



UNIVERSIDADE FEDERAL DA BAHIA - UFBA

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA - IME

PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA - PGMAT

DISSERTAÇÃO DE MESTRADO



APRENDIZADO ATIVO PROFUNDO PARA CLASSIFICAÇÃO
DE FÁCIES SÍSMICAS

NAYGUEL DE CASTRO COSTA

Área de Concentração: ESTATÍSTICA

Salvador - Bahia

JUNHO DE 2023

APRENDIZADO ATIVO PROFUNDO PARA CLASSIFICAÇÃO DE FÁCIES SÍSMICAS

NAYGUEL DE CASTRO COSTA

Dissertação de Mestrado apresentada ao
Colegiado da Pós-Graduação em Matemática
da Universidade Federal da Bahia (UFBa),
como parte dos requisitos para obtenção do
título de Mestre em Matemática. Área de
concentração: Estatística.

Orientador: Prof. Dr. Paulo Canas Rodrigues
Co-orientador: Prof. Dr. Luciano Rebouças
de Oliveira

Salvador - Bahia

Junho de 2023

Aprendizado ativo profundo para classificação de fácies sísmicas

Nayguel de Castro Costa

Dissertação apresentada ao Colegiado do Curso de Pós-graduação em Matemática da Universidade Federal da Bahia, como requisito parcial para obtenção do Título de Mestre em Matemática.

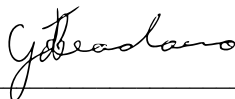
Banca examinadora



Prof. Dr. Paulo Jorge Canas Rodrigues (UFBA)



Prof.ª. Dr.ª. Alexandro Guerra Cerqueira (UFBA)



Prof. Dr. George Teodoro (UFMG)

A Lenisia, Alice, Levi e Nicolas. Como é bom acordar, dormir e viver com vocês.

Agradecimentos

Primeiramente gostaria de expressar minha profunda gratidão a Lenísia, minha esposa, companheira e melhor amiga. A decisão de ingressar neste programa de mestrado representou um grande sacrifício de tempo para mim e, conseqüentemente, um aumento significativo em suas responsabilidades. Agradeço por em um momento crítico de decisão, quando estávamos ajoelhados em oração buscando discernir a vontade de Deus, ela ter me olhado e dito: “façamos assim, sinto que vai ser bom para o seu mestrado”. Além disso, ela contribuiu diretamente revisando o texto.

Agradeço a minha mãe, Joseny, por acreditar em mim como ninguém mais. Agradeço pelo seu entusiasmo, apoio e amor de tantos anos.

Agradeço a minha cunhada, a Dra. Leticia Marques, que me incentivou desde o momento em que decidi me inscrever na seleção de um mestrado em estatística, enveredando por uma área que não dominava. Ela é uma pesquisadora de referência e, por muitas vezes, me ajudou a entender melhor o processo de pesquisa.

Aos meus filhos, Alice, Levi e Nicolas, expressei minha gratidão por serem minha inspiração para me tornar uma pessoa melhor. Eles são as crianças mais sagazes, divertidas e lindas deste mundo.

Agradeço aos meus familiares mais próximos: Pai Chico, Allan, Patricia, Patrick, Dominique, Joana, Antonio, Heleni, Alder, Clara e Arthur. Sua ajuda é constante e essencial em cada aspecto de minha vida.

Agradeço ao meu orientador, Prof. Dr. Paulo Canas Rodrigues, pelo acompanhamento durante todo o processo do mestrado, sempre me orientando em busca da excelência, mas fazendo isso com fino trato, respeito e bom humor. Agradeço também ao meu Co-orientador, Prof. Dr. Luciano Rebouças de Oliveira, por fornecer o conhecimento especializado necessário para a execução deste trabalho e por suas observações técnicas que me permitiram enxergar aspectos que, de outra forma, não teria percebido.

Agradeço aos demais membros da banca, o Prof. Dr. Alexsandro Guerra Cerqueira, Prof. Dr. George Luiz Medeiros Teodoro, a geofísica da Petrobras Dra. Fernanda Figueiredo Farias e Prof. Dr. Paulo Henrique Ferreira da Silva. Suas contribuições são de grande relevância para o trabalho desenvolvido.

Ao longo dessa jornada, muitos professores contribuíram para o meu processo de aprendizado e formação como ser humano. A todos eles, devo um agradecimento especial.

Preciso agradecer ao meu amigo Dr. Jorge Leite. Quando me senti perdido, sem saber como organizar minhas ideias e meu tempo, ele se disponibilizou a ser um consultor, proporcionando conversas e opiniões fundamentais para o meu processo de escrita.

Aos novos amigos que fiz durante a realização deste mestrado, cuja convivência, embora virtual, foi muito agradável, gostaria de expressar minha gratidão. Estudamos juntos, enfrentamos momentos de dificuldade, mas conseguimos superá-los. Em especial agradeço a Sandro e Hellen.

Agradeço a PETROBRAS, por ser uma empresa que investe na qualificação e treinamento de seus funcionários. Em particular, agradeço à minha gerente, Camilla Almeida, por ter me indicado para o CIAG (Curso de Inteligência Artificial Aplicado a Geociências), onde pude adquirir as habilidades de programação fundamentais para a realização deste trabalho.

Agradeço aos colegas de trabalho, Fabricio Miranda e Saulo Carreiro, por nosso trabalho em equipe no CIAG que foi o ponto de partida para minha jornada no tema do Aprendizado Ativo.

Por fim, agradeço também ao carinho e incentivo dos amigos Argemiro Barroso, Luis Claudio Maia, Thiago Barreto, João Vyctor Garcia e Fernando Cezar. Todos estiveram presentes na minha vida ao longo desses meses e me deram o apoio e ânimo necessário para cumprir essa etapa de minha trajetória.

A todos aqueles, que de alguma forma, contribuíram para o meu percurso no mestrado, expresso minha sincera gratidão.

*"Da escola idêntica sai a sociedade
igual."
(Victor Hugo)*

Resumo

A interpretação de fácies sísmicas é um aspecto crítico da exploração de petróleo e gás. Porém, à medida que o volume e resolução dos dados sísmicos aumenta, não se torna prático para os intérpretes humanos analisar minuciosamente cada parte dos dados. Para mitigar esse problema, métodos de interpretação baseados em aprendizado profundo têm ganhado atenção recentemente. No entanto, adquirir um conjunto de dados de treinamento suficientemente grande e anotado, dentro dos cronogramas de projeto, continua sendo um desafio. Para superar esse obstáculo, métodos de aprendizado ativo foram propostos. Esses métodos reduzem a quantidade de anotação nos dados de treinamento ao criar um conjunto de treinamento rotulado otimizado a partir de dados não rotulados.

Neste estudo, uma rede neural profunda do tipo *encoder-decoder* foi desenvolvida para executar a tarefa de classificação de fácies sísmicas. Em seguida, foi aplicado um protocolo de aprendizado ativo com três diferentes estratégias de amostragem. A pesquisa foi feita utilizando o conjunto de dados público Parihaka. Adicionalmente, é introduzida uma estratégia inovadora para avaliar o intervalo de confiança nas curvas de aprendizado ativo, utilizando *bootstrap*. Os resultados mostraram que métricas comparáveis às do modelo de referência foram alcançadas usando menos da metade do conjunto de dados de treinamento rotulado, mesmo empregando métodos simples, como a amostragem aleatória. A amostragem por incerteza mostrou-se a mais eficaz dentre as estratégias de amostragem estudadas, apresentando potencial para não apenas priorizar as imagens mais informativas, mas também identificar imagens não informativas. Esses resultados promissores sugerem que a incorporação de técnicas de aprendizado ativo pode melhorar a praticidade e eficiência da interpretação sísmica baseada em aprendizado profundo, reduzindo a dependência de grandes conjuntos de dados de treinamento anotados.

Palavras-chave: Aprendizado Profundo, Aprendizado Ativo, Interpretação Sísmica, Segmentação de Imagens.

Abstract

Seismic facies interpretation is a critical aspect of oil and gas exploration, yet it is not practical for human interpreters to analyze every part of the data thoroughly as the volume and resolution of seismic data increases. To address this issue, deep learning-based interpretation methods have gained attention. However, acquiring a sufficiently large and accurately labeled training dataset within project timelines remains a challenge. Active learning methods have been proposed to overcome this obstacle. They reduce the number of required training labels by creating an optimized labeled training set from unlabeled data.

In this study, we developed an end-to-end encoding-decoding deep neural network for seismic facies classification and applied an active learning workflow with three distinct query strategies. The research was made using the Parihaka public dataset. Additionally, we introduced a unique bootstrap-based strategy to assess the confidence interval for the active learning curves. Our results showed that comparable outcomes to the baseline model could be achieved using less than half of the labeled training dataset, even when employing rudimentary methods, such as random sampling. Notably, the uncertainty sampling proved to be the most effective among the query strategies studied, as it has the potential to not only prioritize the most informative images but also identify uninformative ones. These promising findings suggest that the incorporation of active learning techniques can enhance the practicality and efficiency of deep learning-based seismic interpretation by reducing the reliance on large, labeled training datasets.

Key words: Deep Learning, Active Learning, Seismic Interpretation, Image Segmentation.

Sumário

1	Introdução	1
1.1	Motivação	2
1.2	Objetivo	3
1.3	Organização do trabalho	3
2	<i>Deep learning</i> aplicado a prospecção sísmica de petróleo	5
2.1	Fundamentos da prospecção sísmica de petróleo	5
2.2	Interpretação sísmica	7
2.3	Segmentação de imagens aplicada em imagens sísmicas	8
2.4	Deep learning para interpretação sísmica	10
3	Redes neurais convolucionais	14
3.1	Aprendizagem	14
3.2	Redes neurais	15
3.2.1	Gradiente descendente	16
3.2.2	Retropropagação do erro	17
3.3	Redes neurais convolucionais	17
3.3.1	Arquitetura	17
3.3.1.1	Camada convolucional	19
3.3.1.2	Camada ReLU	20
3.3.1.3	Pooling	21
3.3.1.4	Convolução transposta	21
3.3.1.5	Skip connection	22
3.3.1.6	Softmax	23
3.3.1.7	Visão geral	23
3.3.2	Classificação de imagens	24
3.4	U-Net	24
3.5	Técnicas de aumento de dados	25

4	Aprendizado ativo para classificação de fácies sísmicas	27
4.1	Aprendizado ativo	27
4.2	Estratégias de amostragem	28
4.2.1	Amostragem por incerteza	28
4.2.1.1	Amostragem baseada na menor confiança	29
4.2.1.2	Amostragem baseada na entropia	29
4.2.2	Amostragem por representatividade	30
4.2.2.1	Amostragem aleatória estratificada	30
4.2.2.2	Amostragem estratificada em passos	30
4.2.2.3	Amostragem por representatividade baseada em agrupamentos	31
4.2.2.4	Amostragem por representatividade baseada na distância	31
4.3	Aplicação do aprendizado ativo à interpretação sísmica	32
5	Metodologia	33
5.1	Dados utilizados	33
5.2	Arquitetura adotada	35
5.3	Treinamento da rede	37
5.3.1	Preparação do conjunto de dados de treinamento	38
5.3.2	Pré-processamento	38
5.3.3	Batch Normalization	40
5.3.4	Função de perda	41
5.3.5	Adam optimizer	41
5.3.6	Escolha de hiperparâmetros	42
5.4	Teste	42
5.4.1	Métricas de desempenho	43
5.5	Algoritmo de aprendizado ativo	44
5.6	Estimativa do intervalo de confiança	45
6	Resultados	47
6.1	Análise da convergência do processo de treinamento	47
6.2	Resultados no conjunto de teste	48
6.3	Aprendizado ativo	53
7	Conclusão	61
A	Formulação do F1-escore	69
B	Algoritmos de aprendizado ativo	72

Lista de Figuras

2.1	Fluxo de uma prospecção sísmica, da aquisição a modelagem. Adaptado de Zhang et al. (2016).	6
2.2	Formas mais comuns de visualização do dado sísmico 3D, extraídas do volume sísmico público F3.	6
2.3	Slice: corte horizontal no cubo sísmico, representa uma “foto” de um tempo(ou profundidade) fixo. Sísmica extraída do dado publico F3.	7
2.4	Análogos vistos na superfície de feições de interesse observadas nos dados sísmicos. Os dados sísmicos conseguem registrar geometrias de feições e ambientes sedimentares em subsuperfície semelhantes aos que podem ser vistos na superfície.	8
2.5	Evolução do reconhecimento de objetos, ou entendimento de cena, da inferência grosseira até a refinada: classificação, detecção ou localização, segmentação semântica, e segmentação de instância. Adaptado de Garcia-Garcia et al. (2017).	10
2.6	Exemplo mostrando as seções sísmicas originais (linha superior), anotações (linha central) e a predição com aprendizado profundo (linha inferior) nas inlines de treinamento, da coluna da esquerda para a direita: #190, #290, #490. Os dados são do conjunto de dados F3. Modificado de Di et al. (2019).	13
3.1	Um único neurônio atuando como unidade de processamento.	16
3.2	Esquerda: conexão esparsa em uma CNN, em que nem todos os neurônios da camada de entrada estão conectados com os da camada de saída. Direita: compartilhamento de pesos em uma camada convolucional, em que os mesmos conjuntos de pesos w_1, w_2, w_3 e w_4 são utilizados em saídas diferentes.	18
3.3	Ilustração do processo de convolução. A entrada possui dimensão $32 \times 32 \times 3$ e é convolvida com seis filtros $5 \times 5 \times 3$, com padding e stride 1, resultando em 6 mapas de ativação de tamanho 28×28	20
3.4	Gráfico da função de ativação ReLU.	21

3.5	Exemplo de <i>pooling</i> . A camada de entrada é um mapa de ativação 8×8 , o filtro <i>pooling</i> tem tamanho 2×2 . O <i>stride</i> é 2, resultando em mapa de ativação de saída com o tamanho 4×4	22
3.6	Representação da convolução transposta com um kernel 2×2 . Uma entrada de tamanho 2×2 recebe a aplicação de um filtro de convolução transposta de tamanho 2×2 , resultando em uma saída com resolução 3×3	22
3.7	Arquitetura típica de uma CNN.	24
3.8	Imagem cujo tamanho não é adequado como entrada da CNN.	26
3.9	Recortes da imagem original no tamanho adequado para servir de entrada para a CNN. Notar trechos de superposição entre as imagens.	26
5.1	Mapa geológico da bacia Taranaki, incluindo o 3D Parihaka - marcado pelo polígono em preto. (Fonte: Kumar (2016))	34
5.2	Geometria do dado sísmico Parihaka, do desafio do workshop de interpretação assistida por ML da SEG, em 2020. A área azul indica os dados de treinamento. A área vermelha indica os dados de teste, não usados em nenhuma etapa do treinamento. As áreas verde e amarela indicam as posições dos dois conjuntos de dados de teste cego fornecidos pela SEG. (Fonte: Chai et al. (2022))	35
5.3	Visualização de um exemplo dos dados de treinamento: seção sísmica na crossline de número 780 (esquerda) e seção anotada correspondente.	36
5.4	Ilustração da arquitetura da rede adotada, proposta por Chai et al. (2022).	37
5.5	Crossline 1, retirada dos experimentos devido a erro de anotação de fácies.	39
5.6	Alguns pares entrada/saída da CNN. Na primeira linha as seções sísmicas de entrada e abaixo as fácies classificadas. O tamanho de cada imagem é 992×576	40
5.7	Fluxo proposto de aprendizado ativo para três diferentes tipos de estratégia de amostragem: amostragem por incerteza, amostragem aleatória e amostragem estratificada. A seta vermelha indica a única etapa em que é necessária a intervenção do intérprete.	44
6.1	Perda e acurácia do modelo ao longo de 50 épocas de treinamento.	47
6.2	(a) Distribuição das amostras em cada fácies sísmica no conjunto de treinamento. (b) Distribuição das amostras em cada fácies sísmica no conjunto de teste.	48
6.3	Imagem 1 (inline). (a) Seção sísmica. (b) Anotações de Fácies da SEG. (c) Fácies preditas utilizando o modelo treinado por 50 épocas.	49

6.4	Imagem 206 (crossline). (a) Seção sísmica. (b) Anotações de Fácies da SEG. (c) Fácies preditas utilizando o modelo treinado por 50 épocas. . . .	49
6.5	Imagem 117 (inline). Seção sísmica e fácies preditas mostrando que os resultados do modelo apresentam boa continuidade e adequação ao dado. .	50
6.6	Matriz de confusão calculada sobre o conjunto de teste. O modelo utilizado para a predição foi treinado por 50 épocas.	51
6.7	Imagem 101 (crossline). (a) Seção sísmica. (b) Anotações de Fácies da SEG. (c) Fácies preditas utilizando o modelo treinado por 50 épocas. . . .	52
6.8	Evolução do F1-escore e IOU em relação as imagens do conjunto de teste. As imagens $[1,14] \cup [105,118]$ foram extraídas de inlines. As demais de crosslines.	53
6.9	Imagem 205 (crossline) - melhor resultado obtido nas métricas dentro do conjunto de teste. (a) Seção sísmica. (b) Anotações de Fácies da SEG. (c) Fácies preditas utilizando o modelo treinado por 50 épocas.	54
6.10	Imagem 106 (inline) - pior resultado obtido nas métricas dentro do conjunto de teste. (a) Seção sísmica. (b) Anotações de Fácies da SEG. (c) Fácies preditas utilizando o modelo treinado por 50 épocas.	54
6.11	Imagem 206 (crossline). (a) Anotações de Fácies da SEG. (b) Fácies preditas utilizando o modelo treinado por 50 épocas. (c) Fácies preditas utilizando o modelo treinado por 15 épocas.	55
6.12	Curvas de desempenho da métrica F1-escore em relação ao percentual de dados anotados no treinamento, empregando diferentes métodos de amostragem: aleatória (azul); por incerteza (laranja); e estratificada em passos (verde). Para cada modelo é apresentado o valor médio e o intervalo de confiança de 95%, calculados a partir das amostras bootstrap.	55
6.13	(a) Variação da amplitude do intervalo de confiança de 95% com o tamanho da amostra <i>bootstrap</i> em relação ao conjunto de teste original. O ponto em destaque representa a escolha feita: 100 imagens (aproximadamente 50% do número de imagens do conjunto de teste). (b) Variação da amplitude do intervalo de confiança de 95% com o número de amostras <i>bootstrap</i> . O gráfico mostra que não ocorre uma variação significativa do tamanho do intervalo com o aumento no número de amostras.	56
6.14	Detalhe dos gráficos da Figura 6.12, ilustrando o desempenho do F1-escore de acordo com o percentual de dados anotados no treinamento para três diferentes métodos de amostragem: aleatória (azul);por incerteza (laranja); e estratificada em passos (verde).	57

6.15	Predições obtidas pelos modelos treinados após receberem cerca de 60% das imagens anotadas de $\mathcal{D}_{oráculo}$. A primeira linha apresenta as predições para a Imagem 1 (inline), enquanto a segunda linha faz o mesmo para a Imagem 206 (crossline). Na primeira coluna, são exibidas as fácies anotadas pelos especialistas da SEG (ground truth), e nas colunas subsequentes são apresentadas as predições para cada uma das estratégias de amostragem utilizadas.	59
6.16	Imagem 206 (crossline). Visualização dos resultados da predição do modelo que utiliza a estratégia de amostragem por incerteza a medida que mais imagens anotadas de $\mathcal{D}_{oráculo}$ são acrescentadas no treinamento.	60
A.1	Conjuntos relevantes na segmentação de uma classe em uma imagem. O retângulo maior, cinza, representa a imagem completa. Os pixels dentro do retângulo delimitado em amarelo correspondem à classe alvo, enquanto os pixels dentro do retângulo delimitado em roxo representam a classe predita. Os termos VP(Verdadeiro Positivo), VN(Verdadeiro Negativo), FP(Falso Positivo), e FN(Falso Negativo) são utilizados para descrever as diferentes categorias de pixels em relação à sua classificação correta ou incorreta. . .	70

Lista de Tabelas

5.1	Hiperparâmetros escolhidos	42
6.1	Resultado das métricas de avaliação após o treinamento no conjunto completo de treinamento por 50 e 15 épocas.	50

Lista de Abreviações

SFC Seismic Facies Classification

DL Deep Learning

ML Machine Learning

SEG Society of Exploration Geophysicists

SGD Stochastic Gradient Descent

FCNN Fully Connected Neural Network

CNN Convolutional Neural Network

ReLU Rectified Linear Unit

RF Receptive Field

AL Active Learning

IOU Intersection Over Union

Capítulo 1

Introdução

O entendimento das estruturas na subsuperfície da terra é um componente essencial para várias aplicações relevantes, tais como monitoramento ambiental, sequestro de carbono e exploração de óleo e gás (Di et al., 2020). No caso da indústria de petróleo e gás, os dados sísmicos são a principal ferramenta utilizada para detectar e delimitar feições de interesse exploratório, de modo que, dificilmente um poço é perfurado sem o auxílio de uma imagem sísmica pra auxiliar a locação. Resumidamente, os métodos sísmicos envolvem a estimativa da geometria e propriedades físicas das camadas por meio do registro das reflexões de ondas sonoras que são propagadas através da terra (Bjorlykke, 2015). Delimitar sequências geológicas a partir de dados sísmicos agrega muito valor no processo de exploração e produção, diminuindo incertezas e reduzindo custos.

As ciências da terra utilizam grandes volumes de dados geofísicos para obter um entendimento suficiente da estrutura interna da terra, sobretudo dados sísmicos de reflexão. O desenvolvimento das tecnologias de aquisição sísmica, em especial o advento da sísmica 3D, permitiu o imageamento de estruturas geológicas a milhares de metros em profundidade com resolução de dezenas de metros (Cartwright and Huuse, 2005). Em contrapartida, os conjuntos de dados se tornaram cada vez mais extensos, tornando mais árdua a tarefa de prever a localização das jazidas de óleo e gás através da análise dos dados sísmicos. Bond et al. (2012) constataram a crescente necessidade de tempo e experiência de especialistas para lidar com a tecnologia sísmica 3D. Yu and Ma (2021) destacaram que, além da questão do tempo prolongado, um desafio da interpretação sísmica tradicional é a ambiguidade resultante da interação dos especialistas com os grandes volumes de dados, ou seja, os resultados não são únicos.

Visão computacional é uma área da ciência da computação que tem por objetivo tomar imagens brutas e, após um processamento adequado, derivar informações relevantes delas automaticamente, para que decisões possam ser tomadas ou dados importantes registrados para utilização futura. A visão computacional tem aplicações em diversas

áreas da medicina, indústria e militar, entre outras. A área tem se desenvolvido rapidamente nos últimos anos (Feng et al., 2019), em grande parte devido a aplicação de técnicas de aprendizado de máquina (do inglês *Machine Learning (ML)*). Esse desenvolvimento tem permitido a realização de tarefas cada vez mais complexas, como a classificação de imagens, localização de objetos em imagens e, em especial, a separação da imagem em múltiplas partes com anotações categóricas, tarefa conhecida como segmentação semântica. Interpretar manualmente dados sísmicos, separando-os em diferentes domínios estratigráficos e estruturais, se aproxima da ideia de segmentar uma imagem.

Aprendizado Profundo (do inglês *Deep Learning (DL)*) é um novo ramo do ML baseado em redes neurais artificiais. Por meio de redes neurais com múltiplas camadas com centenas de neurônios, as redes neurais profundas são capazes de simular funções complexas capazes de executar tarefas de visão computacional, processamento de linguagem natural e de muitos outros campos. Recentemente, o avanço da capacidade de processamento e o surgimento de novos algoritmos tem possibilitado a aplicação de segmentação semântica em dados sísmicos para caracterizar falhas [Jiang (2017), Wu et al. (2019b), Milosavljevic (2019)], domos de sal (Islam, 2020), domínios sismoestratigráficos (Dramsch and Luthje, 2018), etc. Os resultados publicados até o momento demonstram que DL é incapaz de substituir os intérpretes humanos, mas encorajam a investigação dessas técnicas no suporte ao trabalho dos especialistas.

1.1 Motivação

Os dados sísmicos 3D são utilizados comumente em inúmeras áreas das geociências, a exemplo da geologia estrutural e estratigráfica. A aquisição, processamento e interpretação desses dados é a principal ferramenta para a descoberta de novas jazidas de petróleo, além de desempenhar um papel fundamental no gerenciamento dos campos descobertos, ao longo de todos os anos do ciclo de exploração das reservas.

Os métodos tradicionais usados para extrair conhecimento geológico a partir dos dados sísmicos envolvem a experiência e técnica de intérpretes humanos, levando a resultados subjetivos e não únicos (Zhang et al., 2021). Além disso, com o desenvolvimento das tecnologias de aquisição, esses dados tem se tornado cada vez maiores, tornando quase impossível para o humano checar cada parte do dado em um tempo razoável.

O desenvolvimento da área de DL tem proporcionado o surgimento de aplicações nas geociências, em especial a utilização de segmentação semântica nas tarefas de interpretação sísmica. É um consenso, tanto no meio acadêmico quanto na indústria, que a maior dificuldade para levar essas aplicações adiante é ter um conjunto de dados com uma quantidade suficiente de anotações de qualidade.

1.2 Objetivo

O presente trabalho tem como objetivo geral investigar a aplicação de técnicas de aprendizado ativo para reduzir a dificuldade na obtenção de conjuntos de dados com anotação de fácies suficiente para o tratamento do problema de classificação fácies sísmicas (do inglês *Seismic Facies Classification (SFC)*) a partir de volumes sísmico 3D assistidos por algoritmos de DL. Como objetivos específicos tem-se:

- Estudar, descrever e implementar algoritmos de DL suficientemente robustos para realizar a tarefa proposta.
- Descrever e implementar todo o fluxo de carregamento dos dados, pré-processamento, construção do dataset de treino e teste, treinamento da rede e verificação do desempenho dos modelos no volume sísmico 3D público Parihaka.
- Estudar, descrever e avaliar técnicas de aprendizado ativo no contexto de classificação de fácies sísmicas, simulando situações em que não existe uma base de dados anotada suficientemente grande no início do estudo.

1.3 Organização do trabalho

Este trabalho está estruturado em sete capítulos, incluindo este capítulo introdutório, no qual o problema de SFC é apresentado juntamente com a oportunidade da aplicação de técnicas de DL e Aprendizado Ativo (do inglês *Active Learning (AL)*) nessa tarefa. No Capítulo 2, são apresentados conceitos básicos da área de geofísica de exploração, fornecendo um contexto fundamental para leitores não familiarizados com essa área do conhecimento; além disso, é apresentado um breve resumo do estado da arte das aplicações de DL na área de geofísica de exploração. O Capítulo 3 aborda uma fundamentação teórica sucinta, compreendendo conceitos essenciais de DL que são utilizados na arquitetura U-Net escolhida para os experimentos. No Capítulo 4, são apresentados os conceitos de aprendizado ativo, que são essenciais para o entendimento do trabalho: são discutidos os fundamentos teóricos e as principais estratégias de AL com possibilidade de aplicação em SFC. Seguidamente, o Capítulo 5 detalha o conjunto de dados utilizado, a arquitetura da rede neural utilizada, a composição dos conjunto de treino e teste, o pré-processamento dos dados e as métricas de desempenho adotadas para avaliação dos resultados. No Capítulo 6 são descritos os resultados dos experimentos realizados, apresentando análises qualitativas e quantitativas correspondentes; também são apresentados os resultados dos testes de AL realizados, que simularam a interação entre especialistas e algoritmos ao longo de meses de trabalho, ao acrescentar progressivamente no conjunto

de treinamento as anotações existentes. Por fim, no capítulo 7 são apresentadas as considerações finais, destacando as principais contribuições do trabalho, suas limitações e possibilidades de trabalhos futuros.

