UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Elton Alves Trindade

Anisotropic Orthogonal Ensemble Network for 3D Seismic Facies Segmentation

Florianópolis
2021

Elton Alves Trindade

Anisotropic Orthogonal Ensemble Network for 3D Seismic Facies Segmentation

Dissertação submetida ao Programa de Pós-Graduação em Ciência da Computação para a obtenção do título de Mestre em Ciência da Computação.
Orientador: Prof. Mauro Roisenberg, Dr.

Florianópolis

2021

Elton Alves Trindade

Anisotropic Orthogonal Ensemble Network for 3D Seismic Facies Segmentation

O presente trabalho em nível de mestrado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Alexandre Gonçalves Silva, Dr.
Universidade Federal de Santa Catarina

Prof. Rafael de Santiago, Dr.
Universidade Federal de Santa Catarina

Prof. Dr. Felipe Guadagnin, Dr.
Universidade Federal do Pampa

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de Mestre em Ciência da Computação.

_____

Coordenação do Programa de Pós-Graduação

_____

Prof. Mauro Roisenberg, Dr.
Orientador

Florianópolis, 2021.

ACKNOWLEDGEMENTS

"Beneath all the wealth of detail in a geological map lies an elegant, orderly simplicity."— Tuzo Wilson

RESUMO

Os avanços tecnológicos na caracterização de reservatórios de petróleo e gás, como atributos sísmicos e sísmica 3D, enriqueceram a descrição da subsuperfície feita por especialistas. No entanto, o processo de segmentação de imagens deste enorme volume de dados tornou-se uma tarefa complexa. A fim de gerenciar de forma mais eficiente grandes dados sísmicos, este trabalho explora uma rede computacionalmente mais barata, com o uso de redes neurais convolucionais em planos ortogonais 2D redes neurais convolucionais para classificação de fácies e grupos lito-estratigráficos em cubos sísmicos 3D, guiados por uma heurística baseada nas Leis da estratigrafia, que é uma das etapas no processo de caracterização de reservatórios e exploração de óleo e gás. Foi proposta uma transferência de conhecimento de 2D para 3D, na qual dividimos as amostras de nosso dado 3D em 3 planos ortogonais entre si e a subsequente conversão dos parâmetros treinados para uma rede convolucional tri-dimensional equivalente. Cada um dos planos bi-dimensionais é convenientemente convertida em uma convolução tri-dimensional, emulando a visão de um géologo da amostra em 3 vistas ortogonais. As amostras de ambas redes (2D e 3D) foram extraídas em posições X,Y selecionadas aleatoriamente, simulando a extração de informação de poços já perfurados. A arquitetura de uma UNet foi selecionada para a rede proposta, visto que ela é uma das mais amplamente utilizadas para tarefas de segmentação de imagem. A metodologia proposta foi aplicada tanto em dados sintéticos (StanfordVI Reservoir) e dados sísmicos reais (F3-Block). Os experimentos no primeiro obtiveram excelentes resultados (96% na metrica IOU e 97,9% na métrica F1-Score), melhores que uma rede UNet 3D equivalente. Os resultados no dado F3-Block também foram superiores aos obtidos na literatura (ALAUDAH et al., 2019) e de uma rede UNet 3D, obtendo um resultado 7% maior que as demais na métrica MCA (Acurácia média por classe). Em comparação a outros modelos no mesmo benchmark, a rede proposta obteve melhores resultados a um custo computacional viável, sugerindo que essa metodologia é promissora, versátil e de fácil replicação.

Palavras-chave: Redes Neurais Convolucionais. Segmentação de Imagem. Interpretação Sísmica. Inteligência Artificial.

RESUMO ESTENDIDO

Introdução

Um depósito sedimentar pode se apresentar em diferentes escalas e formas, ambas controladas pelo gradiente deposicional em direção à bacia (HESTHAMMER; LANDRO; FOSSEN, 2001). Um geólogo tem como objetivo olhar o dado por todas as perspectivas, de modo a obter uma classificação confiavel de um depósito e/ou suas fácies e evitar a influência do ruído na interpretação. O mapemamento das camadas sedimentares e sua relevância econômica são fundamentais na avaliação de um campo de petróleo, incluindo a análise de risco inerente à atividade.

Interpretação de dados sísmico é o método mais utilizado para estimar as propriedades físicas e geológicas na indústria de óleo e gás. Através da sísmica, geocientistas avaliam o reservatório em relação à sua composição, fluido, dimensões e geometria.

Durante o processo de interpretação, geólogos/geofísicos utilizam análise dos sinais e seu conhecimento geológico para, manualmente, dividir o dado seguindo algum critério (Cronológico, sedimentar, geofísico) em classes. Essa tarefa tem grande potencial de aprimoramento através de Inteligência Artificial(IA), já que esta já é amplamente aplicada em problemas de segmentação de imagens.

Nesse contexto, o presente trabalho propõe uma abordagem de IA que emula o processo de interpretação de um geólogo, para uma tarefa de segmentação de imagem em sísmica 3D, através de uma abordagem convolucional multi-ortogonal.

Objetivos

Essa tese propôe:

- Investigar se cada vista ortogonal obtém uma perfomance distinta;

- Verificar se a influência do conhecimento geológico na arquitetura da rede na performance;

- Comparar os resultados entre uma rede multi-ortogonal e uma rede 3D equivalente, em relação à acurácia e o custo computacional;

Metodologia

De acordo com os estudos de (O'Mahony et al., 2018), a simplificação de um modelo 3D em múltiplos slices é uma idéia que permite operacões convolucionais eficientes e que torna viável o emprego de redes profundas pré-treinadas em grandes Datasets. Considerando isso, e as desvantagens de outras abordagens da literatura, a rede AOE (Anisotropic Ortogonal Ensemble) foi proposta, usando uma estrutura baseada na UNet (RONNEBERGER; FISCHER; BROX, 2015). A UNet foi criada para tarefas de segmetação de imagens médicas, cuja aquisição é baseada em um princípio físico semelhante ao da sísmica de reflexão. Para os experimentos foram utilizados um dado sísmico sintético e um real. Em ambos os casos 10 coordenadas X/Y foram aleatoriamente selecionadas, simulando poços já perfurados na área. Os dados de treinamento 2D foram extraídos ao longo dos mesmos, assim como os dados de treinamento 3D. Os dados de teste foram todas as amostras do dado. Todas as rotinas foram implementadas utilizando um Desktop com um processador Intel(R) Core(TM) i7-8700K CPU @ 3.70GHz, com 64GBRAM e uma placa de vídeo NVIDIA GeForce GTX 1080 Ti 12GB GPU. O código está disponível em https://github.com/eltontrindade/F3-block-dataset.

Resultados e Discussão

Os resultados do treinamento em 2D já demonstram que cada vista ortogonal das amostras possuem performances distintas umas das outras. A rede com amostras com a visão em planta obteve um resultado bem inferior às demais, enquanto que a rede no plano transversal obteve o melhor resultado geral, apesar de a rede longitudinal ter obtido um melhor resultado para a fácies "Canal". Tais diferenças foram interpretadas como reflexo da anisotropia dos depositos (em cada vista a associação de fácies e suas dimensões e geometrias diferem) e da frequência de ocorrências das classes. No dado Stanford VI, as redes AOE obtiveram os melhores resultados nas métricas utilizadas, melhor inclusive que uma rede UNet 3D (em média 1% melhor). Também foram realizados experimentos no dado sísmico real do bloco F3 do mar do norte (F3-block Dataset). Para esse dado, as classes eram unidades litoestratigraficas, que além de possuir características litológicas também possuem uma correlação com a idade geológica de formação. Nesse sentido, foi incorporada um conhecimento a priori baseado na Lei de Steno (Sobreposição vertical de camadas) e a Lei de Walter. A primeira afirma que dada uma sequencia de rochas empilhadas verticalmente, caso não haja nenhum evento tectonico significativo, as camadas inferiores são mais antigas que as superiores. A segunda afirma que camadas sedimentares verticalmente sobrepostas ocorrem lateralmente no tempo. De modo a adicionar essa informação, foi extraida a profundidade para cada amostra, usando apenas a indexação do próprio dado. A rede proposta obteve um resultado melhor que as da literatura em todas métricas, em especial a rede que utilizava a sísmica e a profundidade como informação de treinamento (aproximadamente 10% e 2% respectivamente). Esse resultado corrobora com a influência da adição de um conhecimento geológico a priori pode levar a uma melhor performance da rede. A rede AOE se mostrou melhor que a rede proposta por (ALAUDAH et al., 2019), fato comprovado por um teste estatistico com um poder de 98 % e um nivel de confiança de 95%).

Considerações Finais

Esta tese de mestrado descreve nossa abordagem para um problema de Segmentação de Imagens tri-dimensionais, através de nossa rede AOE. Este modelo foi capaz de obter uma performance melhor com uma menor complexidade computacional. Nossa proposta de planos ortogonais demonstrou-se viável para capturar o contexto 3D do dado e a diferença de resolução e anisotropia dos corpos geológicos, que também são aplicáveis a outras áreas da ciência. Embora os experimentos desta tese foram conduzidos apenas com a UNet como base, essa abordagem é flexivel e pode ser utilizada com qualquer arquitetura, inclusive diferentes arquiteturas dentro do ensemble. Essa flexibilidade, assim como sua eficiência, permite que essa metodologia seja atrativa para a indústria do petróleo, uma vez que o enorme tamanho dos dados sísmicos requer que a rede seja precisa e eficiente em termos de tempo processamento.

Palavras-chave: Redes Neurais Convolucionais. Segmentação de Imagem. Interpretação Sísmica. Inteligência Artificial.

ABSTRACT

Technological advances in oil and gas reservoir characterization, such as 3D seismic and seismic attributes, enriched the subsurface's description made by specialists. Nevertheless, image segmentation process of this now huge volume of data became a complex task. In order to more efficiently manage big seismic data, this work explores a computationally cheaper network with the use of 2D orthogonal planes convolutional neural networks for 3D seismic cube facies classification and lithostratigrafic groups, supported by an heuristic based on the Laws of Stratigraphy, which is one of the steps of reservoir characterization and oil and gas exploration. We proposed a 2D to 3D Transfer-learning in which we split the training samples of our 3D data as 3 orthogonal slices and convert the trained parameters to a 3D counterpart of the network, with each direction of the training planes conveniently converted to a 3D convolution, as if a geologist could see each sample through one of the orthogonal views. The samples for both 2D and 3D network were extracted in random selected X,Y locations, a sampling method that emulates the information extraction of drilled wells. Unet was selected as the backbone of the proposed network, since it it is one of the most applied architectures in image segmentation/classification tasks. Our methodology was applied in the synthetic data of the Stanford VI-E reservoir and the real seismic data of F3-block dataset. The experiments in the former obtained remarkable results (96 % in IOU and 97.9% in FWIU), even better than a 3D network using Unet. The results in the F3-block dataset were also superior to the ones obtained in the literature (ALAUDAH et al., 2019) and also a 3D Unet, with a MCA 7% above the previous mentioned networks. Compared to other models in the same benchmark, the proposed Anisotropic Orthogonal Ensemble Network classifier obtained better results than other architectures of literature at a very feasible computational cost, which suggests that such approach is a promising one and of easy replication.

Keywords: Convolutional Neural Networks, Transfer-learning, Image Segmentation, Seismic Interpretation.

LIST OF FIGURES

CONTENTS

1 INTRODUCTION

A sedimentary deposit may present itself in different scales and shapes, both are mainly controlled by the depositional gradient towards the basin (HESTHAMMER; LANDRO; FOSSEN, 2001). A geologist aims to look at all perspectives to have a confident classification of a deposit and/or its facies and avoid noise influence on the interpretation (HESTHAMMER; LANDRO; FOSSEN, 2001), meaning that in a real-world scenario the seismic data is manually interpreted in 2D images, rather than in a volumetric scale (XIONG et al., 2018). The mapping of the sedimentary layers and its economic relevance is fundamental in an oil field evaluation, including the risk assessment inherent to oil exploration.

Seismic data interpretation (Figure 1) is the most used method to estimate geological and physical properties in the oil and gas industry since it set out to detect shapes and petrophysical properties of underlying rocks in several scales (MONDOL; BJøRLYKKE, 2011). Geoscientists use the seismic data to evaluate the reservoir regarding their composition, fluid, dimensions and geometry (SELLEY; SONNENBERG, 2014).

In order to be more feasible, seismic interpretation is performed mainly by interpreting horizons in 2D seismic lines. Those horizons are mainly interpretation of paleo-surfaces or the boundaries (either top or bottom) of interpreted rock layers.

After the horizon interpretation phase, the amplitudes between those layers, or over the horizons, are extracted in order to identify depositional features, fluids and understand the distribution and connectivity of the sedimentary deposits as separate geobody units.

All those information combined result in a geological model conception, based on which the location of the well is chosen. Additionally, the geological prediction of the lithologies expected to be present in the well location are imperative for the well planning, and during the future development of the field (HAQUE; ISLAM; SHALABY, 2018). This task, however, is quite challenging since it requires the extraction of the most relevant information available and combines it with the complex knowledge of both geology and geophysics (HESTHAMMER; LANDRO; FOSSEN, 2001).

The amount of data available for an oil field and/or concession block is quite large (several TeraBytes) and its analysis is very time-consuming (over several weeks/months). In this scenario, the introduction of automatic routines may aid in tasks, such as interpretation and success probability predictions, since the amount of data is steadily increasing each day. Besides saving time in interpretation tasks, it also has a better capacity of combining several data sources in its output, compared to a human geologist/geophysicist.

During the seismic data interpretation process, the geologists/geophysicists responsible for the area, using all the geological knowledge and signal analysis, manually divide the seismic section according to his criteria (chronological, sedimentary, anomaly) in classes, in a very subjective and manual process. After identifying some geological features and/or surfaces, the interpreter reviews the assigned labels according to the literature distribution of the interpreted features (lateral and vertical association, lateral continuity, shape, among others),
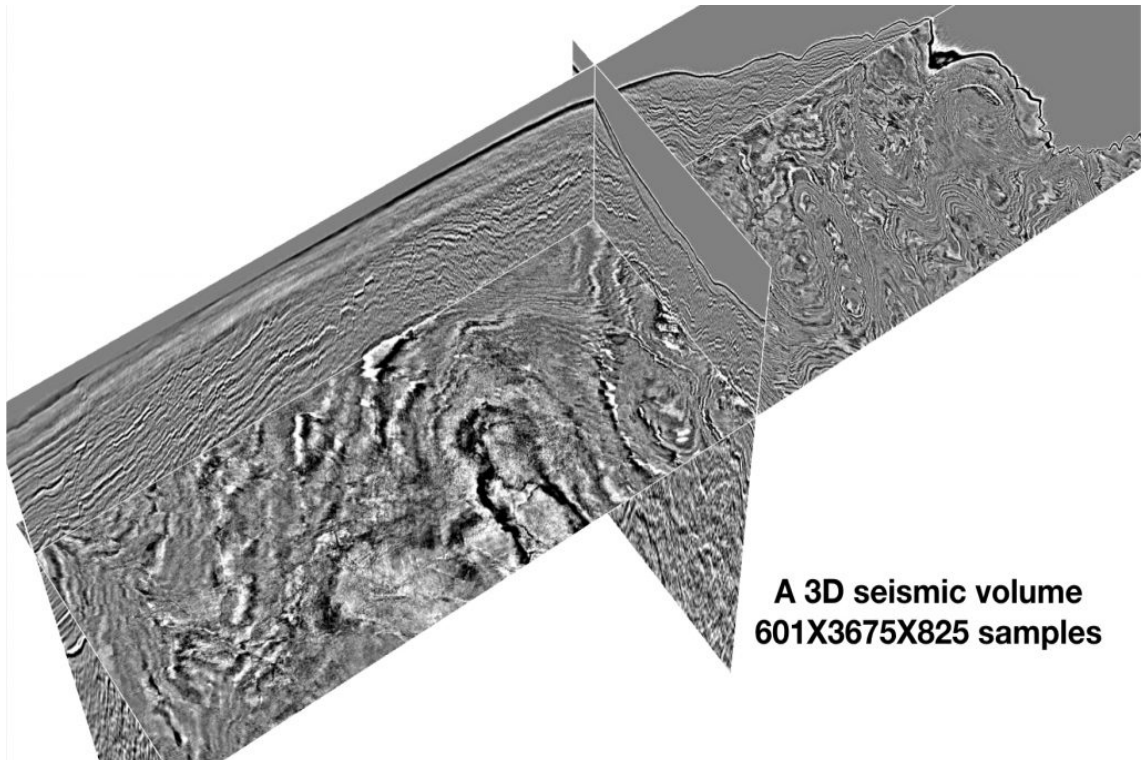
Figure 1 – Seismic Volume simplified in three orthogonal slices, which is the most frequent method geologists/geophysicists survey the data and map geologic features. Source: (WU et al., 2018)

as well as its spatial distribution on a 3D context. This tasks can be treated as image segmentation/classification problems, based on the seismic data amplitudes and geological features shapes, just like other AI (Artificial Intelligence) applications in image datasets.

Segmentation, classification and regression problems are very common in the computer science world, with several methodologies, approaches and insights, which some could be very useful in the seismic mapping. Artificial intelligence has a vast range of applications, from face recognition to e-mail spam filtering. In seismic interpretation, for instance, one main goal is to automatically model the reservoir properties from seismic data. This characterization, if automatically performed in large seismic datasets, can result in important time savings during the process of interpretation (RAMIREZ; LARRAZABAL; GONZáLEZ, 2015).

For instance, estimation of rock physical properties based on seismic data in combination with well logs data can be treated as a regression problem, where the informations obtained through the wells can be extrapolated through the whole data. Currently, the majority of commercial inversion technologies are usually based on the application of linear equations of the convolution type for estimating the functions of the reflectivity along the seismic track and follow-up evaluation of acoustic and elastic parameters of the geological formation (HAMPSON; RUSSELL; BANKHEAD, 2005).

When it comes to image classification and segmentation, CNN's (Convolutional Neural Networks) are the front runners deep learning algorithms, ever since the AlexNet won the ImageNet Challenge (KöPüKLü et al., 2019). Image classification is where machines can look

at an image and assign a (correct) label to it (or the probability of it), whereas image segmentation is based on partitioning an image into multiple parts or regions, usually based on the characteristics of the pixels in the image. Even though they have a bigger computational cost compared to other deep learning techniques, their better overall performance greatly make up for it(LECUN; BENGIO; HINTON, 2015). They have an advantage over the regression problems, since it takes into account the spatial correlation between the samples, which is very useful in geological bodies identification and classification, since those bodies are often amorphous and of heterogeneous distribution.

A seismic facies is a sedimentary unit which has different seismic characteristics than its surrounding units. Seismic facies identification is a fundamental process to identify geological patterns and can be treated as an image segmentation/classification problem, reducing the tedious and time consuming manual interpretation(DI; WANG; ALREGIB, 2018). Although CNN's and other deep learning methods seem suitable to this task, the majority of them classify the images as a whole, whereas in seismic the main goal is distinguishing geological bodies and/or features within the same image, therefore, an image segmentation problem (ZHAO, 2018).

