GabrielNet: Segmentação de Fácies Sísmicas com Arquitetura Multi-Decoder e Ensemble de Redes Neurais Profundas

Juan Perri[1], **Lorena M. B.**[1], **C. Miceli Farias**[1], **Arick Jurdan**[1]

[1] Departamento de Ciência da Computação – UFRJ Analytica
Universidade Federal do Rio de Janeiro (UFRJ)
Rio de Janeiro – RJ – Brasil

juanperri@ufrj.br lorenamb@cos.ufrj.br
cmicelifarias@cos.ufrj.br arickjurdan.20221@poli.ufrj.br

Resumo. *A segmentação de fácies sísmicas, entendidas como padrões de reflexão que representam diferentes unidades geológicas e características deposicionais do subsolo, enfrenta desafios significativos devido ao ruído de aquisição, à escassez de rótulos e ao severo desbalanceamento entre classes. Este artigo propõe o* ***GabrielNet****, uma arquitetura que mitiga esses problemas por meio de um ensemble multi-decoder composto pelos modelos Attention U-Net, U-Net e DeepLabV3+, conectados a um encoder convolucional compartilhado. O pipeline integra um pré-processamento físico-informado baseado em filtragem espectral via FFT e difusão anisotrópica de Perona-Malik, engenharia de atributos sísmicos utilizando a transformada de Hilbert e pós-processamento morfológico para redução de falsas predições. Avaliada sobre o volume sísmico Parihaka, a abordagem alcançou 84,24% de mean IoU e 86,50% de acurácia de pixel, demonstrando elevada coerência geológica e robustez em cenários com restrições computacionais e forte desbalanceamento de classes.*

Abstract. *Seismic facies segmentation, where facies correspond to reflection patterns associated with distinct geological units and depositional characteristics, is challenged by acquisition noise, label scarcity, and severe class imbalance. This paper presents* ***GabrielNet****, a unified architecture composed of a shared convolutional encoder and a multi-decoder ensemble formed by Attention U-Net, U-Net, and DeepLabV3+. The proposed pipeline incorporates physics-informed preprocessing based on low-pass FFT filtering and Perona-Malik anisotropic diffusion, seismic attribute engineering using the Hilbert transform, and morphological post-processing to suppress spurious predictions. Experiments conducted on the Parihaka seismic volume achieved 84.24% mean IoU and 86.50% pixel accuracy, demonstrating high geological consistency and robustness under computational constraints and highly imbalanced conditions.*

1. Introdução

Fácies sísmicas são unidades sedimentares distinguidas pelos padrões de reflexão sísmica (amplitude, continuidade, frequência e geometria), o que permite tratar sua delimitação como uma tarefa de segmentação de imagens. A segmentação de imagens sísmicas é um desafio pois fácies adjacentes exibem texturas e amplitudes muito semelhantes, com fronteiras graduais em vez de bordas nítidas, e o sinal é corrompido por ruído de aquisição que mascara a estrutura dos refletores, fatores que tornam a interpretação manual lenta e dependente de especialistas. Para a resolução desse problema, usamos conceitos de *visão computacional* e *Deep Learning* [LeCun et al. 2015, Goodfellow et al. 2016]. Ainda assim, há desafios como o custo computacional relacionado a placas de vídeo, como a VRAM, e a forma como os dados sísmicos são obtidos, por serem muito ruidosos e desbalanceados [Gutierrez et al. 2025]. Soma-se a isso a escassez de rótulos, já que os

2. Trabalhos Relacionados

CNN e Vision Transformers em segmentação sísmica. A popularização dos *Vision Transformers* motivou tentativas de complementar o *backbone* convolucional com auto-atenção global, como o CSWin-UNet de Liu et al. [Liu et al. 2025], que introduz atenção em janelas cruzadas para ampliar o campo receptivo. Essa escolha, viável em imagens médicas adquiridas sob protocolos controlados, eleva o custo paramétrico e a dependência de pré-treino sem tratar a alta variância dos dados brutos. Em dados sísmicos, onde ruído de alta e baixa frequência corrompe a estrutura local dos refletores, transferir toda a filtragem para as camadas de atenção produz instabilidade no aprendizado [Gutierrez et al. 2025, Atolagbe and Koeshidayatullah 2025]. Nosso modelo inverte essa lógica: o pré-processamento físico-informado reduz a dimensionalidade efetiva antes da inferência, permitindo que três decodificadores convolucionais leves operem sobre representações já estruturadas, sem as penalidades de memória e dados das arquiteturas de atenção global.