Capítulo 2

Deep learning aplicado a prospecção sísmica de petróleo

2.1 Fundamentos da prospecção sísmica de petróleo

A exploração de petróleo consiste no processo de buscar e localizar reservas econômicas de óleo e gás. A prospecção sísmica é fundamental nesse processo, proporcionando altas taxas de sucesso na determinação da localização dos reservatórios. O objetivo deste método é criar uma imagem tridimensional do subsolo para identificar áreas propícias para a existência de depósitos de petróleo e de gás. Isso é feito enviando ondas sísmicas através da terra e registrando a reflexão dessas ondas a partir das camadas subterrâneas. As ondas sísmicas são geradas por meio de explosões controladas ou vibrações mecânicas no solo. São necessários investimentos altos em materiais, tecnologia e recursos humanos para adquirir esses dados, processá-los e extrair às informações importantes. Este fluxo de trabalho é ilustrado na Figura 2.1.

Os mecanismos básicos do método sísmico de reflexão, segundo Jiang (2017), são:

- Uma fonte artificial de energia, que gera ondas mecânicas que se propagam em subsuperfície. Nas interfaces das camadas, parte da onda é refletida viajando de volta para a superfície, e parte é refratada, sendo transmitida até encontrar uma nova interface.
- As reflexões das ondas são registradas em sensores instalados na superfície da terra, ou embarcados em navios.
- Esses dados são então processados por computadores para criar uma imagem tridimensional do subsolo. Esse conjunto de imagens processadas é chamado de cubo sísmico 3D. Os dados 3D normalmente são visualizados por meio de *time/depth slices* (corte horizontal), *inline slice* (corte vertical que é paralelo a linha de receptores)

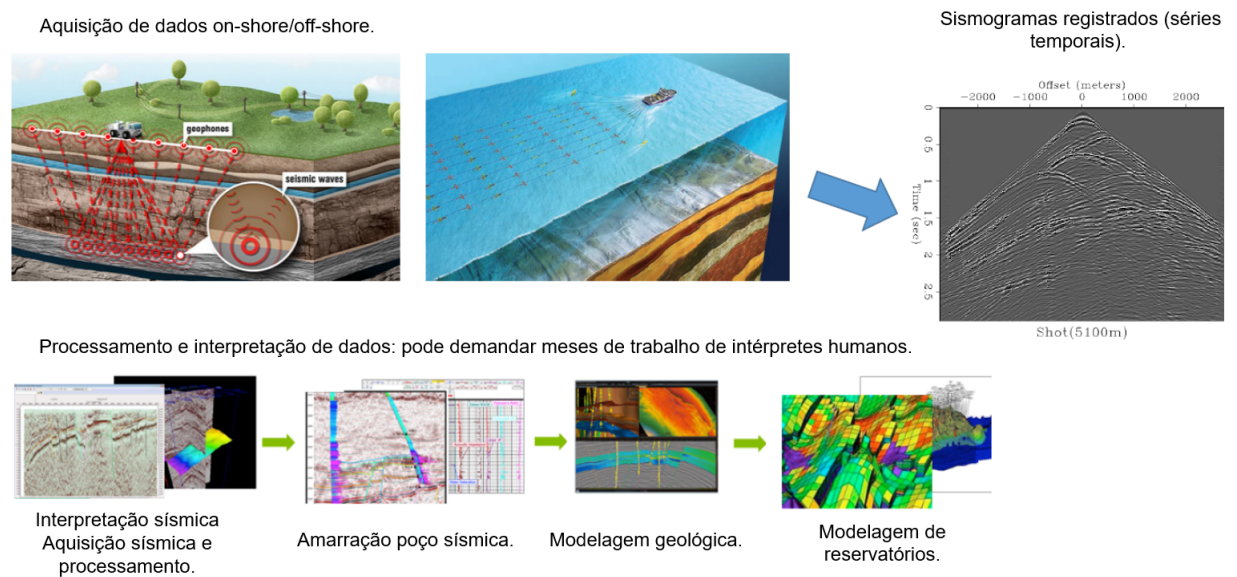
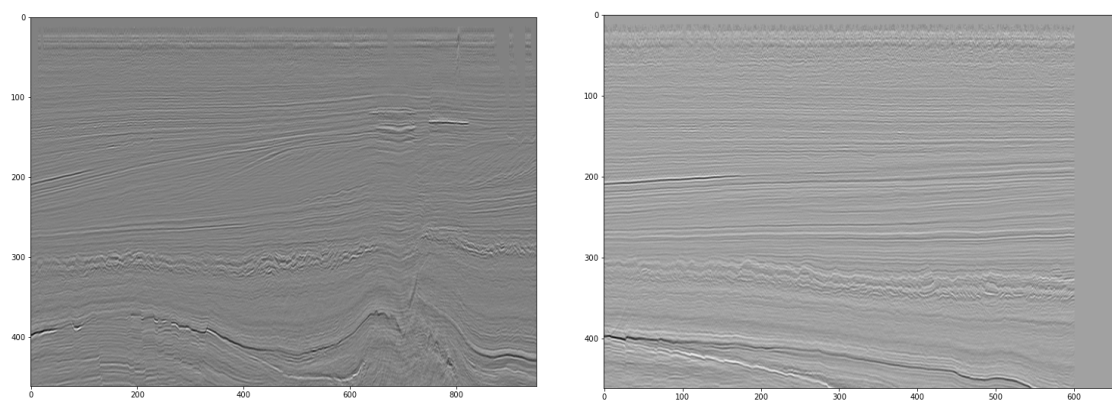


Figura 2.1: Fluxo de uma prospecção sísmica, da aquisição a modelagem. Adaptado de Zhang et al. (2016).

e *crossline slice* (corte vertical perpendicular a *inline*). As Figuras 2.2 e 2.3 apresentam exemplos dessas visualizações, retirados do dado sísmico público F3¹. Os geofísicos trabalham nessas imagens para identificar e delimitar feições geológicas de interesse.



(a) Inline: corte vertical no cubo sísmico paralelo a linha de receptores. (b) Crossline: corte vertical no cubo sísmico perpendicular a inline.

Figura 2.2: Formas mais comuns de visualização do dado sísmico 3D, extraídas do volume sísmico público F3.

¹Pequeno volume de dados sísmicos 3D marinhos disponibilizados para a comunidade científica realizar trabalhos de Machine Learning

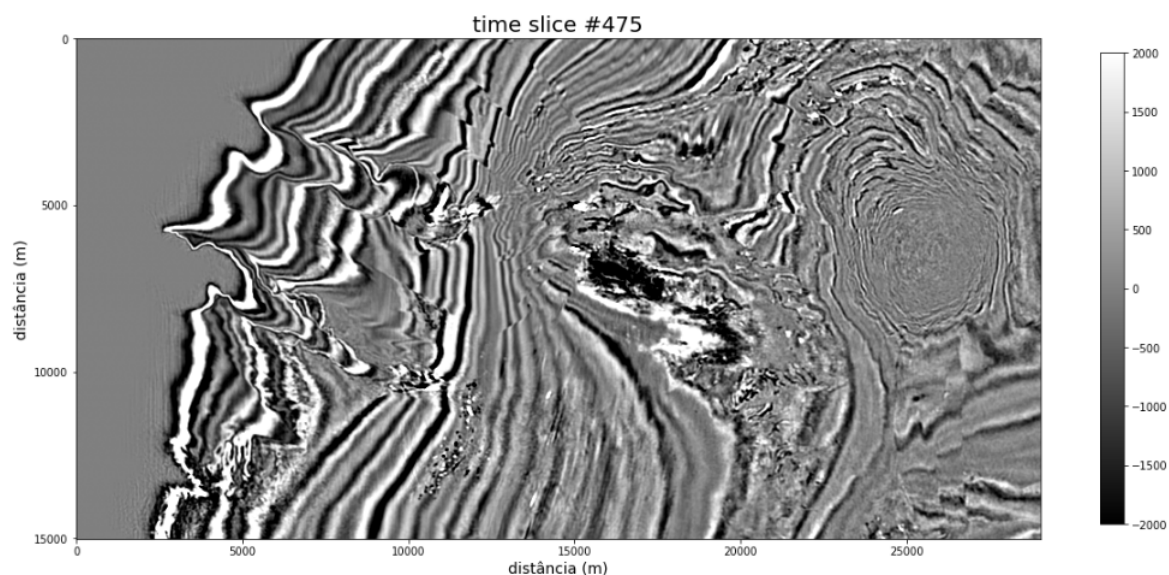


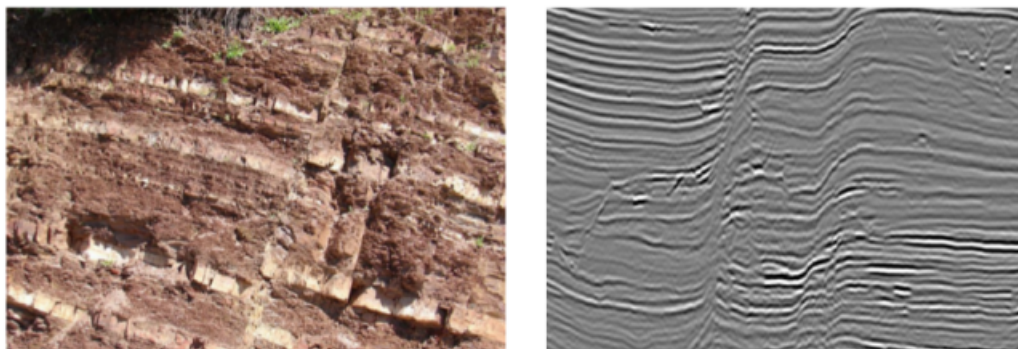
Figura 2.3: Slice: corte horizontal no cubo sísmico, representa uma “foto” de um tempo(ou profundidade) fixo. Sísmica extraída do dado publico F3.

Os dados sísmicos 3D possuem tamanhos variados, comumente na casa das centenas de gigabytes. Normalmente, são gravados em um formato bastante específico chamado SEG-Y. Este formato foi desenvolvido pela *Society of Exploration Geophysicists (SEG)* para facilitar a troca de dados geofísicos (SEG wiki, 2017). Para que os algoritmos sejam utilizados, são necessários pacotes capazes de ler e visualizar esse formato de dados. Muita atenção deve ser dispensada aos sistemas de coordenadas e geometria do levantamento para que as conclusões obtidas tenham alguma validade.

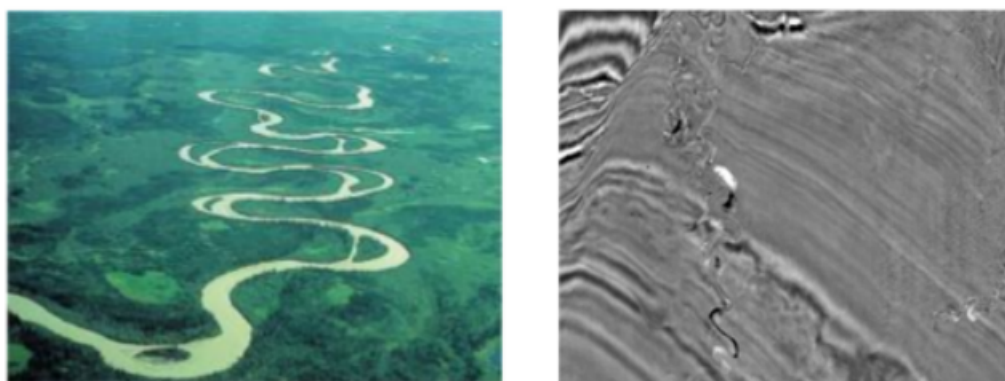
2.2 Interpretação sísmica

Para que os depósitos de óleo e gás possam ser identificados, os dados sísmicos precisam ser interpretados por geocientistas. O objetivo da interpretação sísmica estrutural é elaborar mapas das formações rochosas de interesse em subsuperfície. Entre as feições geológicas que ajudam na identificação e delimitação dos reservatórios encontram-se as falhas, os canais e os domos de sal. A Figura 2.4 mostra algumas dessas feições em ambientes sedimentares e seus respectivos análogos no dado sísmico.

Facies sísmica é comumente definida como a unidade sedimentar cujas propriedades petrofísicas características podem ser discriminadas, na escala sísmica, das unidades que a circundam (Liu et al., 2020). A interpretação sismoestratigráfica tem por objetivo delimitar estas unidades, tradicionalmente rastreando um conjunto de refletores sísmicos que separam intervalos com características particulares de geometria, continuidade lateral, amplitude, frequência e velocidade intervalar (Di et al., 2020). O objetivo da classificação



(a) Esquerda: falhas em um afloramento no Marrocos. Direita: inline-slice mostrando falhas em subsuperfície. Fonte: Jiang (2017).



(b) Esquerda: Foto aérea de um canal. Direita: time slice mostrando uma estrutura de canal em subsuperfície. Fonte: Jiang (2017).

Figura 2.4: Análogos vistos na superfície de feições de interesse observadas nos dados sísmicos. Os dados sísmicos conseguem registrar geometrias de feições e ambientes sedimentares em subsuperfície semelhantes aos que podem ser vistos na superfície.

de fácies sísmicas (SFC) é interpretar os dados sísmicos dentro de várias categorias pre-definidas que, em tese, estão conectadas com heterogeneidades geológicas (Zhang et al., 2021). Este passo é fundamental para a compreensão da história de deposição, definição do sistema petrolífero, e estimação das propriedades do reservatório.

2.3 Segmentação de imagens aplicada em imagens sísmicas

Para extrair as fácies sísmicas de um dado sísmico 3D, é necessário particionar todo o volume em múltiplas partes com seus rótulos categóricos. Especificamente, espera-se designar um rótulo, i.e. uma fácies sísmica, para cada pixel ou voxel do dado sísmico. Dentro do campo da visão computacional este procedimento é chamado de segmentação

semântica.

Os algoritmos de segmentação particionam imagens baseando-se em características como intensidade, cor ou textura, com o objetivo de definir sub-regiões (Jiang, 2017). Entre os métodos básicos de segmentação estão:

- *Baseados em limiares*: que calculam um ou mais limites de acordo com uma regra ou função, designando os pixels para diferentes classes de acordo com esses limites.
- *Baseados em bordas*: As bordas de uma região geralmente apresentam mudanças bruscas de intensidade, o que indica que os gradientes são cruciais na detecção das bordas ou cantos de uma imagem. Portanto, isso forma a base dessa abordagem de segmentação.
- *Baseados em região*: são feitas partições na imagem baseadas em critérios de similaridade, muitas vezes a partir de “sementes” - pixels selecionados na imagem - que são associados a pixels vizinhos de acordo com algum critério de similaridade.
- *Baseados em agrupamentos*: a imagem é dividida em vários grupos baseados na similaridade da cor do pixel, intensidade, localização entre outras propriedades. Um algoritmo popular nessa categoria é o K-means.

Estes algoritmos de segmentação particionam a imagem em regiões separadas, porém não determinam a semântica de cada região. A segmentação semântica divide as imagens em regiões que são rotuladas como uma classe semântica. Para Garcia-Garcia et al. (2017), a segmentação semântica não deve ser entendida como um campo isolado, mas como um passo natural na área de reconhecimento de objetos ou entendimento de cena. A tarefa mais simples é a classificação, que consiste em fazer a predição de uma entrada, predizendo qual é o objeto em uma imagem ou mesmo fornecendo uma lista ordenada no caso de uma imagem com vários objetos. Localização ou detecção é o próximo passo em direção a uma inferência refinada, provendo não apenas as classes, mas também informação adicional a respeito da localização espacial destas classes, e.g., centróides ou *bounding boxes* (retângulos anotados ao redor de objetos da imagem). Daí se entende que a segmentação semântica é a sequência natural para uma inferência refinada, pois tem como objetivo inferir os rótulos de maneira densa, classificando cada pixel. Mais avanços podem ser obtidos, a exemplo da segmentação de instâncias, que tem rótulos separados para diferentes instâncias da mesma classe. A Figura 2.5 exemplifica essa evolução.

O problema da classificação pixel a pixel pode ser formulado da seguinte maneira: utilizando algum critério é estabelecido um espaço de classes $\mathcal{L} = \{l_1, l_2, \dots, l_k\}$, onde cada elemento de um conjunto de variáveis aleatórias $\mathcal{X} = \{x_1, x_2, \dots, x_N\}$ será classificado com um rótulo l que representa uma classe ou objeto, com k estados possíveis. Normalmente,

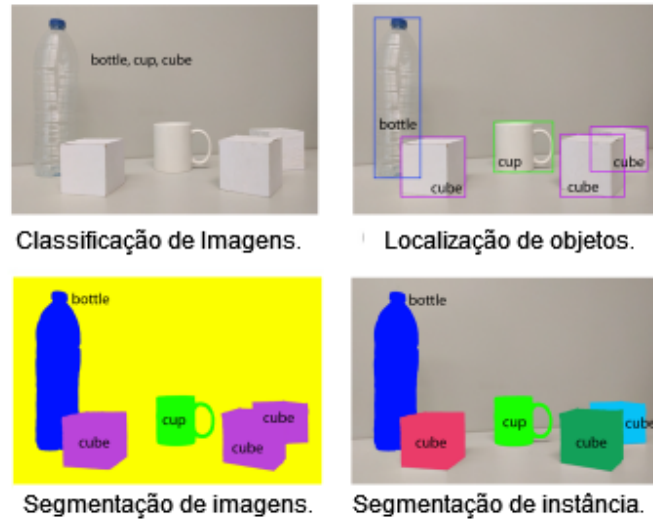


Figura 2.5: Evolução do reconhecimento de objetos, ou entendimento de cena, da inferência grosseira até a refinada: classificação, detecção ou localização, segmentação semântica, e segmentação de instância. Adaptado de Garcia-Garcia et al. (2017).

\mathcal{X} é uma imagem 2D de $W \times H = N$ pixels. No entanto, o conjunto de variáveis aleatórias pode ser estendido para outras dimensões.

2.4 Deep learning para interpretação sísmica

DL é uma subcategoria do campo da inteligência artificial que tem ganhado destaque nos últimos anos devido à sua capacidade de análise de grandes conjuntos de dados e descoberta de padrões e relações complexas. Essa abordagem é fundamentada em redes neurais artificiais profundas, compostas por várias camadas de neurônios interconectados, permitindo que o modelo capture características cada vez mais abstratas e de alto nível à medida que as informações fluem pelas camadas. A principal adição é o uso de camadas de convolução, que extraem características relevantes de forma local e compartilhada.

Uma das grandes vantagens do DL é a capacidade de lidar com dados complexos e de grande escala, como imagens, áudio, texto e sequências temporais. Essa abordagem tem se destacado em diversas áreas, incluindo reconhecimento de padrões (Phinyomark and Scheme, 2018), processamento de linguagem natural (Wu et al., 2019a), visão computacional (Voulodimos et al., 2018), entre outras.

Nos últimos anos, vê-se uma grande gama de uso das técnicas de *deep learning*, em especial as redes neurais convolucionais (do inglês *Convolutional Neural Network (CNN)*) em diversas aplicações na área de visão computacional: segmentação de imagens (Chen et al., 2017), classificação de imagens (Rashid et al., 2019), interpolação/extrapolação de imagens (Yan et al., 2020), detecção de objetos (Li et al., 2020) e reconhecimento de

padrões (Guo et al., 2016), entre outras.

No contexto da interpretação de dados sísmicos, o Deep Learning oferece uma poderosa ferramenta para auxiliar na identificação e interpretação de padrões complexos presentes nos registros sísmicos. Esses padrões podem ser relacionados a variações nas formações geológicas, como camadas rochosas e estruturas subterrâneas, além de auxiliar na detecção de possíveis reservatórios de petróleo e gás. É importante ressaltar que os dados sísmicos costumam ser enormes e complexos, o que dificulta a extração de informações significativas por parte dos geocientistas sem o uso de técnicas computacionais avançadas, como o *deep learning* (Walker, 2017). Essa abordagem permite uma análise mais eficiente e precisa dos dados, possibilitando a identificação de informações sutis, aprimorando a capacidade de interpretação dos geocientistas.

Sendo a geofísica uma área essencialmente de geração, processamento e interpretação de imagens, era natural que logo surgissem publicações explorando essas aplicações. As principais conferências da área de geofísica já apresentam um aumento significativo de artigos com aplicações de *deep learning*, a exemplo da SEG, que já separa uma seção específica para o tema da inteligência artificial e *big data*.

Em particular, o tema de classificação de fácies sísmicas com redes neurais profundas vem sendo explorado nos últimos anos. Xu et al. (2017) implementaram um autoencoder esparso para detecção de chaminés de gás, em uma estrutura semisupervisionada de classificação; Waldeland et al. (2018) utilizaram CNNs para interpretar automaticamente corpos de sal; enquanto Zhao (2018) aplicou uma CNN *encoder-decoder* para classificação de fácies sísmicas em uma linha 2D (8 classes); e Peters et al. (2019) rastrearam horizontes sísmicos a partir de poucas imagens de treinamento utilizando uma rede neural multi-resolução.

Embora uma variedade de trabalhos importantes tenha sido proposta na área de segmentação semântica, os dados sísmicos apresentam desafios que não podem ser imediatamente resolvidos por alguns dos métodos existentes. Em primeiro lugar, ao contrário de imagens em que os limites entre objetos são bem definidos, na sísmica as estruturas em subsuperfície não são identificadas facilmente de maneira unívoca, devido a diversas questões como ruídos, conteúdo de frequência, ambiguidades nas propriedades físicas das rochas etc.; em segundo lugar, os dados sísmicos são em escala de cinza, daí as cores não podem ser usadas para distinguir estruturas (i.e., registros sísmicos tem banda de frequência limitada). Além disso, faltam dados anotados para treinamento e benchmarks bem estabelecidos para testar as várias abordagens de aprendizado. A dificuldade de conseguir um volume suficiente de dados anotados confiáveis levou alguns autores a optar por dados sintéticos, sendo que a geração de dados sísmicos com características semelhantes aos dados reais iniciou, por si só, uma nova frente de pesquisa. Wu et al. (2019b)

estabeleceram um fluxo de trabalho para gerar dados sísmicos sintéticos com anotações de falhas e horizontes para treinamento supervisionado de CNNs.

Outro aspecto importante a se destacar nas aplicações de *deep learning* para segmentar imagens sísmicas é o fato de que muitos trabalhos apresentam apenas resultados qualitativos, simplesmente apresentando algumas imagens das predições, acompanhadas de comentários sobre os resultados obtidos. Mesmo entre os autores que buscam apontar resultados quantitativos, existe uma variação nas métricas utilizadas: Islam (2020) utilizou a métrica *Intersection Over Union*(IOU) para mensurar a efetividade de uma U-Net na classificação binária de domos de sal nos dados públicos do concurso da TGS²; já Dramsch and Luthje (2018) utilizaram a Acurácia como métrica para medir a eficiência na classificação de fácies sísmicas (SFC) no dado F3, para três diferentes arquiteturas de rede; enquanto Zhang et al. (2021) utilizaram a métrica F1-score para avaliar a eficiência de uma CNN, com o apoio de uma transformada *wavelet* contínua, na previsão de litologia de poços a partir do dado sísmico.

A Figura 2.6 mostra um exemplo típico de apresentação qualitativa dos resultados de SFC nos trabalhos revisados.

²TGS é uma companhia de aquisição e processamento de dados geofísicos

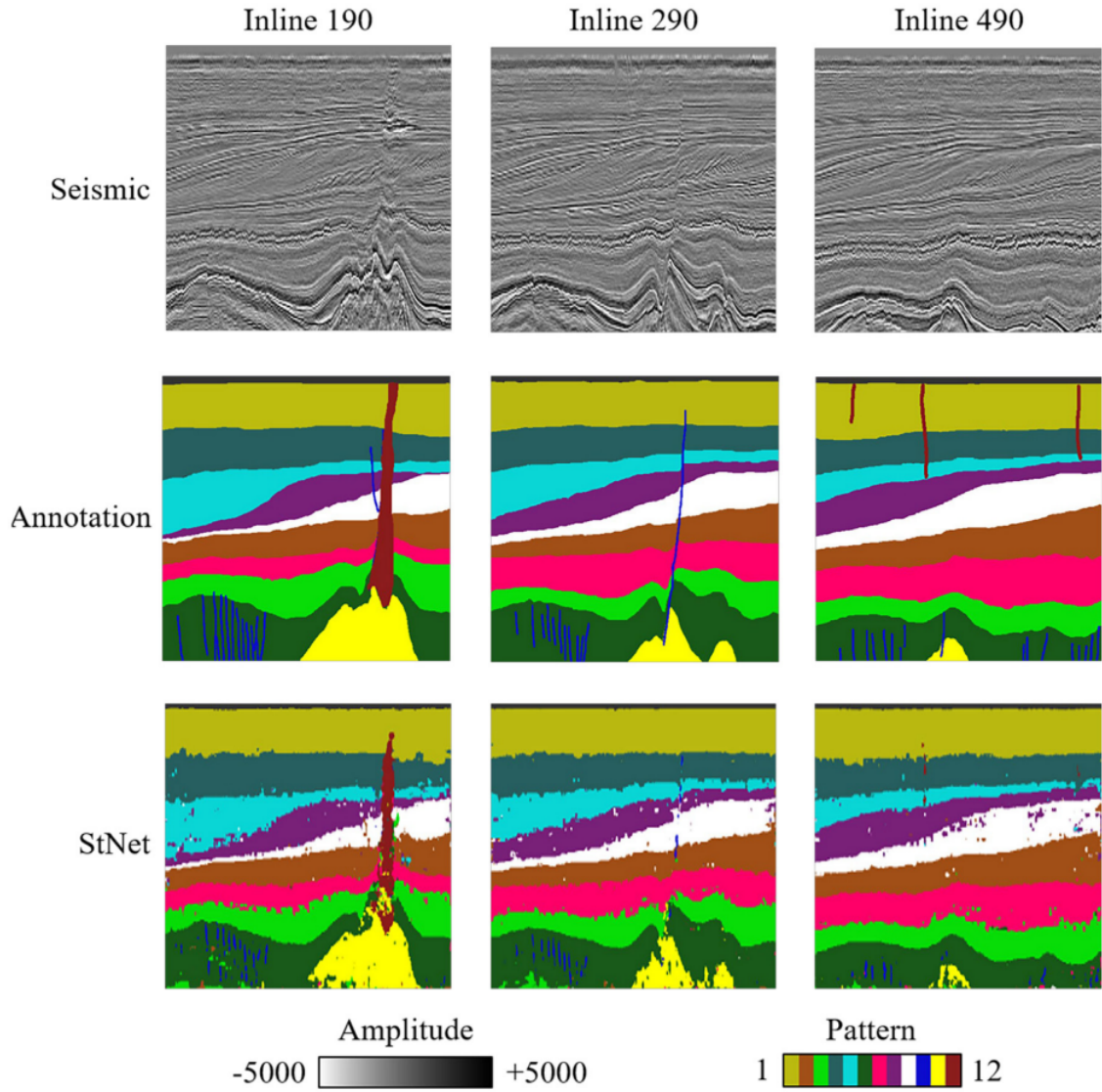


Figura 2.6: Exemplo mostrando as seções sísmicas originais (linha superior), anotações (linha central) e a predição com aprendizado profundo (linha inferior) nas inlines de treinamento, da coluna da esquerda para a direita: #190, #290, #490. Os dados são do conjunto de dados F3. Modificado de Di et al. (2019).

Capítulo 3

Redes neurais convolucionais

Como dito anteriormente, *Deep learning* tem se tornado uma ferramenta cada vez mais utilizada na solução de problemas de reconhecimento de padrões em imagens, aumentando a capacidade preditiva dos algoritmos na detecção, classificação e segmentação de objetos de interesse. As redes neurais profundas, compostas por múltiplas camadas com milhões de neurônios, são capazes de modelar funções complexas, resolvendo assim tarefas desafiadoras que são objeto de estudo na área de visão computacional. Com o aumento da capacidade computacional, essas redes têm superado outras técnicas consolidadas, quando aplicadas nos dados *benchmark* da área (Jordan and Mitchell, 2015).

Neste capítulo, alguns conceitos básicos das redes neurais artificiais serão apresentados. Em seguida, será detalhada a arquitetura das redes neurais convolucionais, que são a base para as aplicações de *deep learning* na área de imagem. Essas noções básicas são fundamentais para o entendimento da metodologia proposta e melhor apreciação dos resultados obtidos na pesquisa.