Image segmentation has been proposed in medical imaging datasets, such as magnetic resonance (MRI) and computed tomography (CT) (YANG et al., 2021), which are comparable to seismic data. Coincidentally, the studies conducted by Su et al. (SU et al., 2015) with 3D shapes showed that building networks using a multi-view 2D image of those shapes dramatically outperform the classifiers designed directly on the 3D representations, which suits what happens in the seismic interpretation. Some geological features have complex three dimensional shapes and, albeit influenced by many environmental factors such as tectonic activity, may be present at different orientations. Since most of the available studies published in computer vision are applied in 2D images, and big image datasets are more available than 3D ones (YANG et al., 2021), it creates a gap to employ Transfer-learning techniques, since it brings a computationally cheaper way of capturing 3D contexts.

Transfer-learning is motivated by the fact that people can intelligently apply knowledge learned previously to solve new problems faster or with better solutions (PAN; YANG, 2010), just like someone who already plays guitar may use this prior knowledge to learn how to play bass faster. Since real data is scarce and of high cost in the oil and gas industry, Transfer-learning alternatives may benefit applications of AI, since information from large trained datasets or from other areas with similar characteristic and much more reliable information could be used to fit the model.

Transfer-learning from 2D CNNs, trained on large-scale datasets, has been proposed in 3D medical image analysis since there is a shortage of available labeled data of 3D datasets (YANG et al., 2021). It benefits from large scale pretraining on natural images, although its conversion to a 3D context might not be trivial (YANG et al., 2021).

In the Yang et al. work (YANG et al., 2021), it is proposed a 2D to 3D Transfer-learning, where three orthogonal (axial, coronal and sagital) networks are concatenated in the

channel dimension and operate at the same time in the 3D input (ACS Convolutions). However, in the proof-of-concept experiment, the kernels are extracted over the same 2D image, rotating the learned filter to each of the orthogonal axis as the ACS kernels. Such inference might be valid for some contexts but that does not suit well in an anisotropic image dataset and geobodies, like seismic facies.

The studies of Liu et al. (LIU et al., 2018) also transferred knowledge from a 2D pre-training in their proposed 3D Anisotropic Hybrid Network (AH-Net), but assigning the learned kernels in just 2 axis, regarding the resolution difference between slices and within slices. This difference is often observed in the seismic interpretation process. Seismic lines have a much lower resolution inter-lines than intra-lines, with the vertical resolution commonly around dozens of meters and the horizontal (or inter-lines) distance over 100 meters. Additionally, the interpreters also analyze a 3D seismic through 2D orthogonal slices, in order to grasp how the desired feature behaves in all directions or through the reflector shape assign more accurately a lithology or other geological feature. Another step aims to check whether what was interpreted in the surrounding lines, and the ones perpendicular, are indeed consistent with each other.

Based on the previous mentioned similarities, this master thesis proposes to combine the advantages of these previous works as an extension of the AH-Net proposed by Liu et al. (LIU et al., 2018) with the principles of the ACS architecture proposed by Yang et al. (YANG et al., 2021), in a model we called Anisotropic Orthogonal Ensemble Network (AOE Net). This network employs a Transfer-learning in which we split the training samples of our 3D data as 3 orthogonal slices (like a geologist/geophysicist usually does) and convert the trained parameters to a 3D counterpart of the network, with each direction of the training planes conveniently converted to a 3D convolution, by the expansion of each 2D kernel in one dimension, similarly to the Anisotropic Hybrid Networks (AH-Nets) proposed by Liu et al.(LIU et al., 2018), but with a specific axis assigned to each parameter 3D conversion.

At the end of the network the outputs of each AH-Net is averaged in an Ensemble, in order to take into account the output of the network from all three orthogonal points of view. This approach may also benefit from parameters learned in 2D images large datasets. This model aims to insert a geological heuristic based on the different views of the geobodies, similar to the human manual interpretation process and its application in anisotropic 3D images while maintaining a low computational cost.

1.1 RESEARCH PROBLEM

Although the main source of data for oil exploration consists of a 3D seismic volume, the interpretation process is mainly conducted in 2D sections. Additionally, that approach is a simplification in order to make the process more feasible for a human interpreter. A 3D interpretation is a more accurate and has a better spatial contextualization than a 2D one, but it is more complex computationally wise and is not supported by most of the interpretation

softwares. This trade-off is present in the oil and gas industry workflow for many years and has a potential for AI optimization techniques.

In order to increase the AI applications in oil and gas industries, the algorithm must find routines that have a good performance and during feasible time, which is not an easy task, since the amount of data can often are be as large as several Terabytes.

Most of the datasets in which AI routines are applied are from natural images, which do not have an expert insight or a innate knowledge approach. Even on medical datasets, which derives from expensive and manual labelling process, the principles that guide an specialist labelling is usually not taken into account during the model conception.

Considering all the issues regarding the high computational cost of a pure 3D network, and the inherent anisotropy of geological features and seismic data, this research proposes a seismic facies classification composed of 3 AH-NET like networks arranged, based on geological knowledge, in an Ensemble (AOE Network). We adopted the UNet(LECUN; BENGIO; HINTON, 2015) backbone since it is the most employed architecture and was developed to a medical image segmentation task. Both seismic data and medical images (like tomography) are acquired through the same physical property. Both compose their images based on the interaction between the target (rock layers and human tissues, respectively) and the waves emitted by the source. The physical contrast between the waves and the target are captured and its response results in the image analyzed by experts.

1.2 OBJECTIVES

1.2.1 General Objectives

In order to investigate the before aforementioned hypothesis, this thesis proposed to explore a multi-orthogonal approach for a 3D network, allied with a specialist knowledge input, to increase the accuracy in 3D seismic facies prediction.

1.2.2 Specific Objectives

This thesis aims to:

- Investigate whether each orthogonal view of the network has a different performance;

- Verify whether the influence of geological knowledge in the network architecture improves the performance;

- Compare the results between a multi-orthogonal network and a 3D equivalent, regarding the accuracy and in computational cost;

1.3 RESEARCH QUESTION

For this thesis, there are the following research question:

- Does an orthogonal planes approach achieve better results than a 3D counterpart?

1.3.1 Hypothesis

Automatic segmentation of geologic features is very useful and time saving in the oil exploration workflow. This task, however, is conducted on very large data, which leads to a very costly and a large time consuming process. Besides, this automatic extraction and inference of geological interpretation does not possess any heuristic or geological knowledge inherent to it. This research aims to investigate if a simplification of a 3D network in orthogonal planes, through an ensemble network, outperforms its 3D equivalent and other state-of-art proposals in seismic facies segmentation, with a lower computational cost.

The research considers the following hypotheses:

- Hypothesis 1: Prior knowledge does not yield better results than a default 3D and state-of-art approach. Alternative: Neural Networks achieves superior results than a default 3D and state-of-art approach.

- Hypothesis 2: Three Orthogonal planes ensemble does not yield better results than a default 3D and state-of-art approach. Alternative: Three Orthogonal planes ensemble achieves superior results than a default 3D and state-of-art approach.

1.4 TEXT STRUCTURE

This thesis is organized in the following topics: After the Introduction, Chapter 2 covers the core concepts regarding both geology/geophysics and machine learning topics related to the research. Chapter 3 contains the review of AI applications and its main contributions to oil and gas tasks and/or image segmentation. In Chapter 4 the methodology of the experiments and the data information is provided and explained in more detail. Chapter 5 contains the results of the experiments proposed, as well as the comparison between different architectures and approaches, throgh the concepctions of the proposed methods. And, lastly, the conclusions and final considerations are presented in Chapter 6.

2 BACKGROUNG

Since our research involves two very distinct science fields (Artificial intelligence and geology), this chapter several key topics are exposed in order to better contextualization of the theoretical foundation of this work.

2.1 GEOLOGY CONCEPTS

2.1.1 Facies

Outcrop and rock description has been the mainstay of the geology, even before it was defined as a science. Even without much formalism, the descriptions were based on some criteria, in order to differentiate one rock formation from another. Along the geological concepts creation, some authors postulated their principles which they observation relied on (CROSS; HOMEWOOD, 1997; TEICHERT, 1958). However, it was needed standard criteria to correlate horizontally and vertically different and similar formations, which could be based on depositional process, age, petrologic features, among others. The term "facies" created by the swiss geologist A. Gressly is the most widely used when dealing with sedimentary rocks (CROSS; HOMEWOOD, 1997). Facies are rock layers with paleontologic and petrologic features that are distinguishable from its surroundings. Classifying lithologies and facies, and its associations, is important to distinctly define rocks of interest and to build a better understanding of the depositional environments encountered, making it more feasible to interpret the depositional environment and how the sedimentary process took place in the area (MOHAMED et al., 2019).

Facies classification is an essential element of geologic investigations and involves using recorded attributes and measurements to assign a class or type to rock samples. Defining a rock facies precisely can build a better understanding of the depositional environment penetrated by a wellbore (WEI et al., 2019). Even though Gressly is widely credited with the first modern use and definition of "facies" (DUNBAR; RODGERS, 1957; TEICHERT, 1958), his contributions to stratigraphic principles are much broader and deserve greater appreciation. He explained the genesis of sedimentary facies by processes operating in depositional environments, and demonstrated regular lateral facies transitions along beds that he interpreted as mosaics of environments along depositional profiles. He recognized the coincidence of particular fossil morphologies with particular facies, and distinguished "facies fossils" from those that had time value and that were useful for biostratigraphy ("index" or "zone" fossils) (WEI et al., 2019).

Teichert (TEICHERT, 1958) summarized Gressly's use of the term "facies" and his derivation of the facies concepts, since Gressly's purpose was not just to propose the term "facies" for descriptive rock attributes independent of time connotation, rather setting apart a rock term ("facies") from a time term ("terrain" or timestratigraphic unit). Without having a language to express these two properties of strata, one cannot differentiate between lateral variations in

lithology ("facies") along one or more beds (timestratigraphic units), and vertical changes and repetitions in lithology through a succession of beds. Gressly defined the term "facies" to single out those observable physical, chemical, and biological properties of rocks that collectively permit objective description, as well as distinctions among rocks of different types (WEI et al., 2019). One fundamental aspect of Gressly approach is its explicit discrimination between objectively observable properties and any connotation of their age (CROSS; HOMEWOOD, 1997).

Having explained that facies are properties of rocks not specific to time, Gressly further recognized that facies are products of genetic processes that operated in the depositional environments in which they accumulated. Just as laterally linked depositional environments change over a geographic area, the facies that are incorporated into the stratigraphic record change gradationally along beds that are parallel original depositional surfaces. He observed that, by walking along beds and following changes in the physical and paleontological components of facies, one can establish the details of a depositional profile (CROSS; HOMEWOOD, 1997). Gressly understood that there are two basic concepts in stratigraphy: the first is that sediments accumulate by a set of processes in depositional environments, and the second is that this happens during the passage of time.

Gressly then newly conceived approach had a few advantages, being the most significant:

- It simplifies the apparent complexity in paleontology and provides a coherent link between paleontological and physical and lithological attributes by establishing a limited number of closely interrelated laws.

- It is the basis for reconstructing successive paleogeographies and depositional profiles through time.

- He developed a new method of stratigraphic correlation, based not upon establishing the equivalency of rock type, but upon establishing equivalency of rocks in a time frame. it is the basis for understanding four-dimensional time-stratigraphic relationships.

These principles are employed by geologist from many areas of study, ranging from basic field research to oil exploration, allowing geoscientists to properly correlate and/or establish geological models even when information is scarce and far apart.

2.1.2 Brief review of Depositional systems

Sedimentary basins are large (at least $10.000 \ km^2$) lower topographic areas where the combination of deposition and subsidence has formed thick accumulations of sediment and sedimentary rock. Sedimentary basins are of great scientific and economic value, since it host the majority of oil and gas accumulations and it contains the geological and biological evolution in its sediments/rocks. A sedimentary environment is a geographic location characterized by a

particular combination of climate conditions and physical, chemical, and biological processes (Figure 2) which reflect in their petrological features (grain size, sorting, mud content, among others).

Marine environments can be classified based on the distance from land (GROTZINGER et al., 2007), which we will focus on two main ones for the oil and gas industry:

- Continental margin and slope environments are found in the deeper waters at and off the edges of the continents, where sediment is deposited by turbidity currents. A turbidity current is a turbulent submarine avalanche of sediment and water that moves downslope. Sediments deposited by turbidity currents are almost always siliciclastic, except for sites where organisms produce a lot of carbonate sediment. In this case, continental margin and slope sediments may be rich in carbonates.

- Deep-sea environments include all the floors of the deep ocean, far from the continents, where the waters are much deeper than the reach of wave-generated currents and other shallow-water currents, such as tides. These environments include the continental slope, which is built up by turbidity currents traveling far from continental margins; the abyssal plains, which accumulate carbonate sediments provided mostly by the skeletons of plankton; and the midocean ridges.

Continental shelves on passive margins (such as in the East Coast of South America) are mainly composed by fine-grained sediments, siliciclastic and/or carbonatic, with episodic coarse-grained deposits caused by mass flow into the basin. Albeit formed in distinct sedimentary environments, deep water deposits and fluvial deposits share some similarities (Figure 3), which is a consequence of both being mainly driven by the same physical principles and physiographic factors (gravity, relief, viscosity, surface, grain size, among others).

Among the several types of sedimentary deposits in this paleo-geographic context, the high density debris flows, better known as turbidites, are the ones with the most economic relevance. Turbidites are formed by turbidity current moving downward due to its density contrast with the sea water, which impacts the intensity, erosion strength and geometry of the sediments flow. Those currents represent the main mechanism sedimentary deposition to oceanic basins (MUTTI et al., 2009; TALLING et al., 2015).

Turbidites usually occur with distinct longitudinal and transverse variation, depending on its position in the slope. They typically are elongated parallel to the steepest slope direction and towards the basin depocentre, with its sinuosity (perpendicular to the flow direction), varies according to the irregularities of the surface or even the decrease of the energy of the flow (Figure 3). Additionally, the shape of the geological bodies, in a transverse view, may range from thinner and larger in area (basin-floor fan) to incised valleys, whose geometries are also driven by the terrain physiography and the sediment flow properties (higher density flows result in more restrict and thicker deposits, whereas lower density ones form shallower and broader deposits). It is noteworthy that, due to the relative sea-level oscillation during geologic time
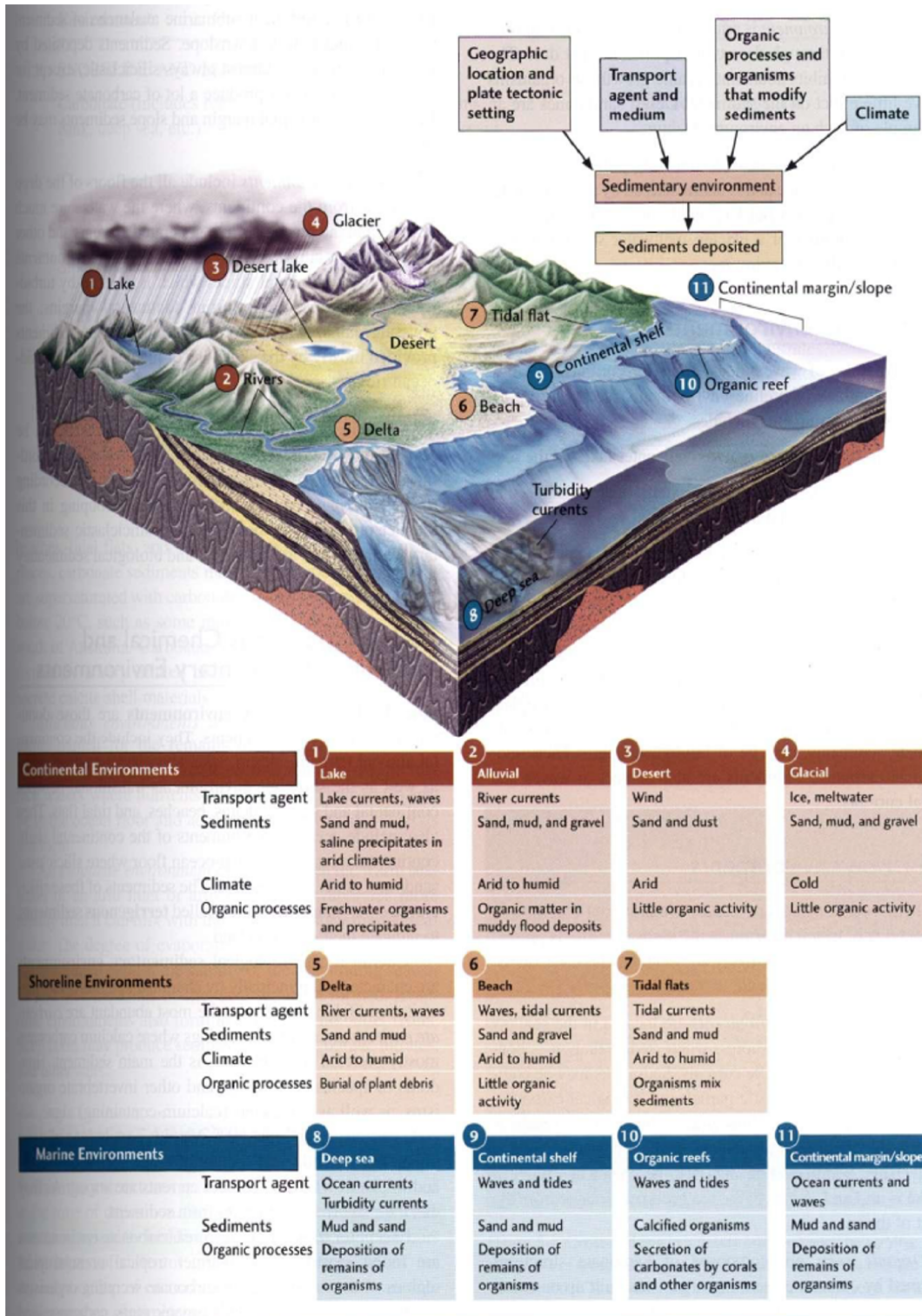
Figure 2 – Agents and components of sedimentary processes, from continental environments to deep sea. Source: (GROTZINGER et al., 2007)

and sediment influx rate, the paleogeography and the depositional components may change, generating overlapping or even obliteration of the primary structures and geometries of the sediments.
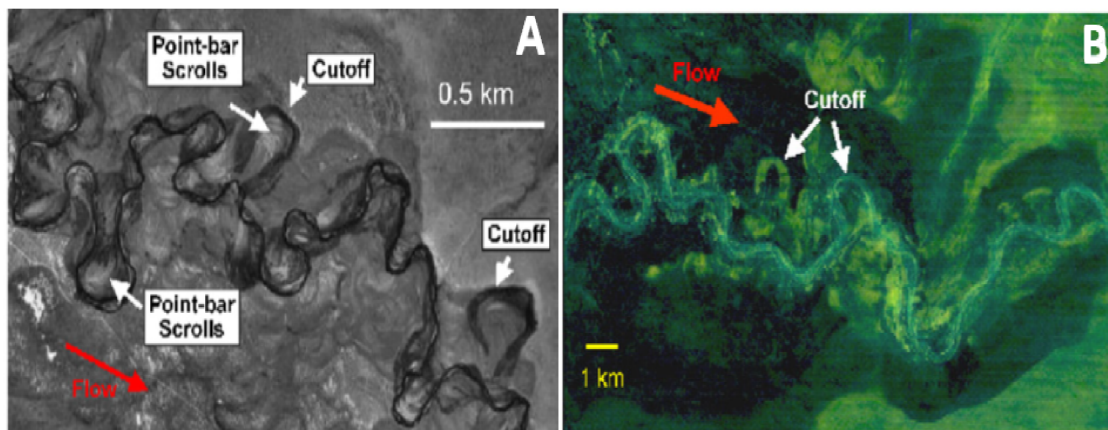
Figure 3 – High sinuosity deposits of Bighorn River (A) and offshore eastern Borneo, Kalimantan, Indonesia (B). Although they are subaerial and subaqueous, respectively, they both show similar shapes. Source (KOLLA; POSAMENTIER; WOOD, 2007)

2.1.3 Stratigraphy concepts

Nicolas Steno (1638–1686) is considered by many as the father of most basic principles of stratigraphy. These principles form the exclusive basis for the interpretation of a large part of the history of Earth (KRAVITZ, 2014). In the 17th century, Nicolas Steno established three cardinal principles of stratigraphic analysis (MACLEOD, 2005):

- Original horizontality – unconsolidated sediments deposited on a solid base must have originally formed horizontal layers since the sediment particles would have 'slithered' to the lowest point. Thus, consolidated strata inclined at some angle must have become tilted after consolidation.

