

UNIVERSIDADE FEDERAL DO MARANHÃO - UFMA
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

HENRIQUE RIBEIRO DE MELLO

**SEGMENTAÇÃO DE ACUMULAÇÕES DE GÁS NATURAL EM
IMAGENS DE REFLEXÃO SÍSMICA ATRAVÉS DE TÉCNICAS 1D,
1.5D E 2D**

SÃO LUÍS - MA

2025

HENRIQUE RIBEIRO DE MELLO

**SEGMENTAÇÃO DE ACUMULAÇÕES DE GÁS NATURAL EM
IMAGENS DE REFLEXÃO SÍSMICA ATRAVÉS DE TÉCNICAS 1D,
1.5D E 2D**

Tese apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da UFMA como parte dos requisitos necessários para obtenção do grau de Doutor em Engenharia Elétrica.

Orientador: Prof. Dr. Anselmo Cardoso de Paiva

Coorientador: Prof. Dr. Francesco Renna

SÃO LUÍS MA

2025

Texto de Exame de Qualificação de autoria de Henrique Ribeiro de Mello, sob o ttulo “**Segmentação de acumulações de gás natural em imagens de reflexão sísmica através de técnicas 1D, 1.5D e 2D**”, apresentado ao Programa de Pós-Graduação em Engenharia Elétrica da UFMA, como parte dos requisitos para obtenção do ttulo de Doutor em Engenharia Elétrica, na área de concentração Ciência da Computação, aprovada em ____ de _____ de _____ pela comissão julgadora constituída pelos doutores:

Prof. Dr. Anselmo Cardoso de Paiva
Orientador

Prof. Dr. Francesco Renna
Coorientador

Prof. Dr. Aristófanés Correa Silva
Membro da Banca Examinadora

Prof. Dr. João Dallyson Sousa de Almeida
Membro da Banca Examinadora

Prof. Dr. Noel Pérez-Pérez
Membro da Banca Examinadora

Prof. Dr. Claudio Souza Baptista
Membro da Banca Examinadora

À minha família, companheira e amigos

ACKNOWLEDGEMENTS

Agradeço primeiramente aos meus pais Fernando e Eleonor, por todo o amor e carinho que me deram ao longo da vida e por sempre me incentivarem em tudo o que me propus a fazer ao longo de toda a minha vida. Sou grato aos amigos com quem convivi durante a minha graduação em Física, pelas horas jogando Mario Kart no centro acadêmico, pelos *coffee breaks* depois de colóquios, palestras e defesas, pelas risadas e pelas aventuras nos congressos. Aos meus amigos do NCA, que conheci apenas depois de um ano e meio de aulas e reuniões remotas por conta da pandemia, e que me fizeram lembrar como é participar de um grupo de pesquisa e dividir dúvidas, conselhos e pedaços de código. Ao meu orientador Anselmo por ter embarcado nessa jornada comigo e por seus conselhos e ideias que tanto me auxiliaram até aqui. Também agradeço ao meu coorientador Francesco por me receber tão bem no Porto durante o meu período de doutorado sanduíche e pelas orientações antes e depois da minha estadia em Portugal. Por fim, agradeço a minha namorada, esposa, e companheira Ana Elisa pelo amor, companheirismo, conversas filosóficas e viagens desde que nos conhecemos.

*“Quem sabe faz a hora
Não espera acontecer.”*

Geraldo Vandré

ABSTRACT

The most common indirect identification method used by the oil and natural gas industry in the search for natural gas accumulations is conducting a survey based on the principle of seismic reflection. Specialists look for sudden changes in signal intensity, which may indicate the presence of natural gas, oil, or other hydrocarbons. In this thesis, we propose three methods to automate this task using seismic reflection data: one method that uses an Ensemble of U-Nets, a well-established architecture in image segmentation, and two others that make use of the Robustly Optimized Bidirectional Encoder Representation from Transformers Pretraining Approach (RoBERTa), a method originally developed in the context of natural language modeling. The methods are developed in such a way that we simulate what happens in practice, where models are trained in a given region and are used to evaluate unknown areas. To do so, we propose a train-test split policy for the public dataset used that simulates the use of the methods in a new region. The U-Net based method requires the annotation to be considerably larger than the original and is able to determine the annotated region more precisely. The method based on the RoBERTa architecture shows more false positives in areas of high acoustic-impedance variation, but it is capable of identifying possible smaller accumulations, as well as delineating much of the structure that stores the natural gas. All three methods present segmentation evaluation metrics superior to the state-of-the-art when considering the model's generalization capability using a single dataset. On the F3 block dataset in the Netherlands, the methods achieved F1-scores between 0.36 and 0.62 on the proposed test set, while sensitivity varies between 0.65 and 0.79.

Palavras-chave: Seismic data, Oil and gas, Natural gas segmentation, Machine Learning, Deep Neural Networks, Time series, Natural Language Processing, Ensemble, Generalization.

RESUMO

O método mais comum de identificação indireta usado pela indústria de petróleo e gás natural na busca de acumulações de gás natural é a realização de um levantamento baseado no princípio da reflexão sísmica. Especialistas procuram por mudanças repentinas na intensidade dos sinais, o que pode indicar a presença de gás natural, petróleo ou outros hidrocarbonetos. Nesta tese, propomos três métodos para automatizar esta tarefa utilizando dados de reflexão sísmica: um método que utiliza um Ensemble de U-Nets, arquitetura consolidada na segmentação de imagens, e outro dois que fazem uso do Robustly Optimized Bidirectional Encoder Representation from Transformers Pretraining Approach (RoBERTa), método originalmente desenvolvido no contexto de modelagem de linguagem natural. Os métodos são desenvolvidos de modo que, para o dataset utilizado, seja feita uma simulação do que acontece na prática da exploração sísmica: modelos treinados em uma determinada região são utilizados para prever a ocorrência de gás natural e petróleo em outras ainda desconhecidas. Para isso, propomos uma política de divisão de treino-teste para o conjunto de dados público. O método baseado em U-Net requer que a anotação seja consideravelmente maior que a original, e consegue determinar mais precisamente a região anotada. O método baseado na arquitetura RoBERTa apresenta mais falsos positivos em regiões de alta variação de impedância acústica, porém é capaz de identificar possíveis acumulações menores, além de delimitar grande parte da estrutura que armazena o gás natural. Os três métodos apresentam métricas de avaliação da segmentação superiores ao estado-da-arte quando se considera a capacidade de generalização do modelo usando um único conjunto de dados. Nos dados do bloco F3, nos Países Baixos, os métodos apresentaram F1-score entre 0.36 e 0.62 no conjunto de teste proposto, enquanto que a sensibilidade varia entre 0.65 e 0.79.

Palavras-chave: Dados sísmicos, petróleo e gás, segmentação de gás natural, aprendizado de máquina, redes neurais profundas, séries temporais, modelos de linguagem natural, ensemble, generalização

LISTA DE ILUSTRAÇÕES

Figura 1 – Aquisição de dados sísmicos por reflexão (Exemplo <i>offshore</i>). Imagem de elaboração própria.	26
Figura 2 – Amplitude de deformações em função do tempo. Imagem adaptada de (DENTITH; MUDGE, 2014)	26
Figura 3 – Exemplo de uma aquisição 3d <i>offshore</i> . A imagem resultante é um cubo 3d abaixo da região de prospecção. Imagem retirada de (CAMESELLE, 2010).	27
Figura 4 – Navio que faz a aquisição sísmica. Imagem retirada de https://www.tgs.com/seismic/marine-acquisition/the-fleet/vessels/ramform-atlas . Acesso em 04/11/2024	27
Figura 5 – Empilhamento de ponto comum. Imagem retirada de https://germangarabito.wordpress.com/2015/04/05/common-point-stack/ . Acesso em 03/04/2025.	29
Figura 6 – Dados após migração temporal, já se comportando como uma fotografia ao mostrar a forma aparente das estruturas geológicas em subsuperfície. Para a visualização mais exata das estruturas, é necessária uma migração espacial. Imagem retirada de https://wiki.aapg.org/index.php?title=Seismic_migration . Acesso em 13/07/2025.	29
Figura 7 – Imagem gerada pela aquisição sísmica de reflexão, que pode ser interpretada como uma seção transversal da subsuperfície. Imagem retirada de (SILVA et al., 2019b).	30
Figura 8 – Tipos comuns de falhas. Imagem adaptada de (ROBINSON; TREITEL, 2000).	31
Figura 9 – Localização de brightspots em uma imagem de reflexão sísmica em escala de cinza. Imagem de elaboração própria.	31
Figura 10 – Exemplo de arquitetura de uma CNN. Imagem adaptada de https://medium.com/codex/convolutional-neural-networks-a-beginners-journey-into-the-architecture . Acesso em 13/07/2025.	34
Figura 11 – Arquitetura U-Net. Elaboração própria.	37
Figura 12 – A arquitetura Transformer. Imagem retirada de (VASWANI et al., 2017)	38

Figura 13 – À esquerda, o mecanismo de atenção de produto escalar com escala. À direita, a camada de atenção multi-cabeça, que consiste de várias camadas de atenção em paralelo. Imagem retirada de (VASWANI et al., 2017)	39
Figura 14 – Arquitetura BERT. Imagem retirada de https://wikidocs.net/224181 . Acesso em 13/07/2025.	41
Figura 15 – Tarefa de modelagem de linguagem mascarada. Imagem retirada de https://wikidocs.net/224181 . Acesso em 13/07/2025.	42
Figura 16 – Tarefa de previsão de próxima sentença. Imagem retirada de https://wikidocs.net/224181 . Acesso em 13/07/2025.	42
Figura 17 – Diagrama do algoritmo Wordpiece	44
Figura 18 – Valor da função de custo (loss) pela probabilidade da classe de referência (ground truth). Gráfico retirado de (LIN et al., 2018)	45
Figura 19 – Wavelets pertencentes à família Daubechies, cuja denominação dbN refere-se à ordem N da wavelet da família Daubechies (db).	47
Figura 20 – Na figura à esquerda, a imagem original. No centro, após a aplicação da operação de dilatação na imagem original. À direita, após a aplicação da operação de erosão.	49
Figura 21 – Operação morfológica de abertura.	49
Figura 22 – Operação morfológica de fechamento.	50
Figura 23 – Operações morfológicas de abertura e fechamento aplicadas à segmentação de gás natural.	51
Figura 24 – Especificação de histograma de uma imagem sísmica de reflexão usando outra imagem sísmica como referência.	52
Figura 25 – Matriz de confusão	53
Figura 26 – Bloco F3 e sua localização geográfica. Imagem retirada de (SILVA et al., 2019a).	64
Figura 27 – Pipeline da segmentação com a arquitetura U-Net. Imagem de elaboração própria.	65
Figura 28 – Processo de escolha de ensemble. Imagem de elaboração própria. . . .	67
Figura 29 – Quantização de traços sísmicos.	70
Figura 30 – Esquema de pré-processamento de um sinal sísmico no método proposto.	70

Figura 31 – Processo de pré-treinamento. Cada t-frase tokenizada possui alguns de seus tokens substituídos por tokens de máscara, representados por tokens -1, e o modelo aprende a prever qual o número do token escondido (mascarado). Esse procedimento faz o modelo aprender a relação entre tokens diferentes antes da tarefa de classificação realizada no treino supervisionado da rede.	71
Figura 32 – Processo de anotação das frases. Imagem de elaboração própria. . . .	72
Figura 33 – Alargamento efetivo das anotações usadas para treinar o modelo. . .	72
Figura 34 – Segmentação de uma imagem sísmica feita durante a etapa de inferência.	73
Figura 35 – Contexto usado pelo método multi-traços. Os traços em azul sobre a imagem sísmica possuem espessura exagerada para melhor visualização	74
Figura 36 – Ilustração de um t-parágrafo. $N + 1$ t-frases tokenizadas lado a lado (no exemplo, $N = 4$), podem ser vistos como um parágrafo formado por $N + 1 = 5$ frases diferentes.	74
Figura 37 – Pipeline da inferência com o método multi-traços.	75
Figura 38 – Exemplo de traços utilizados na inferência nas bordas com o método multi-traços. Em laranja, o traço a ser segmentado e em vermelho os traços vizinhos considerados pelo método.	76
Figura 39 – Vista axial do volume, ilustrando as duas acumulações de gás presentes, uma a norte e outra a sul, e a política aleatória de divisão de fatias entre treino e teste.	79
Figura 40 – Vista axial do volume, ilustrando as duas acumulações de gás presentes, uma a norte e outra a sul, e a política não-aleatória de divisão de fatias entre treino e teste.	80
Figura 41 – Comparação entre a segmentação de U-Nets usando as políticas aleatória e não-aleatória.	85
Figura 42 – Segmentação da Inline 574 das U-Nets que compõe um ensemble. A imagem de baixo ilustra a inferência do ensemble para uma mesma inline.	85
Figura 43 – Segmentação da inline 574 realizada por U-Nets individuais que constituem um ensemble. A imagem inferior constitui a segmentação da mesma inline feita pelo ensemble, cujas imagens passaram apenas pelo pré-processamento com a transformada wavelet.	87

Figura 44 – Segmentação da Inline 574 realizadas por U-Nets individuais que constituem um ensemble. A imagem inferior constitui a segmentação da mesma inline feita pelo ensemble, cujas imagens passaram pelo pré-processamento de DWT e quantização.	87
Figura 45 – O gráfico no topo à esquerda mostra como a fatia axial vista na figura inferior é feita. com aproximadamente dois quilômetros na direção das inlines, das inlines 558 a 635. A figura no topo à direita mostra melhor, no gráfico 3D, a profundidade considerada na criação da figura inferior. A figura inferior mostra a visão axial do bloco F3 depois das previsões do ensemble de U-Nets no conjunto de teste.	88
Figura 46 – Comparação entre a camada de ativação 3 da parte codificadora (encoder) de duas U-Nets em experimentos distintos. A segmentação final de cada um dos modelos, ao redor da área anotada, também é mostrada.	89
Figura 47 – Comparação entre a camada de ativação 8 da parte codificadora (encoder) de duas U-Nets em experimentos distintos. A segmentação final de cada um dos modelos, ao redor da área anotada, também é mostrada.	89
Figura 48 – Comparação entre a camada de ativação 22 da parte decodificadora (decoder) de duas U-Nets em experimentos distintos. A segmentação final de cada um dos modelos, ao redor da área anotada, também é mostrada.	90
Figura 49 – Sequência de inlines 55, 79, 115 e 581 na política aleatória sem operações morfológicas. Todas as imagens possuem um intervalo de TWT entre 400 ms e 800 ms.	92
Figura 50 – Detalhe das inlines 38 e 55 que mostram uma perda de resolução na área onde existe falha na segmentação.	92
Figura 51 – Sequência de inlines 55, 79, 115 e 581 na política aleatória após a aplicação das operações morfológicas. Todas as imagens possuem um intervalo de TWT entre 400 ms e 800 ms.	94
Figura 52 – Segmentação completa da inline 38.	94
Figura 53 – Segmentação completa da inline 55.	95
Figura 54 – Segmentação da inline 558 feita pelo modelo treinado, antes das operações morfológicas.	96

Figura 55 – Segmentação das inlines 587, 588, 589 e 590 (de cima para baixo). A última imagem é a inline 588 sem a segmentação para uma melhor visualização dos brightspots. Os falsos positivos na região à direita aparecem aproximadamente nos mesmos locais.	97
Figura 56 – Inlines 558, 565, 582 e 585 na política não-aleatória após aplicação de operações morfológicas. Todas as imagens apresentam um corte no eixo TWT entre 400 ms e 800 ms.	98
Figura 57 – Segmentação da inline 558 feita pelo modelo treinado, depois das operações morfológicas.	99
Figura 58 – O gráfico no topo à esquerda mostra como a fatia axial vista na figura inferior é feita com aproximadamente dois quilômetros na direção das inlines, das inlines 558 a 635. A figura no topo à direita mostra melhor, no gráfico 3D, a profundidade considerada na criação da figura inferior. A figura inferior mostra a visão axial do bloco F3 depois das previsões no conjunto de teste.	100
Figura 59 – Comparação da segmentação de duas imagens antes e depois da aplicação da técnica de especificação de histograma.	101
Figura 60 – Segmentação com e sem OM. Há um aumento significativo no número de pixels segmentados corretamente (verdadeiros positivos) depois da aplicação das OM, vistos em azul escuro na figura.	104
Figura 61 – Similaridade entre a geologia das regiões nas acumulações norte e sul e sua localização na visão axial do volume de dados. Em retângulos azuis, as inlines 150 (sul) e 558 (norte) estão evidenciadas.	104
Figura 62 – Segmentação das inline 50 e 66 com o modelo 1.5D.	106
Figura 63 – Segmentação das inline 565 e 574 com o modelo 1.5D.	109
Figura 64 – Inlines 558, 565, 582 e 585 na política não-aleatória para o método multi-traços. Todas as imagens apresentam um corte no eixo TWT entre 400 ms e 800 ms.	110
Figura 65 – Comparação entre a segmentação da inline 574 com o modelo 1D e 1.5D.	111
Figura 66 – Comparação entre a segmentação da inline 567 com o modelo 1D e 1.5D.	111

Figura 67 – Comparação entre a segmentação da inline 597 com o modelo 1D e 1.5D.	112
Figura 68 – Segmentação com e sem OM. Existe um aumento na sensibilidade do modelo ao segmentar as duas inlines, porém menor do que no método 2.	113
Figura 69 – Os dois gráficos no topo mostram a segmentação em três dimensões das inlines do reservatório de gás ao norte. A figura inferior mostra a visão axial do bloco F3 com as previsões do método baseado no RoBERTa multi-traços.	114
Figura 70 – Segmentação detalhada ao redor da região de gás natural nas inlines 55, 79, 115 e 581 na política aleatória. A U-Net P e o Método 1 (Ensemble U-Net) são os mais precisos na segmentação de inlines que pertencem à maior acumulação de gás (sul), enquanto que a LSTM e os Método 2 e 3 apresentam falsos positivos ao redor da anotação.	117
Figura 71 – Segmentação detalhada ao redor da região de gás natural nas inlines 558, 565, 582 e 585 na política não-aleatória. Os métodos propostos mostram desempenho melhor na segmentação das inlines mostradas aqui quando a política de divisão treino-teste usada é a política não-aleatória.	119

LISTA DE TABELAS

Tabela 1 – Referências e respectivos métodos dos trabalhos relacionados a este projeto.	62
Tabela 2 – Correspondência entre termos baseados no domínio de linguagens e suas definições originais no domínio sísmico.	69
Tabela 3 – Tempos de inferência e treinamento de cada modelo.	78
Tabela 4 – Parâmetros da arquitetura U-Net.	81
Tabela 5 – Parâmetros para o treinamento da U-Net.	81
Tabela 6 – Métricas de segmentação de treino e teste na política aleatória. . . .	82
Tabela 7 – Métricas de segmentação dos conjuntos de treino e teste usando a política não-aleatória.	83
Tabela 8 – Métricas de segmentação dos conjuntos de treino e teste na política aleatória usando quantização, DWT e ensembling.	83
Tabela 9 – Métricas de segmentação dos conjuntos de treino e teste na política não-aleatória usando quantização, DWT e ensembling.	84
Tabela 10 – Médias das métricas de segmentação tomadas sobre cinco experimentos para cada combinação de métodos no conjunto de testes da política não-aleatória.	86
Tabela 11 – Parâmetros para o pré-treinamento na tarefa de MLM.	90
Tabela 12 – Parâmetros para o treinamento na tarefa de classificação de t-frases.	91
Tabela 13 – Valores médios das métricas de avaliação com seus respectivos erros padrões para a política aleatória sem a aplicação de operações morfológicas no pós-processamento, avaliados pixel a pixel antes da aplicação de operações morfológicas.	91
Tabela 14 – Valores médios das métricas de avaliação com seus respectivos erros padrões para a política aleatória com a aplicação de operações morfológicas no pós-processamento, avaliados pixel a pixel antes da aplicação de operações morfológicas.	93
Tabela 15 – Valores médios das métricas de avaliação com seus respectivos erros padrões para a política não-aleatória sem a aplicação de operações morfológicas no pós-processamento, avaliados pixel a pixel antes da aplicação de operações morfológicas.	95

Tabela 16 – Valores médios das métricas de avaliação com seus respectivos erros padrões para a política não-aleatória com a aplicação de operações morfológicas no pós-processamento, avaliados pixel a pixel antes da aplicação de operações morfológicas.	98
Tabela 17 – Comparação das métricas de inlines selecionadas na política aleatória antes e depois das operações morfológicas (OM). Os melhores resultados para cada inline em cada métrica estão em negrito.	103
Tabela 18 – Comparação das métricas de inlines selecionadas na política não-aleatória antes e depois das operações morfológicas (OM). Os melhores resultados para cada inline em cada métrica estão em negrito.	103
Tabela 19 – Parâmetros para o treinamento.	105
Tabela 20 – Valores médios das métricas de avaliação com seus respectivos erros padrões para a política aleatória sem a aplicação de operações morfológicas no pós-processamento, avaliados pixel a pixel antes da aplicação de operações morfológicas.	106
Tabela 21 – Valores médios das métricas de avaliação com seus respectivos erros padrões para a política aleatória com a aplicação de operações morfológicas no pós-processamento, avaliados pixel a pixel antes da aplicação de operações morfológicas.	107
Tabela 22 – Valores médios das métricas de avaliação com seus respectivos erros padrões para a política não-aleatória sem a aplicação de operações morfológicas no pós-processamento, avaliados pixel a pixel antes da aplicação de operações morfológicas.	108
Tabela 23 – Valores médios das métricas de avaliação com seus respectivos erros padrões para a política não-aleatória com a aplicação de operações morfológicas no pós-processamento, avaliados pixel a pixel antes da aplicação de operações morfológicas.	108
Tabela 24 – Comparação das métricas de inlines selecionadas na política não-aleatória antes e depois das operações morfológicas (OM). Os melhores resultados para cada inline em cada métrica estão em negrito.	113

Tabela 25 – Comparação de métricas entre os métodos propostos e as redes neurais de referência LSTM e U-Net P, para ambas as políticas aleatória e não-aleatória. Os valores em negrito mostram os melhores resultados dos métodos cuja anotação estendida foi feita de modo similar (todos os métodos com exceção do Ensemble de U-Nets).	115
Tabela 26 – Métricas para o Método 1. Por ser necessária a utilização de um alargamento maior da área considerada como sendo gás, e as métricas dependem da marcação utilizada, optou-se por separar estes valores dos anteriores, que usam um alargamento menor das anotações fornecidas pelo especialista.	115
Tabela 27 – Métricas de segmentação das inlines da Figura 70	116
Tabela 28 – Métricas da segmentação das inlines mostradas na Figura 71.	118
Tabela 29 – Referências e respectivos métodos dos trabalhos relacionados a este projeto.	124

LIST OF ABBREVIATIONS AND ACRONYMS

BERT	<i>Bidirectional Encoder Representation from Transformers</i>
FN	Falso Negativo
FP	Falso Positivo
IDH	Indicadores Diretos de Hidrocarbonetos
MLM	<i>Masked Language Modeling</i>
NSP	<i>Next Sentence Prediction</i>
OM	Operações Morfológicas
P&G	Petróleo e Gás
PLN	Processamento de Linguagem Natural
p.p.	Pontos percentuais
PRC	Precisão
px	Pixel
ROI	Região de Interesse
SEN	Sensibilidade
STD	Desvio Padrão
TWT	Two-Way Traveltime
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo

SUMÁRIO

1	INTRODUÇÃO	21
1.1	Objetivos	22
1.2	Contribuições	23
1.3	Organização do trabalho	23
2	FUNDAMENTOS TEÓRICOS	25
2.1	Aquisição e interpretação de dados sísmicos	25
2.1.1	Tratamento dos dados	28
2.1.2	Interpretação	30
2.2	Segmentação de imagens	32
2.3	Redes neurais convolucionais	33
2.3.1	U-Net	36
2.4	Transformer	37
2.5	Modelos de linguagem	39
2.5.1	Bidirectional Encoder Representation from Transformers - BERT	40
2.5.2	Tokenização WordPiece	43
2.5.3	Funções de custo	44
2.5.4	Entropia cruzada binária	44
2.5.5	Focal	45
2.5.6	Dice	46
2.6	Técnicas de pré e pós processamento	46
2.6.1	Transformada Wavelet	46
2.6.2	Operações morfológicas	48
2.6.3	Especificação de histograma	51
2.7	Métricas de Avaliação	52
2.8	Cálculo de erros experimentais	54
2.9	Considerações Finais	54
3	TRABALHOS RELACIONADOS	56
3.0.1	Aprendizado profundo para segmentação e detecção de gás natural	56

3.0.2	Redes neurais profundas aplicadas em dados sísmicos	58
3.0.3	Considerações finais	60
4	MATERIAIS E MÉTODOS	63
4.1	Base de dados	63
4.2	Método 1: Abordagem 2D - Ensemble de U-Net para segmen- tação de imagens	64
4.2.1	Pré-processamento e treinamento	65
4.2.2	Inferência	65
4.3	Método 2: Abordagem 1D - RoBERTa para classificação de traços	66
4.3.1	Pré-processamento	67
4.3.2	Pré-treinamento	69
4.3.3	Treinamento	71
4.3.4	Inferência	72
4.4	Método 3: Abordagem 1.5D - RoBERTa Multi-traços	73
4.5	Considerações Finais	76
5	RESULTADOS E DISCUSSÃO	78
5.1	Método 1: Abordagem 2D - Ensemble de U-Nets	80
5.1.1	Pré-processamento e treinamento	80
5.1.2	Experimento 1: U-Net na política aleatória	82
5.1.3	Experimento 2: U-Net na política não-aleatória	82
5.1.4	Comparação entre diferentes combinações de técnicas	85
5.1.5	Análise dos mapas de características	88
5.2	Método 2: Abordagem 1D - RoBERTa	90
5.2.1	Experimento 3: Política aleatória sem operações morfológicas	91
5.2.2	Experimento 4: Política aleatória com operações morfológicas	92
5.2.3	Experimento 5: Política não-aleatória sem operações morfo- lógicas	95
5.2.4	Experimento 6: Política não-aleatória com operações morfo- lógicas	97
5.2.5	Visualizações axial e tridimensional da segmentação	99

5.2.6	Experimento 7: Uso de especificação de histograma na política não-aleatória.	100
5.3	Comparação entre as políticas de treino e teste nos Experimentos 1-5	101
5.4	Método 3: Abordagem 1.5D - RoBERTa Multi-traços	105
5.4.1	Experimento 8: Política aleatória sem operações morfológicas	105
5.4.2	Experimento 9: Política aleatória com operações morfológicas	107
5.4.3	Experimento 10: Política não-aleatória sem operações morfológicas	107
5.4.4	Experimento 11: Política não-aleatória com operações morfológicas	107
5.4.5	Comparação entre as políticas de treino e teste nos Experimentos 8-11	112
5.5	Comparação entre métodos	114
5.6	Considerações finais	119
6	CONCLUSÃO	120
6.1	Trabalhos Futuros	122
6.1.1	Produções científicas	124
	REFERÊNCIAS	125

1 INTRODUÇÃO

Apesar da demanda crescente de fontes de energia renováveis e políticas de carbono zero, combustíveis fósseis são, ainda hoje, uma importante fonte de energia que possui expectativa de crescimento de consumo (NALLEY; LAROSE, 2021). Antes, porém, que cheguem ao usuário final, existem diversas etapas que devem ser feitas, que vão desde a prospecção da existência de hidrocarbonetos em uma determinada região e eventual exploração até o refinamento e distribuição do combustível.

A determinação da localização e extensão de um reservatório de hidrocarboneto através de dados de subsuperfície obtidos é essencial para a análise de novos poços exploratórios e sua viabilidade econômica. No entanto, a acumulação de petróleo e gás natural pode estar num espaço de 300 metros a 3500 metros abaixo do assoalho oceânico, que por sua vez pode estar entre 100 metros e 2500 metros abaixo do nível do mar. A determinação direta é feita através da perfuração de um poço e verificação da presença ou ausência desses indícios diretamente no local, mas o fato de ser necessária a investigação em áreas tão profundas, fez com que fossem desenvolvidos métodos indiretos de aquisição, isto é, métodos que não precisam que equipamentos estejam a essa profundidade. A perfuração direta é extremamente onerosa - em torno de \$300-\$600 USD por metro de profundidade de perfuração (EIA, 2016), de tal forma que métodos indiretos são utilizados para uma prospecção inicial para, só então, realizar a verificação direta. Os principais métodos indiretos são:

- Levantamento gravimétrico, em que se busca por diferenças na densidade das rochas em subsuperfície,
- Levantamento magnético, que investiga a geologia local com base nas anomalias no campo magnético da Terra causadas pelas propriedades magnéticas das rochas,
- Levantamento sísmico de reflexão e refração, em que se observa o comportamento de reflexão ou refração de ondas sísmicas que se propagam em subsuperfície através dos diversos estratos geológicos.

Dentre os métodos mencionados, o de reflexão é o mais usado pela indústria de Petróleo e Gás (P&G) para identificar regiões onde possam existir reservas de petróleo e gás natural, que consiste na emissão de pulsos de energia mecânica que são transmitidos

em subsuperfície e que, ao encontrar transições no meio de propagação - associadas a mudanças na estrutura física do estrato onde o pulso está propagando -, refletem e são registrados na superfície por sensores denominados hidrofones (em caso de aquisição *offshore*). Os dados do levantamento sísmico são processados e resultam nos chamados traços sísmicos, que correspondem a sequências de intensidade de cada sinal que chega ao hidrofone em função do tempo. Agrupados, os traços sísmicos formam uma imagem da subsuperfície, chamada de imagem sísmica.

Um levantamento sísmico habitual contém inúmeras imagens sísmicas de uma determinada região, que são interpretadas por geofísicos com o auxílio de dados adicionais, tais como o registro de características físicas dos poços (*well logs*), conhecimento prévio de poços próximos etc. A tarefa de delimitar manualmente a extensão de poços de petróleo e/ou gás natural torna-se extenuante à medida em que a quantidade de dados aumenta, além do fato de que detalhes nas imagens podem ser imperceptíveis ao especialista por conta da própria resolução e tamanho de algumas estruturas. O uso de técnicas automáticas para encontrar indícios de hidrocarbonetos mitiga esse problema ao fazer uma espécie de triagem e indicar ao especialista regiões que são mais prováveis de se encontrar petróleo ou gás, e este, então, usa a sua expertise para acatar ou não a indicação automática.

1.1 Objetivos

O objetivo da tese é desenvolver métodos de segmentação de imagens sísmicas automáticos baseados em redes neurais profundas que generalizem para regiões de diferentes perfis geológicos, isto é, que métodos cujos modelos treinados com dados de uma região com determinado contexto geológico seja capaz de segmentar regiões de contextos distintos com assertividade similar. Nesse sentido, propomos metodologias baseadas em redes neurais convolucionais e em arquiteturas tradicionalmente utilizadas em processamento de linguagem natural, abordando o sinal sísmico como uma sequência a ser codificada em um *embedding* produzido através de uma arquitetura baseada em Transformer.

Os objetivos específicos envolvem 1) o desenvolvimento de um método para que sinais sísmicos possam ser processados por modelos de linguagem natural, 2) a adaptação do algoritmo WordPiece para criação de representações de sinais sísmicos tokenizadas, 3) a mudança da divisão de dados de treino e teste comparada à utilizada na literatura

para o mesmo conjunto de dados e 4) o aumento da robustez de uma arquitetura de visão computacional frente à alteração de seus pesos iniciais para comparação com um método baseado em linguagem natural.

1.2 Contribuições

Nesta tese, identificamos um problema na estabilidade da arquitetura U-Net frente a diferentes inicializações na segmentação de imagens de reflexão sísmicas e propomos uma metodologia baseada em um ensemble de U-Nets em conjunto com o uso da técnica de Transformada Wavelet Discreta como uma etapa de pré-processamento nas imagens, cujo intuito é a remoção de ruído e o realce de sinais relevantes para o problema. Mais ainda, propomos a utilização inovadora de modelos de linguagem natural como codificadores de sinais sísmicos. Esta metodologia consiste na adaptação de sinais sísmicos para métodos de tokenização e modelos de classificação textual baseados em redes neurais profundas, em particular a arquitetura BERT. Contribuímos também com a averiguação da capacidade de generalização de modelos para regiões diferentes daquelas cujos dados foram usados para treinamento.

1.3 Organização do trabalho

A tese está organizada da seguinte forma:

- **Capítulo 2** - Arcabouço teórico necessário para o entendimento do problema abordado e do método proposto. Este capítulo apresenta a descrição o tipo de dado utilizado bem como possui considerações das técnicas básicas usadas sobre as quais o método foi desenvolvido.
- **Capítulo 3** - Trabalhos relacionados à tarefa de segmentação ou detecção de estruturas em subsuperfície e seus respectivos desempenhos. Também são referenciados textos relacionados à aplicação de arquiteturas voltadas para tarefas envolvendo sequências e dados sísmicos.
- **Capítulo 4** - Apresentação das metodologias propostas, da base de dados e dos requisitos técnicos.

- **Capítulo 5** - Resultados, discussões e comparações com outras abordagens presentes na literatura.
- **Capítulo 6** - Conclusão

2 FUNDAMENTOS TEÓRICOS

Neste capítulo, serão apresentados conceitos relevantes no entendimento e desenvolvimento deste trabalho. Primeiro, falaremos sobre o processo de aquisição de dados mais comum na indústria de exploração desses recursos e sua interpretação. Em seguida, serão explicados a teoria de séries temporais e sua relação com o PLN, área de estudo na qual dois dos métodos usados estão inseridos. Por fim, será feita uma revisão sobre métodos de aprendizado profundo (*deep learning*), em particular as arquiteturas Bidirectional Encoder Representations from Transformer (BERT) e U-Net.

2.1 Aquisição e interpretação de dados sísmicos

O levantamento sísmico de reflexão funciona através da geração de ondas mecânicas por uma fonte artificial de impacto, como disparos de cápsulas de ar comprimido quando se trata de uma aquisição marítima, chamados “tiros”.

Ao chegar ao assoalho marinho, as ondas se propagam em subsuperfície, formada por estratos compostos por diferentes materiais, como sal, rochas ígneas, petróleo, gás natural, água etc. A mudança de meio de propagação também faz com que parte da onda seja refletida. A Figura 1 ilustra esse processo.

A proporção de onda refletida e transmitida na interseção entre dois estratos diferentes é dada pela Lei de Snell (KIM et al., 2020) e varia de acordo com o ângulo de incidência e os índices de refração dos estratos, que estão relacionados às suas propriedades elásticas, em particular a impedância acústica (LATIMER; DAVIDSON; RIEL, 2000). Os hidrofones detectam uma série de ondas que tomaram caminhos diferentes pela subsuperfície, refletidos pelos topos e fundos de cada uma das camadas da subsuperfície, sendo possível observar sinais oriundos de vários quilômetros de profundidade. A amplitude das ondas refletidas de volta à superfície em função do tempo compreende uma série temporal, chamada de **traço sísmico**, ilustrado na Figura 2.

Através da análise da amplitude de diferentes ondas e de seus tempos de chegada aos detectores, as propriedades da subsuperfície podem ser compreendidas. Em particular, a razão entre as impedâncias acústicas de meios com a presença de hidrocarbonetos e sedimentos é sempre negativa e, em geral, com amplitude do sinal elevada em comparação com as outras transições. As anomalias nas amplitudes devidas a diferença entre

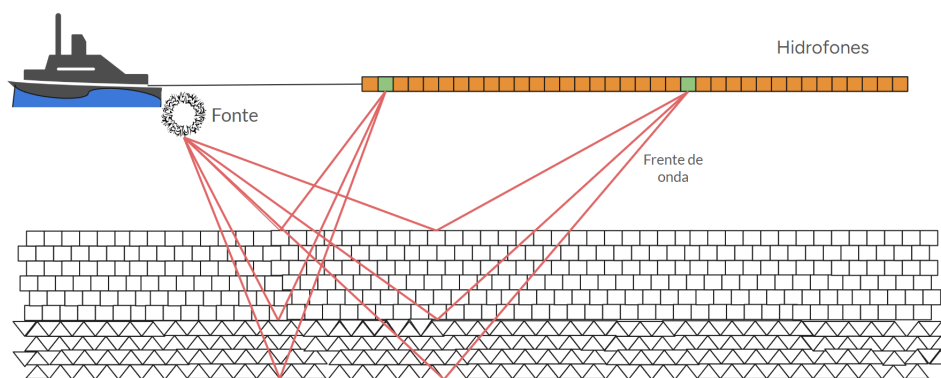


Figura 1 – Aquisição de dados sísmicos por reflexão (Exemplo *offshore*). Imagem de elaboração própria.

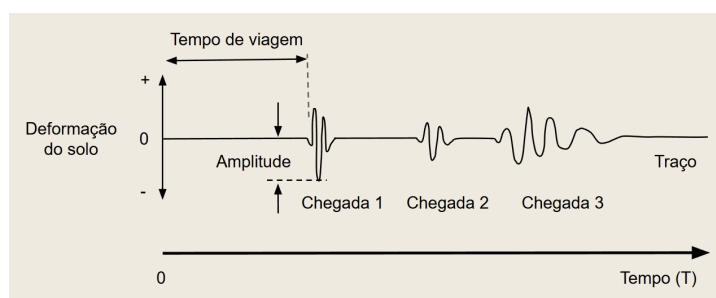


Figura 2 – Amplitude de deformações em função do tempo. Imagem adaptada de (DENTITH; MUDGE, 2014)

impedâncias acústicas dos meios de propagação das ondas são chamadas *bright spots* e são um Indicador Direto de Hidrocarboneto (IDH), isto é, indicam onde na subsuperfície existe uma possibilidade maior de encontrar petróleo e gás natural.

Existem dois tipos de aquisições sísmicas por reflexão: a aquisição 2D e a 3D. Na aquisição 2D, tanto as fontes de impacto quanto os hidrofones são colocados ao longo de uma mesma linha, como é o caso da Figura 1. A imagem resultante pode ser considerada como uma seção transversal da subsuperfície abaixo da linha onde os hidrofones estão posicionados. Sua vantagem em relação à aquisição 3D consiste no seu menor custo efetivo e processamento mais rápido, o que torna este tipo de aquisição própria para os primeiros estágios de exploração de uma determinada área.

O segundo tipo de aquisição, a 3D, consiste em espalhar hidrofones sobre uma área na superfície e mover a fonte de vibração ao longo dessa área, como exemplificado na Figura 3. Na Figura 4 mostra-se uma foto de um navio que faz esse tipo de aquisição. A aquisição resultante deste método pode ser entendida como um cubo formado abaixo da região de prospecção. Esta técnica possui vantagens em relação à aquisição 2D, como a inclusão de uma dimensão adicional, o fato de muito mais informação estar disponível em cada ponto - o que normalmente implica numa melhoria da resolução das características

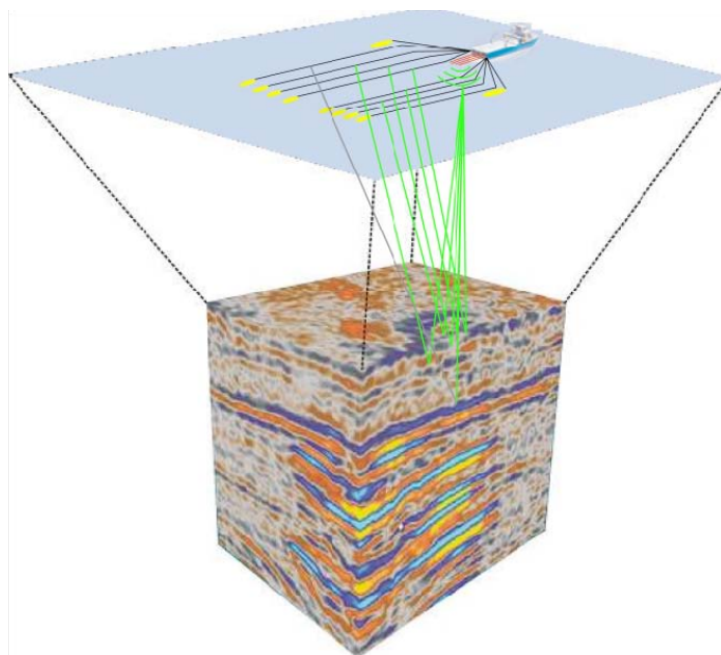


Figura 3 – Exemplo de uma aquisição 3d *offshore*. A imagem resultante é um cubo 3d abaixo da região de prospecção. Imagem retirada de (CAMESELLE, 2010).



Figura 4 – Navio que faz a aquisição sísmica. Imagem retirada de <https://www.tgs.com/seismic/marine-acquisition/the-fleet/vessels/ramform-atlas>. Acesso em 04/11/2024

em subsuperfície, e a eliminação de reflexões “fantasmas” de regiões próximas, i.e., a reflexão ocorreu mas não devido à estrutura mostrada na linha 2D.

O cubo resultante da aquisição 3D pode ser dividido em seções transversais na direção da linha de tiro (o sentido e direção a qual a fonte de vibração se desloca), que são convencionalmente chamadas de *inlines* e em seções perpendiculares à linha de tiro, chamadas *crosslines*. A profundidade do cubo representa o tempo que a onda levou para chegar até a camada refletora e voltar para ser registrada pelos hidrofones.

2.1.1 Tratamento dos dados

Dada a diferença entre as características de cada bacia sedimentar, não existe um método ou fluxograma de métodos único que funcione para analisar os dados de todas as regiões. Por isso, o tratamento de dados sísmicos adquiridos por reflexão sísmica é feito por equipes multidisciplinares formadas por geofísicos, geólogos e engenheiros. As ações mais comuns durante a etapa de tratamento são a deconvolução, o agrupamento de ponto comum (*Common Midpoint Gathering ou CMP*) e a migração - temporal e espacial (KEAREY; BROOKS; HILL, 2009). A etapa de migração é importante principalmente para facilitar a interpretação dos dados, pois realoca os eventos sísmicos para a localização onde eles ocorreram na subsuperfície ao invés da localização onde o evento foi detectado na superfície. Em outras palavras, é através desse processo que os dados são transformados em uma imagem da subsuperfície.