3.1 Aprendizagem

Aprendizagem é um dos pilares do aprendizado de máquina, que é um campo da inteligência artificial que se concentra na criação de modelos computacionais capazes de aprender a partir de dados, se valendo de métodos estatísticos, bem como de elementos de computação, otimização e teoria da informação (Tibshirani and Friedman, 2009). O objetivo principal é desenvolver algoritmos que sejam capazes de aprender a solucionar problemas e realizar tarefas sem necessidade de programação explícita.

Há duas formas de aprendizagem: a supervisionada e a não-supervisionada. A primeira acontece quando o modelo é treinado em dados rotulados, e a segunda quando o modelo é treinado em dados não rotulados. A abordagem de aprendizagem supervisionada, que será empregada neste trabalho, é baseada no conceito de ajuste dos parâmetros

do modelo para minimizar a diferença entre as classificações previstas e os rótulos verdadeiros obtidos por um especialista. Essa diferença é quantificada por uma função de perda (*loss function*), também conhecida como função objetivo, que fornece uma medida de quão precisas são as classificações realizadas pelo modelo. O objetivo da aprendizagem supervisionada é, então, minimizar ou maximizar essa função de perda, dependendo das necessidades específicas da aplicação.

3.2 Redes neurais

Redes neurais são sistemas dinâmicos com um grande número de unidades de processamento (UPs) conectadas, chamadas de neurônios. O neurônio constitui a unidade funcional fundamental da rede, recebendo números reais como entrada, que podem vir dos próprios dados ou das saídas de outros neurônios. A saída é transmitida por meio de conexões, que possuem um peso definido. Antes de um valor ser transmitido, ele é multiplicado pelo peso correspondente. Modificar os valores dos pesos, por sucessivas aplicações de uma regra de aprendizagem, permite que a rede se aproxime de uma função que relaciona os padrões de entrada às saídas desejadas. Uma rede típica é formada por neurônios organizados em camadas, incluindo uma camada de entrada, uma ou mais camadas escondidas e uma camada de saída.

Redes neurais são aproximadores universais (Csáji et al. (2001)), ou seja, capazes de aproximar qualquer função contínua, a despeito de possuir um número finito de neurônios. No entanto, ainda não se conhece um método capaz de determinar a configuração adequada de neurônios e camadas para aproximar uma função, fazendo com que as redes neurais comumente sofram de sobreajuste. O processamento matemático que acontece em um neurônio é mostrado na Figura 3.1 e nas equações abaixo:

$$z_o = \mathbf{w}'\mathbf{x} + b \quad (3.1)$$

$$z = \sigma(z_o) \quad (3.2)$$

Na equação 3.1 os valores do vetor de entrada \mathbf{x} são multiplicados pelos vetores de pesos \mathbf{w} do neurônio e somados a um *bias* (um elemento independente que ajuda a aumentar o grau de liberdade dos ajustes dos pesos), produzindo a saída z . Já na equação 3.2 é aplicada uma função de ativação na saída, que confere uma não-linearidade ao processamento realizado pelo neurônio.

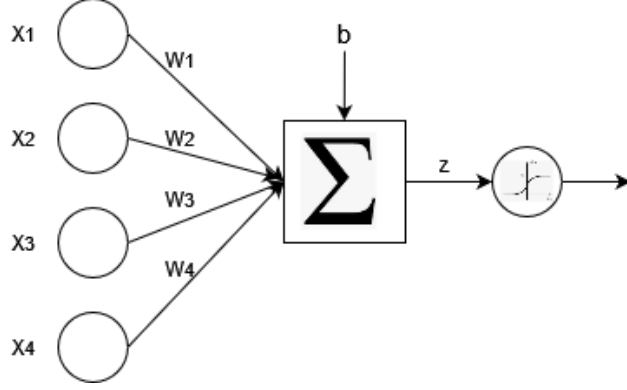


Figura 3.1: Um único neurônio atuando como unidade de processamento.

3.2.1 Gradiente descendente

Ao tratar um problema de classificação em aprendizado de máquina, uma rede neural F aproxima y a partir de x ,

$$\hat{y} = F(\mathbf{x}; \Theta) \quad (3.3)$$

onde Θ é o vetor de parâmetros da rede, \mathbf{x} representa os dados de entrada, e \hat{y} um vetor *one-hot-encoder* que representa as classes preditas. Θ é obtido através de uma aproximação de alta dimensão entre os dois conjuntos $\hat{\mathbf{Y}} = \{\hat{y}_i, i = 1 \dots\}$ e $\mathbf{Y} = \{y_i, i = 1 \dots\}$, isto é, classes previstas e anotadas. A aproximação é obtida por meio da otimização de uma função de perda/custo E (Yu and Ma, 2021):

$$E(\Theta, \mathbf{Y}, \mathbf{X}) = \sum_{i=1}^N \left\| \mathbf{y}_i - F(\mathbf{x}_i; \Theta) \right\|_2^2 \quad (3.4)$$

Quando F é diferenciável, um método gradiente pode ser usado para otimizar Θ . Os algoritmos de gradiente descendente sugerem que deve-se sempre seguir o gradiente de inclinação máxima para buscar o máximo/mínimo global. Porém, essa abordagem possui um problema prático, que é a necessidade de processar toda a base de treinamento para efetuar um único ajuste nos parâmetros, gerando elevado custo computacional. Daí, em *deep learning* frequentemente se adota o algoritmo Gradiente Descendente Estocástico (do inglês *Stochastic Gradient Descent (SGD)*) como otimizador. O SGD extrai aleatoriamente um *batch* (grupo de amostras de dados processadas simultaneamente) do *dataset* e aproxima o verdadeiro gradiente global baseado no gradiente deste *batch*.

Em um conjunto de dados com N amostras, e um *batch* de tamanho m , após calcular o gradiente médio nas m amostras do *batch* selecionado, os parâmetros são atualizados da seguinte maneira:

$$\Theta \leftarrow \Theta - \frac{\eta}{m} \sum \nabla_{\Theta} E \quad (3.5)$$

onde η é a taxa de aprendizagem, que define o quanto os parâmetros serão alterados e ∇ é o operador gradiente.

3.2.2 Retropropagação do erro

O treinamento da rede neural consiste de duas etapas principais:

1. *forward*: a saída da rede é calculada a partir dos dados de entrada.
2. *backward*: o desajuste entre a saída esperada e a calculada é usado para atualizar os parâmetros (pesos) da rede.

O algoritmo de retropropagação do erro, proposto por Rumelhart(1998) permite que milhões de pesos das conexões da rede sejam atualizados por meio dos seus respectivos gradientes. O erro que é calculado na saída ao comparar as classes previstas com as anotadas é propagado através de cada camada e os gradientes de todos os parâmetros intermediários são calculados usando a regra da cadeia. Para os propósitos da retropropagação do erro, as funções específicas de perda e ativação não importam, contanto que suas funções derivadas possam ser avaliadas eficazmente.

3.3 Redes neurais convolucionais

Redes neurais tradicionais, totalmente conectadas (do inglês *Fully Connected Neural Network (FCNN)*), possuem algumas desvantagens em aplicações de visão computacional. Uma delas é que as FCNN ignoram a estrutura 2D natural da imagem, pois admitem como entrada vetores 1D, sendo necessário o achatamento (do inglês *flattening*) da matriz 2D que representa a imagem. Além disso, o custo computacional para treinar redes totalmente conectadas é muito elevado quando se trata de imagens, uma vez que cada pixel deve ser conectado a todos os neurônios da camada seguinte.

Já as redes neurais convolucionais (CNN) lidam bem com essas limitações, pois ao mesmo tempo que dependem de menos parâmetros de ajuste do que as FCNN, conseguem extrair características relevantes da imagem por meio das operações que serão descritas nesta seção.

3.3.1 Arquitetura

Desde 2010, as CNNs tem se desenvolvido rapidamente na área de classificação e segmentação de imagens (Yu and Ma, 2021). Algumas arquiteturas populares de CNNs são a VGGNet (Simonyan and Zisserman, 2015)- que se destaca por sua profundidade e simplicidade, consistindo principalmente de camadas convolucionais e camadas de *pooling*,

a AlexNet (Krizhevsky et al., 2012) - que introduziu o uso de técnicas como a função de ativação ReLU e o treinamento em GPUs, e a ResNet (He et al., 2016) - que apresenta conexões residuais entre camadas, permitindo o treinamento de modelos com desempenho superior em desafios de reconhecimento de imagens. Uma CNN é organizada em camadas com diferentes funções. A camada convolutiva é responsável por extrair características da imagem, a camada de *pooling* por realizar uma sub-amostragem da imagem e a camada *Rectified Linear Unit (ReLU)* para introduzir não linearidade na rede.

As CNNs tem duas características principais decorrentes de sua arquitetura: campo receptivo local (do inglês *Receptive Field (RF)*) e compartilhamento de pesos (LeCun and Bengio, 1995).

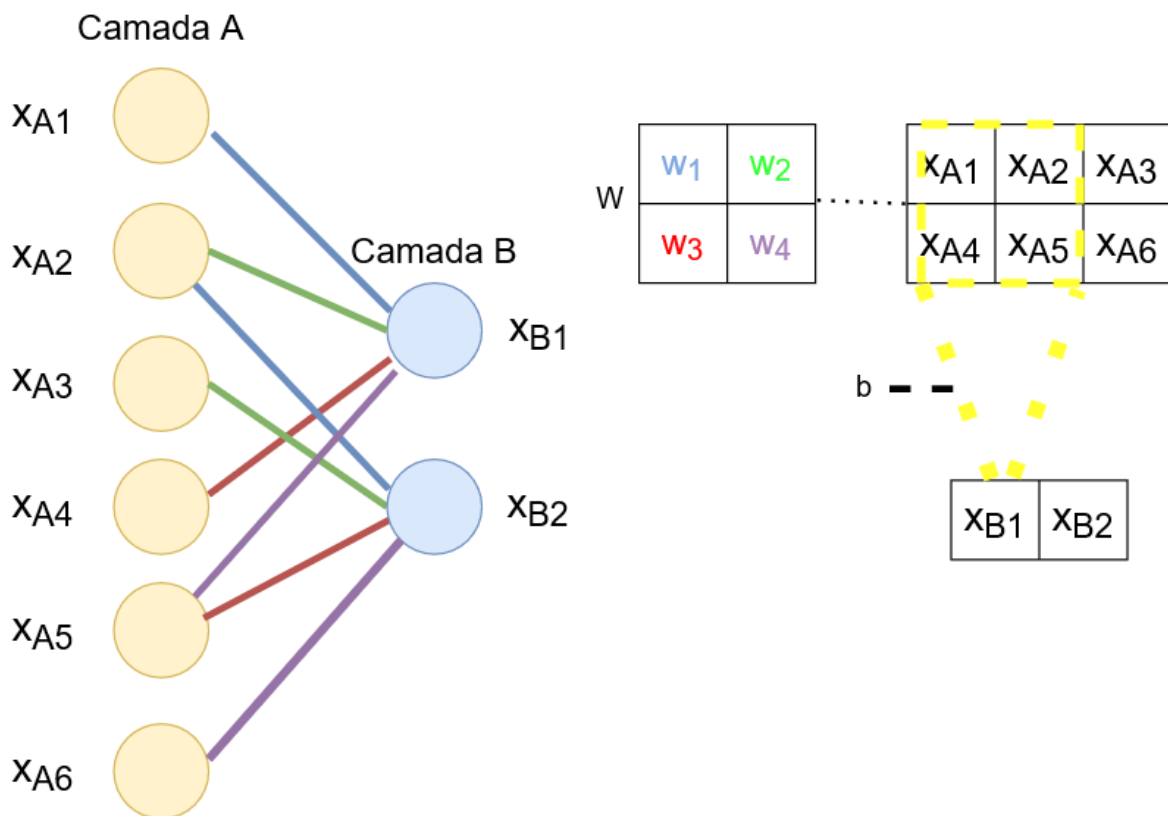


Figura 3.2: Esquerda: conexão esparsa em uma CNN, em que nem todos os neurônios da camada de entrada estão conectados com os da camada de saída. Direita: compartilhamento de pesos em uma camada convolutiva, em que os mesmos conjuntos de pesos w_1 , w_2 , w_3 e w_4 são utilizados em saídas diferentes.

O RF é definido como o tamanho da região da camada l que está conectada a um neurônio da camada $l + 1$. Na FCNN, o RF é composto por todos os neurônios da camada l , na CNN o neurônio da camada $l + 1$ é conectado de maneira esparsa a um pequeno número de neurônios da camada l . A Figura 3.2 exemplifica o conceito do campo receptivo local. Da figura também se observa que existe um compartilhamento de pesos, ilustrado

pelas conexões com a mesma cor.

3.3.1.1 Camada convolucional

Uma camada convolucional consiste em múltiplos *kernels*/filtros convolucionais de tamanho $K_1 \times K_2$ aplicados em uma imagem de tamanho $X_1 \times X_2$. O filtro convolucional é representado por uma matriz tridimensional $K_1 \times K_2 \times K_3$, onde as dimensões (altura, largura, profundidade) são aplicadas quando a convolução é realizada em imagens coloridas, uma vez que tais imagens incorporam as intensidades das três cores primárias no modelo RGB (*Red*, *Green* e *Blue*). O resultado dessa operação de convolução é conhecido como mapa de ativação ou mapa de características (do inglês *feature maps*).

Além do tamanho do *kernel*, a saída da convolução depende de outros dois hiperparâmetros: *padding* e *stride*. O *stride* denota o número de pixels pelos quais o filtro se movimenta após cada operação, enquanto que o *padding* insere simetricamente novos pixels ao redor da imagem de entrada, em geral para preservar a dimensão da imagem original, no que é conhecido como *same-padding*.

O processo de convolução está mostrado na Figura 3.3, considerando uma imagem colorida de tamanho $32 \times 32 \times 3$ e seis filtros $5 \times 5 \times 3$. A convolução é realizada sem *padding* e com *stride* 1. Cada ativação de um neurônio na posição (i, j) do mapa de ativação é o resultado do produto escalar entre o *kernel* e o RF associado, somado a um *bias*. É importante notar que o número de pesos a serem aprendidos nesta camada dependem unicamente do tamanho do *kernel*, e não da imagem. Os mapas de ativação resultantes dessa convolução tem tamanho 28×28 . Os mapas resultantes de cada um dos seis filtros são empilhados resultando em uma matriz $28 \times 28 \times 6$.

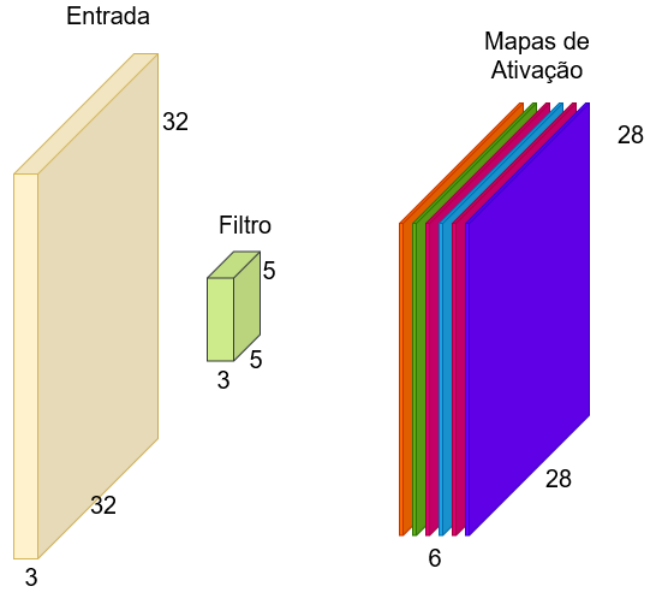


Figura 3.3: Ilustração do processo de convolução. A entrada possui dimensão $32 \times 32 \times 3$ e é convolvida com seis filtros $5 \times 5 \times 3$, com padding e stride 1, resultando em 6 mapas de ativação de tamanho 28×28 .

3.3.1.2 Camada ReLU

Em CNNs, após cada camada convolucional, o mapa de ativação resultante deve ser passado para uma camada não linear. Existem várias funções de ativação descritas na literatura que cumprem esse papel, a exemplo das funções degrau, sigmoideal e tangente hiperbólica. Dentre as diversas opções, a ReLU (Figura 3.4) tem sido largamente usada em algoritmos de *deep learning*. Na Equação 3.6 a função ReLU $f(z)$ é aplicada na saída z , retornando o valor 0 para saídas negativas e o próprio valor de z para as saídas não-negativas.

$$f(z) = \max(0, z) \quad (3.6)$$

A função de ativação ReLU apresenta diversas vantagens no treinamento de redes neurais. Primeiramente, a ReLU pode atuar como um mecanismo de regularização, evitando que a rede aprenda recursos não informativos ou redundantes, o que pode melhorar a generalização do modelo. Além disso, a ReLU evita a saturação em regiões de alta ativação, prevenindo problemas de *vanishing gradients* durante a retropropagação do erro. Também vale ressaltar que a ReLU é uma função computacionalmente eficiente e de baixo custo de implementação, tendo sido demonstrado que o tempo de treinamento pode ser reduzido significativamente em comparação com outras funções de ativação, por exemplo, a tangente hiperbólica, conforme reportado por Krizhevsky et al. (2012).

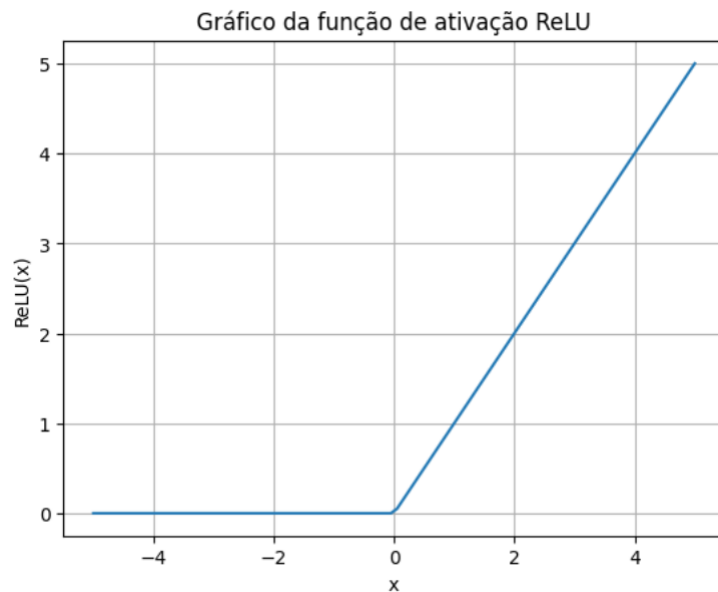


Figura 3.4: Gráfico da função de ativação ReLU.

3.3.1.3 Pooling

O objetivo da camada *pooling* é reduzir a resolução do mapa de ativação de entrada e concentrar a informação. Uma vez que certas características são detectadas, a posição exata dessa característica se torna menos importante para a decisão final, ao passo que pode prejudicar a generalização. A camada *pooling* não possui parâmetros a serem ajustados no treinamento.

Os tipos mais comuns de *pooling* são: *max pooling* e *average pooling*. No primeiro, o valor máximo é selecionado enquanto no segundo é retornado uma média de todas as entradas. A Figura 3.5 ilustra como ocorre a operação de *pooling* espacialmente com redução da dimensão da camada de entrada.

3.3.1.4 Convolução transposta

Os processos de convolução (exceto *same-padding*) provocam uma redução de escala na imagem, intencional e necessária nas CNNs. Porém, para algumas tarefas é necessário recuperar o tamanho original da imagem. Existem várias técnicas para realizar esse aumento de escala, entre elas *Nearest Neighbors*, Interpolação Bilinear e *Bed of Nails* (Long et al., 2015).

A convolução transposta é uma técnica de aumento de escala de imagens em que um mapa de ativação de entrada retorna um mapa de ativação de saída com uma dimensão desejada. Embora sejam operações diferentes, a convolução transposta e o *pooling* têm uma relação complementar em uma arquitetura de CNN. Enquanto o *pooling* pode ser usado para reduzir a resolução e extrair características invariantes de escala da imagem, a

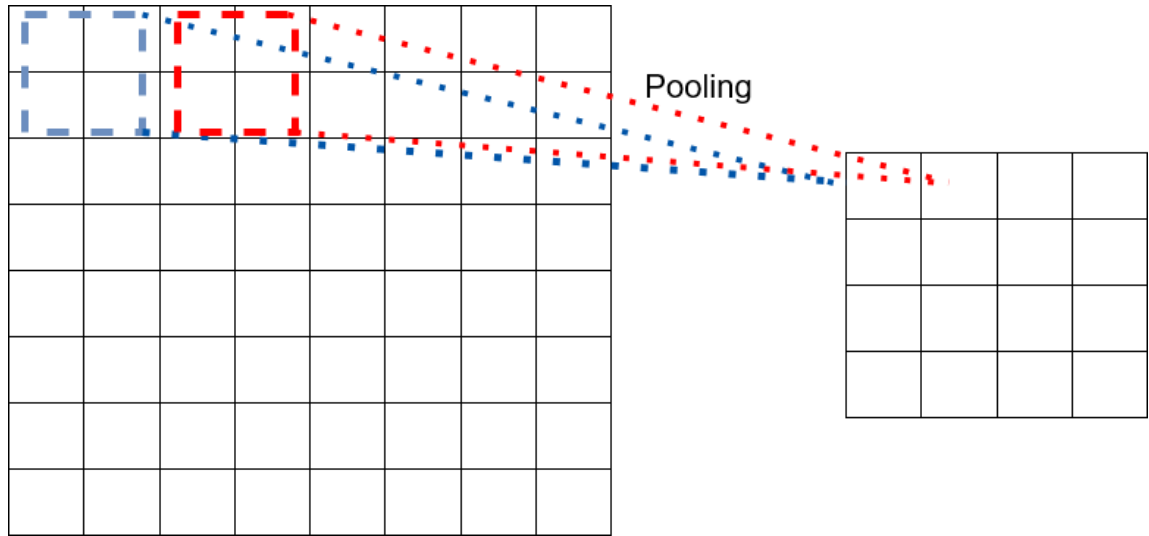


Figura 3.5: Exemplo de *pooling*. A camada de entrada é um mapa de ativação 8×8 , o filtro *pooling* tem tamanho 2×2 . O *stride* é 2, resultando em mapa de ativação de saída com o tamanho 4×4 .

convolução transposta pode ser usada para aumentar a resolução e reconstruir informações espaciais. A operação é descrita de maneira simplificada na Figura 3.6

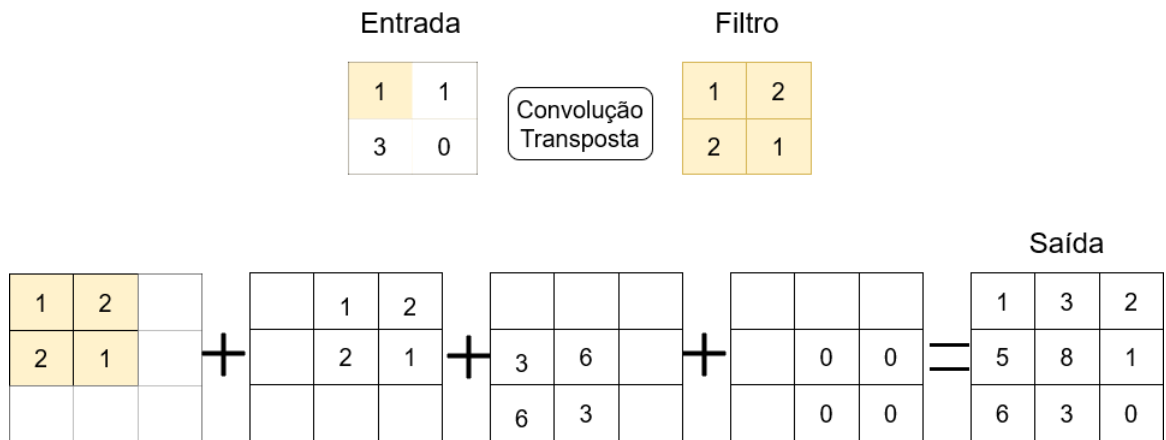


Figura 3.6: Representação da convolução transposta com um kernel 2×2 . Uma entrada de tamanho 2×2 recebe a aplicação de um filtro de convolução transposta de tamanho 2×2 , resultando em uma saída com resolução 3×3 .

3.3.1.5 Skip connection

As *skip connections* são operadores que levam às ativações das camadas iniciais para camadas mais a frente na arquitetura da rede. Como o nome sugere, as *skip connections* “pulam” algumas das camadas da rede neural e usam essa saída como entrada de uma camada mais a frente. Li et al. (2017) mostraram que a superfície da função de perda

em uma rede neural com *skip connections* é mais suave, levando a uma convergência mais rápida do que a que acontece nas redes sem *skip connections*.

De acordo com o que foi apresentado no capítulo anterior, o processo de atualização dos pesos da rede neural, por meio do algoritmo de retropropagação do erro, requer a multiplicação de termos com o gradiente. Quanto mais complexa a arquitetura da rede, maior é o número de multiplicações realizadas (conforme a regra da cadeia) ao longo do processo de retropropagação, camada por camada. Como resultado, os gradientes se tornam muito pequenos, pois os elementos da multiplicação possuem módulo menor que um. Em algumas situações, os gradientes chegam a zero, impedindo a atualização das camadas iniciais. Este fenômeno é conhecido na literatura como “gradientes de fuga” (do inglês *vanishing gradients*).

As *skip connections* mitigam este problema, garantindo o fluxo contínuo de gradientes das camadas iniciais até as camadas finais. Além disso, nas arquiteturas encoder-decoder, *skip connections* mais longas ajudam na recuperação da informação espacial que muitas vezes é perdida nas etapas de redução do tamanho da imagem.

3.3.1.6 Softmax

A camada softmax é geralmente usada na camada de saída da rede para classificar a entrada em diferentes categorias. Em termos simples, a camada softmax transforma um vetor de valores de entrada em uma distribuição de probabilidade, onde cada elemento na saída representa a probabilidade da entrada pertencer a uma das classes possíveis. Essa transformação é realizada pela aplicação de uma função exponencial normalizada aos valores de entrada (Goodfellow et al., 2016).

A camada softmax é especialmente útil em tarefas de classificação, pois permite que a rede neural convolucional atribua probabilidades para cada classe possível. Esse fato é importante para avaliar a confiança do modelo em suas previsões e também para selecionar a classe com a maior probabilidade como a classe final de saída.

3.3.1.7 Visão geral

Os blocos mais comuns de uma CNN consistem em sequências de camadas convolucionais e *pooling*. Na Figura 3.7, é apresentada a sequência típica de camadas de uma CNN. A fim de estimar as probabilidades de cada classe, é necessário que a última camada seja “achatada” (do inglês *flattening*), ou seja, que o mapa de características resultante das camadas convolucionais seja convertido em um vetor unidimensional e, em seguida, totalmente conectado a saída. Normalmente, a função de ativação utilizada na última camada é uma *softmax*, que confere a essa operação um caráter de cálculo de probabilidades.

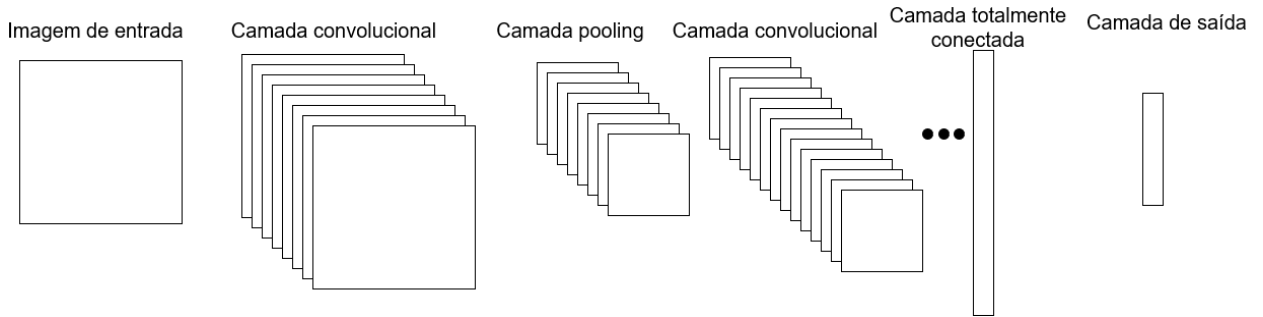


Figura 3.7: Arquitetura típica de uma CNN.