- Original continuity – layers of unconsolidated sediments deposited on a solid base would have formed continuous sheets of material. Thus, bands of consolidated sediments whose ends have been broken must have experienced this breakage and erosion after consolidation.

- Superposition – Since each layer of unconsolidated sediment deposited on a solid base must form after the basal layer has been deposited, layers of sediment that overly other layers are younger than the other layers.

The Law of Superposition Steno proposed states that in an undisturbed sequence of rocks, the bottom layer is oldest, and the top layer is youngest. Based on that, geologists are able to determine the relative order, or continuum of geological events in time (SLATT, 2006). Those principles are fundamental in geological mapping and correlation between outcrops, since those laws rule how the different rock formation are expected to behave spatially and in depth, and were taken into account in the 20th century advances in stratigraphy.

A "sedimentary facies" (or just "facies") is a term used to identify a sedimentary rock/layer unit regarding its distinct and specific descriptive features, originated from physical, biological, and/or chemical processes, during formation and from which an interpretation of its origin may be made (LOPEZ, 2013). In other words, specific environment and sedimentological conditions results in different facies. Sedimentary facies can be associated and grouped by lateral (or horizontal) association, or "facies sequence" when assigned to distinct vertical stacking, succession, or sequence of facies that reflects a particular depositional environment or linked environments in the stratigraphic record (MIDDLETON, 1973).

Those concepts were later used in sedimentology as foundation to the Walther's Law, which states that various deposits of the same facies and similarly the sum of the rocks of different ones are formed beside each other in space, though in a cross section they are vertically stacked (WALTHER, 1894). As a depositional environment shifts, so too must the sedimentary facies in any location change. As time progresses, the positions of facies also progress laterally in space and time, causing the laterally related environments to become superimposed forming vertical successions, preserving the originally laterally continuous environments in vertical sections (MIDDLETON, 1973). The combination of both lateral facies tracts and vertical successions can be used to map broader three-dimensional depositional systems tracts that migrate through time and space as a function of cyclic forcing (FRIEDMAN; SANDERS; KOPASKA-MERKEL, 1992). Identifying and mapping these depositional systems tracts and how they change through time are a fundamental component of sequence stratigraphy and very useful in finding targets for oil and gas reservoirs. Such insight is imperative in oil and gas industry, since the hard data is obtained mainly through wells, which have a horizontal resolution of inches. By understanding the vertical stack, geologist have a better understanding of which facies to expect in the whole oil field area, since this correlation is likely to occur horizontally.

2.1.4 Seismic data

In seismic reflection, a controlled sound wave is generated on the ground surface or underwater in marine environment and detected on the surface using geo- or hydro-phones. As seismic waves from the source through the earth, portions of that energy are reflected back to the surface as the energy waves traverse through different geological layers affected by the contrast of properties of the rocks (AMINZADEH; DASGUPTA, 2013). The reflections, which are a result of a significant property contrast between layers/rocks, are captured at the surface by the receptors (geo- or hydro-phones), where their delay time (from source to reflectors) and amplitude is recorded (Figure 4).

Since the amplitude is a measure of the contrast of properties between layers, it can be used for prediction of heterogeneity's in reservoir rocks, net pay prediction, and fluid contacts. Lateral changes in amplitude from trace to trace along the same events or rock interface across an area could be an indicator of changes in deposition environment, porosity, rock type, or fluid saturation. (AMINZADEH; DASGUPTA, 2013). One relevant aspect of the seismic amplitude

is its ambiguos nature: the same response might be caused by several rock associations. For instance, a the same reservoir, with the same petrological characteristic, may have different amplitudes if the surrounding rocks very along it, since the seismic waves are heavily dependent on the impedance contrast.
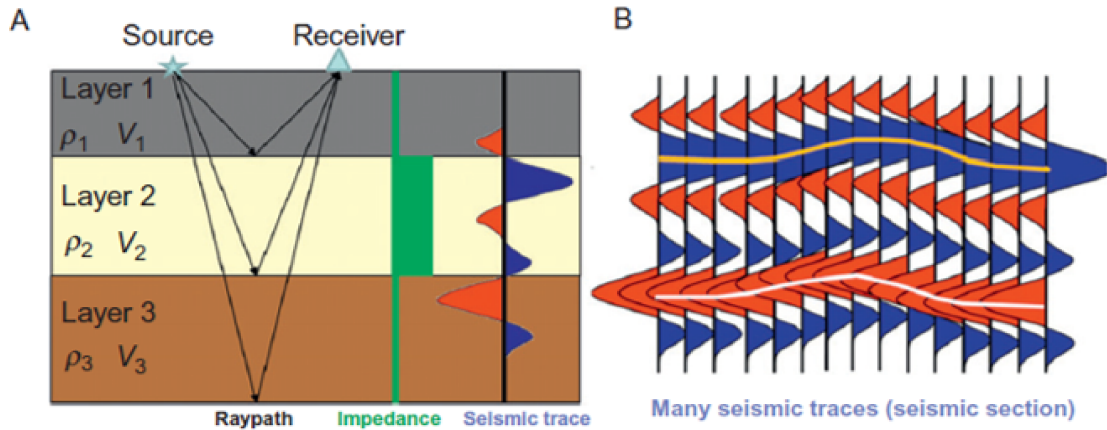
Figure 4 – A simplified view of the seismic reflection data acquisition : (A) A geological model with three layers with different properties (velocity and porosity) and the resultant measures of the receiver (impedance and seismic trace) (b) the seismic section if the layers had heterogeneous velocities.(BONESS, 2013)

Geophysical methods provide indirect measurement of the subsurface properties, which combined with measurements such as core analysis data and well testing help in the reservoir characterization process.

Well data has high vertical resolution, but poor lateral definition, which implies that well log data are often correlated with wells hundreds to thousands of meters far apart. In contrast, geophysical data has a broader coverage (dozens of km) and is uniformly sampled, albeit with lower vertical resolution when compared to well data.

The correlation between the well information with seismic data allow interpreters to extrapolate their geological model and contribute to a better spatial resolution of the reservoir modelling, providing a 3D distribution of the rock properties measured at the wells (AMIN-ZADEH; DASGUPTA, 2013).

Seismic data is a very cost-effective method when compared to other expenses (such as a well drilling) and provides a more accurate estimation and valuable information of the depth and arrangement of the the underlying rocks, some of which may contain oil or gas deposits.

The interpretation of geophysical measurements along with depositional, diagenetic and other geologic data, and reservoir information and assumptions are blended together to form input parameters for the initial reservoir characterization model (NOLEN-HOEKSEMA, 1990).

2.2 BRIEF REVIEW OF DEEP LEARNING

The pioneers convolutional neural networks (CNN's) studies, similarly to the perceptron, had their origin in biological experiments. In the late 50's and early 60's, the researchers Hubel and Wiesel (HUBEL; WIESEL, 1959) studied the nervous system of cats and the image processing mechanism of their visual system. In their study it was noted that neurons from different stages of visual cortex were activated by patterns strongly orientated by light, but seemed to ignore more complex patterns, which in response were activated in latter stages neurons (RAWAT; WANG, 2017). In 1979, a multi-layer neural network named neocognitron was proposed by Fukushima (FUKUSHIMA; MIYAKE, 1982), based on Hubel and Wiesel experiments. This network achieved good results in simple input pattern recognition, regardless of distortion or position variation of the object (RAWAT; WANG, 2017). Such results set the foundation the framework of the 21's century CNN's (LECUN; BENGIO; HINTON, 2015).

Convolutional Neural Networks (CNN's) are multi-layer artificial neural networks designed to handle inputs with more than a single dimension with two main structures: convolution and pooling layers. The basic operators of convolutional layers are kernels (or filters), which are matrix of values (or weights) that are updated during the training to identify better the features. The input is convolved with the kernels, aiming to extract the most relevant features for each kernel, with the overlapping area of the kernel with the input called receptive field. The output of each convolutional layer is summed with a bias term and a non-linear activation function, such as ReLU (Rectified Linear Unit), which is one of the most popular and resembles the neurological impulse of the brain. The training process becomes more efficient, memory and time-wise, with the addition of a pooling layer between convolutional layers, reducing the input redundancy and the amount of data passed through the deeper layers of the network. Pooling operators, besides reducing the feature maps resolution, allows the network to be invariant to distortions and translations. Several convolutional and pooling layers are sequentially organized in the network architecture, with last one followed by a fully-connected layer and a feature vector, where the network output is evaluated, just like the traditional neural networks (Figure 5).

Although they were already a fact since the early 90's, the computational power then available was not enough to implement it. The application of CNN became a new trend when a CNN greatly outperformed the state-of-art architectures during the Imagenet challenge in 2012. Even though several variations of its core concepts have been proposed, with the birth of classical networks (AlexNet, VGG, ResNet), the majority of those competitions utilized 2D datasets, with a great amount of labelled data. 3D datasets are not abundant, being comprised mostly by medical images and video data (time as the third dimension). The former is the most frequent in the literature, since its annotation is more feasible with a specialist input and its vital application in medical treatments, such as early cancer nodules identification.

Initially conceived to 2D images applications, CNN's architectures can be converted to a n-dimensional context, like a 3D seismic data. However, the size of a 3D dataset is much

bigger than a conventional 2D RGB images, along with an additional dimension for each con-
volutional filter, which increases the computational complexity in one order. The geometrical
comparison between 2D and 3D networks contexts is exposed in Figure 6.
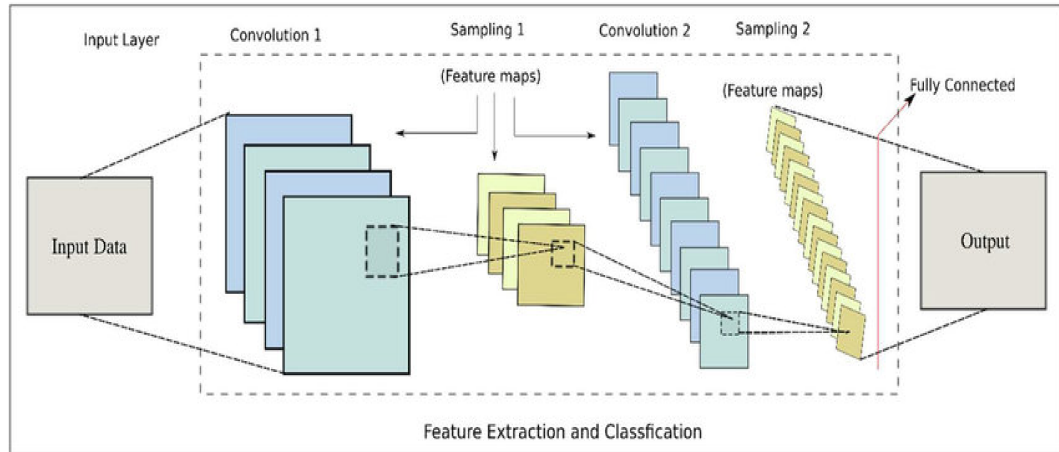
Figure 5 – Basic workflow of deep learning.(JAN et al., 2017)
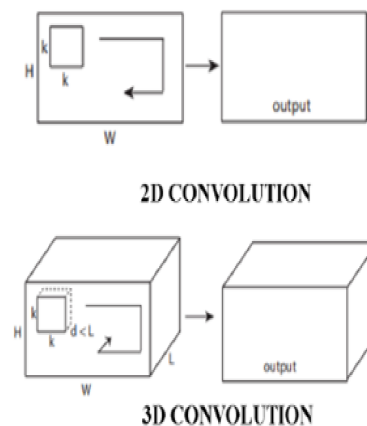
Figure 6 – Two-dimensional Convolution vs Three-dimensional Convolution. Modified from (TRAN et al., 2015)

2.2.1 Transfer-learning

Neural networks that have learned one task can be reused on related tasks in a pro-
cess that is called "transfer" (PRATT; JENNINGS, 1998). Similar to some human learning
processes, like previous knowledge on how to play guitar helps you lean how to play bass or
another instrument, with the degree of the impact depending upon the correlations between
the two instruments. Typically, the domain from which information is extracted is called the
"source" and it is transferred to the "target" domain (Figure 7).

Since the oil field data is very large its unfeasible to train from scratch over and over
again, which can benefit heavily from Transfer-learning techniques, either from the same area
or from a similar one, considering that you can transfer the knowledge from a previous trained

network instead of running a new one from scratch, saving many hours, even days, of processing time without a significant loss of accuracy. Such techniques have shown remarkable results even when the shared parameters were exhaustively trained in a different semantic context than the ones you are targeting.

Shortly after ImageNet competition revolutionized the Neural Networks field, some authors described the good generalization ability of pre-trained networks trained with ImageNet datasets (LECUN et al., 1998; DONAHUE et al., 2014). Transfer-learning with a pre-trained network has two common approaches (LUNDSTR OM, 2017):

- Feature extractor or frozen weights: the output from some layer in the network is used as features for a trainable classifier (usually the layer before the classification layer is chosen);

- Weight initialization or fine-tunning: it uses some of the pre-trained layers as a weight initialization and then trains the entire network for the new task;

In the work of (AZIZPOUR et al., 2015) the influence of 5 factors affecting the transferability were investigated: which layers to cut, whether fine-tuning should be used, the underlying architecture, source/target similarity and the benefit of additional data. This study came to a conclusion that fine-tuning generally helps, which was also stated by (YOSINSKI et al., 2014) on experiments with the ImageNet. In their study fine-tuning and frozen weights were studied layer by layer of the network, regarding which layer could be cut. They concluded that fine-tuning is unsusceptible to the layer choice, whereas the frozen weights are heavily dependant of the layers localization.

Transfer-learning aims at improving the performance of target learners on target domains by transferring the knowledge contained in different but related source domains. In this way, the dependence on a large number of target domain data can be reduced for constructing target learners (ZHUANG et al., 2021). However, there are instances where labeled training data is scarce and expensive, which creates a need to obtain high-performance learners from datasets from different domains (WEISS; KHOSHGOFTAAR; WANG, 2016). This knowledge transfer can be divided into two main categories: homogeneous and heterogeneous Transfer-learning. Homogeneous Transfer-learning approaches are developed and proposed for handling the situations where the domains are of the same feature space or slightly different, whereas in heterogeneous Transfer-learning the domains are different (Figure 8), which requires some adaptations and makes it more complex than the former (ZHUANG et al., 2021).

According to (PAN; YANG, 2010), the Transfer-learning approaches can be categorized into four groups:

- Instance-based: mainly based on the instance weighting strategy;

- Transform the original features to create a new feature representation;
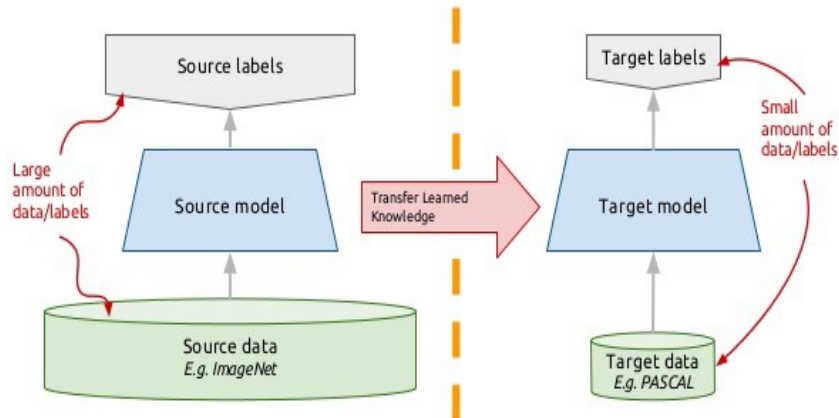
Figure 7 – Transfer-learning main workflow components. Source:(MCGUINESS, 2017)

- Transfer the knowledge at the model/parameter level;

- Mainly focus on the problems in relational domains;

The learning process can also be divided in two classes (SHLEZINGER et al., 2020): data-based and model based. The former, also known as "hand-designed method", has its mapping guided by prior knowledge of the underlying statistics relating the input x and the label. The latter, in contrast, learn their mapping through data. Roughly speaking, data-based interpretation covers the above-mentioned instance-based and feature-based Transfer-learning approaches, but from a broader perspective. Model-based interpretation transfers the knowledge at the model/parameter level.

Most Transfer-learning proposals heavily focus on transferring the knowledge via the adjustment and the transformation of data (Figure 8). If there is an available dataset that is drawn from a domain that is related to, but does not an exactly match a target domain of interest, then homogeneous transfer learning can be used to build a predictive model for the target domain, if the input feature space is the same (WEISS; KHOSHGOFTAAR; WANG, 2016).

The methodology of homogeneous transfer learning is directly applicable to a big data environment, such as seismic data. In this study, we focus more on homogeneous Transfer-learning, which aims to reduce the distribution difference between the source-domain and the target-domain instances.