A Figura 5 mostra um exemplo dos dados originais, onde o eixo horizontal representa a distância à direita e à esquerda em relação ao ponto de onde foi emitido o tiro o eixo vertical é o tempo que demora para se detectar o sinal refletido. A imagem à direita é uma versão mais granular no espaço da imagem à esquerda. Após o processamento de migração temporal, os dados ficam com a aparência ilustrada na Figura 6. A imagem se encontra no modo denominado wiggle, onde cada traço é representado individualmente, e foi incluída deste modo para facilitar a comparação entre o CMP e a imagem migrada temporalmente. Nas imagens seguintes, a imagem migrada será apresentada no modo escala de cinza, que é, quantitativamente, o mesmo tipo de visualização mostrada na Figura 6.

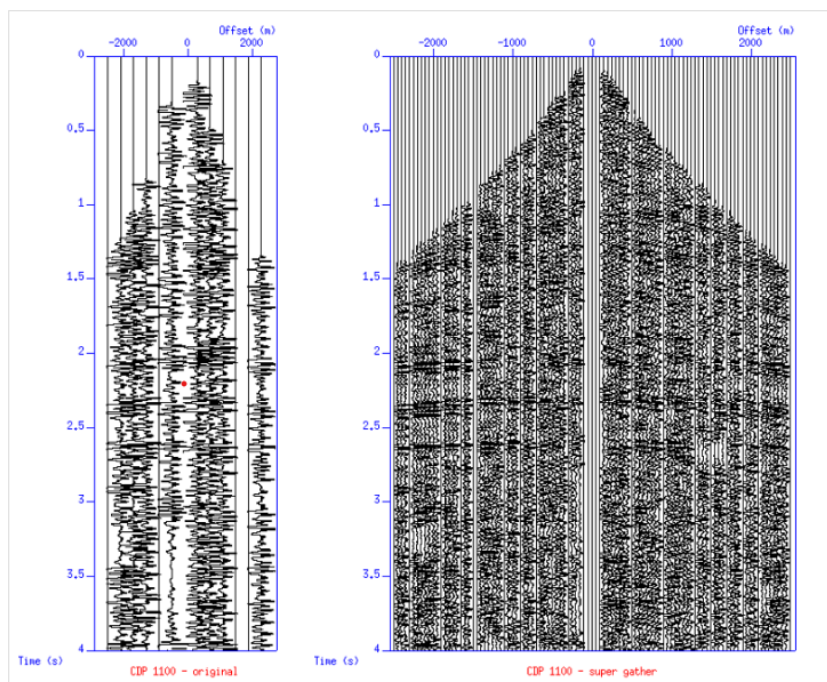


Figura 5 – Empilhamento de ponto comum. Imagem retirada de <https://germangarabito.wordpress.com/2019/04/13/cmp-super-gather-formation/>. Acesso em 03/04/2025.

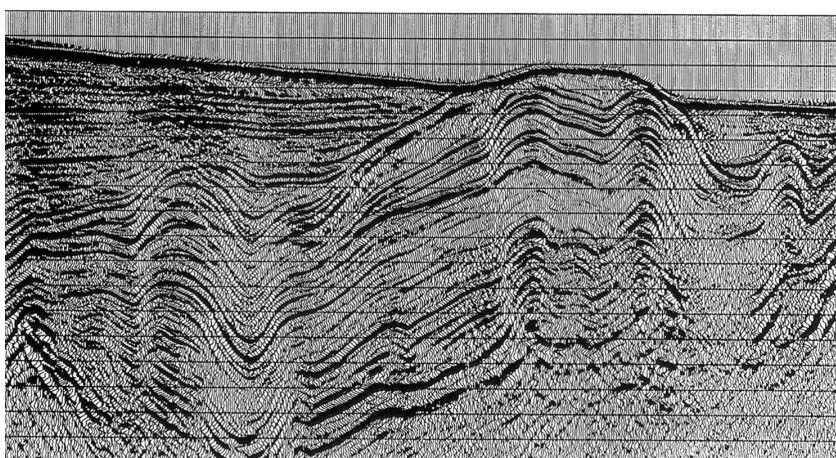


Figura 6 – Dados após migração temporal, já se comportando como uma fotografia ao mostrar a forma aparente das estruturas geológicas em subsuperfície. Para a visualização mais exata das estruturas, é necessária uma migração espacial. Imagem retirada de https://wiki.aapg.org/index.php?title=Seismic_migration. Acesso em 13/07/2025.

Após essas operações, os dados estão prontos para serem interpretados, e resultam em imagens como a da Figura 7. Para esta figura específica, o eixo horizontal representa o tempo de propagação ida-e-volta da onda originada na fonte de vibração, foi refletida por alguma estrutura em subsuperfície e chegou nos detectores na superfície. Esse tempo é chamado, em inglês, de *Two-Way Traveltime* (TWT). O eixo horizontal é normalmente medido em número de traços ou em quilômetros, sendo a conversão do

primeiro para o segundo feita conhecendo-se quantos traços são empilhados na etapa de CMP e a distância entre os detectores.

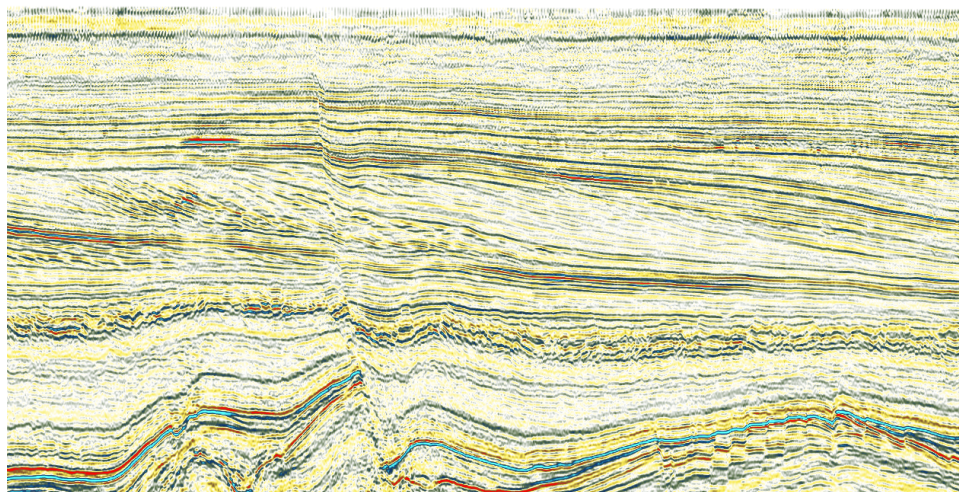


Figura 7 – Imagem gerada pela aquisição sísmica de reflexão, que pode ser interpretada como uma seção transversal da subsuperfície. Imagem retirada de (SILVA et al., 2019b).

2.1.2 Interpretação

A maior parte dos reservatórios estão associados a formações geológicas que possuem estruturas convexas para cima (trapa anticlinal) e falhas, como indicado na Figura 8. Alguns outros casos também podem apresentar outras estruturas, como a trapa estratigráfica, mas as duas primeiras são mais comuns. Durante a fase de interpretação dos dados sísmicos, o especialista busca inferir a geologia do local. Na prospecção de reservatórios de gás natural e petróleo, essa interpretação é feita com o objetivo de identificar regiões em subsuperfície que possam conter hidrocarbonetos - entre eles o gás natural. Um dos principais indicadores de hidrocarbonetos é a presença de um *brightspot*, que são pontos onde ocorre um aumento local de amplitude, que pode ser causado pela presença de gás ou óleo nos poros de alguma rocha. Em escala de cinza, e tratando normalmente amplitudes negativas como mais próximas ao branco, esses locais apresentam pontos bem claros em relação ao resto da imagem, como visto na Figura 9.

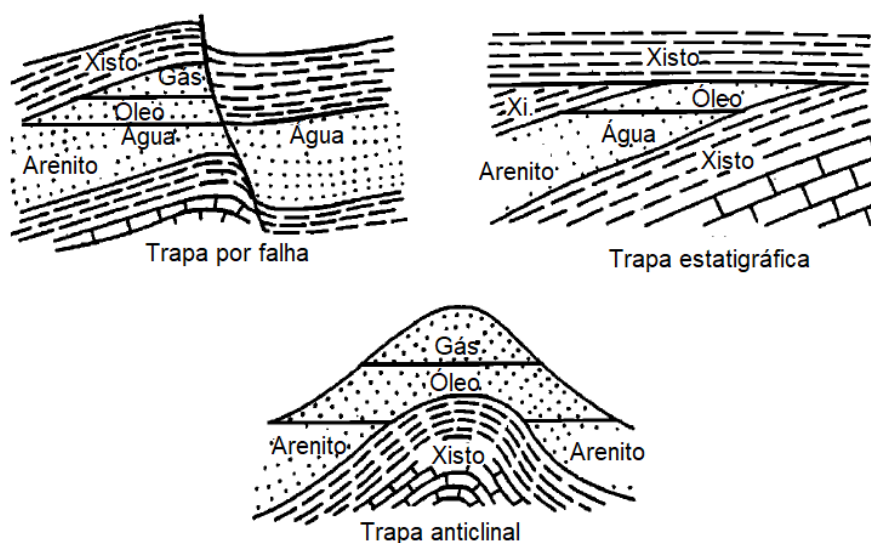


Figura 8 – Tipos comuns de falhas. Imagem adaptada de (ROBINSON; TREITEL, 2000).

O geocientista normalmente faz essa interpretação com o uso de softwares especializados através da análise de cada linha individualmente se for um levantamento 2D ou de volumes inteiros no caso de um levantamento 3D. Esse trabalho é minucioso e consome muito tempo do geofísico devido a enorme quantidade de dados gerados em levantamentos sísmicos modernos.

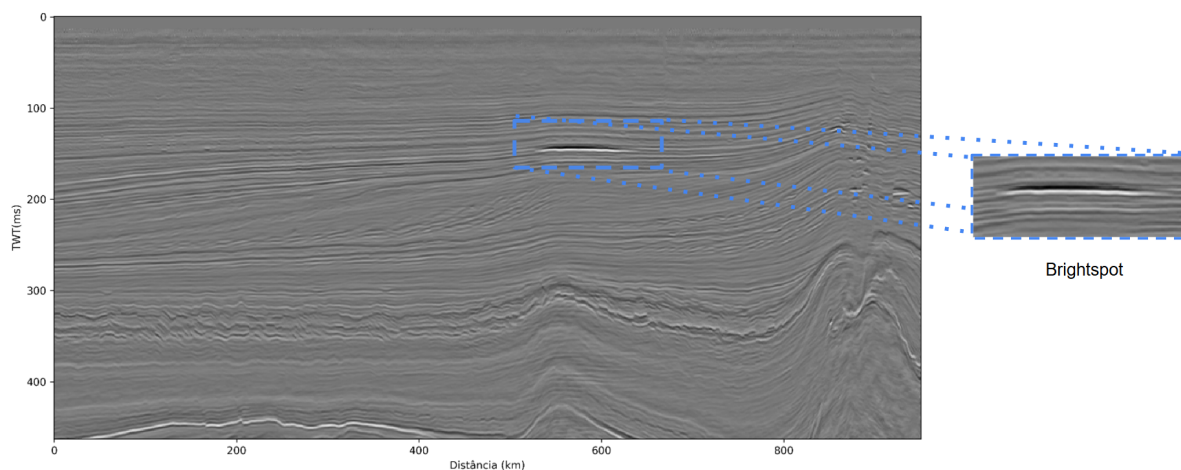


Figura 9 – Localização de brightspots em uma imagem de reflexão sísmica em escala de cinza. Imagem de elaboração própria.

Com o objetivo de auxiliar o especialista nessa tarefa, nos últimos anos têm sido pesquisadas diversas aplicações de algoritmos de aprendizagem de máquina em uma grande variedade de tarefas relacionadas ao campo do conhecimento da geofísica. Para tarefas de prospecção de gás e petróleo, a segmentação é a técnica de visão computacional mais visada, pois destaca grupos de pixels que podem indicar a presença de gás natural

e petróleo na imagem sísmica, direcionando o foco do especialista para áreas com maior potencial de possuir reservas.

2.2 Segmentação de imagens

A segmentação de imagens é uma tarefa fundamental em visão computacional com aplicações em diversas áreas. Seu principal mecanismo é o particionamento de uma imagem $I : \Omega \rightarrow \mathbb{R}^2$ (com $\Omega \subset \mathbb{R}^2$) em múltiplos segmentos R_i de forma que

$$\bigcup_{i=1}^n R_i = \Omega, \quad R_i \cap R_j = \emptyset \quad \forall i \neq j \quad (2.1)$$

e cada região R_i é homogênea com respeito a um certo critério de similaridade f de modo que

$$\forall x, y \in R_i, \quad f(x) \approx f(y) \quad (2.2)$$

Uma imagem é segmentada em uma ou mais regiões distintas com o objetivo isolá-las de forma a corresponderem a objetos reais, como um carro em uma imagem de alguma região urbana, um animal em uma floresta ou uma região que contém hidrocarbonetos em uma imagem sísmica. A abordagem mais intuitiva para essa tarefa envolve o uso de softwares por um anotador especialista, isto é, a pessoa que irá operar o software que, em geral, possui conhecimento especializado caso a identificação e delimitação do objeto o exija. O especialista identifica e seleciona o contorno ou os pixels do objeto a ser segmentado manualmente com base na própria imagem e informações adicionais caso estejam disponíveis.

No caso particular das imagens sísmicas, que possuem diferentes características de resolução entre o eixo vertical (tempo/profundidade) e o eixo horizontal (espaço), a aparência das características geológicas é afetada, o que dificulta ainda mais o processo de segmentação. Os dados sísmicos também possuem muitos tipos de ruído, como o ruído aleatório ou devido a reflexões múltiplas, que acabam por obscurecer algumas características geológicas ou criar falsos padrões. Mais ainda, é possível que existam múltiplas interpretações para o mesmo dado sísmico graças à própria natureza das imagens sísmicas cujas fronteiras geológicas podem apresentar transições graduais de dezenas de metros.

Os fatores mencionados contribuem para que a segmentação manual por um es-

pecialista seja, muitas vezes, um processo lento e laboroso que prolonga o período entre o levantamento de uma determinada região com potencial de produção e a extração do hidrocarboneto efetivamente. Por isso, é cada vez mais comum que esta etapa seja auxiliada por algoritmos de aprendizado de máquina, cujo objetivo é direcionar a atenção do especialista em determinada região de maior potencial de produção de hidrocarbonetos.

O tipo de técnica de aprendizagem de máquina usada para segmentar a imagem pode ser escolhida baseando-se em duas formas de se estudar os dados sísmicos: como uma imagem (imagem sísmica de reflexão) ou como um conjunto de séries temporais (traços sísmicos). No primeiro caso, técnicas baseadas em redes neurais convolucionais (CNN, em inglês) (LECUN; KAVUKCUOGLU; FARABET, 2010) são o estado da arte em diversas áreas de pesquisa, incluindo não só aplicações na indústria de P&G (ZHANG et al., 2020a; CHANG et al., 2018; FENG et al., 2020), como também no georreferenciamento (FENG; PASSONE; SCHUSTER, 2021), na área médica (DINIZ et al., 2022a; LASSAU et al., 2021) e outras.

Vistos como séries temporais, podemos utilizar um conjunto bastante amplo de técnicas já bem amadurecidas e estudadas para classificação, como, por exemplo, a própria CNN (WANG; YAN; OATES, 2017), máquina de vetor de suporte (ZHANG et al., 2010) e, mais recentemente, técnicas baseadas em Transformers (WEN et al., 2022). Ambas as interpretações possuem vantagens e desvantagens, que serão discutidas nas próximas seções.

2.3 Redes neurais convolucionais

Redes neurais convolucionais (CNNs, na sigla em inglês) são arquiteturas que combinam três conceitos: campos receptivos locais, pesos compartilhados e sub-amostragem espacial. Com campos receptivos locais, os neurônios da rede podem extrair características visuais elementares como bordas, pontas, quinas etc. Essas características são então combinadas em camadas posteriores para encontrar características de ordem superior, como, por exemplo, texturas. A Figura 10 ilustra o comportamento da extração de características em uma CNN.

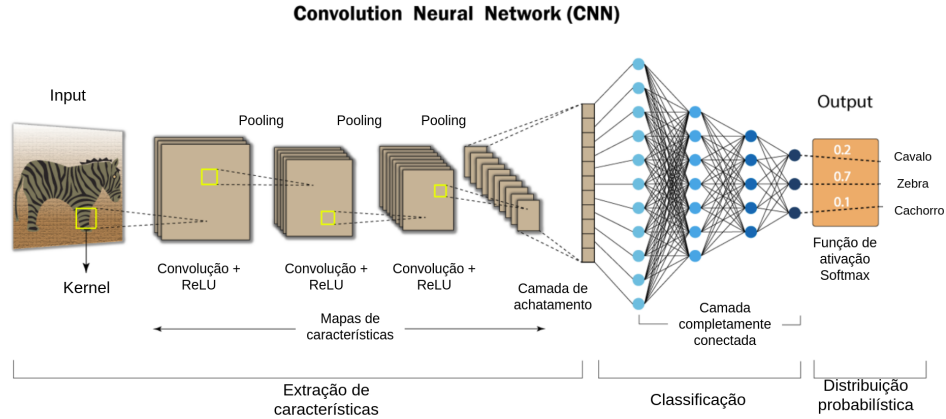


Figura 10 – Exemplo de arquitetura de uma CNN. Imagem adaptada de <https://medium.com/codex/understanding-convolutional-neural-networks-a-beginners-journey-into-the-architecture>. Acesso em 13/07/2025.

As características de uma parte da imagem podem ser úteis em toda a imagem. Por exemplo, uma característica típica de uma acumulação de gás natural no centro da imagem pode aparecer em uma outra acumulação no lado esquerdo da imagem. O conhecimento de características que aparecem por toda a imagem é obtido através do uso de pesos compartilhados entre as unidades que fazem a extração em diferentes campos receptivos. O conjunto de saídas dessas unidades é chamado de mapa de características.

A implementação sequencial de um mapa de características passaria por toda a imagem de entrada com uma única unidade que possui um campo receptivo local, que armazena os estados dessa unidade nas localizações correspondentes no mapa de características. Essa operação é equivalente a uma convolução, seguida de uma adição de viés e uma função de ativação.

Podemos definir formalmente a convolução $C(x, y)$ de uma função $f(x, y)$ por um kernel $g(x, y)$ como

$$C(x, y) = (f * g)(x, y) = \sum_{i, j = -\infty}^{\infty} f(i, j)g(x - i, y - j) \quad (2.3)$$

onde i e j são índices que percorrem todas as localizações de cada pixel que forma a imagem. O kernel g pode ser entendido como o mapa de características. Dada uma função de ativação $\phi(\cdot)$ e um viés B , o resultado final para uma operação de uma camada convolucional se dá por

$$C_f(x, y) = \phi(C(x, y) + B) \quad (2.4)$$

As funções de ativação ϕ são usadas em nodos de redes neurais para aprender características através de transformações não-lineares. Algumas propriedades desejáveis de funções de ativação são:

- adicionar não-linearidade no processo de otimização, melhorando a convergência da rede neural;
- não aumentar exageradamente a complexidade do modelo
- preservar a distribuição dos dados para otimizar o treinamento da rede neural
- evitar problemas durante a etapa de gradiente descendente, como a dissipação ou explosão de gradiente

Algumas funções de ativação estão descritas a seguir. A função sigmoide, definida como

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2.5)$$

transforma quaisquer valores reais x em valores entre 0 e 1. Por conta destes valores de saída, é comumente usada em modelos que prevêem probabilidades.

A tangente hiperbólica

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.6)$$

muitas vezes é usada no lugar da sigmoide, especialmente por proporcionar um maior desempenho a redes neurais com muitas camadas. Sua função é transformar qualquer valor real x em um valor entre -1 e 1.

Como último exemplo de função de ativação, citaremos a Unidade Linear Retificadora (ReLU, na sigla em inglês). Seu uso implica na redução da probabilidade de ocorrência da dissipação do gradiente, já que quando $x > 0$, o gradiente apresenta valor constante. Também é mais eficiente computacionalmente em relação à sigmoide e tangente hiperbólica por não haver necessidade de se calcular exponenciais.

$$ReLU(x) = \max(0, x) \quad (2.7)$$

Devido ao custo computacional de treinar uma rede neural, a operação de redução (*pooling*) é usada para reduzir o número de pesos na rede e também controlar o sobreajuste na rede neural. Existem diversos tipos de redução (GHOLAMALINEZHAD; KHOSRAVI, 2020), e nesta tese será utilizada a redução máxima (*max-pooling*) definida pela equação

$$y_{i,j} = \max(x_{m,n} | (m,n) \in R(i,j)) \quad (2.8)$$

onde $R(i,j)$ é o campo receptor na localização centralizada em (i,j) e consiste em todos os índices m e n que localizam pixels dentro da janela $X_{m,n}$. A escolha desta operação se dá por sua capacidade de ser robusta a ruído, além de ser útil para extrair características mais definidas como quinas, pontos etc, comumente encontradas em imagens sísmicas.

2.3.1 U-Net

A U-Net é uma arquitetura baseada em redes neurais convolucionais formada por duas partes principais: o codificador e o decodificador. O codificador é construído como uma série de camadas convolucionais, cujas dimensões ficam menores a medida que a rede se aprofunda, a fim de capturar informações contextuais importantes e diminuir a resolução da imagem de entrada codificando-a em uma série de mapas de características fundamentais em um espaço latente que facilitam a identificação de acumulações de gás em imagens sísmicas. A parte decodificadora é expansiva: parte do conjunto final de mapas de características obtidos pelo codificador e usa este conjunto de filtros para procurar a localização das características que se deseja encontrar, no fim chegando à resolução da imagem de entrada.

Um elemento fundamental da arquitetura U-Net, que a diferencia de um autoencoder (BANK; KOENIGSTEIN; GIRYES, 2021) é o uso de conexões diretas, chamadas *skip connections*, entre camadas do codificador e do decodificador de mesmo tamanho. Essas conexões servem para recuperar informações aprendidas nas camadas iniciais, durante a codificação, e que possam ter sido desprezadas em camadas posteriores. A ilustração da arquitetura U-Net na Figura 11 mostra como cada um dos componentes descritos se encaixa na arquitetura.

Na última camada do decodificador, a função de ativação utilizada para a segmentação binária é a sigmoide definida na Equação 2.5.

O objetivo desta função é fazer com que quaisquer valores de saída da última

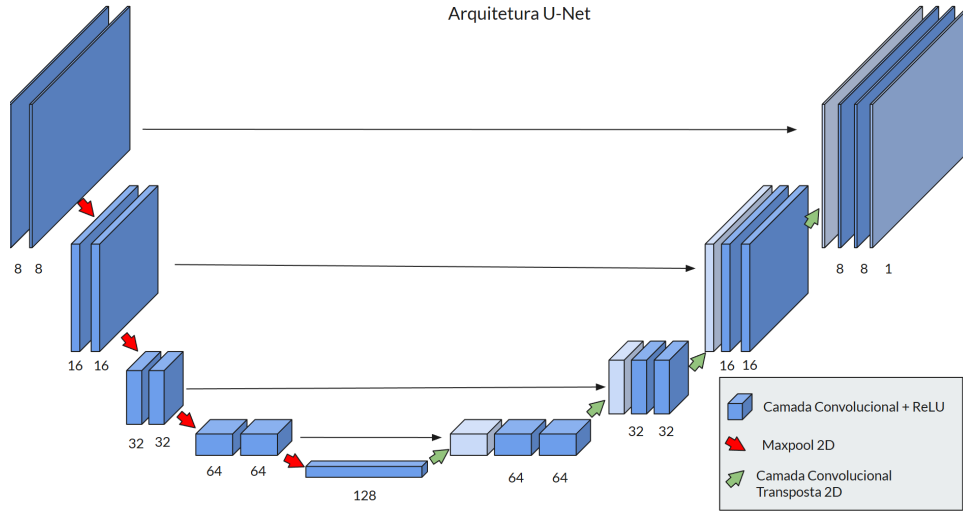


Figura 11 – Arquitetura U-Net. Elaboração própria.

camada convolutiva esteja entre 0 e 1. Intuitivamente, interpretamos o resultado desta camada como a probabilidade do pixel correspondente pertencer à classe positiva. Para que a imagem final seja composta por pixels 0 ou 1 apenas, é necessário que se aplique um limiar sobre a saída da função sigmoide. Neste trabalho usa-se o limiar 0.5, ou seja, caso um pixel apresente saída maior ou igual a 0.5, ele é classificado como a classe positiva. Caso seja menor que 0.5, será classificado como pertencente à classe negativa. Assim, a imagem de saída da arquitetura U-Net possui a mesma dimensão da imagem de entrada, porém com valores de pixel binários correspondentes à classe inferida pela rede.

2.4 Transformer

A arquitetura Transformer foi desenvolvida para permitir maior paralelização das computações ao mesmo tempo em que fosse capaz de aprender dependências entre palavras independentemente de suas distâncias relativas nas sequências de entrada ou saída.

Como pode ser visto na Figura 12, a arquitetura é composta de duas partes menores. A primeira parte - codificadora -, à esquerda na Figura, mapeia uma sequência de representações simbólicas $(x_1, x_2, \dots, x_n)$ para uma sequência de representações contínuas $\mathbf{z} = (z_1, z_2, \dots, z_n)$. A segunda parte, decodificadora, gera, para cada $\mathbf{z}$, uma sequência de símbolos $(y_1, y_2, \dots, y_n)$ de saída um elemento de cada vez. A cada passo o modelo consome símbolos já gerados anteriormente como entrada adicional ao gerar o próximo.

O codificador é composto por uma pilha de $N = 6$ camadas idênticas, que pos-

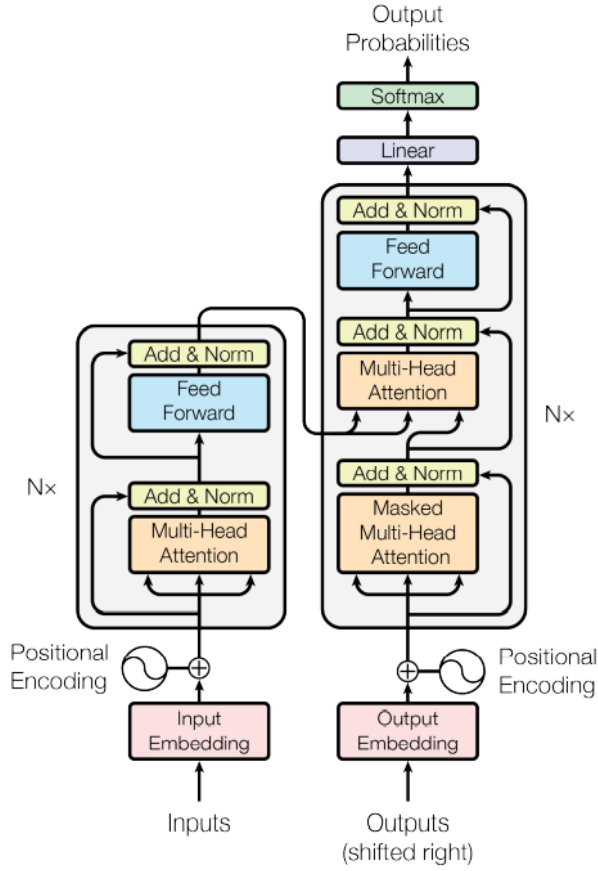


Figura 12 – A arquitetura Transformer. Imagem retirada de (VASWANI et al., 2017)

suem duas subcamadas cada. A primeira é uma camada de auto-atenção multi-cabeça e a segunda é camada totalmente conectada simples, com uma conexão residual (HE et al., 2016) ao redor de cada subcamada, seguida de uma normalização de camada (BA; KIROU; HINTON, 2016).

Já o decodificador é composto, também, por $N = 6$ camadas iguais entre si. Além das duas subcamadas do codificador, ele apresenta uma terceira subcamada, que faz um processo de atenção de múltiplas cabeças (*multi-head attention*) sobre a saída do codificador. Conexões residuais e normalizações de camadas também são usadas.

A função de atenção mapeia uma consulta (query) Q , e um conjunto chave-valor (key-value), K e V , todos vetores, a uma saída também vetorial. A saída é calculada através de uma soma ponderada de todos os valores onde os pesos são calculados através de uma função de compatibilidade da consulta com a chave correspondente:

$$\text{Atenção}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2.9)$$

onde d_k é a dimensão das consultas Q e chaves K. O fluxograma para essa operação pode ser encontrado à esquerda na Figura 13 e é chamada de **Atenção de Produto Escalar com Escala**. Na Transformer, porém, são realizadas, para uma única 3-tupla (Q,K,V), h funções de atenção como essa, onde cada valor é projetado linearmente e em seguida fornecido como entrada para a função de atenção. As h saídas são então concatenadas de modo que a **Atenção Multihead**, ilustrada na Figura 13, é definida como

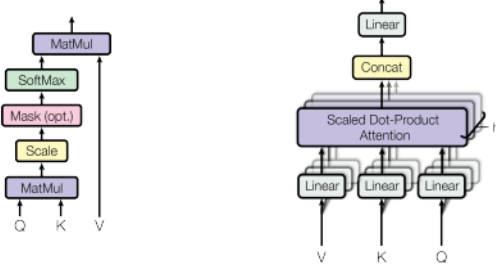


Figura 13 – À esquerda, o mecanismo de atenção de produto escalar com escala. À direita, a camada de atenção multi-cabeça, que consiste de várias camadas de atenção em paralelo. Imagem retirada de (VASWANI et al., 2017)

$$\text{AtençãoMC} = \text{Concat}(\text{head}_1, \dots, \text{head}_n)W^O \quad (2.10)$$

onde $\text{head}_i = \text{atenção}(QW_i^Q, KW_i^K, VW_i^V)$, as projeções são matrizes de parâmetros $W_i^Q, W_i^K \in \mathcal{R}^{d_{\text{model}} \times d_k}$, $W_i^V \in \mathcal{R}^{d_{\text{model}} \times d_v}$ e $W_i^O \in \mathcal{R}^{d_{\text{model}} \times h d_v}$ e d_v é a dimensão dos valores V.

2.5 Modelos de linguagem

De forma bastante simplificada, podemos considerar linguagem como uma sequência de palavras, que podem ser codificadas como uma sequência de símbolos. Um modelo de linguagem busca capturar as propriedades estatísticas da linguagem natural. Ao aprender a distribuição de probabilidade de sequências de palavras, um modelo de linguagem pode prever a probabilidade de uma determinada palavra seguir uma sequência de palavras anteriores.

A capacidade de um modelo de linguagem aprender essa distribuição depende fundamentalmente de sua compreensão do contexto e significado do texto. Como esses modelos não conseguem processar texto diretamente, é necessário que as palavras e outros símbolos linguísticos sejam convertidos em vetores numéricos em um processo chamado de tokenização e *embedding*. A tokenização é a etapa onde palavras são frequentemente

quebradas em partes menores, como radicais, sufixos e prefixos com base em processos bem definidos, como *Byte-pair Encoding* (BPE) (SENNRICH; HADDOW; BIRCH, 2016) e *WordPiece* (WU et al., 2016), e atribuir a cada uma dessas partes únicas um valor numérico inteiro distinto, chamado token. Durante o processo de *embedding*, cada token é convertido em um vetor cuja dimensão é previamente definida.

Entre as diversas arquiteturas desenvolvidas para fazer o *embedding* de tokens, a arquitetura Bidirectional Encoder Representation from Transformers, ou BERT (DEVLIN et al., 2018), e suas variações possuem grande desempenho em muitas áreas de aplicação atuais, como na restauração de dados em sistemas de energia (HU et al., 2023), no reconhecimento de entidades nomeadas (NER, em inglês) em dados biomédicos (KOPRDI et al., 2024) e no entendimento de séries temporais (ZHANG et al., 2025).

2.5.1 Bidirectional Encoder Representation from Transformers - BERT

Graças a sucessivos casos de melhora no estado-da-arte com a Transformer, pesquisadores desenvolveram a Representação Bidirecional de Codificadores de Transformers, BERT na sigla em inglês, cuja arquitetura é formada exclusivamente por codificadores de Transformer empilhados (Figura 14). Essa arquitetura foi desenvolvida para ser pré-treinada em tarefas não-supervisionadas de modo a aprender representações de palavras, para que estas sejam utilizadas em tarefas posteriores, como classificação, inferência de linguagem e outras. Isso faz com que este seja um modelo capaz de usar a grande quantidade de dados não-rotulados disponíveis de diferentes linguagens.

O pré-treinamento é realizado em duas tarefas simultâneas. Na primeira, chamada de Modelagem de Linguagem Mascarada (MLM), algumas palavras dos dados de entrada da rede são mascaradas para que a própria rede aprenda a inferir que palavra é aquela dado o contexto da frase - em ambas as direções. Esse processo é útil para que uma palavra não “veja” a si mesma de modo que o modelo previsse trivialmente a palavra alvo. Por exemplo, na Figura 15, a frase original é “Paris is a beautiful city. I love Paris”. Neste exemplo, a palavra “city” é escondida do modelo e, após achar a representação de cada uma das palavras existentes na frase, ele deve prever qual é a palavra que foi mascarada, com o auxílio de uma camada *feed forward* e função de ativação softmax.

A segunda tarefa é chamada de Previsão de Próxima Sentença (NSP, em inglês), ilustrada na Figura 16. Nela, o modelo aprende a entender relações entre duas sentenças, o

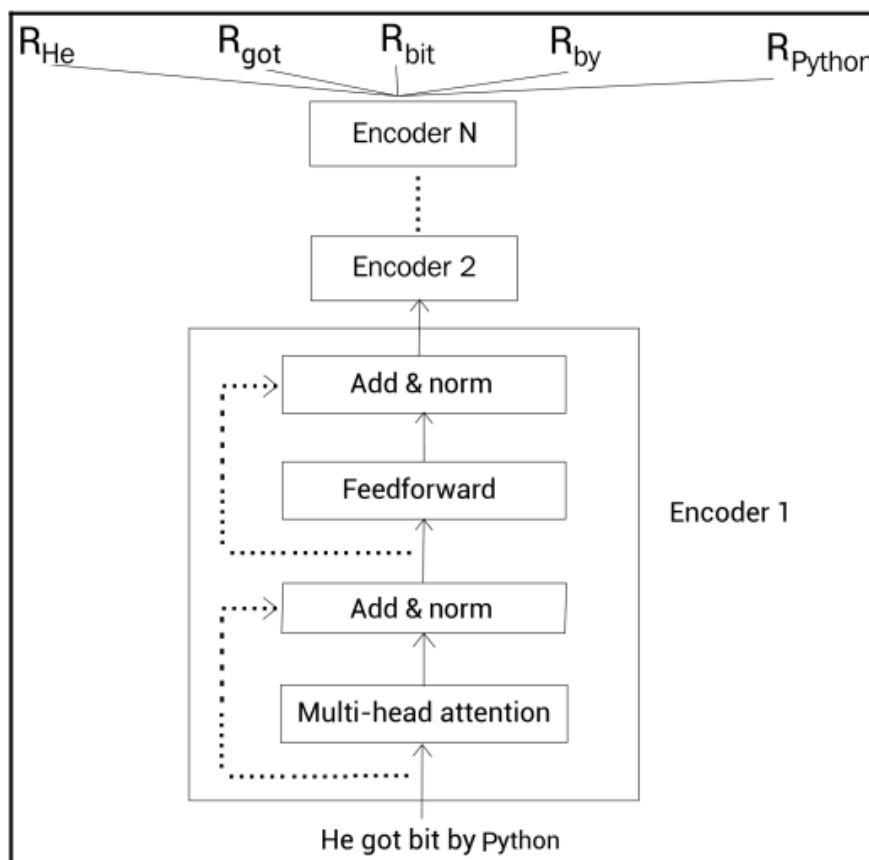


Figura 14 – Arquitetura BERT. Imagem retirada de <https://wikidocs.net/224181>. Acesso em 13/07/2025.

que interfere diretamente no desempenho do modelo em tarefas como *Question Answering* e Inferência de Linguagem Natural.

Como exemplo desta tarefa, na Figura 16, existem duas frases em sequência: “She cooked pasta” e “It was delicious”, separadas por um token [SEP], utilizado para denotar separações entre sentenças e fins de frases nesta arquitetura. Aqui usa-se a representação do token [CLS], que denota o começo de uma sequência de frases, para inferir se a segunda frase na sequência é ou não uma frase que pode vir depois da primeira. No exemplo, a segunda frase possui 90% de chance de seguir a primeira.

Os autores de (LIU et al., 2019) argumentam que o BERT era uma modelo significativamente subtreinado, isto é, com grande margem para melhoria de desempenho a depender de escolhas no projeto da arquitetura. No trabalho, os autores mostram a importancia de algumas escolhas no projeto que foram negligenciadas e desenvolvem o Robustly Optimized BERT Pretraining Approach (RoBERTa).

As mudanças cruciais feitas no RoBERTa são

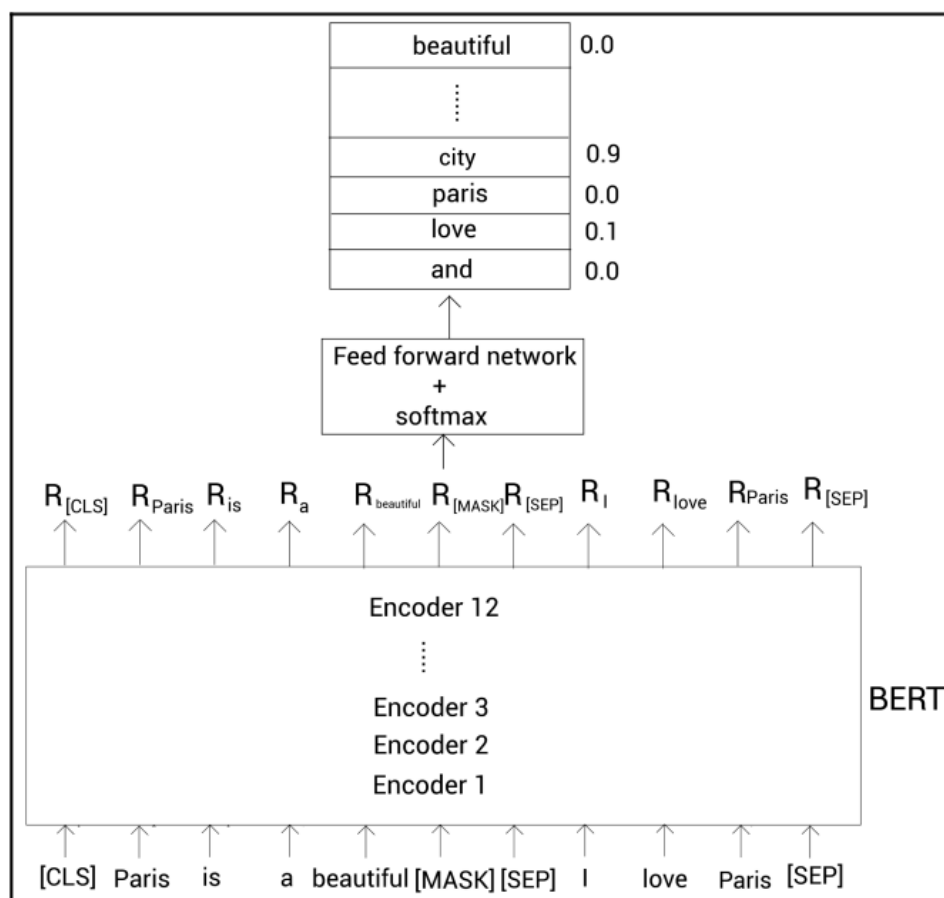


Figura 15 – Tarefa de modelagem de linguagem mascarada. Imagem retirada de <https://wikidocs.net/224181>. Acesso em 13/07/2025.

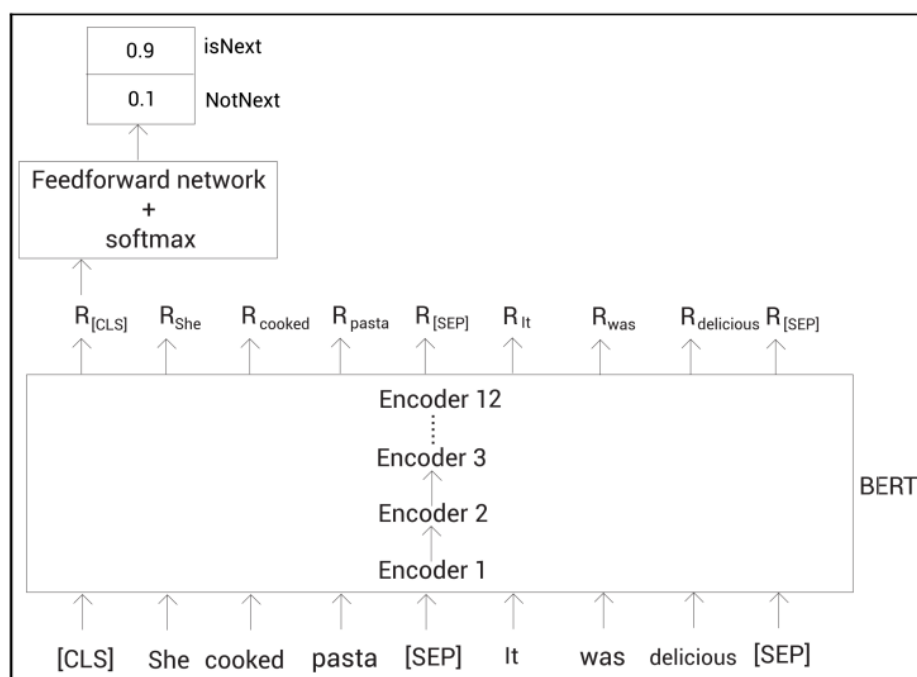


Figura 16 – Tarefa de previsão de próxima sentença. Imagem retirada de <https://wikidocs.net/224181>. Acesso em 13/07/2025.

- A não inserção aleatória de sequências pequenas durante a etapa de MLM.
- A remoção da etapa de NSP.
- O uso do parâmetro $\beta_2 = 0.98$ no otimizador ADAM para melhorar a estabilidade do modelo ao treinar com batches grandes.

Com isso, a arquitetura passa a ser mais robusta e com melhores resultados que o BERT original. Por conta disso, esta foi uma das arquiteturas escolhidas nesta tese.