3.3.2 Classificação de imagens

A classificação de imagens é uma tarefa corriqueira dentro do domínio da visão computacional, tendo sido estudada e aprimorada ao longo dos anos. Tem por missão identificar os objetos classificáveis existentes em uma imagem. Entre as arquiteturas CNN que se destacaram nessa tarefa podem ser citadas: *i)* LeNet (LeCun et al., 1998), aplicada originalmente para reconhecer dígitos manuscritos no dataset MNIST; *ii)* AlexNet (Krizhevsky et al., 2012), que venceu o concurso ImageNet 2010, classificando 1.2 milhões de imagens em 1000 classes; e *iii)* VGG Net (Simonyan and Zisserman, 2015), consideravelmente mais profunda que as antecessoras, conseguiu superar os resultados da AlexNet, se tornando uma das arquiteturas CNN mais populares na última década. Neste trabalho foi utilizada a rede U-Net, que é fruto do desenvolvimento das arquiteturas pioneiras mencionadas.

3.4 U-Net

A U-Net é uma arquitetura de rede neural convolucional proposta por Ronneberger et al. (2015) para tarefas de segmentação de imagem. É uma arquitetura muito popular em áreas como medicina, biologia e engenharia, devido a sua capacidade de produzir resultados precisos com dados anotados limitados.

A principal característica da U-Net é ter um caminho de compressão e outro de expansão da imagem, no que ficou conhecido como estratégia *encoding-decoding*. Na rede *encoder*, uma sequência de camadas de convolução e *pooling* reduz o tamanho da imagem de entrada e extrai características em diferentes escalas. A rede *decoder* é a versão espelhada da *encoder*, usando convoluções transpostas para reconstruir a resolução espacial da imagem. A inovação chave da U-Net é a *skip connection*, que permite que a rede *decoder* acesse os atributos de alta resolução da *encoder* e produza segmentações mais precisas. A camada central da arquitetura de uma U-Net é conhecida como *bottleneck*. É nesta camada que a rede aprende as características mais importantes da imagem e onde

ocorre a maior perda de informação (Ronneberger et al., 2015), tendo assim um papel crítico na segmentação.

De maneira geral, a U-Net apresenta resultados impressionantes em diversos problemas de estudo. Sua estrutura em forma de “U” garante que cada pixel na saída contenha toda a informação da entrada, tornando-a uma abordagem adequada para o mapeamento entre diferentes domínios, como dados sísmicos para fácies sísmicas. No entanto, para alcançar esse resultado, os dados de entrada e saída nos conjuntos de treinamento e teste devem ter o mesmo tamanho, ou passar por algum processamento para ajustar as dimensões (Yu and Ma, 2021).

3.5 Técnicas de aumento de dados

O aumento de dados (do inglês *data augmentation*) é uma técnica de aprendizado de máquina que visa aumentar o número de exemplos de treinamento disponíveis para um modelo, aplicando diferentes operações aos dados existentes. Essas operações incluem translação, achatamento, alongamento, zoom, giro, distorção, espelhamento, rotação, adição de ruído, entre outros. Esta técnica apresenta diversas vantagens, tais como a prevenção do *overfitting*, o aumento da acurácia do modelo e redução do custo operacional de anotação e preparação do conjunto de dados bruto (Goodfellow et al., 2016).

Por vezes, o tamanho original da imagem não é adequado para ser apropriadamente processado pela arquitetura de uma CNN (em outros casos o tamanho da imagem é grande demais para a infraestrutura disponível de processamento), sendo necessário fragmentar a imagem em pedaços menores, denominados *patches*, para que cada parte seja analisada e prevista separadamente. Uma forma alternativa para se obter aumento do conjunto de treinamento, é ajustar o parâmetro *stride* (quantidade de pixels que um recorte de imagem é movido ao longo de um eixo) de modo a que haja superposição de partes da imagem entre *patches* vizinhos, como mostram as Figuras 3.8 e 3.9. Neste trabalho, essa estratégia foi utilizada para aumentar o conjunto de imagens de treinamento retiradas a partir das seções sísmicas.



Figura 3.8: Imagem cujo tamanho não é adequado como entrada da CNN.



Figura 3.9: Recortes da imagem original no tamanho adequado para servir de entrada para a CNN. Notar trechos de superposição entre as imagens.

Capítulo 4

Aprendizado ativo para classificação de fácies sísmicas

Muito esforço tem sido dedicado para abordar problemas típicos no reconhecimento de padrões sísmicos, incluindo a delineação de corpos de sal e a detecção de falhas, como evidenciado nos trabalhos de Zeiler and Fergus (2013), Badrinarayanan et al. (2017) e Islam (2020). No entanto, é notável a falta de pesquisa relativa à interpretação sísmica estratigráfica assistida por CNNs. Embora seja aparentemente simples aplicar CNNs para esse fim, Di et al. (2020) destacam duas principais limitações: em primeiro lugar, é necessária uma grande quantidade de dados de treinamento para construir um modelo de aprendizado de máquina para interpretação sísmica confiável. Em segundo lugar, a aplicação de CNNs requer seções sísmicas (ou partes delas) anotadas para treinamento, o que nem sempre é viável, devido ao trabalho intensivo envolvido na anotação, bem como, às incertezas presentes na complexa tarefa de interpretação sísmica. Nesse sentido, as técnicas de aprendizado ativo (AL, do inglês *active learning*) têm-se mostrado promissoras para reduzir o processo trabalhoso e intensivo de anotação dos dados de treinamento. Este capítulo apresenta uma descrição da abordagem de AL e delinea o processo de treinamento de uma Rede Neural Profunda para a classificação de fácies sísmicas usando AL.

4.1 Aprendizado ativo

A proposta do AL é imitar a habilidade dos seres humanos de aprender ao longo da tarefa (Kweon and Kim (2019)), com o objetivo de treinar modelos de alta precisão com uma quantidade limitada de dados. Um modelo de AL seleciona observações mais informativas para anotação, com base em uma estratégia de amostragem que determina que parte dos dados deve ser adicionada ao conjunto de treinamento. O dataset atualizado

é então utilizado na próxima etapa de treinamento, e esse processo é repetido até que se obtenham resultados satisfatórios.

A aplicação das técnicas de AL é especialmente relevante em problemas que demandam um alto custo de anotação, isto é, que dependem de uma grande quantidade de tempo de especialistas. Por essa razão, muitos trabalhos que aplicam essas técnicas se concentram na área de segmentação de imagens médicas. Em particular, estudos realizados por Zhang et al. (2019a,b) mostraram que treinar modelos apenas com a parte mais informativa do conjunto de dados anotado pode produzir resultados similares aos obtidos com todo o conjunto de dados de treinamento anotado. Smailagic et al. (2020) propuseram uma nova metodologia de AL aplicada à análise de imagens médicas, que obteve acurácia igual à obtida pela rede neural base, utilizando apenas 25% de anotação do conjunto de dados.

Assim como a análise de imagens médicas, a interpretação de dados sísmicos é uma tarefa que consome tempo do intérprete. O tempo necessário para a interpretação completa de um volume de dados sísmicos 3D varia muito e depende de diversos fatores como tamanho do dado, qualidade sísmica, complexidade geológica e número de fácies a serem identificadas. Entretanto, a tarefa sempre é realizada numa escala de tempo de semanas, meses de trabalho, dificilmente menos do que isso. Na prática, os intérpretes trabalham em ciclos de interpretação e avaliação, a fim de aprimorar gradualmente os resultados. Nesse sentido, é desejável que abordagens computacionais, a exemplo do AL, emulem esse processo iterativo.

4.2 Estratégias de amostragem

Nesta seção é fornecida uma breve descrição das estratégias de amostragem mais utilizadas em AL. Normalmente essas técnicas são comparadas com uma amostragem aleatória simples, que consiste em simplesmente escolher aleatoriamente um conjunto de imagens para ser acrescentado ao modelo na próxima rodada de treinamento. Em geral, essa abordagem se mostra um ponto de partida surpreendentemente difícil de ser batido (Nath et al., 2021).

4.2.1 Amostragem por incerteza

O método de amostragem por incerteza consiste na seleção dos pontos onde as predições do modelo são mais incertas na classificação (Burmeister et al., 2022). Duas abordagens bem conhecidas para capturar a incerteza do modelo são o método da menor confiança (Sharma et al., 2019) e a entropia de Shannon (Shannon, 1948).

4.2.1.1 Amostragem baseada na menor confiança

Esse tipo de amostragem baseia-se na medida da confiança do modelo em sua predição para uma determinada observação do conjunto de dados. Ele seleciona os pontos onde a probabilidade prevista pelo modelo para a classificação correta é a mais baixa. A intuição por trás dessa abordagem é que o modelo tem menos certeza sobre o rótulo correto para esses pontos de dados e, portanto, a rotulagem desses pontos pode ajudar a melhorar o desempenho do modelo.

Para o caso em que a observação em questão é uma imagem bidimensional, a incerteza baseada na menor confiança é calculada da seguinte maneira: para cada imagem $x \in \mathbb{R}^{w \times h}$ pertencente ao conjunto não anotado de dados de treinamento, a predição $P_\theta(\hat{y}|x)$ do modelo atual θ é calculada, sendo $\hat{y} \in \mathbb{R}^K$ o vetor de classes predito, K o número de classes, e w e h , respectivamente, a largura e altura da imagem. Em seguida, para cada pixel, é calculada a distância entre a predição e um valor máximo de incerteza α . Quanto mais próximo de α estiver a probabilidade predita, mais incerto é o modelo. A incerteza da imagem calculada pela menor confiança, $x_{MC}^*(x)$, é obtida a partir da soma negativa das medidas de distância de cada classe em todos os pixels:

$$x_{MC}^*(x) = - \sum_{l=1}^K \sum_{i=1}^w \sum_{j=1}^h |\alpha - P_\theta(y_{l,i,j}|x)| \quad (4.1)$$

com

$$\alpha = \begin{cases} 0.5, & \text{para um problema de classificação binário} \\ 1/K, & \text{para qualquer número de classes, onde } K = \# \text{ classes} \end{cases}$$

4.2.1.2 Amostragem baseada na entropia

Outra forma de medir incerteza é através da entropia de Shannon (Shannon, 1948): uma medida da incerteza, ou “desordem”, em uma distribuição de probabilidade. No contexto da amostragem por incerteza, a entropia de Shannon é usada para medir a incerteza da distribuição de predição do modelo para um determinado ponto dos dados. Ela seleciona os pontos onde a distribuição de probabilidade prevista pelo modelo é mais “espalhada,” ou tem a entropia mais alta. A intuição por trás dessa abordagem é que, se o modelo tem incerteza sobre o rótulo correto para um ponto dos dados, a distribuição de probabilidade prevista pelo modelo será mais espalhada entre os diferentes rótulos.

Nesta abordagem, a incerteza é calculada da seguinte maneira: para cada imagem $x \in \mathbb{R}^{w \times h}$ é calculada a soma das medidas de distância entre a predição e α e multiplicado pelo logaritmo das medidas de distância:

$$x_E^*(x) = - \sum_{l=1}^K \sum_{i=1}^w \sum_{j=1}^h |\alpha - P_\theta(y_{l,i,j}|x)| \times \log |\alpha - P_\theta(y_{l,i,j}|x)| \quad (4.2)$$

com

$$\alpha = \begin{cases} 0.5, & \text{para um problema de classificação binário} \\ 1/K, & \text{para qualquer número de classes, onde } K = \# \text{ classes} \end{cases}$$

4.2.2 Amostragem por representatividade

A amostragem por representatividade é uma técnica que busca maximizar a diversidade e representatividade do conjunto de dados de treinamento. Essa técnica pode ser classificada em dois grupos: amostragem independente do modelo e amostragem dependente do modelo. No primeiro caso, as amostras são selecionadas sem levar em conta as informações do modelo. No segundo caso, a seleção das amostras é baseada em características específicas do modelo, visando maximizar o desempenho da aprendizagem de máquina. Dentre as estratégias mais utilizadas para a amostragem dependente do modelo, destacam-se a amostragem baseada em agrupamento (clustering) e a amostragem baseada em distância.

4.2.2.1 Amostragem aleatória estratificada

Esta estratégia de amostragem é usada quando o conjunto de dados contém classes com diferentes proporções. A ideia é selecionar uma amostra aleatória simples proporcional ao tamanho das classes, de modo a preservar a distribuição original do conjunto de dados. Isso é feito de forma aleatória, garantindo que todas as imagens tenham a mesma chance de serem selecionadas. A amostragem aleatória estratificada é amplamente utilizada em problemas de classificação, onde a distribuição das classes é uma informação importante.

4.2.2.2 Amostragem estratificada em passos

A amostragem estratificada em passos consiste na seleção de um subconjunto de elementos de uma população, tomando-se cada n -ésimo elemento para ser adicionado ao conjunto de treinamento, onde n é o passo ou salto. Por exemplo, se o passo é dois, um item é escolhido, o próximo é deixado de lado e assim sucessivamente. Essa abordagem é frequentemente utilizada em problemas de análise de imagens, nos quais os dados usualmente são coletados em uma sequência contínua, ou grid, e a seleção da n -ésima amostra pode reduzir a quantidade de dados que precisa ser processada, sem grande prejuízo para os resultados. No contexto específico da atividade de interpretação sísmica, essa técnica

é similar às estratégias utilizadas pelos especialistas para lidar com grandes conjuntos de dados a serem interpretados. Normalmente, os intérpretes criam uma malha de interpretação, saltando 10, 20, 50 linhas em uma determinada direção (*inlines* ou *crosslines*), dependendo do tamanho da área e do tempo disponível, interpolando as interpretações no final.

4.2.2.3 Amostragem por representatividade baseada em agrupamentos

A amostragem baseada em agrupamentos visa a segmentação do conjunto de dados em grupos (clusters) com características semelhantes, de modo a permitir a seleção de uma amostra representativa de cada grupo. Nas aplicações de visão computacional, essa e outras abordagens de amostragem por representatividade dependentes do modelo utilizam vetores de características que representam o conteúdo da imagem de uma forma comprimida. A fim de obter esses vetores de características, é realizada uma predição em cada imagem $x \in \mathbb{R}^{w \times h}$ do conjunto de dados de treinamento utilizando o modelo atual θ . A partir disso, é extraído um vetor de características de uma das camadas internas do modelo, normalmente da camada *bottleneck*, no caso da U-Net. Adicionalmente, outras técnicas de redução de dimensionalidade, como max-pooling ou análise dos componentes principais (PCA, do inglês *principal component analysis*) são aplicadas.

Posteriormente, utilizando-se os vetores de características obtidos, é realizado um agrupamento dos dados. As imagens não anotadas são selecionadas a partir dos clusters obtidos, de modo que cada cluster seja igualmente representado no conjunto de treinamento em conformidade com o seu tamanho relativo.

4.2.2.4 Amostragem por representatividade baseada na distância

Essa estratégia de amostragem é empregada quando o conjunto de dados contém amostras similares ou redundantes. A ideia subjacente é selecionar uma amostra representativa de cada região do espaço de características do conjunto de dados (Sharma et al., 2019). A seleção é realizada levando em conta a distância euclidiana média entre cada vetor de características das imagens não anotadas e cada vetor de características do conjunto de dados de treinamento, definido como:

$$s(x) = \frac{1}{N} \sum_{i=1}^N d(f_i, f_x), \quad (4.3)$$

onde N é o tamanho do conjunto de dados de treinamento, f_i é o vetor de características da i -ésima imagem do conjunto de dados de treinamento, f_x é o vetor de características de uma imagem não anotada x e d é a distância euclidiana. As imagens não anotadas são ordenadas de acordo com as maiores distâncias médias em relação aos vetores de

características do conjunto de dados de treinamento, sendo consideradas como as mais representativas aquelas com as maiores distâncias. Em seguida, são inseridas as M imagens mais representativas no conjunto de dados a ser utilizado na próxima etapa de treinamento.

4.3 Aplicação do aprendizado ativo à interpretação sísmica

A aplicação de técnicas de AL na interpretação sísmica é uma área de pesquisa emergente, com poucos estudos publicados na literatura científica. Recentemente, Mustafa and AlRegib (2023) publicaram um artigo pioneiro que propõe uma metodologia para aprendizado ativo em classificação de fácies sísmicas. Até onde a investigação realizada indica, esse é o primeiro estudo que caracteriza dados sísmicos não anotados em ordem de relevância para o modelo de aprendizado de máquina, o que representa um avanço significativo para a área. A publicação dos resultados deles ocorreu após a definição de metodologia deste trabalho e, embora existam algumas similaridades entre os dois estudos, eles foram desenvolvidos de maneira independente.

Uma das principais diferenças entre os dois trabalhos é o escopo das simulações de aprendizado ativo. Mustafa and AlRegib (2023) realizaram apenas cinco iterações de aprendizado ativo, começando com apenas uma das *crosslines* (retirada do conjunto de dados F3) no conjunto de treinamento e adicionando uma nova *crossline* a cada iteração. No final, eles utilizaram aproximadamente 1% do conjunto de dados de treinamento disponíveis. Nas próximas seções, será detalhado o protocolo de aprendizado ativo empregado neste estudo. Resumidamente, o processo de simulação é iniciado com a utilização de 1% do conjunto de dados de treinamento, e são realizadas 20 iterações para cada método, culminando no emprego de 100% dos dados anotados. É importante ressaltar que o número total de *inlines* e *crosslines* utilizados é significativamente maior, proporcionando uma base mais sólida para a comparação dos métodos de AL. Isso permite uma avaliação mais precisa do comportamento das métricas ao longo das interações entre o modelo e intérprete, garantindo maior confiabilidade nos resultados obtidos.

No domínio da classificação de imagens, existem várias estruturas para realização de AL, especialmente na literatura médica, como já foi enfatizado em seções anteriores. Uma das contribuições deste trabalho é a introdução de uma estrutura de AL aplicada à segmentação de imagens sísmicas extraídas de volumes sísmicos tridimensionais.

Capítulo 5

Metodologia

5.1 Dados utilizados

O conjunto de dados utilizados neste estudo foi obtido a partir do desafio do workshop de interpretação com machine learning, que fez parte da reunião anual da SEG realizada em 2020, conhecido como dataset Parihaka. Esse volume de dados foi adquirido na Bacia Taranaki (Figura 5.1), uma bacia rifte onshore-offshore de Idade Cretácea, localizada na costa oeste da Nova Zelândia. O desenvolvimento dessa bacia se deu no contexto dos estresses extensionais resultantes da abertura do Gondwana. A Bacia cobre aproximadamente 100.000 km^2 , contendo em sua maioria sedimentos marinhos, com sedimentação terrestre significativa em Idades mais recentes (Kumar, 2016). A maioria da produção de óleo e gás da Nova Zelândia ocorre dentro dessa Bacia, com centenas de poços e dezenas de campos de óleo e gás descobertos (Alves, 2020).

O conjunto de dados é composto por um volume sísmico 3D fornecido pela *New Zealand Petroleum and Minerals* (NZPM) e um volume de rótulos contendo a interpretação de fácies sísmicas, anotado pela Chevron Inc., San Ramon, CA, USA. A quantidade de área coberta pelos dados é de aproximadamente 294 km^2 , com números de *inlines* variando de 1 a 590 e número de *crosslines* variando de 1 a 782. Cada traço contém 1006 pontos de tempo, com um intervalo de amostragem $dt = 3ms$. A Figura 5.2 apresenta a geometria dos dados adquiridos e a forma como eles foram divididos para o desafio da SEG.

Na Figura 5.3 é apresentado um exemplo do conjunto de treinamento, com uma seção sísmica e a máscara correspondente com as fácies sísmicas interpretadas. Cada pixel no volume de rótulos é anotado com um valor de 1 a 6, denotando sua classificação de fácies (essas e outras informações foram retiradas do site oficial do workshop de interpretação com ML da SEG 2020, <https://public.3.basecamp.com/p/JyT276MM7krjYrMoLqLQ6xST>):

- 1) Embasamento/Outro (Cor Vermelha): Baixa relação sinal-ruído (SNR), poucos refle-

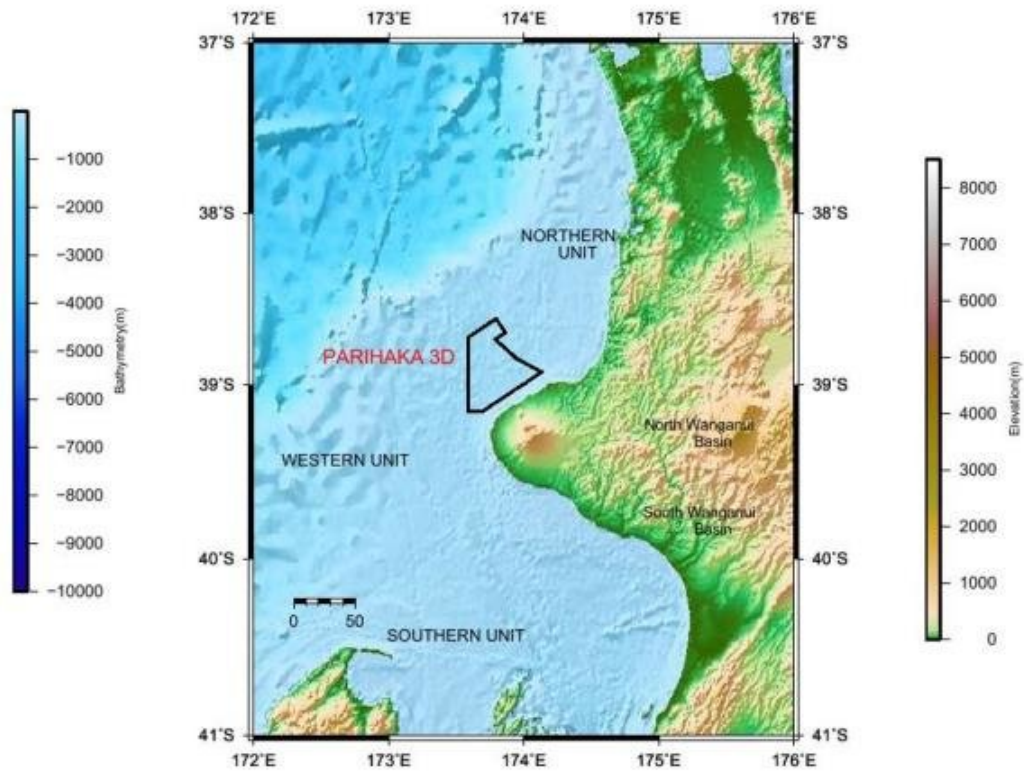


Figura 5.1: Mapa geológico da bacia Taranaki, incluindo o 3D Parihaka - marcado pelo polígono em preto. (Fonte: Kumar (2016))

tores internos. Pode conter vulcões em alguns lugares. 2) Mudstone de Talude A (Cor Verde): Mudstones do talude ao piso da Bacia, limites superior e inferior de alta amplitude e refletores internos contínuos de baixa amplitude. 3) Depósito de Transporte de Massa (Cor Laranja): Mistura de fácies caótica e reflexões paralelas de baixa amplitude. 4) Mudstone de Talude B (Cor Azul): Mudstones e Arenitos do talude ao piso da bacia, refletores paralelos de alta amplitude e superfícies de baixa continuidade. 5) Vale Inciso (Cor Fúcsia): Canais/vales incisos de grande amplitude e relevo relativamente baixo. 6) Sistema de Canyon Submarino (Cor Amarela): De base erosiva em forma de U com alto relevo local.

Estas fácies possuem características de amplitude, continuidade e geometria que resultam em diferentes respostas sísmicas. Observa-se um desbalanceamento na distribuição de classes, sendo as fácies 3, 5 e 6 relativamente menos frequentes. O desafio proposto pela SEG consistiu na aplicação de modelos de DL no conjunto de dados de treinamento, a fim de prever a ocorrência de cada uma das seis fácies em volumes de teste não anotados. Os resultados foram avaliados primariamente por meio da média dos valores de interseção sobre união (do inglês *Intersection Over Union (IOU)*), uma métrica amplamente utilizada na avaliação do desempenho de modelos de segmentação de imagens, para as seis

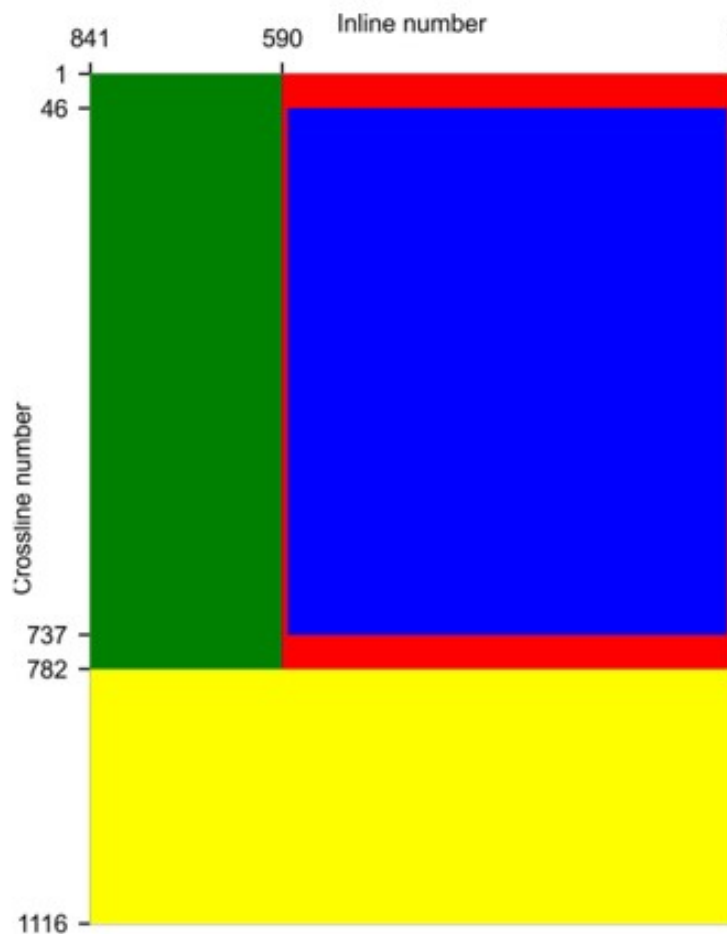


Figura 5.2: Geometria do dado sísmico Parihaka, do desafio do workshop de interpretação assistida por ML da SEG, em 2020. A área azul indica os dados de treinamento. A área vermelha indica os dados de teste, não usados em nenhuma etapa do treinamento. As áreas verde e amarela indicam as posições dos dois conjuntos de dados de teste cego fornecidos pela SEG. (Fonte: Chai et al. (2022))

fácies. Detalhes adicionais sobre esta e outras métricas de avaliação serão apresentados em seções posteriores deste capítulo.