2.2.2 Deep learning applied to oil and gas industry

At present, AI has been widely used in many industries (such as communication, financial, search engines), and, although has been recently widely applied, it still has a long way
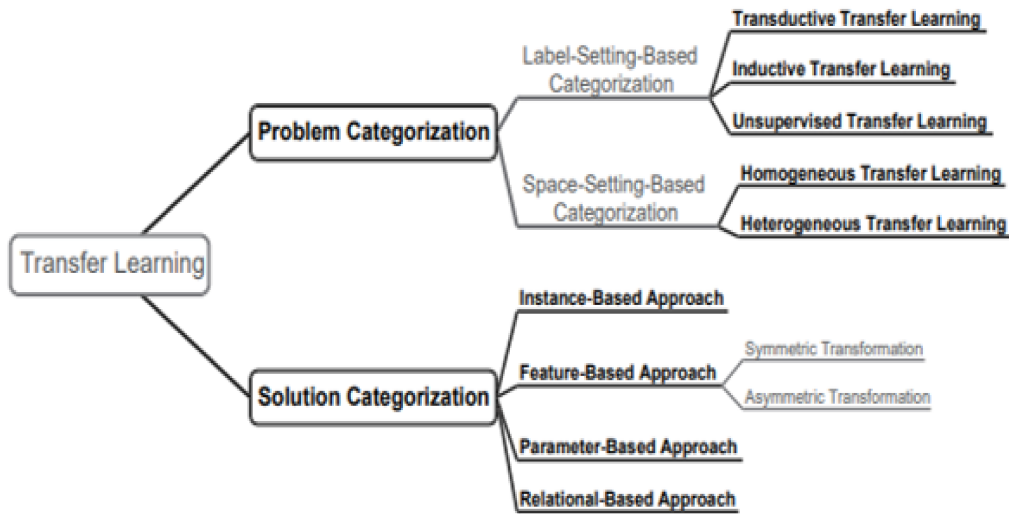
Figure 8 – Transfer-learning categories. Source:(MCGUINESS, 2017)

to go in the unique topics of oil industry (LI et al., 2020b), which are of unique nature and complexity. Data scientist, when trying to implement deep learning algorithms for oil exploration, usually deal with some technical issues, such as the large data amount, diversity of subjects involved, unstructured data. With the development and application of big data and continuous improvement of various related algorithms, AI plays an increasingly important role in the field of oilfield development. New technologies and systems involved with AI will be constantly proposed, which will surely become an important way to reduce costs and improve efficiency. Nowadays, a competitive oilfield should be one able to fully perceive, automatically control, predict trends and optimize decisions (DENNEY, 2006; ANDERSON; BARVIK; RABITOY, 2019).

The exploration and production (E&P) activity cycle generally starts with analysis of seismic and other geophysical data that leads to drilling of exploratory well(s) in the first phase. In the case of a promising discovery, the project enter second phase of action, where more data and a better resolution seismic data are acquired to identify suitable locations for drilling delineation/appraisal wells. Depending on the reservoir evaluation and economic viability, appropriate development plans are initiated in the final phase by planning and drilling the production/injection wells framework.

At first, most ideas were applied to well log information, where rock properties patterns were used to extrapolate some useful properties, such as porosity (SHIBILI; WONG, 1998; CHEN et al., 2020), velocity (FABIEN-OUELLET; SARKAR, 2019; YANG; MA, 2019) and rock classification (ANJOS et al., 2020; YANG; MA, 2019). This approach was deemed very promising, since the rock sampling process was expensive and it was composed by 2 centimeter samples, which is quite small to a reservoir scale, or core sampling, that, although it is typically over 6m, is a much more complex operation and may cost millions of dollars.

Since sampling all the wells drilled in an area is not feasible, machine-learning methods were used to transfer the information obtained by the above previous methods in one or few wells to the many others drilled in the same area, tying the information through the well log data, which is much cheaper and obtained by the default.

It is worth mentioning that it is imperative the collection and high quality of hard data from wells and its adequate geological description in order to have a robust result. In the field of exploration, Artificial Neural Networks have already achieved good results in reducing exploration risks like improving the success rate of exploration wells (CADEI et al., 2020) and reducing the drilling cost (HOJAGELDIYEV, 2018).

Seismic data interpretation is the most used method to estimate geological and physical properties in the oil and gas industry since it set out to detect shapes and petrophysical properties of underlying rocks in several scales (MONDOL; BJøRLYKKE, 2011). Geoscientists use the seismic data to evaluate the reservoir regarding their composition, fluid, dimensions and geometry (SELLEY; SONNENBERG, 2014), mainly interpreting horizons in orthogonal seismic lines associated with seismic attributes extracted over the interpreted surfaces, in order to identify depositional facies and understand the distribution and connectivity of the sedimentary deposits as separate geobody units, which are imperative for the well planning during future development of the field (HAQUE; ISLAM; SHALABY, 2018). This task, however, is quite challenging since it requires the extraction of the most relevant information available and combines it with the complex knowledge of both geology and geophysics (HESTHAMMER; LANDRO; FOSSEN, 2001).

Seismic facies mapping is a fundamental process to identify geological patterns and can be treated as an image segmentation problem, reducing the tedious and time consuming manual interpretation (DI; WANG; ALREGIB, 2018). One main goal is to automatically model the reservoir properties from seismic data. This characterization, if automatically performed in large seismic datasets, can result in important time savings during the process of interpretation (RAMIREZ; LARRAZABAL; GONZáLEZ, 2015). Accurate prediction of reservoir presence and estimation of reservoir properties (net-to-gross, porosity, permeability, geometry, continuity, etc.) has been critical for the economic evaluation of an oil field.

Since sedimentary deposit may present itself in different scales and shapes, both are mainly controlled by the depositional gradient towards the basin (HESTHAMMER; LANDRO; FOSSEN, 2001), a geologist/geophysicist aims to look at all perspectives to have a confident classification of a deposit and/or its facies and avoid noise influence on the interpretation. However, it is much more feasible and agile for interpreters to inspect a huge seismic volume in 2D slices, or seismic sections, and that's how the majority of interpretation softwares work. In other words, in a real-world scenario the seismic data is manually interpreted in 2D images, rather than in a volumetric scale (XIONG et al., 2018), using all the prior geological knowledge of all the principles that rule the depositional system, including inferences regarding the 3D correlation with the adjacent slices.

3 RELATED WORK

The theoretical foundation of this research is divided between both Geology and Computer Science, which have very few publication merging those two subjects, specially in a 3D context. After first searching for AI application for geological spatial contexts, it was conducted a review of the latest studies regarding three dimensional CNN's for image segmentation, which were conducted mainly on medical images and video datasets.

Medical images are usually obtained through tomography, which measures a physical property contrast similar to a seismic data. Based on that, another survey used as filter parameters papers more recent than 2018 and containing the keyword "Medical image" and "image segmentation". Although the datasets were relatively large (around a few GB), they had a small number of classes and anotation compared to a seismic data. Additionally, most of the paper addressed the problem in a pure 3D or 2D approach, which would be either time-costly (for a 3D approach) or without enough spatial relevance (for a 2D approach). Based on this issue, another search was conducted looking for studies of 2D to 3D transfer learning.

3.1 3D IMAGE SEGMENTATION

According to the universal approximation theory, a neural network can approximate any continuous function with a sufficient number of parameters and training samples (HORNIK; STINCHCOMBE; WHITE, 1989). However, training a large number of network parameters requires a great amount of manually labeled real data, which usually is unfeasible for most researchers and interpreters in geophysics (SHENGRONG et al., 2019).

Convolutional Neural Networks (CNN's) achieve the better performance overall compared to other architectures when applied to seismic data (HUANG; DONG; CLEE, 2017; SHI; WU; FOMEL, 2019; Wang et al., 2018; ZHAO, 2018) and are also the most popular networks employed in medical image segmentation, with two main approaches (ZHAO, 2018): a pixel patch-based model (where the output from the network is a single value representing the facies label of the seismic sample at the center of the input patch) and an encoder-decoder model (where output is an image class labels with the same dimension as the input image).

The former has 2 disadvantages in 3D interpretation (SHI; WU; FOMEL, 2019): it requires a sliding cube/window to scan the whole dataset to assign a class to every centered pixel and it could have issues establishing the boundaries of geological objects. Therefore, the model proposed in this study is an encoder-decoder type, which is also the most popular in medical image segmentation experiments (YANG et al., 2021; Shan et al., 2018).

In their work (ZHAO, 2018) the authors state that the encoder-decoder model generates superior classification, despite the bigger training time and labeled data picking process. Besides, the encoder-decoder model output resulted in cleaner geobodies, such effect was correlated with the ability of the network of capturing the large scale spatial correlation between classes.

Similarly to seismic data, 3D medical images problems have a large size and very few labeled data. Researchers tackled those limitations by training a 2D network through slices of the dataset and then initializing a 3D counterpart of the previous network with the learned parameters (XIONG et al., 2018; YANG et al., 2021). This technique was applied by (Shan et al., 2018) to a medical image denoising problem, where a 3×3 convolutional filter is expanded to a $3 \times 3 \times 3$ kernel in a 3D network, with 3 adjacent slices with the same size as the input. The proposed network outperformed all the other networks in the experiment and led to the conclusion that a 3D network with 2D pre-training shows better results than its counterpart trained from scratch.

Moreover, the studies of Li et al. (LI et al., 2020a) the authors addressed the issue with the "black box" nature of most deep-learning methods, where there's very low interpretabillity of the results of the algorithm. The authors state that improving the interpretabillity between the results and the inputs is necessary to improvement for AI applications, especially in complex problems, such as geological context.

3.2   3D IMAGE SEGMENTATION THROUGH 2D PROBLEM REDUCTION

The works of (Wang et al., 2018) and (LIU et al., 2018) approached a 3D image classification problem by a prior simplification to a 2D context, taking into account the balance between receptive field, model complexity and memory consumption. (Wang et al., 2018) realized that a small receptive field induces the model to majorly use local features, whereas a larger receptive field focus more global features. The small receptive field has the downside of global features and the larger one has a much higher memory consumption, restricting the model complexity and representation ability. Based on those remarks, (LIU et al., 2018) proposed the 3D Anisotropic Hybrid Network (AH-Net) to learn features from medical images, based on the fact that the image resolution in XY plane/slice (or within-slice resolution) is higher (up to 10 times) than that of the Z resolution (or between-slice resolution). They use a ResNet backbone to train the data in a 2D context and then convert the 3×3 convolutions to $3 \times 3 \times 1$ and $1 \times 1 \times 3$ counterpart, so that the convolutions are more suited to extract the features within slices and between slices, respectively (Figure 9), which has shown better results than other 3D networks in a liver lesion segmentation task (Figure 10). Moreover, the authors suggest that more designs of the anisotropic kernels might yield a performance increase (LIU et al., 2018), since their approach was based on the different resolution between the slices depending of the axis. Regarding this, an expert knowledge might be helpful for the network design. In this thesis, a geological inference was taken into account during the algorithm architecture conception.

In the (YANG et al., 2021) work, it is proposed a 2D to 3D Transfer-learning, similar to AH-Nets, where each ACS kernel (axial, coronal and sagital), extracted in three orthogonal views from the data, are concatenated in the channel dimension and operate at the same time in the 3D input (Figure 11). This approach allows any 2D network architecture to be seamlessly converted to an ACS equivalent (YANG et al., 2021).
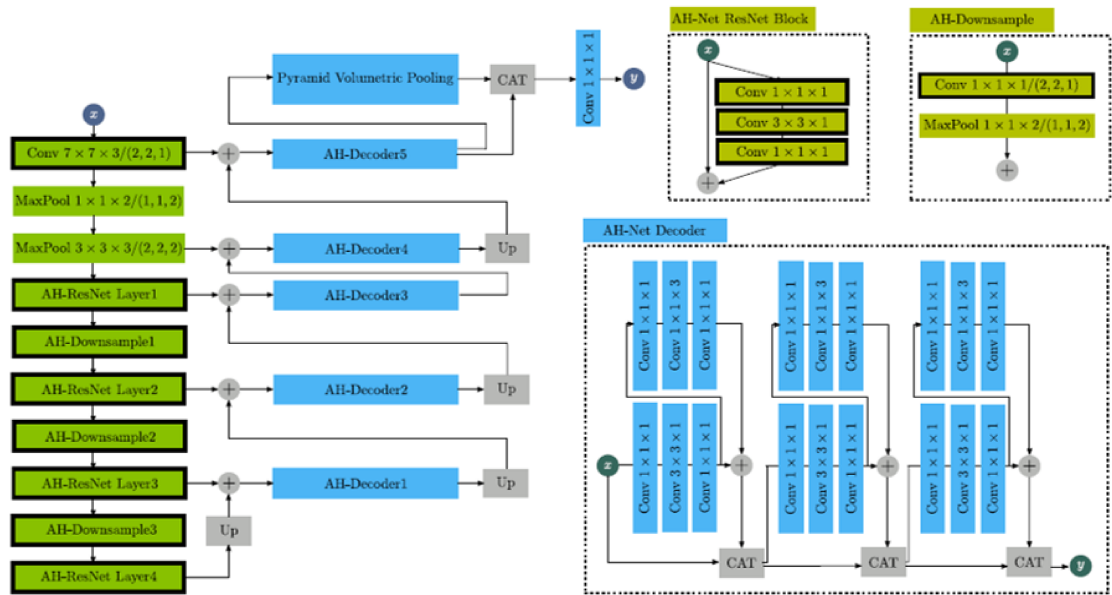
Figure 9 – The architecture of 3D AH-Net. Feature summation is used instead of concatenation to support more feature maps with less memory consumption. The weights of the blocks with black borders are transformed from the 2D MC-GCN. Source (LIU et al., 2018)

.

Method	Lesion		Liver	
	DG	DPC	DG	DPC
leHealth	0.794	**0.702**	0.964	0.961
H-DenseNet	0.829	0.686	0.965	0.961
hans.meine	0.796	0.676	0.963	0.960
medical	0.783	0.661	0.951	0.951
deepX	0.820	0.657	0.967	**0.963**
superAI	0.814	0.674	-	-
GCN	0.788	0.593	0.963	0.951
3D AH-Net	**0.834**	0.634	**0.970**	**0.963**

Figure 10 – The liver lesion segmentation (LITS) challenge result with the dice global (DG) and dice per case (DPC). Source (LIU et al., 2018).

In their work, the authors compared the performance of different architectures with and without the 2D to 3D Transfer-learning, where it was observed that the Transfer-learning approach achieved superior accuracy in lung nodules segmentation tasks, even better than other 3D networks with pre-training 12. However, in the proof-of-concept experiment, the kernels are extracted in the same view from the data, rotating this kernel to match the three orthogonal views from the data.

The authors claim that any view of a 3D image is still a 2D spatial image, making it unnecessary an axis assignment for the learned parameters. Such inference might be valid for some contexts but that does not suit well to an anisotropic image dataset and geobodies, like seismic facies.

Another important aspect of the results of the experiments conducted by (YANG et al., 2021) is the comparison between the perfomance of the 3D, a semantic image segmentation neural network model proposed by Chen et al. (CHEN et al., 2018), trained from scratch (3D

DeepLab r.) with their counterparts. The random initialized network outperformed the other 3 networks with weight transfer (I3D, MED3D and Video), showing that transfer-learning does not always guarantee a better performance.
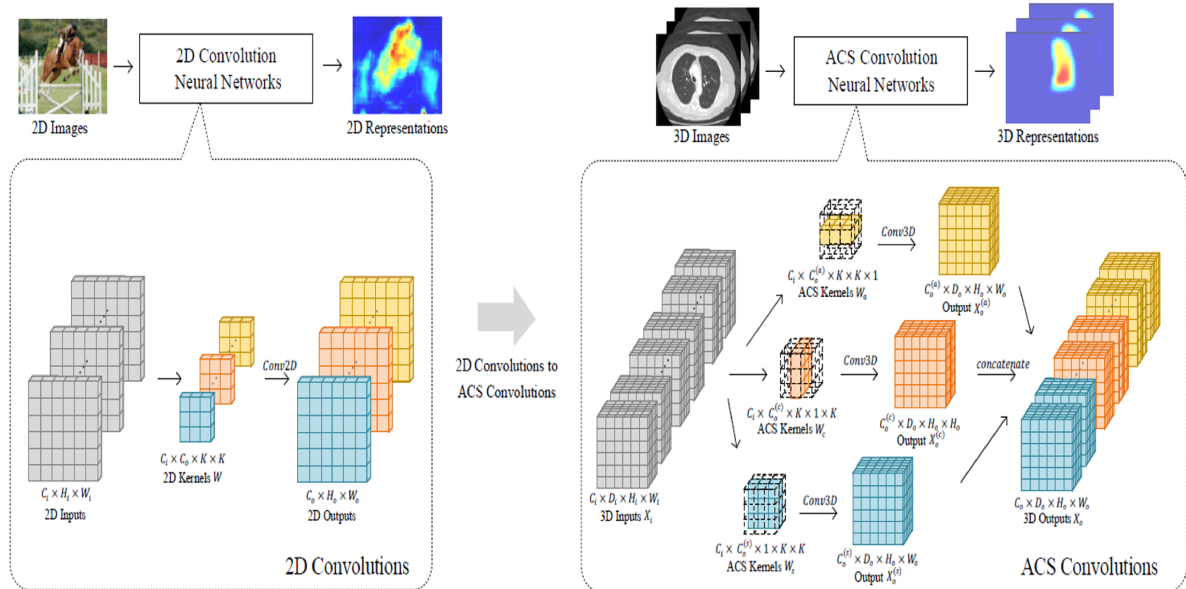
Figure 11 – Demonstration of how any 2D convolutions parameters are transfered to its ACS convolutions model counterpart. Source (YANG et al., 2021)

Models	Lesion		Liver	
	Dice Global	Dice per Case	Dice Global	Dice per Case
H-DenseUNet	82.4	72.2	96.5	96.1
Models Genesis	-	-	-	91.13 ±1.51
P₃SC₁ r.	72.3	59.0	93.9	94.2
2.5D DeepLab r.	72.6	56.7	93.1	92.7
2.5D DeepLab p.	73.3	59.8	92.9	92.0
3D DeepLab r.	75.3	62.2	94.8	94.8
3D DeepLab p. I3D	76.4	57.7	94.1	93.4
3D DeepLab p. Med3D	66.8	53.9	92.0	93.6
3D DeepLab p. Video	65.2	55.8	92.5	93.2
ACS DeepLab r.	75.2	62.1	95.0	94.9
ACS DeepLab p.	**78.9**	**65.3**	**96.7**	**96.2**

Figure 12 – Performance comparison between several architectures in 2D, 2,5D and 3D, with (.p) or without pre-training (.r), where the ACS variants, with and without transfer-learning, outperformed most of the networks in the experiment. Source (YANG et al., 2021)

3.3 AI APPLIED TO SEISMIC

In seismic interpretation and seismic data analysis, it is imperative to effectively identify certain geologic formations from very large seismic datasets (DI; WANG; ALREGIB, 2018). Machine learning is based on parallel processing used for approximation, clustering and pattern recognition purposes in large multivariant datasets (MOHAMED et al., 2019) and can explore the hidden connections among different physical quantities. Artificial intelligence applications in the oil and gas industry have been reported since the early '90s. Most of them, however, have been majorly in lithology prediction with well logs, using techniques such as SVM (Support Vector machines) (AN-NAN; LU, 2009), fuzzy logic (LAMMOGLIA; OLIVEIRA; FILHO, 2014) and random forest (BESTAGINI; LIPARI; TUBARO, 2017).

The application of geophysics was mainly applied in the exploration to locate the hydrocarbon reservoirs and to evaluate potential targets for drilling. More recently the the evaluation of optimum development and more accurate drilling locations have been increasingly emphasized, since it maximize the start-up production rate by drilling the sweetest spots of the reservoir first (AMINZADEH; DASGUPTA, 2013).

Even though many deep learning techniques are suitable for geological problems, the high demand for labeled data restricts their application in a real-world scenario with limited annotations (YANG et al., 2021). The main issue regarding geological modeling of an oil field is to spatially correlate the knowledge obtained through the well to the seismic volumes, since well data is rare and sparse (ZHAO, 2018).

Recently, more studies have been focused on algorithms to make predictions in a 3D volume, like the study developed by Shi et al. (SHI; WU; FOMEL, 2019), which aimed to identify salt boundaries in the SEAM 3D seismic dataset (FEHLER; KELIHER, 2011), using an encoder-decoder architecture (Figure 13). This methodology showed good results in salt bodies segmentation, both in synthetic and real data. However, the author pondered about the high computational cost of such method, since the input had dimensions 128x128x128, limiting the batch size of training to only 2 samples.