2.5.2 Tokenização WordPiece

Algoritmos de tokenização são métodos usados para transformar dados de texto em tokens individuais que carregam algum significado dentro do contexto do problema para o qual utilizam a tokenização. Tokens podem ser caracteres, palavras, sub-palavras - subdivisões de uma palavra, como um radical ou um sufixo - ou alguma outra sequência de caracteres.

O WordPiece (DEVLIN et al., 2018) é um algoritmo de tokenização usado durante o pré-treinamento do BERT e arquiteturas baseadas neste, como RoBERTa, DistilBERT, MobileBERT etc.

Primeiro, é criado um vocabulário com cada caractere, de modo que o tamanho inicial do vocabulário coincida com o número total de caracteres no alfabeto brasileiro. Cada par de caracteres e/ou sub-palavras no vocabulário é usado para calcular um score através da equação

$$score(a, b) = \frac{f_{ab}}{f_a f_b} \quad (2.11)$$

onde f_{ab} é a frequência de um par consecutivo de caracteres e/ou subpalavras a e b e f_a , f_b são as frequências dos caracteres a e b no corpus, respectivamente. O par com score mais alto é adicionado ao vocabulário e é considerado como uma sub-palavra. Este processo se repete até que um número pré-determinado de subpalavras e caracteres seja acrescentado ao vocabulário. A Figura 17 mostra o diagrama contendo o algoritmo.

A escolha deste algoritmo de tokenização para este trabalho se dá pelo fato de que o WordPiece possui uma vantagem significativa em relação a outro algoritmo comumente usado em NLP chamado Byte-Pair Encoding (BPE) (SENNRICH; HADDOW; BIRCH,

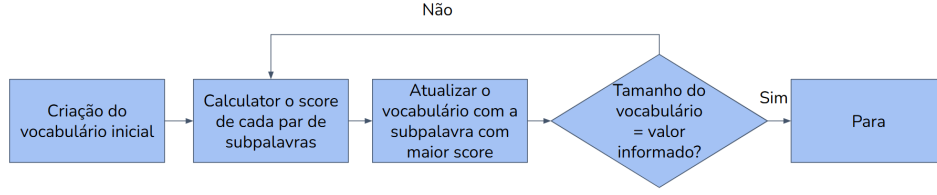


Figura 17 – Diagrama do algoritmo Wordpiece

2016). Ao dividir a frequência de pares de caracteres/subpalavras consecutivos pelo produto de cada parte, o algoritmo prioriza a mescla de pares cujas partes individuais são menos frequentes no texto.

2.5.3 Funções de custo

Durante o treinamento de uma rede neural, a função de custo tem a função de otimizar os parâmetros do modelo minimizando a diferença entre os valores esperados e previstos pelo modelo.

Funções de custo possuem uma série de propriedades como convexidade, diferenciabilidade, suavidade, além de precisarem ser monotônicas. Nesta tese, usamos três funções de custo, a depender da metodologia usada no treinamento. São elas a Focal, a entropia cruzada binária e a Dice.

2.5.4 Entropia cruzada binária

A entropia cruzada binária é uma função de custo amplamente usada em problemas de classificação e segmentação. Seja y_i o valor de referência de uma determinada amostra i e p_i a probabilidade da i -ésima amostra pertencer à classe positiva de acordo com o modelo. A função de custo de entropia cruzada binária L_{BC} é definida pela equação

$$\mathcal{L}_{BC}(y, p) = - \sum_{i=1}^N y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \quad (2.12)$$

Seu papel é calcular a diferença da distribuição dos valores de referência com a distribuição dos valores fornecidos pelo modelo. Durante o treinamento, caso a otimização seja convergente, essa diferença será cada vez menor.

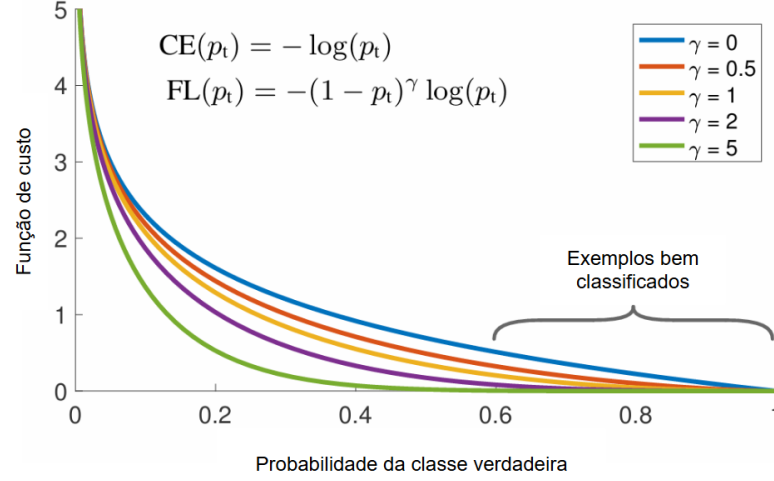


Figura 18 – Valor da função de custo (loss) pela probabilidade da classe de referência (ground truth). Gráfico retirado de (LIN et al., 2018)

2.5.5 Focal

A função de custo Focal é usada para treinar modelos cujos dados são altamente desbalanceados (LIN et al., 2018). Para um determinado valor de referência (*ground-truth*) $y \in \{0, 1\}$, e $p \in [0, 1]$ a estimativa de um modelo para a probabilidade de o valor de entrada pertencer à classe $y = 1$, a função de custo Focal $\mathcal{L}_F$ é definida como

$$\mathcal{L}_F(p_t) = -\alpha(1 - p_t)^\gamma \log(p_t), \quad (2.13)$$

onde $\gamma > 0$ é o parâmetro de foco, α é o parâmetro que equilibra os exemplos positivos e negativos e

$$p_t = \begin{cases} p, & \text{if } y = 1 \\ 1 - p, & \text{do contrário} \end{cases} \quad (2.14)$$

A Figura 18 ilustra a função de custo Focal com diferentes valores para γ , com $\alpha = 1$. Quando $\gamma = 0$, a função coincide com a entropia cruzada binária, portanto sua interpretação é similar à da entropia cruzada. No gráfico, vemos que quanto maior a confiança do modelo em relação à classificação, menor é o valor da função de custo, sempre abaixo do valor para $\gamma = 0$. Isto significa que o fator $(1 - p_t)^\gamma$ diminui a contribuição de exemplos fáceis durante o treinamento e foca rapidamente nos exemplos mais difíceis.

2.5.6 Dice

A função de custo Dice é dada pela equação a seguir:

$$\mathcal{L}_{Dice}(y, p) = 1 - \frac{2 \sum_i y_i p_i}{\sum_i y_i y_i + \sum_i p_i p_i} \quad (2.15)$$

em que as somas são feitas sobre todos os pixels da imagem. Esta função tem como objetivo avaliar o quanto de uma imagem foi corretamente segmentada durante o treinamento do modelo.

2.6 Técnicas de pré e pós processamento

Com o objetivo de melhorar o desempenho dos modelos utilizados nesta tese, é necessário transformar os dados brutos - de entrada ou de saída, após a inferência - para realce de características importantes, remoção de ruído, e melhora do resultado da segmentação. Para isso, são utilizadas técnicas de pré e pós processamento em diferentes etapas do fluxograma de treinamento ou inferência. Nesta seção, serão descritos três métodos: Transformada Wavelet, Operações Morfológicas e Especificação de Histograma.

2.6.1 Transformada Wavelet

Wavelets são um conjunto de funções ortogonais que subdividem uma determinada função em diferentes componentes baseados em frequência e analisam cada componente com uma resolução correspondente à sua escala. O uso de bases ortogonais garante que a representação de um determinado traço sísmico não seja redundante. Possuem várias funções em diferentes áreas, como mecânica quântica e engenharia elétrica. Nesta tese, a Transformada Wavelet funciona como um filtro para remover características das imagens que não são relevantes para a segmentação de gás natural.

Ao contrário da transformada de Fourier, que faz uso de funções seno e cosseno, que se estendem ao infinito, a transformada wavelet consegue aproximar bem funções que possuem picos muito abruptos pois wavelets são funções localizadas no espaço.

Como traços sísmicos são uma série temporal discreta, usamos a Transformada Wavelet Discreta (TWD). Dada uma wavelet mãe $\psi(x)$, podemos definir sua forma dimensionada e transladada como

$$\psi_{m,k}(x) = \psi(2^m x - k) \quad (2.16)$$

De forma semelhante, a wavelet pai $\phi(x)$ possui uma forma dimensionada e transladada definida por

$$\phi_{m,k}(x) = \phi(2^m x - k) \quad (2.17)$$

Qualquer função $f(x)$ pode ser aproximada pela combinação das versões dimensionadas e escaladas das wavelets mãe e pai através da equação

$$f(x) = \sum_{k=-\infty}^{k=\infty} a_k \phi_{m,k}(x) + \sum_{m < n} \sum_{k=-\infty}^{k=\infty} b_{m,k} \psi_{m,k}(x) \quad (2.18)$$

Os coeficientes a_k são conhecidos como os coeficientes de aproximação por armazenar informação sobre características gerais do sinal. Os coeficientes $b_{j,k}$ são os coeficientes de detalhe por capturar detalhes menores e ruído. Em sinais e imagens com um número finito de amostras, as wavelets pai atuam como um filtro passa-baixa, enquanto as wavelets mãe, como um filtro passa-alta.

Cada conjunto de funções ortogonais é chamado de uma família Wavelet. Dentre as diversas famílias wavelet possíveis, a família de wavelets Daubechies se destaca em diversas aplicações, incluindo em análise de sinais sísmicos (Adhikari et al., 2020; AL-HASHMI; RAWLINS; VERNON, 2013). Por este motivo, a família de wavelets Daubechies será usada como pré-processamento para as imagens que servirão para treinar modelos da arquitetura U-Net.

A Figura 19 ilustra as wavelets de alguns conjuntos da família de wavelets Daubechies.

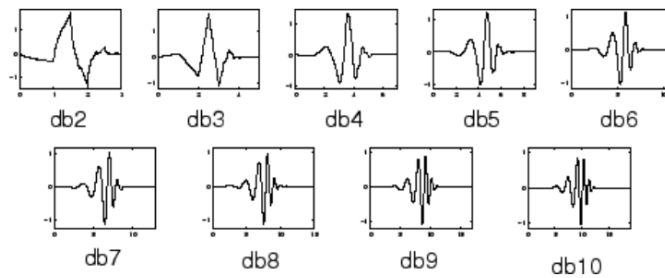


Figura 19 – Wavelets pertencentes à família Daubechies, cuja denominação dbN refere-se à ordem N da wavelet da família Daubechies (db).

2.6.2 Operações morfológicas

Operações morfológicas são normalmente aplicadas como pós-processamento na segmentação de imagens com o intuito de diminuir o número de pixels da classe negativa classificados como positivos e aumentar o número de pixels da classe positiva corretamente classificados. As operações morfológicas são ferramentas úteis para resolver uma grande variedade de problemas em processamento de imagens digitais, e tem como base a Teoria de Conjuntos.

As operações morfológicas consistem em interações entre uma imagem binária A e um conjunto estruturante B , também chamado de elemento estruturante.

Sejam A e B subconjuntos de $\mathbb{Z}^2$. A translação de A por x é denotada de A_x e é definida como

$$A_x = \{c : c = a + x, \text{ para } a \in A\} \quad (2.19)$$

A reflexão de B , denotada por $\hat{B}$, é definida como

$$\hat{B} = \{x : x = -b, \text{ para } b \in B\} \quad (2.20)$$

O complemento de A é A^C e a diferença entre os conjuntos A e B é denotada por $A - B$.

São duas as operações básicas, a *erosão* e a *dilatação*. A dilatação de uma imagem A por um elemento estruturante B é definida como

$$A \oplus B = \{x : \hat{B}_x \cap A \neq \emptyset\} \quad (2.21)$$

O resultado da operação é o conjunto de todos os pontos gerados pela reflexão de B ao redor de sua origem e após transladar essa reflexão de x . Um exemplo dessa operação em uma imagem pode ser encontrado na segunda imagem da Figura 20. No jargão de processamento de imagens, podemos dizer que o elemento estruturante B atua como uma máscara convolucional.



Figura 20 – Na figura à esquerda, a imagem original. No centro, após a aplicação da operação de dilatação na imagem original. À direita, após a aplicação da operação de erosão.

A erosão de uma imagem é definida como

$$A \ominus B = \{x : B_x \subseteq A\} \quad (2.22)$$

O resultado dessa operação é o conjunto de todos os pontos que são alcançados pelo centro de B quando transladados por x no conjunto A . No âmbito do processamento de imagem, isso é equivalente a fazer a imagem perder pixels da borda, como mostra a última imagem da Figura 20.

As duas operações morfológicas básicas discutidas podem ser usadas separadamente ou em conjunto, em paralelo ou em série. Quando usadas conjuntamente e em série, dão origem a duas transformações importantes: abertura e fechamento. A primeira é definida como

$$A \circ B = (A \ominus B) \oplus B, \quad (2.23)$$

ou seja, é uma erosão seguida de dilatação. É útil para remover ruídos de imagens, como indicado na Figura 21.



Figura 21 – Operação morfológica de abertura.

Sua operação dual, o fechamento, é definida como

$$A \bullet B = (A \oplus B) \ominus B, \quad (2.24)$$

e realiza as mesmas duas operações básicas na ordem inversa da abertura, ou seja, dilatação seguida de erosão. Seu intuito no processamento de imagens é, principalmente, fechar falhas e buracos na imagem, como pode ser visto na figura 22.



Figura 22 – Operação morfológica de fechamento.

A Figura 23 exemplifica a aplicação das operações morfológicas definidas pelas equações 2.6.2 e 2.6.2. No exemplo, a imagem A é a segmentação feita pelo modelo e o elemento estruturante B é uma matriz 3x3 cujos elementos são todos iguais a 1. A primeira sequência de imagens é a aplicação da operação de fechamento, e a segunda sequência é a aplicação da operação de abertura baseada no resultado da primeira operação. A operação de fechamento fecha os buracos na segmentação original feita pelo modelo, enquanto que, aplicada em seguida, a operação de abertura reduz as eventuais e esparsas previsões erradas.

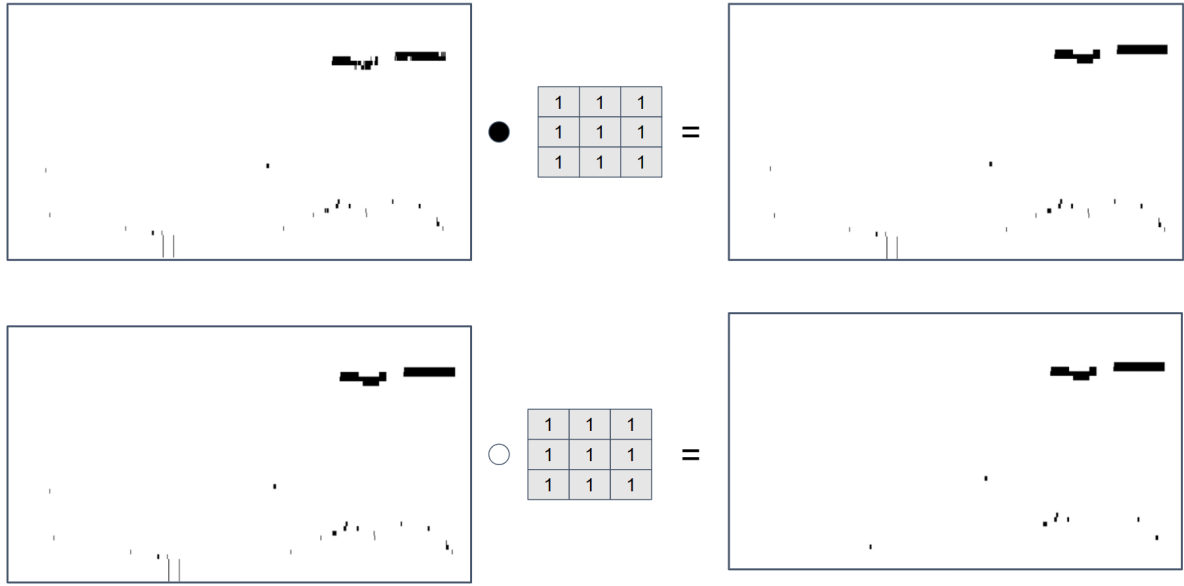


Figura 23 – Operações morfológicas de abertura e fechamento aplicadas à segmentação de gás natural.

2.6.3 Especificação de histograma

Em processamento de imagens, histogramas são usados para visualizar a distribuição de intensidade de pixel de uma imagem. No contexto de segmentação de gás natural, diferentes distribuições de intensidade das imagens usadas para treinar e testar o modelo podem induzir a erros na inferência do conjunto de teste. Para contornar este problema, podemos usar a técnica de especificação de histograma (THOMAS; FLORESTA; PISTORIUS, 2011).

Para uma imagem sísmica de altura H e largura W , contendo L pixels de intensidade distintos, a probabilidade $p_r(r_k)$ de ocorrência de uma intensidade de nível r_k é

$$p_r(r_k) = \frac{n_k}{HW}, \quad k = 0, 1, \dots, L - 1 \quad (2.25)$$

A imagem resultante possui níveis de intensidade de pixel s_k , definidos por

$$s_k = (L - 1) \sum_{j=0}^k p_r(r_j) \quad (2.26)$$

Então, cada pixel com intensidade r_k na imagem original é mapeado a um pixel correspondente de intensidade s_k . Outra imagem com valores de intensidade de pixel z_q são mapeados em s_k pela Equação

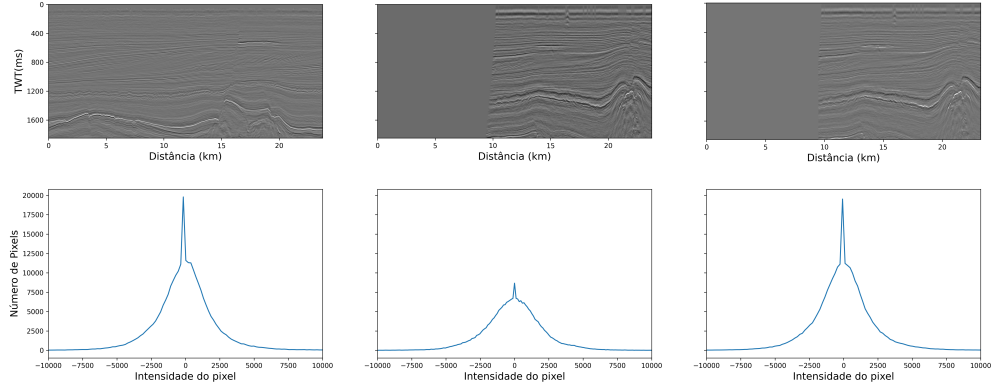


Figura 24 – Especificação de histograma de uma imagem sísmica de reflexão usando outra imagem sísmica como referência.

$$G(z_q) = (L - 1) \sum_{i=0}^q p_z(r_i) \quad (2.27)$$

de modo que para um determinado valor q ,

$$G(z_q) = s_k \quad (2.28)$$

onde $p_z(r_i)$ é o i -ésimo valor do histograma especificado para combinar com a imagem original. Nós podemos obter z_q fazendo a transformação inversa

$$z_q = G^{-1}(s_k) \quad (2.29)$$

O resultado é a troca dos pixels de níveis de intensidade r_k da imagem original por níveis z_q extraídos da imagem de referência. A Figura 24 ilustra como a técnica de especificação de histograma funciona nas imagens sísmicas do conjunto de dados utilizado. Por conta de um problema durante a aquisição ou processamento dos dados sísmicos usados nesta tese, se torna necessário usar a especificação de histograma. Note que o contraste da imagem aumenta após a aplicação da técnica.

2.7 Métricas de Avaliação

Com o intuito de avaliar quantitativamente os resultados obtidos e compará-los com a literatura, usaremos diferentes métricas que servem para quantificar certas características das previsões, como, por exemplo, se o que previu estava ou não correto e

em que quantidade, quantos dos rótulos verdadeiros foram encontrados pelo modelo etc.

A escolha das métricas foi baseada na literatura, sendo utilizadas as seguintes: sensibilidade, precisão e F1 score. A Figura abaixo mostra a matriz de confusão para uma classificação.

Matriz de confusão		Condição prevista	
		Positivo	Negativo
Condição verdadeira	Positivo	Verdadeiro Positivo (VP)	Falso Negativo (FN)
	Negativo	Falso Positivo (FP)	Verdadeiro Negativo (VN)

Figura 25 – Matriz de confusão

Se para uma determinada amostra de classe positiva, sua classificação também for positiva, esta será considerado um verdadeiro positivo (VP); caso seja negativa, será um falso negativo (FN). Por outro lado, se a amostra for negativa mas classificada incorretamente como positiva, será um caso falso positivo (FP); se classificada como negativa, é considerada um verdadeiro negativo (VN).

Denomina-se precisão a taxa de acertos da classe positiva pela quantidade de previsões totais desta classe. A expressão matemática para essa quantidade é definida como

$$PRC = \frac{VP}{VP + FP} \quad (2.30)$$

Seus valores vão de 0, quando o modelo não é capaz de acertar nenhuma amostra da classe positiva, até 1, quando acerta todas as suas previsões para a classe positiva. Para o nosso problema, uma maior precisão está ligada à menor presença de previsões de IDH em regiões que são não-IDH.

A sensibilidade de um modelo é dada por

$$SEN = \frac{VP}{VP + FN} \quad (2.31)$$

ou seja, de todos os exemplos positivos, quantos o modelo foi capaz de prever corretamente. Esta métrica também possui valores entre 0 e 1, onde 0 indica um modelo ruim, que não é capaz de acertar a classe positiva, e 1 para quando nenhum elemento da

classe negativa é erroneamente predito. Para nós, isso significa que uma maior sensibilidade significa que o modelo é capaz de identificar toda a região de IDH como tal.

Por fim, temos o F1-score, que é baseado nas duas métricas anteriores.

$$F1 = 2 \frac{PRC \times SEN}{PRC + SEN} = \frac{2VP}{2VP + FP + FN} \quad (2.32)$$

cujos valores também vão de 0 a 1. Como vemos, esta métrica leva em consideração, ao mesmo tempo, todas as previsões erradas do modelo e também o quão bom ele é em prever a classe positiva.

2.8 Cálculo de erros experimentais

Para cada métrica Y , utilizada para quantificar o desempenho do modelo, contendo N valores para N diferentes realizações experimentais $\{y_1, y_2, \dots, y_N\}$, usamos a definição de média

$$\bar{Y} = \frac{1}{N} \sum_{i=0}^N y_i \quad (2.33)$$

onde y_i é o valor da métrica para a realização experimental i .

O desvio padrão σ é utilizado é definido como

$$\sigma = \sqrt{\sum_{i=0}^N \frac{(y_i - \bar{Y})^2}{N - 1}} \quad (2.34)$$

Por fim, pode-se definir o erro padrão ξ como

$$\xi = \frac{\sigma}{\sqrt{N}} \quad (2.35)$$

Desse modo, podemos escrever o resultado de um determinado experimento, avaliado através da métrica Y como $\bar{Y} \pm \xi$.

2.9 Considerações Finais

Este capítulo versou sobre o arcabouço teórico necessário para a compreensão dos métodos desenvolvidos nesta tese. Foram discutidos os fundamentos da aquisição de dados sísmicos via técnica de reflexão sísmica, com o intuito de estabelecer um melhor entendimento da natureza e característica dos dados analisados. As arquiteturas

de redes neurais Transformer, BERT e U-Net foram descritas, evidenciando suas aplicações em problemas de classificação e segmentação nos domínios de processamento de linguagem natural e visão computacional. Adicionalmente, apresentamos as técnicas de pré e pós processamento, que visam melhorar os dados fornecidos ao modelo - no caso de pré-processamento - e também as segmentações finais - pós-processamento. Por fim, vimos funções de custo usadas como objetivo de treinamento e métricas de avaliação utilizadas para avaliar segmentações. Esta discussão estabeleceu as bases para os métodos desenvolvidos, apresentados nos capítulos subsequentes.

No próximo capítulo, será apresentada uma revisão da literatura, com trabalhos relacionados cujos temas apresentam interseção com o objeto desta tese. Desta forma, buscamos prover uma visão de como se encontra o estado da arte atualmente no que diz respeito à segmentação de acumulações de gás natural até o presente momento e posicionando as contribuições desta tese dentro desse panorama.

3 TRABALHOS RELACIONADOS

Este capítulo apresenta uma revisão sistemática de trabalhos que possuem interseções temáticas com esta tese. Serão analisadas abordagens baseadas em aprendizado profundo aplicadas à segmentação e detecção de gás natural em dados sísmicos, destacando as metodologias empregadas, os resultados alcançados e suas limitações. Também serão examinados trabalhos que abordam problemas correlatos na área de interpretação sísmica automatizada para reforçar a eficácia das metodologias neste contexto.

A análise destes trabalhos pertirá identificar as lacunas metodológicas existentes e justificar as contribuições propostas nesta tese. Ao final deste capítulo, mostraremos como os métodos desenvolvidos nesta pesquisa se relacionam com as soluções existentes e em que pontos iremos avançar no problema de segmentação automática de gás natural em dados de reflexão sísmica.

3.0.1 Aprendizado profundo para segmentação e detecção de gás natural

Pesquisas com dados sísmicos, em especial relacionadas à prospecção a reservatórios de gás natural e petróleo são majoritariamente feitas internamente pelos empreendimentos que lidam com a exploração direta destes recursos. Por isso, a quantidade de trabalhos publicados relativos a este tópico não são tão comuns quanto na área médica ou no georeferenciamento, por exemplo. Ainda assim, diversos estudos podem ser encontrados e, neste capítulo, trataremos de alguns deles.

Em (SANTOS et al., 2019), os autores utilizam uma rede neural LSTM (HOCHREITER; SCHMIDHUBER, 1997) para fazer detecção de IDHs (artefatos em imagens sísmicas que podem indicar gás natural ou petróleo) utilizando o conjunto de dados do bloco F3. Os autores selecionaram uma região de interesse para fazer o treino e os testes do modelo proposto, afim de evitar rótulos de gás que poderiam confundir a rede durante o treinamento. Os autores obtiveram especificidade de 96.87% e sensibilidade de 94.77%, calculadas levando-se em conta a classificação de cada traço e não a segmentação final.

A abordagem de (ZINI; RIZK; AWAD, 2020), faz uso de transferência de aprendizagem durante a etapa de treino do modelo. É usada uma rede *autoencoder* para aprender as características da imagem de maneira não-supervisionada. Em seguida, a parte codificadora tem seus pesos congelados e apenas o decodificador é treinado na tarefa de

segmentação. Por fim, obtém um valor de F1-score de 0.956 na segmentação para dados do bloco F3 (??). Apesar do resultado, as anotações de regiões com gás natural são consideravelmente amplas, o que pode levar à previsão de falsos positivos que não são detectados desta forma por não serem regiões com gás natural, embora estejam marcadas como tal.

Em (CHENIN; BEDLE, 2022) utilizaram-se técnicas de aprendizado de máquina não-supervisionado, juntamente com diferentes atributos físicos extraídos diretamente dos traços sísmicos, para diferenciar reservatórios com gás de baixa densidade, que não são economicamente viáveis, dos de alta densidade, que normalmente são explorados comercialmente. No trabalho, são utilizados dados da bacia do Golfo do México e da costa da Austrália. O método, porém, não generaliza bem para levantamentos diferentes, isto é, no caso do artigo, um modelo treinado nos dados do Golfo do México não apresenta um bom desempenho nos dados provenientes da Austrália e vice-versa. Os autores argumentam que isso pode ter relação com a diferença de saturação de fluídos entre as duas regiões e história de enterramento de matéria orgânica.

Uma arquitetura composta inteiramente de perceptrons, denominada MLP-Mixer, foi utilizada para a delimitação de uma região de interesse em imagens sísmicas para a posterior aplicação de uma U-Net na etapa de segmentação de gás natural (CIPRIANO et al., 2022). O método apresenta um bom desempenho, com F1-score de 80%, sensibilidade de 82,7% e precisão de 80,4% no conjunto de teste da base utilizada.

Em (DINIZ et al., 2022b), a arquitetura Transformer é usada para a segmentação de reservatórios de gás em uma base de dados privada. A ideia dos autores é utilizar o método Time2Vec (KAZEMI et al., 2019) para criar um *embedding* dos traços sísmicos que serão passados para a rede. O método requer a criação de uma região de interesse previamente delimitada por especialistas, de modo a não comprometer o desempenho da rede com informações de áreas que não são de interesse econômico da companhia. Os resultados para diferentes campos de exploração em uma mesma região, apresentados com relação aos segmentos de traços classificados, ao invés da segmentação final, são promissores, porém apresentam uma variação de valores de métricas considerável entre si.

Outra abordagem como LSTM convolucionais e Otimização por Enxame de Partículas (PSO, em inglês) foi descrita em (JÚNIOR et al., 2023). Os autores fazem a segmentação de patches ao invés de utilizar a imagem inteira. Para um determinado

patch, outros ao seu redor são selecionados e empilhados, de modo que a entrada para a rede proposta possui múltiplos canais (o patch a ser segmentado e patches ao seu redor). A PSO foi utilizada para escolha da altura e largura ótimas dos patches, além da quantidade de patches ao redor que se deveria extrair. No conjunto de testes, o método apresentado apresentou um F1 score de 84.22%, sensibilidade de 98.06% e precisão de 84.22% em um conjunto de dados privado.

Um método para selecionar de maneira mais assertiva um dado modelo para segmentar gás natural em dados de regiões diferentes também foi proposto em (SEPULVEDA et al., 2023). Ao invés de usar um único modelo para inferir novos dados, os autores proõem o treinamento de diversos modelos, cada um com um conjunto de dados similares em geologia do local, intensidade dos pixels, contraste e outras propriedades das imagens extraídas por um autoencoder. Cada um desses grupos é criado a partir de algoritmos não-supervisionados de clusterização. Sempre que a inferência nova imagem sísmica é necessária, descobre-se o cluster com imagens mais parecidas com a nova imagem e usa-se o modelo treinado especificamente para este cluster. Com isso, há um aumento de até 10% na sensibilidade e 7% na precisão quando comparado ao caso de um único modelo treinado com todos os dados disponíveis.

O problema de generalização na segmentação de gás natural foi tratado por (SANTOS et al., 2025) através da implementação de um método de aprendizado auto-supervisionado que usa Transformer generativas pré-treinadas para a extração de características das imagens e posteriormente é usada uma rede neural recorrente para a análise dos traços sísmicos. O F1-score no conjunto de dados utilizado aumenta em até 20% em comparação com a rede recorrente sem as características extraídas pela rede Transformer.

3.0.2 Redes neurais profundas aplicadas em dados sísmicos

Além da aplicação direta de aprendizado profundo para a segmentação de gás natural, também é possível aplicar soluções baseadas em aprendizado profundo em outros problemas que usam dados sísmicos.

Em (ZHANG et al., 2020b), os autores utilizam duas redes neurais convolucionais inspiradas na Segnet (BADRINARAYANAN; KENDALL; CIPOLLA, 2016) e na U-Net (RONNEBERGER; FISCHER; BROX, 2015) para a segmentação de fácies sísmicas, que analisa a sequência estratigráfica do ambiente sedimentar e providencia uma base para a

previsão de áreas favoráveis ao surgimento de reservatórios. Aqui os autores aumentaram os dados através de rotações em diferentes ângulos, inversão, adição de ruído gaussiano e sal-e-pimenta, além de borramento, afim de ter um número maior de amostras para treino, apesar da U-Net ter sido projetada para situações onde o número de imagens disponíveis é pequeno.

Ambas as redes são treinadas usando dois subconjuntos diferentes dos dados, resultando em quatro modelos. Estes são usados em conjunto para formar um ensemble e fazer a classificação. Os autores argumentam que dessa forma - ao utilizar um ensemble - os erros de classificação e as descontinuidades nas bordas de cada fácies diminuem, melhorando o desempenho do método. No entanto, os autores não fazem comparações com outros trabalhos, não ficando claro em que aspecto o modelo proposto por eles é melhor que os já existentes. Além disso, utilizam apenas a acurácia como métrica, o que pode ser enganoso, já que principalmente para conjuntos de dados com classes desbalanceadas, essa não é uma métrica adequada.

Observando uma abordagem um pouco diferente da baseada em imagens, a eficácia de modelos de linguagem, BERT em particular, no tratamento de dados sísmicos pode ser notada em (HARSUKO; ALKHALIFAH, 2022). Os autores fazem uso da arquitetura BERT como uma forma de armazenar características de um conjunto de dados sísmicos para tarefas subsequentes. Cada traço sísmico é considerado como sendo uma palavra, e o pré-treinamento é feito através das duas tarefas explicadas na seção 2.5.1. O modelo é depois treinado em outras tarefas como a remoção de ruído e estimativa de velocidade.

No contexto de *deep learning*, é sabido que, em geral, inferências *zero-shot* (isto é, sem um nova etapa de treinamento) de um determinado modelo em domínios diferentes do qual ele foi originalmente treinado, possui desempenho aquém do observado no domínio fonte. Para contornar esse problema, em (MA et al., 2019) é utilizada uma abordagem em duas partes: na primeira, o treino é realizado com dados de domínios diferentes com rótulos de domínio. No segundo passo, um pequeno subconjunto de dados do domínio fonte é selecionado com base no classificador de domínio para o processo de ajuste fino de parâmetros. Através deste método, os autores conseguiram melhorar a transferência de aprendizado para o domínio alvo em comparação com o estado-da-arte, e esta abordagem pode ser usada como ponto de partida na etapa de adaptação de domínio no decorrer desta pesquisa.

Com o objetivo de monitorar impactos de terremotos em tempo real, o trabalho

de (WANG et al., 2025) utiliza um *embedder* multi-escala, *patching* latente e blocos pré-treinados da arquitetura GPT-2 (RADFORD et al., 2019). O modelo teve um desempenho estado-da-arte em seis tarefas diferentes relacionadas a monitoramento sísmico: estimativa de azimute reverso, estimativa de distância epicentral, estimativa de magnitude, escolha de fase e classificação de polaridade de primeiro movimento. Denominado de SeisMoLLM, o método teve um desempenho particularmente superior aos outros modelos usados como base de comparação em tarefas relacionadas a localização e magnitude, mostrando a robustez de LLMs em tarefas relacionadas a séries temporais, e em particular, dados sísmicos.

3.0.3 Considerações finais

A Tabela 29 traz um resumo dos trabalhos relacionados mencionados neste capítulo, apontando os autores e as técnicas utilizadas.

Para dois dos métodos propostos nesta tese, a ideia de usar o traço todo como uma palavra, o paradigma criado em (HARSUKO; ALKHALIFAH, 2022) não é viável pois resultaria numa classificação sobre todo o traço sísmico, o que removeria a localização temporal da região de gás natural. Os autores, no entanto, mostram que o uso do BERT para problemas envolvendo dados sísmicos possui um bom desempenho, sendo um indicativo para que abordagens similares a essa podem funcionar na detecção de gás natural e outros hidrocarbonetos.

Por outro lado, o trabalho de (DANG et al., 2021) é um ponto de partida para dois métodos desenvolvidos nesta tese. Ao contrário da aplicação em séries temporais das bases de dados KPI e Yahoo, usadas pelos autores, considerar cada amostra como sendo uma palavra diferente introduz uma granularidade muito baixa no manejo de dados sísmicos. Para identificar visualmente uma região de IDH, como explicado no capítulo 2, observam-se as estruturas geológicas presentes no dado, então é natural agregar os sinais de modo com que eles representem, ainda que vagamente, tais estruturas.

Os estudos descritos neste capítulo avançaram o conhecimento sobre a aplicação de diversos métodos e arquiteturas de deep learning na segmentação de gás natural e petróleo e dados sísmicos de forma geral. No entanto, possuem algumas limitações. A maior limitação para o estudo de problemas que envolvem dados sísmicos, independentemente de tarefa (segmentação, detecção etc) ou objeto de estudo (gás, petróleo, domos de sal

etc) é a falta de dados públicos disponíveis, em especial aqueles com anotações para determinado problema. Devido à natureza econômica das atividades que fazem uso destes dados, poucas empresas disponibilizam seus dados coletados e anotados para a comunidade científica. Essa escassez justifica, de certa forma, a escassez de estudos nessa área do conhecimento e também, dos estudos, a pouca variedade que utiliza dados abertos. Nesta tese, utilizaremos uma base pública anotada por um geocientista, cujas anotações serão disponibilizadas à quem requisitar.

Isto posto, um dos objetivos de se utilizar aprendizagem de máquina para a segmentação de reservatórios de gás natural é utilizar o modelo já treinado em novas regiões que ainda não possuem anotações para determinar seu potencial econômico através da extração do recurso natural. Os estudos desenvolvidos até o momento de escrita desta tese carecem de um estudo mais detalhado sobre a generalização dos modelos desenvolvidos.

Um outro ponto a ser destacado é a ausência, até o momento da publicação da pesquisa desenvolvida no período de doutoramento, da aplicação direta de modelos de Grandes Modelos de Linguagem - *Large Language Models* (LLM) - na segmentação de hidrocarbonetos em imagens sísmicas, em particular gás natural, apesar de todo o histórico recente de resultados estado-da-arte nas mais diversas aplicações.

No estudo desenvolvido e relatado nesta tese, propomos um modelo de segmentação de imagens sísmicas baseado em um LLM e um outro baseado em uma arquitetura utilizada em problemas de visão computacional. Estudamos a capacidade de generalização dos modelos treinados com uma base de dados pública limitada, sendo esta base de dados amplamente usada em estudos que não apresentam seus resultados em bases de dados privadas.

Tabela 1 – Referências e respectivos métodos dos trabalhos relacionados a este projeto.

Trabalhos relacionados	Métodos desenvolvidos	Algoritmos	Resultados
(MA et al., 2019)	Classificação de domínio e seleção de dados	BERT	-
(SANTOS et al., 2019)	Identificação de IDHs	LSTM	SEN = 0.97 auc= 0.988
(ZHANG et al., 2020b)	Segmentação de fácies sísmicas	CNNs	Acurácia = 0.96
(ZINI; RIZK; AWAD, 2020)	Autoencoder para pré-treinamento de modelo e treino do decodificador para segmentação	Autoencoder	Acurácia = 0.957
(DANG et al., 2021)	BERT para detecção de anomalias em séries temporais	BERT	
(XU et al., 2021)	Detecção de anomalias em séries temporais	Transformer	
(HARSUKO; ALKHALIFAH, 2022)	Remoção de ruído e estimativa de velocidade	BERT	—
(CIPRIANO et al., 2022)	Detecção de gás natural	MLP-Mixer e U-Net	F1 = 0.8418 SEN = 0.8685 Acurácia = 0.996
(DINIZ et al., 2022b)	Detecção de gás natural	Transformer e Time2vec	SEN = 0.8685 Acurácia = 0.996
(SEPULVEDA et al., 2023)	Generalização da inferência de detecção de gás	Autoencoder e K-Means	F1 = [0.17, 0.54] SEN = [0.19, 0.73]
(Júnior et al., 2023)	Segmentação de gás natural	LSTM Convolucional e PSO	F1 = 0.8422 SEN = 0.9806 Acurácia = 0.9942
Método Proposto	Segmentação de gás natural e estudo sobre a generalização para outros reservatórios	RoBERTa e U-Net	

4 MATERIAIS E MÉTODOS

O capítulo anterior apresentou uma revisão dos trabalhos relevantes que servem como referência e ponto de partida para as contribuições desta tese. Neste capítulo, descreveremos a base de dados utilizada e os métodos desenvolvidos para a segmentação de gás natural em imagens sísmicas de reflexão.

Desenvolvemos três abordagens distintas cuja nomenclatura é baseada na quantidade de informação disponibilizada ao modelo para fazer a segmentação. O primeiro método proposto é uma abordagem bidimensional (2D), que consiste na incorporação de técnicas para realçar características de interesse para a segmentação de gás natural bem como alcançar a estabilidade do desempenho frente à mudança de pesos iniciais no treinamento. O segundo método uma abordagem unidimensional (1D) e adapta sinais sísmicos para uma rede neural de arquitetura RoBERTa - baseada em Transformers -, e utiliza informação contida apenas no segmento de sinal para classificá-lo em gás e não-gás. O terceiro método proposto é uma variação do primeiro método que também faz uso da informação em traços próximos daquele que se quer classificar, ou seja, é uma abordagem intermediária em comparação com os outros dois métodos, e denominada 1.5D.

Todos os métodos propostos foram avaliados no conjunto de dados sísmicos do bloco F3, nos Países Baixos. As seções subsequentes descrevem os dados desse bloco, bem como as explicam detalhadamente cada um dos métodos utilizados.

4.1 Base de dados

Dada a natureza comercial que envolve a extração de gás natural, a maior parte dos dados estão sob a posse de empresas privadas. Desta forma, optamos por utilizar neste trabalho a base de dados do bloco F3, localizado nos Países Baixos como indicado na figura 26, que está disponível no Open Seismic Repository dGB Earth Sciences (SILVA et al., 2019b). O bloco consiste em um único volume 3D que contém os dados de reflexão sísmica de 384 km^2 no setor neerlandês do Mar do Norte, onde quatro poços foram perfurados em diferentes localidades. Os dados são migrados temporalmente e possuem um intervalo temporal de 1848 ms , a uma taxa de amostragem de 4 ms e resolução espacial de 25 m . As anotações de *ground-truth* usadas para treinar os modelos foram feitas por um geofísico, cuja interpretação usou como base os *bright spots* que aparecem nas imagens,

além dos dados adicionais disponibilizados no repositório, como a impedância acústica do volume e perfilagem de poços com dados como porosidade, densidade, impedância etc.

