Literature Review on Deep Learning for the Segmentation of Seismic Images

Bruno A. A. Monteiro[a,*], Gabriel L. Canguçu[a], Leonardo M. S. Jorge[a], Rafael H. Vareto[a], Bryan S. Oliveira[b], Thales H. Silva[a], Luiz Alberto Lima[d], Alexei M. C. Machado[b,c], William Robson Schwartz[a] and Pedro O. S. Vaz-de-Melo[a]

[a]*Department of Computer Science, Universidade Federal de Minas Gerais, Av. Pres. Antônio Carlos, 6627, Belo Horizonte, Brazil 31270-901 (e-mail: brunoalemaogeo@gmail.com, gabriel-lima@ufmg.br, leonardosouza2012@gmail.com, rafaelvareto@dcc.ufmg.br, thaleshenrique@ufmg.br, william@dcc.ufmg.br, olmo@dcc.ufmg.br)*

[b]*Department of Computer Science, Pontifical Catholic University of Minas Gerais, R. Dom Jose Gaspar, 500, Belo Horizonte, Brazil 30535-901 (e-mail:bsoliveira@proton.me, alexeimcmachado@gmail.com)*

[c]*Department of Anatomy and Imaging, Universidade Federal de Minas Gerais, Av. Alfredo Balena, 192, Belo Horizonte, Brazil, 30130-100*

[d]*Petróleo Brasileiro S.A., Av. República do Chile 65, Rio de Janeiro, Brazil 20031-170 (e-mail:lual@petrobras.com.br)*

ARTICLE INFO

Keywords:
Literature Review
Seismic Image
Segmentation
Seismic Facies
Semantic Segmentation
Deep Learning

ABSTRACT

This systematic literature review summarizes the current state of deep learning (DL) for semantic segmentation in seismic data, focusing on facies segmentation. It presents the architectures, cost functions and learning paradigms used in seismic data segmentation, and their results. Various DL approaches and methodologies are discussed, highlighting challenges in evaluation procedures and data availability. It also identifies research opportunities, such as robust benchmarking. Furthermore, it discusses the potential of approaches such as few-shot and semi-supervised learning, transfer learning, and ensemble techniques to address the challenge of limited labeled data. A comprehensive discussion on the raised challenges can improve segmentation quality and provide more reliable seismic facies interpretation.

1. Introduction

The comprehension of subsurface geology and its geometry is essential to understanding reservoir properties and supporting strategic decisions in the field (Chevitarese et al., 2018; Abid et al., 2022; Wang et al., 2023). The interpretation of geological features such as geobodies and lithostratigraphic facies or structures such as horizons and faults in seismic data can be performed as a visual recognition task. However, interpreting seismic data is time-consuming, expensive and subject to observer-dependent expertise (Waldeland et al., 2018; Wang and Chen, 2021; Wang et al., 2023).

Seismic methodologies have been used since the early 1900s, primarily for assessing water depths and identifying icebergs (Mondol, 2010). Over time, these technological advances enabled their application in discovering oil reservoirs. Currently, seismic surveying holds a prominent position as the primary geophysical technique in the field of oil exploration. By recording acoustic wave propagation times, this methodology provides means to discern subsurface rock formations, allowing for estimations of their geometrical configurations and intrinsic physical attributes. These signals are measures of the two-way travel time, i.e., the amount of time it takes for a seismic wave to leave the wave generator, strike a reflecting surface, and then travel back to the receiver.

Seismic data are therefore a representation of the acoustic response of a seismic wave's reflection caused by a physical change in the propagation medium or a change in the rock characteristics (Bjørlykke, 2015; Herron, 2011; Nanda, 2021). The seismic acquisition process typically collects

brunoalemaogeo@gmail.com (Bruno A. A. Monteiro)
ORCID(s): 0000-0001-7288-5504 (Bruno A. A. Monteiro); 0009-0004-0206-3790 (Leonardo M. S. Jorge); 0000-0002-0431-5945 (Rafael H. Vareto); 0000-0001-8077-3377 (Alexei M. C. Machado); 0000-0003-1449-8834 (William Robson Schwartz); 0000-0002-9749-0151 (Pedro O. S. Vaz-de-Melo)

List of abbreviations: ABL: Active Boundary Loss; ABUnet: Atrous Bidirectional UNet; ADV: Adversarial loss; ASPP: Spatial Pyramid Pooling architecture; BNN: Bayesian Neural Network; CEL: Cross-entropy loss; CNN: Convolutional Network; CON: Contrastive loss; CS: Cosine similarity; FWIoU: Frequency Weighted Intersection over Union; GAN: Generative Adversarial Network; GPU: Graphic Processing Unit; LSTM: Long Short-term Memory; MCA: Mean class accuracy; mIoU: Mean Intersection over Union; ML: Machine Learning; MSE: Mean Squared Error; PA: Pixel Accuracy; RMV: Relevance Vector Machine; SGD: Stochastic Gradient Descent; SLIC: Simple Linear Iterative clustering; SSL: Self-supervised Learning; SVM: Support Vector Machine; TVL: Total Variation Loss.