More recently, the studies of Alaudah et al. (ALAUDAH et al., 2019) proposed a benchmark seismic facies classifier based on the F3-block dataset (BARONI et al., 2018). The authors employed a 2D classifier using UNet as the network architecture, for both patch-based (32x32 images) and section based models (each inline/crossline as a sample). Comparing the perfomance on both patch and section-based samples is a relevant analysis, since the patch-based models are more feasible memory-wise compared to section-based ones, since a section can easily be hundreds of times bigger than a single patch, limiting the batch size of the experiments (Figure 14). Albeit more memory-consuming, the authors pointed out that section-based experiments are able to learn the spatial correlation between the classes, which yielded better results in their study, even though the main goal was to propose a benchmark classifier. The experiments showed good overall results, they were obtained on 2D samples, without any 3D spatial information taken into account. Even better results were obtained through data augmen-
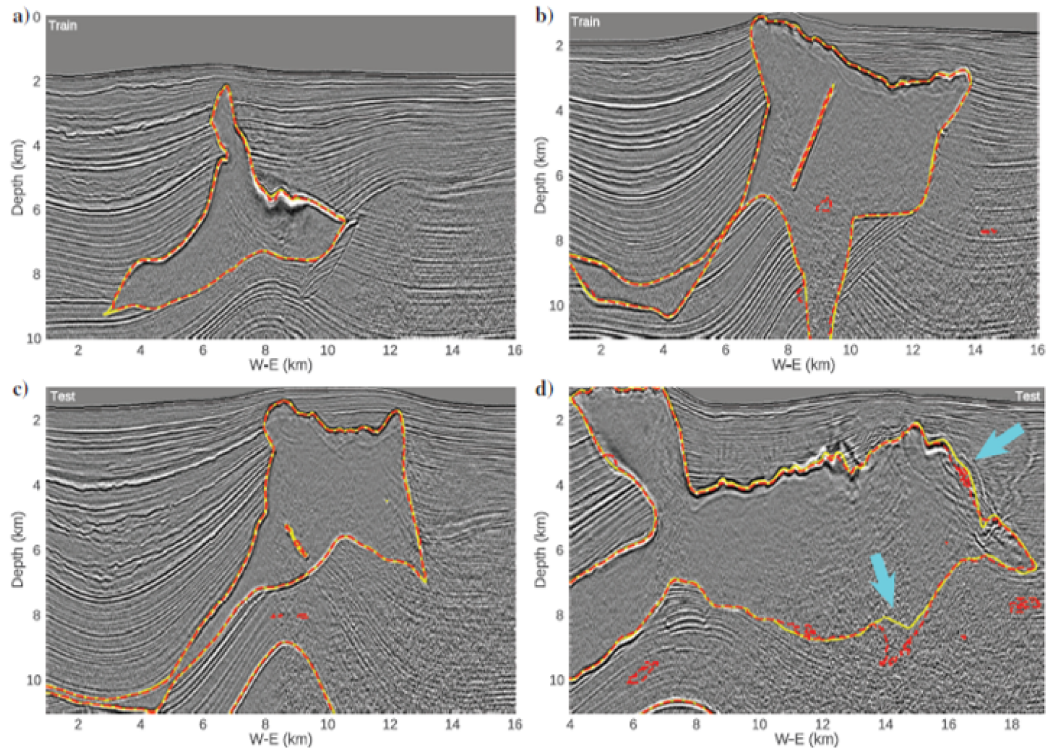
Figure 13 – Results of the 3D network proposed by Shi et al. applied to salt body segmentation (red dashed= network prediction, solid yellow curve= ground truth, a and b= training set, c and d= testing set). Source (SHI; WU; FOMEL, 2019).

tation, like flipping and rotating each sample, since the class transitions could be steeper where the halokinesis (the movement of salt bodies) was more prominent. On the downside, the results were obtained only using the seismic amplitude as a feature, not taking into account any other seismic attribute or geological insight.

Due to the orientation of the geobodies and the direction of the seismic acquisition, the same feature might have different resolution/quality depending on the view of the interpreter. Processing artifacts and/or seismic acquisition noises influence may vary between slices, especially in high amplitude contrasts and steep reflectors (HESTHAMMER; LANDRO; FOSSEN, 2001), which are not taken into account in a 2D approach.

In contrast, a full 3D training has a much higher computational cost and greatly reduce the training and test batch size. Therefore, this thesis proposes an approach that aims to incorporate 3D context but at a computational complexity of the same order as a 2D network (Table 1).

Table 1 – Table comparing the networks from the literature compared to the proposed method (AOE Network).

	Data Samples dimension	Transfer-learning	Orthogonal Approach	Applied to Seismic data
(LIU et al., 2018)	2D and 3D	Yes	Yes	No
(SHI; WU; FOMEL, 2019)	3D	No	No	Yes
(YANG et al., 2021)	2D and 3D	Yes	Yes	No
(ALAUDAH et al., 2019)	2D	No	No	Yes
Proposed Network	2D and 3D	Yes	Yes	Yes

| Metric | PA | Class Accuracy | | | | | | MCA | FWIU |
Model		Zechstein	Scruff	Rijnland/Chalk	Lower N. S.	Middle N. S.	Upper N. S.		
Patch-based model	0.788	0.264	0.074	0.499	0.992	0.804	0.754	0.565	0.640
Patch-based + aug.	0.852	0.434	0.221	0.707	0.974	0.884	0.916	0.689	0.743
Patch-based + aug + skip	0.862	0.458	0.286	0.673	0.974	0.912	0.926	0.705	0.757
Section-based model	0.879	0.219	0.539	0.744	0.951	0.872	0.973	0.716	0.789
Section-based + aug.	0.901	0.714	0.423	0.812	0.979	0.940	0.956	0.804	0.844
Section-based + aug + skip	0.905	0.602	0.674	0.772	0.941	0.938	0.974	0.817	0.832

(a) Seismic data

(b) Ground truth labels

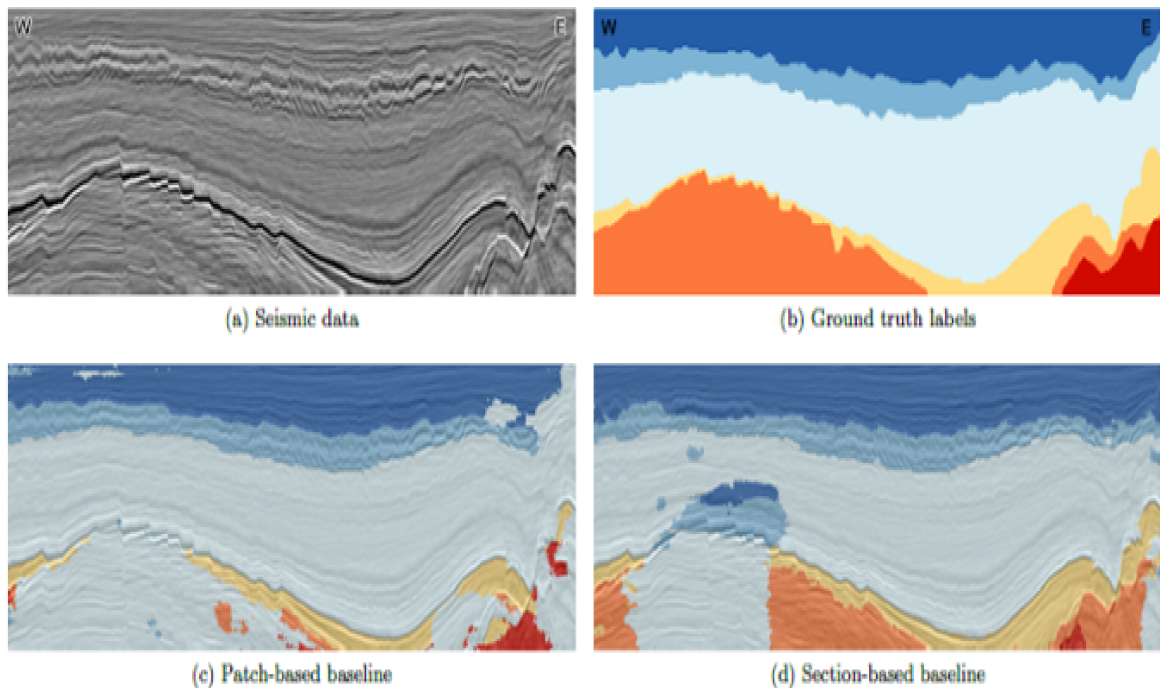(c) Patch-based baseline

(d) Section-based baseline

Figure 14 – Results of the benchmark network proposed by (ALAUDAH et al., 2019) applied to the F3-block dataset. The results show a better accuracy in shallower layers compared to the deeper ones, notably the salt diapirs (Zechstein Group) and Scruff Group. Source (ALAUDAH et al., 2019)

4 METHODOLOGY

This section describes the steps and the relevant information regarding the issues and benefits taken into account during the experiments, until the conception of the proposed network.

4.1 DATASETS DESCRIPTION

The proposed method was tested on 2 of the most widely used 3D seismic volume available: The synthetic model of Stanford VI (CASTRO; CAERS; MUKERJI, 2005) and the real seismic data of the F3-block (BARONI et al., 2018). Both datasets, although with the purpose of segmentation and classification of geological bodies, have different scales and class distribution.

4.1.1 Stanford VI Dataset

The reservoir is 3.75 km wide (East-West) and 5.0 Km long (North-South), with a shallowest top depth of 2.5 Km and deepest top depth of 2.7 km. The reservoir is 200 m thick and consists of three layers with thicknesses of 80 m, 40 m, and 80 m respectively. In terms of a grid, the Stanford VI-E reservoir is represented in a 3D regular stratigraphic grid of $150 \times 200 \times 200$ cells and the dimensions of the grid correspond to 25 m in the x and y directions and 1 m in the z direction (CASTRO; CAERS; MUKERJI, 2005).

The stratigraphy of the Stanford VI-E reservoir (Figure 15) corresponds to a prograding fluvial channel system, where deltaic deposits represented in layer 3 were deposited first and followed by meandering channels in layer 2 and sinuous channels in layer 1 (Figure 15). The model contains 4 different facies: floodplain (shale deposits), point bar (sand deposits that occur along the convex inner edges of the meanders of channels), the channel (sand deposits), and the boundary (shale deposits) (CASTRO; CAERS; MUKERJI, 2005). Each of those classes have different mineralogy and petrological characteristics that reflect the different mechanisms responsible for their deposition (flow strength, declivity, river depth). This difference yield a different physical properties and, by consequence, a different seismic response.

4.1.2 F3-block Dataset

The F3-block is an area located on the shelf of the North Sea in the Netherlands (BA-RONI et al., 2018). The North Sea is rich in hydrocarbon deposits, which is why this area is very well examined (thanks of the drilling and seismic profiling) and described in the literature (BARONI et al., 2018). Within the shelf of the North Sea Group ten geological units have been distinguished. Alaudah et al. (ALAUDAH et al., 2019) based on the data of 26 wells and 40 different interpreted horizons, merged the various lithostratigraphic described in the literature
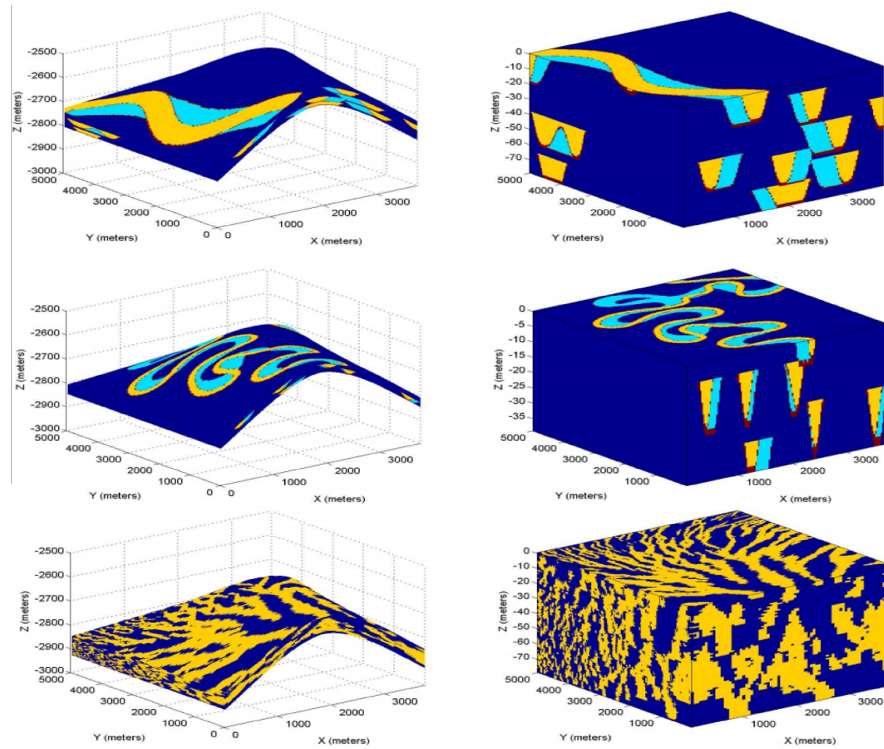
Figure 15 – Overview of Stanford VI-E reservoir facies model. Notice the high imbalance in class frequency and the significant variance between each cube view.
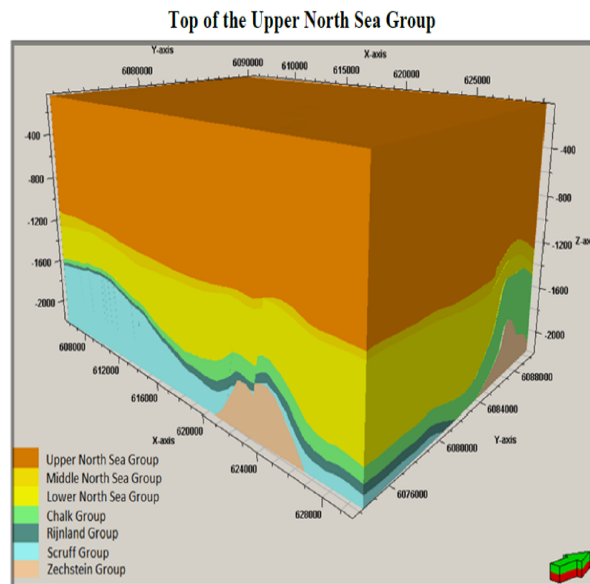
Figure 16 – Overview of F3-Block dataset class distribution. Source (BARONI et al., 2018)

and assigned them in 6 different classes (from the oldest one): Zechstein Group, Scruff Groups Rijnland/Chalk Group, Lower and Middle North Sea groups, Upper North Sea Group (DUIN et al., 2006). As shown in Figure 16, the classes of F3-block are more continuous than the ones from the Stanford VI dataset, especially laterally. Such a pattern is expected since the class correspond to lithostratigraphic groups, which are more regional features than sedimentary facies.

4.2   DESCRIPTION OF THE PROPOSED ARCHITECTURE

According to the studies of O'Mahony (O'Mahony et al., 2018), the simplification of a 3D model into multiple 2D views is a straightforward idea that allows efficient convolutional and pooling operations and deep networks pre-trained on large labeled image datasets to be leveraged. Such approach is the same in the oil and gas industry, where interpreters inspect a seismic data through 2D sections in several directions, mainly orthogonal lines (cross-lines and in-lines) and a view from the top (time-slices) since a full interpretation in a 3D volume is not very feasible. Due to the subjetive nature and uncertainty in the geology field, the individual interpretation of each specialist is often checked by other geologist, in order to have another point of view of the same task.

Based on those remarks, and on the downsides of other approaches in the literature, the AOE Network network is proposed in this study (Figure 17). It was employed an UNet-like architecture (LECUN; BENGIO; HINTON, 2015) as the backbone of our AOE Network network. Although the UNet was especially designed for image segmentation tasks in medical images, it performs well not only on biomedical segmentation but also on many kinds of semantic segmentation tasks, such as satellite image (LIU et al., 2018). This methodology has already been applied with success, showing better overall performance and with a lower computational cost (YANG et al., 2021). The present work was mainly based on the methodology proposed by both (YANG et al., 2021) and (LIU et al., 2018).
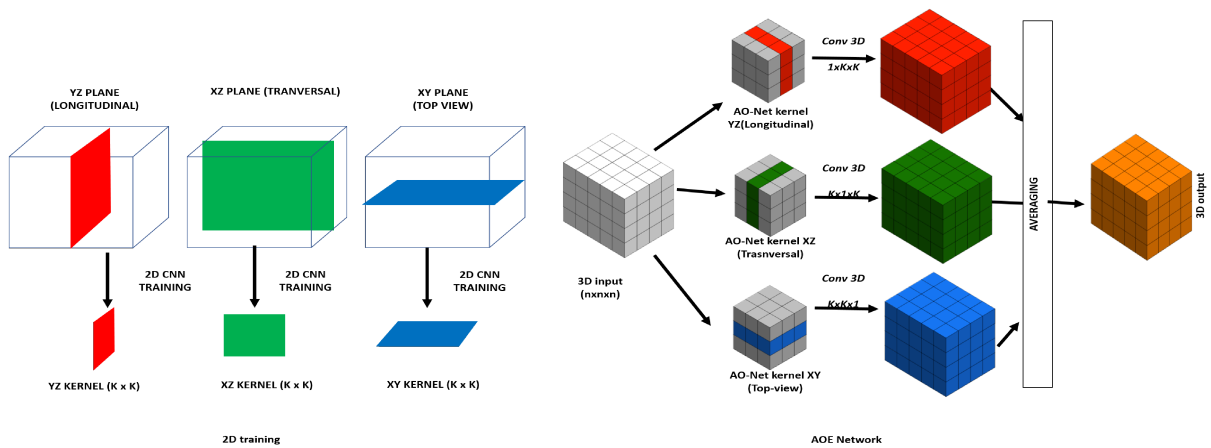
Figure 17 – Design of the proposed AOE Network network, which accepts any 2D model to be transferred to its 3D counterpart.

4.2.1   Data sampling

In the oil industry, the drilled well is the main source of a real data during the exploration phase, since seismic and other methods of bigger coverage are indirect measurements of the subsurface physical properties. The well information is very spatially limited, which requires methods of extrapolation of the physical and geological information found. Thus, most

of the properties models of an area are spatial extrapolations from the information of well logs, core and fluid samples and physical properties extracted from those wells.