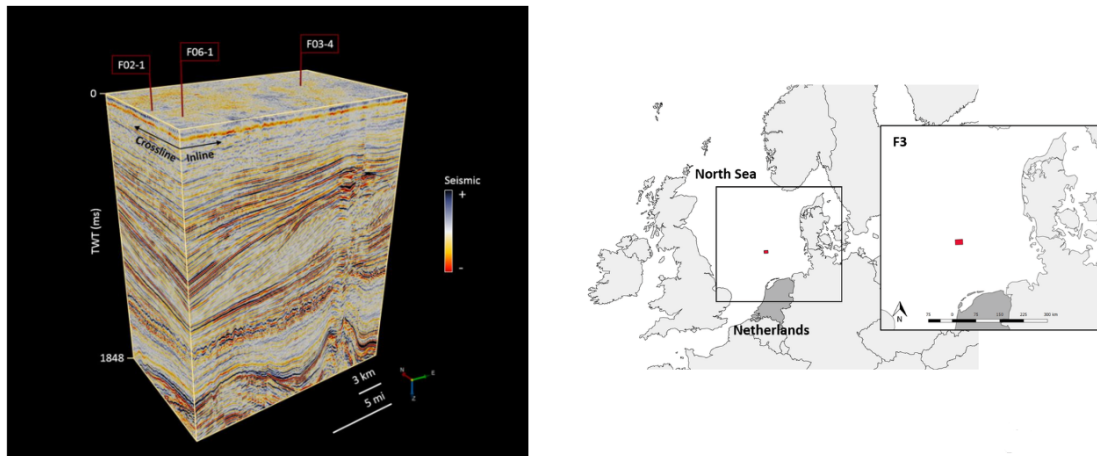


Figura 26 – Bloco F3 e sua localização geográfica. Imagem retirada de (SILVA et al., 2019a).

4.2 Método 1: Abordagem 2D - Ensemble de U-Net para segmentação de imagens

A U-Net foi escolhida como modelo para segmentação e estudo por ser um método utilizado em diversos estudos e aplicações apresentando um bom desempenho. Sua arquitetura também é simples o suficiente para que seja conduzido um estudo sobre generalização e o que é aprendido por cada camada da rede de maneira intuitiva, o que pode levar a insights sobre o comportamento de arquiteturas voltadas para Visão Computacional no contexto de imagens sísmicas. O método possui três etapas: pré-processamento, treinamento e inferência no conjunto de testes. No pré-processamento, é feito um realce de estruturas relevantes para o problema, em que as imagens são redimensionadas, e têm seus pixels discretizados, sobre os quais é aplicada a transformada wavelet discreta. Em seguida, um limiar de valores é feito a fim de eliminar ruído e partes da imagem sísmica que não são relevantes para a tarefa de segmentação de reservatórios de gás natural, efetivamente atuando como um filtro. O treinamento é feito com as imagens filtradas e, após esta etapa, é feita a inferência nas imagens sísmicas do conjunto de teste. A ilustração da pipeline se encontra na Figura 27.

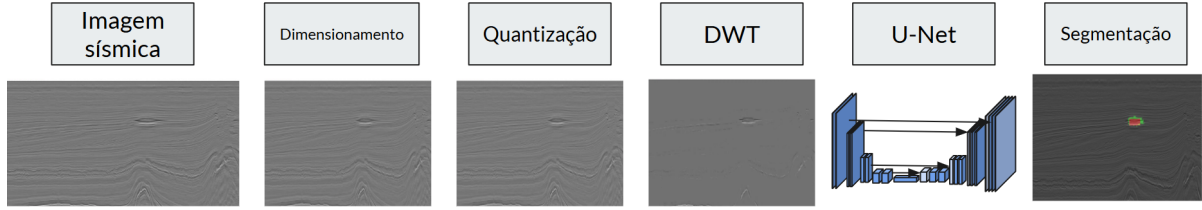


Figura 27 – Pipeline da segmentação com a arquitetura U-Net. Imagem de elaboração própria.

4.2.1 Pré-processamento e treinamento

Empiricamente, é verificado e demonstrado, durante a discussão dos resultados, que a quantização melhora os resultados da segmentação quando alinhada à DWT e ao pós processamento de ensembling. Uma possível explicação para isso é que, por conta da natureza dos dados contínuos e da similaridade entre diferentes regiões da imagem, regiões similares podem ser entendidas como diferentes pelo modelo, tendo em vista a discrepância entre as amplitudes causada por fatores geofísicos e de processamento da imagem sísmica.

Transformamos as amplitudes de tal modo que para uma dada amplitude a , temos $a \in \mathbb{N} | a < 20$. Foi verificado empiricamente, através de experimentos com $a \in \{10, 20, 30, 50, 80\}$, que 20 é o valor que resulta em um melhor desempenho do método.

Após a quantização, aplicamos a DWT, conforme discutido na seção 2.6.1, utilizando uma função de onda da família de wavelets Daubechies por vez. Durante os experimentos, diversas funções da família Daubechies foram usadas. Um limiar é aplicado aos coeficientes da transformada para que as partes cujo sinal é mais fraco - e, portanto, com chances ínfimas de pertencerem a uma região contendo gás natural - sejam removidas durante a reconstrução da imagem para o treinamento.

Após o pré-processamento, a rede é treinada em batches com as imagens redimensionadas e pré-processadas, com data augmentation para espelhar a imagem verticalmente afim de evitar sobreajuste dos dados, visto que as imagens usadas no treinamento são altamente correlacionadas.

4.2.2 Inferência

Foi observado empiricamente que um único modelo U-Net é instável perante diferentes inicializações - a cada novo treinamento da arquitetura, com mesmos dados de treino e validação, o modelo resultante se comportava diferente com relação às segmen-

tações produzidas. Para lidar com esta questão, utilizamos um conjunto (ensemble) de U-Nets. Para avaliar experimentalmente o impacto do uso do Ensemble e compará-lo com modelos de U-Net única, a inferência é feita de duas formas. A primeira consiste na inferência de uma U-Net única, onde cada inline é pré-processada e fornecida para a U-Net realizar a segmentação. A segunda forma considera segmentações individuais de diversas U-Nets treinadas com os mesmos hiperparâmetros, porém com pesos inicializados diferentemente uns dos outros. Uma única inferência é feita considerando um conjunto (ensemble) de 10 modelos diferentes de U-Net. Cada pixel é avaliado de acordo com a Equação 4.1:

$$\hat{y} = \arg \max_{v \in \{0,1\}} \sum_{i=1}^{10} \mathbb{I}(M_i(x) = v) \quad (4.1)$$

onde $\hat{y}$ é a classificação de um único pixel da imagem, $M_i(x)$ é a previsão da i -ésima U-Net com a entrada x e $\mathbb{I}(\cdot)$ é a função indicadora, que é 1 se o argumento é verdadeiro e 0 se é falso. Ao fazer isso, espera-se que o conjunto de U-Nets forneça uma segmentação melhorada ao remover previsões positivas falsas e, ao mesmo tempo, manter as previsões positivas verdadeiras da maior parte dos modelos pertencentes ao conjunto.

Para evitar tempo extra de computação no treinamento durante a criação dos ensembles, para cada conjunto de hiperparâmetros - família wavelet, limiar da DWT etc - optamos por treinar 30 U-Nets diferentes. Cada ensemble é feito escolhendo aleatoriamente 10 destas 30 U-Nets. Este processo é feito para simular o treinamento de 10 U-Nets diferentes sempre que for necessário criar um ensemble. A Figura 28 ilustra este método de escolha de U-Nets para a formação de ensembles.

4.3 Método 2: Abordagem 1D - RoBERTa para classificação de traços

Os dados passam por duas etapas de pré-processamento: quantização (ou discretização) dos valores dos pixels das imagens sísmicas e tokenização. Dados sísmicos são, por natureza, dados de amplitude contínua. No entanto, modelos de linguagem, como o RoBERTa, lidam apenas com *entidades* discretas - caracteres e palavras. Para usarmos técnicas de codificação de sequências em dados sísmicos, foi necessário, portanto, discretizar os valores dos traços sísmicos que compõem as imagens sísmicas. Em seguida, criamos tokens baseados nas amplitudes e sequências de amplitudes com o algoritmo TracePiece. Por fim, o modelo foi treinado na tarefa de *Masked Language Modeling* e posteriormente

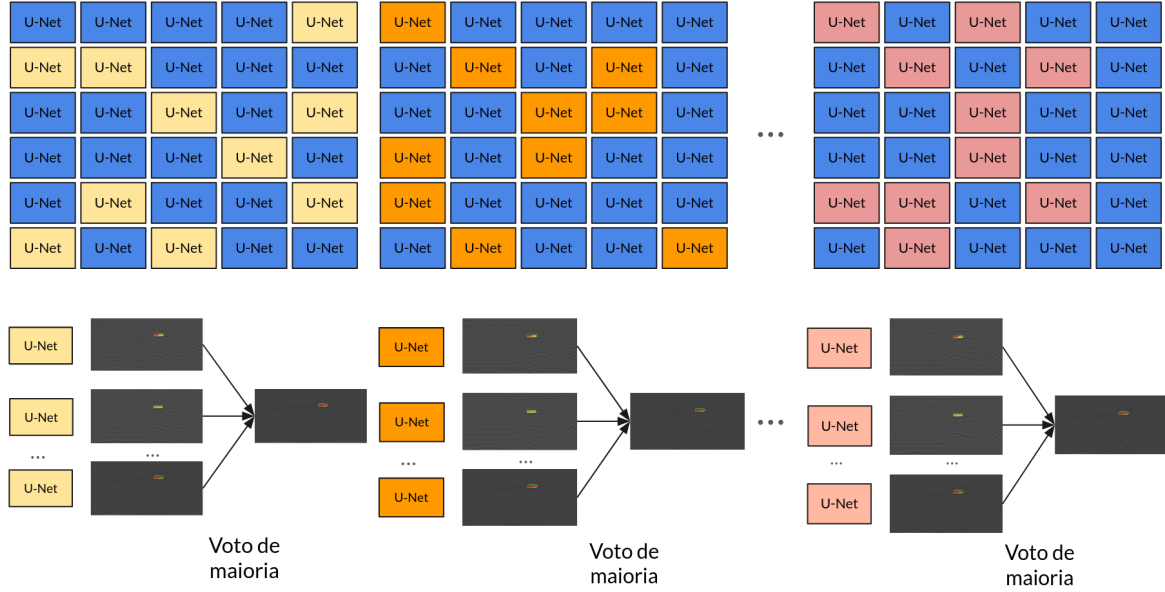


Figura 28 – Processo de escolha de ensemble. Imagem de elaboração própria.

treinado na tarefa de classificação de traços.

4.3.1 Pré-processamento

Um sistema de escrita é baseado em caracteres ou símbolos que representam um som ou grupo de sons. Isso implica que existe um número finito de caracteres em qualquer sistema de escrita. Para simular esta característica linguística, o primeiro passo necessário para usar um modelo de linguagem natural é quantizar (ou discretizar) os objetos que fornecemos de entrada ao modelo. Um traço sísmico é composto por um conjunto de amplitudes contínuas $a \in \mathbb{R}$. Nós quantizamos essas amplitudes de maneira uniforme de tal forma que elas se tornem números pertencentes ao conjunto $a \in \mathbb{N} | a < 50$, isto é, as amplitudes se tornam números inteiros que vão de 0 a 49, resultando em 50 amplitudes diferentes. Assim como no Método 1, neste Método foram experimentados diferentes valores para a constante de quantização $a \in \{10, 20, 30, 50, 70, 80\}$, e, para este método, o valor 50 forneceu melhores desempenhos no decorrer dos experimentos.

Em seguida, realizamos a etapa de tokenização. Essa etapa é central no processamento de linguagem natural, pois converte frases em formas que os algoritmos são capazes de processar. Atualmente existem diversos algoritmos de tokenização para linguagem. Destes, os mais populares são o WordPiece (WU et al., 2016) e o Byte-Pair Encoding (BPE) (SENNRICH; HADDOW; BIRCH, 2016). Em (TORAMAN et al., 2023), é demonstrado que para as linguagens aglutinativas estudadas, isto é, linguagens em que as

palavras são unidas para formar outras palavras, estes dois algoritmos de tokenização possuem melhor desempenho, com o tokenizador WordPiece possuindo um desempenho ligeiramente melhor que o BPE.

O objetivo é simular uma linguagem em um domínio em que sequências de sinais não possuem um significado *a priori*, como uma sequência de caracteres combinados formam uma palavra. Sendo assim, um algoritmo baseado no WordPiece parece ser o melhor método para simular uma linguagem, pois ele pode, em teoria, separar o traço sísmico em pedaços com algum significado. Por exemplo, a frase “Reservatórios de gás são difíceis de segmentar” pode ser dividida em “ Reserv -atórios de gás são difíceis de seg -mentar”. Aqui, o hífen indica uma sub-palavra encontrada dentro da palavra original. A sub-palavra “Reserv” pode ser encontrada em palavras como “reservado”, “reservas” e “reservatórios”. Ao dividir palavras em pedaços menores, modelos de linguagem podem aprender a semelhança entre palavras parecidas i.e. palavras que compartilham pedaços, ao invés de considerá-las palavras completamente independentes.

Séries temporais são parecidas com texto em linguagem natural no sentido de que ambas possuem unidades fundamentais que podem ser unidas para criar formas mais complexas: no domínio de linguagem natural, caracteres podem ser unidos em uma sequência e formar palavras, que por sua vez podem ser unidos para formar frases; no domínio de séries temporais, amplitudes quantizadas individuais, a que chamaremos t-caracteres, podem ser unidas para formar t-palavras de tamanho l_w , que por sua vez podem formar t-frases de tamanho l_s . Para mimetizar a tokenização WordPiece, a tokenização TracePiece divide um traço sísmico em pedaços pequenos, a que chamaremos de t-subpalavras, para preservar pares ou sequências de amplitudes que se repetem com frequência de acordo com uma fórmula. No método 1.5D, estenderemos a analogia também para o conceito de parágrafo, que é um conjunto de frases. No domínio de séries temporais, o chamaremos de t-parágrafo. A Tabela 2 inclui os termos originais usados no domínio sísmico e suas analogias ao domínio de PLN.

O algoritmo TracePiece segue um padrão similar ao WordPiece, discutido na seção 2.5.2. Criamos inicialmente um vocabulário com cada t-caractere, isto é, com cada amplitude quantizada. Neste trabalho em particular, usamos 50 amplitudes diferentes. Para cada par de t-caracteres a e b presentes no vocabulário, o valor de um score é calculado através da Equação 2.11:

Tabela 2 – Correspondência entre termos baseados no domínio de linguagens e suas definições originais no domínio sísmico.

Termo			Definição
Caractere caractere)	temporal	(t-	Amplitudes individuais quantizadas de um traço sísmico
Palavra temporal (t-palavra)			Sequência de amplitudes quantizadas de tamanho arbitrário l_w .
Sub-palavra subpalavra)	temporal	(t-	Pedaços de sequências de amplitudes quantizadas de tamanho arbitrário l_{sw} , tal que $l_w \leq l_{sw}$, obtidas pelo método de tokenização TracePiece.
Frase temporal (t-frase)			Sequência de t-palavras de tamanho l_s .
Parágrafo parágrafo)	temporal	(t-	Sequência de t-frases que contém l_p frases.
Vocabulário			Número de t-palavras e t-subpalavras que o modelo pode reconhecer.

$$score(a, b) = \frac{f_{ab}}{f_a f_b}$$

O par com maior score é adicionado ao vocabulário e agora é considerado uma t-subpalavra e, dentro do algoritmo TracePiece descrito, será considerado um t-caractere com o propósito de calcular seu score com outros t-caracteres. Esse processo se repete até que um tamanho de vocabulário predefinido seja atingido. Uma vez que tenhamos um vocabulário completo, cada t-palavra é dividida em t-subpalavras. Por fim, a cada t-subpalavra é designado um número. A Figura 29 mostra o processo de quantização de um traço sísmico. No exemplo, usamos $l_s = 7$ e $l_w = 5$. Na Figura 30 todo o pré-processamento é ilustrado com $l_s = 5$ e $l_w = 3$. Depois da quantização, a t-frase é criada, contendo $l_s \times l_w$ amplitudes quantizadas individuais, ou t-caracteres. A t-frase é então dividida em pequenos pedaços numerados através do algoritmo TracePiece, como descrito no parágrafo anterior. Se o tamanho da frase for reduzido de l_s para um número menor, um padding é feito para que o tamanho da entrada do modelo seja sempre $l_s \times l_w$.

4.3.2 Pré-treinamento

O objetivo do pré-treinamento em um modelo de linguagem é fazer com que ele aprenda padrões nos dados que serão úteis em diferentes tarefas envolvendo os mesmos dados. Por exemplo, o modelo pré-treinado usado nesta tese pode ser utilizado como

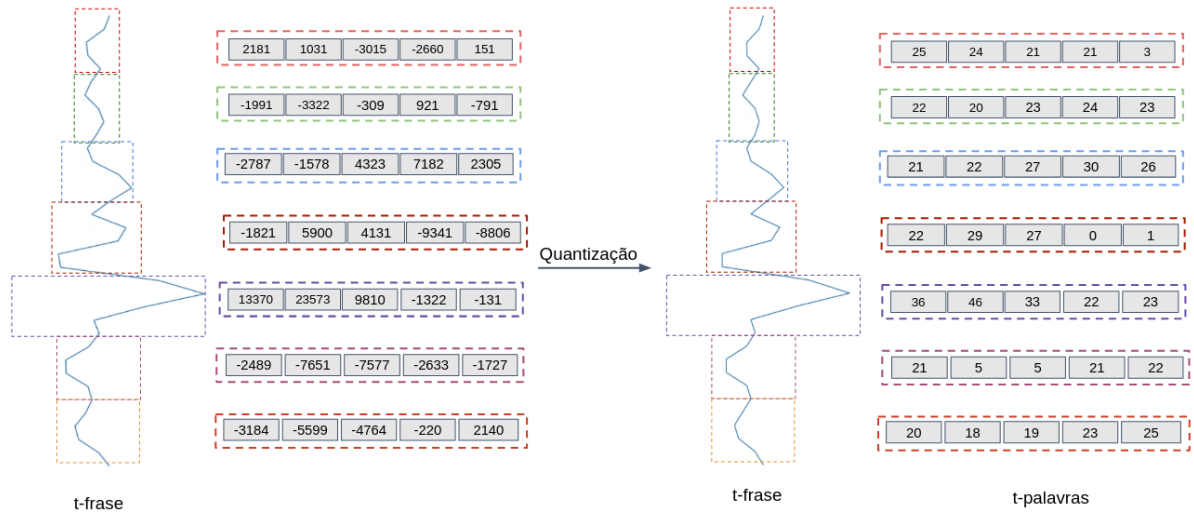


Figura 29 – Quantização de traços sísmicos.

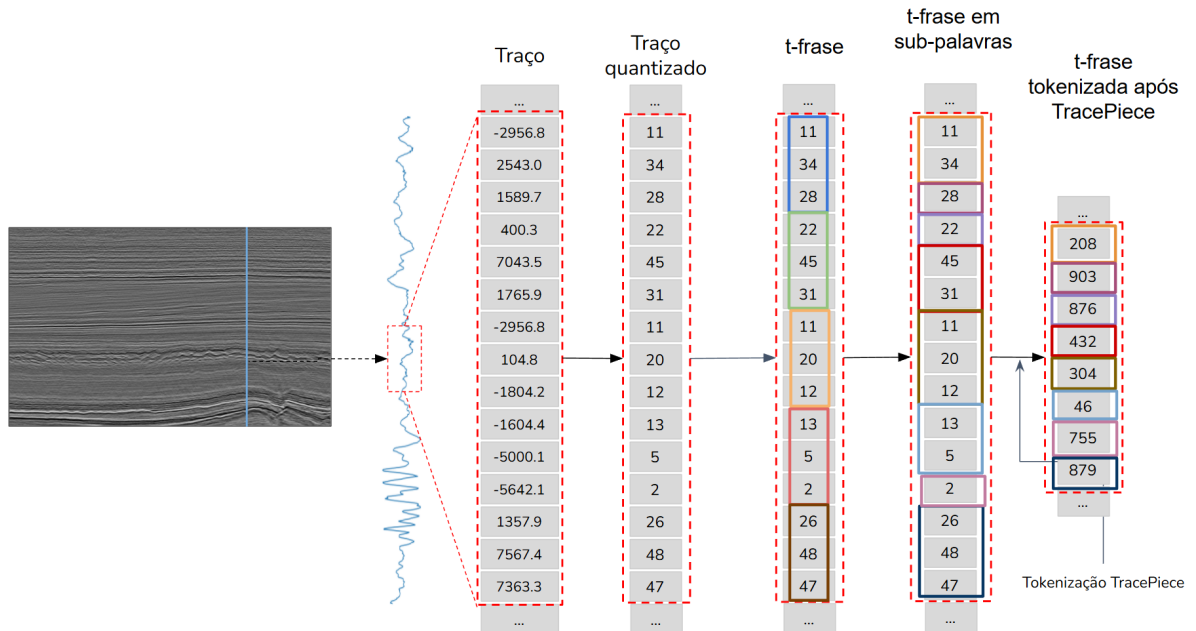


Figura 30 – Esquema de pré-processamento de um sinal sísmico no método proposto.

ponto de partida pro treinamento de modelos em tarefas como segmentação de fácies sísmicas, segmentação de gás natural, segmentação de domos de sal, entre outras tarefas envolvendo imagens de reflexão sísmica. Ainda que essas aplicações sejam diferentes, a relação entre as amplitudes para um dado conjunto de dados é a mesma.

Na etapa de pré-treinamento, frases com máscaras são criadas a partir das t-frases tokenizadas originais. O processo de mascaramento é feito de acordo com o procedimento discutido em (DEVLIN et al., 2018). Para cada t-frase, 12% de todos os tokens são substituídos por um token de máscara e escondidos da rede neural. Outros 1.5% são substituídos por um token aleatório que esteja presente no vocabulário. Esse processo é

necessário para forçar os Transformers que compõem o RoBERTa a manter uma distribuição da representação contextual de todos os tokens de entrada. Em (DEVLIN et al., 2018), um estudo usando proporções diferentes de tokens com máscara e aleatoriamente inseridos é feito, e as porcentagens mencionadas aqui são as melhores para pre-treinar modelos BERT. Em seguida, RoBERTa é treinado a prever quais são os tokens escondidos, como indicado na Figura 31. Como o modelo apenas recebe as t-frases como entrada na etapa de pré-treinamento, sem notação alguma, ele vai aprender como as t-palavras e t-subpalavras se relacionam umas com as outras e em qual contexto elas aparecem, com o objetivo de prever o token escondido.

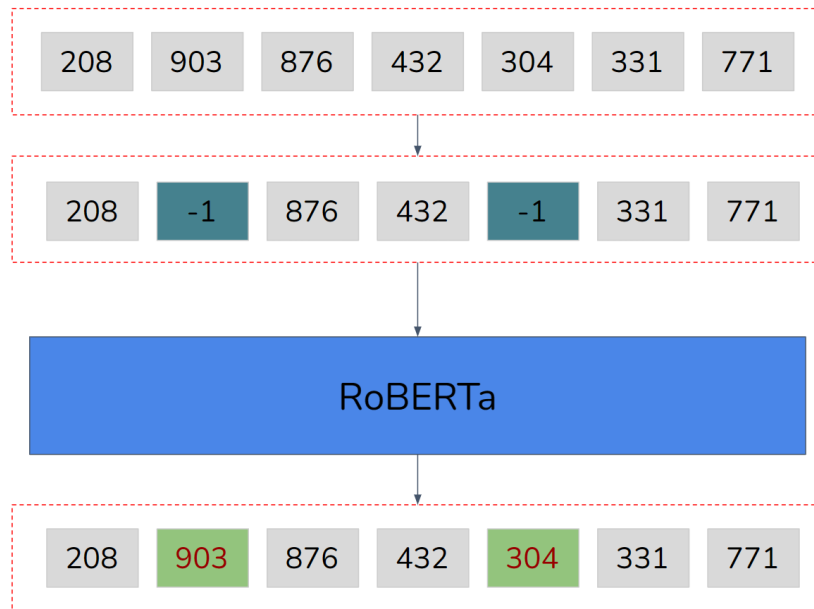


Figura 31 – Processo de pré-treinamento. Cada t-frase tokenizada possui alguns de seus tokens substituídos por tokens de máscara, representados por tokens -1, e o modelo aprende a prever qual o número do token escondido (mascarado). Esse procedimento faz o modelo aprender a relação entre tokens diferentes antes da tarefa de classificação realizada no treino supervisionado da rede.

4.3.3 Treinamento

Para anotar as t-frases de cada traço, são selecionadas l_s sequências de t-palavras, cada t-palavra de tamanho l_w . As t-frases são anotadas usando um método de janela deslizando, como ilustrado na Figura 32. Para que seja considerado como uma t-frase de gás, o token anotado como gás precisa estar dentro de uma das palavras que pertencem a essa t-frase. Esse processo equivale ao alargamento das anotações de gás, designadas inicialmente a uma única amplitude no traço sísmico (e portanto, um único t-caractere).

Efetivamente, esse alargamento aumenta o número de amostras anotadas como gás usadas para treinar o modelo de linguagem na tarefa de classificação.

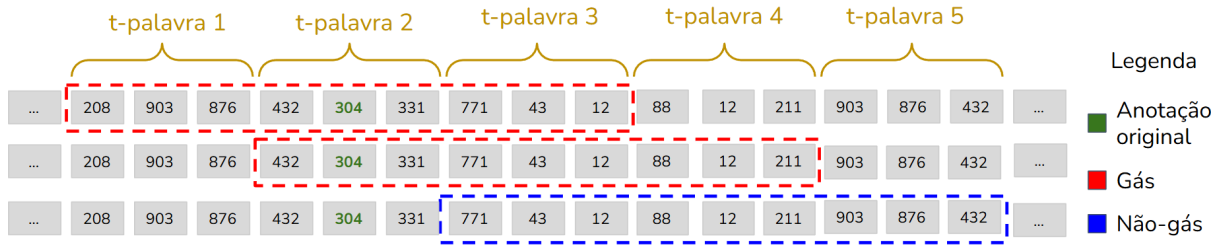


Figura 32 – Processo de anotação das frases. Imagem de elaboração própria.

A Figura 33 mostra um exemplo com a imagem sísmica original, a anotação feita pelo especialista e a anotação efetiva usada para treinar os modelos após o alargamento das anotações.



Figura 33 – Alargamento efetivo das anotações usadas para treinar o modelo.

O processo de alargamento das anotações não é ideal na perspectiva geológica, pois um único pixel representa um tempo equivalente a dezenas de metros a depender da resolução do equipamento. No entanto, essa é uma etapa necessária no pré-processamento para gerar mais amostras da classe gás e diminuir o desequilíbrio entre o número de amostras das duas classes. A arquitetura usada para classificar as t-frases é a mesma que a usada no pré-treinamento, com a adição de uma camada completamente conectada em cima da saída do RoBERTa e com uma camada de ativação Softmax. Os pesos da rede, com exceção da camada completamente conectada, são inicializados como sendo os mesmos obtidos após o pré-treinamento. Os pesos da última camada são inicializados aleatoriamente e nenhum dos pesos é mantido congelado, isto é, todos podem ser atualizados durante o treinamento na tarefa de classificação.

4.3.4 Inferência

A inferência de uma imagem sísmica usando o modelo treinado da forma descrita nas seções anteriores pode ser vista na Figura 34. Cada traço é tokenizado usando o

mesmo vocabulário aprendido pelo algoritmo TracePiece antes do treinamento. Após a etapa de tokenização, a inferência pode ser feita. Para cada traço, uma janela de tamanho $l_s \times l_w$ é tomada, e o modelo a classifica em gás ou não-gás. O valor previsto é então atribuído para cada pixel que forma essa janela na imagem original. A janela então desliza um pixel para baixo e repete o processo até que todo o traço sísmico seja classificado. É permitido aos resultados da classificação que se sobreponham para que classificações erradas sejam corrigidas pelas subsequentes. Ao final do processo de inferência, todo o traço sísmico está segmentado. Ao realizar a inferência da forma descrita em cada traço da imagem, o resultado será uma imagem completamente segmentada.

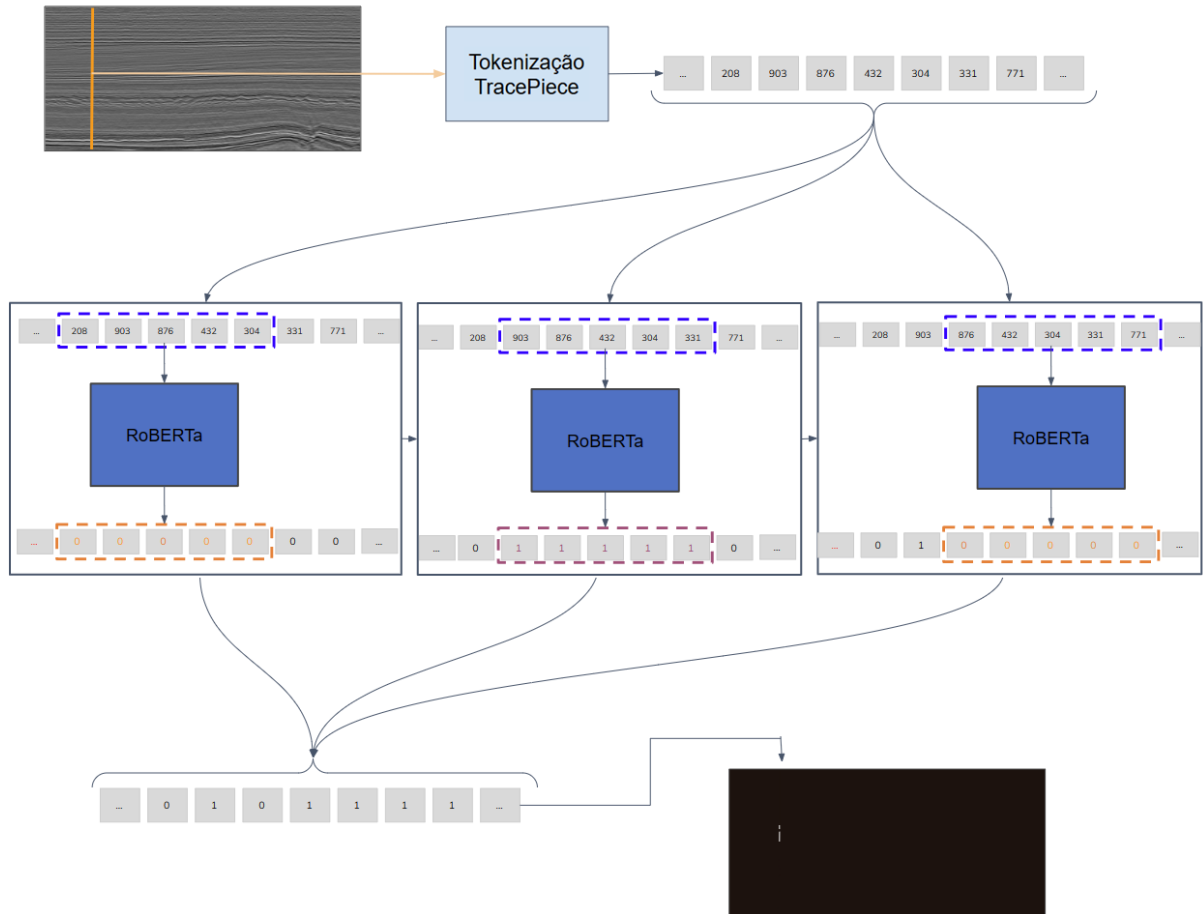


Figura 34 – Segmentação de uma imagem sísmica feita durante a etapa de inferência.

4.4 Método 3: Abordagem 1.5D - RoBERTa Multi-traços

A abordagem 1.5D, a que chamamos multi-traços, é semelhante àquela desenvolvida na seção anterior para a abordagem 1D. As etapas de tokenização com o algoritmo TracePiece e pré-treinamento com MLM são as mesmas. A diferença entre as duas abordagens é o contexto usado pelo RoBERTa para fazer sua classificação de traços, que pode

ser vista na Figura 35. Para uma determinada t-frase, ilustrada pelo quadrado laranja, $t_p - 1$ traços no mesmo intervalo de TWT ao seu redor também são considerados na classificação, formando um t-parágrafo. A Figura 36 mostra a construção de um t-parágrafo de forma mais detalhada.

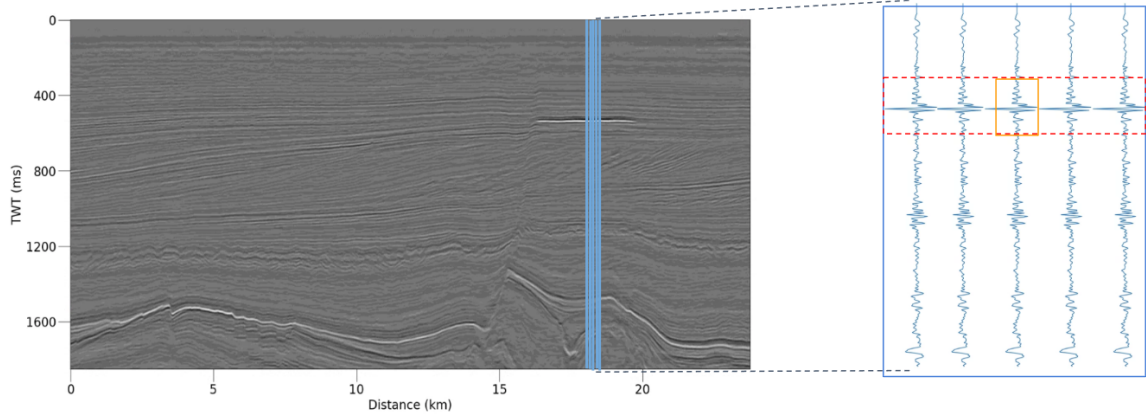


Figura 35 – Contexto usado pelo método multi-traços. Os traços em azul sobre a imagem sísmica possuem espessura exagerada para melhor visualização

O processo de anotação das t-frases é exatamente o mesmo ilustrado na Figura 32, já que aqui o valor da anotação válido para o conjunto de t-frases é o referente ao traço do meio, isto é, aquele que precisa ser classificado, da maneira detalhada na Figura 32.

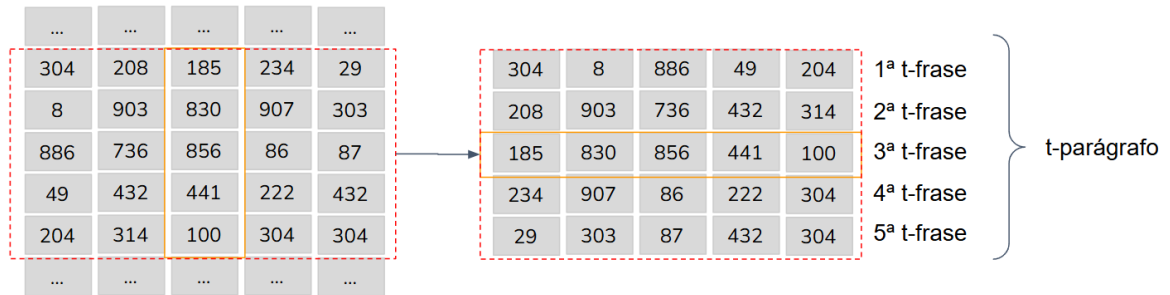


Figura 36 – Ilustração de um t-parágrafo. $N + 1$ t-frases tokenizadas lado a lado (no exemplo, $N = 4$), podem ser vistos como um parágrafo formado por $N + 1 = 5$ frases diferentes.

Por fim, como ilustrado na Figura 37, infere-se uma determinada imagem sísmica, com o modelo treinado, passando pela etapa de tokenização do traço e uma janela deslizante de modo a segmentar toda a imagem traço a traço e levando em consideração os traços vizinhos ao que se deseja segmentar.

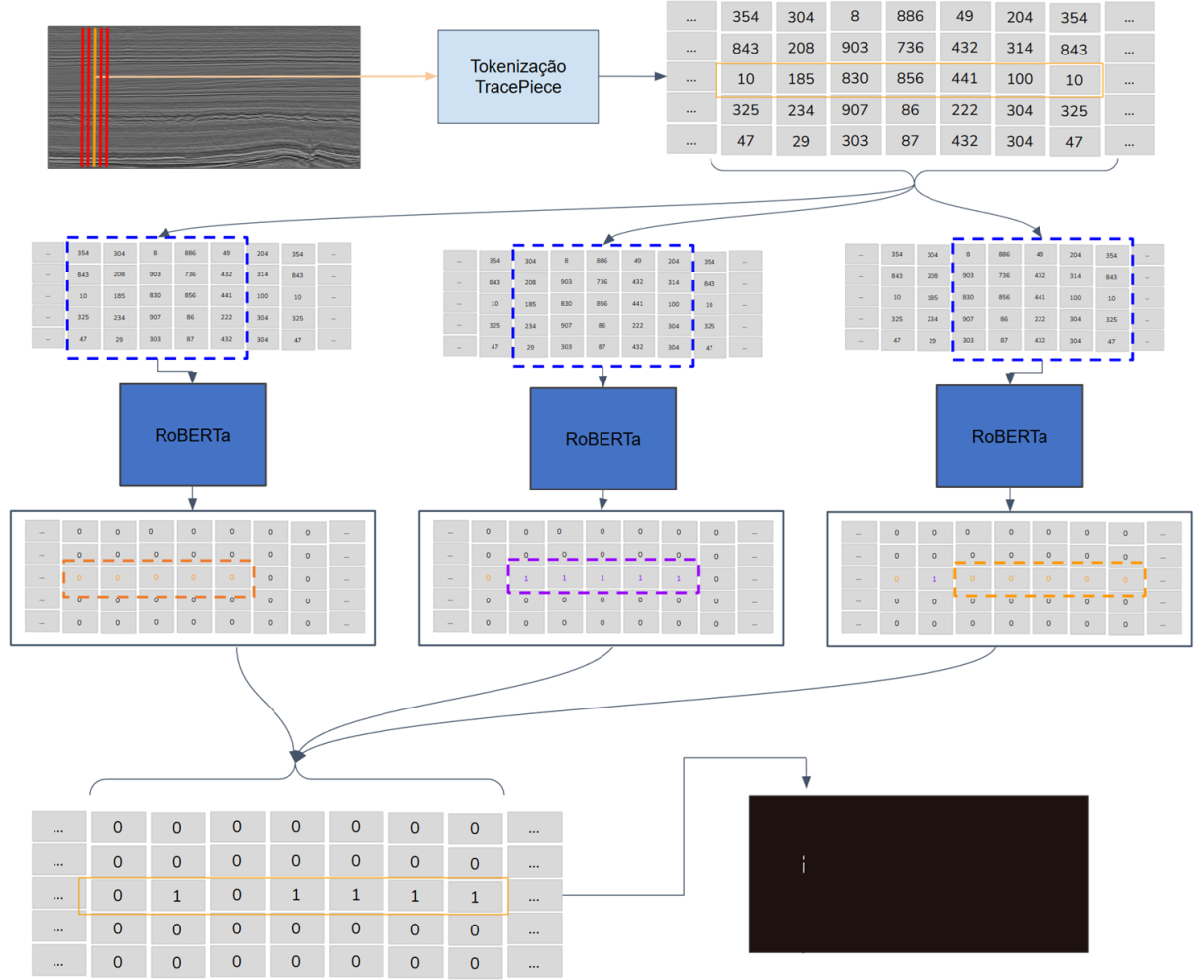


Figura 37 – Pipeline da inferência com o método multi-traços.

O número efetivo de traços N_{ef} vizinhos a serem considerados na inferência dependem da posição j do traço a ser inferido, da largura L da imagem sísmica e do parâmetro N .

$$N_{ef}(j, r) = \begin{cases} j + N/2, & \text{se } j < N/2 \text{ (borda esquerda)} \\ N, & \text{se } N/2 \leq j \leq L - 1 - N/2 \text{ (interior)} \\ L - 1 - j + N/2, & \text{se } j > L - 1 - N/2 \text{ (borda direita)} \end{cases} \quad (4.2)$$

Ou seja, conforme o traço analisado está mais próximo das bordas, menos traços são levados em consideração (pois a janela observada ultrapassa os limites da imagem). A Figura 38 ilustra os traços utilizados na inferência com o Método 3 para $j = 1$ e $j = L - 1$

no caso em que $N = 4$.

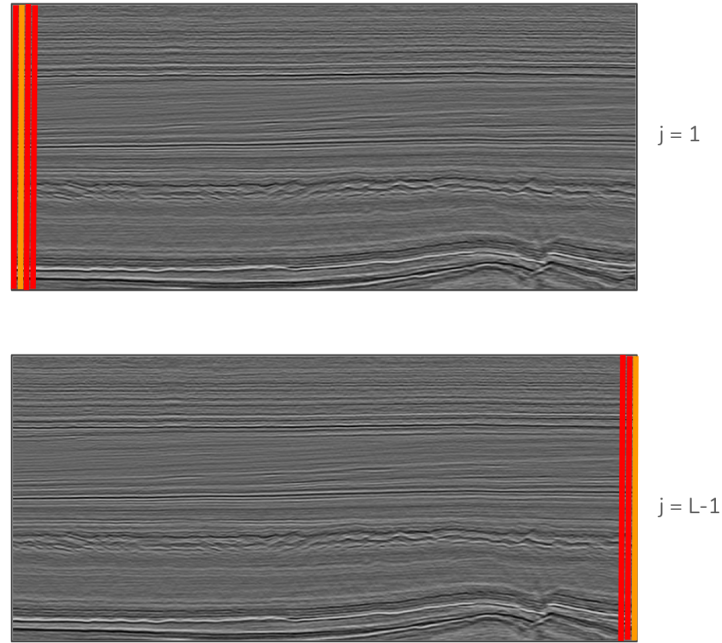


Figura 38 – Exemplo de traços utilizados na inferência nas bordas com o método multi-traços. Em laranja, o traço a ser segmentado e em vermelho os traços vizinhos considerados pelo método.

4.5 Considerações Finais

Este capítulo apresentou os três métodos desenvolvidos para a segmentação de gás natural em dados sísmicos de reflexão, explorando diferentes estratégias de representação e processamento da informação. Foram detalhadas todas as etapas do método de cada abordagem, desde o pré-processamento dos dados até a inferência das segmentações finais.

O Método 1, que é bidimensional, incorpora técnicas para realçar *brightspots* e remover ruídos. Assim, preparam-se as imagens sísmicas antes do treinamento e segmentação pela U-Net. Também é implementado uma estratégia de ensembles a fim de estabilizar o desempenho dos modelos treinados. Os Métodos 2 e 3, fundamentados em arquiteturas de modelos de linguagem, empregam técnicas de tokenização e adaptação de traços sísmicos para o formato exigido pela arquitetura RoBERTa. Os métodos se diferem quanto ao contexto utilizado para realizar a previsão: apenas um único segmento de traço no Método 2 e um determinado número de segmentos de traços vizinhos no Método 3.