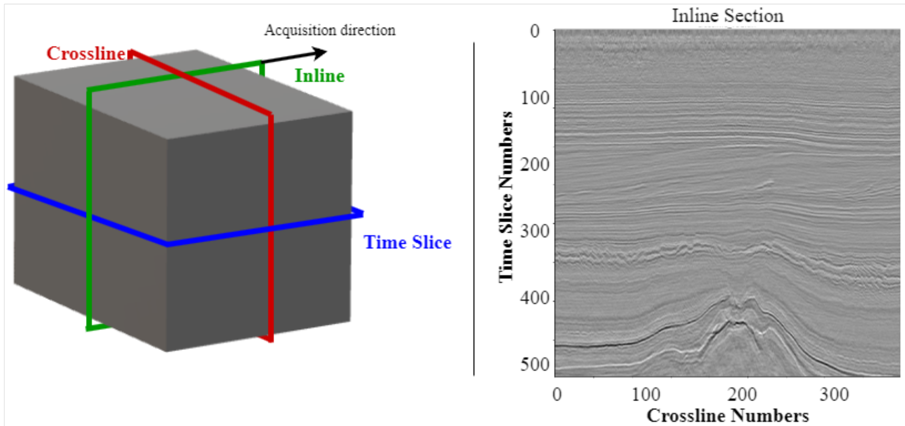

Figure 1: Seismic data. On the left, a schematic of the three-dimensional seismic data acquisition and axis nomenclature; on the right, an example of an inline section from the seismic dataset F3 Netherlands (Silva et al., 2019).

three-dimensional data that, after many processing steps, result in an understandable cubic data volume describing the Earth's subsurface. As depicted in Fig. 1, these data cubes are represented on the space spanned by three perpendicular axes: an *inline* axis, parallel to the acquisition direction; a *crossline* axis, perpendicular to the inline axis with which it defines a horizontal plane; and a *time* or *depth vertical* axis (EnergyGlossary, 2023). Therefore, an inline section or image is in a plane perpendicular to the crossline axis; a crossline section is perpendicular to the inline axis; and the *time slice* is a plane that is perpendicular to the vertical axis EnergyGlossary (2023).

In geology, facies refer to distinct bodies of rock that possess specific characteristics, which can include observable attributes such as the rock's overall appearance, composition, or how it was formed, as well as any changes that may have occurred in these attributes across a geographic area (Reading, 1978). Facies' features also include chemical, physical, and biological properties that set them apart from neighboring rocks (Parker, 1984). Generally defined by depositional structures, geological structures, or even changes in fluid type, a seismic facies is a three-dimensional sedimentary unit composed of units of reflection patterns distinct from other neighboring facies (Roksandić, 1978). The lithology represented by the seismic facies can be more precisely identified using rock core samples and well profile data, but the costs involved in drilling and exploring these wells can be prohibitive (Li et al., 2022). An alternative strategy is leveraging computational methods, particularly deep learning techniques, to identify facies in seismic data. This approach significantly enhances efficiency and reduces the costs associated with the identification.

Problem statement: Deep learning for seismic facies segmentation presents a unique challenge in terms of literature organization, as different approaches may use diverse data formatting, training strategies, and evaluation metrics. This lack of standardization burdens researchers attempting to navigate the growing body of literature in this domain. The absence of standardized practices for organizing and presenting deep learning research can lead to several difficulties, including limited reproducibility, inefficient literature search, and barriers to collaboration. Managing the literature in deep learning for seismic facies segmentation is crucial for overcoming these challenges and fostering a more productive research environment.

Therefore, the objective of this paper is to offer a thorough and critical examination of the literature related to the application of deep learning methods in facies identification. To achieve this, we conduct a systematic review encompassing literature on deep learning methods specifically applied to the segmentation of seismic data, with a concentrated emphasis on facies segmentation. We highlight the most relevant aspects discussed in those papers, such as the main goals, used datasets, employed architectures, metrics, and results. Most importantly, we provide an organized critical rundown of the literature, describing some of the challenges that researchers face in the field and opportunities for further exploration.