The samples for our training set were obtained using this methodology, with random selection of 10 location (random x and y coordinates), where the 3D cubes centered in those locations were extracted along the z dimension with a 50 % overlap between adjacent cubes. The work of Guazzeli et al. (GUAZZELLI; ROISENBERG; RODRIGUES, 2020) proposed a classifier based on this approach that a 3D cube could be simplified by three orthogonal planes (Figure 18), which significantly decreased the memory usage with a good performance. In similar strategy, the samples of the 2D training set were extracted, even though the 3D samples were also sampled in our network.
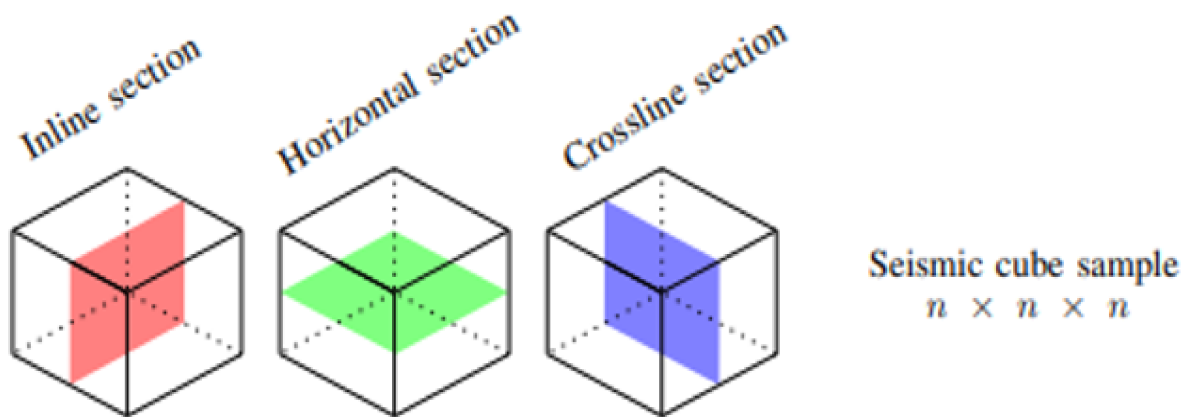
Figure 18 – 3D cube simplification in three orthogonal planes for the 2D training. Adapted from (GUAZZELLI; ROISENBERG; RODRIGUES, 2020)

The whole 3D volume is also divided in sub-volumes, of dimensions $32 \times 32 \times 32$, to be split between training, test/validation set. The training set is composed by 30 % of the samples and the test/validation set by the other 70 %, both randomly selected. Since the datasets dimension may not be divisible by the samples dimensions (32x32x32), a padding operation was conducted on the edges of the data, which was subtracted in the final metrics calculation.

4.2.2   Network backbone

As already mentioned, this methodology can be implemented with different network architectures. The UNet was the network selected as the backbone of our algorithm, since it was developed for medical image tasks, which are based on similar physical principles as the seismic. The UNet (RONNEBERGER; FISCHER; BROX, 2015) is based on a fully-connected network, supplementing a usual contracting network by successive layers, where pooling operators are replaced by upsampling operators (Figure 19).

The contracting path (or encoder) is a a traditional stack of convolutional and max pooling layers to capture the context in the image, by encoding it into feature representations at multiple different levels.

The expansive path (or decoder) approximately symmetric to the contracting path and semantically projects the features (lower resolution) learnt by the encoder onto the pixel space (higher resolution) to get a dense classification. The main operations are upsampling and concatenation followed by regular convolution operations.
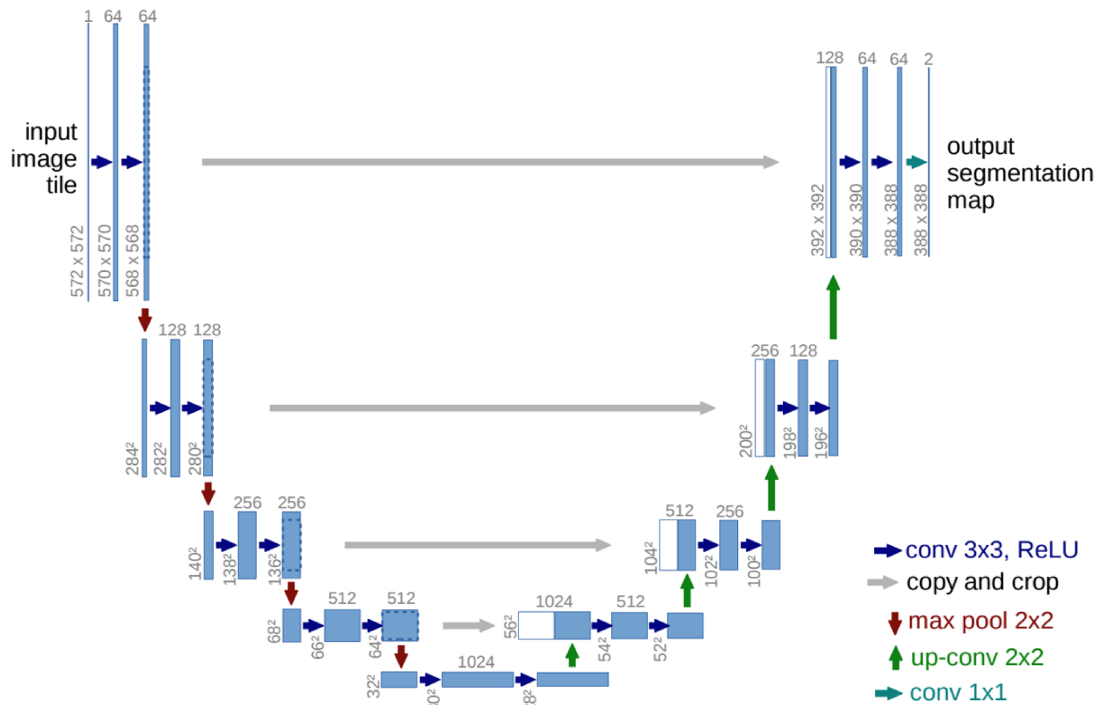
Figure 19 – UNet architecture for a 32x32 pixels image. The number of channels is denoted on top of the box.(Blue box = multi-channel feature map, White boxes = copied feature maps, Source = (RONNEBERGER; FISCHER; BROX, 2015))

4.2.3 Network architecture

The first step of the proposed network consists of a 2D network trained from scratch, with the inputs being images extracted in 3 orthogonal slices (Longitudinal, Transverse and Top views) of a sub-volume training sample, each one individually (Figure 17). This approach is very similar to the daily routine of an interpreter of an oil company, where the interpreter evaluates the data through 2D seismic section and maps, extracted from 3D volumes, since it is more feasible. In this case, the network resembles three interpreters (geologist and/or geophysicist) trying to identify facies using only one point of view. Since the samples size are smaller, the network trained on all samples for 50 epochs, which takes roughly 10 minutes.

Afterwards, the pre-trained weights are transferred into the hybrid 3D counterpart of the 2D network used during training. Each orthogonal view is "inflated" in one dimension, perpendicular to the plane (1xKxK, KxKx1, 1xKxK, where K is the filter size), also called ACS kernels (YANG et al., 2021). This transfer-learning allows the network to capture 3D spatial

information without a relevant increment of memory usage, with the filters being of the same order of complexity of regular 2D networks.

Additionally, the correct positioning of the weights to its orthogonal positions is fundamental to assign each of the trained parameters to their respective views of the data. This was inspired by the AH-Net proposed by (LIU et al., 2018), although their original motivation for this approach was mainly the resolution discrepancy between intra-slice and inter-slice of medical images.

The dataset is also sampled in 3D cubes, which are convolved by those ACS kernels in three different parts, one for each of the orthogonal views of the 2D pre-training (Figure 17). This section works as if 3 different interpreters look at the same data through different views (cross-lines, inlines and top-view), with their own conclusions (or classes), at first, independent from each other. At the last layer of the network, the output of the 3 ACS convolutions is the average of the probabilities for each class, which is the final output of the proposed AOE Network Network.

With this weight transfer, the 3D part of the network does not need to random initialize its weights, which leads to a faster convergence. It also offers a great flexibility, since it can be seamlessly applied to any network architecture.

4.2.4 Metrics

In order to properly compare the results obtained in this research with other approaches in the literature, it was proposed to evaluate the performance of our network based on the following metrics:

- Pixel Accuracy(PA): The percent of pixels in the image (or of each class) which were correctly classified.

  Pixel Accuracy(PA) =
  $$\frac{TP + TN}{TP + TN + FP + FN}$$

  (TP= True Positives, TN= True Negatives, FP= False Positives, FN= False Negatives)

- Mean Class Accuracy(MCA): The mean of the Pixel Accuracy (PA) of each class.

  Mean Class Accuracy(MCA)=
  $$\frac{1}{nc} \sum_{i} PA_i$$

  (nc= number of classes, i= class index)

- F1-Score/Dice: The harmonic mean of precision and recall.

  F1-Score/Dice =
  $$\frac{TP}{TP + 0.5(FP + FN)}$$

- Frequency Weighted Intersection over Union(FWIU): Intersection over Union is defined as the number of elements of the intersection of F_i (the set of pixels classified as class i) and G_i (set of pixels that belong to class i) over the number of elements of their union set.

Intersection over Union(FWIU) =

$$\frac{F_i \cap G_i}{F_i \cup G_i}$$

In order to prevent this metric from being overly sensitive to small classes, each class is pondered by their frequency.

Frequency Weighted Intersection over Union (FWIU)=

$$\frac{1}{\sum G_i} \cdot \sum_i G_i \cdot \frac{F_i \cap G_i}{F_i \cup G_i}$$

4.2.5 Implementation

All the routines were implemented in PyTorch and the experiments were conducted using a Desktop with a Intel(R) Core(TM) i7-8700K CPU @ 3.70GHz processor, with 64GB RAM and a NVIDIA GeForce GTX 1080 Ti 12GB GPU. The code is available at https://github.com/eltontrindade/F3-block-dataset.

5 EXPERIMENTS

5.1 EXPERIMENTS ON THE STANFORD VI DATASET

In a attempt to verify the feasibility of the network, the experiments first investigated how each view of the sample (Longitudinal, Transverse and Top view) performed when used individually during the image segmentation task in the Stanford VI dataset. Using the UNet architecture as the backbone, this thesis implemented the following networks:

- UNet 2D: A 2D network using UNet as the backbone was executed to obtain the 2D weights for the transfer-learning.

- ACS Network: with a 3D input $X_i \in R^{C_i \times D_i \times H_i \times W_i}$ and a 3D output $X_o \in R^{C_o \times D_o \times H_o \times W_o}$, and 2D filters $W \in R^{C_o X C_i \times K \times K}$, the filters are divided and rearranged in 3 (three) parts (ACS kernels) by the channel output, in order to obtain the 3D representation associated to each view plan: $W_a \in R^{C_{o(a)} \times C_i \times K \times K \times 1}$, $W_c \in R^{C_{o(c)} \times C_i \times K \times 1 \times K}$, $W_s \in R^{C_{o(s)} \times C_i \times 1 \times K \times K}$, with $Co_{(a)} + Co_{(c)} + Co_{(s)} = C_{(o)}$ (C= Channel, K= kernel dimension, D $\times H \times W$ = data dimension, i = input, o = output). The inputs are convolved with each ACS kernel and are concatenated by the channel to generate the output X_o.

- 2,5D Network : Using the same approach proposed by Liu et al. (LIU et al., 2018) for the AH-Net like networks, each ACS convolution is replaced by convolutions with $1 \times K \times K$ kernels.

- 3D UNet: The same architecture of the classic UNet, where, instead of ACS convolutions or the previous approach, fully 3D convolutionals were employed, without any weight transfer.

The Table 2 shows the comparison between the several networks complexity and memory-wise.

The experiments consisted of randomly selected 10 coordinates (x,y) in the Stanford VI-E dataset and extracting cubes of dimensions 32x32x32 along the z-axis, centered in those coordinates, which compose the 3D training set. The 2D training data are the 3 orthogonal planes (Longitudinal, Transverse and Top view) of the 3D training set cubes. The whole dataset is used as the test set, including the 3D training set.

For a comparative analysis, we also train a 3D version of the UNet but randomly initialized. To verify if our network combines the prediction of each AH-Net of the Ensemble, we also compared the results with the ones from each network separately. As already mentioned, the dataset is not balanced, so in order to mitigate that issue, the F1-Score/Dice per batch was used as the metric for the model adjustment. As the studies of Sudre et al. (SUDRE et al., 2017) have shown, this metric performs well even with imbalanced datasets, since it penalizes the false positives predictions.

Table 2 – Time and Space complexity comparison between the networks.(C_o,C_i, are the number of channels of output and input respectively, D=Depth,H=Height,W=Width of input and K the kernel size)

Network	Time Complexity	Space Complexity
2D UNet	$O(DHWC_oC_iK^2)$	$DHWC_o$
AOE Network	$O(DHWC_oC_iK^2)$	$DHWC_o$
3D UNet	$O(DHWC_oC_iK^3)$	$DHWC_o$

Regarding the metric for the models evaluation, it was adopted the 2 most used ones in the literature: Frequency Weighted Intersection Over Union (FWIU) and Dice/F1-Score, both weighted by the number of samples of each class. The batch-size during training was 64 and 4, for the 2D and 2,5D/3D training respectively, and the test set batch size for the 2,5D/3D equal to 8. The remaining parameters were the same for both networks (learning rate=0.001, filter size=3, momentum=0.9, number of epochs=50).

5.1.1 2D Networks

The results of the 2D training already show that each view of the samples has a very distinct performance from one another (Figure 20). The network fed with the Top view planes performed significantly worse than the other two, while the network with the Transverse plane samples was the best one. Besides having the better performance overall, the networks also have different proficiency in classification regarding each class. For instance, the Transverse network, even though has the best performance overall, has slightly worse accuracy for the "channel" facies (Figure 20) compared to the Longitudinal network.

This divergence may be a consequence of the environmental conditions that controls their deposition and by consequence their shapes. For instance, in a section parallel to the biggest slope of the basin, mass flow deposits and turbidity currents should form a elongated shaped deposits, whereas in a perpendicular view those deposits should be thicker and of limited lateral continuity (Figure 21). Besides, since each network works on only a 2D from the data, each of the 2D slices may have a different class frequency and a different class transition from each other (a 2D section in the middle of the river, along its axis, would rarely show a floodplain or other deposits that usually deposit far from the river, for example).

Such information might suggest that a simple combination of the predictions might

yield even better results since the knowledge of the networks could be complementary.

Precision: 0.8032959343025907
Recall: 0.7513627485795454
F1-Score: 0.7694348857528215
IOU: 0.6442203284560122

Axial AH-Net (Longitudinal)
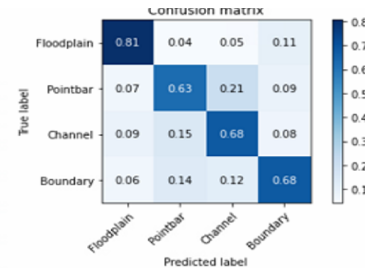
Precision: 0.8494678289989326
Recall: 0.8240700461647728
F1-Score: 0.8332770330433932
IOU: 0.7322612956011058

Coronal AH-Net (Transversal)
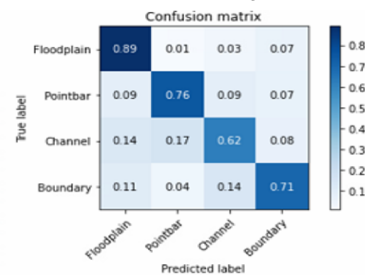
Precision: 0.7394677121838487
Recall: 0.6816350763494318
F1-Score: 0.7029342088000528
IOU: 0.570873091529111

Sagital AH-Net (Top View)
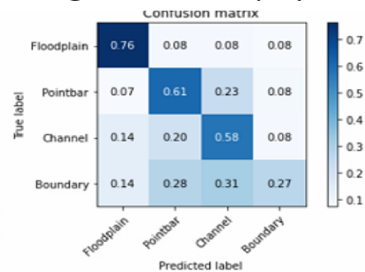
Figure 20 – Confusion Matrix of the 2D UNet for each orthogonal plane.

5.1.2   3D UNet

In the full 3D context training the networks, as expected, obtained much better results in the facies segmentation problem (Figure 22) than its 2D version. The UNet randomly initialized obtained good results (IoU=95,1% and F1-Score =97,4%), especially in identifying the class "floodplain" (99% accuracy) and "channel" (97% accuracy). Such result was expected, since geobodies have a complex 3D distribution and more information was given to the network (32 $\times$ more pixels per sample).
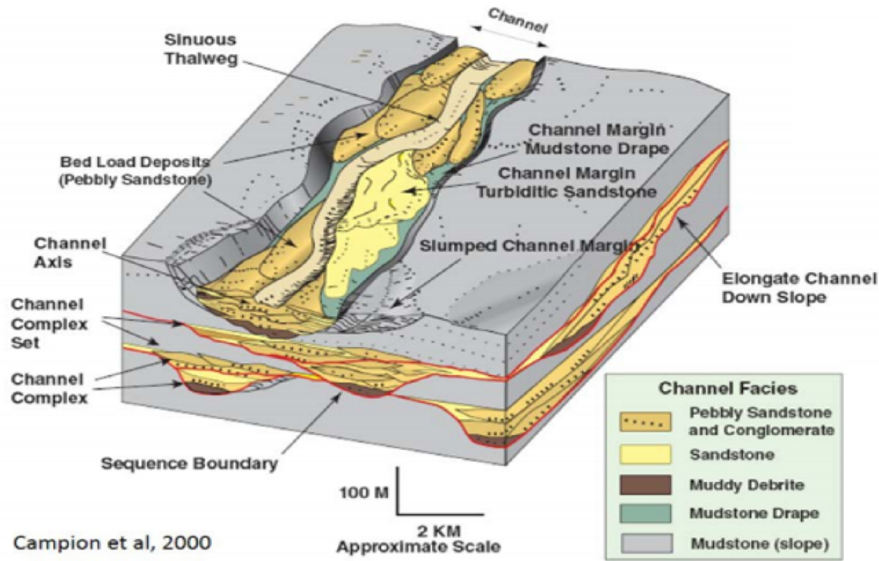
Figure 21 – 3D projection of a confined channel complexes, showing how drastically the shapes and facies frequencies and association vary in a depositional system depending on the point-of- view
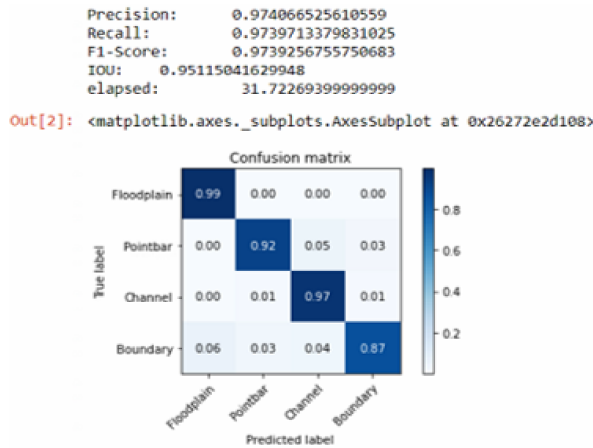
Figure 22 – Confusion Matrix of the 3D UNet for the classification of Stanford VI reservoir.

5.1.3 AH-Net

The performance of the Transfer-learning for AH-Net versions of each orthogonal plane individually was also evaluated. The three networks greatly outperformed its 2D counterpart, achieving over 95% on both IoU and F1-Score in the Transverse and Longitudinal networks, as shown in Figure 23. As in the 2D training, the Top view network fared worse, but it still obtained and IoU of 89.6% and an F1-Score of 94.2%. Additionally, two of the networks even slightly surpassed the 3D UNet in performance (approximately 0.5% better in both metrics) and roughly three times faster. Table 3 has a broader comparison, with the accuracy of each network for each class.