No próximo capítulo, os resultados dos três métodos propostos serão apresentados e analisados. As avaliação quantitativa e qualitativa dos desempenhos permitirá

estabelecer as virtudes e falhas de cada um dos métodos, bem como identificar as contribuições desta pesquisa para o estado da arte na segmentação de gás natural em dados sísmicos.

5 RESULTADOS E DISCUSSÃO

Este capítulo apresenta os resultados obtidos na segmentação de acumulações de gás natural em dados sísmicos de reflexão mediante a aplicação dos três métodos propostos no capítulo anterior. O desempenho dos métodos será avaliado através de métricas quantitativas descritas no Capítulo 2 e também serão feitas análises qualitativas das segmentações produzidas.

Dividimos os resultados em seções correspondentes a cada modelo, e cada seção é composta de diferentes experimentos, cujas características variam de acordo com o tipo de divisão de dados realizada, como será explicado adiante, e também com o uso ou não de operações morfológicas nas segmentações dos modelos. Esses experimentos também permitem avaliar aspectos como a estabilidade do desempenho, especificamente no caso do Método 1, e da capacidade de generalização do modelo considerando apenas o conjunto de dados do bloco F3.

As métricas usadas para avaliar os modelos são a precisão (PRC), sensibilidade (SEN) e F1-score. As diferentes abordagens foram implementadas em linguagem Python 3.7¹, utilizando-se as bibliotecas Huggingface Transformers² e Pytorch³. Os experimentos foram executados em uma máquina com CPU Intel Core i5-10400, 16GB de RAM, placa gráfica NVIDIA GeForce RTX 3060 e sistema operacional Ubuntu. Nessa configuração, o tempo de treinamento e inferência de cada modelo é o indicado na Tabela

Tabela 3 – Tempos de inferência e treinamento de cada modelo.

Modelo	Treinamento	Inferência
U-Net (individual)	40 minutos	10 segundos/imagem
RoBERTa (MLM)	6 dias	-
RoBERTa (Classificação)	6 horas	6 minutos/imagem

A escolha da divisão entre dados de treino e teste para os experimentos é feita ao levar em consideração que o conjunto de dados avaliado possui duas acumulações de gás natural separadas por aproximadamente 10 quilômetros de distância na direção norte-sul. Presumimos, então, a hipótese de que a formação geológica da acumulação norte seja distinta daquela da acumulação sul. Isto facilita a discussão sobre a generalização dos modelos treinados, já que na prática desejamos usá-los em regiões distintas daquelas

¹ <https://python.org/>

² <https://huggingface.co/>

³ <https://pytorch.org>

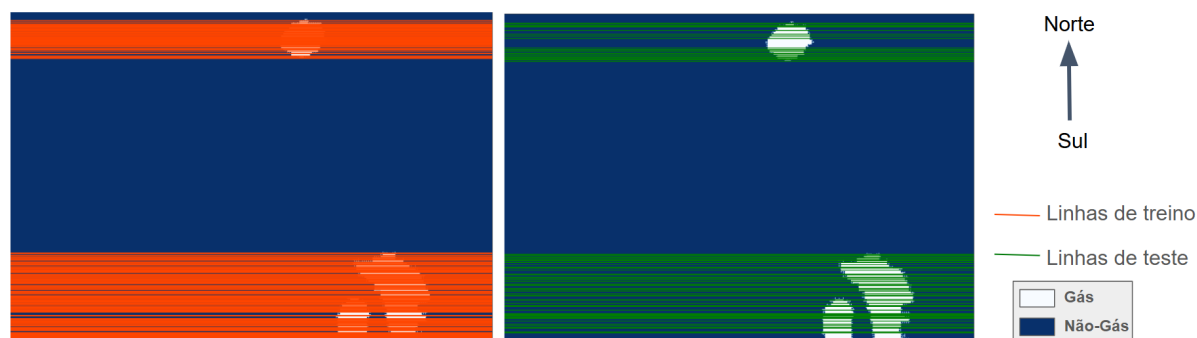


Figura 39 – Vista axial do volume, ilustrando as duas acumulações de gás presentes, uma a norte e outra a sul, e a política aleatória de divisão de fatias entre treino e teste.

cujos dados foram usados para treinar os algoritmos de segmentação. Devido à falta de dados anotados disponíveis publicamente, o bloco F3 torna-se a principal fonte de dados de imagens sísmicas pública e com dados adicionais suficientes para uma interpretação por um especialista. Em adição a esse cenário, existe o fato de que inlines próximas são altamente correlacionadas. Por isso, consideramos que a divisão de dados mais sensível a se fazer é usar uma das acumulações exclusivamente para o treino e validação do modelo e outra para teste. A este tipo de divisão, damos o nome de política de divisão de treino-teste não-aleatória, cuja nomenclatura foi reduzida para política não-aleatória no decorrer o texto. Para compararmos os resultados com alguns modelos utilizados na literatura, usaremos o modo de separação anteriormente utilizado na literatura, que denominaremos de política de divisão de treino-teste aleatória, ou política aleatória.

Na política aleatória, foram selecionadas inlines pertencentes a ambas acumulações de gás de maneira aleatória, de modo que os conjuntos de treino e teste possuem inlines de ambas as regiões, contando 178 inlines para treino e 76 para teste. Através do método para criação de t-frases explicado no Capítulo 4, foram criadas 3.724.530 t-frases para treinar e validar o modelo, das quais 18.530 foram classificadas como gás e 3.706.000 como não-gás. A Figura 39 mostra uma vista axial do volume de dados, ilustrando todas as inlines utilizadas para treino e teste em cores diferentes.

Na política não-aleatória, todas as 178 inlines pertencentes à acumulação sul do bloco foram usadas para treino e validação do modelo e 76 inlines pertencentes à acumulação norte foram usadas para teste, como visto na Figura 40. O número de t-frases criadas para treino, validação e teste são os mesmos da política aleatória.

Foram usadas apenas inlines ao invés de apenas crosslines ou uma combinação dos dois tipos para que a diferença a respeito dos aspectos geológicos das linhas que

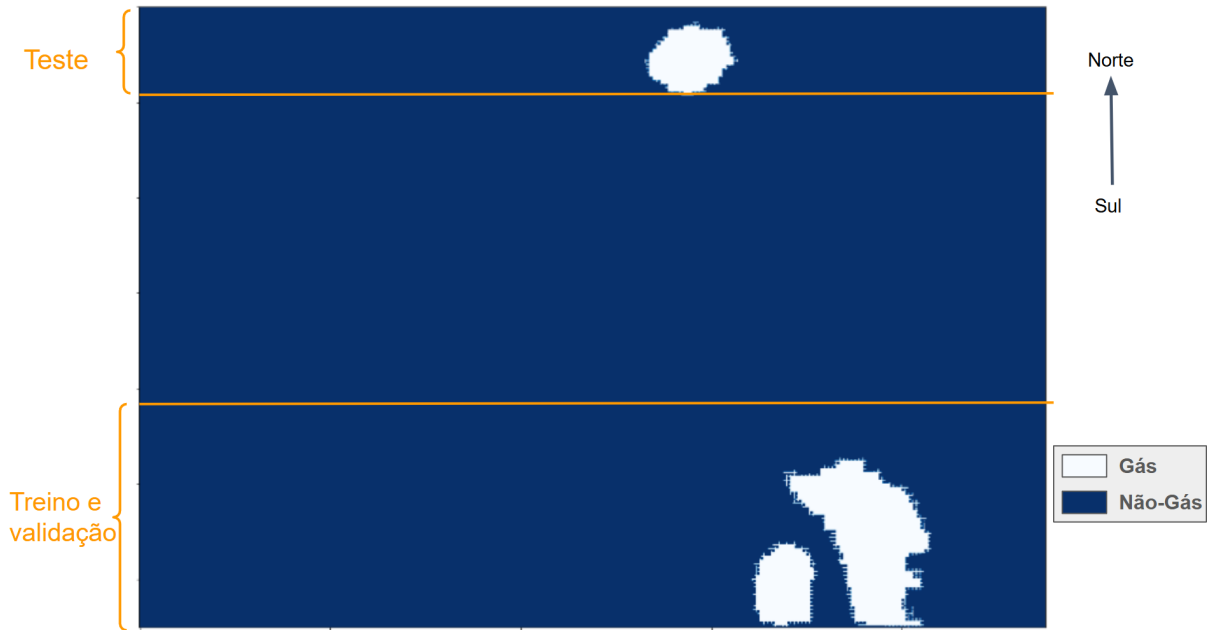


Figura 40 – Vista axial do volume, ilustrando as duas acumulações de gás presentes, uma a norte e outra a sul, e a política não-aleatória de divisão de fatias entre treino e teste.

contém acumulação pudesse ser maximizada. Em outras palavras, para que informação sobre os arredores da acumulação norte não fosse incluída no treinamento.

5.1 Método 1: Abordagem 2D - Ensemble de U-Nets

Nesta seção, serão definidos os parâmetros utilizados nos experimentos feitos com a arquitetura U-Net, bem como os resultados da segmentação nas políticas aleatória e não-aleatória. É verificado experimentalmente que as segmentações feitas por esta arquitetura com os parâmetros mencionados são instáveis e consideravelmente diferentes entre si a depender da inicialização dos pesos. Por isso, além da discussão sobre os resultados da segmentação, nesta seção também será relatada a metodologia para obter resultados consistentes com a arquitetura U-Net. A distribuição inicial de pesos nas camadas lineares e convolucionais é feita obedecendo à distribuição uniforme Glorot, usada com o intuito de manter a variância das ativações e gradientes aproximadamente constantes ao longo das diferentes camadas.

5.1.1 Pré-processamento e treinamento

Em todos os experimentos, o pré-processamento para o treinamento da U-Net é feito do seguinte modo: primeiro, a imagem é redimensionada para 512 x 512. Em seguida,

aplica-se a quantização dos pixels que formam a imagem e, por fim, a transformada Wavelet usando a família Daubechies 2. As imagens são aumentadas usando unicamente o espelhamento de 20% das imagens de treino. Do contrário, devido a alta correlação entre as imagens, é observado um sobreajuste na posição da imagem em que a acumulação se encontra.

Cada U-Net pertencente ao ensemble foi treinada de acordo com os parâmetros da arquitetura mostrada na Tabela 4 e de treinamento na Tabela 5. As sementes aleatórias não foram definidas previamente, sendo alteradas em cada treinamento individual das U-Nets. Isso permite com que os pesos iniciais sejam distintos em todos os experimentos realizados, de modo a verificar a independência dos resultados com relação ao estado inicial do modelo.

Tabela 4 – Parâmetros da arquitetura U-Net.

Camada	Número de mapas de características
1 (codificador)	16
2 (codificador)	32
3 (codificador)	64
4 (codificador)	128
5 (bottleneck)	256
6 (decodificador)	128
7 (decodificador)	64
8 (decodificador)	32
9 (decodificador)	16

Tabela 5 – Parâmetros para o treinamento da U-Net.

Parâmetro	Valor
Épocas	20
Learning rate	5×10^{-4}
Tamanho do batch	8

De uma camada para a seguinte, usou-se a técnica de Maxpooling tamanho de kernel igual a 2. Foram usadas 8 imagens por batch e a função de custo utilizada foi uma média entre a função de custo Dice e a função de custo de Entropia Binária definidas pelas Equações 2.15 e 2.12 :

$$\mathcal{L} = \frac{1}{2}\mathcal{L}_{BC} + \frac{1}{2}\mathcal{L}_{Dice} \quad (5.1)$$

5.1.2 Experimento 1: U-Net na política aleatória

Como no caso da U-Net as métricas variam consideravelmente, os resultados abaixo mostram conjuntos de métricas relativos a diferentes experimentos. Cada experimento consiste no treinamento e teste de uma U-Net, utilizando exatamente o mesmo conjunto de treinamento e teste, assim como o mesmo pré-processamento. A Tabela 6 mostra os resultados médios das métricas de avaliação e seus respectivos desvios padrão para cada um de 5 experimentos (exp) diferentes.

Tabela 6 – Métricas de segmentação de treino e teste na política aleatória.

Exp #	Conjunto de treino			Conjunto de teste		
	F1 $\pm$ STD	PRC $\pm$ STD	SEN $\pm$ STD	F1 $\pm$ STD	PRC $\pm$ STD	SEN $\pm$ STD
0	0.86 $\pm$ 0.15	0.83 $\pm$ 0.17	0.89 $\pm$ 0.12	0.80 $\pm$ 0.22	0.78 $\pm$ 0.21	0.87 $\pm$ 0.19
1	0.89 $\pm$ 0.16	0.89 $\pm$ 0.15	0.89 $\pm$ 0.15	0.84 $\pm$ 0.23	0.84 $\pm$ 0.20	0.87 $\pm$ 0.22
2	0.91 $\pm$ 0.15	0.91 $\pm$ 0.14	0.91 $\pm$ 0.14	0.87 $\pm$ 0.18	0.88 $\pm$ 0.17	0.88 $\pm$ 0.18
3	0.84 $\pm$ 0.12	0.84 $\pm$ 0.15	0.86 $\pm$ 0.09	0.82 $\pm$ 0.15	0.81 $\pm$ 0.18	0.86 $\pm$ 0.10
4	0.85 $\pm$ 0.14	0.81 $\pm$ 0.16	0.92 $\pm$ 0.08	0.81 $\pm$ 0.18	0.76 $\pm$ 0.20	0.90 $\pm$ 0.10

Quando a política de divisão de treino-teste é aleatória, os valores das métricas de treino e teste são próximos. Mais ainda, existe uma diferença baixa, de no máximo 0.07, entre os valores de F1 para diferentes experimentos no conjunto de treino ou no conjunto de teste. A variação da precisão é um pouco maior, de 0.10 no conjunto de treino e de 0.12 no conjunto de testes. A sensibilidade é a métrica mais estável, com diferença de no máximo 0.05 no conjunto de treino e de 0.04 no conjunto de teste.

5.1.3 Experimento 2: U-Net na política não-aleatória

Conforme observado em outras abordagens, na política não-aleatória a variação dos valores das métricas é consideravelmente maior. A Tabela 7 ressalta essa diferença em relação à política aleatória. Primeiramente nota-se que os valores das três métricas de avaliação aumentaram no conjunto de treino. A exclusão das inlines pertencentes à acumulação norte, que possui geologia diferente, além da região de corte e diferença de distribuição de intensidade dos pixels, permitiu um sobreajuste maior aos dados de treinamento, que agora são ainda mais correlatos do que antes. Por outro lado, vemos que os valores das métricas no conjunto de teste diminuíram, além de variarem significativamente entre os experimentos realizados. Na política aleatória, a diferença máxima entre os F1 no conjunto de testes é de 0.07, enquanto que na política não-aleatória é de 0.57.

Tabela 7 – Métricas de segmentação dos conjuntos de treino e teste usando a política não-aleatória.

Conjunto	Run #	Métricas		
		F1 ($\pm$ STD)	Precision ($\pm$ STD)	Recall ($\pm$ STD)
Treino	0	0.92 ± 0.16	0.96 ± 0.09	0.92 ± 0.17
	1	0.90 ± 0.14	0.87 ± 0.16	0.95 ± 0.11
	2	0.86 ± 0.14	0.83 ± 0.15	0.88 ± 0.12
	3	0.88 ± 0.14	0.86 ± 0.15	0.91 ± 0.09
	4	0.91 ± 0.14	0.90 ± 0.15	0.93 ± 0.11
Teste	0	0.53 ± 0.18	0.45 ± 0.18	0.69 ± 0.21
	1	0.45 ± 0.19	0.57 ± 0.23	0.49 ± 0.25
	2	0.68 ± 0.17	0.61 ± 0.21	0.87 ± 0.10
	3	0.70 ± 0.19	0.62 ± 0.22	0.86 ± 0.06
	4	0.13 ± 0.19	0.14 ± 0.15	0.21 ± 0.31

O fato de a diferença entre as métricas no conjunto de testes depender fortemente do conjunto de pesos iniciais do treinamento faz com que a confiança no modelo treinado seja muito reduzida: ao treinar o modelo, não se saberá o nível de desempenho que a versão treinada terá em um determinado conjunto de dados.

Tabela 8 – Métricas de segmentação dos conjuntos de treino e teste na política aleatória usando quantização, DWT e ensembling.

Conjunto	Exp #	Métricas		
		F1 ($\pm$ STD)	PRC ($\pm$ STD)	SEN ($\pm$ STD)
Treino	0	0.87 ± 0.17	0.87 ± 0.18	0.89 ± 0.15
	1	0.86 ± 0.18	0.88 ± 0.18	0.87 ± 0.17
	2	0.87 ± 0.18	0.88 ± 0.17	0.88 ± 0.16
	3	0.86 ± 0.18	0.87 ± 0.18	0.88 ± 0.16
	4	0.87 ± 0.17	0.87 ± 0.17	0.87 ± 0.15
Teste	0	0.87 ± 0.16	0.82 ± 0.17	0.96 ± 0.10
	1	0.88 ± 0.16	0.83 ± 0.17	0.96 ± 0.06
	2	0.87 ± 0.16	0.81 ± 0.18	0.97 ± 0.07
	3	0.87 ± 0.16	0.81 ± 0.18	0.97 ± 0.05
	4	0.87 ± 0.16	0.80 ± 0.17	0.97 ± 0.04

A Tabela 8 mostra cinco outros experimentos já com o uso de quantização, DWT e ensembling para a política aleatória. Os valores de todas as métricas são consistentes, com variação de no máximo 0.01 no F1 de treino e teste.

A Tabela 9 mostra cinco experimentos, também com o uso de todos os pré e pós-processamentos usados - quantização, DWT e ensembling - para a política não-aleatória. Assim como ocorre com o uso da U-Net base discutida anteriormente, na política não-

Tabela 9 – Métricas de segmentação dos conjuntos de treino e teste na política não-aleatória usando quantização, DWT e ensembling.

Conjunto	Exp #	Métricas		
		F1 ($\pm$ STD)	PRC ($\pm$ STD)	SEN ($\pm$ STD)
Treino	0	0.92 ± 0.15	0.93 ± 0.12	0.92 ± 0.14
	1	0.91 ± 0.15	0.93 ± 0.13	0.91 ± 0.15
	2	0.92 ± 0.15	0.93 ± 0.13	0.91 ± 0.14
	3	0.92 ± 0.15	0.94 ± 0.11	0.91 ± 0.15
	4	0.92 ± 0.15	0.94 ± 0.11	0.91 ± 0.16
Teste	0	0.62 ± 0.15	0.49 ± 0.14	0.88 ± 0.08
	1	0.62 ± 0.15	0.49 ± 0.10	0.91 ± 0.05
	2	0.64 ± 0.17	0.59 ± 0.18	0.75 ± 0.12
	3	0.61 ± 0.20	0.56 ± 0.20	0.72 ± 0.16
	4	0.60 ± 0.16	0.58 ± 0.18	0.69 ± 0.12

aleatória as métricas de treino aumentam em comparação com as da política aleatória por apresentar, durante o treinamento, imagens muito mais semelhantes entre si já que não incluem inlines da acumulação norte. Nota-se também uma estabilidade maior nas segmentações por conta de precisões e sensibilidades próximas entre as previsões de inlines de um mesmo experimento. No conjunto de teste, observa-se uma melhora significativa tanto em valores absolutos das métricas de avaliação quanto na estabilidade dos resultados.

A Figura 41 mostra a segmentação de 10 U-Nets diferentes na inline 571, pertencente à acumulação de gás ao norte. Cinco destas previsões, mostradas no lado esquerdo, representam a segmentação de cada U-Net treinada na política aleatória. As cinco previsões do lado direito são a segmentação de cada U-Net treinada na política não-aleatória. Conforme já discutido, na política aleatória os resultados são mais uniformes entre diferentes experimentos, enquanto que na política não-aleatória há uma variação muito maior de desempenho dependendo do experimento e sua respectiva inicialização de pesos da rede.

Uma ilustração mais detalhada dessa variação pode ser encontrada na Figura 42. Nela vemos a segmentação na inline 574 de 10 U-Nets base e seu ensemble. Muitos dos modelos não segmentam corretamente a região anotada, sendo que duas U-Nets mostradas na imagem não classificam nenhum pixel na imagem como gás. Isto resulta em uma segmentação por ensemble com muitos falsos negativos. A última U-Net à esquerda possui muitos falsos positivos, incluindo segmentação praticamente contínua numa região acima da região anotada. Essa forma tomada pelos falsos positivos denotam particular-

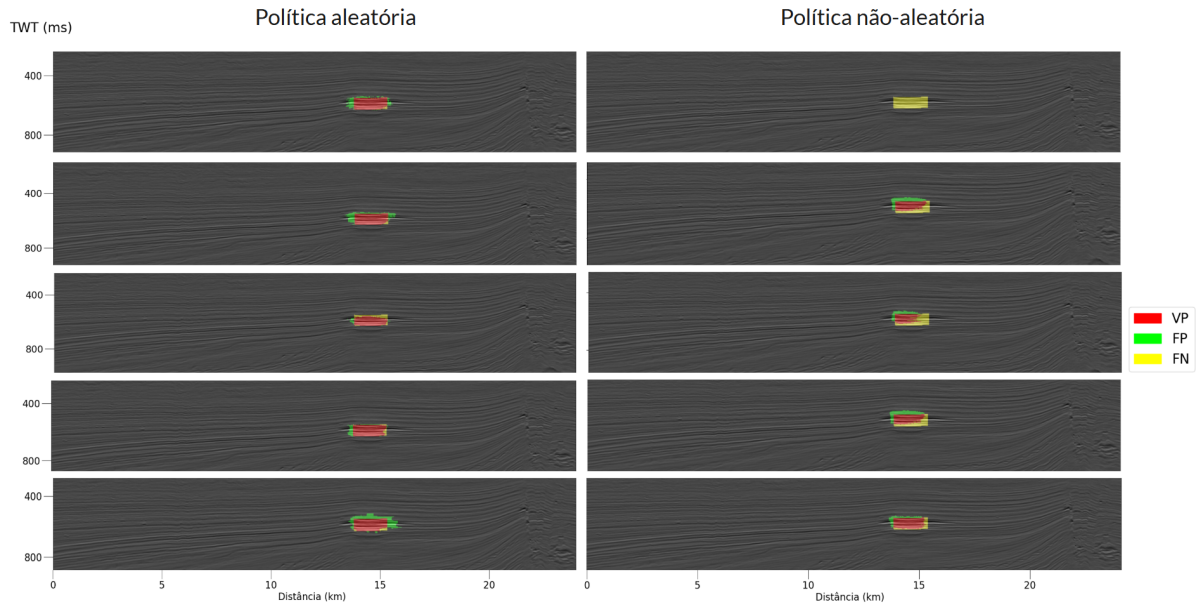


Figura 41 – Comparação entre a segmentação de U-Nets usando as políticas aleatória e não-aleatória.

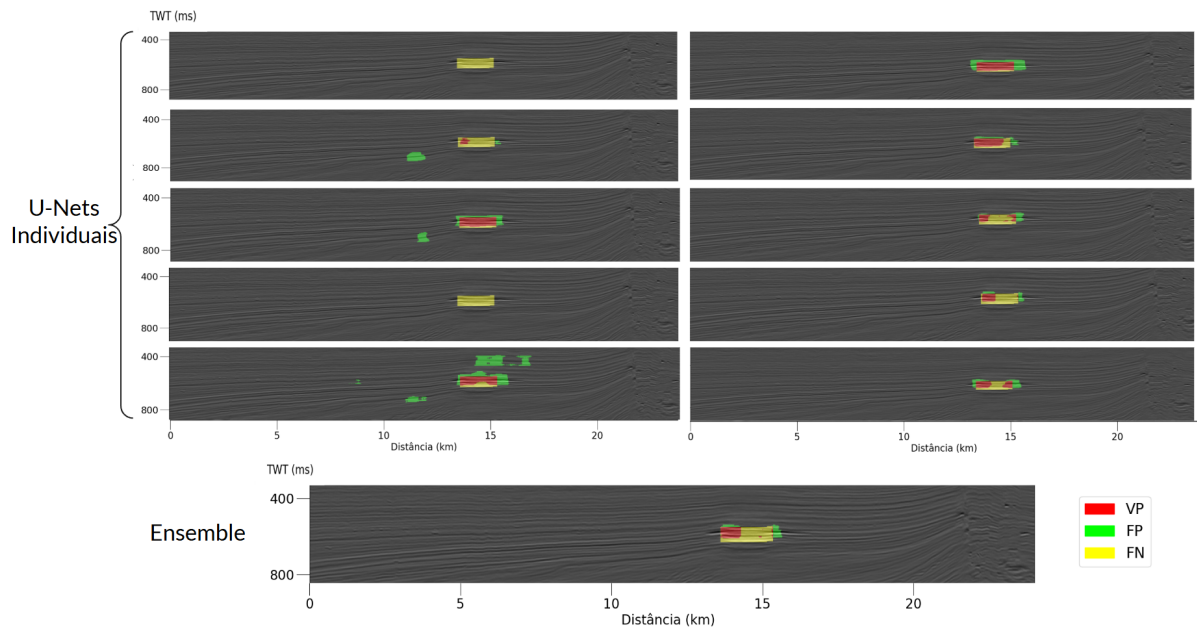


Figura 42 – Segmentação da Inline 574 das U-Nets que compõe um ensemble. A imagem de baixo ilustra a inferência do ensemble para uma mesma inline.

mente um sobreajuste nos dados por ser um formato muito parecido com as das regiões de gás (retangular).

5.1.4 Comparação entre diferentes combinações de técnicas

A Tabela 10 mostra as métricas médias sobre diferentes experimentos considerando cada combinação de técnicas de pré e pós-processamento. O caso “Base” se refere ao experimento sem pré ou pós-processamento. Adicionar DWT garante consistência dos

resultados, evidenciado no decréscimo do desvio padrão do F1-score em uma ordem de magnitude e no decréscimo dos desvios padrão da precisão e da sensibilidade em valores comparados com o caso em que nenhum pré-processamento ou pós-processamento é realizado.

Fazer o ensemble das segmentações das U-Nets treinadas sem nenhum tipo de pré-processamento nas imagens também induz uma melhora na estabilidade, especialmente a precisão e na sensibilidade, porém remove muitos verdadeiros positivos para melhorar a precisão, levando a uma diminuição na sensibilidade em comparação com o caso Base. Ao usar a quantização e DWT simultaneamente, o F1-score é o mesmo que no caso de sem pré-processamento, mas existe um aumento na sensibilidade e um decréscimo na precisão. Ao usar todas as três técnicas, existe um aumento efetivo no F1-score graças a um aumento significativo na sensibilidade, ao passo que o desvio padrão de todas as métricas é o menor entre virtualmente todos os casos, com exceção do experimento onde foi realizado apenas o uso de ensemble como pós-processamento.

Tabela 10 – Médias das métricas de segmentação tomadas sobre cinco experimentos para cada combinação de métodos no conjunto de testes da política não-aleatória.

Método #	Conjunto de teste		
	F1 ($\pm$ STD)	PRC ($\pm$ STD)	SEN ($\pm$ STD)
Base	0.49 $\pm$ 0.21	0.56 $\pm$ 0.08	0.62 $\pm$ 0.24
DWT	0.49 $\pm$ 0.06	0.47 $\pm$ 0.05	0.63 $\pm$ 0.15
Quantização	0.42 $\pm$ 0.13	0.50 $\pm$ 0.10	0.53 $\pm$ 0.21
Ensemble	0.48 $\pm$ 0.09	0.74 $\pm$ 0.02	0.47 $\pm$ 0.09
Quantização + DWT	0.49 $\pm$ 0.14	0.43 $\pm$ 0.14	0.65 $\pm$ 0.10
Quantização + DWT + Ensemble	0.62 $\pm$ 0.01	0.54 $\pm$ 0.04	0.79 $\pm$ 0.08

Na Figura 43, vemos o impacto do treinamento apenas com a adição da DWT no pré-processamento das imagens. Nota-se que as segmentações são mais uniformes entre si, como denotado pela redução no desvio padrão do F1 score e precisão na Tabela 10. Porém, ainda existem muitos falsos negativos em diversos experimentos diferentes.

Na Figura 44, tanto a quantização quanto a DWT foram aplicadas nas imagens passadas para a rede neural. Neste caso, vemos que 4 das 10 apresentam uma marcação contínua de falsos positivos entre 5 km e 10 km. As áreas anotadas também são marcadas em toda a sua extensão, com exceção da terceira U-Net à esquerda, que possui alguns falsos negativos. Ao fazer o ensemble dos resultados, os falsos positivos são descartados pelo voto da maioria e, de igual modo, os falsos negativos da terceira U-Net à esquerda

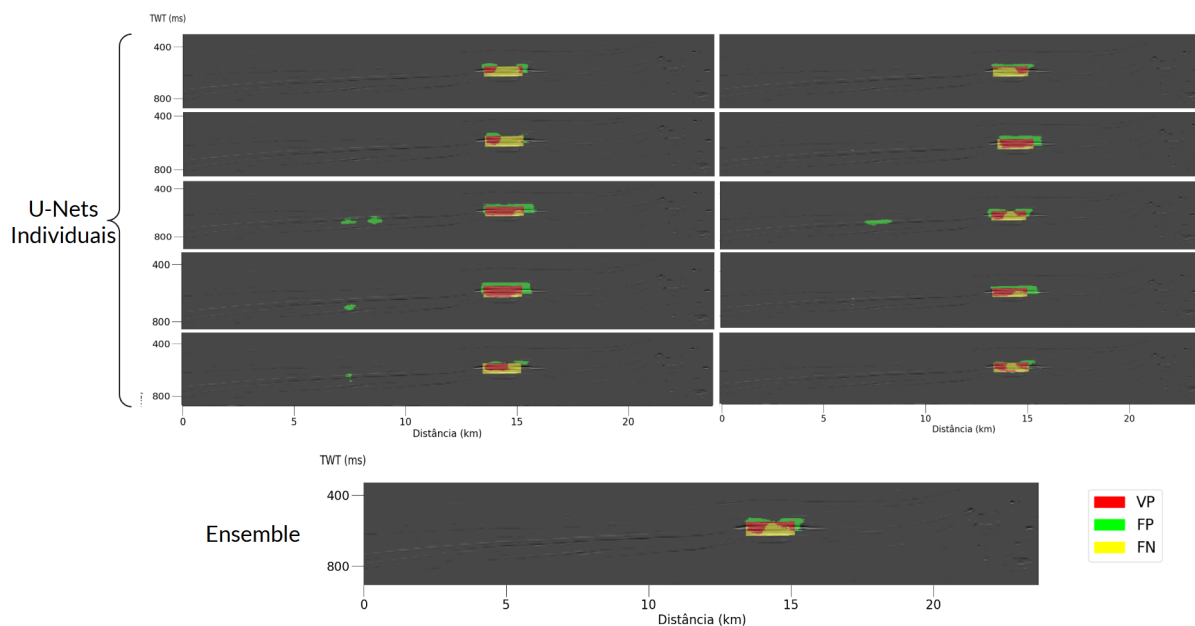


Figura 43 – Segmentação da inline 574 realizadas por U-Nets individuais que constituem um ensemble. A imagem inferior constitui a segmentação da mesma inline feita pelo ensemble, cujas imagens passaram apenas pelo pré-processamento com a transformada wavelet.

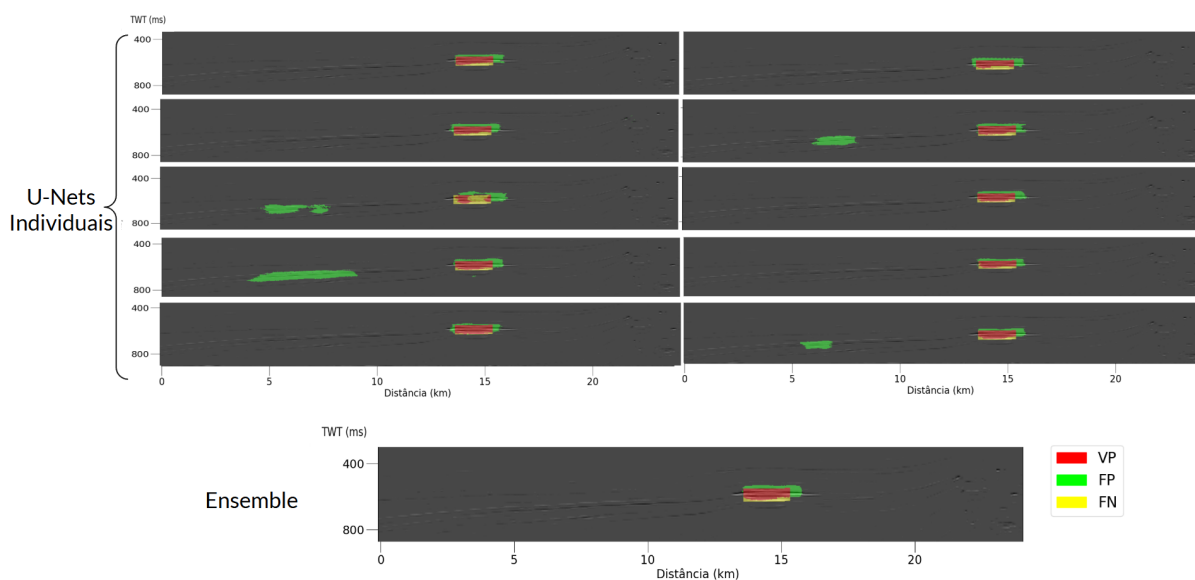


Figura 44 – Segmentação da Inline 574 realizadas por U-Nets individuais que constituem um ensemble. A imagem inferior constitui a segmentação da mesma inline feita pelo ensemble, cujas imagens passaram pelo pré-processamento de DWT e quantização.

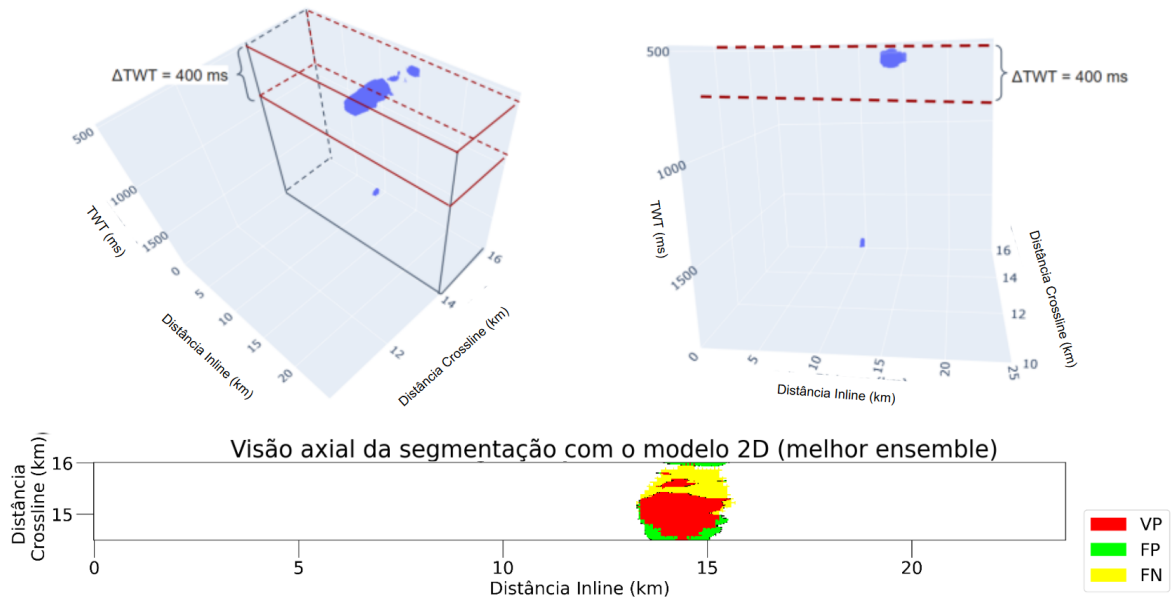


Figura 45 – O gráfico no topo à esquerda mostra como a fatia axial vista na figura inferior é feita, com aproximadamente dois quilômetros na direção das inlines, das inlines 558 a 635. A figura no topo à direita mostra melhor, no gráfico 3D, a profundidade considerada na criação da figura inferior. A figura inferior mostra a visão axial do bloco F3 depois das previsões do ensemble de U-Nets no conjunto de teste.

não são levados em consideração. No final, a segmentação gerada pelo ensemble segmenta apenas a região de interesse, com alguns falsos positivos ao redor resultantes da marcação da estrutura que armazena o gás natural.

Podemos ver como fica a segmentação tridimensional da U-Net, já considerando-se a quantização, DWT e o ensembling, na Figura 45. Nota-se uma continuidade muito maior e mais próxima das bordas laterais da acumulação de gás, porém com pouca capacidade de segmentar a região de corte. Também podemos ver uma baixa incidência de falsos positivos nas regiões em que existe uma transição abrupta de tipo de sedimento que não está relacionada com gás natural, ainda que seu efeito no traço sísmico seja semelhante.

5.1.5 Análise dos mapas de características

Procuramos entender em que ponto os modelos diferentes de U-Nets (mesma arquitetura porém treinados com pesos iniciais distintos) passaram a divergir em relação às características extraídas das imagens. Para isso, analisamos alguns mapas de características em diferentes profundidades do codificador e decodificador.

Na Figura 46, comparamos diversos filtros (ou mapa de características) de uma mesma camada da parte codificadora de duas U-Nets, uma com bom desempenho na

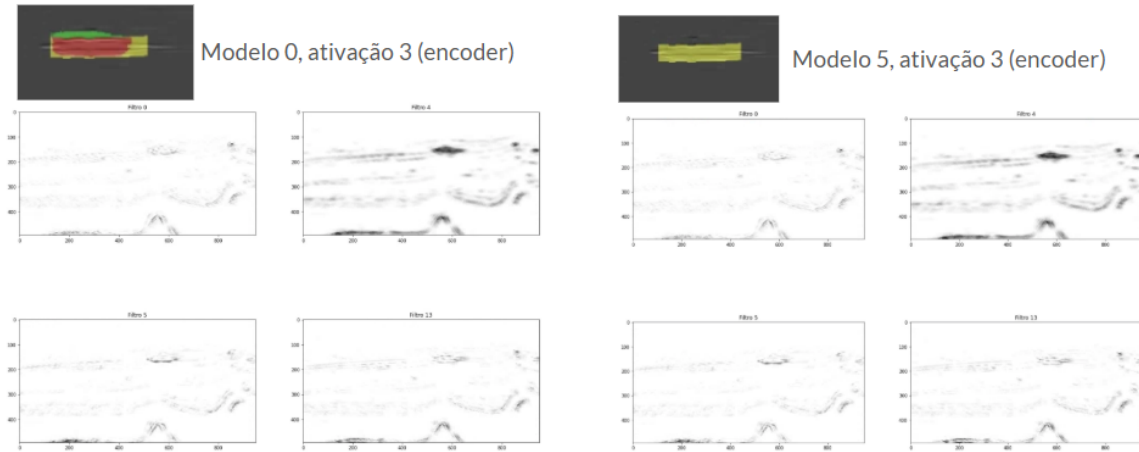


Figura 46 – Comparação entre a camada de ativação 3 da parte codificadora (encoder) de duas U-Nets em experimentos distintos. A segmentação final de cada um dos modelos, ao redor da área anotada, também é mostrada.

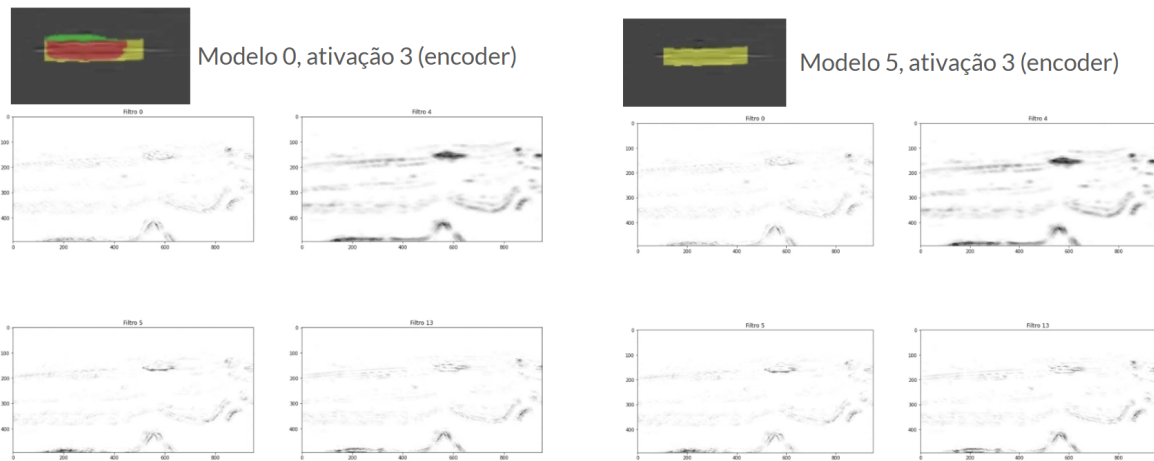


Figura 47 – Comparação entre a camada de ativação 8 da parte codificadora (encoder) de duas U-Nets em experimentos distintos. A segmentação final de cada um dos modelos, ao redor da área anotada, também é mostrada.

segmentação e outra que não segmentou nada. Nesta ativação na primeira camada do codificador, vemos que não há uma grande diferença qualitativa entre os diferentes filtros. O mesmo acontece também na camada de ativação 8, um pouco mais adiante no codificador, como mostra a Figura 47.