5.2 Arquitetura adotada

A literatura que trata de novas variantes de redes neurais convolucionais para tarefas de segmentação é ampla e em constante expansão, com novas publicações sendo produzidas a todo momento com resultados variados, dependendo em grande medida do tipo de dataset utilizado. Nesse contexto, não é objetivo deste trabalho realizar uma comparação sistemática dos resultados obtidos por diferentes arquiteturas, tendo em vista que os dados sísmicos empregados demandam um tempo considerável para o treinamento de

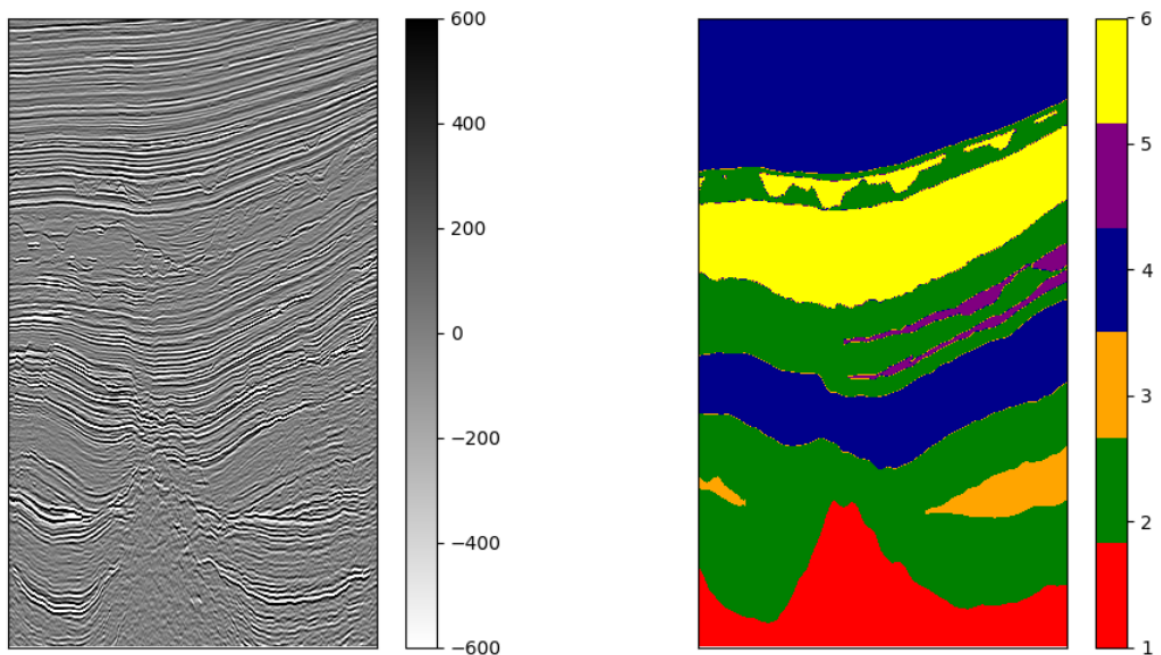


Figura 5.3: Visualização de um exemplo dos dados de treinamento: seção sísmica na crossline de número 780 (esquerda) e seção anotada correspondente.

modelos. Ademais, durante a aplicação das técnicas de aprendizado ativo, esses treinamentos são repetidos exaustivamente, com elevado custo computacional resultante das simulações. Diante disso, adotou-se a arquitetura proposta por Chai et al. (2022), a qual demonstrou ser muito efetiva para SFC nos inúmeros testes que eles realizaram em seu trabalho, apresentando as melhores métricas de Acurácia e IOU dentre os trabalhos similares revistos.

A Figura 5.4 exibe a arquitetura empregada para predições de fácies sísmicas: que consiste em uma versão aprimorada de uma U-Net padrão. A arquitetura da CNN aqui apresentada é otimizada em relação à U-Net padrão devido a escolhas chave de hiperparâmetros como taxa de aprendizagem, função de *pooling* e tamanho do *kernel* convolucional, conforme proposto por Chai et al. (2022) e Falk et al. (2018). Essa arquitetura também foi testada em outras aplicações, como na interpolação de dados sísmicos (Chai et al., 2020) e na deconvolução multitraços em dados sísmicos (Chai et al., 2021).

Na Figura 5.4, à medida que os retângulos diminuem de tamanho eles passam a possuir um maior número de mapas de características e uma resolução espacial menor. A rede *encoder* à esquerda recebe a seção sísmica original como entrada, e extrai mapas de características de diferentes resoluções passo a passo. O número de pixels nas dimensões temporal (vertical) e espacial (horizontal) da imagem de entrada é denotado por N_1 e N_2 , respectivamente. A rede *decoder* à direita sucessivamente compõe as fácies sísmicas, começando nos mapas de característica de larga escala até chegar nos de fina escala. A rede

encoder é composta de sucessivas aplicações de convoluções 2D *same-padding* (retângulos amarelos), cada uma seguida da aplicação de um operador de ativação e um operador *pooling* (retângulo vermelho). O fator de redução da dimensão da imagem na camada *pooling* é de dois. O lado *decoder* contém um operador de convolução transposta 2D (retângulo azul), seguido por um operador de concatenação com os mapas de características da rede *encoder* com a resolução correspondente (seta roxa) e repetidas camadas convolucionais 2D, cada uma seguida de um operador de ativação. No final, uma camada convolucional 1×1 seguida por um operador *softmax* é usada para projetar os resultados dos mapas de ativação no resultado desejado de SFC. Como função de ativação foi utilizada a ReLU, que entrega como saída $\max(\mathbf{x}, 0)$, em que \mathbf{x} é um tensor.

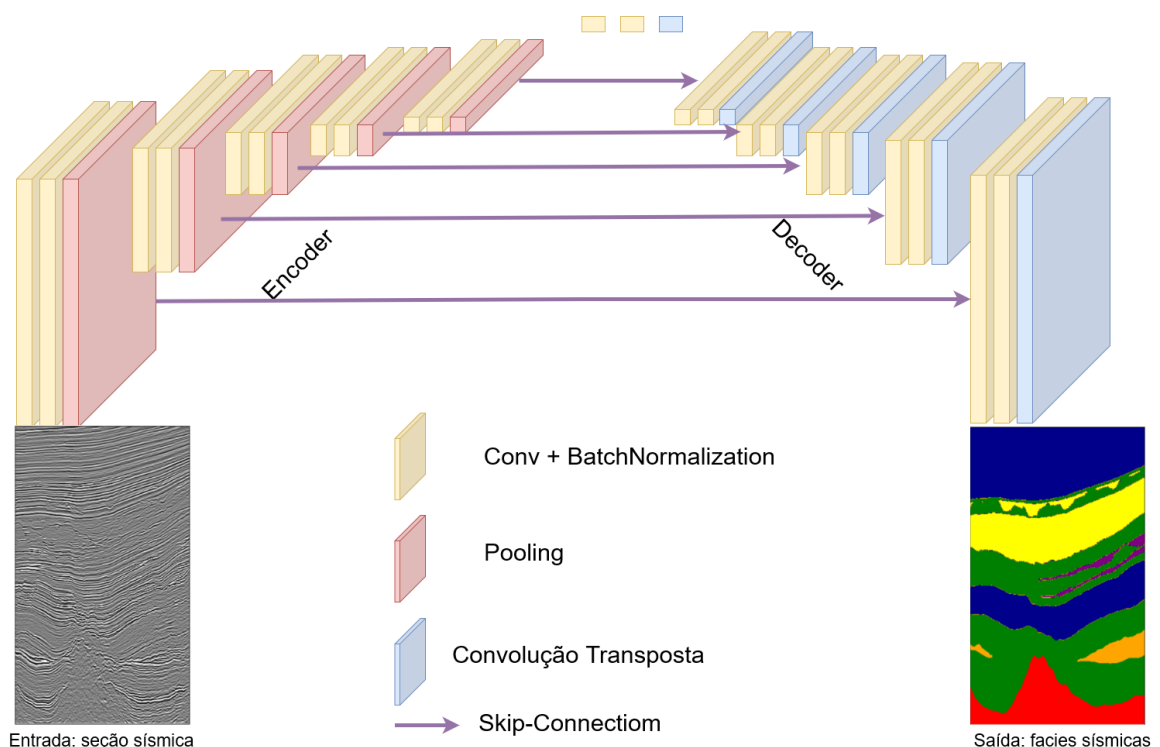


Figura 5.4: Ilustração da arquitetura da rede adotada, proposta por Chai et al. (2022).

Nas próximas seções serão fornecidos mais detalhes a respeito dos motivos da escolha do tamanho do *kernel* dos operadores de convolução e do número de repetições dos blocos de operadores *pooling*/convolução transposta. Os códigos foram desenvolvidos no framework de DL PyTorch, que oferece uma ampla variedade de ferramentas para o desenvolvimento de redes neurais profundas.

5.3 Treinamento da rede

Para a arquitetura selecionada, alguns parâmetros considerados (por exemplo, o tamanho do *kernel*) e as estratégias de treinamento tem como limitação principal para

sua definição o hardware disponível, em particular a unidade de processamento gráfico (GPU), cuja especificação disponível foi uma NVIDIA Tesla V100 (32GB).

As estratégias de treinamento inicialmente adotadas seguiram a metodologia descrita por Chai et al. (2022), a fim de validar a eficácia dos códigos produzidos e comparar as métricas com as divulgadas pelos autores, além de obter valores de referência para iniciar as simulações de aprendizado ativo (AL). Conforme as etapas de AL foram iniciadas, algumas adaptações se fizeram necessárias devido a limitações de infraestrutura e dos prazos para a entrega dos resultados da pesquisa.

5.3.1 Preparação do conjunto de dados de treinamento

A preparação dos conjuntos de dados de treinamento, validação e teste requer um cuidado especial em qualquer aplicação de DL. No caso de interpretação de volumes sísmicos 3D, ela é ainda mais relevante, já que seções sísmicas adjacentes possuem alta correlação entre si. É necessário garantir que os conjuntos de treino, validação e teste possuam representação suficiente de todas as categorias de fácies sísmicas e, ao mesmo tempo, minimizar a correlação entre os diferentes conjuntos o tanto quanto possível.

Considerando que a SEG não forneceu os volumes com as fácies anotadas para o conjunto de teste (regiões verde e amarela na Figura 5.2), foi feita uma divisão entre treino e teste da seguinte forma:

1. *Conjunto de Treinamento (área azul)*: Inlines [8,583] e crosslines [46,737].
2. *Conjunto de teste (área vermelha)*: Inlines $[1,7] \cup [584,590]$ e crosslines $[2,45] \cup [738,782]$.

A razão entre a área selecionada para treino e teste (86% versus 14%) é considerada adequada. É importante salientar que, já que a abordagem deste trabalho foi quantitativa, os dados sísmicos sem suas contrapartes anotadas do conjunto de teste cego não foram utilizados nos experimentos. Adicionalmente, a *crossline* de número 1 não foi considerada no treinamento, ou no cálculo das métricas de avaliação, devido a um evidente erro de anotação de uma classe (Figura 5.5), onde a fácies 6 está incorretamente anotada, supostamente exagerando a sua distribuição. Nota-se que a porção superior da fácies 6, em amarelo, não tem uma geometria condizente com um sistema de canyon submarino, possivelmente uma fácies de mudstone de talude registrada com o número errado.

5.3.2 Pré-processamento

A CNN foi treinada com uma janela de amostragem cujo tamanho foi o mais perto possível do tamanho das *inlines* e *crosslines* do volume sísmico, ao invés da opção usual

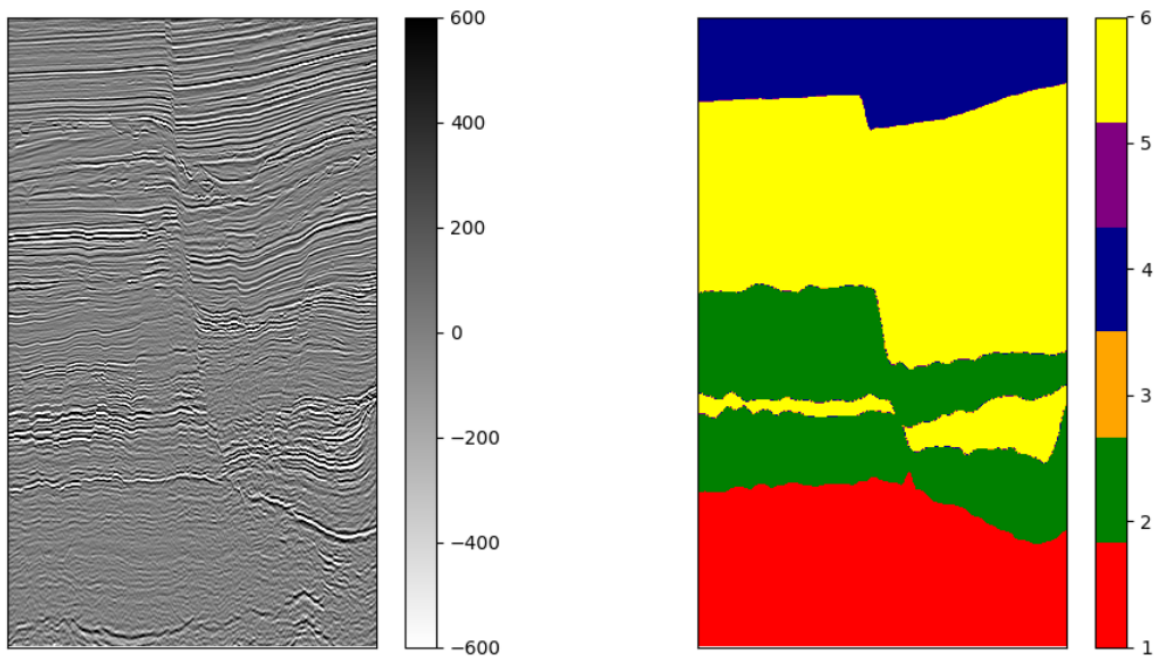


Figura 5.5: Crossline 1, retirada dos experimentos devido a erro de anotação de fácies.

por janelas menores de 256×256 pixels. Esse processo tem a vantagem de permitir que a rede considere as relações existentes entre diferentes unidades litoestratigráficas, e aproveite uma maior quantidade de informações de espaço e profundidade para diferenciar as categorias de fácies (Zhang et al., 2021). Por outro lado, o uso de uma janela menor de amostragem limita a CNN, que praticamente aprende apenas características locais (Li et al., 2021).

Entretanto, existem alguns limites para a escolha do tamanho da janela (*patch*) que devem ser considerados. Primeiramente, as crosslines e inlines possuem um número diferente de traços, o que impõe que a largura máxima do *patch* seja a menor entre as larguras originais das seções. Além disso, a entrada da U-Net é condicionada ao tamanho do fator de redução (S) e ao número de operações de *pooling*. Neste caso, como $S = 2$ e existem 5 etapas de *pooling*, a altura/largura da seção sísmica de entrada deve ter um valor múltiplo de $S^5 = 32$.

Considerando as vantagens de se utilizar a maior seção sísmica possível como entrada da rede e ao mesmo tempo respeitando os limites já discutidos, optou-se por um tamanho de amostra de 992×576 , que se aproxima do tamanho das seções sísmicas originais (1006×692 das inlines e 1006×576 das crosslines). O tamanho da imagem, em conjunto com a GPU em uso, determina o tamanho do lote de amostras (*batch size*) utilizado pelo algoritmo para treinar o modelo, durante uma etapa de atualização dos pesos no treinamento. Assim, para aproveitar a total capacidade da GPU disponível foi selecionado o valor 16 como *batch size*.

Foram coletadas amostras 2D (*patches*) de tamanho 992×576 de cada *inline* e *crossline* do volume sísmico 3D, com superposição entre imagens vizinhas controladas pelo parâmetro *stride*: $[8, 8]$ para o conjunto de treinamento e $[230, 14]$ para o conjunto de teste. Esse procedimento resultou em 29724 imagens para o conjunto de treinamento e 260 para o conjunto de teste. Em cada época de treinamento foram selecionadas de forma aleatória 10000 imagens dentre as quase 30000 disponíveis.

O único processamento de amplitude aplicado foi a normalização das seções sísmicas de entrada por meio dos valores para o intervalo $[-1, 1]$, dividindo cada pixel pelo valor absoluto máximo. Na Figura 5.6 alguns pares dos *patches* de treinamento são apresentados.

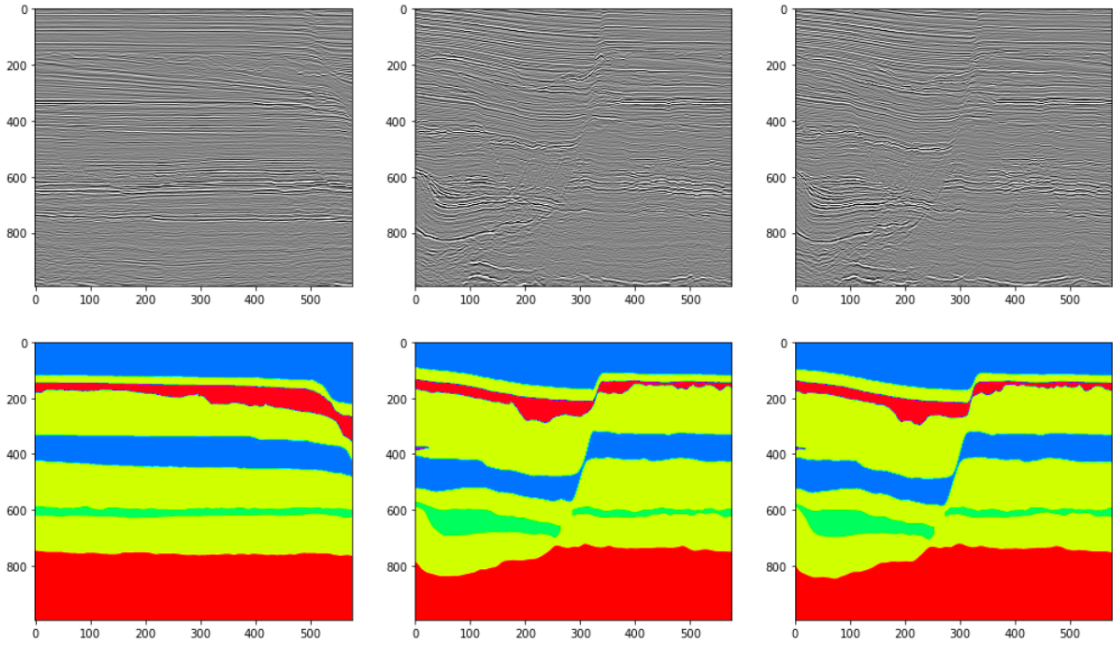


Figura 5.6: Alguns pares entrada/saída da CNN. Na primeira linha as seções sísmicas de entrada e abaixo as fácies classificadas. O tamanho de cada imagem é 992×576 .

5.3.3 Batch Normalization

Batch Normalization é uma técnica de treinamento de redes neurais que ajuda a acelerar o processo de treinamento, reduzir a variância dos parâmetros de saída, prevenir o overfitting e melhorar a precisão do modelo (Ioffe and Szegedy, 2015). Fundamentalmente, a *Batch Normalization* adiciona à arquitetura da rede camadas de normalização após a convolução e antes da função de ativação não linear. Para a entrada $x = [x_1 \dots x_m] \in B$, a *Batch Normalization* transforma as ativações de cada *batch* de acordo com a Equação 5.1:

$$BN(x_i) = \gamma \frac{x_i - \mu[x]}{\sqrt{Var[x]}} + \beta, \quad (5.1)$$

onde $\mu[x]$ e $Var[x]$ são a média e a variância da amostra do *batch* B. Dois parâmetros treináveis, γ e β , são introduzidos como fator de escala e de deslocamento, respectivamente. Nos experimentos realizados, foi adicionada uma camada de *Batch Normalization* após cada camada convolucional.

5.3.4 Função de perda

A escolha da função de perda é considerada essencial para o treinamento de uma CNN. Em aplicações de segmentação de imagem, a função de perda Entropia Cruzada (CE, do inglês *cross entropy loss*) é normalmente selecionada. Para problemas multiclasse, recomenda-se o uso da função de perda Categórica de Entropia Cruzada (CCE, do inglês *categorical cross entropy loss*). A perda CCE calcula o desajuste entre os rótulos \mathbf{Y}_{ref} e as previsões \mathbf{Y}_{pre} :

$$CCE(\mathbf{Y}_{ref}, \mathbf{Y}_{pre}) = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C \mathbf{Y}_{ref}^{i,j} \log(\mathbf{Y}_{pre}^{i,j}), \quad (5.2)$$

onde N representa o número de amostras e C o número de classes. $\mathbf{Y}_{ref}^{i,j}$ utiliza um esquema *one-hot-encoder* para descrever os rótulos (no caso de estudo, a fácies 2, por exemplo, é representada por $[0, 1, 0, 0, 0, 0]$), $\mathbf{Y}_{pre}^{i,j}$ é uma matriz com os valores preditos para cada classe, e os índices i e j iteram sobre todos os pixels e classes, respectivamente.

5.3.5 Adam optimizer

Durante o treinamento de modelos de ML, o processo de atualização dos parâmetros é realizado com base nos gradientes calculados a partir de *batches* aleatórios extraídos do conjunto de dados. Consequentemente, a função de perda é estocástica e pode ser otimizada pelo algoritmo Adam. Otimizador Adam, cujo acrônimo em inglês significa *Adaptive Moment Estimation*, é um método que combina as técnicas de gradiente descendente estocástico (SGD) e de gradiente descendente com momento (SGDM), e foi proposto por Kingma and Ba (2015). Esse algoritmo utiliza um estimador adaptativo de primeira e segunda ordem do momento das variáveis de gradiente, armazenando a média móvel e a variância móvel do gradiente ao longo do tempo. A média móvel é utilizada para calcular o momento do gradiente, enquanto a variância móvel é usada para ajustar a taxa de aprendizagem de cada parâmetro de rede. O Adam também emprega um mecanismo de correção de viés para lidar com o viés introduzido pela inicialização do estimador.

O Adam foi escolhido devido a sua boa adaptabilidade, velocidade de convergência e eficiência computacional, como mencionado por Ruder (2017); Reddi et al. (2018); Liu et al. (2019). É importante ressaltar que a biblioteca Pytorch oferece uma implementação pronta para uso do algoritmo Adam.

5.3.6 Escolha de hiperparâmetros

A seleção adequada dos valores dos hiperparâmetros é fundamental para o bom desempenho de modelos de DL em tarefas de classificação e predição. No entanto, a escolha desses fatores não é trivial e depende de diversos fatores, como a natureza dos dados, a complexidade do modelo e a estratégia de otimização empregada. Embora existam estudos anteriores que investigam a seleção de hiperparâmetros em diferentes contextos (Jadon, 2020; Zhang, 2018), é importante ressaltar que a aplicação dessas escolhas em dados sísmicos requer uma avaliação cuidadosa de sua validade para o problema específico de SFC.

Nesse sentido, Chai et al. (2022) conduziram uma análise sistemática dos hiperparâmetros no volume sísmico 3D Parihaka, avaliando a estratégia de *pooling*, o número de repetições nos blocos de convolução + ativação, o tamanho do kernel dos filtros e a taxa de aprendizagem. Com base em seus extensos testes, foram selecionados os melhores parâmetros que se mostraram adequados para o problema de SFC em questão, e que estão apresentados na Tabela 5.1.

Tabela 5.1: Hiperparâmetros escolhidos

Hiperparâmetro	Escolha/valor
Estratégia de Pooling	Max Pooling
Número de Blocos de Convolução + Ativação	3
Tamanho do Kernel	11×11
Taxa de Aprendizagem	0.0001

5.4 Teste

Na etapa de teste, o modelo resultante do treinamento é aplicado nas 206 imagens contidas no conjunto de teste. A avaliação dos resultados é feita de forma qualitativa, observando a continuidade e o caráter geológico das fácies preditas ao longo de uma seção, e também de maneira quantitativa, usando métricas conhecidas da área de visão computacional.

5.4.1 Métricas de desempenho

A fim de acompanhar a evolução do treinamento, o valor da função de perda é monitorado em cada época, com a expectativa de que os valores sejam cada vez menores a medida que o treino prossegue. Além da evolução da função de perda, outras métricas são usadas para avaliar objetivamente o desempenho das predições no conjunto de teste. As métricas escolhidas são amplamente adotadas na literatura de visão computacional (Xu et al., 2019): F1-escore (F1) e Interseção sobre União (IOU).

O conjunto de pixels que pertencem a classe i é designado como T_i , enquanto o conjunto de pontos classificados na classe i é designado P_i , onde $i \in [1, n_c]$ e n_c é o número de classes (fácies sísmicas) a serem classificadas. Para o conjunto de dados utilizado, $n_c = 6$. O conjunto de pontos corretamente classificados para a fácies i é dado pela intersecção entre T_i e P_i . O F1-escore para a fácies i é calculado da seguinte maneira ¹:

$$F1_{i, i \in [1, n_c]} = \frac{2 \times |T_i \cap P_i|}{|T_i| + |P_i|}, \quad (5.3)$$

em que $|\cdot|$ denota o número de pontos em um conjunto. Para representar o resultado global da métrica, a média do F1-escore (MF1) é definida sobre todas as classes:

$$MF1 = \frac{1}{n_c} \sum_{i=1}^{n_c} F1_i. \quad (5.4)$$

IOU_i é definida como o número de pontos de interseção entre T_i e P_i sobre o número de pontos de sua união (Alaudah et al., 2019):

$$IOU_{i, i \in [1, n_c]} = \frac{|T_i \cap P_i|}{|T_i \cup P_i|}. \quad (5.5)$$

IOU é igual a 1 se e somente se todos os pontos são classificados corretamente. Para uma média da IOU sobre todas as classes é usado o $MIOU$:

$$MIOU = \frac{1}{n_c} \sum_{i=1}^{n_c} IOU_i. \quad (5.6)$$

A desigualdade de classes no conjunto de dados tem o potencial de afetar o desempenho das métricas. Por isso, como estratégia comum, é recomendável ponderar cada classe de acordo com seu número de pontos (Alaudah et al., 2019), a fim de evitar resultados altamente sensíveis às classes com poucos pontos. Essa ponderação é realizada modificando a métrica da seguinte maneira:

¹Essa forma bem elegante de representar o F1-escore não é a mais comum na literatura. No Apêndice A ela é demonstrada.

$$wIoU = \frac{1}{\sum_{i=1}^{n_c} [T_i]} \sum_{i=1}^{n_c} [T_i] \frac{[T_i \cap P_i]}{[T_i \cup P_i]}. \quad (5.7)$$

5.5 Algoritmo de aprendizado ativo

A Figura 5.7 ilustra a estrutura geral do algoritmo de AL usado no contexto de SFC. As métricas obtidas a partir do conjunto completo de treinamento foram aproveitadas como referência (*baseline*) para o desempenho dos experimentos de AL. Uma malha primária de interpretação sísmica foi simulada por meio da seleção equidistante de 5% dos pares sísmica-rótulo disponíveis. Em seguida, foram realizadas 20 iterações de AL, cada uma composta por uma etapa de anotação e uma etapa de treinamento.

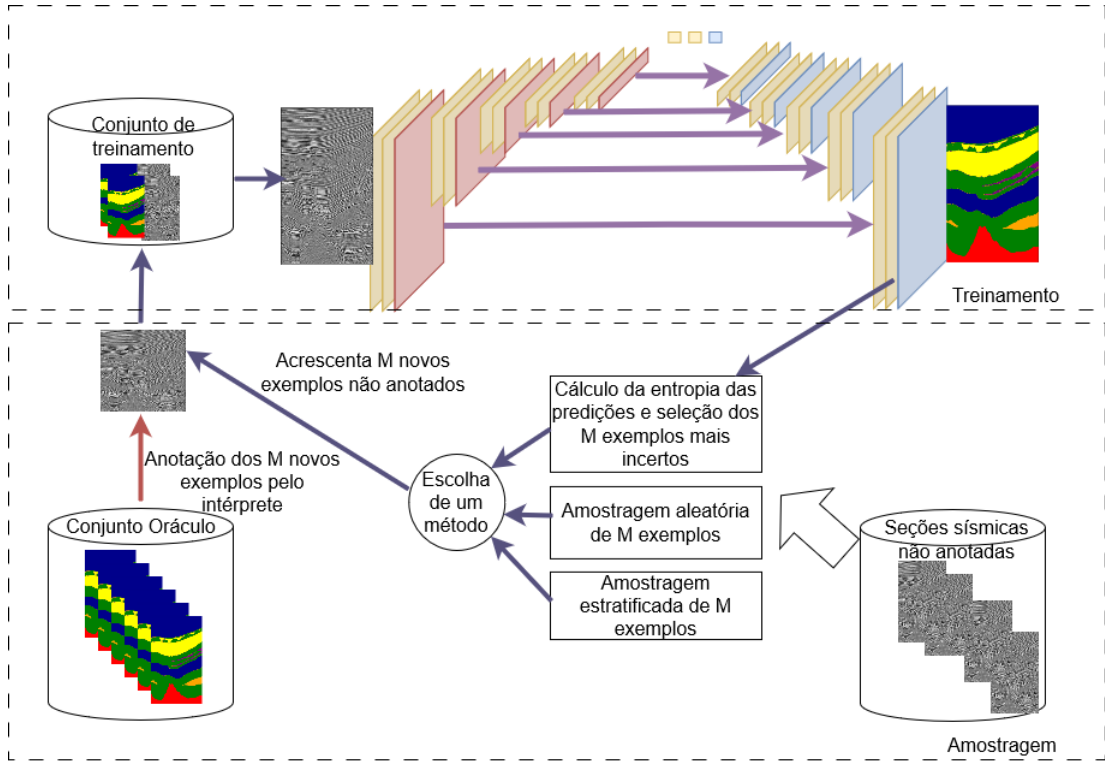


Figura 5.7: Fluxo proposto de aprendizado ativo para três diferentes tipos de estratégia de amostragem: amostragem por incerteza, amostragem aleatória e amostragem estratificada. A seta vermelha indica a única etapa em que é necessária a intervenção do intérprete.

