se claramente como a definição da frequência de corte para filtragem tem uma influência determinante nos resultados reportados.

Para uma primeira avaliação do potencial da realimentação, utiliza-se um procedimento para validação dos dados baseado no valor de predição associado a

cada ponto. Esses valores são contínuos, entre 0 e 1. Optou-se por usar como base predições em 4 cenários diferentes (percf06, percf085, percf095 e percf099), ao redor dos poços usados para treinamento. Foram selecionadas apenas predições em que todos os modelos concordem, considerando um limiar de 0.02. A Figura 6 ilustra esse processo de seleção, com o dado filtrado indicando os pontos efetivamente aprovados para realimentação. Nota-se que, mesmo que uma pequena proporção de dados atende aos critérios estabelecidos, eles representam 30 vezes mais dados do que o anotado inicialmente, permitindo o uso de redes maiores.

foram capazes de extrair padrões relevantes para o problema, porém sem atingir ainda resultados que justifiquem sua aplicação prática imediata.

Naturalmente, o processo de seleção de dados para realimentação funciona como um exemplo de como o processo é conduzido. Para melhores resultados, é importante que especialistas avaliem e validem os dados a entrarem no processo de realimentação.

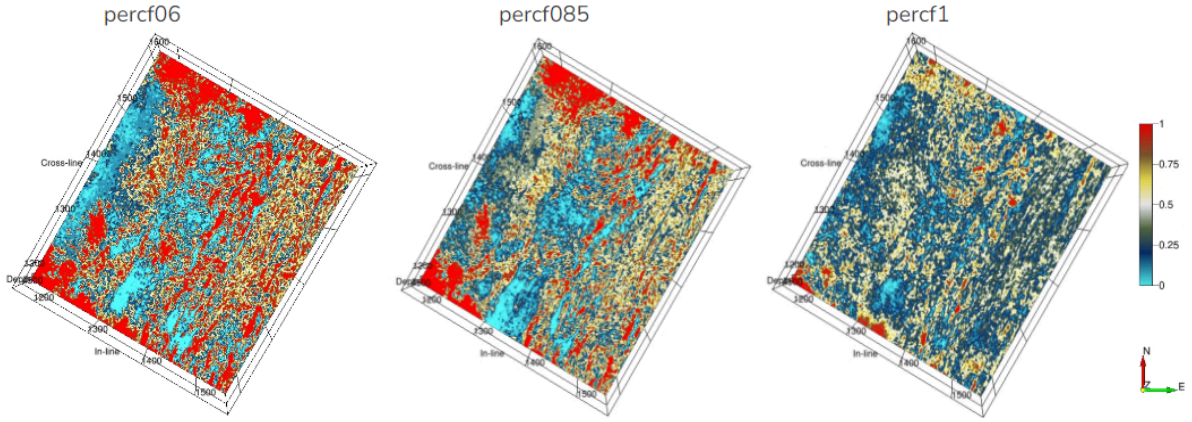

Figura 5 - Predições realizadas para modelos treinados com a arquitetura 6x2, nos dados de treinamento do cenário D, usando apenas dados originais (sem aumento de dados), para frequências associadas a 3 percentuais diferentes (0.6, 0.85 e 1).

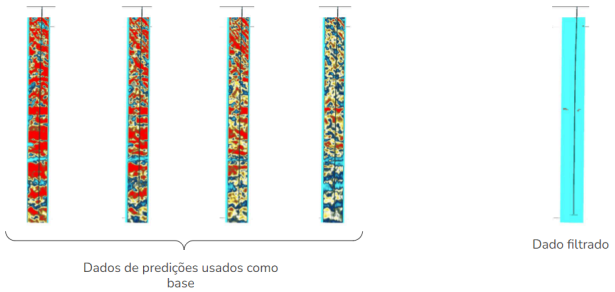

Figura 6 - Diferentes predições ao redor de um dos poços de treinamento e o dado filtrado para ser usado na realimentação. Apenas poucos pontos atendem ao critério adotado (dado filtrado).

Os resultados do treinamento com base em redes treinadas com esses dados encontram-se na Tabela 1. Nesta tabela, apresentam-se as arquiteturas de rede testadas, usando o mesmo padrão de definição citado no primeiro teste, bem como as mesmas métricas de qualidade empregadas anteriormente. Nota-se como a complexidade das arquiteturas das redes utilizadas, medidas pelo número total de neurônios presente, é muito maior que as originais, devido ao maior número de dados anotados permitidos pela realimentação. Já os coeficientes de qualidade indicam que ambos os modelos

Tabela 1 - Resultados para a rede treinada a partir da realimentação.

Arquitetura	ACC	MCC
20x12x6x4x2	0.64	0.40
10x8x4x2	0.62	0.36

Discussão e Conclusões

O objetivo do presente trabalho foi a construção de uma metodologia para predição de fraturas usando aprendizado de máquina e dados sísmicos pré-empilhamento com azimute completo.