To the best of our knowledge, this is the first literature review specifically focused on the literature of segmentation of seismic facies based on deep learning. Existing studies in the domain predominantly gravitate toward fault detection and structural interpretation or geological hazard analysis. The review presented by Wang et al. (2018) is primarily centered on interpreting faults and salt domes. Their research compiled extensive interpretation workflows employing conventional and advanced image-processing techniques and machine-learning algorithms. Ma and Mei (2021) presented an overview on the application of deep learning methods for the analysis of geological hazards focusing on convolutional and recurrent neural networks. More recently, An et al. (2023) conducted a comprehensive survey of the literature on deep learning methods for fault interpretation. Similar to our findings, they also observed the prevalence of the UNet and its variants as the most frequently employed network architectures. Additionally, they highlighted the scarcity of publicly available datasets. Among a total of 73 cataloged seismic datasets, only three field datasets and four synthetic datasets are accessible for benchmarking purposes. In short, this survey proposes the following contributions:

1. A systematic review of the literature on deep learning approaches for seismic data segmentation, especially facies delimitation;
2. A description of the state-of-the-art on deep learning approaches used for seismic data segmentation, architectures, cost functions, and results;
3. A critical review of the literature describing the merits and shortcomings of the area, and from that, a discussion on the opportunities for future studies.

2. Why Deep Learning?

2.1. Traditional Techniques and Shallow Machine Learning

The main approaches used for the interpretation of seismic images before the development of deep learning techniques include seismic attribute analysis (Chopra and Marfurt, 2005; Gao, 2007), pre-stack (Qian et al., 2018) and post-stack (Song et al., 2017) waveform classification methods, as also proposed in a recent study by Su-Mei et al. (2022).

In the most general sense, the definition of seismic attributes refers to all quantities derived from seismic data (Chopra and Marfurt, 2005). There are dozens of distinct seismic attributes calculated from seismic data and applied to interpreting geological structure, stratigraphy, and rock/pore fluid properties. Compared to manual interpretation, attribute analysis greatly improves efficiency. However, the burden associated with its interpretation is still significant due to multiple seismic properties such as amplitudes, frequencies, and phases (Li et al., 2022). These properties can vary according to the geological characteristics of the subsurface and may affect the interpretation of the data. The complex interactions between seismic features and the target geological entities make it difficult to evaluate the accuracy of the categorization results.

The waveform classification method generally follows two approaches: post-stack and pre-stack seismic facies classification. While the former essentially involves identifying different waveform signal patterns (Qian et al., 2018), the latter exploits richer information, which can theoretically result in higher resolution and more accurate interpretation. Compared to seismic attribute analysis, waveform classification approaches can overcome their disadvantages by carefully illustrating the lateral variation of the seismic trace, which expresses the flat distribution of anomalies (de Matos et al., 2007). However, this technique cannot be used in regions with significant changes in formation thickness, and efficient quality control rules must be implemented to provide high-quality data.

The applicability and dependability of these traditional methodologies cannot satisfy the demands of contemporary seismic facies analysis due to the growing complexity of seismic data. In addition, the outcomes depend on the experts' subjective evaluation (Zhang et al., 2019). Unfortunately, the fact that these conventional approaches still call for the manual selection of seismic attributes and that the interaction of those attributes directly impacts the performance is even more problematic. Moreover, since different geological environments differ greatly from one another, the combination of seismic properties chosen in one location cannot be immediately applied to another region (Li et al., 2022).

The development of machine learning (ML) techniques encouraged many researchers to carry out automatic or semi-automatic facies identification. There is a very rich literature on shallow ML (supervised and unsupervised) methods for facies classification. Zhao et al. (2015) reviews some of the most commonly used techniques. For instance, Principal Component Analysis has been used by Wolf and Pelissier-Combescure (1982) to cluster and select seismic attributes. Delfiner et al. (1987) employed discriminant functions to determine seismic facies classes and Liu et al. (2020b) were able to identify facies using a relevance vector machine (RVM) and developed a facies discriminant method based on a multi-kernel RVM.

2.2. Deep Learning Techniques

Over the past few years, much research involving seismic interpretation tasks has explored deep learning techniques to aid interpretation (Waldeland et al., 2018; Alaudah et al., 2019; Abid et al., 2022; Nasim et al., 2022; Tolstaya and Egorov, 2022; Li et al., 2022; Wang et al., 2023). Deep learning algorithms are powerful methods to discover representation from data in their raw form and map it to diverse tasks (LeCun et al., 2015). These techniques have achieved remarkable results over the past couple of decades in a wide range of tasks such as speech recognition (Deng et al., 2013; Hinton et al., 2012a), biometrics (Vareto et al., 2023), healthcare (Jiang et al., 2017), agricultural applications (Eli-Chukwu, 2019) and natural language processing (Chowdhary and Chowdhary, 2020). Important advances took place since the late 80s when Rumelhart et al. (1986) introduced backpropagation, and LeCun et al. (1989) demonstrated that large backpropagation networks could be employed to image recognition with little preprocessing. In 1998, LeCun et al. (1998) showed the potential of convolutional neural networks (CNNs) as feature extractors, reducing manual-designed extractions.