No entanto, no início do decodificador, já na camada 22, vemos que existe uma mudança considerável nos filtros. No modelo 0, vemos que existem duas regiões onde o modelo considera que as características sejam relevantes. Uma mais acima e numa região central da imagem, que é onde está localizada a acumulação de gás natural, e outra mais a direita para baixo, onde existe um ponto com *brightspots* não associados à gás natural. Já o modelo 5 foca somente nessa segunda região e não “vê” a acumulação de

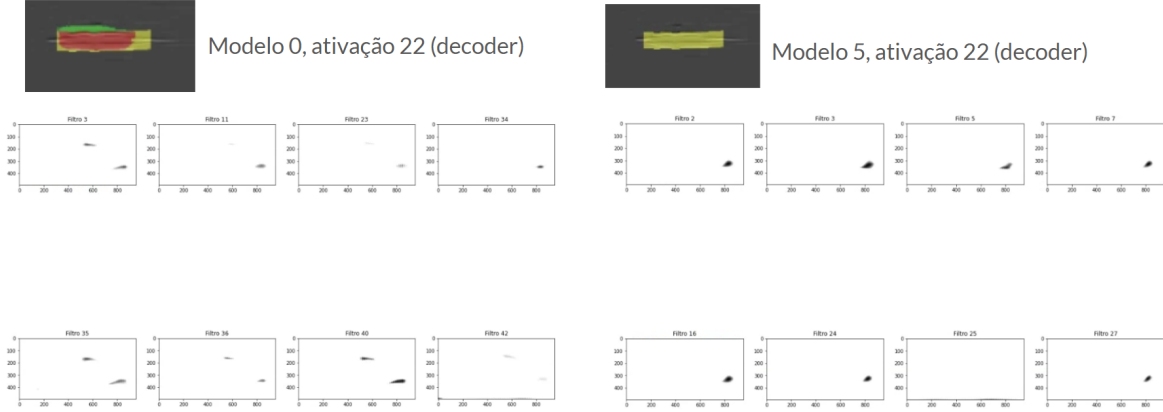


Figura 48 – Comparação entre a camada de ativação 22 da parte decodificadora (decoder) de duas U-Nets em experimentos distintos. A segmentação final de cada um dos modelos, ao redor da área anotada, também é mostrada.

gás. Como isto acontece logo no começo do decodificador, consideramos a hipótese de que o modelo é particularmente sensível às características no espaço latente do *bottleneck* da arquitetura U-Net. Isto reflete a dificuldade inerente em lidar com dados de imagens sísmicas, que são susceptíveis a ruídos e interferências em diversas etapas de sua aquisição e pré-processamento.

5.2 Método 2: Abordagem 1D - RoBERTa

Todos os experimentos na abordagem 1D compartilham o mesmo pré-treinamento. Durante a etapa de pré-treinamento, o RoBERTa foi treinado na tarefa de MLM com os hiperparâmetros da Tabela 11. Foram usadas todas as inlines disponíveis para treino para simular a inclusão da maior quantidade de dados disponíveis para essa tarefa não-supervisionada.

Tabela 11 – Parâmetros para o pré-treinamento na tarefa de MLM.

Parâmetro	Valor
Hidden layers	8
Attention heads	12
Tamanho do vocabulário	200
Batch size	1024
Épocas	200

Foi treinado um modelo na tarefa de classificação em que todos os pesos foram otimizados. Os parâmetros usados nesta tarefa são descritos na Tabela 12. Em ambas as políticas de treino-teste, foram tomados como pesos iniciais aqueles obtidos durante o pré-treinamento.

Tabela 12 – Parâmetros para o treinamento na tarefa de classificação de t-frases.

Parâmetro	Valor
Hidden layers	8
Attention heads	12
Tamanho do vocabulário	200
Batch size	1024
Épocas	200
Learning rate	5×10^{-5}
l_s	5
l_w	7
input size	35 tokens
Função de custo	Focal

5.2.1 Experimento 3: Política aleatória sem operações morfológicas

Tabela 13 – Valores médios das métricas de avaliação com seus respectivos erros padrões para a política aleatória sem a aplicação de operações morfológicas no pós-processamento, avaliados pixel a pixel antes da aplicação de operações morfológicas.

Métrica de Avaliação	Com ou sem região de corte	
	Com corte	Sem corte
F1-Score	0.53 ± 0.02	0.54 ± 0.02
PRC	0.41 ± 0.02	0.41 ± 0.01
SEN	0.83 ± 0.03	0.83 ± 0.03

A Tabela 13 mostra as métricas de avaliação da segmentação no Experimento 1. O modelo treinado na política aleatória é capaz de identificar possíveis acumulações de gás natural com uma sensibilidade média de 0.83 ± 0.03 , resultando em demarcações mais claras e contínuas da região de interesse. A baixa precisão deve-se principalmente aos falsos positivos ao redor da região anotada. Neste experimento, observamos também que considerar ou não a região com imagens corrompidas não altera significativamente as métricas de avaliação. Isto é esperado já que os traços pertencentes às imagens corrompidas também são usados no treinamento do modelo, que aprende a segmentar corretamente levando os problemas nos dados dessas imagens em consideração.

A Figura 49 mostra a consistência das inferências em várias inlines do conjunto de teste, que contém seções transversais altamente correlacionadas com aquelas usadas no treinamento. Ambas as regiões destacadas pelo especialista são identificadas sem um único falso positivo a uma distância maior que 1 km das anotações originais. Os falsos negativos indicados nessas inferências acontecem em regiões com uma diminuição de

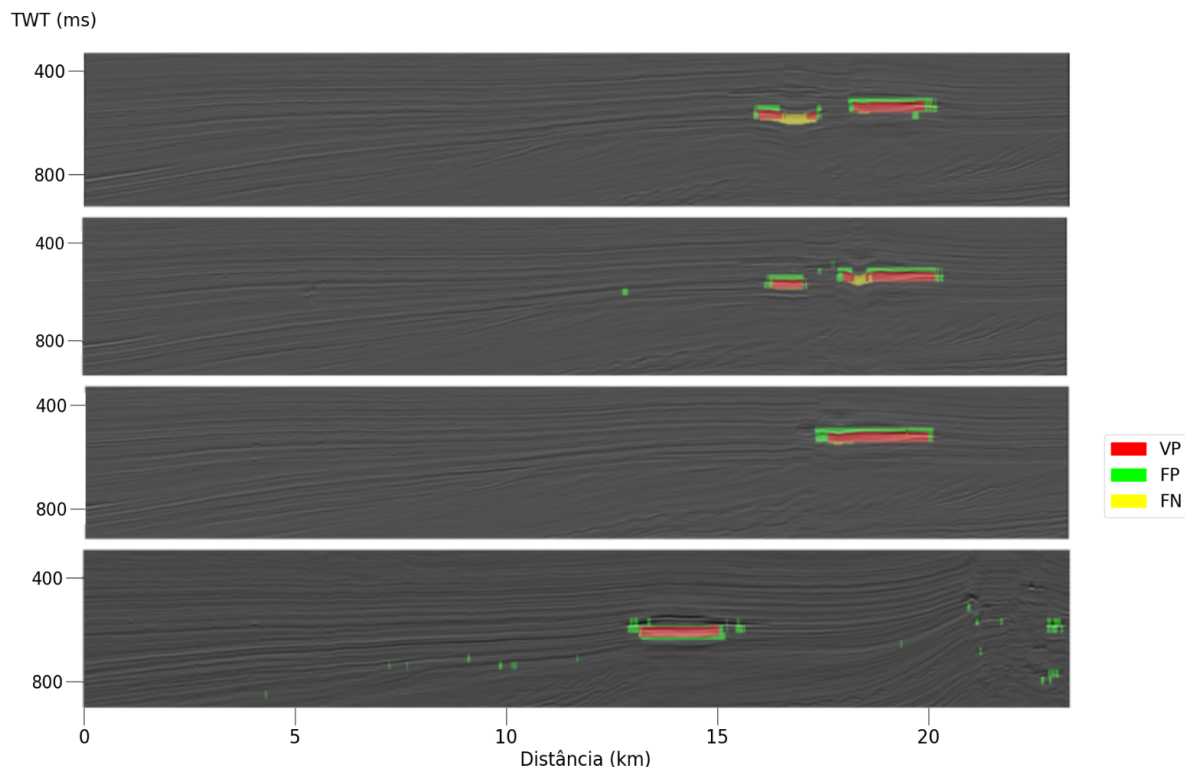


Figura 49 – Sequência de inlines 55, 79, 115 e 581 na política aleatória sem operações morfológicas. Todas as imagens possuem um intervalo de TWT entre 400 ms e 800 ms.

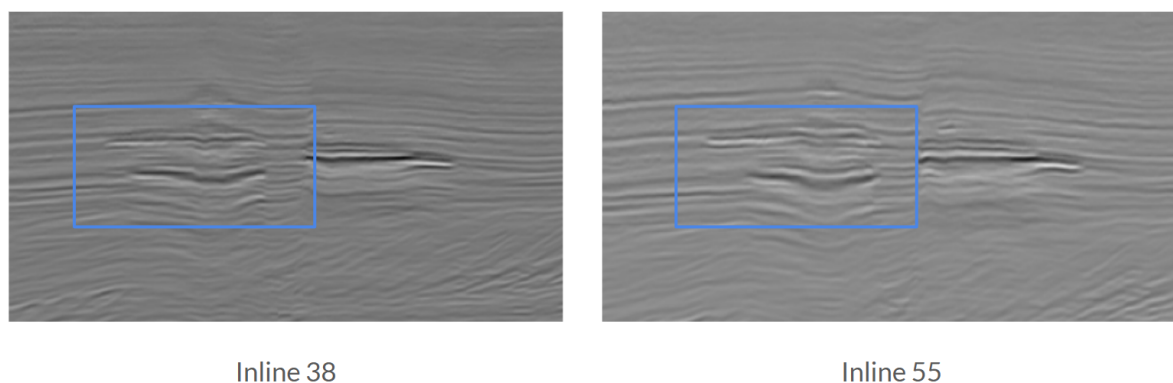


Figura 50 – Detalhe das inlines 38 e 55 que mostram uma perda de resolução na área onde existe falha na segmentação.

resolução no estrato. A Figura 50 permite a visualização desta diferença na resolução.

5.2.2 Experimento 4: Política aleatória com operações morfológicas

O uso das operações morfológicas é justificado pela presença de falsos positivos esparsamente presentes nas imagens sísmicas, além de alguns falsos negativos dentro da região anotada como gás. As operações de abertura e fechamento são aplicadas para remover os falsos positivos ou tornar os falsos negativos em verdadeiros positivos dependendo de suas posições na previsão original. Na Tabela 14 observamos a manutenção dos

valores das métricas dentro da margem de erro e, portanto, sem diferença significativa em comparação com a segmentação original sem as operações morfológicas. Este pode ser considerado um dos efeitos do uso de dados de ambas as regiões tanto para o treino quanto para o teste. Pelo fato de as imagens sísmicas nesta política serem altamente correlacionadas, o resultado acaba por ser similar ao que acontece quando há um vazamento de dados - dados de teste presentes no treino - isto é, o modelo apresenta um desempenho tão bom quanto aquele nas imagens de treinamento.

Tabela 14 – Valores médios das métricas de avaliação com seus respectivos erros padrões para a política aleatória com a aplicação de operações morfológicas no pós-processamento, avaliados pixel a pixel antes da aplicação de operações morfológicas.

Métrica de Avaliação	Com ou sem região de corte	
	Com corte	Sem corte
F1-Score	0.54 ± 0.02	0.55 ± 0.02
PRC	0.41 ± 0.02	0.42 ± 0.01
SEN	0.87 ± 0.03	0.87 ± 0.02

Após o uso das operações morfológicas, a segmentação das inlines mostradas na Figura 49 continuam consistentes entre si porém desta vez com um número menor de falsos positivos e também de falsos negativos, como visto na Figura 51.

As Figuras 52 e 53 mostram a segmentação feita pelo modelo RoBERTa treinado na política aleatória. Os falsos positivos ocorrem, em sua maioria, abaixo de uma profundidade de 1600 ms, na região do chamado horizonte Terciário. Por causa da diferença na composição das rochas e sedimentos no local desta transição, as amplitudes das reflexões sísmicas se tornam mais expressivas e acabam por serem entendidos como pixels pertencentes à classe de gás. No entanto, essas previsões são pontuais e podem ser desconsideradas na prática.

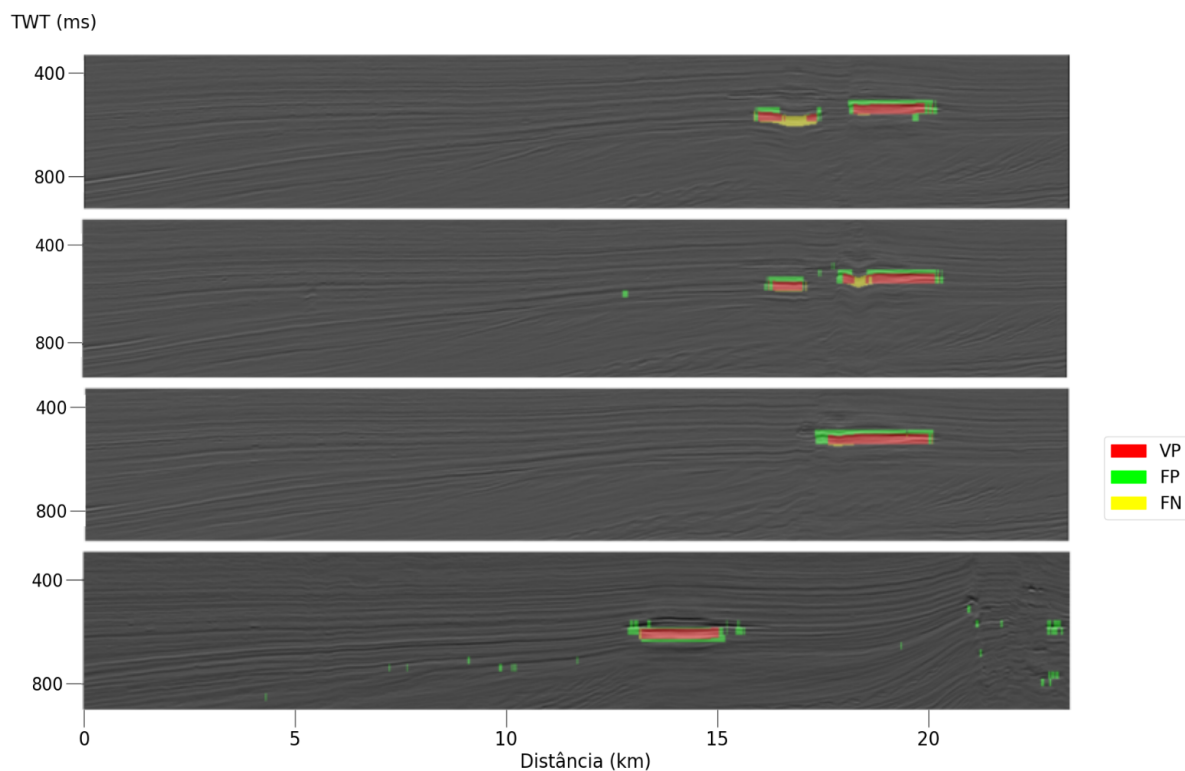


Figura 51 – Sequência de inlines 55, 79, 115 e 581 na política aleatória após a aplicação das operações morfológicas. Todas as imagens possuem um intervalo de TWT entre 400 ms e 800 ms.

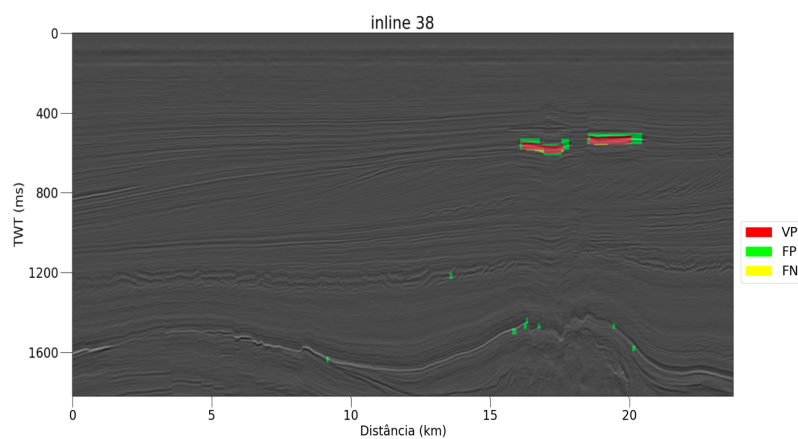


Figura 52 – Segmentação completa da inline 38.

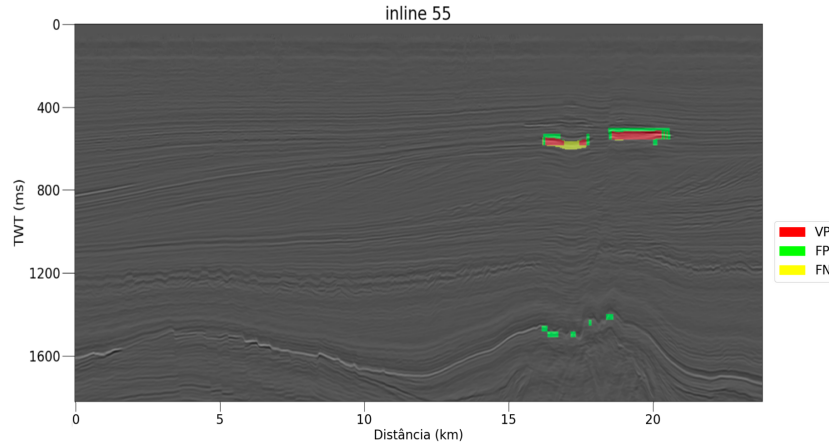


Figura 53 – Segmentação completa da inline 55.

5.2.3 Experimento 5: Política não-aleatória sem operações morfológicas

Em relação aos Experimentos 1 e 2, o Experimento 3 apresenta valores das métricas de avaliação consideravelmente menores na Tabela 15. Esta diferença demonstra o efeito da escolha de política de divisão de testes. Neste Experimento, o modelo desconhece as inlines da acumulação norte usada para teste, e portanto nos fornece uma visão mais clara da capacidade de generalização do modelo para uma região desconhecida pelo modelo.

Tabela 15 – Valores médios das métricas de avaliação com seus respectivos erros padrões para a política não-aleatória sem a aplicação de operações morfológicas no pós-processamento, avaliados pixel a pixel antes da aplicação de operações morfológicas.

Métrica de Avaliação	Com ou sem região de corte	
	Com corte	Sem corte
F1-Score	0.20 ± 0.01	0.31 ± 0.01
PRC	0.15 ± 0.01	0.228 ± 0.009
SEN	0.38 ± 0.03	0.52 ± 0.02

Na Figura 54, vemos a segmentação feita pelo modelo na inline 588 sem o pós-processamento com operações morfológicas. É possível observar a existência de falsos positivos especialmente na porção direita da anotação, além de pequenas linhas de falsos positivos entre alguns verdadeiros positivos. A maior parte dos falsos positivos estão localizados ao redor da anotação. Pela própria natureza da anotação, as extremidades de uma acumulação são de difícil demarcação exata e falsos positivos nessas regiões podem indicar, na verdade, a presença de gás natural não anotados durante pelo especialista. Os falsos positivos que surgem na região superior à direita estão associados a *brightspots*

produzidos por mudanças abruptas na impedância acústica nestes pontos, e podem estar associados a pequenos reservatórios de gás natural que não são economicamente viáveis ou que contêm apenas água, de modo que o especialista não os anotou como gás natural mas o modelo os interpretou como tal.

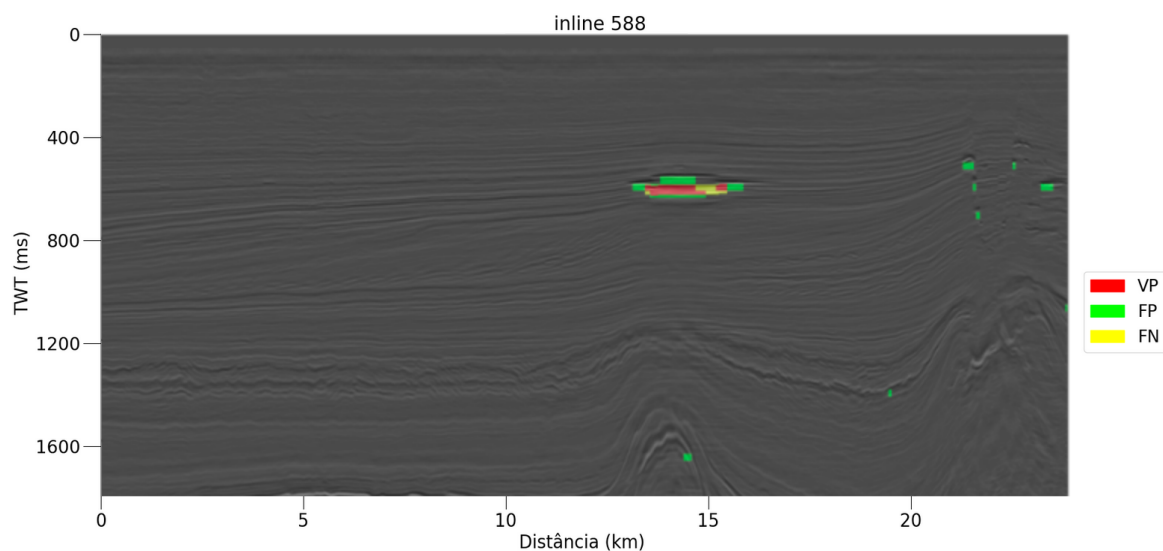


Figura 54 – Segmentação da inline 558 feita pelo modelo treinado, antes das operações morfológicas.

A Figura 55 mostra a ocorrência dos falsos positivos à direita das imagens virtualmente nos mesmos locais, sendo consistentes entre si e indicando a possibilidade de algum tipo de hidrocarboneto, ainda que economicamente inviável, ou água nessas regiões. A última imagem sísmica é a inline original, onde é possível ver a presença de *brightspots* nos locais onde o modelo prevê a presença de gás natural.

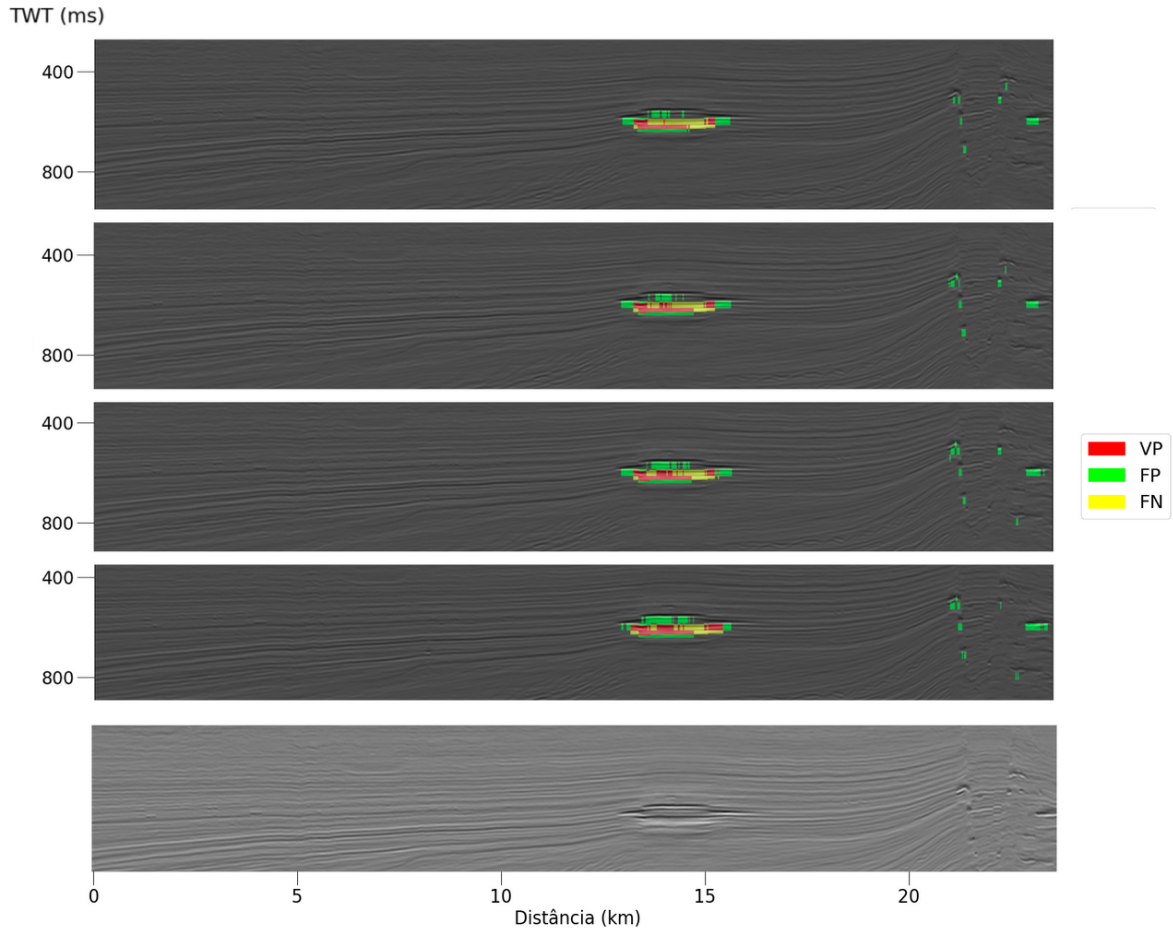


Figura 55 – Segmentação das inlines 587, 588, 589 e 590 (de cima para baixo). A última imagem é a inline 588 sem a segmentação para uma melhor visualização dos brightspots. Os falsos positivos na região à direita aparecem aproximadamente nos mesmos locais.

Neste Experimento também observamos uma diferença maior nas métricas ao considerar ou não a região de corte. Como as imagens pertencentes à região de corte possuem distribuições de pixels diferentes daquelas usadas para treino, o modelo apresenta dificuldades para segmentar corretamente imagens dessa região.

5.2.4 Experimento 6: Política não-aleatória com operações morfológicas

Na política não-aleatória, a vantagem de usar as operações morfológicas fica mais evidente, já que, pela própria definição da política, o modelo apresenta mais dificuldades para fazer uma boa classificação dos pixels. O pós-processamento com operações morfológicas fornece um aumento de 25%, ou 0.13 p.p. na sensibilidade do modelo desconsiderando a região de corte em comparação com a mesma segmentação sem as operações morfológicas. O ganho na precisão é menor mas também aumenta aproximadamente 18% ou 4 p.p. em comparação ao caso sem o pós processamento. Com isso, vemos que o uso

dessas operações implica em uma melhora geral do modelo e melhora sua capacidade de segmentar uma maior extensão do reservatório de gás natural.

Tabela 16 – Valores médios das métricas de avaliação com seus respectivos erros padrões para a política não-aleatória com a aplicação de operações morfológicas no pós-processamento, avaliados pixel a pixel antes da aplicação de operações morfológicas.

Métrica de Avaliação	Com ou sem região de corte	
	Com corte	Sem corte
F1-Score	0.25 ± 0.01	0.36 ± 0.01
PRC	0.18 ± 0.01	0.26 ± 0.01
SEN	0.50 ± 0.03	0.65 ± 0.02

O modelo apresenta desempenho similar em outras inlines do conjunto de testes, como pode ser observado na Figura 56. Essa sequência de inferências ilustra a tendência da segmentação da estrutura que aprisiona o gás natural, ainda que esta estrutura não apresente uma anotação de gás. Com as operações morfológicas, os falsos positivos à direita das inlines se consolidam ao passo que o número de falsos negativos diminuem na região anotada. Isto fica mais claro na Figura 57, onde as OM foram aplicadas na segmentação da inline da Figura 54, onde é observada uma redução de falsos positivos pontuais na região entre 15km e 20km por volta de TWT = 1400 e também um aumento de verdadeiros positivos.

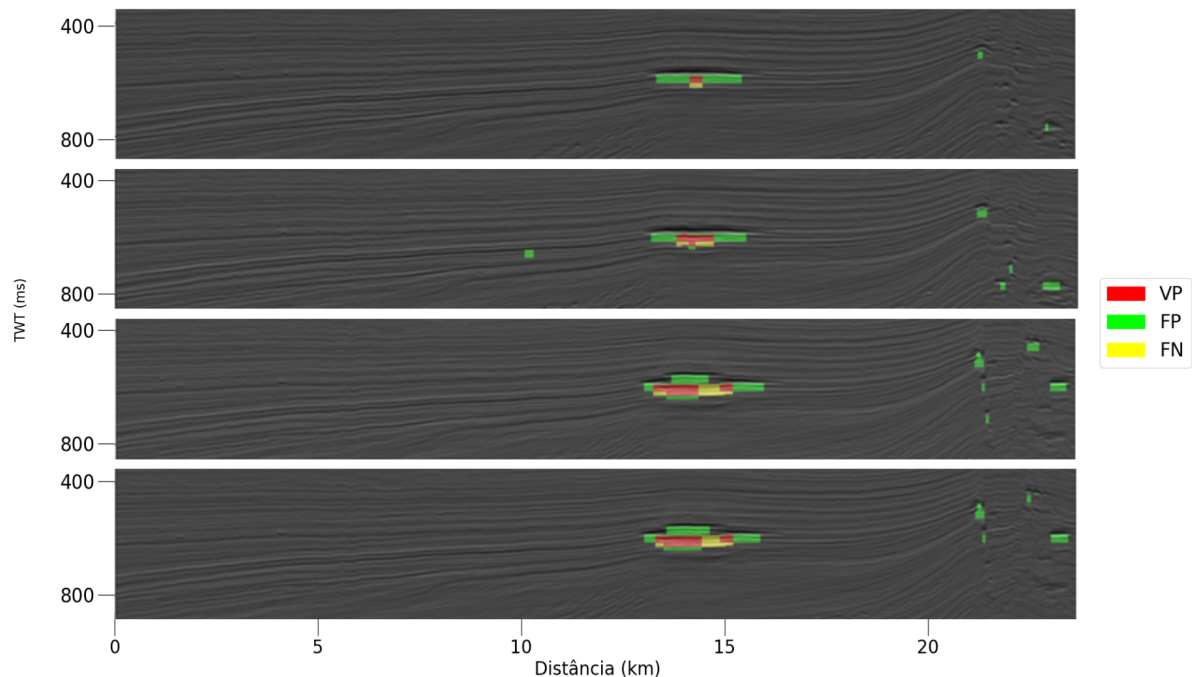


Figura 56 – Inlines 558, 565, 582 e 585 na política não-aleatória após aplicação de operações morfológicas. Todas as imagens apresentam um corte no eixo TWT entre 400 ms e 800 ms.

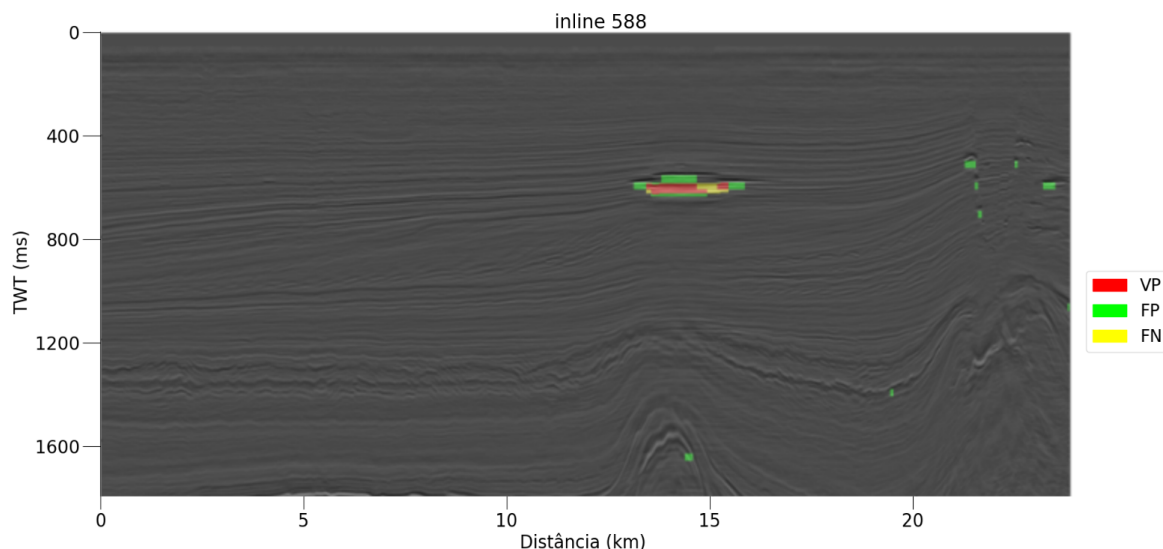


Figura 57 – Segmentação da inline 558 feita pelo modelo treinado, depois das operações morfológicas.

5.2.5 Visualizações axial e tridimensional da segmentação

Para uma melhor visualização da capacidade de segmentação do modelo treinado, vemos na Figura 58 uma visão axial da região de teste. Nela, observamos três gráficos. O primeiro gráfico à esquerda mostra toda a segmentação da região de teste ao longo de toda a profundidade (em ms) dos dados, feita através do empilhamento de todas as segmentações bidimensionais nas imagens sísmicas. No gráfico à direita, temos uma visão um pouco mais frontal, mas ainda inclinada, do volume de teste. Na visualização inferior, consideramos um corte de 400 ms de espessura no volume, com as inlines de 558 a 635. Se qualquer traço possuir pelo menos uma previsão de gás dentro desse volume, consideramos nessa visualização que foi previsto gás natural nessa área. Notamos que a maior parte da acumulação de gás está contida dentro da região prevista pelo modelo. A partir da inline 600, explicitada na imagem, ocorre o problema de corte na imagem e, por consequência, diminui a capacidade do modelo de prever corretamente a presença de gás natural, resultando em um número maior de falsos negativos.

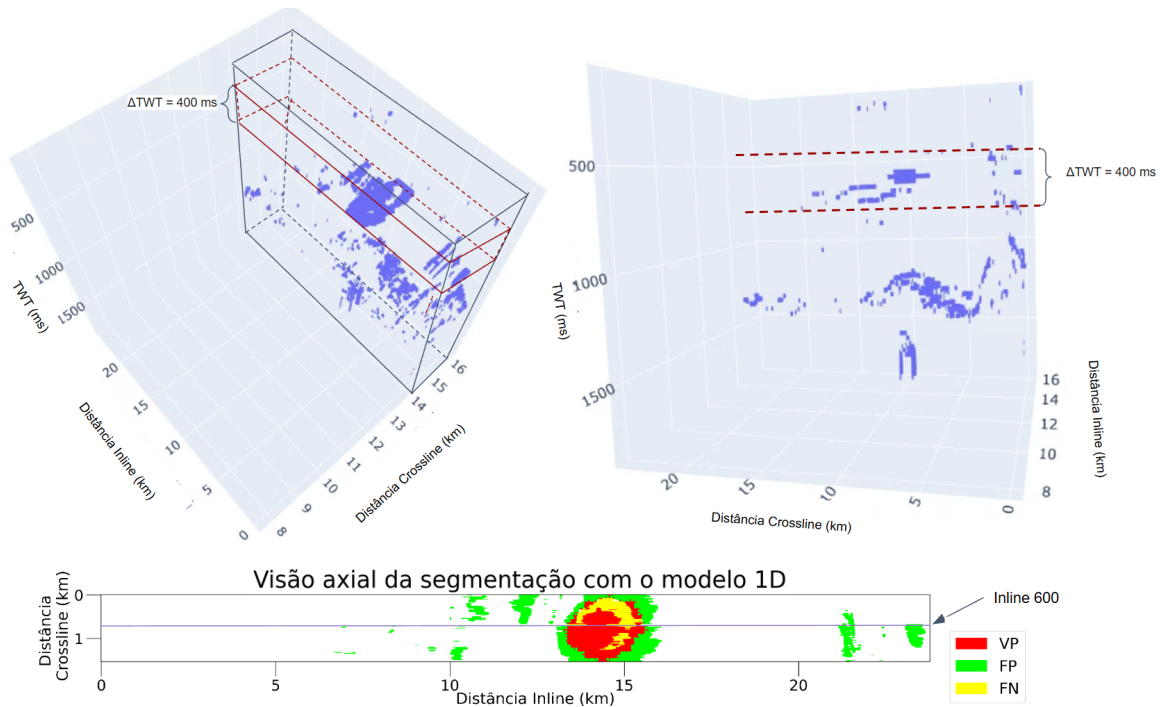


Figura 58 – O gráfico no topo à esquerda mostra como a fatia axial vista na figura inferior é feita com aproximadamente dois quilômetros na direção das inlines, das inlines 558 a 635. A figura no topo à direita mostra melhor, no gráfico 3D, a profundidade considerada na criação da figura inferior. A figura inferior mostra a visão axial do bloco F3 depois das previsões no conjunto de teste.

5.2.6 Experimento 7: Uso de especificação de histograma na política não-aleatória.

Durante os experimentos, foi observado que uma parte dos dados possui uma distribuição de intensidade de pixels e contraste diferente do resto, além de uma resolução menor. Esta informação foi relatada a um especialista que confirmou o problema nos próprios dados disponibilizados publicamente. A diferença entre as distribuições não é percebida nas métricas de avaliação na política aleatória pois os dados desta região também são usados no treinamento e o modelo também aprende a segmentar os traços que estão localizados nela. No entanto, na política não-aleatória este problema torna-se aparente já que o modelo desconhece a distribuição dos dados nesta região. Sendo esta a única fonte para estes dados disponível, foi necessário utilizar um método para contornar este problema. Como mencionado na seção 2.6.3, a técnica de especificação de histograma pode ser usada para aproximar a distribuição de intensidade de pixels entre duas imagens diferentes. Nas imagens sísmicas corrompidas localizadas na região de corte, a técnica é aplicada de forma que sua distribuição de intensidade de pixel fique mais próxima daquela das imagens usadas no conjunto de treino, tornando o contraste mais parecido

entre as imagens dos dois conjuntos. Em particular, o objetivo principal dessa técnica é diminuir o número de falsos positivos ao mesmo tempo em que busca manter o máximo de verdadeiros positivos possível.

Na Figura 59, observamos a ocorrência desse problema em duas inlines. Na inline 604, vemos uma diminuição de falsos positivos, mas em contrapartida também ocorre um aumento de falsos negativos na região de gás, isto é, a aplicação da especificação de histograma também levou a uma piora da segmentação nesta imagem. Por outro lado, na inline 631, existe uma melhora significativa na redução de falsos positivos ao mesmo tempo em que a segmentação da estrutura que armazena o gás se mantém. Esse é um caso extremo em que a precisão aumenta uma ordem de magnitude, demonstrando o potencial que a técnica de especificação de histograma tem para a redução de falsos positivos. Na prática, essa redução é desejada para que o geocientista possa focar seus esforços apenas em áreas com um maior potencial de apresentar acumulações de gás.

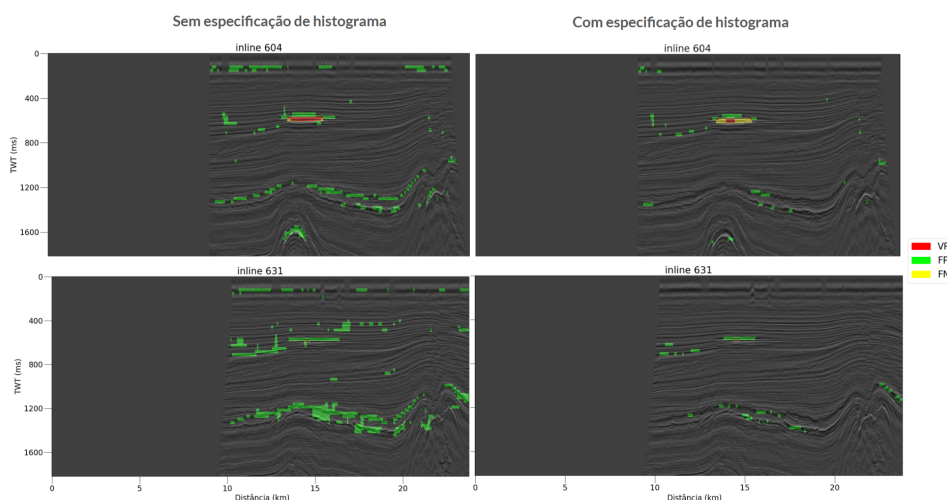


Figura 59 – Comparação da segmentação de duas imagens antes e depois da aplicação da técnica de especificação de histograma.

5.3 Comparação entre as políticas de treino e teste nos Experimentos 1-5

Observamos um maior valor de F1-score, precisão e sensibilidade na política aleatória em comparação com a não-aleatória. Conforme discutido no começo deste capítulo no início da seção 5, este é um resultado esperado por conta da alta correlação entre inlines próximas que são usadas no treino e no teste. Deste modo, após o treinamento, o modelo é capaz de ter um desempenho parecido no conjunto de teste. Além disso, também é capaz de aprender a segmentar os sinais da região de corte, já que parte das inlines dessa região são usadas no treinamento.

As previsões da política não-aleatória, no entanto, possuem uma maior probabilidade de representar o potencial real de generalização de um modelo, visto que os testes são feitos em imagens sísmicas retiradas de áreas geográficas o mais distante possível umas das outras. Ainda que as métricas de segmentação possuam valores menores em relação às da política aleatória, as previsões do modelo identificam a região da acumulação de gás, já que a maior parte dos pixels classificados como gás estão próximos a essa região, sendo efetivamente útil para ajudar o especialista a focar seus esforços no estudo dessa área.

Na política aleatória, a diferença entre as métricas antes e depois das OM está dentro da margem de erro padrão, ou seja, não apresenta segmentação com diferença estatística significativa. O fato da inclusão de imagens altamente correlacionadas estarem presentes tanto no conjunto de treino quanto no de teste leva a uma segmentação ótima no conjunto de teste, dada uma boa segmentação no conjunto de treinamento. Isto resulta em poucos falsos positivos ou falso negativos para serem alterados. Por outro lado, na política não-aleatória, as OM melhoram consideravelmente as métricas médias, com um aumento de 0.12 e de 0.13 na sensibilidade média das imagens de teste na política não-aleatória, passando de 0.38 para 0.50 considerando a região de corte e de 0.52 para 0.65 sem a região de corte. Isto significa que a região de gás está muito mais claramente delineada na imagem segmentada. A precisão também aumenta em 0.03 nos dois casos da política não-aleatória. Esse aumento menor pode ser explicado pelo aumento do número de falsos positivos na região próxima à marcação do especialista, contrabalanceado pela remoção de falsos positivos em regiões afastadas da marcação.