Precision: 0.9763803251663276
Recall: 0.9760816169507576
F1-Score: 0.9761786086428037
IOU: 0.9552394036144002

Axial AH-Net (Longitudinal)

Confusion matrix

	Floodplain	Pointbar	Channel	Boundary
Floodplain	0.99	0.00	0.00	0.01
Pointbar	0.00	0.96	0.01	0.02
Channel	0.01	0.02	0.95	0.02
Boundary	0.06	0.04	0.02	0.88
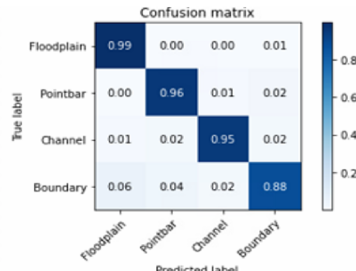
Precision: 0.9761267926520222
Recall: 0.9759863651160038
F1-Score: 0.9760431937664539
IOU: 0.9551501984944583

Coronal AH-Net (Transversal)

Confusion matrix

	Floodplain	Pointbar	Channel	Boundary
Floodplain	0.99	0.00	0.00	0.01
Pointbar	0.00	0.95	0.02	0.02
Channel	0.01	0.01	0.97	0.01
Boundary	0.06	0.04	0.04	0.86
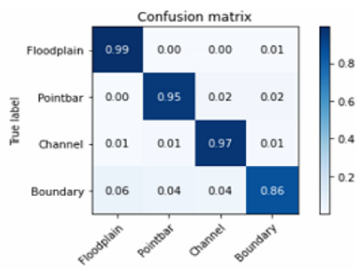
Precision: 0.9428749591479164
Recall: 0.9429053104285038
F1-Score: 0.9424351660019666
IOU: 0.8962367760193536

Sagital AH-Net (Top View)

Confusion matrix

	Floodplain	Pointbar	Channel	Boundary
Floodplain	0.99	0.00	0.00	0.01
Pointbar	0.05	0.86	0.04	0.05
Channel	0.07	0.03	0.88	0.03
Boundary	0.11	0.06	0.06	0.77
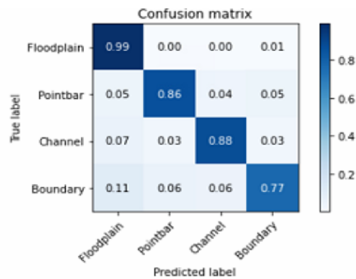
Figure 23 – Confusion Matrix of the AH-NET for the classification of Stanford VI reservoir.

5.1.4 AOE Network

Although the AH-Nets individually obtained a remarkable performance, we also combined those networks in an Ensemble architecture in order to increase the accuracy in the predictions taking into account the point of view of each one of them. The AOE Network results are exposed in Figure 25. The accuracy of every class improved upon the merging of the AH-Net outputs, showing that such approach is indeed a promising one. We also have the results of all networks compared based on the accuracy of each class of the dataset in Table 3. Figure 24 compares the prediction of the AOE Network with the ground truth in a 3D projection.

5.1.5 Overall analysis

The initial experiments showed that the combination of the prediction of three orthogonal views indeed results in better performance with a lower computational cost. This result shows that AI aplications to facies classification is a promising one since these lithofacies become stacked into stratigraphic units because the environments shift through time, with lithofacies accumulating along any given vertical axis. The nature of these environmental shifts is often predictable, which means that the resulting lithofacies successions are equally predictable (MIALL, 2015).

This conclusion was taken further in the future analysis, since this methodology should also be compared to other approaches in a real data.
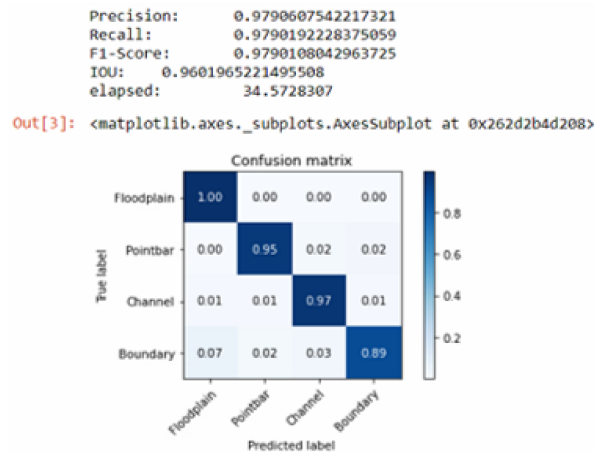
Figure 24 – Confusion Matrix of the AOE Network for the classification of Stanford VI reservoir.

Table 3 – Comparative results of the networks based on the accuracy for each class and overall F1-Score and IoU. The best results are highlighted in red.

	UNet3D	Longitudinal AH-Net (1xKxK)	Transverse AH-Net (Kx1xK)	Top view AH-Net (KxKx1)	AOE Network
FLOODPLAIN	0.990	0.990	0.990	0.990	**1.000**
POINTBAR	0.920	**0.960**	0.950	0.860	0.950
CHANNEL	**0.970**	0.950	**0.970**	0.880	**0.970**
BOUNDARY	0.870	0.880	0.860	0.770	**0.890**
Overall F1-Score	0.973	0.976	0.976	0.942	**0.979**
Overall IoU	0.951	0.955	0.955	0.896	**0.960**

5.2 EXPERIMENTS ON REAL SEISMIC DATA

After the initial promising results, another set of experiments were conducted, this time on a real world scenario of an oil field: the F3-block (BARONI et al., 2018). The experiments compared the performance of the proposed AOE Network with the 3D UNet, trained from scratch, and the results obtained by Alaudah et. al (ALAUDAH et al., 2019). So far the approaches were based on samples of a single feature: the seismic amplitude of dataset, which
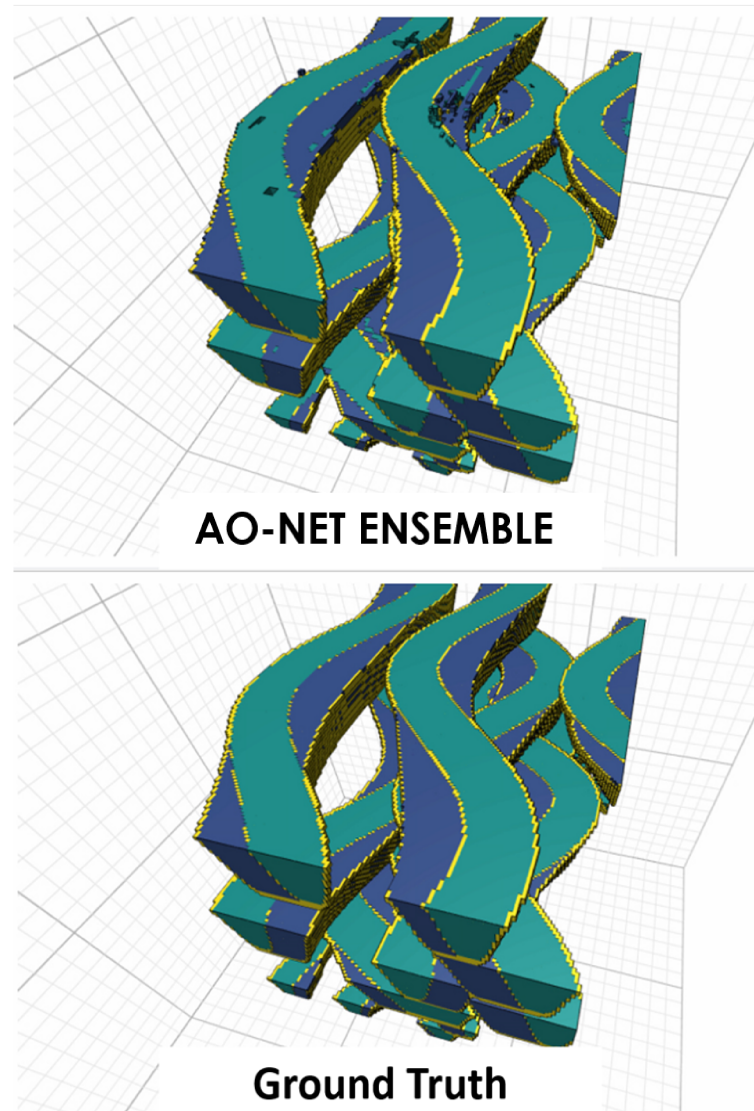
Figure 25 – Comparison between the AOE Network prediction and the ground truth.

is the most fundamental information in seismic interpretation. However, the class distribution is not merely random, since the classes of the F3-block dataset are chronostratigrafic units, which inherently have a spatial correlation between each other, especially vertically (or in depth). In other words, the relative age of the units have a important consequences in their association and proximity probability, which is an information that might benefit the network overall accuracy. On this section this improvement was investigated, pondering its higher cost with the accuracy increase.

The proposed method was also tested in the F3-block dataset (BARONI et al., 2018), since the studies of Alaudah et al. (ALAUDAH et al., 2019) established a benchmark performance using a geological model based on this data. Whereas the previous dataset was a synthetic reservoir model, the F3-block dataset is a real case of a seismic volume of a oil field.

The task in this experiment is to correctly classify the lithostratigraphic units of the data. In their paper (ALAUDAH et al., 2019), the authors recognized that a network fed with

samples of a whole seismic line (larger image) greatly outperforms a patch-based model, stating that such approach can easily learn the relationships between different classes and can take the depth information into account when labeling the samples. Since AOE Network is a patch based network, were conducted experiments with only the seismic as input and one with the addition of a depth feature to each pixel of the volume, in order to take that information into account during the segmentation. This approach is supported by the Law of Superposition, firstly proposed by Nicolas Steno, which states that in undeformed stratigraphic sequences, the oldest strata will be at the bottom of the sequence. Based on the Law of Superposition, depth was used as an additional feature for each sample, since the classes of the dataset were lithostratigrafic units with chronological relation with each other. Even though there are some geological deformation in the data, in general those classes relative positions are strongly related to their depth of occurrence. The experiments were conducted in the same data in which (ALAUDAH et al., 2019) proposed its benchmark classifier:

- 2D Training set: The F3-block dataset has 10 locations corresponding to actual wells drilled. On those (x,y) coordinates planes of dimensions 32x32 were sampled along the z-axis, centered in those coordinates, in which the 3 orthogonal planes (Longitudinal, Transverse, and Top view) were extracted, like in the previous experiment, totalling 525 samples for each orthogonal view.

- 3D Training set: Aiming to reproduce the experiments of the benchmark proposed by (ALAUDAH et al., 2019) the 3D training set includes all the data in the range of inlines [300,700] and crosslines [300,1000]. 399 voxels of dimensions 32x32x32 were extracted from the training set, using padding operation to cover the whole dataset and to avoid overlap between samples. The padding was set as equal to the border values of each dimension. Bigger voxels (64 and 128 dimensions) were experimented but did not obtain noticeably better results and implicated and much larger training time and memory limitations, since it restrained the batch size.

- 3D Test set: It's composed by the samples in the range of inlines [100,299] and crosslines [300,1000] (Test Set 1) and inlines [100,700] and crosslines [1001,1200] (Test set 2).

    The results were evaluated using the same metrics used by (ALAUDAH et al., 2019):

- PA (Pixel Accuracy): The percentage of accurate predictions over all samples;

- CA (Class Accuracy): The percentage of correctly predicted samples for each class;

- MCA(Mean Class Accuracy): The average Class Accuracy over all classes;

- FWIU (Frequence Weighted Intersection over Union): The IoU for each class, weighted by the size of its corresponding class.

The batch-size during training/test was the same as the experiments in the Stanford VI data set (2D= 64 for both training and test set; 2,5D/3D= 4 for training and 8 for test).

During the experimentation, better results were obtained with a filter size 7 compared to smaller ones (3 and 5). The remaining parameters used were a learning rate of 0.001, momentum of 0.9 and number of epochs of 50. It's worth mentioning that it was not employed any operation of data augmentation during the experiments.

5.2.1 Results

The experiments obtained a good result on the task, with superior performance when compared to the patch-based from Alaudah et al (ALAUDAH et al., 2019), although the section-based one obtained superior results in the 4 most frequent classes and in the FWIU and PA accuracy metrics (Tabela 4). The proposed network, however, had a better performance in the deeper and less frequent classes (over 30% better in some classes) and with only one class with accuracy below 50% (Scruff=46%), whereas the benchmark network obtained subpar results (below 30%).

Additionally, a network with both seismic and depth as a feature was also executed, on the same samples as the single channel one. That approach was aided by a geological insight, since the classed are chronostratigraphic units and have a vertical distribution strongly conditioned by age, as the Law of Superposition proposes, which means that the vertical succession of the units should probably match their age relation. It's worth mentioning that this feature was applied by using the depth of each sample as an approximation of this relative age, without any processing or other costly acquisition necessary.

The AOE Network networks with 2 features (seismic and depth) showed better performance overall and for most classes in the metrics selected (Table 4), showing that a geoheuristic is indeed a efficient approach. As expected, the less frequent classes showed a higher error rate, but the network still managed to have a good performance in more complex context. The results are far superior than the patch-based one from the benchmark (over 10% in PA and over 17% in FWIU) and slighlty better than the section-based one (around 1.4% and 2.5%, in PA and FWIU, respectively). Additionally the proposed method seems to be less biased by the imbalance of the dataset, showing a higher mean class accuracy (over 7% better than the benchmark model). These results emphasize the effect of the geoheuristic, based on the Law of Superposition, applied to the classifier modeling and the task domain evaluation. In Figure 26 the output from the network (with the depth and seismic as input features) is compared with the ground truth labels, where it's evident the potential of this approach, since even high deformed areas near the salt diapir (bottom-right) showed good accuracy.

A comparison with a native 3D counterpart, like in the previous experiment with the Stanford VI dataset, was also conducted, when 12 iteractions of training both networks were executed for 20 epochs. The results of two metrics (FWIU and F1-Score) were compared in a test of difference between matched pairs of experiments through 12 iterations. The results from

the AOE Network were superior to the ones obtained by the 3D UNet, with a 95% confidence level (colocar a figura), with an implied power of over 98% for both metrics (Figure 27).
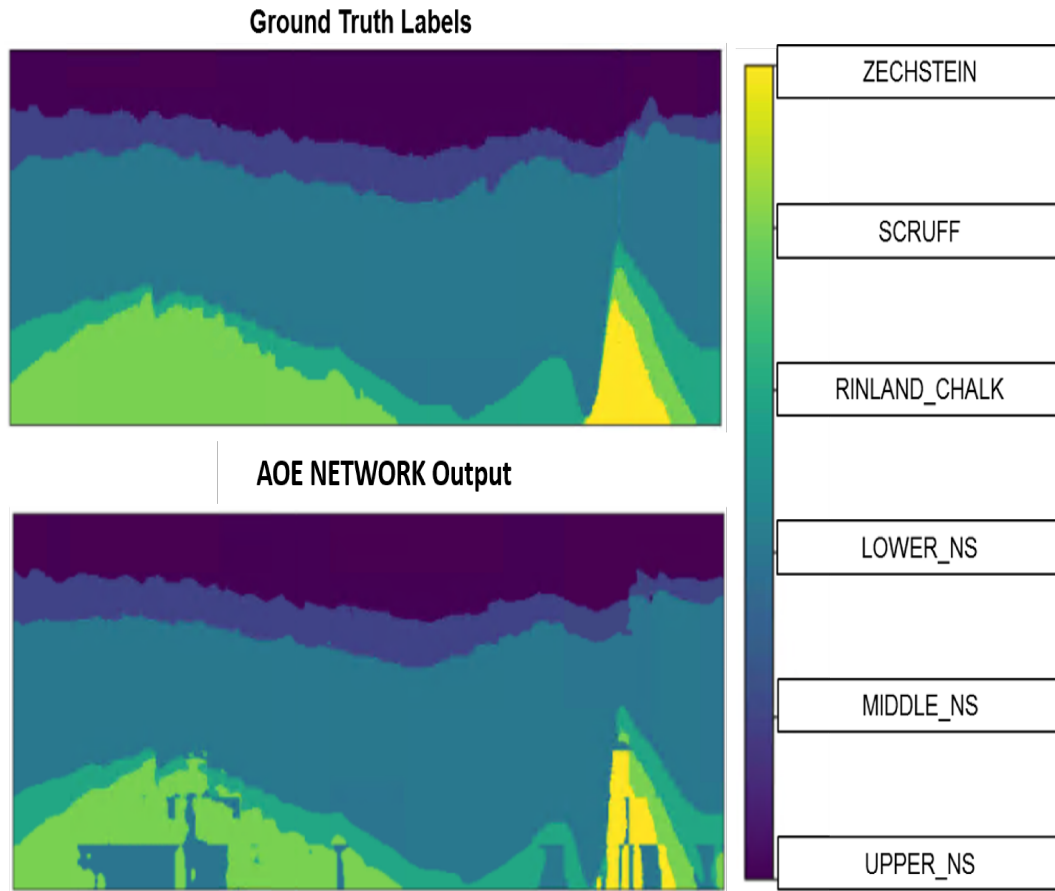
Figure 26 – Comparison between the AOE Network prediction and the ground truth on inline 186 of test set

Table 4 – Comparative results of the 4 networks tested. The best results are highlighted in red.

Metric	Class Accuracy						PA	MCA	FWIU
	Zechstein	Scruff	Rijnland/Chalk	Lower N. S	Middle N. S.	Upper N. S.			
Alaudah et al(Patch-based)	0.264	0.074	0.499	**0.992**	0.804	0.754	0.788	0.565	0.640
Alaudah et al (Section-Based)	0.219	**0.539**	**0.744**	0.951	0.872	0.973	0.879	0.716	0.789
AOE Network (Seismic only)	0.59	0.46	0.65	0.93	0.85	0.96	0.861	0.74	0.759
AOE Network (Seismic and depth)	**0.69**	0.49	0.67	0.96	**0.92**	**0.98**	**0.893**	**0.785**	**0.814**

Figure 27 – Result of the T-Test for the mean comparison between the experiments with the 3D UNet and the proposed Anisotropic Orthogonal Ensemble Network, for both F1/Dice Score (left) and FWIU (right), showing th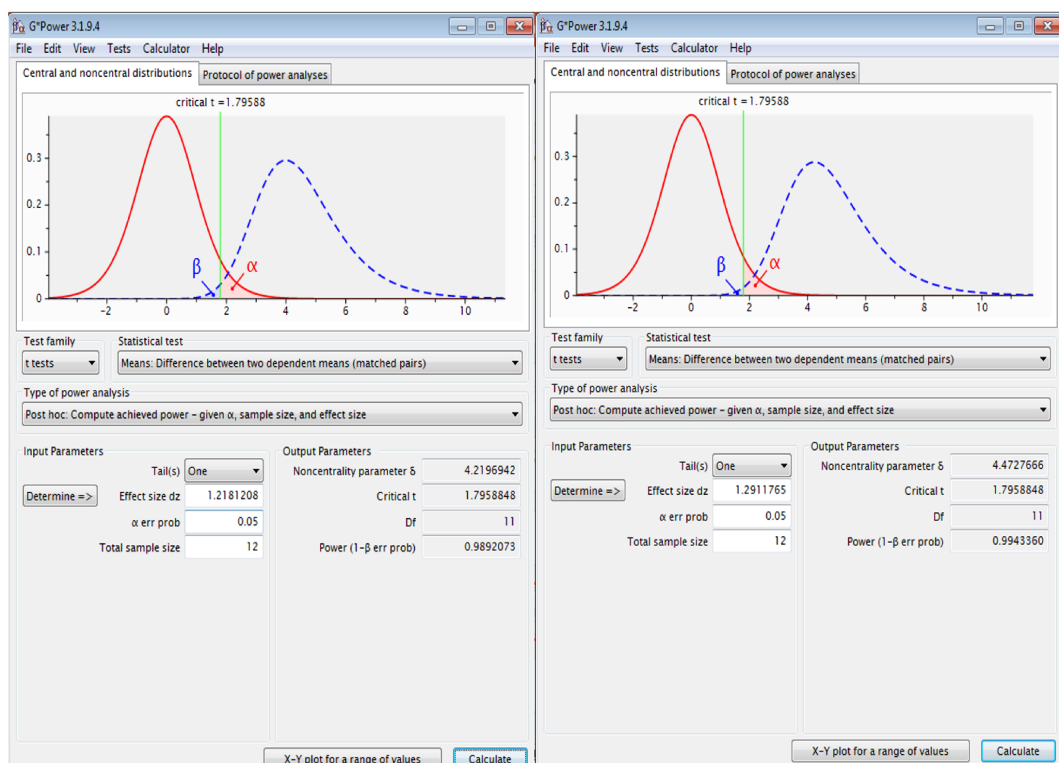at in both parameters the proposed architecture was statistically superior than the 3D UNet in both metrics with a 95% significance level.