A breakthrough in the deep learning domain happened in 2012 when Krizhevsky et al. (2012a) published AlexNet, an innovative neural network that won the ImageNet competition (Deng et al., 2009) of that year by a significant margin. AlexNet further popularized the use of graphic processing units (GPUs) and CNNs, making it one of the most influential papers in the history of computer vision. The subsequent years would reveal additional improvements with the introduction of ZFNet (2013) (Zeiler and Fergus, 2014), VGGNet (2014) (Simonyan and Zisserman, 2014), GoogLeNet (2014) (Szegedy et al., 2015) and ResNet (2015) (He et al., 2016). Each model continued to build on the success of the previous ones by changing the depth and size of the networks (Zeiler and Fergus, 2014), experimenting on different hyperparameters, or applying theoretical techniques, such as residual connections (He et al., 2016) and dropout (Hinton et al., 2012b) to handle overfitting and vanishing gradients. The success of these models relies on their ability to learn multi-level representations through non-linear transformations from the raw data that support pattern recognition tasks (LeCun et al., 2015; He et al., 2016). Furthermore, these achievements have also occurred due to greater computing power in the form of GPUs.

Despite being a geological assignment, identifying seismic features such as lithostratigraphic and seismostratigraphic facies from a computational perspective is primarily a computer vision routine, particularly related to image segmentation tasks. Semantic segmentation is one of the most common computer vision strategies for visual pattern recognition (Long et al., 2015; Ronneberger et al., 2015; Lateef and Ruichek, 2019). It consists of labeling a class with semantic meaning for each region or pixel in an image (or video) (Lateef and Ruichek, 2019; Minaee et al., 2022), and has been applied in numerous specific contexts, such as analysis of medical images (Shen et al., 2022; He et al., 2021), autonomous driving (Li et al., 2018), and remote sensing (Li et al., 2021). Since before the deep learning era, many algorithms have been proposed to tackle this task, e.g., thresholding (Otsu, 1979), watershed transformations (Najman and Schmitt, 1994; Neubert and Protzel, 2014), region-growing (Nock and Nielsen, 2004), Felzenszwalb segmentation (Felzenszwalb and Huttenlocher, 2004), and Simple Linear Iterative clustering (SLIC) (Achanta et al., 2012). Other techniques relied on multiscale classification (Dos Santos et al., 2012), depending on a stepped process that dealt with data representation (defining objects of interest), feature extraction (describing texture or contours), and training a machine learning algorithms for classification, such as Support Vector Machines (SVM) (Tzotsos et al., 2011) or shallow Neural Networks (Ouma et al., 2008). However, shallow machine learning approaches often have limited capacity to analyze raw data input (LeCun et al., 2015). They may require specific knowledge of the application field, which can be a significant issue in domain-specific tasks.

Deep learning has also brought advances to computer vision tasks, constantly pushing the state-of-the-art for image classification (Krizhevsky et al., 2012b; Szegedy et al., 2015; Chen et al., 2023), object detection and instance segmentation (Girshick, 2015; He et al., 2017; Li et al., 2018), segment anything task (Kirillov et al., 2023) and semantic segmentation (Ronneberger et al., 2015; Badrinarayanan et al., 2017). Long et al. (2015) made a breakthrough in semantic segmentation, proposing the use of fully-connected networks. Since then, many other architectures brought improvements to this concept, e.g., utilizing U-shaped encoder-decoder architectures (Ronneberger et al., 2015), preserving pooling positional indices (Badrinarayanan et al., 2017), and applying atrous convolutions (Chen et al., 2017b).

Unsurprisingly, research in the oil and gas field has been eager to use the aforementioned algorithms to address seismic interpretation problems (Waldeland et al., 2018; Alaudah et al., 2019; Wrona et al., 2021). This interest comes mostly as an alternative to aid and accelerate manual interpretations, since the massive volume of information contained in seismic data, especially 3D surveys, makes manual evaluation by experts costly and time-consuming (Waldeland et al., 2018; Liu et al., 2020a). Deep learning applications in seismic interpretation include the employment of CNNs to classify salt domes (Waldeland et al., 2018; Shi et al., 2019), delineate geological faults/horizon (Bi et al., 2021), relative geologic time estimation (Bi et al., 2020), and facies segmentation (Liu et al., 2020a; Su-Mei et al., 2022; Monteiro et al., 2022).

3. Systematic Review and Data Extraction

This work employs a systematic review to comprehensively analyze the existing literature on deep learning models utilized for the semantic segmentation of seismic images. The review entails selecting and analyzing relevant studies, extracting their data, and synthesizing their findings. It organizes essential components of any research effort, including examining and synthesizing existing material of the state of knowledge in a study domain. More importantly, this procedure can address inquiries that individual research endeavors might not be able to answer and identify shortcomings in primary research that should be resolved in subsequent studies (Page, 2021). Theses aspects include a discussion about the lack of experimental standardization and the choice of deep learning architectures for the domain. Finally, a systematic literature review and data collection method aims to reduce bias, improve the validity of the findings, and contribute to advancing knowledge in the area.

Our study followed the steps suggested by Khan et al. (2003) and the Covidence Blog (Covidence) to conduct a systematic review. The first essential step is to formulate the questions that will be addressed, ensuring their relevance concerning the intended application of the review and the need to bridge existing knowledge gaps. This process assists in centering attention on the identified problem and maintaining a clear project scope.