O desempenho do modelo em algumas inlines antes e depois das OM em ambas as políticas de treino-teste podem ser vistas nas tabelas 17 e 18. Na política aleatória, houve um pequeno aumento na sensibilidade em todas as inlines mostradas, enquanto que a precisão foi maior em dois casos - as linhas 79 e 581. O aumento ou diminuição dos valores foi de no máximo 0.03. Por outro lado, a aplicação das OM melhoraram consideravelmente as métricas de desempenho em duas inlines na política não-aleatória: 582 e 585.

O efeito das OM nas métricas das inlines 565 e 585 é mais evidente daquele nas linhas mais ao sul, 558 e 565. Isto acontece porque apresentam diferenças geomorfológicas comparadas com algumas linhas usadas no treinamento, que pertencem à acumulação de gás do sul, ilustradas na Figura 61. Dessa forma, a segmentação original feita pelo

Tabela 17 – Comparação das métricas de inlines selecionadas na política aleatória antes e depois das operações morfológicas (OM). Os melhores resultados para cada inline em cada métrica estão em negrito.

Inline #	Antes das OM			Depois das OM		
	F1-Score	PRC	SEN	F1-Score	PRC	SEN
55	0.58	0.48	0.75	0.59	0.48	0.78
79	0.59	0.46	0.84	0.58	0.44	0.86
115	0.65	0.49	0.96	0.66	0.51	0.97
581	0.51	0.35	0.98	0.48	0.32	0.99

Tabela 18 – Comparação das métricas de inlines selecionadas na política não-aleatória antes e depois das operações morfológicas (OM). Os melhores resultados para cada inline em cada métrica estão em negrito.

Inline #	Antes das OM			Depois das OM		
	F1-Score	PRC	SEN	F1-Score	PRC	SEN
558	0.17	0.09	0.6	0.17	0.09	0.6
565	0.31	0.21	0.69	0.28	0.17	0.69
582	0.30	0.22	0.46	0.35	0.25	0.61
585	0.26	0.22	0.34	0.37	0.27	0.55

RoBERTa não preenche completamente a região de gás, daí o efeito maior das OMs. A Figura 60 mostra o efeito da aplicação das OMs nas linhas 565 e 585, onde o efeito é mais evidente. As OMs expandem as previsões originais, completando espaços entre pixels classificados como positivos na segmentação original. Essa expansão aumenta efetivamente a continuidade da segmentação e, portanto, a sensibilidade.

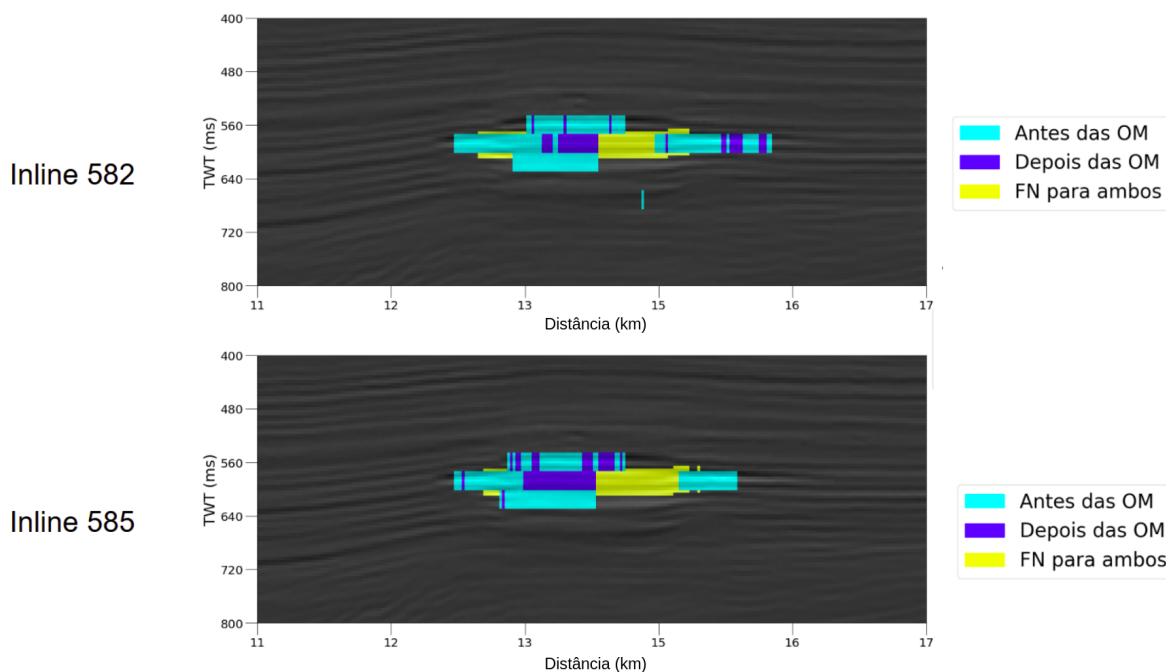


Figura 60 – Segmentação com e sem OM. Há um aumento significativo no número de pixels segmentados corretamente (verdadeiros positivos) depois da aplicação das OM, vistos em azul escuro na figura.

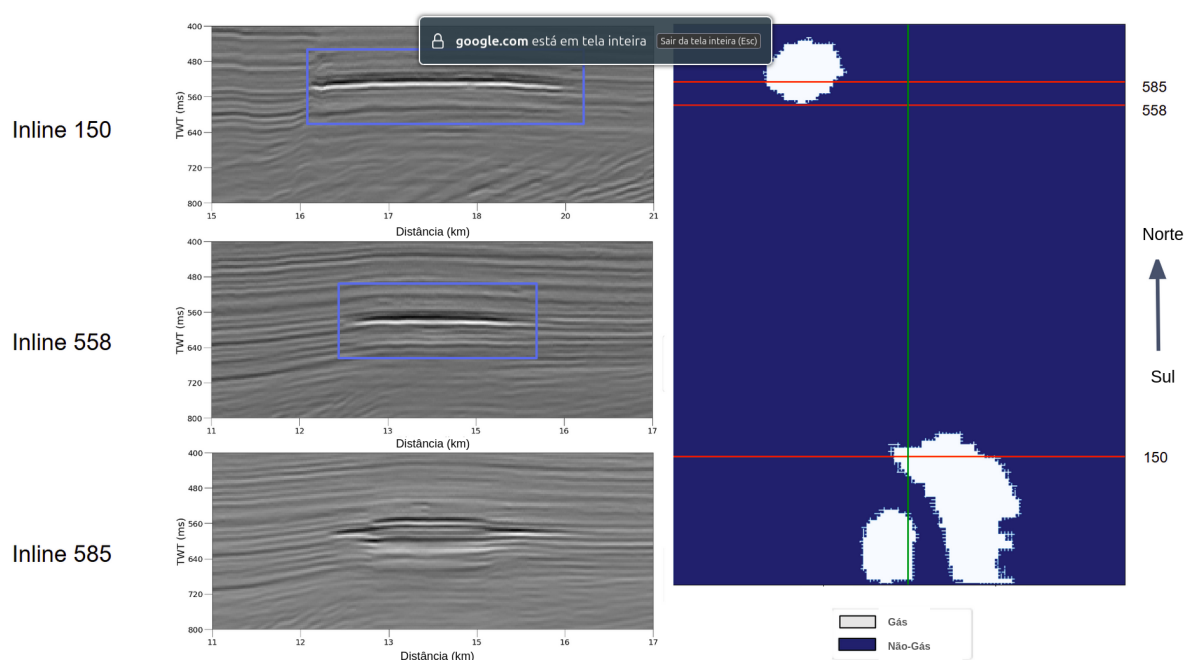


Figura 61 – Similaridade entre a geologia das regiões nas acumulações norte e sul e sua localização na visão axial do volume de dados. Em retângulos azuis, as inlines 150 (sul) e 558 (norte) estão evidenciadas.

A remoção, do conjunto de teste, da região com problemas de imagem aumenta os valores das métricas de avaliação em ambas as políticas de divisão treino-teste. No entanto, o maior aumento ocorre na política não-aleatória: o aumento na sensibilidade média é

de 0.15 e de 0.08 na precisão média. Por outro lado, na política aleatória a sensibilidade média aumenta em 0.01 e a precisão média em 0.05. Esta pequena melhora em comparação com a política não-aleatória é explicada pelo fato de que, no caso aleatório, as imagens sísmicas pertencentes à região de corte também foram usados para treinar o modelo, de modo que este aprendeu os padrões e características dessas imagens corrompidas. No caso não-aleatório, o modelo não teve acesso a esses dados e, portanto, desconhece as características e padrões dessas imagens.

5.4 Método 3: Abordagem 1.5D - RoBERTa Multi-traços

Nesta seção, mostraremos os resultados do método 1.5D, que também utiliza o modelo RoBERTa, porém considera um contexto maior: as t-frases ao redor da qual se deseja classificar. O modelo foi treinado com os parâmetros da Tabela 19.

Tabela 19 – Parâmetros para o treinamento.

Parâmetro	Valor
Hidden layers	8
Attention heads	12
Tamanho do vocabulário	200
N	4
Batch size	8
Épocas	100

5.4.1 Experimento 8: Política aleatória sem operações morfológicas

Na tabela 20 observamos que as métricas caem consideravelmente em relação ao Experimento 1. Apenas a sensibilidade se mantém acima de 0.6. Um dos principais problemas apresentados por este método é a dificuldade em segmentar regiões com perdas de resolução e formato diferente da maior parte dos dados de treinamento. O mesmo já acontecia em menor grau no Método 1D, onde a Figura 50 ilustra a região mencionada. Como exemplo dessa dificuldade, podemos ver as linhas sísmicas 50 e 66, pertencentes à acumulação sul, na Figura 62. Ambas são inlines usadas somente para o teste do modelo na política aleatória. Das duas regiões anotadas, a região à esquerda possui uma baixa taxa de verdadeiros positivos, enquanto que a região à direita possui a maior parte de seus pixels corretamente classificados. Além disso, ambas as inlines também possuem falsos positivos isolados em uma estrutura abaixo de 1400 ms, caracterizada também por

uma diferença entre as refletâncias dos materiais que a formam, mas não associados à presença de hidrocarbonetos como gás natural e petróleo.

Tabela 20 – Valores médios das métricas de avaliação com seus respectivos erros padrões para a política aleatória sem a aplicação de operações morfológicas no pós-processamento, avaliados pixel a pixel antes da aplicação de operações morfológicas.

Métrica de Avaliação	Com ou sem região de corte	
	Com corte	Sem corte
F1-Score	0.26 ± 0.01	0.29 ± 0.02
PRC	0.19 ± 0.01	0.22 ± 0.01
SEN	0.57 ± 0.02	0.59 ± 0.03

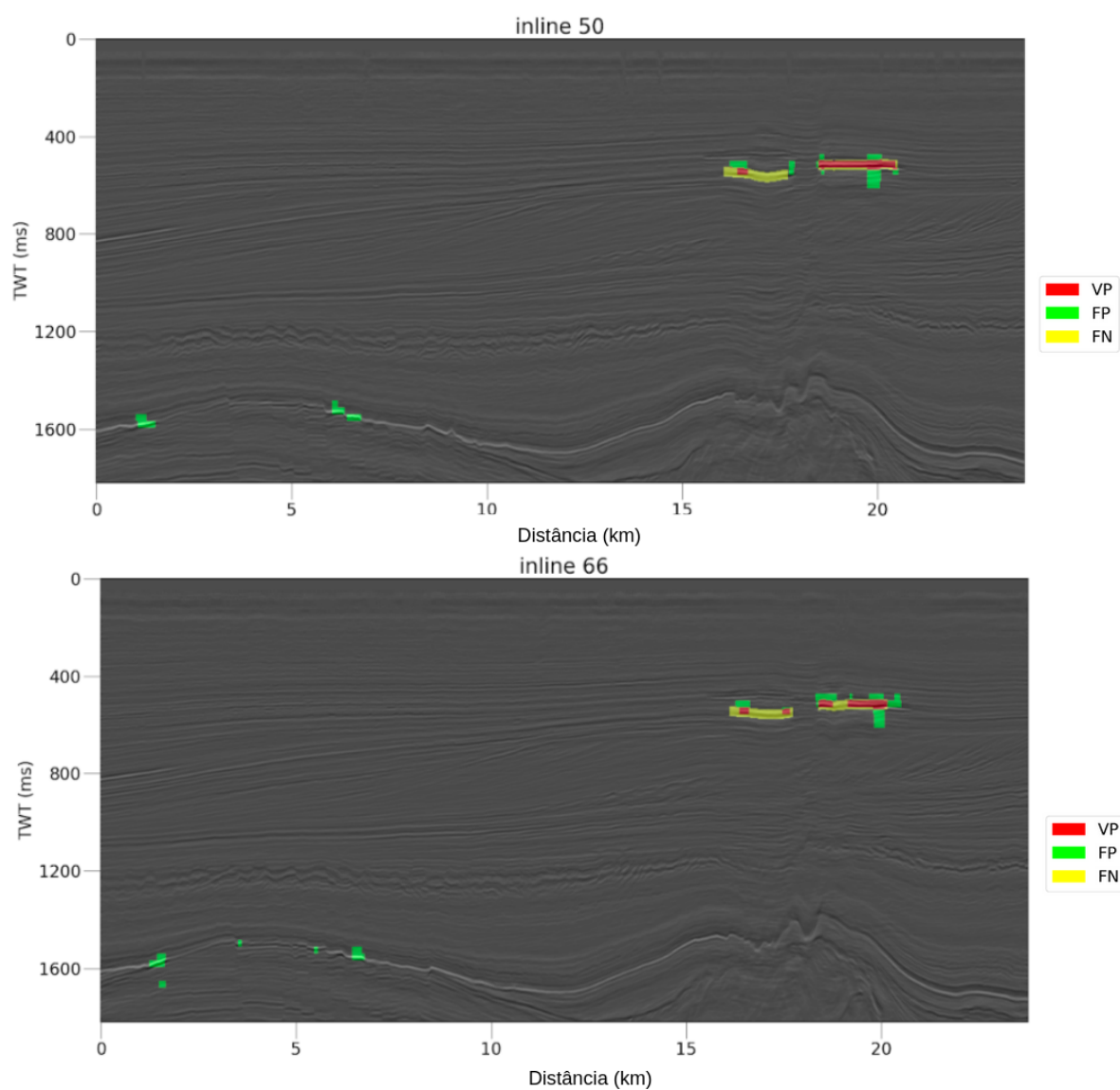


Figura 62 – Segmentação das inline 50 e 66 com o modelo 1.5D.

5.4.2 Experimento 9: Política aleatória com operações morfológicas

Tabela 21 – Valores médios das métricas de avaliação com seus respectivos erros padrões para a política aleatória com a aplicação de operações morfológicas no pós-processamento, avaliados pixel a pixel antes da aplicação de operações morfológicas.

Métrica de Avaliação	Com ou sem região de corte	
	Com corte	Sem corte
F1-Score	0.26 ± 0.02	0.29 ± 0.02
PRC	0.19 ± 0.02	0.22 ± 0.01
SEN	0.64 ± 0.03	0.65 ± 0.03

Pelos valores presentes na tabela 21 observamos que, assim como no Método 1D, na política aleatória os valores das métricas de avaliação no conjunto de teste são similares antes ou depois da aplicação das operações morfológicas por conta da alta correlação entre as imagens usadas no treino e no teste do modelo, o que providencia uma segmentação que tende a ser mais contínua e, portanto, que não se diferencia muito daquela obtida após o pós-processamento com operações morfológicas.

5.4.3 Experimento 10: Política não-aleatória sem operações morfológicas

Neste Experimento, análogo ao Experimento 3, observamos através da Tabela 22 que o método multi-traços apresenta valores maiores em todas as métricas de avaliação tanto no caso que leva em consideração a região de corte quanto aquele que não a considera. Em particular, a sensibilidade do modelo desconsiderando a região de corte, em comparação ao Experimento 3, é 27% ou 14 p.p. maior. Há também um aumento na precisão de 7 p.p. com a região de corte e aproximadamente 13 p.p. sem ela. Por consequência, também há um aumento maior no F1-score de 7 p.p. com a região de corte incluída e 12 p.p. sem a mesma. Os valores mencionados indicam que a adição de um contexto maior na política não-aleatória fornece ao modelo um conhecimento maior do traço a ser segmentado, que se torna capaz de segmentar melhor a região de interesse.

5.4.4 Experimento 11: Política não-aleatória com operações morfológicas

Na Tabela 23 observamos que o uso das operações morfológicas amplificam a melhora da segmentação do Método 1.5D em comparação ao Método 1D na política não-aleatória.

Tabela 22 – Valores médios das métricas de avaliação com seus respectivos erros padrões para a política não-aleatória sem a aplicação de operações morfológicas no pós-processamento, avaliados pixel a pixel antes da aplicação de operações morfológicas.

Métrica de Avaliação	Com ou sem região de corte	
	Com corte	Sem corte
F1-Score	0.25 ± 0.02	0.39 ± 0.01
PRC	0.23 ± 0.02	0.34 ± 0.01
SEN	0.35 ± 0.03	0.57 ± 0.03

Tabela 23 – Valores médios das métricas de avaliação com seus respectivos erros padrões para a política não-aleatória com a aplicação de operações morfológicas no pós-processamento, avaliados pixel a pixel antes da aplicação de operações morfológicas.

Métrica de Avaliação	Com ou sem região de corte	
	Com corte	Sem corte
F1-Score	0.27 ± 0.02	0.43 ± 0.02
PRC	0.23 ± 0.02	0.36 ± 0.01
SEN	0.40 ± 0.04	0.66 ± 0.03

Na acumulação de gás ao norte, o modelo apresenta um desempenho pior que o método 1D. Como pode ser visto na Figura 63, apesar de segmentar a maior parte da região corretamente, apresenta muitos falsos positivos, inclusive em regiões mais afastadas do local anotado.

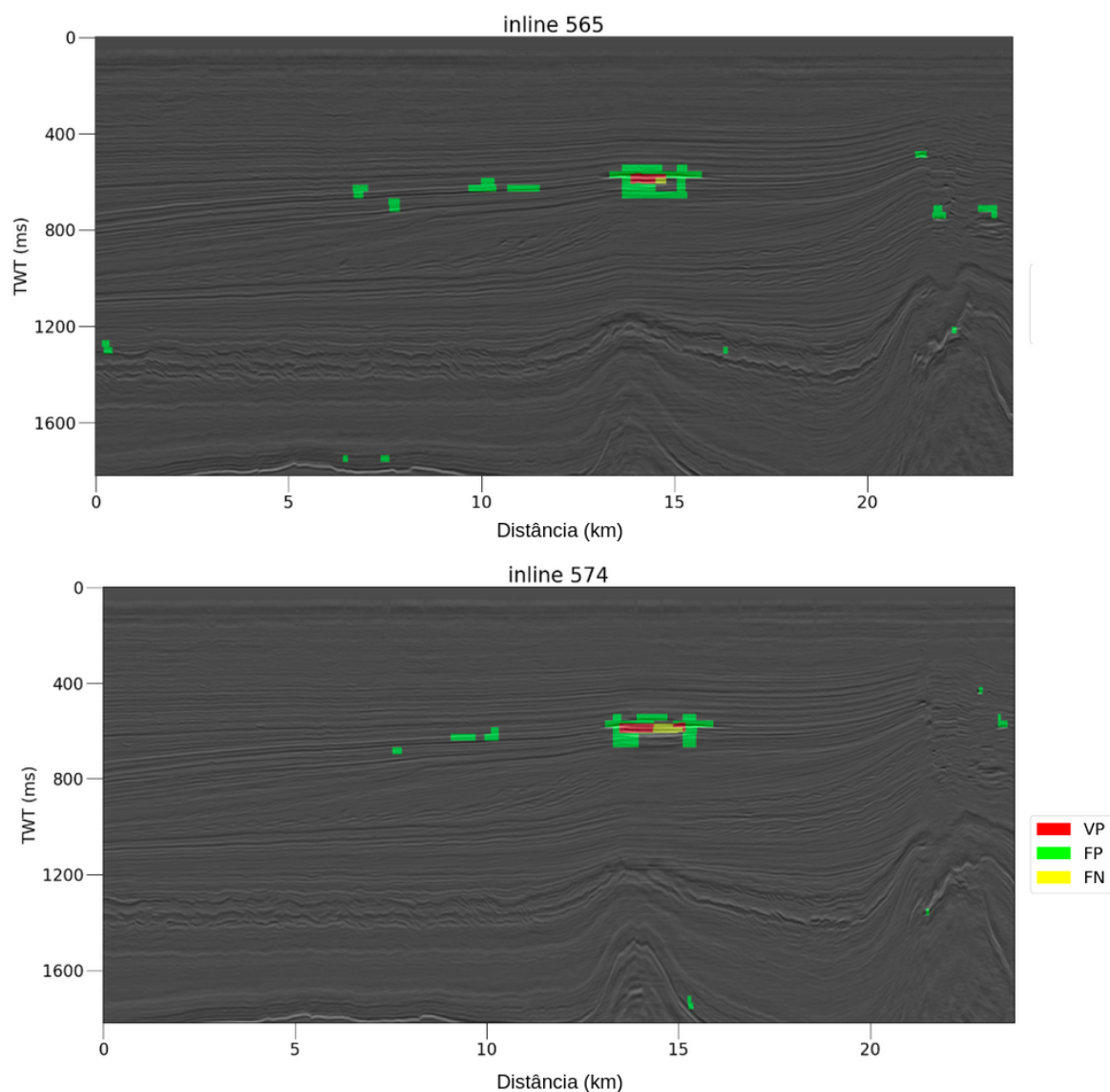


Figura 63 – Segmentação das inline 565 e 574 com o modelo 1.5D.

Por outro lado, como já mencionado, o método 1.5D providencia uma melhora à segmentação. A Figura 64 mostra algumas inlines da acumulação norte segmentadas usando o RoBERTa na abordagem multi-traços. Podemos comparar essas imagens com as da Figura 56 e observar que os falsos positivos nas distancias acima de 20 km não estão mais presentes na segmentação multi-traços. Além disso, as segmentações preenchem mais a área anotada de modo menos linear, como acontece no método 1D que tende a segmentar exatamente em cima das estruturas que armazenam o gás. Esses resultados demonstram que o contexto, isto é, as estruturas ao redor do traço a ser segmentado possuem uma influência considerável na indicação da presença de gás natural. Com informação sobre a geologia na vizinhança da acumulação de gás, o modelo é capaz de ser mais preciso e mais sensível à presença de gás natural.

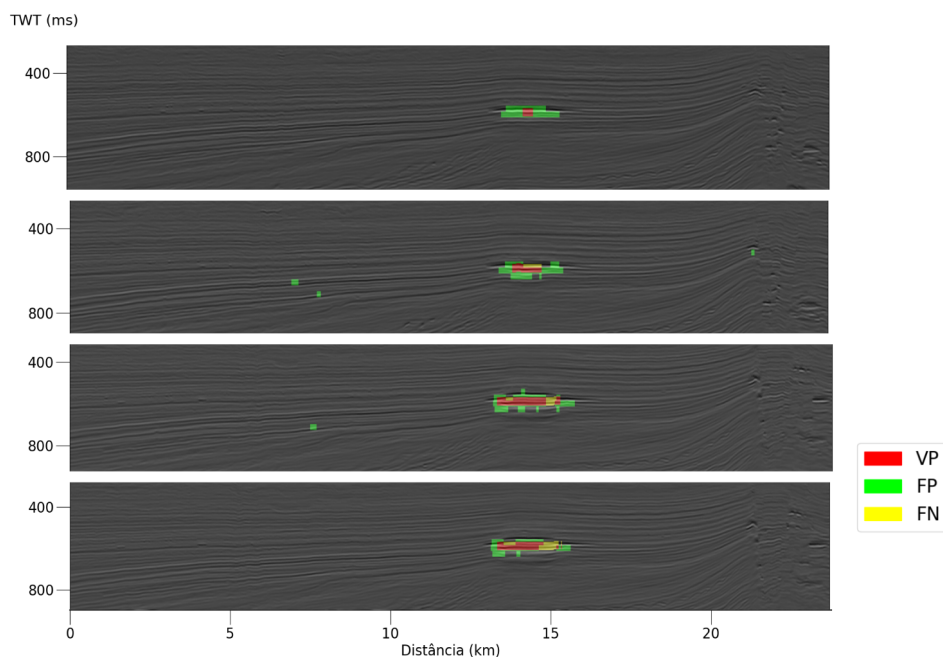


Figura 64 – Inlines 558, 565, 582 e 585 na política não-aleatória para o método multi-traços. Todas as imagens apresentam um corte no eixo TWT entre 400 ms e 800 ms.

Ao comparar uma inline segmentada por inteiro, vemos mais claramente o efeito de se considerar o contexto na maioria dos casos. A Figura 65, por exemplo, mostra que a vasta maioria dos falsos positivos são removidos, aumentando a precisão, e os verdadeiros positivos são mantidos, conservando sua sensibilidade. Na Figura 66 temos um aumento da sensibilidade por conta da faixa das anotações logo abaixo da estrutura que armazena o gás, que não é segmentada no método 1D. Aqui também há uma redução no número de falsos positivos, em especial na camada temporal ao redor de $TWT = 1200$ ms.

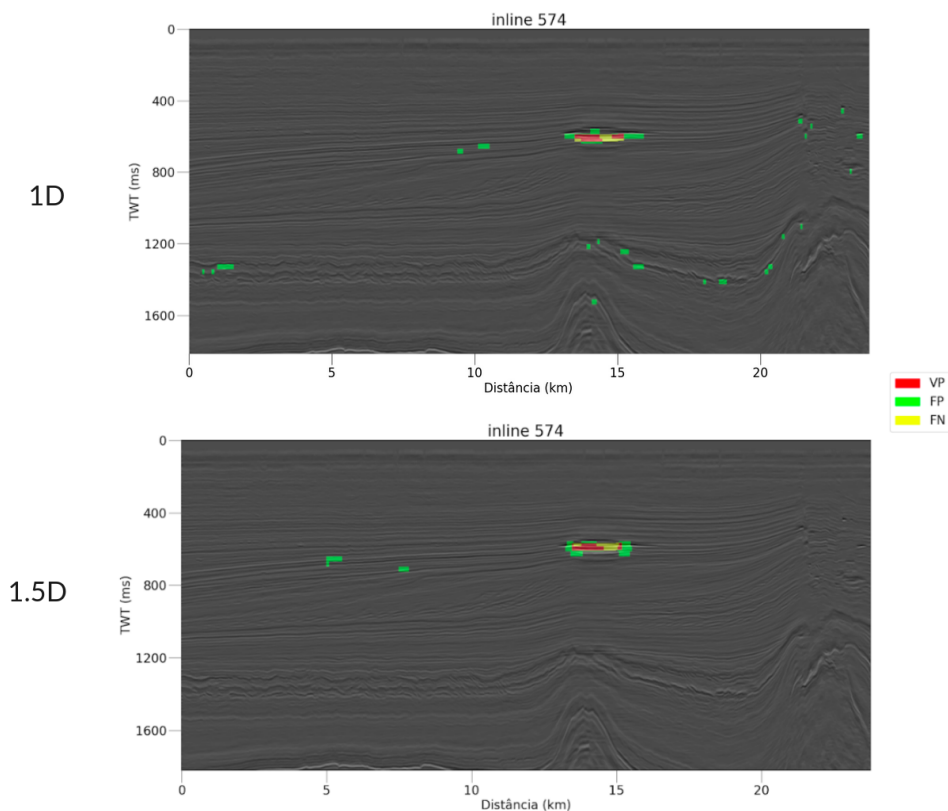


Figura 65 – Comparação entre a segmentação da inline 574 com o modelo 1D e 1.5D.

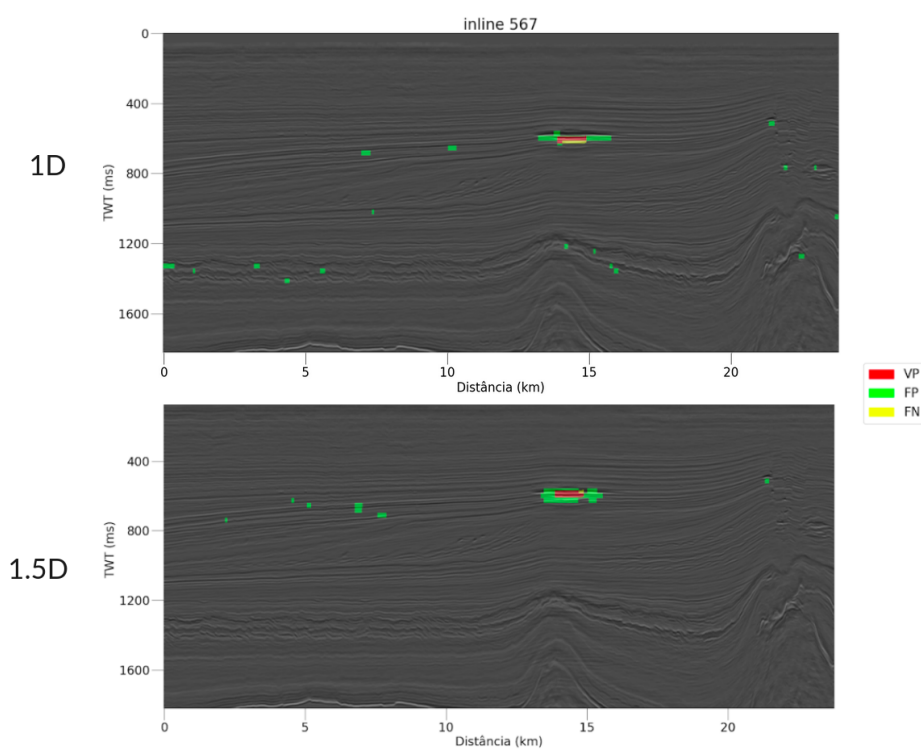


Figura 66 – Comparação entre a segmentação da inline 567 com o modelo 1D e 1.5D.

Apesar da melhora de desempenho de modo geral, o modelo multi-traços apre-

sentam uma piora para algumas inlines mais próximas da região de corte. Por exemplo, na Figura 67 vemos uma piora considerável na segmentação ao considerar o contexto ao redor da t-frase a ser classificada. Ao mesmo tempo em que a maior parte dos falsos positivos foram eliminados, também a maior parte dos verdadeiros positivos não foram detectados. Uma possível causa para esta piora próximo à região de corte é a mudança gradual na qualidade e propriedades dos dados. Ao considerar mais de um único traço na avaliação, a deterioração dos dados contribui mais para uma classificação errônea que leva em conta o ruído adicionado aos traços adjacentes, em oposição ao ruído de um único traço como no método 1D.

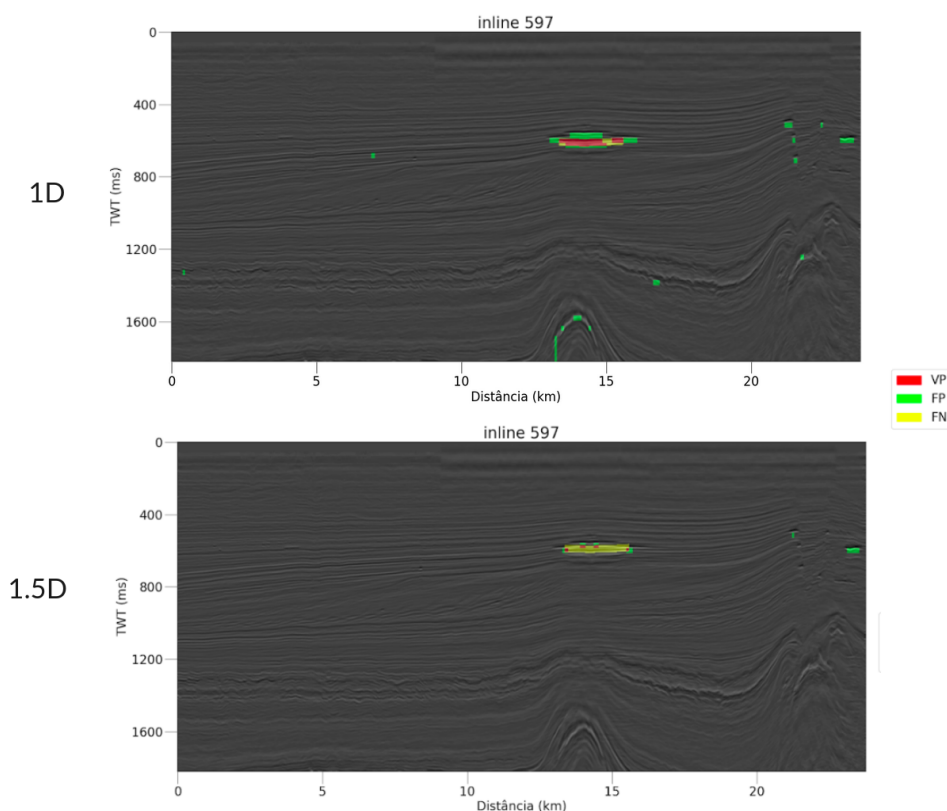


Figura 67 – Comparação entre a segmentação da inline 597 com o modelo 1D e 1.5D.

5.4.5 Comparação entre as políticas de treino e teste nos Experimentos 8-11

Assim como fizemos para o Método 1D, podemos comparar a segmentação em diferentes inlines antes e depois das operações morfológicas para o Método 1.5D. Na Tabela 24 vemos que há um claro aumento nas métricas das inlines selecionadas em comparação com os valores mostrados na Tabela 18. Na Figura 68, vemos o resultado do uso das OMs nas inlines 582 e 585. Diferentemente do que acontece no Método 2, aqui o modelo treinado é capaz de segmentar corretamente boa parte da área anotada, de modo

que as OM's ainda contribuem para um aumento de sensibilidade, porém menos do que no Método 2. A área próxima a 15km no eixo horizontal também não é segmentada neste método, sugerindo uma limitação do modelo RoBERTa nesta região tanto considerando-se apenas o traço a ser segmentado (Método 2) quanto o contexto em que se encontra (Método 3).

Tabela 24 – Comparação das métricas de inlines selecionadas na política não-aleatória antes e depois das operações morfológicas (OM). Os melhores resultados para cada inline em cada métrica estão em negrito.

Inline #	Antes das OM			Depois das OM		
	F1-Score	PRC	SEN	F1-Score	PRC	SEN
558	0.19	0.11	0.7	0.26	0.15	1.0
565	0.35	0.23	0.70	0.37	0.24	0.75
582	0.52	0.42	0.69	0.58	0.46	0.80
585	0.43	0.39	0.48	0.55	0.48	0.65

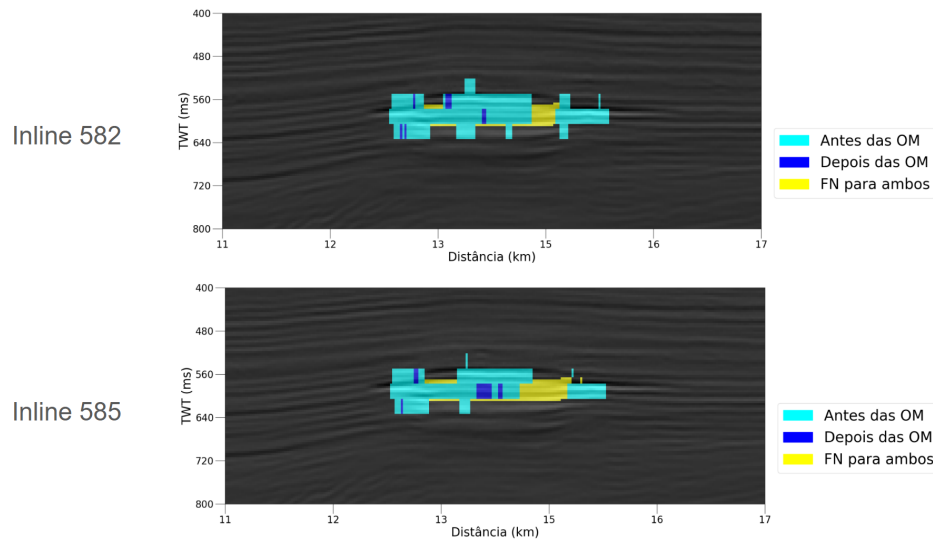


Figura 68 – Segmentação com e sem OM. Existe um aumento na sensibilidade do modelo ao segmentar as duas inlines, porém menor do que no método 2.

Na Figura 69 observamos a segmentação em três dimensões e também sua visão axial. O método multitraços também apresenta problemas para segmentar a região com corte. No entanto, segmenta quase a totalidade da região anterior ao corte (antes da inline de número 600).

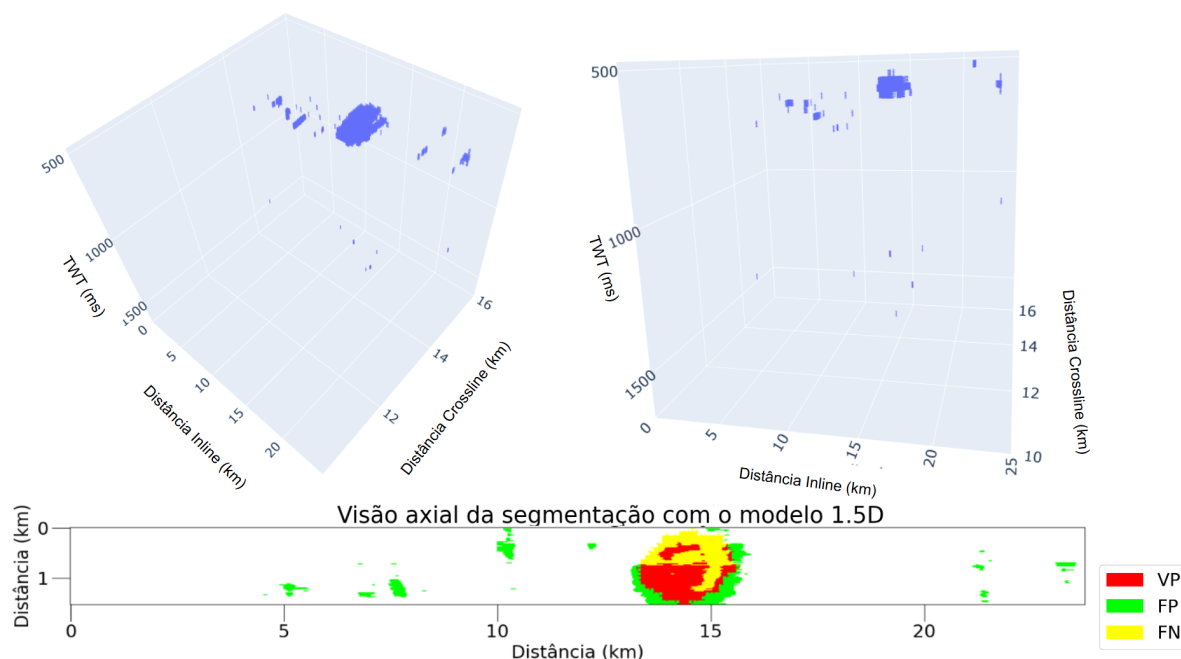


Figura 69 – Os dois gráficos no topo mostram a segmentação em três dimensões das inlines do reservatório de gás ao norte. A figura inferior mostra a visão axial do bloco F3 com as previsões do método baseado no RoBERTa multi-traços.

5.5 Comparação entre métodos

A Tabela 25 de dois dos métodos propostos com outros dois modelos da literatura: LSTM e U-Net. Como não foram encontrados os pesos exatos utilizados para cada um dos modelos na literatura na aplicação em segmentação de gás natural, optou-se por treinar modelos com base nas arquiteturas e parâmetros mencionados em artigos relacionados. Os parâmetros da LSTM utilizados são os mesmos de (SANTOS et al., 2019), enquanto que a arquitetura U-Net utilizada é a utilizada no artigo original (RONNEBERGER; FISCHER; BROX, 2015). A Tabela 26 mostra o mesmo conjunto de métricas, porém apenas para o Método 1. A separação dos resultados deste método em relação aos outros se deve ao fato da espessura da anotação efetiva utilizada ser maior do que a utilizada nos outros métodos, o que afeta diretamente os valores numéricos das métricas e a interpretação dada a cada um.

Para evitar confusão entre a nomenclatura do Método 1 com a U-Net utilizada para comparação, esta última será chamada de U-Net Padrão (U-Net P). A LSTM e a U-Net P são treinadas com as anotações do especialista aumentadas de forma idêntica à usada nos métodos 2 e 3, e por pelo menos 200 épocas, parando o treinamento antes caso não houvesse melhora da métrica de validação do modelo (*early stopping*).

Na política aleatória usada, os métodos com maior desempenho são o Ensemble

Tabela 25 – Comparação de métricas entre os métodos propostos e as redes neurais de referência LSTM e U-Net P, para ambas as políticas aleatória e não-aleatória. Os valores em negrito mostram os melhores resultados dos métodos cuja anotação estendida foi feita de modo similar (todos os métodos com exceção do Ensemble de U-Nets).

Método	Aleatória			Não-aleatória		
	F1-Score	PRC	SEN	F1-Score	PRC	SEN
U-Net P	0.69 $\pm$ 0.05	0.78 $\pm$ 0.05	0.68 $\pm$ 0.05	0.033 $\pm$ 0.009	0.07 $\pm$ 0.02	0.028 $\pm$ 0.007
LSTM	0.57 $\pm$ 0.02	0.42 $\pm$ 0.01	0.92 $\pm$ 0.02	0.17 $\pm$ 0.02	0.21 $\pm$ 0.02	0.15 $\pm$ 0.02
Método 2 (RoBERTa 1D)	0.55 $\pm$ 0.02	0.42 $\pm$ 0.01	0.87 $\pm$ 0.02	0.36 $\pm$ 0.01	0.26 $\pm$ 0.01	0.65 $\pm$ 0.02
Método 3 (RoBERTa 1.5D)	0.29 $\pm$ 0.02	0.22 $\pm$ 0.01	0.65 $\pm$ 0.02	0.43 $\pm$ 0.01	0.36 $\pm$ 0.01	0.66 $\pm$ 0.03

Tabela 26 – Métricas para o Método 1. Por ser necessária a utilização de um alargamento maior da área considerada como sendo gás, e as métricas dependem da marcação utilizada, optou-se por separar estes valores dos anteriores, que usam um alargamento menor das anotações fornecidas pelo especialista.