As análises realizadas indicam que as redes de arquitetura simples (4x4x2 e 8x4x2), treinadas utilizando apenas os dados anotados disponíveis para treinamento, não são capazes de realizar predições confiáveis para o volume sísmico como um todo. Uma grande taxa de fraturas em regiões em que essa ocorrência não era esperada foram preditas, além de uma correlação muito grande das predições com os horizontes sísmicos. Por outro lado, as métricas de qualidade obtidas para predições referentes a dados nas imediações daqueles utilizados para treinamento apresentam valores promissores. Assim, a metodologia está no caminho

certo, porém ainda precisa ser mais bem trabalhada para permitir seu uso em situações práticas.

Percebe-se também que o processamento do dado anotado tem um impacto significativo na resposta da aplicação, principalmente no que diz respeito à seleção de frequência de corte para filtragem. Já o aumento de dados não mostrou influência significativa na qualidade das redes nesta etapa. O efeito do aumento de dados deve ser verificado novamente com as redes maiores caso opte-se por realimentar a rede.

O processo de realimentação de uma rede com dados previstos nas proximidades dos poços de treinamento se mostra como uma opção promissora para melhorar o modelo. Um primeiro estudo desse processo foi realizado, indicando que redes mais complexas podem ser empregadas. Esse procedimento pode tornar possível a aplicação, inclusive, de camadas convolucionais. No entanto, mais estudos com relação ao processo de validação dos dados para realimentação é necessário, principalmente a partir do ponto de vista de um especialista no tema. Além disso, é importante seguir-se trabalhando com a metodologia desenvolvida e realizando novas análises para que resultados com relevância prática sejam obtidos.

Agradecimentos

Agradecemos à Petrobras pelo financiamento, pela oportunidade de trabalhar com dados de grande qualidade, pelo apoio técnico valioso e pela colaboração durante todo o processo de realização deste estudo. Também agradecemos à Universidade Federal do Rio de Janeiro (UFRJ) pelo apoio técnico fornecido. Por fim, agradecemos ao Serviço Brasileiro de Apoio às Micro e Pequenas Empresas (Sebrae) pelo acompanhamento empresarial durante todo o projeto.

Referências

Babasafari, A. A., Chinelatto, G. F., & Vidal, A. C. (2022). Fault and fracture study by incorporating borehole image logs and supervised neural network applied to the 3D seismic attributes: a case study of pre-salt carbonate reservoir, Santos Basin, Brazil. Petroleum Science and Technology, 40(12), 1492–1511.

Cheng, Z., Bian, L., Chen, H., Wang, X., Ye, D., & He, L. (2022). Multiscale fracture prediction technique via deep learning, seismic gradient disorder, and aberrance: Applied to tight sandstone reservoirs in the Hutubi block, southern Junggar Basin. Interpretation, 10(4), T647–T663.

Cheng, S., Cheng, Z., Jiang, C., Mao, W., & Zhang, Q. (2023). An effective self-supervised learning method for various seismic noise attenuation. arXiv (Cornell University).

Clark, R. A.; Benson, P. M.; Carter, A. J. & Moreno, C. A. G. (2009). Anisotropic P-wave attenuation measured from a multi-azimuth surface seismic reflection survey. Geophysical Prospecting, 57 (5): 835-845. doi: 10.1111/j.1365-2478.2008.00772.x

Dershowitz, W. S. & Herda, H. H. (1992). Interpretation of fracture spacing and intensity. Proceedings of the 33rd US Symposium on Rock Mechanics, Santa Fe, 3-5 June 1992.

Fang, X., Zheng, Y., & Fehler, M. C. (2017). Fracture clustering effect on amplitude variation with offset and azimuth analyses. Geophysics, 82(1), N13-N25.

Hall, S. A., & Kendall, J. M. (2003). Fracture characterization at Valhall: Application of P-wave amplitude variation with offset and azimuth (AVOA) analysis to a 3D ocean-bottom data set. Geophysics, 68(4), 1150-1160.

Harwood, S.; Rowbotham, P. S.; Bush, I.; Williamson, P. R.; Houllevigue, H. & Kendall, J. M. (1998). Uncertainties in the analysis of AVOA for determination of fracture orientation. Revue de l'Institut Français du Pétrole, 53 (5): 621-627.

Sheng, H., Wu, X., Si, X., Li, J., Zhang, S., & Duan, X. (2023). Seismic Foundation Model (SFM): a new generation deep learning model in geophysics. arXiv (Cornell University).

Wang, Q., Shen, Y., Fu, L., & Li, H. (2020). Seismic data interpolation using deep internal learning. Exploration Geophysics (Melbourne, Vic. Online)/Exploration Geophysics, 51(6), 683–697

Wu, X., Liang, L., Shi, Y., Fomel, S., 2019. FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. Geophysics, 84(3), pp. IM35-IM45.

Yuan, Z., Jiang, Y., Li, J., & Huang, H. (2021). A convolutional neural network for prestack fracture detection. arXiv (Cornell University).

Zheng, Y., Zhang, Q., Yusifov, A., & Shi, Y., 2019. Applications of supervised deep learning for seismic interpretation and inversion. The Leading Edge, 38(7), 526-533.