We structured this review to address the knowledge gap regarding the *automated segmentation of geological features in seismic data using deep learning*. From a geological perspective, it focuses on continuous features, such as lithostratigraphic and seismostratigraphic facies, channeled structures, and salt bodies, and, from a computational standpoint, on computer vision techniques mainly related to the image segmentation task.

In the following sections, we describe the methods used in the review, including database selection, search keywords, inclusion and exclusion criteria, and data extraction methods. We also review strategies for evaluating the quality of the included research and approaches for data synthesis and presentation.

3.1. Research methods

After formulating the questions that will be addressed, the second stage of a systematic review is to search for relevant publications, usually in digital libraries. As this is a very laborious task, search engines are employed as look-up tools for returning publications according to a provided query and defined restrictions that may not have been filtered out in the queries. In this study, we used Harzing's Publish or Perish (Harzing, A.W.), a software that searches for and analyzes academic citations. It is able to search across many search engines utilizing specific keyword queries, date constraints, and other settings. The outcome is shown on the screen and may be saved as a RIF file that can be imported into reference managers.

To ensure the reproducibility of the workflow, all the applied search methods are described. The selected search engine was Google Scholar[1], as it encompasses an extensive repository of scholarly papers spanning a wide range of disciplines. All searches were restricted by date, as in this review we were only interested in deep learning approaches that received major attention after the publication of the AlexNet (Krizhevsky et al., 2012b) in 2012. Therefore, we were interested in papers made public between 2012 and May 2023. Despite that, only papers after 2017 were found within the aims of this review. To provide in-scope results, a series of queries were framed using geological and deep learning-related keywords. These queries are listed in

[1] https://scholar.google.com

Table 1
Searching queries and number of papers returned.

Query	**# Papers**
seismic deep learning; depositional elements; architectural lobes; channels; semantic segmentation	151
geosciences; deep learning; depositional elements; architectural elements; lobes; meandering channels; semantic segmentation; seismic	54
("facies turbidites") AND ("deep learning")	1
("facies turbidites" OR "architectural elements") AND ("semantic segmentation") AND ("geology")	15
("depositional elements" OR "architectural elements" OR "seismic" OR "facies turbidites") AND ("semantic segmentation") AND ("geology")	521
("depositional elements" OR "architectural elements" OR "facies turbidites") AND ("geology") AND ("seismic") AND ("semantic segmentation")	6
("depositional elements" OR "architectural elements" OR "lobes" OR "meandering channels" OR "facies turbidites") AND ("geology") AND ("seismic") AND ("semantic segmentation")	32
("depositional elements" OR "architectural elements" OR "lobes" OR "meandering channels" OR "facies turbidites") AND ("geology") AND ("seismic") AND ("semi-supervised")	39
("depositional elements" OR "architectural elements" OR "lobes" OR "meandering channels" OR "facies turbidites") AND ("geology") AND ("seismic") AND ("self-supervised")	1
("depositional elements" OR "architectural elements" OR "lobes" OR "meandering channels" OR "facies turbidites") AND ("geology") AND ("seismic") AND ("transformer")	44
("depositional elements" OR "architectural elements" OR "lobes" OR "meandering channels" OR "facies turbidites") AND ("geology") AND ("seismic") AND ("active learning")	21
("depositional elements" OR "architectural elements" OR "lobes" OR "meandering channels" OR "facies turbidites") AND ("geology") AND ("seismic") AND ("GANs") AND ("generative")	30
("depositional elements" OR "architectural elements" OR "lobes" OR "meandering channels" OR "facies turbidites") AND ("geology") AND ("seismic") AND ("diffusion models")	39
Classification; Segmentation; Seismic Facies; Neural networks; Machine Learning	80
Total entries	1034
Total unique entries	**898**

Table 1 as well as the number of papers found by them. These searches resulted in 858 unique entries, which were subsequently entered into the Rayyan software (Ouzzani et al., 2016), a collaborative tool for managing and accelerating systematic reviews.

3.2. Title and abstract screening

Title and abstract screening is an important stage in systematic reviews and research synthesis. This procedure entails evaluating the significance of possibly relevant research based on their titles and abstracts. It is an efficient method for identifying works that are likely to match the survey's scope while rejecting unrelated studies.

Given the papers found in the previous stage, the current step comprises screening through them. We screened the title and abstracts of all the 898 unique entries, providing a flag (“included”, “maybe”, or “excluded”) signaling individual concerns regarding whether the study should proceed to the next stage. We registered at least three reviewers per paper, with no maximum amount of opinions. If at least two reviewers *excluded* a paper, it was disregarded in the next stage. Otherwise, if at least two reviewers approved it, it was kept for the next stage. In the case of divergence, a second look at the paper was conducted, considering the opinion of the senior members involved in this step to resolve the conflicts. Some of the excluding reasons in this stage are stated below. At the end of this procedure, 51 papers were selected to advance in the process.