6 CONCLUSIONS

This master thesis describes our approach to a 3D image segmentation problem through a Anisotropic Orthogonal Ensemble Network. This model is able to achieve a better performance with a lower computational complexity in both syntetic and real seismic data. This result is accomplished firstly by a quick 2D training in orthogonal planes extracted in 10 random locations, where a cube is simplified by three orthogonal 2D simplification, just as a geologist looking at seismic sections. Afterwards the learned weights are converted to their 3D counterpart and used as the initial weights for the 3D Ensemble Network, each set of weights assigned to the geometrically correct axis.

The majority of the related work in the literature either evaluates similar problems fully by either 2D or 3D networks, since the former demands less computational power and the latter is more accurate for 3D objects. Some works addressed some proposals of 2D to 3D space adaptation for 3D objects segmentation tasks, but they lacked a deeper analysis of the resolution and object shape anisotropy in a 3D data. In our work, the seismic data resolution and the sedimentary concepts were taken into account, since geological features usually have a distinct morphology base on the source rock and the deposition mechanism.

Our proposal of orthogonal planes showed a viable approach to both capture the 3D context and address the difference of resolution and shape distribution of the classes in a 3D seismic, which can also be applied to other areas, such as medical images.

In the initial experiments our network outperformed even a native 3D network (3D UNet), confirming the potential of this approach. Additionally the difference between the performance of the three orthogonal networks individually showed that the orientation of the samples and training reflect on the network accuracy.

The experiments in the F3-block dataset also showed a great performance in lithostratigraphy units segmentation, even surpassing the benchmark proposed by Alaudah et. al. (ALAUDAH et al., 2019). For this experiment it was also evaluated a version of the network with depth as an additional feature. This approach showed that this additional feature increased the performance of the network.

Although the experiments were only conducted with UNet as the backbone of the network, this approach is flexible and support several network architectures, including different ones in the same ensemble. This flexibility, as well as the resource efficiency, enables this methodology to be very attractive to oil and gas industry, since the huge size of the data makes it imperative to have a accurate and time-efficient network.

7 FUTURE PERSPECTIVES

Although with good results, this work still has room to improve. For instance, the proposed network averages the output of all three networks of the ensemble, without any inference whether the networks performs equally in the same task or in accuracy for each label.

Most geological features are irregularly shaped and have heterogeneous distribution, but they might have different shapes and classes association depending on the point of view. Also, other network architectures could be tested in our network, which can be easily applied. Other networks might be more efficient and/or more accurate in Seismic facies segmentation or even a combination of different ones in the ensemble.

Since the author of this work is a geologist, the application in this paper was mainly focused for a geological context. However, this methodology can be explores by other subjects, such as medical images.

Due to the limitations of this work, we propose the following activities:

- Experiments with other architectures other than UNet as the backbone of the network;

- Analysis of the same methodology in other domains (Medical Images, Video analysis);

- The addition of a layer where the output of each orthogonal network is weighted by a trainable variable;

8 PUBLICATIONS

During the studies conducted during this research, the article entitled *Multi-view 3D Seismic Facies Classifier* (Qualis-CAPES: A1) was accepted and presented at the The 36th ACM/SIGAPP Symposium On Applied Computing (March 22-March 26, 2021), .

BIBLIOGRAPHY

ALAUDAH, Y. et al. A machine learning benchmark for facies classification. **Interpretation**, v. 7, n. 3, p. SE175–SE187, 05 2019.

AMINZADEH, F.; DASGUPTA, S. N. Chapter 3 - fundamentals of petroleum geophysics. In: AMINZADEH, F.; DASGUPTA, S. N. (Ed.). **Geophysics for Petroleum Engineers**. [S.l.]: Elsevier, 2013, (Developments in Petroleum Science, v. 60). p. 37–92.

AN-NAN, J.; LU, J. Studying the lithology identification method from well logs based on de-svm. In: **2009 Chinese Control and Decision Conference**. [S.l.: s.n.], 2009. p. 2314–2318.

ANDERSON, S.; BARVIK, S.; RABITOY, C. Innovative digital inspection methods. Offshore Technology Conference, 05 2019.

ANJOS, C. et al. Deep learning for lithological classification of carbonate rock micro-ct images. 07 2020.

AZIZPOUR, H. et al. Factors of transferability for a generic convnet representation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 38, 11 2015.

BARONI, L. et al. Netherlands f3 interpretation dataset. Zenodo, set. 2018. Disponível em: https://doi.org/10.5281/zenodo.1471548.

BESTAGINI, P.; LIPARI, V.; TUBARO, S. A machine learning approach to facies classification using well logs. In: . [S.l.: s.n.], 2017. p. 2137–2142.

BONESS, N. Seismic data for reserves estimation. 2013. Available at http://www.spe.org/training/courses/SDRE.php.

CADEI, L. et al. Machine learning advanced algorithm to enhance production optimization: An ann proxy modelling approach. In: . [S.l.]: International Petroleum Technology Conference, 2020.

CASTRO, S. A.; CAERS, J.; MUKERJI, T. The stanford vi reservoir. **Stanford Center for Reservoir Forecasting**, p. 8–14, 05 2005.

CHEN, L. et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 40, n. 4, p. 834–848, 2018.

CHEN, W. et al. Deep learning reservoir porosity prediction based on multilayer long short-term memory network. **Geophysics**, v. 85, p. WA213–WA225, 04 2020.

CROSS, T.; HOMEWOOD, P. Amanz gressly's role in founding modern stratigraphy. **Geological Society of America Bulletin**, v. 109, p. 1617–1630, 12 1997.

DENNEY, D. To support digital oil fields. **Journal of Petroleum Technology**, p. 71–72, 10 2006.

DI, H.; WANG, Z.; ALREGIB, G. Why using cnn for seismic interpretation? an investigation. In: . [S.l.: s.n.], 2018. p. 2216–2220.

DONAHUE, J. et al. Decaf: A deep convolutional activation feature for generic visual recognition. In: **Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32**. [S.l.]: JMLR.org, 2014. (ICML'14), p. I–647–I–655.

DUIN, E. et al. Subsurface structure of the netherlands - results of recent onshore and offshore mapping. **Netherlands J. Geosci.**, v. 85, 12 2006.

DUNBAR, C. O.; RODGERS, J. Book. **Principles of stratigraphy**. [S.l.]: Wiley New York, 1957. 356 p. : p. ISBN 0471225398.

FABIEN-OUELLET, G.; SARKAR, R. Seismic velocity estimation: A deep recurrent neural-network approach. **GEOPHYSICS**, v. 85, p. 1–35, 11 2019.

FEHLER, M.; KELIHER, P. Seam phase i: Challenges of subsalt imaging in tertiary basins, with emphasis on deepwater gulf of mexico: Seg. In: ____. [S.l.: s.n.], 2011. p. 1–168. ISBN 978-1-56080-287-7.

FRIEDMAN, G. M.; SANDERS, J. E.; KOPASKA-MERKEL, D. C. **Principles of Sedimentary Deposits: Stratigraphy and Sedimentology**. [S.l.]: Maxwell Macmillan International, 1992.

FUKUSHIMA, K.; MIYAKE, S. Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position. **Pattern Recognition**, v. 15, n. 6, p. 455–469, 1982. Disponível em: http://www.sciencedirect.com/science/article/pii/0031320382900243.

GROTZINGER, J. et al. **Understanding Earth.** 5th ed.. ed. New York: W. H. Freeman, 2007. 110-111 p.

GUAZZELLI, A. B.; ROISENBERG, M.; RODRIGUES, B. Efficient 3d semantic segmentation of seismic images using orthogonal planes 2d convolutional neural networks. **2020 International Joint Conference on Neural Networks (IJCNN)**, p. 1–8, 2020.

HAMPSON, D. P.; RUSSELL, B. H.; BANKHEAD, B. Simultaneous inversion of pre-stack seismic data. In: ____. **SEG Technical Program Expanded Abstracts 2005**. [s.n.], 2005. p. 1633–1637. Disponível em: https://library.seg.org/doi/abs/10.1190/1.2148008.

HAQUE, A. E.; ISLAM, M. A.; SHALABY, M. Three-dimensional facies analysis using object-based geobody modeling: A case study for the farewell formation, maui gas field, taranaki basin, new zealand. p. 1, 01 2018.

HESTHAMMER, J.; LANDRO, M.; FOSSEN, H. Use and abuse of seismic data in reservoir characterisation. **Marine and Petroleum Geology - MAR PETROL GEOL**, v. 18, p. 635–655, 05 2001.

HOJAGELDIYEV, D. Artificial intelligence in hse. In: . [S.l.]: Abu Dhabi International Petroleum Exhibition and Conference, 2018.

HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Multilayer feedforward networks are universal approximators. **Neural Networks**, v. 2, n. 5, p. 359–366, 1989.

HUANG, L.; DONG, X.; CLEE, T. A scalable deep learning platform for identifying geologic features from seismic attributes. **The Leading Edge**, v. 36, p. 249–256, 03 2017.

HUBEL, D.; WIESEL, T. Receptive fields of single neurones in the cat's striate cortex. **The Journal of physiology**, v. 148, p. 574–591, 04 1959.

JAN, B. et al. Deep learning in big data analytics: A comparative study. **Computers and Electrical Engineering**, v. 75, 12 2017.

KOLLA, V.; POSAMENTIER, H.; WOOD, L. Deep-water and fluvial sinuous channels—characteristics, similarities and dissimilarities, and modes of formation. **Marine and Petroleum Geology**, v. 24, p. 388–405, 06 2007.

KRAVITZ, G. The geohistorical time arrow: From steno's stratigraphic principles to boltzmann's past hypothesis. **Journal of Geoscience Education**, v. 62, p. 691–700, 11 2014.

KöPüKLü, O. et al. Resource efficient 3d convolutional neural networks. In: **2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)**. [S.l.: s.n.], 2019. p. 1910–1919.

LAMMOGLIA, T.; OLIVEIRA, J. K. de; FILHO, C. R. de S. Lithofacies recognition based on fuzzy logic and neural networks: A methodological comparison. In: . [S.l.: s.n.], 2014.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, p. 436–44, 05 2015.

LECUN, Y. et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, v. 86, n. 11, p. 2278–2324, 1998.

LI, F. et al. Addcnn: An attention-based deep dilated convolutional neural network for seismic facies analysis with interpretable spatial-spectral maps. **IEEE Transactions on Geoscience and Remote Sensing**, PP, p. 1–12, 06 2020.

LI, H. et al. Applications of artificial intelligence in oil and gas development. **Archives of Computational Methods in Engineering**, p. 1–13, 01 2020.

LIU, S. et al. 3d anisotropic hybrid network: Transferring convolutional features from 2d images to 3d anisotropic volumes. In: . [S.l.]: Springer International Publishing, 2018. p. 851–858. ISBN 978-3-030-00933-5.

LOPEZ, G. I. Walther's law of facies. In: _____. **Encyclopedia of Scientific Dating Methods**. Springer Netherlands, 2013. p. 1,2. ISBN 978-94-007-6326-5. Disponível em: https://doi.org/10.1007/978-94-007-6326-5_30-1.

LUNDSTR OM, D. **Data-efficient Transfer Learning with Pre-trained Networks**. 57 p. Dissertação (Mestrado) — Linköping University, Computer Vision, 2017.

MACLEOD, N. Principles of stratigraphy. **Encyclopedia of Geology**, 01 2005.

MCGUINESS, K. Transfer learning. In: . [S.l.]: Dublin City University, 2017.

MIALL, A. **Stratigraphy: A Modern Synthesis**. [S.l.: s.n.], 2015. 87-91 p. ISBN 978-3-319-24302-3.

MIDDLETON, G. V. Johannes walther's law of the correlation of facies. **GSA Bulletin**, v. 84, n. 3, p. 979–988, 03 1973. Disponível em: https://doi.org/10.1130/0016-7606(1973)84<979:JWLOTC>2.0.CO;2.

MOHAMED, I. et al. Formation lithology classification: Insights into machine learning methods. **Pure and Applied Geophysics**, Curran Associates, Inc, 09 2019.

MONDOL, N.; BJøRLYKKE, K. Petroleum geoscience, springer. In: ____. [S.l.: s.n.], 2011. cap. Seismic Exploration, p. 375–402.

MUTTI, E. et al. Turbidites and turbidity currents from alpine 'flysch' to the exploration of continental margins. **Sedimentology**, v. 56, p. 267–318, 2009.

NOLEN-HOEKSEMA, R. C. The future role of geophysics in reservoir engineering. **The Leading Edge**, v. 9, n. 12, p. 89–97, 1990. Disponível em: https://doi.org/10.1190/1.1439708.

O'Mahony, N. et al. Convolutional neural networks for 3d vision system data : A review. In: **2018 12th International Conference on Sensing Technology (ICST)**. [S.l.: s.n.], 2018. p. 160–165.

PAN, S.; YANG, Q. A survey on transfer learning. **IEEE Transactions on Knowledge and Data Engineering**, v. 22, n. 10, p. 1345–1359, 11 2010.

PRATT, L.; JENNINGS, B. A survey of connectionist network reuse through transfer. In: ____. **Learning to Learn**. USA: Kluwer Academic Publishers, 1998. p. 19–43. ISBN 0792380479.

RAMIREZ, C.; LARRAZABAL, G.; GONZáLEZ, G. Salt body detection from seismic data via sparse representation. **Geophysical Prospecting**, v. 64, 06 2015.

RAWAT, W.; WANG, Z. Deep convolutional neural networks for image classification: A comprehensive review. **Neural Computation**, v. 29, p. 1–98, 06 2017.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-Net: Convolutional networks for biomedical image segmentation. **Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015**, May 2015.

SELLEY, R.; SONNENBERG, S. Elements of petroleum geology: Third edition. **Elements of Petroleum Geology: Third Edition**, p. 124–128, 01 2014.

Shan, H. et al. 3-d convolutional encoder-decoder network for low-dose ct via transfer learning from a 2-d trained network. **IEEE Transactions on Medical Imaging**, v. 37, n. 6, p. 1522–1534, 2018.

SHENGRONG, L. et al. Seismic fault detection using an encoder–decoder convolutional neural network with a small training set. **Journal of Geophysics and Engineering**, v. 16, 02 2019.

SHI, Y.; WU, X.; FOMEL, S. Saltseg: Automatic 3d salt segmentation using a deep convolutional neural network. **Interpretation**, v. 7, p. SE113–SE122, 04 2019.

SHIBILI, S.; WONG, P. Use of interpolation neural networks for permeability estimation from well logs. **The Log Analyst**, v. 39, n. 06, 11 1998. ISSN 0024-581X.

SHLEZINGER, N. et al. Model-based deep learning. 12 2020.

SLATT, R. **Stratigraphic reservoir characterization for petroleum geologists, geophysicists, and engineers**. [S.l.: s.n.], 2006. v. 6. 139-140 p. ISBN 0444528180.

SU, H. et al. Multi-view convolutional neural networks for 3d shape recognition. In: **Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)**. USA: IEEE Computer Society, 2015. (ICCV '15), p. 945–953. ISBN 9781467383912. Disponível em: https://doi.org/10.1109/ICCV.2015.114.

SUDRE, C. H. et al. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. **Lecture Notes in Computer Science**, Springer International Publishing, p. 240–248, 2017. ISSN 1611-3349. Disponível em: http://dx.doi.org/10.1007/978-3-319-67558-9_28.

TALLING, P. et al. Key future directions for research on turbidity currents and their deposits. **Journal of Sedimentary Research**, v. 85, p. 153–169, 01 2015.

TEICHERT, C. Concepts of facies1. **AAPG Bulletin**, v. 42, n. 11, p. 2718–2744, 11 1958. ISSN 0149-1423. Disponível em: https://doi.org/10.1306/0BDA5C0C-16BD-11D7-8645000102C1865D.

TRAN, D. et al. Learning spatiotemporal features with 3d convolutional networks. In: . [S.l.: s.n.], 2015. p. 4489–4497.

WALTHER, J. **Einleitung in die Geologie als historische Wissenschaft: Beobachtungen über die Bildung der Gesteine und ihrer organischen Einschlüsse**. [S.l.]: G. Fischer, 1894. v. 3.

Wang, Z. et al. 2d multi-spectral convolutional encoder-decoder model for geobody segmentation. In: **2018 International Conference on Computational Science and Computational Intelligence (CSCI)**. [S.l.: s.n.], 2018. p. 1193–1198.

WEI, Z. et al. Characterizing rock facies using machine learning algorithm based on a convolutional neural network and data padding strategy. **Pure and Applied Geophysics**, v. 176, p. 1–13, 03 2019.

WEISS, K.; KHOSHGOFTAAR, T.; WANG, D. A survey of transfer learning. **Journal of Big Data**, v. 3, 05 2016.

WU, X. et al. Convolutional neural networks for fault interpretation in seismic images. **SEG International Exposition and Annual Meeting**, p. 1946–1950, 2018.

XIONG, W. et al. Seismic fault detection with convolutional neural network. **GEOPHYSICS**, v. 83, p. 1–28, 06 2018.

YANG, F.; MA, J. Deep-learning inversion: a next generation seismic velocity-model building method. **GEOPHYSICS**, v. 84, n. 4, p. R583–R599, 02 2019.

YANG, J. et al. Reinventing 2d convolutions for 3d images. **IEEE Journal of Biomedical and Health Informatics**, PP, p. 1–1, 01 2021.

YOSINSKI, J. et al. How transferable are features in deep neural networks? In: **Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2**. Cambridge, MA, USA: MIT Press, 2014. (NIPS'14), p. 3320–3328.

ZHAO, T. Seismic facies classification using different deep convolutional neural networks. **SEG Technical Program Expanded Abstracts 2018**, p. 2046–2050, 09 2018.

ZHUANG, F. et al. A comprehensive survey on transfer learning. **Proceedings of the IEEE**, v. 109, p. 43–76, 2021.