Seja $\mathcal{D}_{treino}^i = \{(x_j, y_j)\}_{j=1}^{\lfloor \mathcal{D}_{treino}^i \rfloor}$ o conjunto de treinamento para o i -ésimo ciclo de AL. \mathcal{D}_{treino}^i é composto pelas seções sísmicas ($x_j \in \mathbb{R}^{N_1 \times N_2}$) e suas anotações ($y_j \in \mathbb{R}^{N_1 \times N_2}$). O conjunto completo com todas as seções sísmicas anotadas é denotado como $\mathcal{D}_{oráculo} = \{(y_j)\}_{j=1}^{\lfloor \mathcal{D}_{oráculo} \rfloor}$. Na etapa de anotação, uma estratégia de amostragem é aplicada para selecionar um número fixo de exemplos a serem anotados, cerca de 5% do

total, denotado como $M \in \mathcal{D}_{\text{não-annotadas}}$. Os exemplos selecionados são, então, anotados e adicionados a $\mathcal{D}_{\text{treino}}^{i+1}$. Uma vez que todo o conjunto de treinamento já está anotado em $\mathcal{D}_{\text{oráculo}}$, basta retirar desse conjunto as seções anotadas correspondentes às seções sísmicas escolhidas. Em uma situação real, essa etapa demandaria a intervenção do especialista, que conduziria às anotações das M imagens.

Após cada iteração de AL, a etapa de treinamento é realizada utilizando o novo conjunto de dados, que agora conta com exemplos adicionais. Os resultados das métricas são calculados no conjunto de teste e em seguida o processo é reiniciado. Foram avaliadas três diferentes estratégias de amostragem: amostragem aleatória simples, amostragem por incerteza baseada na entropia e amostragem estratificada em passos. Dessa forma, além da amostragem aleatória, que é comumente utilizada como referência para comparação, foram selecionadas representantes de duas abordagens amplamente conhecidas de AL: a amostragem por incerteza e a amostragem por representatividade.

5.6 Estimativa do intervalo de confiança

A maioria dos estudos revisados apresenta os resultados dos algoritmos de AL sem estimativas de intervalo de confiança. Essa prática é compreensível, especialmente no caso de modelos de DL para segmentação semântica de imagens, em razão do alto custo computacional envolvido na etapa de treinamento. Entretanto, alguns estudos se esforçaram para fornecer mais que estimativas pontuais. Burmeister et al. (2022) repetiram cada iteração de AL três vezes com diferentes sementes para os processos aleatórios de inicialização de pesos e seleção inicial do conjunto de treinamento, apresentando a média e o desvio padrão do F1-escore dessas três rodadas como métrica de avaliação. Meirelles et al. (2022) realizaram cinco execuções para calcular o intervalo de confiança, equivalente a um desvio padrão, para cada abordagem de sua aplicação, seguindo a mesma estratégia utilizada por Belharbi et al. (2021). Todas essas abordagens estimaram as incertezas através de técnicas cuja natureza é simplificada, com base nos processos aleatórios internos das redes convolucionais, o que exige repetições no treinamento, etapa mais custosa dentro do fluxo de AL.

Neste trabalho é proposta uma abordagem inovadora para estimar o intervalo de confiança das métricas avaliadas em um processo de AL. Ao contrário dos trabalhos revisados na área, o método proposto estima as incertezas do modelo através da avaliação da variação dos resultados do modelo treinado para diferentes conjuntos de teste, ao invés de realizar modificações na etapa de treinamento. Esta abordagem oferece a grande vantagem de ser substancialmente mais rápida, uma vez que o processo de predição demanda muito menos processamento do que o de treinamento, permitindo aumentar a quantidade

de repetições, resultando em maior robustez estatística. Para estabelecer variabilidade no conjunto de teste, foi empregada a técnica de reamostragem *bootstrap*.

Um intervalo de confiança *bootstrap* não paramétrico permite estimar a incerteza em torno de uma estatística amostral, como a média ou a mediana, sem pressupor a distribuição subjacente dos dados. Essa técnica consiste na criação de várias amostras, com reposição, a partir dos dados originais. Em cada amostra *bootstrap*, a estatística de interesse é estimada e, em seguida, calcula-se a variabilidade das estatísticas *bootstrap*. A partir dessa variabilidade, é possível construir um intervalo de confiança para a estatística de interesse (Davison and Hinkley, 1997). Dessa forma, o intervalo de confiança *bootstrap* é uma ferramenta útil para avaliar a incerteza em torno de uma estimativa amostral, mesmo em casos onde a distribuição dos dados não é conhecida.

A determinação do tamanho da amostra e do número de amostras *bootstrap* é uma etapa crítica nesse tipo de análise, uma vez que influencia diretamente na precisão e confiabilidade das estimativas obtidas. Em geral, o tamanho da amostra deve ser grande o suficiente para garantir que as estimativas sejam representativas da população subjacente (Efron, 1987). No entanto, não há uma regra geral para determinar o tamanho da amostra e o número de amostras *bootstrap*, sendo necessário muitas vezes realizar uma análise exploratória para determinar a adequação desses valores aos dados em questão.

No contexto do AL para segmentação semântica de imagens, a determinação desses parâmetros é ainda mais desafiadora devido ao elevado custo computacional da análise. Assim, é importante considerar o tempo total para o cálculo da métrica em todas as amostras *bootstrap* ao longo de todos os pontos da simulação de AL para definir o número de amostras. Foram realizados testes, apresentados no Capítulo 6, para definir a melhor escolha. Dessa forma, foram criadas 80 amostras *bootstrap*, cada uma com 100 imagens do conjunto de teste, para estabelecer os intervalos de confiança de 95% para cada métrica avaliada.

Capítulo 6

Resultados

6.1 Análise da convergência do processo de treinamento

O modelo foi treinado durante 50 épocas utilizando o conjunto completo de treinamento. A evolução do treinamento foi acompanhada pela acurácia e pela função de perda a cada época. Os gráficos da Figura 6.1 ilustram a evolução das métricas quantitativas ao longo das épocas, indicando um aumento consistente na acurácia e uma diminuição na perda. Esses resultados são coerentes entre si e com a análise visual. O treinamento teve uma duração média de 13 minutos por época, totalizando cerca de 11 horas para 50 épocas. A CNN utilizada possui cerca de 7,8 milhões de parâmetros ajustáveis, o que, juntamente com o tamanho do *patch* escolhido e o número de imagens do conjunto de treinamento, justifica o tempo necessário para o treinamento.

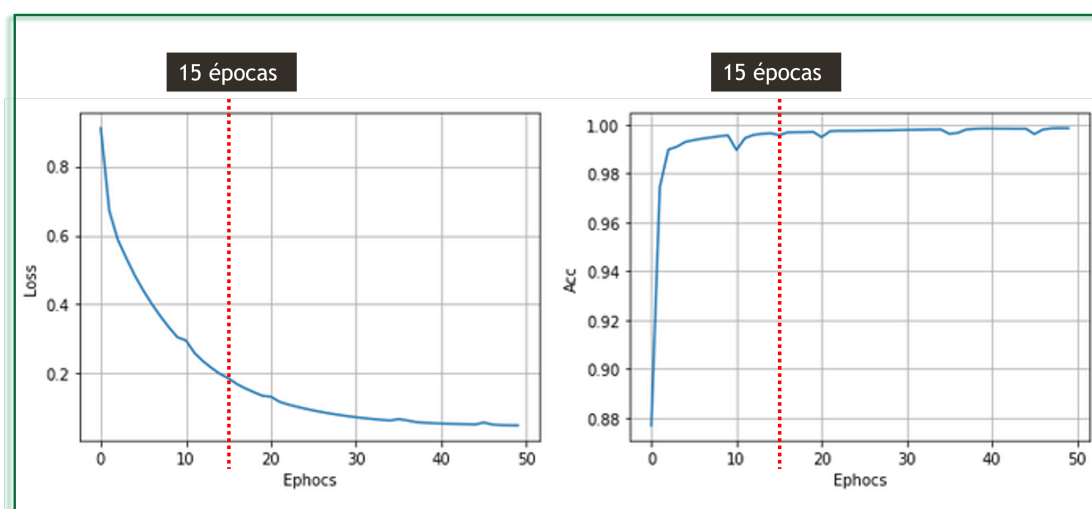


Figura 6.1: Perda e acurácia do modelo ao longo de 50 épocas de treinamento.

Em *deep learning* (DL), é possível que a perda seja muito baixa e que a rede

não generalize bem. O conjunto de teste permite avaliar a capacidade de generalização, uma vez que ele não é utilizado para atualizar os parâmetros da rede neural convolucional (CNN) durante o treinamento. Com o objetivo de acelerar o processo de treinamento, que é repetido várias vezes durante o processo de aprendizado ativo, foi decidido interromper o treinamento na 15^o época, reduzindo o tempo total de treinamento para três horas e meia. Na Seção 6.2, serão discutidas as consequências dessa redução no número de épocas na avaliação dos resultados no conjunto de teste.

6.2 Resultados no conjunto de teste

Para avaliar os modelos propostos, foram utilizadas 206 imagens com dimensões de 992×576 pixels, obtidas das *inlines* e *crosslines* reservadas para teste. As imagens $[1,14] \cup [105,118]$ foram extraídas de *inlines*, e as demais de *crosslines*, de modo a percorrer o conjunto de teste com o menor número possível de *patches*. Em trabalhos de classificação de fácies sísmicas (SFC) multiclasse, a semelhança na distribuição das classes nos conjuntos de treinamento e teste é um critério importante. Na Figura 6.2, é possível observar que o conjunto de teste utilizado atende a esse critério, apresentando uma distribuição de classes semelhante à do conjunto de treinamento.

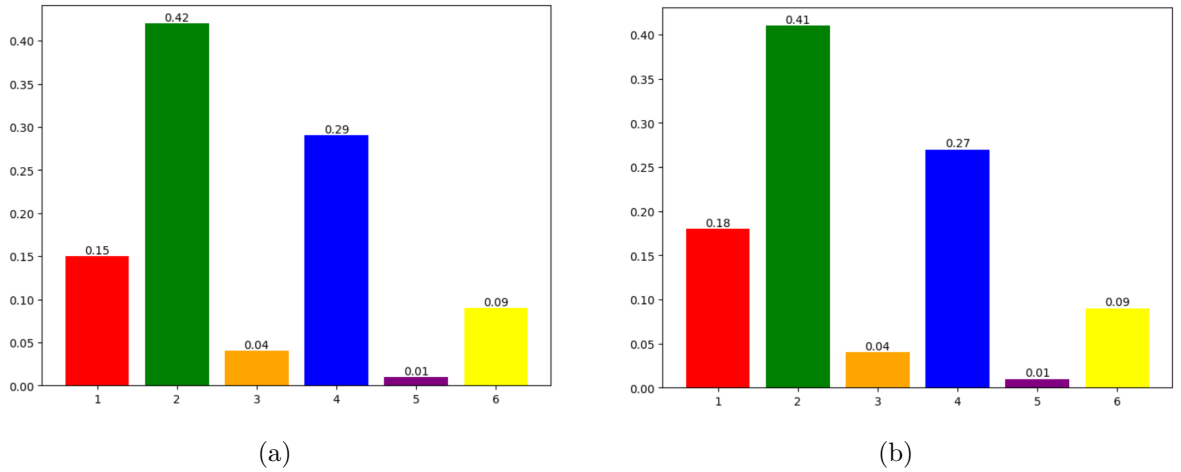


Figura 6.2: (a) Distribuição das amostras em cada fácies sísmica no conjunto de treinamento. (b) Distribuição das amostras em cada fácies sísmica no conjunto de teste.

As Figuras 6.3 a 6.5 apresentam exemplos de seções sísmicas, com suas respectivas anotações de fácies (*ground truth*) e as predições de fácies obtidas pela CNN treinada por 50 épocas. Os exemplos foram escolhidos de forma aleatória e sem perda de generalização. É importante destacar que essas seções sísmicas não foram apresentadas ao modelo durante a etapa de treinamento. A continuidade das fácies preditas ao longo do conjunto de

teste demonstra que o modelo DL utilizado produz resultados consistentes com a interpretação dos especialistas, tanto na direção das *inlines* quanto na das *crosslines*. As formas, localizações e limites das classes são claramente identificadas e a continuidade espacial e temporal das fácies preditas estão em geral de acordo com os processos de sedimentação geológica conhecidos. A arquitetura avaliada é capaz de classificar fácies menos presentes no volume sísmico, como vales incisos e sistemas de canyon submarinos.

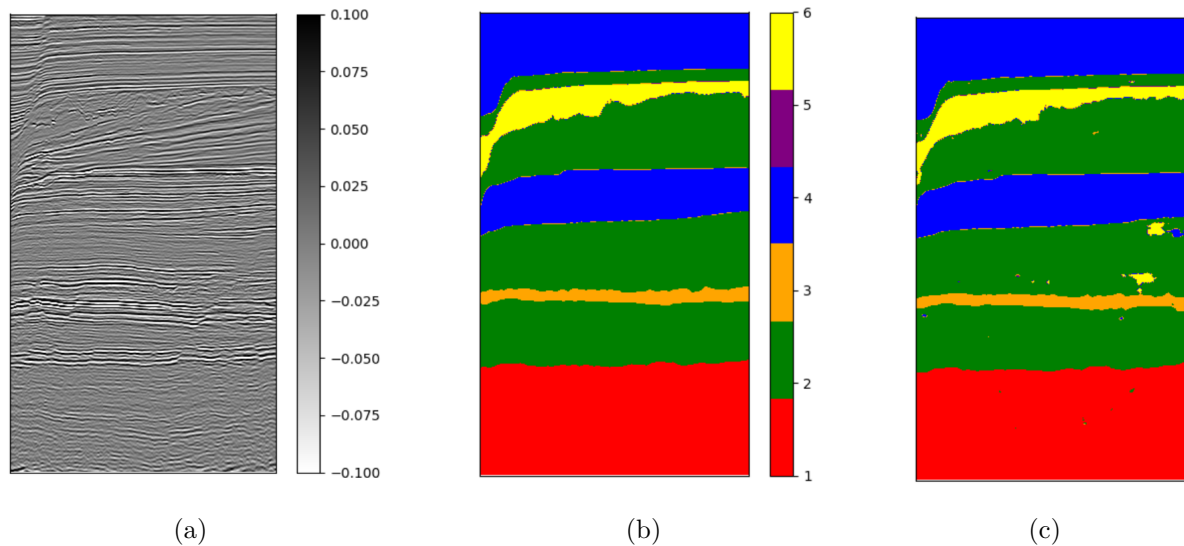


Figura 6.3: Imagem 1 (inline). (a) Seção sísmica. (b) Anotações de Fácies da SEG. (c) Fácies preditas utilizando o modelo treinado por 50 épocas.

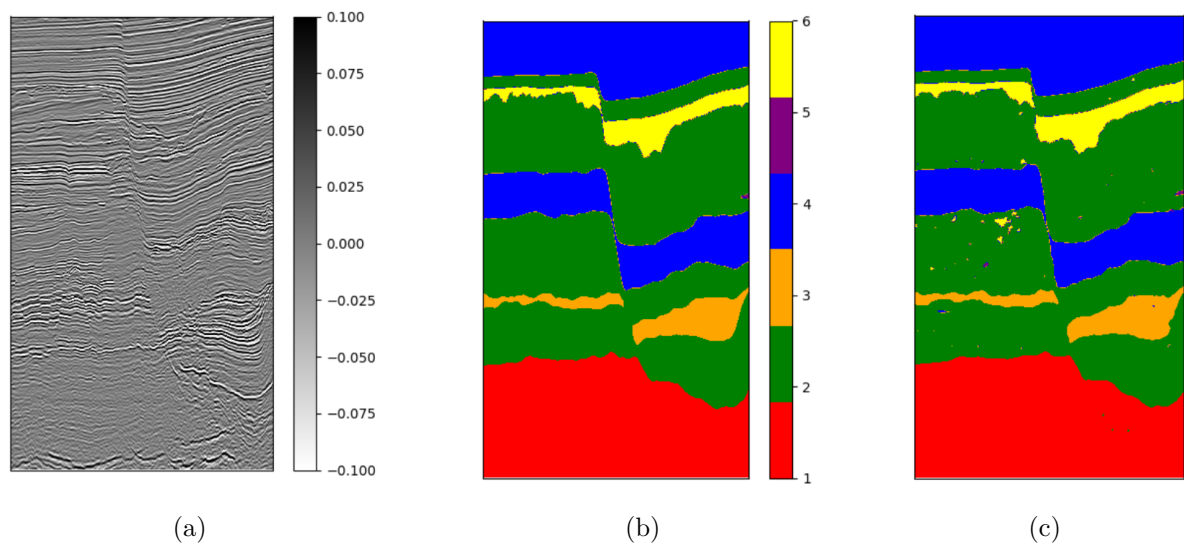


Figura 6.4: Imagem 206 (crossline). (a) Seção sísmica. (b) Anotações de Fácies da SEG. (c) Fácies preditas utilizando o modelo treinado por 50 épocas.

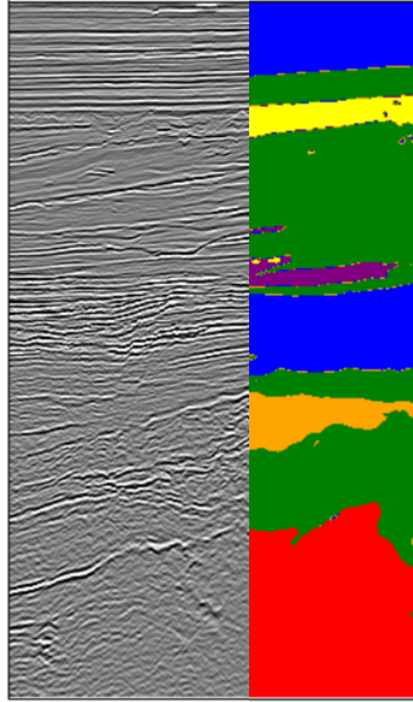


Figura 6.5: Imagem 117 (inline). Seção sísmica e fácies preditas mostrando que os resultados do modelo apresentam boa continuidade e adequação ao dado.

Os resultados obtidos foram encorajadores também do ponto de vista quantitativo, como pode ser observado na matriz de confusão apresentada na Figura 6.6 e nas métricas resumidas na Tabela 6.1. Para avaliar de forma abrangente a performance do modelo, foram selecionadas as métricas *IOU* e F1-escore. Os valores elevados nessas métricas indicam um bom nível de precisão, consistência espacial nas predições e equilíbrio nos erros de sobre-segmentação e subsegmentação. Observou-se que as fácies 3 e 5 apresentaram resultados piores do que as demais, fato que pode ser notado na matriz de confusão. Essas são as duas fácies menos representadas no conjunto de dados (conforme Figura 6.2).

Tabela 6.1: Resultado das métricas de avaliação após o treinamento no conjunto completo de treinamento por 50 e 15 épocas.

Métrica	50 épocas	15 épocas
F1-escore	0.896	0.835
F1-score (ponderado)	0.945	0.881
IOU	0.823	0.745
IOU (ponderado)	0.904	0.858

Na Figura 6.7 é mostrado um exemplo de predição em uma seção com presença significativa da fácies 5, que é a mais rara do conjunto de dados. É possível observar a

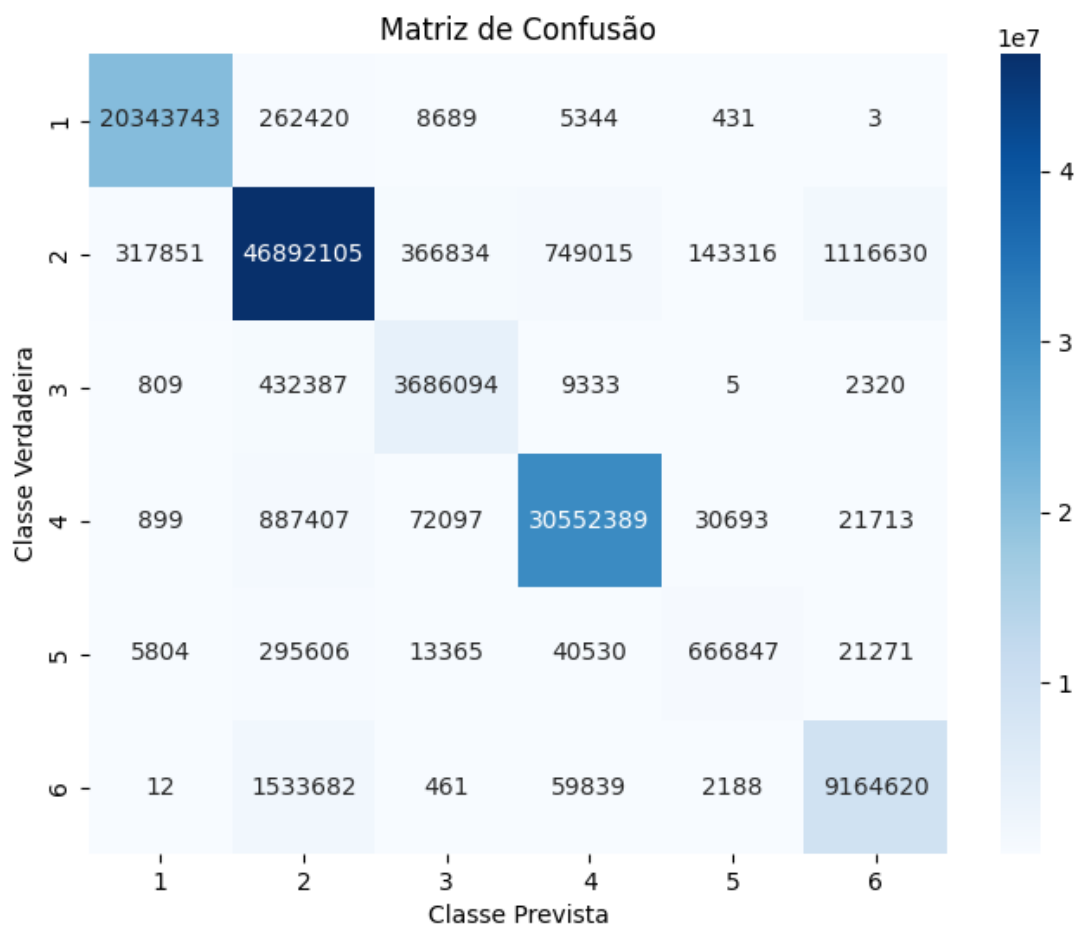


Figura 6.6: Matriz de confusão calculada sobre o conjunto de teste. O modelo utilizado para a predição foi treinado por 50 épocas.

geometria mais complexa dessa classe em relação as demais, o que dificulta sua predição. O mesmo ocorre para a fácies 3. No entanto, mesmo com essas dificuldades, os resultados alcançados foram encorajadores e demonstram a eficácia da CNN utilizada na tarefa de SFC.

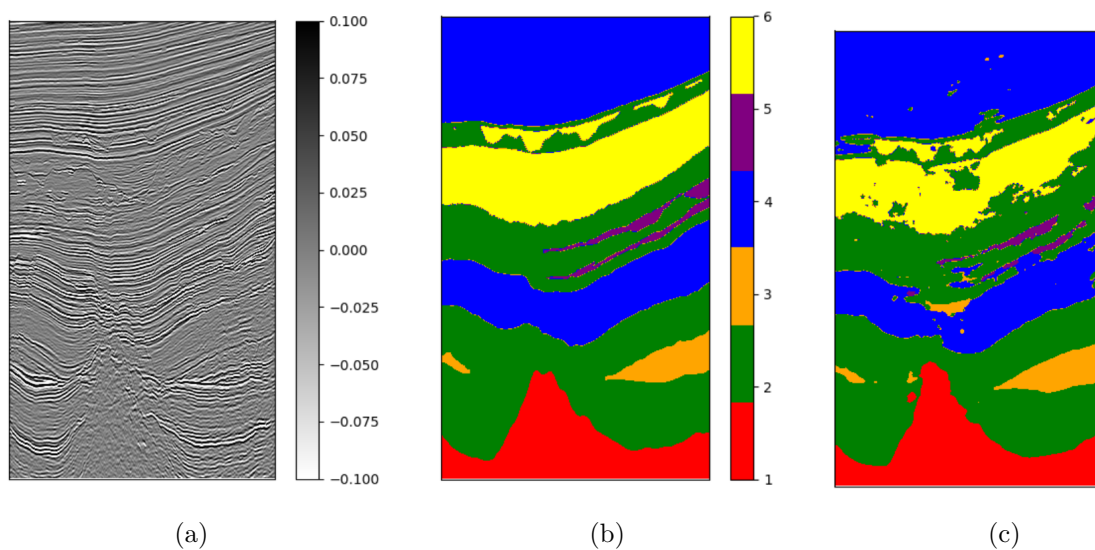


Figura 6.7: Imagem 101 (crossline). (a) Seção sísmica. (b) Anotações de Fácies da SEG. (c) Fácies preditas utilizando o modelo treinado por 50 épocas.

Na Tabela 6.1, são apresentados os valores da média simples e ponderada das métricas escolhidas. Como o conjunto de dados é desbalanceado, e o modelo tem pior desempenho para as fácies menos frequentes, os resultados ponderados das métricas são sempre maiores. A partir de agora, os valores ponderados das métricas serão sempre utilizados nos resultados apresentados, o que é uma escolha usual para dados desbalanceados e atende ao objetivo principal de avaliar a eficiência global dos modelos dentro de um protocolo de aprendizado ativo (AL).

Na Figura 6.8, observa-se uma variação no desempenho da CNN ao longo das imagens do conjunto de teste, o que é esperado, uma vez que a complexidade geológica e, conseqüentemente, as dificuldades para o imageamento sísmico, variam no espaço e no tempo. As Figuras 6.9 e 6.10 apresentam as imagens com os melhores e o piores resultados de segmentação respectivamente. Observa-se que o contexto geológico difere bastante entre elas: no melhor caso, tem-se uma grande falha normal e praticamente uma ausência de vales incisos, enquanto no pior caso há variação do mergulho das camadas mais rasas para as mais profundas, acompanhado de um sistema de vales incisos.

Conforme mencionado na Seção 6.1, para viabilizar a realização de dezenas de treinamentos durante a simulação de AL, foi adotada a estratégia de interromper o treinamento em 15 épocas. No entanto, essa escolha resultou em uma diminuição nos valores das métricas, conforme apresentado na Tabela 6.1, bem como em uma redução na qualidade nas predições, como ilustrado na Figura 6.11. Apesar da queda na qualidade das predições, é possível notar que as principais características e feições geológicas foram preservadas, o que sustentou a decisão de interromper o treinamento em 15 épocas. Além

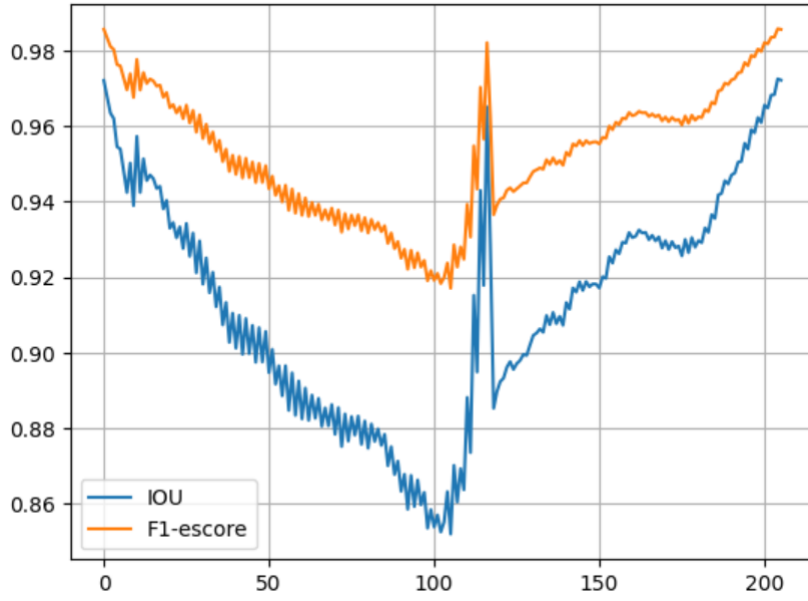


Figura 6.8: Evolução do F1-escore e IOU em relação às imagens do conjunto de teste. As imagens $[1,14] \cup [105,118]$ foram extraídas de inlines. As demais de crosslines.

disso, é improvável que existam fácies dentro de outras. Portanto, para mitigar a perda de qualidade na predição dos modelos treinados durante 15 épocas é possível aplicar um pós processamento que filtre a imagem, modificando a classe das regiões de pequenas dimensões que diferem da classe de fundo. Para o conjunto de dados utilizado, não seria recomendado a aplicação desse pós processamento nas classes três, cinco e seis, que representam formações geológicas de natureza canalizada ou intrusiva.