Exclusion criteria. The following reasons were used to exclude papers: Methods using well log data, approaches not using seismic data, classical seismic attribute approaches, methods employing only shallow machine learning, surveys, introductory papers, dissertations, and out-of-scope publications.

3.3. Full-text screening

Since only 51 studies were accepted in the previous phase, the full-text screening step was conducted parallel with data extraction. With that, we looked deeper into the eligible papers and better understood the approaches used for automated seismic segmentation. At this point, we also presented eligibility criteria for including and excluding papers during the full-text screening stage. Similar to the previous stage, at least two and at most three reviewers' opinions were considered for each study, and only those with two approvals were selected to compose this review. After trialing the publications by full-text screening, 25 publications were selected, from which relevant metadata and key elements were extracted.

Inclusion criteria. The following reasons were considered for including papers: Innovative methods; exceptional results; channel identification; approaches that are not fully supervised (few-shot, semi/self-supervised); domain adaptation; open code; and open data (including labels).

Exclusion criteria. The following reasons were considered to exclude papers: Training depending on borehole data; training depending on explicit geological conditioning; fuzzy metrics; qualitative assessment only; and insufficient validation/testing methods.

3.4. Data Extraction

This is the final stage of the systematic review, in which the obtained data must be organized and synthesized, and the insights brought by this study are presented. Here, we present an overview of the essential aspects, such as the main goals, the year of publication, the learning paradigm, the input format, the deep learning architectures and loss functions employed. The datasets are further examined in Section 4. The results, along with the approaches presented in each analyzed study, are presented in greater detail in Section 5. We organize these aspects to identify patterns and trends in the literature, as well as to highlight any knowledge gaps or contradictions.

Year: The first organizational aspect is the year of publication. In the search methods, we filtered studies published since 2017. Table 2 displays the number of relevant publications available in the selected years, including those considered for full-text screening and those chosen for the final data analysis. While there is a slight increase of interest in this topic since 2018, differently from other deep learning applications, the number of relevant papers has not grown substantially.

Table 2
Count of related papers selected for full-text screening and data analysis.

Year	Full-text screening	Data Analysis
2018	7	2
2019	11	6
2020	2	2
2021	11	4
2022	15	9
2023	4	2
Total	**50**	**25**

Main goals: Secondly, the studies were organized according to their main goals. We initially split the publications into four goals: (i) lithofacies segmentation; (ii) seismic facies segmentation, (iii) salt body identification; and (iv) synthetic data generation. These goals are broad and were selected based on the main application of each paper.

Addressing lithofacies means that the labels specified for the dataset are connected to geological formations, with genetic mean and typically few vertical strata repetitions. This is a reflex of depositional control, such as sediment supply, eustasy, and basin flexure.

When referring to seismic facies, we address facies that are directly connected to the reflectivity pattern and its geometry, which might appear in various depths of the data with multiple repetitions without implying physically connected geological strata. Since the segmentation of facies is a dense prediction task, it has usually been dealt with fully-convolutional networks (Alaudah et al., 2019; Su-Mei et al., 2022; Abid et al., 2022). Despite that, it sometimes includes approaches based on the classification of each pixel through the employment of sliding windows techniques (Guazzelli et al., 2020).

The salt body identification goal refers to delimiting salt structures in a binary manner. These structures are usually well-marked, but their correct delimitation is crucial for the comprehension of reservoir disposal and for obtaining accurate velocity models. The usual approaches consist of performing segmentation or classification of patches of salt (Waldeland and Solberg, 2017; Shi et al., 2019).

Finally, the generation of synthetic data (e.g. channel structure simulation) is important since it makes more feasible controlled testing of algorithms and simulations for

Table 3
Count of chosen articles' major goals and their references.

Main Goal	References	Count
Facies Segmentation	Chevitarese et al. (2018), Silva et al. (2019), Alaudah et al. (2019), Zhang et al. (2019), Wang et al. (2019), Guazzelli et al. (2020), Liu et al. (2020a), Li et al. (2020), Wang and Chen (2021), Zhang et al. (2021), Trinidad et al. (2022), Abid et al. (2022), Chen et al. (2022b), Tolstaya and Egorov (2022), Su-Mei et al. (2022), Wang et al. (2021), Li et al. (2022), Nasim et al. (2022), Monteiro et al. (2022), Wang et al. (2023), Li et al. (2023)	21
Salt body Identification	Waldeland and Solberg (2017), Mukhopadhyay and Mallick (2019), Shi et al. (2019)	3
Synthetic Data Generation	Chen et al. (2022a)	1

Table 4
Count of adopted learning paradigms per year.