Modelos de fundação para interpretação sísmica. Gao et al. [Gao et al. 2026] propõem o *Segment Any Geobodies* (SAG), um modelo de fundação baseado no *Segment Anything Model* (SAM) [Kirillov et al. 2023] e ajustado via Low-Rank Adaptation (LoRA) [Hu et al. 2022] sobre um conjunto, *multi-geobody* para interpretação universal de estruturas sísmicas. O SAG demonstra generalização *cross-survey* e elimina a necessidade de treinar modelos separados por tipo de *geobody*. Entretanto, apesar da eficiência paramétrica do LoRA (6,4 MB treináveis), o encoder congelado herdado do SAM-base totaliza 523 MB e opera sobre imagens de 1024×1024 pixels, impondo custo de inferência incompatível com restrições computacionais típicas da exploração sísmica. Além disso, o pré-processamento do SAG limita-se a normalização de amplitude, sem filtragem espectral ou tratamento de ruído. Logo, a carga de filtragem é inteiramente transferida para as camadas de atenção, estratégia que o próprio trabalho reconhece ser problemática em dados de baixa relação sinal-ruído. Por fim, o SAG não incorpora nenhuma etapa de pós-processamento morfológico, deixando agrupamentos espúrios de pixels. O GabrielNet contorna essas questões ao inverter essas escolhas. Com um *pipeline* físico-informado, o GabrielNet reduz a dimensionalidade do problema antes da rede, e filtros morfológicos suprimem artefatos nas predições, produzindo resultados coerentes com a topologia geofísica da área de estudo.

3. Proposta

O GabrielNet responde aos desafios apontados na Seção 1 com um *pipeline* em que cada etapa ataca um obstáculo específico do dado sísmico. O ruído de aquisição e a alta dimensionalidade são tratados antes da rede, por um pré-processamento físico-informado que combina filtragem espectral via FFT e difusão anisotrópica de Perona-Malik [Perona and Malik 1990], preservando as bordas dos refletores. A escassez de rótulos é mitigada pela engenharia de atributos derivados da transformada de Hilbert, que enriquece a entrada com informação física sem exigir anotações adicionais, e pelo viés indutivo de localidade das CNNs, adequado a regimes de poucos dados. O desbalanceamento de classes é endereçado pela arquitetura *multi-decoder* sobre *encoder* compartilhado, em que decodificadores com focos complementares são combinados por *soft*

voting na inferência, e por um pós-processamento morfológico que suprime predições espúrias. O conjunto opera sob custo de memória compatível com as restrições computacionais realistas da exploração sísmica, em contraste com as arquiteturas de atenção global discutidas na Seção 2. As subseções a seguir detalham cada etapa.

Filtragem e Pré-processamento de Dados. Para a etapa inicial de redução de ruído, aplicou-se a Transformada Rápida de Fourier (FFT), implementando um filtro passa-baixa no domínio da frequência. Em seguida, a transformada inversa foi utilizada para reconstruir a imagem espacial limpa. Esta etapa é fundamental, visto que Redes Neurais Convolucionais (CNNs) são altamente sensíveis a ruídos estocásticos. Adicionalmente, para lidar com ruídos de alta frequência sem comprometer as bordas estruturais das fácies, aplicou-se o filtro de difusão anisotrópica de Perona-Malik.

Engenharia de atributos sísmicos. Para enriquecer a representação de entrada sem ampliar o custo computacional do *encoder*, derivam-se dois atributos da transformada de Hilbert aplicada ao eixo de profundidade. A partir do sinal analítico $z(t) = u(t) + i\,\mathcal{H}\{u(t)\}$, extraem-se o envelope ($|z|$, normalizado por *z-score*) e o cosseno da fase instantânea ($\cos(\angle z)$). Sendo $|z|$ sensível à força de reflexão e $\cos(\angle z)$ à continuidade lateral. Esses atributos são empilhados com a amplitude pré-processada, produzindo um tensor de entrada $(H \times W \times 3)$ compatível com *encoders* pré-treinados na ImageNet.

Arquitetura multi-decoder. O GabrielNet adota uma estrutura *encoder-decoder* com ramificação tripla. Dado o tensor de entrada $\mathbf{X} \in \mathbb{R}^{H\times W\times 3}$ (amplitude pré-processada e os dois atributos de Hilbert), um *encoder* convolucional compartilhado E extrai mapas de salto $\{\mathbf{s}_\ell\}_{\ell=1}^{4}$ e a representação de *bottleneck* $\mathbf{b}$. Três decodificadores paralelos produzem, cada um, logits $\mathbf{Z}_i = D_i(\mathbf{b}, \{\mathbf{s}_\ell\}) \in \mathbb{R}^{H\times W\times C}$, cujas distribuições softmax $P_i = \mathrm{softmax}(\mathbf{Z}_i)$ são combinadas apenas na inferência por *soft voting*, isto é, pela média aritmética $P_{\text{final}} = \frac{1}{3}\sum_{i=1}^{3} P_i$, com a classe de cada pixel dada por $\hat{y} = \mathrm{argmax}_c\, P_{\text{final}}$. Em oposição a arquiteturas baseadas em atenção global, que dependem de grandes volumes de dados, os decodificadores convolucionais operam sobre representações já estruturadas pelo pré-processamento, explorando o viés indutivo de localidade adequado a regimes de poucos dados.