6.3 Aprendizado ativo

A Figura 6.12 exibe as curvas de desempenho dos três métodos de amostragem no conjunto de teste. Todos os modelos possuem um valor fixo de 15 épocas por iteração. São apresentados os valores médios do F1-escore e os intervalos de confiança *bootstrap* de 95%, calculados após cada iteração de AL sobre o conjunto de teste. A curva representada na cor azul corresponde à estratégia de amostragem aleatória, na qual o intérprete seleciona e anota um conjunto de seções escolhidas aleatoriamente para serem adicionadas ao conjunto de dados anotados que serão treinados no próximo ciclo. Essa estratégia extremamente simples serve de referência para os experimentos e, como esperado, o F1-escore melhora à medida que mais dados são adicionados dessa forma. Essa curva é comparada com as curvas resultantes da amostragem estratificada em passos e da amostragem baseada na maior entropia *softmax* (Equação 4.2). Embora o F1-escore seja a métrica utilizada para apresentar os resultados, é importante ressaltar que o formato das cur-

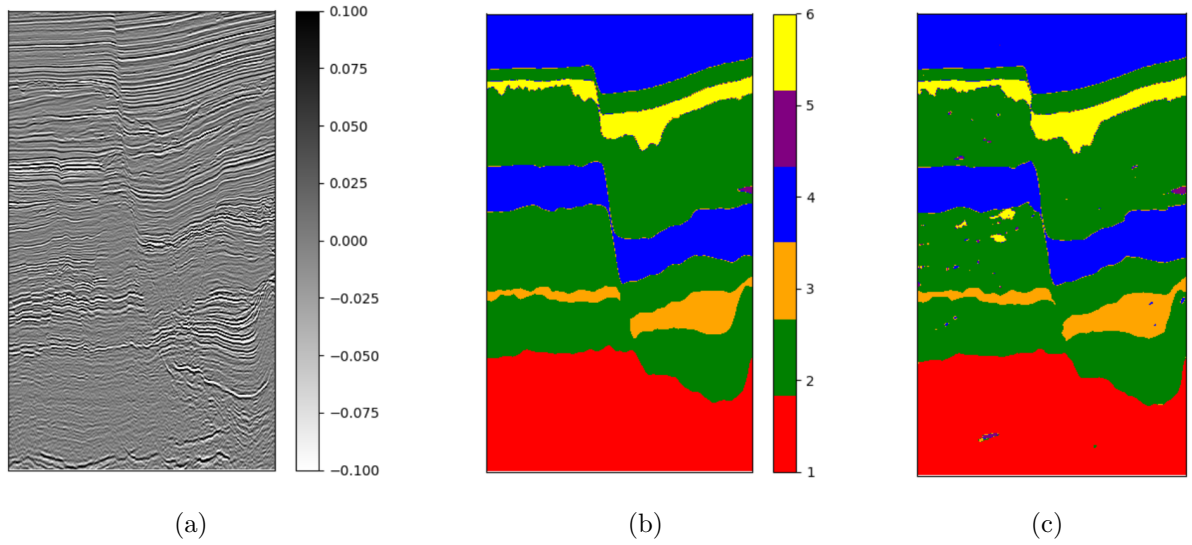


Figura 6.9: Imagem 205 (crossline) - melhor resultado obtido nas métricas dentro do conjunto de teste. (a) Seção sísmica. (b) Anotações de Fácies da SEG. (c) Fácies preditas utilizando o modelo treinado por 50 épocas.

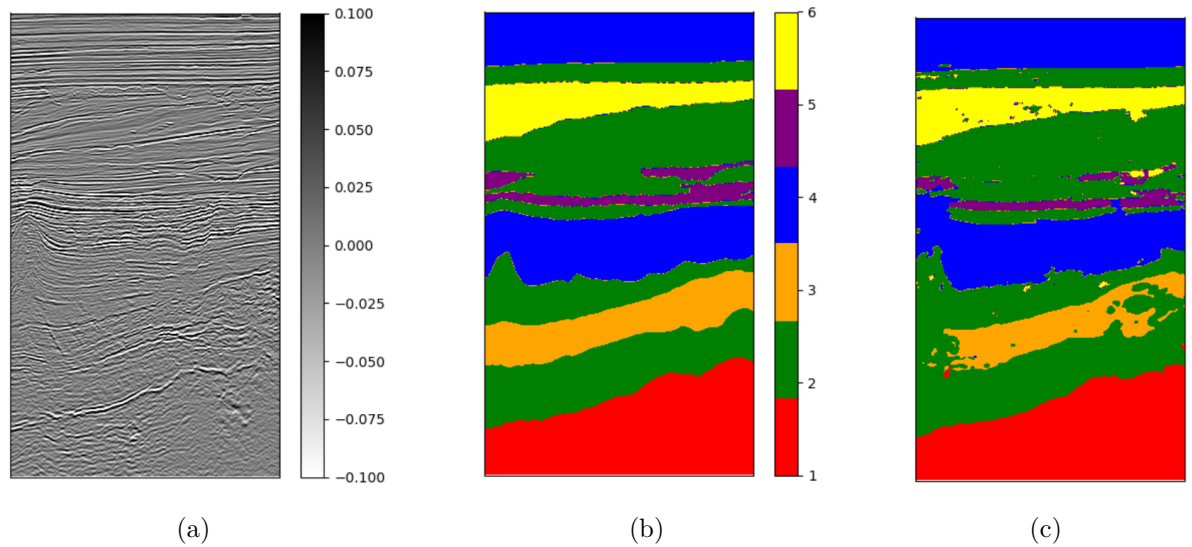


Figura 6.10: Imagem 106 (inline) - pior resultado obtido nas métricas dentro do conjunto de teste. (a) Seção sísmica. (b) Anotações de Fácies da SEG. (c) Fácies preditas utilizando o modelo treinado por 50 épocas.

vas seria semelhante se a métrica *IOU* fosse adotada, dado que essas duas métricas são positivamente correlacionadas.

Como mencionado na Seção 5.6, não há um critério bem definido para escolher o tamanho e o número de amostras *bootstrap*. A escolha adotada baseou-se principalmente na limitação de tempo disponível, uma vez que o processo de inferir as fácies sísmicas no conjunto de teste e calcular a métrica F1-escore ponderada sobre todo o conjunto levou

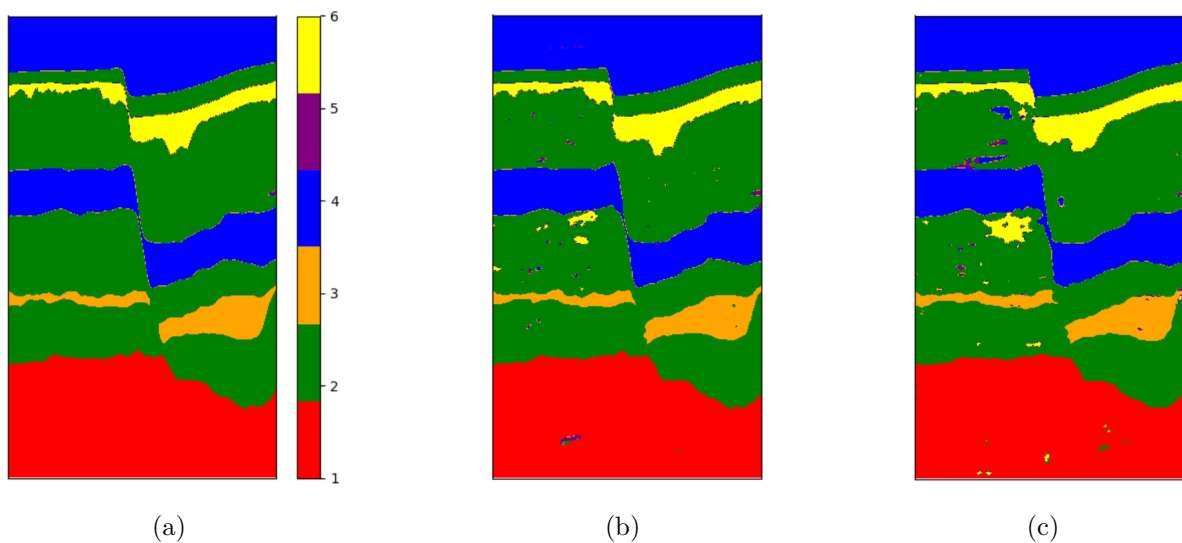


Figura 6.11: Imagem 206 (crossline). (a) Anotações de Fácies da SEG. (b) Fácies preditas utilizando o modelo treinado por 50 épocas. (c) Fácies preditas utilizando o modelo treinado por 15 épocas.

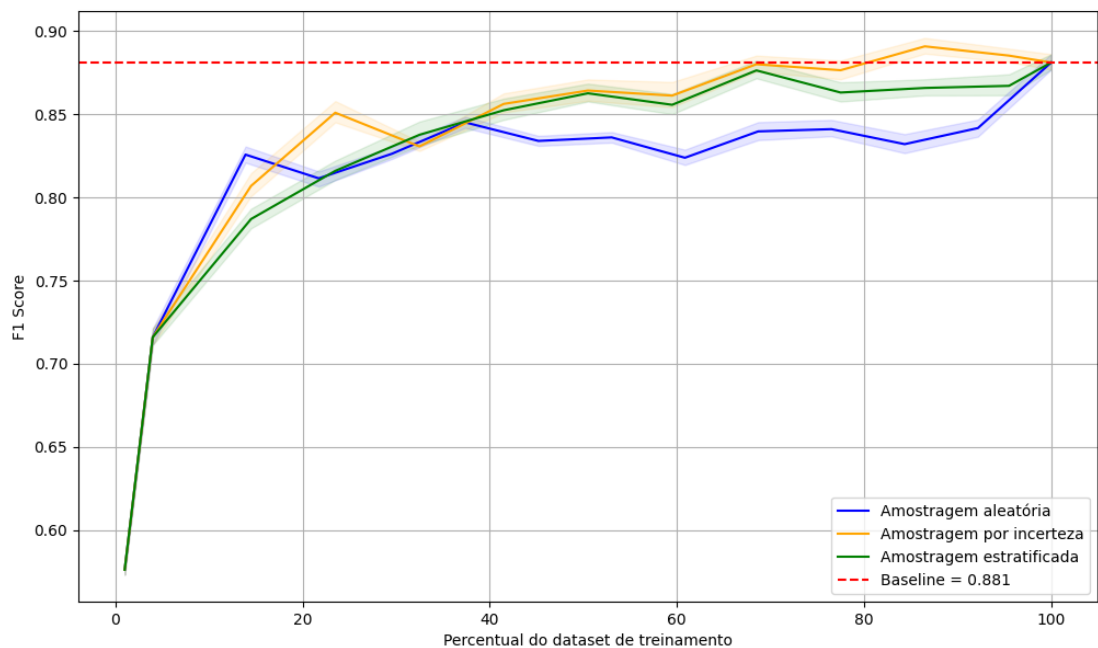


Figura 6.12: Curvas de desempenho da métrica F1-escore em relação ao percentual de dados anotados no treinamento, empregando diferentes métodos de amostragem: aleatória (azul); por incerteza (laranja); e estratificada em passos (verde). Para cada modelo é apresentado o valor médio e o intervalo de confiança de 95%, calculados a partir das amostras bootstrap.

aproximadamente dois minutos e 40 segundos. Esse tempo seria gasto para cada amostra *bootstrap*, tornando inviável a utilização do valor usual de 10000 amostras *bootstrap*, comum em simulações na área de estatística. Para verificar a variação da amplitude do intervalo de confiança com o tamanho e número de amostras *bootstrap*, realizou-se um teste com o modelo treinado com 50% dos dados de treinamento anotados amostrados por incerteza. Optou-se por usar amostras com 100 imagens (aproximadamente 50% do conjunto de teste) e gerar 80 amostras *bootstrap*. A Figura 6.13 apresenta os resultados desse teste, sugerindo que a escolha adotada é adequada para a aplicação da pesquisa.

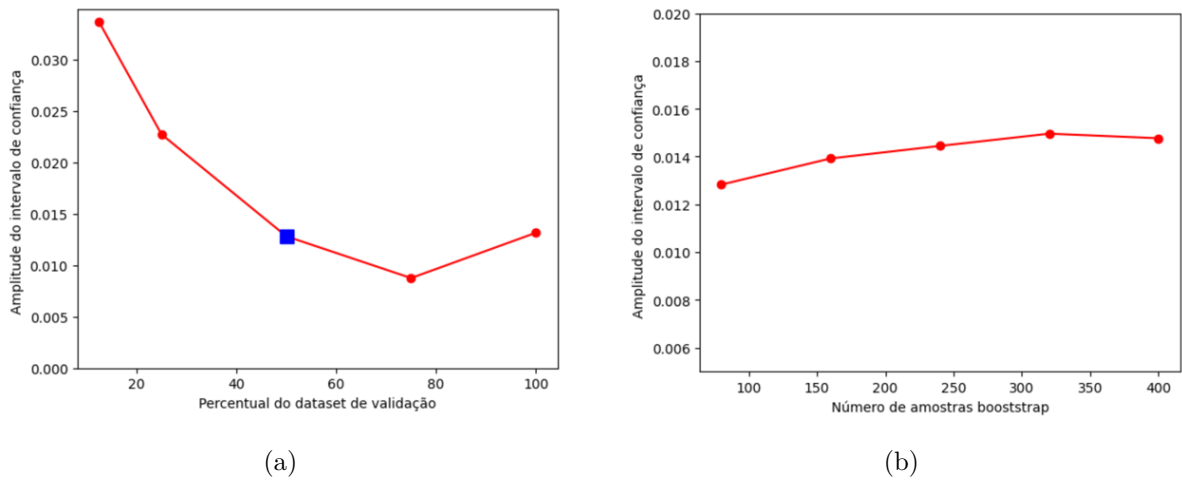


Figura 6.13: (a) Variação da amplitude do intervalo de confiança de 95% com o tamanho da amostra *bootstrap* em relação ao conjunto de teste original. O ponto em destaque representa a escolha feita: 100 imagens (aproximadamente 50% do número de imagens do conjunto de teste). (b) Variação da amplitude do intervalo de confiança de 95% com o número de amostras *bootstrap*. O gráfico mostra que não ocorre uma variação significativa do tamanho do intervalo com o aumento no número de amostras.

Conforme ilustrado na Figura 6.12, observa-se que nenhuma estratégia de amostragem conseguiu superar a amostragem aleatória por uma larga margem. O valor médio do F1-escore obtido pela amostragem por incerteza demonstrou um desempenho ligeiramente superior em relação à amostragem estratificada por passos, a qual, por sua vez, superou de modo geral a amostragem aleatória. As curvas mantiveram-se em proximidade até que aproximadamente 40% das imagens de $\mathcal{D}_{oráculo}$ foram adicionadas no treinamento, momento no qual a amostragem aleatória foi superada pelas outras estratégias. Superar a amostragem aleatória nos primeiros estágios é importante, uma vez que o intérprete deseja ter um modelo aceitável com o mínimo possível de anotações. Todas as curvas do gráfico sugerem que resultados próximos do valor de referência poderiam ser alcançados utilizando-se menos da metade do conjunto de treinamento anotado. Esse fato provavelmente decorre do considerável grau de redundância presente entre seções sísmicas adjacentes.

Ao analisar os gráficos em maior detalhe (Figura 6.14), observa-se que, para alcançar um valor médio de F1-escore de 0,863 (98% do valor de referência: 0,881), a amostragem por incerteza requer 50% das imagens de $\mathcal{D}_{\text{oráculo}}$, enquanto a amostragem estratificada e aleatória demandam 65% e 95% dos dados respectivamente, para atingir o mesmo patamar. No entanto, ao examinar os intervalos de confiança, observa-se que a amostragem por incerteza supera estatisticamente a amostragem estratificada somente quando os modelos já incorporaram cerca de 80% dos dados anotados. Essa constatação indica que a amostragem estratificada em passos, apesar de sua simplicidade, pode competir com estratégias mais complexas e computacionalmente intensivas, como as orientadas pelo modelo. Esse resultado está em consonância com a discussão apresentada por Burneister et al. (2022), em que os autores destacam os bons resultados obtidos pela amostragem estratificada em passos aplicada em um conjunto de imagens médicas 2D extraídas de uma imagem 3D, ressaltando a capacidade dessa estratégia de evitar a seleção de imagens vizinhas, possivelmente redundantes, e sugerindo que essa deve ser a estratégia usada como referência a ser superada em trabalhos de AL aplicados nesse tipo de dado. Essas observações aparentemente se aplicam igualmente a volumes sísmicos 3D.

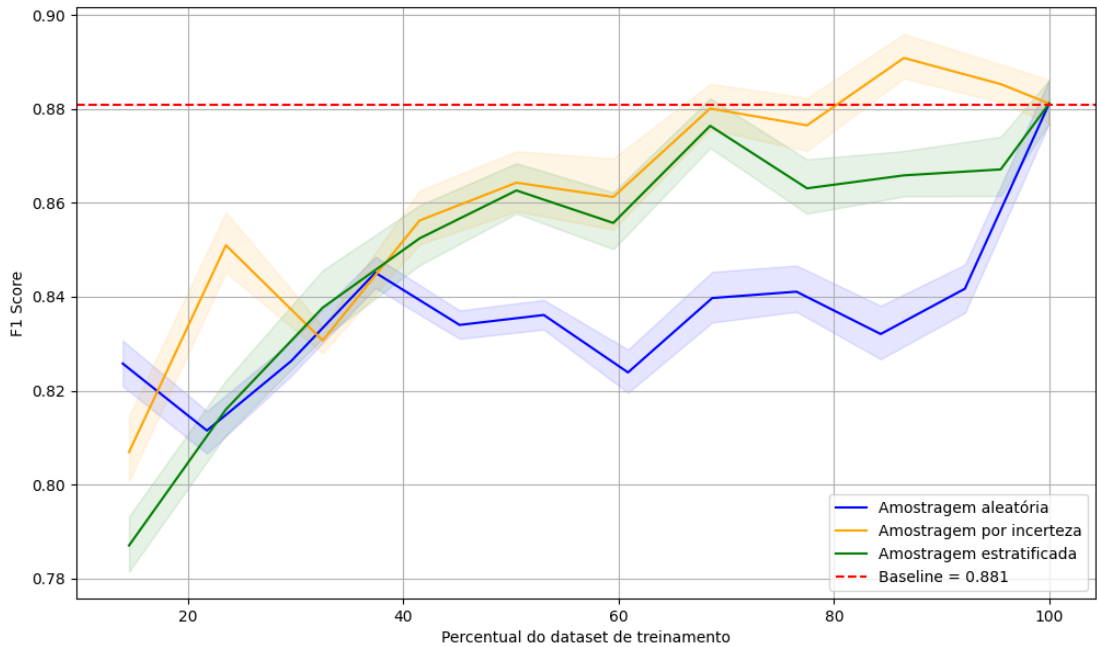


Figura 6.14: Detalhe dos gráficos da Figura 6.12, ilustrando o desempenho do F1-escore de acordo com o percentual de dados anotados no treinamento para três diferentes métodos de amostragem: aleatória (azul); por incerteza (laranja); e estratificada em passos (verde).

Em última análise, o método de amostragem por incerteza demonstrou os melhores

resultados, apresentando médias mais elevadas de F1-escore e intervalos de confiança mais estreitos. Esses resultados indicam uma maior estabilidade diante das variações no conjunto de teste, o que reflete uma maior confiabilidade e capacidade de generalização. É interessante notar que a amostragem por incerteza conseguiu superar o valor da métrica de referência antes mesmo de atingir o ponto em que $\mathcal{D}_{treino} = \mathcal{D}_{oráculo}$, sugerindo que algumas imagens podem não fornecer informações relevantes para o conjunto de teste em questão.

Uma vantagem do método de amostragem por incerteza em relação aos outros dois métodos testados é sua capacidade de guiar a seleção de novas imagens para a iteração seguinte com base no próprio modelo. Conforme o modelo melhora sua capacidade de classificar os dados, ele naturalmente aprimora sua habilidade em identificar exemplos não anotados que merecem ser adicionados ao conjunto de treinamento. No entanto, é importante salientar que a decisão de selecionar as imagens com base na incerteza do modelo pressupõe idealisticamente que todas as amostras contêm informações independentes e igualmente relevantes, o que dificilmente é o caso para seções 2D extraídas de um volume sísmico 3D. Esse aspecto pode potencialmente levar à seleção de imagens não representativas ou redundantes, fato que pode ser explorado em uma pesquisa futura, para alcançar resultados ainda mais promissores.

Uma avaliação visual dos resultados obtidos pelos diferentes modelos de AL é ilustrada na Figura 6.15. O distanciamento entre a curva de F1-escore da amostragem aleatória e as demais curvas, visto na Figura 6.14, é refletido nas seções das fácies previstas pela perceptível qualidade inferior da predição da amostragem aleatória em comparação às demais estratégias. Por outro lado, a diferença entre as estratégias de amostragem por incerteza e a amostragem estratificada em passos é relativamente pequena, tanto em termos das métricas avaliadas, quanto na análise visual das seções previstas.

Na Figura 6.16, a evolução qualitativa dos modelos de AL após cada iteração é exemplificada pela visualização das predições do modelo amostrado por incerteza na Imagem 206. Nos estágios iniciais, com apenas 1% das imagens anotadas de $\mathcal{D}_{oráculo}$, a predição não consegue definir adequadamente nenhuma das fácies sísmicas, apresentando um resultado pouco coerente do ponto de vista geológico. Um salto significativo em termos de qualidade é observado com apenas 4% das imagens anotadas, porém o modelo ainda não é capaz de identificar adequadamente a falha geológica presente na seção nem delimitar corretamente as fácies mais complexas. À medida que mais dados são anotados, especificamente com 25% das imagens anotadas, o modelo já demonstra uma capacidade considerável de predição, apresentando uma qualidade satisfatória na delimitação das fácies. A partir das iterações subsequentes, as melhorias na qualidade das predições tornam-se mais sutis, e em alguns casos, há inclusive deterioração na qualidade em algu-

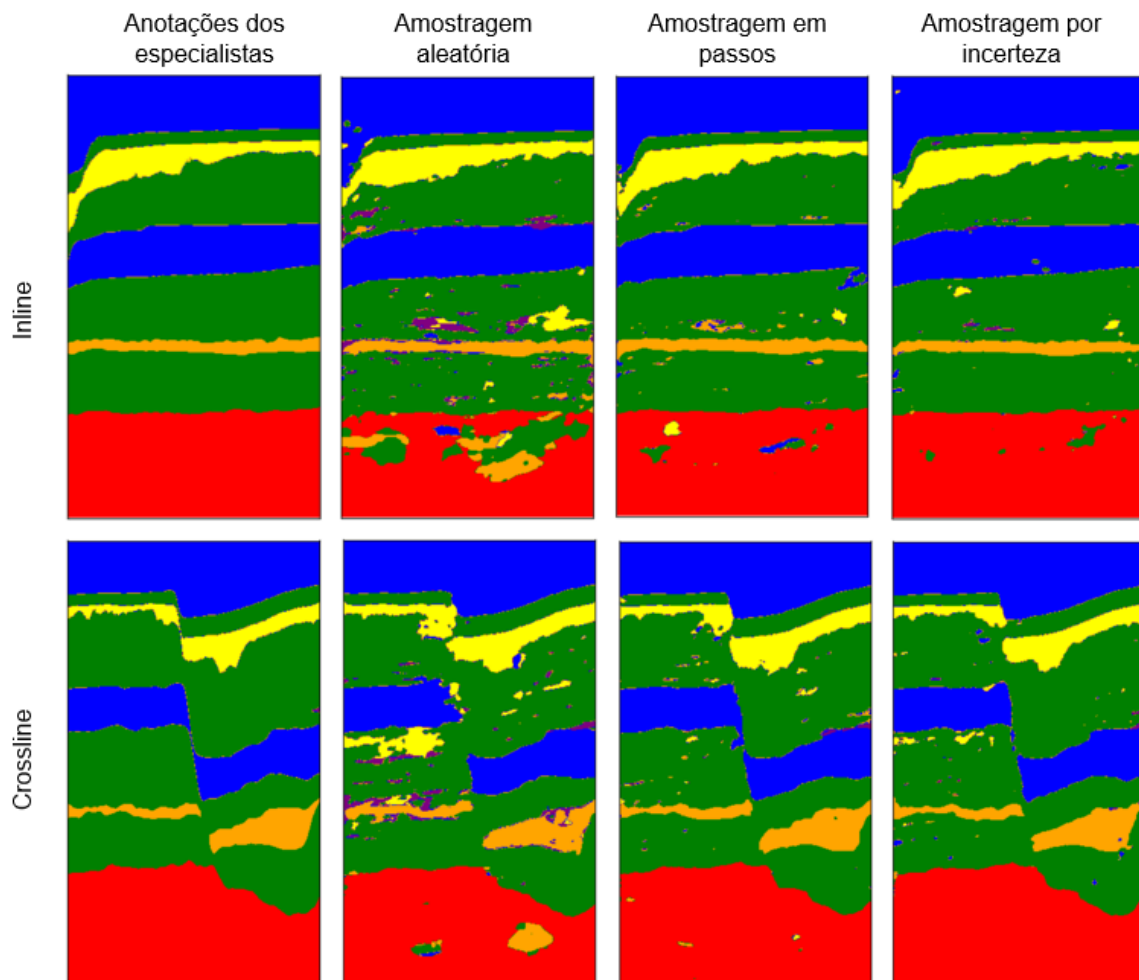


Figura 6.15: Predições obtidas pelos modelos treinados após receberem cerca de 60% das imagens anotadas de $\mathcal{D}_{\text{oráculo}}$. A primeira linha apresenta as predições para a Imagem 1 (inline), enquanto a segunda linha faz o mesmo para a Imagem 206 (crossline). Na primeira coluna, são exibidas as fácies anotadas pelos especialistas da SEG (ground truth), e nas colunas subsequentes são apresentadas as predições para cada uma das estratégias de amostragem utilizadas.

mas regiões. Essa observação pode ser constatada ao realizar uma comparação cuidadosa na Figura 6.16 entre os resultados da predição com 25% e 50% das imagens anotadas.