Year	Learning Paradigm	References	Count
2018	Supervised	Waldeland et al. (2018), Chevitarese et al. (2018)	2
2019	Supervised	Mukhopadhyay and Mallick (2019), Silva et al. (2019), Alaudah et al. (2019), Shi et al. (2019), Wang et al. (2019)	5
2020	Supervised	Zhang et al. (2019), Guazzelli et al. (2020)	2
	Supervised and semi-supervised	Liu et al. (2020a)	1
2021	Supervised	Trinidad et al. (2022)	1
	Supervised (Few/some-shot)	Zhang et al. (2021), Wang and Chen (2021), Li et al. (2020)	3
2022	Supervised	Abid et al. (2022), Tolstaya and Egorov (2022), Wang et al. (2021)	3
	Semi-supervised	Su-Mei et al. (2022), Li et al. (2022)	2
	Self-supervised (Few/some-shot)	Monteiro et al. (2022)	1
	Unsupervised	Chen et al. (2022b), Nasim et al. (2022), Chen et al. (2022a)	3
2023	Semi-supervised	Wang et al. (2023)	1
	Unsupervised	Li et al. (2023)	1

reservoir modeling and characterization (Lee and Mukerji). This was a secondary goal of this search, and only one paper was selected for the final analysis.

Table 3 displays the number of papers according to their goals. Due to the scope of this review, there are many more papers regarding the segmentation of facies (21), compared to the other two goals.

Learning Paradigm: We also identified the machine learning paradigms used in the papers. The paradigms were divided into three primary categories: supervised, semi-supervised, and unsupervised. We highlight specific cases, such as the few-shot regime and self-supervised learning. This was an attempt to preserve the typical structure and add specificity regarding the approaches proposed in each work. Table 4 shows the frequency of papers that use specific learning paradigms according to the publication year. Note that the supervised approach is the most common.

Table 4 also shows that there has been a growing interest in alternative approaches that do not solely rely on fully-supervised methodologies. Notice a growing number of publications adopting semi-supervised and unsupervised techniques, as well as applications employing few-shot and self-supervised methods. This trend is also evident in the literature, as handling unlabeled data is a significant challenge in various deep learning fields (Chen et al., 2020b; Balestriero et al., 2023). These approaches are particularly important for seismic interpretation problems because manually generated labels by experts are costly and time-consuming (Waldeland et al., 2018; Liu et al., 2020a).

Input format: Another relevant aspect that varies greatly among the studies is the format of the data inputted into the network. Overall, seismic data can be viewed as a 3D volume describing the amplitude of the wave signal across a 3D topology (latitude, longitude, and depth).

Table 5
Count of input styles adopted per year.

Year	Input Style	References	Count
2018	2D patches	Chevitarese et al. (2018)	1
	3D patches	Waldeland et al. (2018)	1
2019	2D patches	Mukhopadhyay and Mallick (2019), Silva et al. (2019), Alaudah et al. (2019)	3
	2D slices	Alaudah et al. (2019), Wang et al. (2019)	2
	3D patches	Shi et al. (2019)	1
2020	2D slices	Guazzelli et al. (2020)	1
	2D patches	Zhang et al. (2019)	1
	3D patches	Liu et al. (2020a)	1
2021	2D patches	Wang and Chen (2021)	1
	2D slices	Li et al. (2020); Trinidad et al. (2022)	2
	3D patches	Zhang et al. (2021)	1
2022	2D patches	Abid et al. (2022), Nasim et al. (2022), Chen et al. (2022a)	3
	2D slices	Chen et al. (2022b), Tolstaya and Egorov (2022), Wang et al. (2021), Li et al. (2022), Monteiro et al. (2022)	5
	3D patches	Su-Mei et al. (2022)	1
2023	2D slices	Wang et al. (2023)	1
	3D patches	Li et al. (2023)	1

Nevertheless, due to its large size, which easily surpasses hundreds of gigabytes, this volume is typically divided into smaller segments when employed as input for deep learning architectures. For the sake of comparison, the following input formats were found:

- **Small 2D patches**: small crops that do not attempt to represent global context or structures of the data - usually crops smaller than 200×200 pixels;
- **Large 2D slices**: large slices/sections (inlines or crosslines) that attempt to preserve global context information, such as depositional patterns and their geometry;
- **3D sub-volumes**: small volumes, usually as large as the computational power allows, which typically cannot preserve most of the global features; typically smaller than 128×128×128 pixels.

As shown in Table 5, there is no clear preference for a given input format. Nevertheless, there has been some increase in the adoption of 2D slices over the years as an attempt to preserve contextual information from geological layers.

Deep Learning Architectures: Researchers in deep learning applied to Geosciences, similar to their counterparts in other domains, have been conducting experiments with various deep learning architectures to address the challenges associated with seismic segmentation. Table 6 exhibits the frequency of each architecture used in the reviewed studies. Observe that numerous architectures were used, but there is a clear preference for the UNet (Ronneberger et al., 2015). Other networks that appear more frequently are the DeepLabV3+ (Chen et al., 2017b), ResNet (He et al., 2016), and the Danet-FCN (Chevitarese et al., 2018). A detailed description of the architectures can be found in Section 5.2.