O ***encoder*** **compartilhado** E consiste em quatro blocos convolucionais com *Max Pooling* intercalado (32, 64, 128 e 256 canais) seguidos de um *bottleneck* de 512 canais. Cada bloco aplica duas sequências Conv2D(3×3) $\rightarrow$ BatchNorm $\rightarrow$ ReLU com inicialização Do encoder custom), e suas ativações alimentam, por conexões de salto, os três decodificadores. O **Decodificador 1 (Attention U-Net)** aplica *attention gates* em cada nível: o portão combina o sinal de salto $\mathbf{s}_\ell$ com o sinal ascendente $\mathbf{g}$, gerando um mapa $\alpha \in [0,1]^{H'\times W'}$ que suprime regiões irrelevantes antes da concatenação, favorecendo o realce das fácies minoritárias. O **Decodificador 2 (U-Net)** usa concatenação direta nas conexões de salto, fornecendo uma linha de base robusta e gradientes estáveis durante o treinamento conjunto. O **Decodificador 3 (DeepLabV3+)** aplica o bloco ASPP (*Atrous Spatial Pyramid Pooling*) ao *bottleneck* com taxas de dilatação $\{6, 12, 18\}$ e uma convolução 1×1, capturando contexto multiescala antes de quatro etapas de *upsampling* bilinear progressivo.

Pós-processamento Morfológico. Como refinamento final sobre a máscara de segmentação gerada pelo *ensemble*, aplicaram-se operações de morfologia matemática (como abertura e fechamento). Esta técnica clássica de processamento de imagens foi adotada estrategicamente para eliminar falsos positivos isolados (”falsas ilhas”) e suavizar as fronteiras das predições. Esta etapa de pós-filtragem mostrou-se determinante para mitigar o viés do modelo frente ao extremo desbalanceamento da Classe 4 no *dataset*, resultando em melhorias diretas na métrica de avaliação.

4. Experimentos

Dataset. Os dados provêm do volume sísmico 3D de domínio público *Parihaka*, disponibilizado pela *New Zealand Petroleum and Minerals* (NZPM) sob licença CC-BY-SA 4.0, com a rotulação de fácies realizada pela Chevron [Chevron U.S.A. Inc. 2020]. A partir desse volume, foram extraídos *patches* 2D de resolução 224×224 pixels em escala de cinza, armazenados em formato `.npy` com forma (N, H, W) para as imagens e máscaras, sobre $C = 6$ classes litológicas. A análise exploratória revelou dois desafios centrais. Primeiro, um severo desequilíbrio de classes: a classe majoritária representa mais de 60% dos voxels, enquanto a classe 4 (fácies de maior interesse geológico) ocupa menos de 1%, problema recorrente em interpretação de geocorpos [Gao et al. 2026]. Segundo, a análise espectral dos traços evidenciou concentração de energia entre 3 Hz e 35 Hz e dominância de componentes de baixa frequência, o que motivou o pré-processamento físico-informado descrito na Seção 3.

Métricas A separação de classes é avaliada pela métrica *Intersection over Union* (IoU), escolhida por ser amplamente adotada em detecção de objetos e por penalizar de forma rigorosa tanto falsos positivos quanto falsos negativos. A IoU quantifica a razão entre a área de interseção e a área de união das regiões predita e de referência para cada classe: valores próximos de 1 indicam alta sobreposição entre predição e rótulo, enquanto valores próximos de 0 indicam predições deslocadas ou ausentes; a média sobre as classes (*mean* IoU) resume o desempenho global. Em complemento, reportam-se acurácia de pixel, precisão e *recall* por classe.