Método	Aleatória			Não-aleatória		
	F1-Score	PRC	SEN	F1-Score	PRC	SEN
Método 1 (Ensemble U-Net)	0.92 $\pm$ 0.01	0.93 $\pm$ 0.01	0.91 $\pm$ 0.02	0.62 $\pm$ 0.01	0.54 $\pm$ 0.01	0.79 $\pm$ 0.02

U-Net e a U-Net P. Ambas apresentam desempenho melhor que a LSTM e que os métodos 2 e 3 tanto em F1-score quanto em Precisão, ao passo que a LSTM identifica corretamente a maior parte dos pixels anotados como gás, como visto pela sensibilidade média de 0.92 ± 0.02 . No entanto, os métodos propostos nesta tese apresentam desempenhos melhores que ambos os modelos de baseline na política não-aleatória de acordo com todas as três métricas utilizadas, mostrando uma capacidade maior de generalização quando comparada com a LSTM e a U-Net P. Neste ponto é importante relembrar que o Método 1 utiliza um aumento maior da anotação do especialista em relação aos outros métodos propostos e as arquiteturas de comparação, de forma que a comparação numérica direta entre o Método 1 com os Métodos 2 e 3 não deve ser considerada absoluta.

A Tabela 27 mostra métricas de segmentações individuais para quatro inlines ilustradas na Figura 70. Pode ser visto que todos os modelos apresentam bom desempenho na segmentação quando a política aleatória é usada, fornecendo uma segmentação

praticamente contínua nos pixels anotados como gás. A U-Net P apresenta uma segmentação melhor nas inlines pertencentes à acumulação mais ao sul (inlines 55, 79 e 115), enquanto que seu desempenho é muito menor nas acumulações ao norte, representada na Figura pela linha 581. Além da inline 581, grande parte das outras inlines da acumulação norte também possuem baixas sensibilidades. Isto faz com que a sensibilidade média da U-Net P na política aleatória seja de 0.68, já que baixas sensibilidades nas inlines pertencentes à acumulação ao norte são consideradas para o cálculo da média. O Método 1, baseado em um Ensemble de U-Nets, segmenta com alta precisão e sensibilidade todas as inlines usadas no exemplo, mesmo aquela na acumulação norte.

O Método 2 possui falsos positivos ao redor da área anotada porém distribuídos de modo praticamente equidistante do *brightspot*. Possui sensibilidade similar aos outros métodos, com exceção do Método 3. Este, por sua vez, não apresenta um bom desempenho na política aleatória, com muitos falsos negativos - especialmente na inline 55 - e falsos positivos espalhados ao redor da área anotada de maneira disforme. Sua sensibilidade, de modo geral, é baixa e menor que todos os outros modelos e métodos testados.

Tabela 27 – Métricas de segmentação das inlines da Figura 70

Inline #	Método	F1-Score	PRC	SEN
55	LSTM	0.56	0.41	0.90
	U-Net P	0.93	0.93	0.93
	Ensemble U-Net	0.92	0.87	0.97
	RoBERTa 1D	0.58	0.48	0.76
	RoBERTa 1.5D	0.23	0.29	0.19
79	LSTM	0.59	0.43	0.96
	U-Net P	0.91	0.94	0.89
	Ensemble U-Net	0.92	0.87	0.99
	RoBERTa 1D	0.58	0.44	0.86
	RoBERTa 1.5D	0.35	0.30	0.42
115	LSTM	0.61	0.44	0.98
	U-Net P	0.94	0.92	0.96
	Ensemble U-Net	0.91	0.84	0.99
	RoBERTa 1D	0.67	0.51	0.97
	RoBERTa 1.5D	0.68	0.29	0.69
581	LSTM	0.63	0.46	0.97
	U-Net P	0.00	0.00	0.00
	Ensemble U-Net	0.94	0.89	0.99
	RoBERTa 1D	0.48	0.32	0.99
	RoBERTa 1.5D	0.26	0.17	0.64

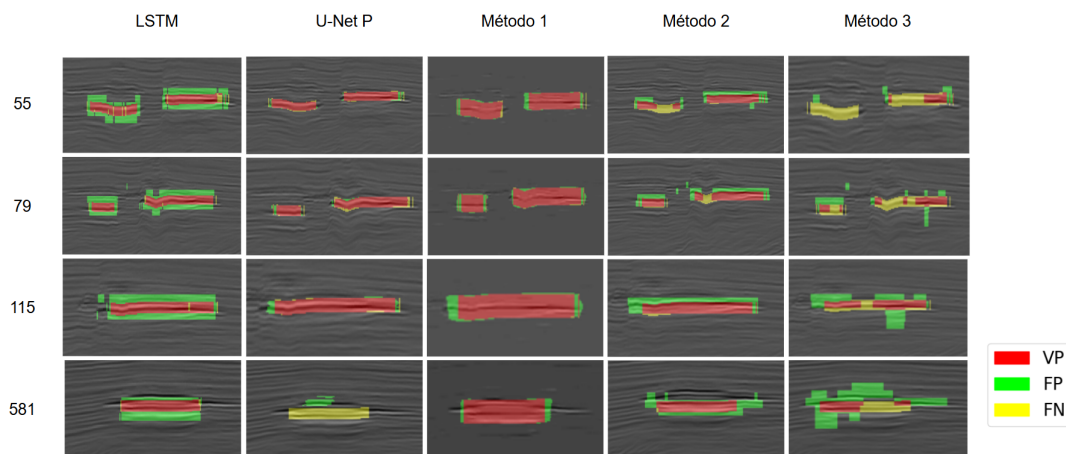


Figura 70 – Segmentação detalhada ao redor da região de gás natural nas inlines 55, 79, 115 e 581 na política aleatória. A U-Net P e o Método 1 (Ensemble U-Net) são os mais precisos na segmentação de inlines que pertencem à maior acumulação de gás (sul), enquanto que a LSTM e os Método 2 e 3 apresentam falsos positivos ao redor da anotação.

Quando analisamos a Tabela 28, vemos que os Métodos propostos possuem desempenho consistentemente melhor que os métodos de presentes na literatura: para cada inline analisada, os métodos propostos apresentam maiores valores das métricas de avaliação. Isto demonstra a confiabilidade dos métodos na identificação de pixels classificados como gás, o que fornece uma assistência melhor ao geocientista ao indicar a região aproximada de um reservatório de gás natural com maior confiança.

A Figura 71 ilustra os resultados da Tabela 28. Nas inlines 582 e 585, a U-Net P prevê gás apenas em uma pequena região acima da região anotada, com nenhum pixel verdadeiro positivo. Já nas inlines 558 e 565, a U-Net P classificou corretamente alguns pixels. A LSTM também apresentou, em sua maioria, falsos negativos, porém com alguns poucos verdadeiros positivos a mais em todas as inlines mostradas na Figura com exceção da inline 558. O Método 1 proposto (Ensemble U-Net), foi o mais capaz de identificar corretamente a região de gás natural e apresenta uma grande sensibilidade, sendo capaz de segmentar a maior parte da região anotada. No entanto, vale ressaltar que, para isso, é necessário ampliar a anotação para que o treinamento seja feito, o que leva a considerar como gás regiões relativamente mais distantes do local anotado pelo especialista. Os Métodos 2 e 3 extrapolam suas segmentações na direção horizontal. Isto é esperado tendo em vista que essas áreas são estruturalmente similares àquela anotada. No entanto, o Método 3 se diferencia do Método 2 ao ter falsos positivos mais próximos da região anotada e em apenas parte da extensão do *bright spot*.

Em síntese, os três métodos possuem vantagens e desvantagens. O Método 1 apre-

Tabela 28 – Métricas da segmentação das inlines mostradas na Figura 71.

Inline #	Método	F1-Score	PRC	SEN
558	LSTM	0.00	0.00	0.00
	U-Net P	0.08	0.06	0.11
	Ensemble U-Net	0.27	0.17	0.67
	RoBERTa 1D	0.16	0.09	0.60
	RoBERTa 1.5D	0.26	0.15	1.00
565	LSTM	0.04	0.07	0.03
	U-Net P	0.15	0.18	0.14
	Ensemble U-Net	0.55	0.43	0.76
	RoBERTa 1D	0.28	0.17	0.69
	RoBERTa 1.5D	0.37	0.24	0.75
582	LSTM	0.17	0.23	0.14
	U-Net P	0.00	0.00	0.00
	Ensemble U-Net	0.73	0.76	0.70
	RoBERTa 1D	0.35	0.25	0.61
	RoBERTa 1.5D	0.58	0.46	0.80
585	LSTM	0.11	0.12	0.09
	U-Net P	0.00	0.00	0.00
	Ensemble U-Net	0.70	0.76	0.64
	RoBERTa 1D	0.37	0.27	0.55
	RoBERTa 1.5D	0.55	0.48	0.65

senta um tempo de treinamento 7,5 vezes mais rápido que os Métodos 2 e 3, considerando-se o tempo de treinamento das 30 U-Nets no Método 1 e o pré-treinamento com o treinamento na tarefa de classificação de traços nos Métodos 2 e 3. Além disso, também é 36 vezes mais rápido para inferir em novas imagens. A segmentação das imagens é mais mais contínua e suave do que em comparação aos métodos baseados em um modelo de linguagem.

Por outro lado, os Métodos 2 e 3 parecem ser capazes de focar nas estruturas que armazenam o gás. Por isso, são capazes de segmentar também regiões menores que foram consideradas “não-gás” pelo primeiro Método, como as estruturas vistas na Figura 56 que estão mais afastadas da região anotada. Estes pontos são, muito possivelmente, pequenas acumulações de gás natural ou petróleo, mas que por serem menores podem ter sido consideradas de pouco valor econômico e não foram devidamente exploradas e, portanto, não foram anotadas como gás natural. O fato dos Métodos 2 e 3 serem capazes de determinar essas (potenciais) acumulações menores podem ser um fator a se considerar na escolha do Método, caso se deseje investigar esse tipo de formação.

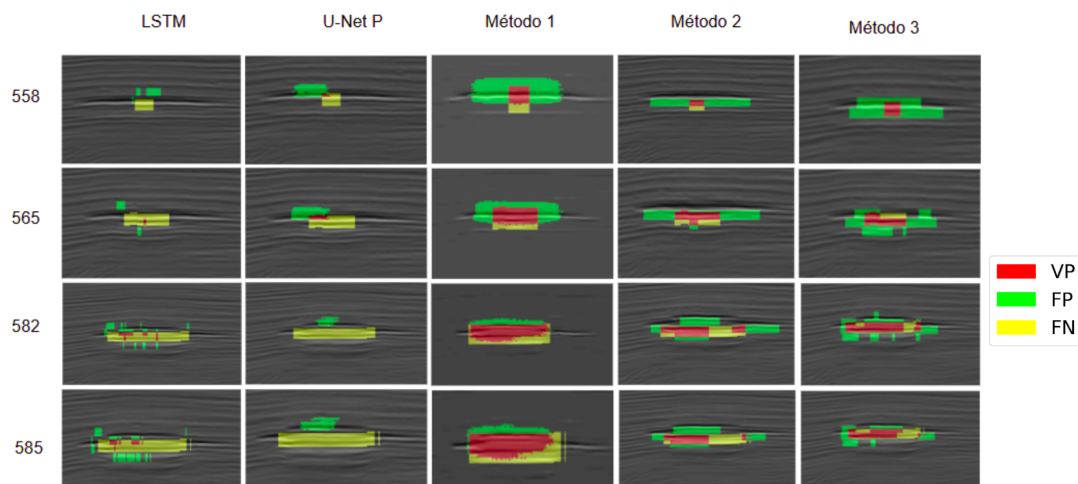


Figura 71 – Segmentação detalhada ao redor da região de gás natural nas inlines 558, 565, 582 e 585 na política não-aleatória. Os métodos propostos mostram desempenho melhor na segmentação das inlines mostradas aqui quando a política de divisão treino-teste usada é a política não-aleatória.

5.6 Considerações finais

Neste capítulo analisamos quantitativamente as segmentações através das métricas F1, Precisão e Sensibilidade em experimentos em duas políticas de divisão de dados em treino e teste. Fizemos, também, uma análise qualitativa das segmentações de todos os métodos propostos, avaliando os aspectos positivos e negativos de cada um.

O Método 1 fornece segmentações mais coesas e contínuas sem a necessidade de pós processamento, ao passo que os Métodos 2 e 3 detectam indícios de hidrocarbonetos (gás natural ou petróleo) em regiões menores, além de delinearem com maior precisão as estruturas capazes de aprisionar o gás natural. Os três métodos superam os modelos estado-da-arte atuais presentes na literatura quando buscamos usar uma divisão de dados que simule a prática na indústria: treinar o modelo com dados de uma região conhecida e explorada (fonte das anotações) e aplicar em regiões que não foram vistas pelo modelo.

No capítulo a seguir sintetizaremos as principais conclusões desta pesquisa, destacando as contribuições para a segmentação de gás natural, suas limitações e perspectivas para trabalhos futuros.

6 CONCLUSÃO

Para avaliar se uma região possui reservatórios de petróleo e gás, e também sua capacidade comercial, empresas atuantes nesse setor realizam levantamentos através de diferentes métodos, sendo o mais comum e amplamente utilizado, o levantamento por reflexão sísmica. Através da análise desses dados, uma vez transformados em imagens, o geocientista é capaz de procurar por formações geológicas que indiquem ou não a presença de hidrocarbonetos. Neste contexto, métodos computacionais podem auxiliá-los nessa tarefa, reduzindo o tempo de avaliação dos dados e identificação de tais estruturas.

Baseado nos resultados apresentados nesta tese, observamos que, dado a natureza dos dados sísmicos, um dos métodos mais comumente empregados em tarefas de segmentação de imagem, a U-Net, apresenta uma instabilidade na segmentação: esta possui grande variação a cada vez que uma U-Net, com hiperparâmetros fixos, é treinada. Isto afeta a confiança no modelo, já que não é possível garantir, em uma dada inferência feita em uma área de exploração desconhecida, se o modelo que faz a segmentação é o que apresenta bons resultados em novas acumulações ou não.

Para resolver esse problema, empregamos duas técnicas de pré-processamento na imagem. Primeiro fazemos uma quantização, isto é, uma discretização dos valores dos sinais de cada traço sísmico presente na imagem sísmica. Depois, utilizamos uma técnica denominada Transformada Wavelet Discreta (DWT, na sigla em inglês). Com ela, removemos características que não são importantes para a identificação de regiões com gás natural ao decompor a imagem e fazer um limiar dos coeficientes para depois reconstruir a imagem que será passada à rede. Também fazemos uso de um *ensemble* de U-Nets com o intuito de estabilizar ainda mais a segmentação final na imagem sísmica e, ao mesmo tempo, aumentar a sensibilidade e precisão do modelo.

Observamos que aplicar somente a DWT aumenta a estabilidade entre as segmentações feitas por diferentes modelos. No entanto, o desempenho da segmentação é similar àquele do método base, isto é, quando se usa as imagens sem pré-processamento algum. Ao quantizar as imagens e aplicar a DWT nelas antes de passá-las à rede e depois realizar um ensemble por maioria dos resultados de diferentes modelos U-Net, temos um desempenho maior de 21% em relação ao da base, além de maior estabilidade.

Mostramos também que é possível modelar um problema de segmentação de ima-

gens utilizando modelos de linguagem. Nesta tese, utilizamos a arquitetura RoBERTa e fazemos classificações unidimensionais de pequenos segmentos de traços sísmicos levando em consideração apenas a informação contida no próprio segmento a ser classificado. O método é análogo a uma classificação de frases no âmbito de processamento de linguagem natural. Estendemos este segundo método em um terceiro, que, além do segmento de traço sísmico sobre o qual se deseja segmentar, considera também segmentos de traços próximos à mesma profundidade. Seu equivalente em processamento de linguagem natural é a classificação de parágrafos.

O modelo baseado na arquitetura RoBERTa é pré-treinada na tarefa de Masked Language Modeling para que a rede seja capaz de identificar relações entre amplitudes sísmicas independentemente da tarefa final desejada, seja ela uma classificação, segmentação, previsão etc. Após a etapa de pré-treinamento, o modelo é treinado na tarefa de classificação, tendo como pesos iniciais aqueles determinados no final do pré-treinamento. Na última etapa, de reconstrução, a classificação dos traços é feita e a imagem segmentada final é reconstruída a partir destas classificações. Aplica-se ainda duas operações morfológicas para a remoção de ruído e aumento da consistência das predições.

Os métodos baseados em modelos de linguagem se mostraram promissores visto que são capazes de identificar estruturas onde ocorrem acumulações de gás natural. Também auxiliam o especialista a encontrar estruturas menores que podem passar despercebidas em uma primeira análise.

Discutimos o conceito de generalização de modelos treinados com dados de uma determinada acumulação de gás serem usados em outros locais cuja presença e quantidade de gás natural ainda é desconhecida. Propomos um método de divisão de dados em treino e teste diferente da utilizada em todos os trabalhos publicados até o momento de escrita desta tese que fazem uso da base de dados F3. Visto que essa base é uma das mais bem documentadas e disponíveis publicamente, é uma das mais utilizadas para o treino de algoritmos cuja finalidade é segmentar ou detectar estruturas em subsuperfície.

A fim de realizar esse estudo, testamos os algoritmos propostos através de duas políticas de divisão de dados, as quais nomeamos política aleatória e política não-aleatória. Argumentamos que a divisão de uma acumulação de gás para treino e outra para teste reproduz melhor a situação real em que se têm dados sobre uma determinada área em subsuperfície e se deseja inferir a presença ou ausência de gás natural em outra região, possivelmente de geologia distinta, em que não se têm dados marcados de antemão.

Ao comparar o desempenho dos três métodos, vemos que dois dos três modelos - 1D e 2D -, apresentam melhores resultados no conjunto de testes quando são treinados na política aleatória, o que é esperado já que as inlines desse conjunto pertencem às duas regiões de gás natural presentes no volume. Sendo assim, existe uma forma de vazamento de dados do conjunto de treino para o conjunto de teste. Por outro lado, com a política não-aleatória obtemos resultados mais próximos do que encontraríamos ao aplicar o modelo treinado em uma região distinta daquela com cujos dados foi treinado. Na política não-aleatória, os modelos treinados apresentam desempenho estado-da-arte na segmentação de acumulações de gás natural no conjunto de dados F3.

Deste modo, as principais contribuições deste trabalho são a adaptação de um método de processamento de linguagem natural para a segmentação de gás natural em imagens sísmicas; a discussão sobre a estabilidade de uma arquitetura muito utilizada em problemas de segmentação quando aplicada no contexto de imagens sísmicas; e a discussão sobre a generalização de modelos treinados com dados sísmicos altamente correlacionados quando usado um determinado conjunto de dados de regiões próximas para treino e teste, o que frequentemente é o caso em levantamentos sísmicos por reflexão.

Nos experimentos realizados, o modelo se mostra eficaz em identificar corretamente a região que coincide com a marcação do especialista, possuindo, ainda sem o pós-processamento, com sensibilidade variando entre 0.66 e 0.79 dependendo do método e da anotação utilizada. A alta sensibilidade implica que grande parte da acumulação de gás marcada pelo especialista é encontrada pelo método. As segmentações são estado-da-arte quando comparadas a trabalhos recentes na política não-aleatória de divisão de dados de treino e teste, além de não necessitarem da escolha prévia de uma região de interesse.

6.1 Trabalhos Futuros

A segmentação de regiões de gás natural no conjunto de dados F3 produzida pelos três métodos propostos são robustas e também generalizáveis quando analisadas em regiões de gás distintas daquelas usadas para treinamento dos modelos. Todos os três modelos apresentam alta sensibilidade, sendo capazes de segmentar a maior parte das regiões anotadas como gás natural pelo especialista. Ainda assim, existem pontos que podem ser trabalhados afim de melhorar seu desempenho, em particular no que tange a

diminuição de falsos positivos da RoBERTa e a estabilidade da U-Net.

O uso de arquiteturas mais atuais e robustas como o ModernBERT (WARNER et al., 2024) e o NeoBERT (BRETON et al., 2025) pode levar a um aumento de desempenho em tarefas com dados sísmicos visto seu maior desempenho em tarefas de linguagem natural quando comparadas às arquiteturas de gerações anteriores como BERT e RoBERTa. Arquiteturas *encoder-decoder*, podem ser treinadas para observar, diretamente através dos pesos do mecanismo de atenção, em quais t-palavras de uma t-frase o modelo focou mais para obter uma determinada classificação de gás/não-gás.

O uso de Grad-CAM (SELVARAJU et al., 2019) em conjunto com a arquitetura U-Net pode facilitar a interpretação dos pontos de atenção do modelo em relação à imagem. Esta técnica deve ajudar na compreensão de possíveis estruturas geológicas que estejam confundindo o modelo, em particular aquelas que podem estar causando a instabilidade da U-Nets individuais.

Do ponto de vista de dados de treinamento, um outro ponto de melhoria é a inclusão de crosslines, além de inlines, nos treinamentos dos modelos 1.5D e 2D, já que irão adicionar informações de uma terceira dimensão e isso dará uma visão mais geral do modelo. Idealmente as crosslines não devem possuir informações sobre outras regiões de gás natural que estarão presentes no conjunto de teste para evitar vazamento de dados do treinamento para o conjunto de teste. No caso 1D, essa inclusão não se faz necessária visto que os traços presentes nas crosslines são os mesmos das inlines e, como este modelo não usa traços adjacentes, nenhuma nova informação é observada com a inclusão de crosslines (todos os traços delas já estão presentes nas inlines).

Ainda é possível fazer o treinamento apenas com os dados da acumulação norte e testar na acumulação sul - inversamente ao que foi feito nos experimentos descritos nesta tese. Isto pode fornecer indícios de como o modelo generaliza, o grau de overfitting ao treinar em cada uma das acumulações e, também, que tipo de geologia é mais facilmente entendida por cada um dos modelos.

Não obstante, sabe-se que, em geral, quando um modelo é treinado em um conjunto de dados, seu uso para inferência em outro conjunto vindos de outro local ou cujo levantamento foi feito com equipamento e processamentos diferentes, não é garantia de bons resultados, ainda que este tenha apresentado bom desempenho no conjunto no qual foi treinado. Por exemplo, os modelos propostos neste trabalho, treinados no bloco F3, não necessariamente reproduziram resultados semelhantes quando aplicado em um con-

junto de dados da Bacia do Parnaíba no Maranhão ou na bacia de Penobscot em Nova Escócia, no Canadá, que possui diferenças quanto à forma de aquisição e tratamento dos dados, além de diferentes características geológicas da região onde são feitos os levantamentos. Dessa forma, o próximo passo deste trabalho seria implementar uma etapa de adaptação de domínio ao método, onde se muda os dados, como, por exemplo, feito em (TAKEMOTO et al., 2019), ou onde a adaptação é feita durante o treinamento do modelo, como em (MA et al., 2019).

As técnicas desenvolvidas nesta tese podem ser usadas em quaisquer problemas que envolvam a identificação, detecção ou segmentação de regiões com determinada característica em imagens sísmicas de reflexão. Atualmente, tendo a transição energética e o processo de descarbonização em vista, técnicas de armazenamento de dióxido de carbono em rochas sedimentares estão sendo desenvolvidas, em particular em rochas reservatórios em campos de hidrocarbonetos. Deste modo, pesquisas sobre técnicas que automatizam este processo de identificação de áreas necessárias para formação e armazenamento de petróleo e gás natural seguirão sendo importantes no futuro.

6.1.1 Produções científicas

Tabela 29 – Referências e respectivos métodos dos trabalhos relacionados a este projeto.

Tipo	Artigo	Qualis	Status
Periódico	de Mello, H.R., de Paiva, A.C., Silva, A.C. et al. Application of natural language modeling techniques in natural gas segmentation in seismic reflection images. Neural Comput & Applic 37, 23832409 (2025). https://doi.org/10.1007/s00521-024-10557-9	A1	Publicado

REFERÊNCIAS

- Adhikari, B.; Dahal, S.; Karki, M.; Mishra, R. K.; Dahal, R. K.; Sasmal, S.; Klausner, V. Application of wavelet for seismic wave analysis in Kathmandu Valley after the 2015 Gorkha earthquake, Nepal. **Geoenvironmental Disasters**, v. 7, n. 1, p. 2, dez. 2020.
- AL-HASHMI, S.; RAWLINS, A.; VERNON, F. A wavelet transform method to detect p and s-phases in three component seismic data. **Open Journal of Earthquake Research**, v. 2, n. 1, p. 1–20, 2013. Disponível em: <<https://www.scirp.org/journal/paperinformation?paperid=28378>>.
- BA, J. L.; KIROS, J. R.; HINTON, G. E. **Layer Normalization**. arXiv, 2016. Disponível em: <<https://arxiv.org/abs/1607.06450>>.
- BADRINARAYANAN, V.; KENDALL, A.; CIPOLLA, R. **SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation**. 2016.
- BANK, D.; KOENIGSTEIN, N.; GIRYES, R. **Autoencoders**. 2021. Disponível em: <<https://arxiv.org/abs/2003.05991>>.
- BRETON, L. L.; FOURNIER, Q.; MEZOUAR, M. E.; CHANDAR, S. **NeoBERT: A Next-Generation BERT**. 2025. Disponível em: <<https://arxiv.org/abs/2502.19587>>.
- CAMESELLE, A. **Mapping of Sedimentary Bodies by 3D Seismic Reflection Data. Application to the pre-Messinian Ebro Margin**. 2010.
- CHANG, D.; YANG, W.; YONG, X.; YANG, Q. Seismic fault detection using deep learning technology. In: _____. **International Geophysical Conference, Beijing, China, 24-27 April 2018**. [s.n.], 2018. p. 1770–1773. Disponível em: <<https://library.seg.org/doi/abs/10.1190/IGC2018-434>>.
- CHENIN, J.; BEDLE, H. Unsupervised machine learning, multi-attribute analysis for identifying low saturation gas reservoirs within the deepwater gulf of mexico, and offshore australia. **Geosciences**, v. 12, n. 3, 2022. ISSN 2076-3263. Disponível em: <<https://www.mdpi.com/2076-3263/12/3/132>>.
- CIPRIANO, C.; JUNIOR, D.; DINIZ, P.; MARIN, L.; PAIVA, A.; DINIZ, J.; SILVA, A. Detection and delimitation of natural gas in seismic images using mlp-mixer and u-net. In: . [S.l.: s.n.], 2022. p. 578–585.
- DANG, W.; ZHOU, B.; WEI, L.; ZHANG, W.; YANG, Z.; HU, S. Ts-bert: Time series anomaly detection via pre-training model bert. In: PASZYNSKI, M.; KRANZLMÜLLER, D.; KRZHIZHANOVSKAYA, V. V.; DONGARRA, J. J.; SLOOT, P. M. A. (Ed.). **Computational Science – ICCS 2021**. Cham: Springer International Publishing, 2021. p. 209–223. ISBN 978-3-030-77964-1.
- DENTITH, M.; MUDGE, S. T. **Geophysics for the Mineral Exploration Geoscientist**. [S.l.]: Cambridge University Press, 2014.
- DEVLIN, J.; CHANG, M.-W.; LEE, K.; TOUTANOVA, K. **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**. arXiv, 2018. Disponível em: <<https://arxiv.org/abs/1810.04805>>.

DINIZ, J.; FERREIRA, J.; CORTES, O. C.; SILVA, A.; PAIVA, A. An automatic approach for heart segmentation in ct scans through image processing techniques and concat-u-net. **Expert Systems with Applications**, p. 116632, 02 2022.

DINIZ, P.; JUNIOR, D. A. D.; DINIZ, J. a. O. B.; PAIVA, A. C. de; SILVA, A. C. d.; GATTASS, M.; QUEVEDO, R.; MICHELON, D.; SIEDSCHLAG, C.; RIBEIRO, R. Time2vec transformer: A time series approach for gas detection in seismic data. In: **Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing**. New York, NY, USA: Association for Computing Machinery, 2022. (SAC '22), p. 6672. ISBN 9781450387132. Disponível em: <<https://doi.org/10.1145/3477314.3507275>>.

EIA, E. **Trends in U.S. Oil and Natural Gas Upstream Costs**. 2016. Url<https://www.eia.gov/analysis/studies/drilling/pdf/upstream.pdf>.

FENG, R.; HANSEN, T. M.; GRANA, D.; BALLING, N. An unsupervised deep-learning method for porosity estimation based on poststack seismic data. **GEOPHYSICS**, v. 85, n. 6, p. M97–M105, 2020. Disponível em: <<https://doi.org/10.1190/geo2020-0121.1>>.

FENG, S.; PASSONE, L.; SCHUSTER, G. Superpixel-based convolutional neural network for georeferencing the drone images. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, PP, p. 1–1, 03 2021.

GHOLAMALINEZHAD, H.; KHOSRAVI, H. **Pooling Methods in Deep Neural Networks, a Review**. 2020. Disponível em: <<https://arxiv.org/abs/2009.07485>>.

HARSUKO, R.; ALKHALIFAH, T. **StorSeismic: A new paradigm in deep learning for seismic processing**. arXiv, 2022. Disponível em: <<https://arxiv.org/abs/2205.00222>>.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: **2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2016. p. 770–778.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural Comput.**, MIT Press, Cambridge, MA, USA, v. 9, n. 8, p. 17351780, nov 1997. ISSN 0899-7667. Disponível em: <<https://doi.org/10.1162/neco.1997.9.8.1735>>.

HU, Y.; YE, K.; KIM, H.; LU, N. **BERT-PIN: A BERT-based Framework for Recovering Missing Data Segments in Time-series Load Profiles**. 2023. Disponível em: <<https://arxiv.org/abs/2310.17742>>.

JÚNIOR, D. A. D.; Batista da Cruz, L.; Otávio Bandeira Diniz, J.; SILVA, A. C.; Cardoso de Paiva, A.; GATTASS, M.; RODRIGUEZ, C.; QUISPE, R.; RIBEIRO, R.; RIGUETE, V. Detection of potential gas accumulations in 2d seismic images using spatio-temporal, pso, and convolutional lstm approaches. **Expert Systems with Applications**, v. 215, p. 119337, 2023. ISSN 0957-4174. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417422023557>>.

KAZEMI, S. M.; GOEL, R.; EGHBALI, S.; RAMANAN, J.; SAHOTA, J.; THAKUR, S.; WU, S.; SMYTH, C.; POUPART, P.; BRUBAKER, M. **Time2Vec: Learning a Vector Representation of Time**. arXiv, 2019. Disponível em: <<https://arxiv.org/abs/1907.05321>>.

KEAREY, P.; BROOKS, M.; HILL, I. **Geofísica de exploração**. [S.l.]: Oficina de textos, 2009. 91 p.

KIM, K.; EE, J.-H.; KIM, K.; KIM, U.-R.; LEE, J. Mechanical snells law. **Journal of the Korean Physical Society**, Korean Physical Society, v. 76, n. 4, p. 281285, fev. 2020. ISSN 1976-8524. Disponível em: <<http://dx.doi.org/10.3938/jkps.76.281>>.

KOPRDI, M.; PRODANOVI, N.; LJAJI, A.; BAARAGIN, B.; MILOEVI, N. From zero to hero: Harnessing transformers for biomedical named entity recognition in zero- and few-shot contexts. **Artificial Intelligence in Medicine**, v. 156, p. 102970, 2024. ISSN 0933-3657. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0933365724002124>>.

LASSAU, N.; AMMARI, S.; CHOUZENOUX, E.; GORTAIS, H.; HERENT, P.; DEVILDER, M.; SOLIMAN, S.; MEYRIGNAC, O.; TALABARD, M.-P.; LAMARQUE, J.-P.; DUBOIS, R.; LOISEAU, N.; TRICHELAI, P.; BENDJEBBAR, E.; GARCIA, G.; BALLEYGUIER, C.; MERAD, M.; STOCLIN, A.; JEGOU, S.; GRISCELLI, F.; TETELBOUM, N.; LI, Y.; VERMA, S.; TERRIS, M.; DARDOURI, T.; GUPTA, K.; NEACSU, A.; CHEMOUNI, F.; SEFTA, M.; JEHANNO, P.; BOUSAID, I.; BOURSIN, Y.; PLANCHET, E.; AZOULAY, M.; DACHARY, J.; BRULPORT, F.; GONZALEZ, A.; DEHAENE, O.; SCHIRATTI, J.-B.; SCHUTTE, K.; PESQUET, J.-C.; TALBOT, H.; PRONIER, E.; WAINRIB, G.; CLOZEL, T.; BARLESI, F.; BELLIN, M.-F.; BLUM, M. G. B. Integrating deep learning ct-scan model, biological and clinical variables to predict severity of covid-19 patients. **Nature Communications**, v. 12, n. 1, p. 634, Jan 2021. ISSN 2041-1723. Disponível em: <<https://doi.org/10.1038/s41467-020-20657-4>>.

LATIMER, R. B.; DAVIDSON, R.; RIEL, P. van. An interpreter's guide to understanding and working with seismic-derived acoustic impedance data. **The Leading Edge**, v. 19, n. 3, p. 242–256, 2000. Disponível em: <<https://doi.org/10.1190/1.1438580>>.

LECUN, Y.; KAVUKCUOGLU, K.; FARABET, C. Convolutional networks and applications in vision. In: **Proceedings of 2010 IEEE International Symposium on Circuits and Systems**. [S.l.: s.n.], 2010. p. 253–256.

LIN, T.-Y.; GOYAL, P.; GIRSHICK, R.; HE, K.; DOLLÁR, P. **Focal Loss for Dense Object Detection**. 2018. Disponível em: <<https://arxiv.org/abs/1708.02002>>.

LIU, Y.; OTT, M.; GOYAL, N.; DU, J.; JOSHI, M.; CHEN, D.; LEVY, O.; LEWIS, M.; ZETTLEMOYER, L.; STOYANOV, V. **RoBERTa: A Robustly Optimized BERT Pretraining Approach**. arXiv, 2019. Disponível em: <<https://arxiv.org/abs/1907.11692>>.

MA, X.; XU, P.; WANG, Z.; NALLAPATI, R.; XIANG, B. Domain adaptation with BERT-based domain classification and data selection. In: **Proceedings of the 2nd Workshop on Deep Learning Approaches for Low-Resource NLP (DeepLo 2019)**. Hong Kong, China: Association for Computational Linguistics, 2019. p. 76–83. Disponível em: <<https://aclanthology.org/D19-6109>>.

NALLEY, S.; LAROSE, A. **International Energy Outlook 2021 (IEO2021)**. 2021. Url<https://www.eia.gov/outlooks/ieo/pdf/IEO2021ReleasePresentation.pdf>.

RADFORD, A.; WU, J.; CHILD, R.; LUAN, D.; AMODEI, D.; SUTSKEVER, I. Language models are unsupervised multitask learners. **OpenAI**, 2019. Accessed: 2024-11-15. Disponível em: <https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf>.

ROBINSON, E. A.; TREITEL, S. **Geophysical Signal Analysis**. Society of Exploration Geophysicists, 2000. Disponível em: <<https://library.seg.org/doi/abs/10.1190/1.9781560802327>>.

RONNEBERGER, O.; FISCHER, P.; BROX, T. **U-Net: Convolutional Networks for Biomedical Image Segmentation**. 2015.

SANTOS, L. F.; SILVA, R. M. G. E.; GATTASS, M.; SILVA, A. C. Direct hydrocarbon indicators based on long short-term memory neural network. In: _____. **SEG Technical Program Expanded Abstracts 2019**. [s.n.], 2019. p. 2373–2377. Disponível em: <<https://library.seg.org/doi/abs/10.1190/segam2019-3215628.1>>.

SANTOS, L. F. T.; GATTASS, M.; RODRIGUEZ, C.; HURTADO, J.; MIRANDA, F.; MICHELON, D.; RIBEIRO, R. Improving generalization through self-supervised learning using generative pre-training transformer for natural gas segmentation. **Computers Geosciences**, v. 196, p. 105809, 2025. ISSN 0098-3004. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098300424002929>>.

SELVARAJU, R. R.; COGSWELL, M.; DAS, A.; VEDANTAM, R.; PARIKH, D.; BATRA, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. **International Journal of Computer Vision**, Springer Science and Business Media LLC, v. 128, n. 2, p. 336359, out. 2019. ISSN 1573-1405. Disponível em: <<http://dx.doi.org/10.1007/s11263-019-01228-7>>.

SENNRICH, R.; HADDOW, B.; BIRCH, A. Neural machine translation of rare words with subword units. In: ERK, K.; SMITH, N. A. (Ed.). **Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**. Berlin, Germany: Association for Computational Linguistics, 2016. p. 1715–1725. Disponível em: <<https://aclanthology.org/P16-1162/>>.

SEPULVEDA, L. F. M.; GATTASS, M.; SILVA, A. C.; QUEVEDO, R.; MICHELON, D.; SIEDSCHLAG, C.; RIBEIRO, R. Generalization of deep learning models for natural gas indication in 2d seismic data. **Pattern Recognition**, v. 141, p. 109642, 2023. ISSN 0031-3203. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0031320323003436>>.

SILVA, R.; BARONI, L.; FERREIRA, R.; CIVITARESE, D. S.; SZWARCMAN, D.; BRAZIL, E. V. Netherlands dataset: A new public dataset for machine learning in seismic interpretation. 03 2019.

SILVA, R. M.; BARONI, L.; FERREIRA, R. S.; CIVITARESE, D.; SZWARCMAN, D.; BRAZIL, E. V. **Netherlands Dataset: A New Public Dataset for Machine Learning in Seismic Interpretation**. arXiv, 2019. Disponível em: <<https://arxiv.org/abs/1904.00770>>.

TAKEMOTO, N.; COIMBRA, T.; ARAÚJO, L.; TYGEL, M.; AVILA, S.; BORIN, E. Enriching synthetic data with real noise using neural style transfer. In: . [S.l.: s.n.], 2019. p. 1–1.

THOMAS, G.; FLORES-TAPIA, D.; PISTORIUS, S. Histogram specification: A fast and flexible method to process digital images. **IEEE Transactions on Instrumentation and Measurement**, v. 60, n. 5, p. 1565–1578, 2011.

TORAMAN, C.; YILMAZ, E. H.; Ş AHI

NUÇ, F.; OZCELIK, O. Impact of tokenization on language models: An analysis for turkish. **ACM Transactions on Asian and Low-Resource Language Information Processing**, Association for Computing Machinery (ACM), v. 22, n. 4, p. 1–21, mar 2023. Disponível em: <<https://doi.org/10.1145%2F3578707>>.

VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, L.; POLOSUKHIN, I. **Attention Is All You Need**. arXiv, 2017. Disponível em: <<https://arxiv.org/abs/1706.03762>>.

WANG, X.; LIU, F.; SU, R.; WANG, Z.; BAI, L.; OUYANG, W. **SeisMoLLM: Advancing Seismic Monitoring via Cross-modal Transfer with Pre-trained Large Language Model**. 2025. Disponível em: <<https://arxiv.org/abs/2502.19960>>.

WANG, Z.; YAN, W.; OATES, T. Time series classification from scratch with deep neural networks: A strong baseline. In: **2017 International Joint Conference on Neural Networks (IJCNN)**. [S.l.: s.n.], 2017. p. 1578–1585.

WARNER, B.; CHAFFIN, A.; CLAVIÉ, B.; WELLER, O.; HALLSTRÖM, O.; TAGHADOUINI, S.; GALLAGHER, A.; BISWAS, R.; LADHAK, F.; AARSEN, T.; COOPER, N.; ADAMS, G.; HOWARD, J.; POLI, I. **Smarter, Better, Faster, Longer: A Modern Bidirectional Encoder for Fast, Memory Efficient, and Long Context Finetuning and Inference**. 2024. Disponível em: <<https://arxiv.org/abs/2412.13663>>.

WEN, Q.; ZHOU, T.; ZHANG, C.; CHEN, W.; MA, Z.; YAN, J.; SUN, L. **Transformers in Time Series: A Survey**. arXiv, 2022. Disponível em: <<https://arxiv.org/abs/2202.07125>>.

WU, Y.; SCHUSTER, M.; CHEN, Z.; LE, Q. V.; NOROUZI, M.; MACHEREY, W.; KRIKUN, M.; CAO, Y.; GAO, Q.; MACHEREY, K.; KLINGNER, J.; SHAH, A.; JOHNSON, M.; LIU, X.; KAISER, .; GOUWS, S.; KATO, Y.; KUDO, T.; KAZAWA, H.; STEVENS, K.; KURIAN, G.; PATIL, N.; WANG, W.; YOUNG, C.; SMITH, J.; RIESA, J.; RUDNICK, A.; VINYALS, O.; CORRADO, G.; HUGHES, M.; DEAN, J. **Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation**. arXiv, 2016. Disponível em: <<https://arxiv.org/abs/1609.08144>>.

XU, J.; WU, H.; WANG, J.; LONG, M. **Anomaly Transformer: Time Series Anomaly Detection with Association Discrepancy**. arXiv, 2021. Disponível em: <<https://arxiv.org/abs/2110.02642>>.

ZHANG, D.; ZUO, W.; ZHANG, D.; ZHANG, H. Time series classification using support vector machine with gaussian elastic metric kernel. In: **2010 20th International Conference on Pattern Recognition**. [S.l.: s.n.], 2010. p. 29–32.

ZHANG, H.; LIU, Y.; QIU, Y.; LIU, H.; PEI, Z.; WANG, J.; LONG, M. **TimesBERT: A BERT-Style Foundation Model for Time Series Understanding**. 2025. Disponível em: <<https://arxiv.org/abs/2502.21245>>.

ZHANG, Y.; LIU, Y.; ZHANG, H.; XUE, H. Seismic facies analysis based on deep learning. **IEEE Geoscience and Remote Sensing Letters**, v. 17, n. 7, p. 1119–1123, 2020.

_____. Seismic facies analysis based on deep learning. **IEEE Geoscience and Remote Sensing Letters**, v. 17, n. 7, p. 1119–1123, 2020.

ZINI, J. E.; RIZK, Y.; AWAD, M. A deep transfer learning framework for seismic data analysis: A case study on bright spot detection. **IEEE Transactions on Geoscience and Remote Sensing**, v. 58, n. 5, p. 3202–3212, 2020.