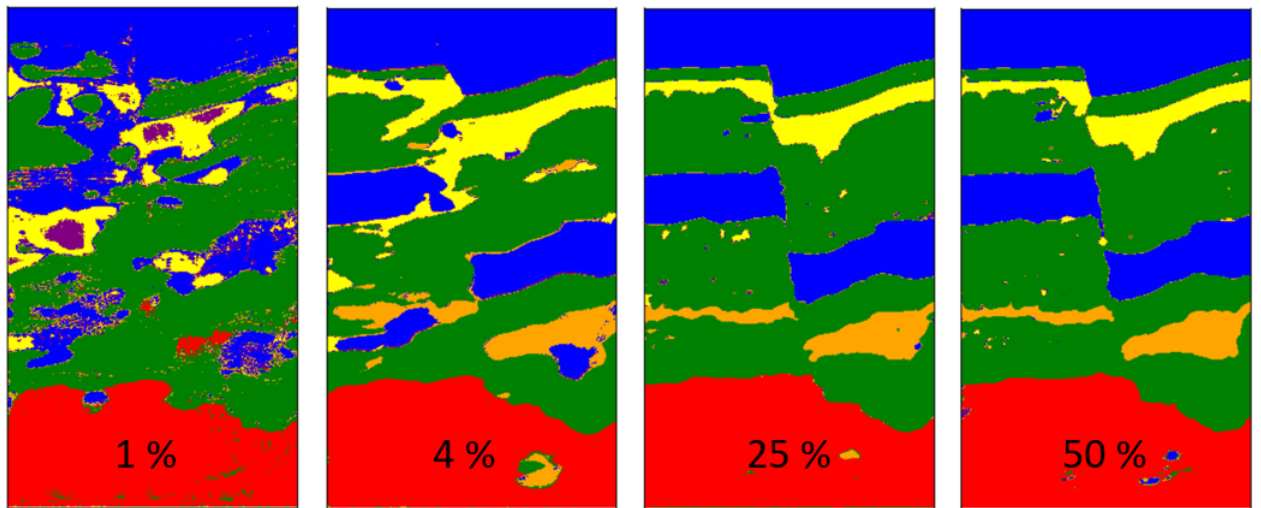


Figura 6.16: Imagem 206 (crossline). Visualização dos resultados da predição do modelo que utiliza a estratégia de amostragem por incerteza a medida que mais imagens anotadas de $\mathcal{D}_{\text{oráculo}}$ são acrescentadas no treinamento.

Capítulo 7

Conclusão

A proposta deste trabalho foi estabelecer uma metodologia para avaliar o impacto das técnicas de aprendizado ativo na tarefa de classificação de fácies sísmicas assistida por redes neurais profundas. O volume de dados sísmicos tridimensionais Parihaka, acompanhado de suas fácies sísmicas anotadas por especialistas foi utilizado para testar a técnica.

Os resultados confirmaram que redes neurais profundas construídas para a tarefa de segmentação de imagens são capazes de classificar fácies sísmicas, o que é muito útil no mapeamento de sequências sedimentares, elementos estruturais e cálculos de volumes de rocha, entre outras aplicações na área de interpretação sísmica. No entanto, esses resultados positivos são dependentes de um conjunto de treinamento com uma quantidade suficientemente grande de dados anotados.

As simulações de aprendizado ativo demonstraram que a aplicação de estratégias simples de amostragem aceleram a obtenção de bons resultados na classificação de fácies sísmicas. A estratégia de amostragem por incerteza baseada na entropia obteve resultados comparáveis com os de um modelo treinado com o conjunto completo de dados, usando apenas 50% das imagens anotadas. A técnica de amostragem por representatividade em passos também apresentou resultados próximos à amostragem por incerteza e, por tratar-se de uma técnica simples e independente do modelo, sugere-se que seja utilizada como estratégia de amostragem de referência a ser superada por novas metodologias de aprendizado ativo propostas.

Outra contribuição importante deste trabalho foi a avaliação da performance das técnicas de aprendizado ativo por meio de um método de estimativa do intervalo de confiança. Ao utilizar a geração de amostras *bootstrap* no conjunto de teste, foi possível viabilizar a construção do intervalo, contornando as limitações de técnicas como teste cruzada ou repetições do treinamento com *dropout* aleatório, quando aplicadas em conjuntos de dados extensos.

A estratégia proposta tem o potencial de reduzir significativamente o esforço dos

intérpretes de dados sísmicos, ao mesmo tempo em que assegura níveis de precisão na interpretação. Este trabalho estabelece um fundamento para futuras pesquisas, que podem explorar diferentes abordagens, como o uso de técnicas de aprendizado não supervisionado na etapa de amostragem, consideração da dependência entre seções sísmicas, utilização de atributos sísmicos adicionais, avaliação da manutenção dos pesos treinados entre iterações (*warm-starting*), utilização de técnicas de pós processamento capazes de melhorar os resultados das predições com um baixo custo computacional, aplicação em diferentes tipos de dados sísmicos e contextos geológicos, bem como a utilização de outras arquiteturas de redes neurais. O aprendizado ativo não precisa se restringir apenas a área de classificação de fácies sísmicas, podendo ser aplicado em diversas tarefas de interpretação e processamento sísmico.

Referências Bibliográficas

- A. Abdel. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging*, 15(29), 2015.
- Y. Alaudah, P. Michalowicz, M. Alfarraj, and G. AlRegib. A machine learning benchmark for facies classification. *Interpretation*, 7(3), 2019.
- G. Alves. Feasibility studies and elastic inversion applied to geophysical characterization of the offshore taranaki basin, new zealand. 2020.
- V. Badrinarayanan, R. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2017.
- S. Belharbi, I. Ayed, L. McCaffrey, and E. Granger. Deep active learning for joint classification & segmentation with weak annotator. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2021.
- K. Bjorlykke. *Petroleum Geoscience*. Springer, 2015.
- C. Bond, R. Lunn, Z. Shipton, and A. Lunn. What makes an expert effective at interpreting seismic images? *Geology*, 40(1), 2012.
- J. Burmeister, M. Rosas, J. Hagemann, J. Kordt, J. Blum, S. Shabo, B. Bergner, and C. Lippert. Less is more: a comparison of active learning strategies for 3d medical image segmentation. In *International Conference on Machine learning*, 2022.
- J. Cartwright and M. Huuse. 3d seismic technology: the geological ‘hubble’. In *Basin Research*, volume 17(1), pages 1–20. doi: 10.1111/J.1365-2117.2005.00252.X, 2005.
- X. Chai, G. Tang, S. Wang, R. Peng, W. Chen, and J. Li. Deep learning for regularly missing data reconstruction. *IEEE Transactions on Geoscience and Remote Sensing*, 58(3), 2020.
- X. Chai, G. Tang, K. Lin, , Z. Yan, H. Gu, R. Peng, X. Sun, and W. Cao. Deep learning for mutitrace sparse-spike deconvolution. *Geophysics*, 86(3), 2021.

- X. Chai, W. Nie, K. Lin, G. Tang, T. Yang, J. Yu, and W. Cao. An open-source package for deep-learning-based seismic facies classification: Benchmarking experiments on the seg 2020 open data. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 2022.
- C. Chen, L. Chen, Y. Xu, J. Zhang, and H. Wang. A review on deep learning techniques for image segmentation. *Neurocomputing*, 268, 2017.
- B. C. Csáji et al. Approximation with artificial neural networks. *Faculty of Sciences, Eötvös Loránd University, Hungary*, 24(48):7, 2001.
- A. Davison and D. Hinckley. *Bootstrap methods and their application (Vol.1)*. Cambridge university press, 1997.
- H. Di, D. Gao, and G. AlRegib. Developing a seismic texture analysis neural network for machine-aided seismic pattern recognition and classification. *Geophysical Journal International*, (218), 2019.
- H. Di, Z. Li, H. Maniar, and A. Abubakar. Seismic stratigraphy interpretation by deep convolutional neural networks: A semi-supervised workflow. 2020.
- J. Dramsch and M. Luthje. Deep learning seismic facies on state-of-the-art cnn architectures. In *SEG International Exposition and 88th Annual Meeting*, page 2036. SEG, 2018.
- B. Efron. Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82, 1987.
- T. Falk, D. Mai, R. Bensch, O. Çiçek, A. Abdulkadir, Y. Marrakchi, A. Böhm, J. Deubner, Z. Jackel, K. Seiwald, A. Dovzhenko, O. Tietz, C. Dal Bosco, S. Walsh, D. Saltukoglu, T. Tay, M. Prinz, K. Palme, M. Simons, I. Diester, T. Brox, and O. Ronneberger. U-net: deep learning for cell counting, detection, and morphometry. *Nature Methods*, 16(4), 2018.
- X. Feng, Y. Jiang, X. Yang, M. Du, and X. Li. Computer vision algorithms and hardware implementations: A survey. 2019.
- A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez. A review on deep learning techniques applied to semantic segmentation. *arXiv:1704.06857*: <https://doi.org/10.48550/arXiv.1704.06857>, 2017.
- I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. MIT press, 2016.
- X. Guo, Y. Liu, and J. Zhu. Deep learning for pattern recognition: A review. *Neurocomputing*, 187, 2016.

- K. He, X. Zhang, S. Ren, and J. Sun. Identity mappings in deep residual networks. In *Computer Vision (ECCV)*, pages 630–645. Springer, 2016.
- S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv:1502.03167*, 2015.
- M. Islam. Using deep learning based methods to classify salt bodies in seismic images. 2020.
- S. Jadon. A survey of loss functions for semantic segmentation. In *IEEE Conf. Comput. Intell. Bioinf. Comput. Biol. (CIBCB)*, 2020.
- Y. Jiang. Detecting geological structures in seismic volumes using deep convolutional neural networks. 2017.
- M. Jordan and T. Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 2015.
- D. Kingma and J. Ba. Adam: A method for stochastic optimization. In *International conference on learning representations*, 2015.
- A. Krizhevsky, I. Sutskever, and Hinton. Imagenet classification with deep convolutional neural networks. In *Neuro information processing systems*, pages 1097–1105, 2012.
- P. Kumar. Application of geometric attributes for interpreting faults from seismic data: An example from taranaki basin, new zealand. pages 2077–2081, 09 2016. doi: 10.1190/segam2016-13944184.1.
- Y. Kweon and J. Kim. Active learning for deep neural networks with expected model output changes. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- Y. LeCun and Y. Bengio. Convolutional networks for images, speech, and time series. *Handb. Brain Theory Neural Network*, 1995.
- Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. In *IEEE*, 1998.
- F. Li, H. Zhou, Z. Wang, and X. Wu. Addcnn: An attetion-based deep dilated convolutional neural network for seismic facies analysis with interpretable spatial-spectral maps. *IEEE Transactions on Geoscience Remote Sense*, 59, 2021.
- H. Li, G. Xu, C. Studer, and T. Goldstein. Visualizing the loss landscape of neural nets. *arXiv:1712.09913*, 2017.

- L. Li, H. Qi, J. Dai, X. Ji, and Y. Wei. Deep learning for object detection: A comprehensive review. *Neurocomputing*, 396, 2020.
- M. Liu, W. Li, M. Jervis, and P. Nivlet. 3d seismic facies classification using convolutional neural network and semi-supervised generative adversarial network. In *SEG International Exposition and Annual Meeting*, 2019.
- X. Liu, X. Chen, J. Li, X. Zhou, and Y. Chen. Facies identification based on multikernel relevance vector machine. *IEEE Transactions on Geoscience and Remote Sensing*, 58(10), 2020.
- J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.
- A. Meirelles, T. Kurc, J. Saltz, and G. Teodoro. Effective active learning in digital pathology: A case study in tumor infiltrating lymphocytes. *Computer Methods and Programs in Biomedicine*, 220, 2022.
- A. Milosavljevic. Identification of salt deposits on seismic images using deep learning method for semantic segmentation. *International Journal of Geo-Information*, 9(24), 2019.
- A. Mustafa and G. AlRegib. Active learning with deep autoencoders for seismic facies interpretation. *Geophysics*, 88, 2023.
- V. Nath, D. Yang, B. Landman, D. Xu, and H. Roth. Diminishing uncertainty within the training pool: Active learning for medical image segmentation. *IEEE Transactions on Medical Imaging*, 40(10), 2021.
- B. Peters, J. Granek, and E. Haber. Multi-resolution neural networks for tracking seismic horizons from few training images. In *Interpretation*, volume 7, pages 1A–T725. Society of Exploration Geophysicists/American Association of Petroleum Geologists, 2019.
- A. Phinyomark and E. Scheme. Emg pattern recognition in the era of big data and deep learning. *Big Data Cogn. Comput.*, 21(2(3)), 2018.
- M. Rashid, M. Islam, M. Ahmed, S. Al-Mamum, and H. Wang. Deep learning for image classification: A review. *Neurocomputing*, 338, 2019.
- S. Reddi, S. Kale, and S. Kumar. On the convergence of adam and beyond. In *International conference on learning representations*, 2018.

- O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science*, pages 234–241. Springer, 2015.
- S. Ruder. An overview of gradient descent optimization algorithms. *arXiv*, arXiv:1609.04747, 2017.
- SEG wiki. "seg-y". <https://wiki.seg.org/wiki/SEG-Y>, 2017. Online; acessado em 10 de Novembro 2021.
- C. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 1948.
- D. Sharma, Z. Shanis, C. Reddy, S. Gerber, and A. Enquobahrie. Active learning technique for multimodal brain tumor segmentation using limited labeled images. In *Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data*, pages 148–156. Springer, 2019.
- K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *International conference on learning representations*, 2015.
- A. Smailagic, P. Costa, A. Gaudio, K. Khandelwal, M. Mirshekari, J. Fagert, D. Walawalkar, S. Xu, A. Galdran, P. Zhang, A. Campilho, and H. Noh. O-medal: Online active deep learning for medical image analysis. *WIREs Data Mining Knowledge Discovery*, 10(1353), 2020.
- R. Tibshirani and J. Friedman. *The Elements of Statistical Learning*. Springer, 2009.
- A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis. Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience - Special Issue*, 2018.
- A. Waldeland, A. Jensen, L. Gelius, and A. Solberg. Convolutional neural networks for automated seismic interpretation. *The Leading Edge*, 37, 2018.
- R. T. Walker. Interpreting seismic data with deep learning. *The Leading Edge*, 36, 2017.
- S. Wu, K. Roberts, S. Datta, J. Du, Z. Ji, Y. Si, S. Soni, Q. Wang, Q. Wei, Y. Xiang, B. Zhao, and H. Xu. Deep learning in clinical natural language processing: a methodical review. *Journal of the American Medical Informatics Association*, 27(3), 2019a.
- X. Wu, L. Liang, Y. Shi, and S. Fomel. Faultseg3d: Using synthetic data sets to train an end-to-end convolutional neural network for 3d seismic fault segmentation. *Geophysics*, 2019b.

- P. Xu, W. Lu, and B. Wang. Multi-attribute classification based on sparse autoencoder: A gas chimney detection example. In *79th European Association Geoscientists and Engineers conf. and Exhibition*, 2017.
- T. Xu, S. Niu, L. Lu, and B. Chen. Localized feature selection for visual pattern recognition. *Pattern Recognition Letters*, 124, 2019.
- S. Yan, D. Xu, H. Zhang, and L. Zhang. Deep learning for image interpolation and extrapolation: A review. *Signal Processing: Image Communication*, 88, 2020.
- S. Yu and J. Ma. Deep learning for geophysics: current and future trends. *Reviews of Geophysics*, 59(3), 2021.
- M. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. 2013.
- C. Zhang, C. Frogner, and T. Poggio. Automated geophysical feature detection with deep learning. In *GPU Technology Conference*, 2016.
- H. Zhang, T. Chen, Y. Liu, Y. Zhang, and J. Liu. Automatic seismic facies interpretation using supervised deep learning. *Geophysics*, 86(1), 2021.
- X. Zhang. *A study of deep learning frameworks: TensorFlow and PyTorch*. University of Wisconsin-Madison, 2018.
- Z. Zhang, J. Li, Z. Zhong, Z. Jiao, and X. Gao. A multi-resolution coarse-to-fine segmentation framework with active learning in 3d brain mri. In *Intelligence Science and Big Data Engineering. Visual Data Engineering*, pages 285–298. Springer, 2019a.
- Z. Zhang, J. Li, Z. Zhong, Z. Jiao, and X. Gao. A sparse annotation strategy based on attention-guided active learning for 3d medical image segmentation. <https://arxiv.org/ftp/arxiv/papers/1906/1906.07367.pdf>, 2019b.
- T. Zhao. Seismic facies classification using different deep convolutional neural networks. In *SEG International Exposition and Annual Meeting*, 2018.

Apêndice A

Formulação do F1-escore

A definição do F1-escore, também conhecido como Dice escore, pode ser encontrada amplamente em fontes acadêmicas, como websites, livros e artigos científicos (por exemplo, Abdel (2015)). Geralmente, é expresso em termos de Precisão e do Recall. No entanto, para manter uma consistência na formulação adotada neste trabalho, o F1-escore foi apresentado na Equação 5.3 com uma notação diferente da convencional. A demonstração da Equação 5.3 é fornecida neste apêndice.

A Figura A.1 sintetiza os conceitos fundamentais para compreender as definições da Precisão, Recall e, conseqüentemente, F1-escore. A Precisão é uma métrica que avalia a confiança na predição de uma classe pelo modelo, sendo definida da seguinte maneira:

$$Precisão = \frac{VP}{VP + FP}. \quad (A.1)$$

utilizando a notação empregada na Seção 6.1, reescreve-se a Precisão da seguinte forma:

$$Precisão = \frac{|T_i \cap P_i|}{|P_i|}. \quad (A.2)$$

em que o conjunto de pixels que pertencem a classe i é designado como T_i , e o conjunto de pontos classificados na classe i é designado P_i , onde $i \in [1, n_c]$ e n_c é o número de classes a serem classificadas.

O Recall é uma métrica que avalia a eficiência da predição de um modelo em detectar os elementos da classe alvo. É definido da seguinte maneira:

$$Recall = \frac{VP}{VP + FN} \quad (A.3)$$

utilizando a notação empregada na Seção 6.1, reescreve-se o Recall da seguinte forma:

$$Recall = \frac{|T_i \cap P_i|}{|T_i|}. \quad (A.4)$$

O F1-escore é definido como a média harmônica entre a Precisão e o Recall:

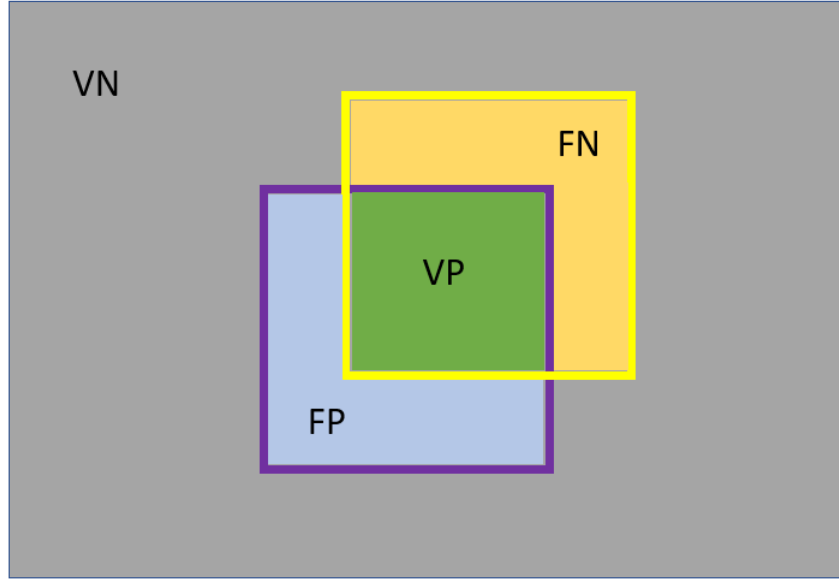


Figura A.1: Conjuntos relevantes na segmentação de uma classe em uma imagem. O retângulo maior, cinza, representa a imagem completa. Os pixels dentro do retângulo delimitado em amarelo correspondem à classe alvo, enquanto os pixels dentro do retângulo delimitado em roxo representam a classe predita. Os termos VP(Verdadeiro Positivo), VN(Verdadeiro Negativo), FP(Falso Positivo), e FN(Falso Negativo) são utilizados para descrever as diferentes categorias de pixels em relação à sua classificação correta ou incorreta.

$$F1 = 2 \times \frac{Precisão \times Recall}{Precisão + Recall}. \quad (A.5)$$

bastando então substituir a Precisão e o Recall pelas Equações A.2 e A.4:

$$F1_{i,i \in [1,n_c]} = \frac{2 \times \frac{|T_i \cap P_i|}{|P_i|} \times \frac{|T_i \cap P_i|}{|T_i|}}{\frac{|T_i \cap P_i|}{|P_i|} + \frac{|T_i \cap P_i|}{|T_i|}}. \quad (A.6)$$

$$F1_{i,i \in [1,n_c]} = \frac{2 \times \frac{(|T_i \cap P_i|)^2}{|P_i| \times |T_i|}}{\frac{|T_i \cap P_i| \times (|P_i| + |T_i|)}{|P_i| \times |T_i|}}. \quad (A.7)$$

Retornando assim a Equação 5.3:

$$F1_{i,i \in [1,n_c]} = \frac{2 \times |T_i \cap P_i|}{|T_i| + |P_i|}. \quad (A.8)$$

A comparação entre as formulações do F1-escore e do IoU nessa notação ressalta as similaridades e diferenças entre essas duas métricas. No numerador do F1-escore, é realizado o cálculo do dobro do número de pixels classificados corretamente, enquanto no denominador é realizada a soma da quantidade de pixels preditos e da quantidade de

pixels na classe alvo. Por outro lado, o IoU considera a união desses dois conjuntos como denominador, refletindo a medida de sobreposição entre a predição e a classe alvo.

Apêndice B

Algoritmos de aprendizado ativo

Algorithm 1: Aprendizado Ativo para Classificação de Fácies Sísmicas: amostragem aleatória simples.

Input: Conjunto de dados sísmicos de treino e teste,

$$\mathcal{D}_{\text{oráculo}} = \{(x_j, y_j)\}_{j=1}^{\lfloor \mathcal{D}_{\text{oráculo}} \rfloor}, \mathcal{D}_{\text{teste}} = \{(x_k, y_k)\}_{k=1}^{\lfloor \mathcal{D}_{\text{teste}} \rfloor}.$$

Output: Modelos \mathcal{F} treinados para cada iteração i com os respectivos valores de F1-escore medidos no conjunto de teste.

```
// Passo 1:  Guarde o tamanho do conjunto completo de treino na
             variável  $L = \lfloor \mathcal{D}_{\text{oráculo}} \rfloor$ 
// Passo 2:  Monte o conjunto inicial de treinamento com  $M$ 
             seções sísmicas  $x$  e suas anotações  $y$  retiradas de maneira
             aleatória de  $\mathcal{D}_{\text{oráculo}}$ ,  $\mathcal{D}_{\text{treino}}^0 = \{(x_j, y_j)\}_{j=1}^M$ .
// Passo 3:  Inicie o modelo,  $\mathcal{F}^0$ .
// Passo 4:  Loop
for  $i = 1$  to  $N_{\text{ciclos}}$  do
    // (a) Treine  $\mathcal{F}^{i-1}$  por  $N_{\text{epocas}}$  em  $\mathcal{D}_{\text{treino}}^{i-1}$ .
    // (b) Utilize o modelo  $\mathcal{F}^{i-1}$  para prever as fácies sísmicas
           em  $\mathcal{D}_{\text{teste}}$ .
    // (c) Calcule  $F1^{i-1}$  em  $\mathcal{D}_{\text{teste}}$ .
    // (d) Retire aleatoriamente  $M$  novas seções sísmicas e suas
           respectivas anotações de  $\mathcal{D}_{\text{oráculo}}$ .
    // (e) Forme o conjunto de treinamento para o novo ciclo,
           acrescentando as imagens e anotações obtidas em (d) de modo
           que  $\mathcal{D}_{\text{treino}}^i = \{(x_j, y_j)\}_{j=1}^{M \times (i+1)}$ 
    // (f) Se  $M \times (i+1) > L$ , SAIR do Loop.
end
```

Algorithm 2: Aprendizado Ativo para Classificação de Fácies Sísmicas:
amostragem estratificada em passos.

Input: Conjunto de dados sísmicos de treino e teste,

$$\mathcal{D}_{oráculo} = \{(x_j, y_j)\}_{j=1}^{\lfloor \mathcal{D}_{oráculo} \rfloor}, \mathcal{D}_{teste} = \{(x_k, y_k)\}_{k=1}^{\lfloor \mathcal{D}_{teste} \rfloor}.$$

Output: Modelos \mathcal{F} treinados para cada iteração i com os respectivos valores de F1-escore medidos no conjunto de teste.

```
// Passo 1:  Guarde o tamanho do conjunto completo de treino na
//           variável  $L = \lfloor \mathcal{D}_{oráculo} \rfloor$ 
// Passo 2:  Monte o conjunto inicial de treinamento com  $M$ 
//           seções sísmicas  $x$  e suas anotações  $y$  retiradas de maneira
//           aleatória de  $\mathcal{D}_{oráculo}$ ,  $\mathcal{D}_{treino}^0 = \{(x_j, y_j)\}_{j=1}^M$ .
// Passo 3:  Inicie o modelo,  $\mathcal{F}^0$ .
// Passo 4:  Loop
for  $i = 1$  to  $N_{ciclos}$  do
    // (a) Treine  $\mathcal{F}^{i-1}$  por  $N_{epocas}$  em  $\mathcal{D}_{treino}^{i-1}$ .
    // (b) Utilize o modelo  $\mathcal{F}^{i-1}$  para prever as fácies sísmicas
    //     em  $\mathcal{D}_{teste}$ .
    // (c) Calcule  $F1^{i-1}$  em  $\mathcal{D}_{teste}$ .
    // (d) Selecione um subconjunto de elementos de  $\mathcal{D}_{oráculo}$ ,
    //     tomando cada  $n$ -ésimo elemento, onde  $n$  é o salto e deve ser
    //     escolhido de maneira que o subconjunto possua  $M$  seções
    //     sísmicas com suas respectivas anotações.
    // (e) Forme o conjunto de treinamento para o novo ciclo,
    //     acrescentando as imagens e anotações obtidas em (d) de modo
    //     que  $\mathcal{D}_{treino}^i = \{(x_j, y_j)\}_{j=1}^{M \times (i+1)}$ 
    // (f) Se  $M \times (i + 1) > L$ , SAIR do Loop.
end
```

Algorithm 3: Aprendizado Ativo para Classificação de Fácies Sísmicas:
amostragem por incerteza baseada na entropia.

Input: Conjunto de dados sísmicos de treino e teste,

$$\mathcal{D}_{oráculo} = \{(x_j, y_j)\}_{j=1}^{\lfloor \mathcal{D}_{oráculo} \rfloor}, \mathcal{D}_{teste} = \{(x_k, y_k)\}_{k=1}^{\lfloor \mathcal{D}_{teste} \rfloor}.$$

Output: Modelos \mathcal{F} treinados para cada iteração i com os respectivos valores de F1-escore medidos no conjunto de teste.

```
// Passo 1:  Guarde o tamanho do conjunto completo de treino na
//           variável  $L = \lfloor \mathcal{D}_{oráculo} \rfloor$ 
// Passo 2:  Monte o conjunto inicial de treinamento com  $M$ 
//           seções sísmicas  $x$  e suas anotações  $y$  retiradas de maneira
//           aleatória de  $\mathcal{D}_{oráculo}$ ,  $\mathcal{D}_{treino}^0 = \{(x_j, y_j)\}_{j=1}^M$ .
// Passo 3:  Inicie o modelo,  $\mathcal{F}^0$ .
// Passo 4:  Loop
for  $i = 1$  to  $N_{ciclos}$  do
    // (a) Treine  $\mathcal{F}^{i-1}$  por  $N_{epocas}$  em  $\mathcal{D}_{treino}^{i-1}$ .
    // (b) Utilize o modelo  $\mathcal{F}^{i-1}$  para prever as fácies sísmicas
    //     em  $\mathcal{D}_{teste}$ .
    // (c) Calcule  $F1^{i-1}$  em  $\mathcal{D}_{teste}$ .
    // (d) Utilize o modelo  $\mathcal{F}^{i-1}$  para calcular a entropia de
    //     Shannon em todas as seções sísmicas restantes em  $\mathcal{D}_{oráculo}$ .
    //     Ordene  $\mathcal{D}_{oráculo}$  das seções com maior entropia para as de
    //     menor entropia.
    // (e) Retire as primeiras  $M$  seções sísmicas com suas
    //     respectivas anotações de  $\mathcal{D}_{oráculo}$ .
    // (f) Forme o conjunto de treinamento para o novo ciclo,
    //     acrescentando as imagens e anotações obtidas em (d) de modo
    //     que  $\mathcal{D}_{treino}^i = \{(x_j, y_j)\}_{j=1}^{M \times (i+1)}$ 
    // (g) Se  $M \times (i + 1) > L$ , SAIR do Loop.
end
```