Implementação O modelo foi implementado em TensorFlow/Keras treinado disponíveis em: `https://github.com/perrijuan/Hackathon-IA-para-Oleo-e-Gas`. e otimizado com Adam (taxa inicial $\eta = 10^{-3}$) e *batch* de 32 amostras, por até 30 épocas. A taxa de aprendizado é reduzida automaticamente por *ReduceLROnPlateau* (fator 0,2; paciência = 4), o melhor modelo é selecionado por *ModelCheckpoint* (mínimo de $\mathcal{L}_{\text{val}}$) e o treino é interrompido por *EarlyStopping* (paciência = 5). Para mitigar o desequilíbrio, um gerador balanceado garante 50% de *patches* com a classe minoritária em cada *batch*. Para reduzir o *overfitting*, aplica-se *data augmentation* em tempo real (espelhamento horizontal, ruído gaussiano $\sigma = 0{,}10$, translação e *zoom* de até $\pm 10\%$).

Resultados. Resultados publicados sobre o mesmo volume *Parihaka* corroboram a dependência de dados das arquiteturas baseadas em atenção global. [Sheng et al. 2025] reportam que um Transformer treinado do zero atinge apenas 48,58% de *mean* IoU e

78,22% de acurácia de pixel na classificação de fácies, abaixo inclusive da U-Net de referência (58,51% e 87,79%). O desempenho do Transformer só se torna competitivo após pré-treino auto-supervisionado em mais de 2,2 milhões de imagens sísmicas, com a melhor variante do *Seismic Foundation Model* alcançando 79,80% de *mean* IoU. Embora os protocolos de avaliação difiram (divisões de treino e validação distintas), o contraste é ilustrativo: o GabrielNet obtém **84,24% de *mean* IoU** e **86,50% de acurácia de pixel** sem qualquer pré-treino em larga escala, substituindo essa exigência pelo *pipeline* físico-informado e pelo viés indutivo convolucional, e retendo **86% de precisão na fácies minoritária**, o que indica maior confiabilidade interpretativa nas classes de maior interesse geológico.

5. Conclusão

Este trabalho apresentou o GabrielNet, um *Modelo* de segmentação de fácies sísmicas que combina pré-processamento físico-informado, engenharia de atributos derivados da transformada de Hilbert e um *ensemble multi-decoder* ancorado em um *encoder* convolucional compartilhado. A proposta atende às restrições computacionais realistas da exploração sísmica, se diferenciando das alternativas tradicionais ao privilegiar vieses indutivos convolucionais em detrimento de arquiteturas baseadas em atenção global, que demandam grandes volumes de dados anotados.

Para trabalhos futuros, sugerimos a avaliação da capacidade de generalização do GabrielNet em volumes de outras bacias e geologias, permitindo comparação com demais modelos de fundação [Gao et al. 2026]. Pretende-se ainda otimizar o *recall* da classe minoritária por meio de *weighted soft voting* (ponderando os decodificadores conforme seu desempenho por classe) e de limiar de decisão dinâmico para as fácies raras. A extensão do *pipeline* para inferência 3D, agregando predições 2D consecutivas, e a incorporação de restrições multimodais (e.g., dados de poço) constituem direções promissoras para ampliar a aplicabilidade do método em fluxos reais de interpretação sísmica.

Referências

Atolagbe, J. and Koeshidayatullah, A. (2025). Toward user-guided seismic facies interpretation with a pre-trained large vision model. *IEEE Access*, 13:42965–42976.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *European Conference on Computer Vision (ECCV)*, pages 801–818.

Chevron U.S.A. Inc. (2020). Labeled geological model of Parihaka seismic data for machine learning. SEG 2020 Annual Meeting Machine Learning Interpretation Workshop. Disponibilizado sob licença CC BY-SA 4.0; dados sísmicos base fornecidos pela New Zealand Petroleum and Minerals (NZPM).

Gao, H., Wu, X., Si, X., Sheng, H., Liu, M., and Wu, H. (2026). A foundation model empowered by a multi-modal prompt engine for universal seismic geobody interpretation across surveys. *Information Fusion*, 125:103437.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press, Cambridge, MA. `http://www.deeplearningbook.org`.

Gutierrez, G., Astudillo, C., Napoli, O., Miranda, D., Souza, A., Navarro, J. P., Villas, L., and Borin, E. (2025). On the performance evaluation of deep learning models for seismic facies segmentation. *Geophysical Prospecting*.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., and Chen, W. (2022). LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations (ICLR)*.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., and Girshick, R. (2023). Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4015–4026.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

Liu, X., Gao, P., Yu, T., Wang, F., and Yuan, R.-Y. (2025). CSWin-UNet: Transformer UNet with cross-shaped windows for medical image segmentation. *Information Fusion*, 113:102634.

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., and Rueckert, D. (2018). Attention U-Net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.

Perona, P. and Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pages 234–241. Springer.

Sheng, H., Wu, X., Si, X., Li, J., Zhang, S., and Duan, X. (2025). Seismic foundation model: A next generation deep-learning model in geophysics. *Geophysics*, 90(2):IM59–IM79.