Loss functions: In deep learning architectures, a loss function, also known as cost or objective function, is a mathematical function that quantifies the discrepancy between the predicted output of the model and the true (or desired) output. The purpose of the loss function is to measure the error incurred by the model during training and guide the gradient descent algorithm. The choice of an appropriate loss function depends on the specific task and the problem's nature. Therefore, different loss functions have been employed for seismic interpretation. By minimizing the loss function, the model learns to make more accurate predictions and to improve its performance on the given task.

Table 7 shows the loss functions used in the studies considered in the literature review. Observe that the most frequently used are Cross-Entropy, Adversarial Loss, and Contrastive Loss, together with their variations, which are

Table 6
Count of adopted architectures and their references.

Backbone	**References**	**Count**
UNet	Zhang et al. (2019), Shi et al. (2019), Wang et al. (2019), Li et al. (2020), Wang and Chen (2021), Wang et al. (2021), Trinidad et al. (2022), Tolstaya and Egorov (2022), Su-Mei et al. (2022), Wang et al. (2023)	10
DeepLabV3+	Zhang et al. (2021), Abid et al. (2022), Li et al. (2022)	3
Danet-FCN	Chevitarese et al. (2018), Silva et al. (2019)	2
Resnet	Monteiro et al. (2022), Li et al. (2023)	2
ConvNet Pixel Classification	Waldeland et al. (2018), Guazzelli et al. (2020)	2
3D VGG	Liu et al. (2020a)	1
Bayesian Neural Network	Wang and Chen (2021)	1
Bayesian SegNet	Mukhopadhyay and Mallick (2019)	1
EarthAdaptNet	Nasim et al. (2022)	1
EfficientNet	Tolstaya and Egorov (2022)	1
GAN	Liu et al. (2020a)	1
Hrnetv2-W32	Chen et al. (2022b)	1
LSTM	Trinidad et al. (2022)	1
SAGAN	Chen et al. (2022a)	1
Segmentation ConvNet	Alaudah et al. (2019)	1
SegNet	Zhang et al. (2019)	1

Table 7
Count of adopted cost functions and their references.

Cost Function	**References**	**Count**
Cross-Entropy	Waldeland et al. (2018), Shi et al. (2019), Li et al. (2020), Wang and Chen (2021), Abid et al. (2022), Chen et al. (2022b), Su-Mei et al. (2022), Monteiro et al. (2022)	8
Adversarial Loss	Liu et al. (2020a), Chen et al. (2022a)	2
Contrastive Loss + Cluster Loss	Li et al. (2023)	1
Contrastive Loss + Cross-Entropy	Li et al. (2022)	1
CORAL Loss	Nasim et al. (2022)	1
Cross-Entropy (supervised) and Mean Squared Error (unsupervised)	Wang et al. (2023)	1
Cross-Entropy + Dice + Total variation Loss	Tolstaya and Egorov (2022)	1
Dice	Wang et al. (2021)	1
Focal loss	Trinidad et al. (2022)	1
Not reported	Chevitarese et al. (2018), Mukhopadhyay and Mallick (2019), Wang et al. (2019), Silva et al. (2019), Alaudah et al. (2019), Guazzelli et al. (2020)	6

further discussed in Section 5.3. It is worth noting that some loss functions are only listed once, indicating that they are used less frequently. Furthermore, six occurrences are labeled as "Not reported", meaning that the study did not specify the loss function employed in the learning architecture.

Evaluation metrics: Evaluation metrics in deep learning are quantitative measures used to assess the models' performance. These metrics provide insights into how well a model performs on a given task, such as classification, regression, or segmentation. They help in comparing different

Table 8
Evaluation metrics with its appearance reference and its count.

Metric	References	Count
Pixel accuracy	Chevitarese et al. (2018), Mukhopadhyay and Mallick (2019), Alaudah et al. (2019),Shi et al. (2019), Zhang et al. (2019), Guazzelli et al. (2020), Liu et al. (2020a), Wang and Chen (2021), Trinidad et al. (2022) Abid et al. (2022), Chen et al. (2022b), Tolstaya and Egorov (2022), Su-Mei et al. (2022), Wang et al. (2021), Li et al. (2022), Nasim et al. (2022), Wang et al. (2023)	17
Mean intersection over union	Wang et al. (2019), Zhang et al. (2021), Wang and Chen (2021), Li et al. (2020), Abid et al. (2022), Chen et al. (2022b), Tolstaya and Egorov (2022), Li et al. (2022), Nasim et al. (2022), Monteiro et al. (2022)	10
Mean class accuracy	Alaudah et al. (2019), Guazzelli et al. (2020), Trinidad et al. (2022), Abid et al. (2022), Chen et al. (2022b), Tolstaya and Egorov (2022), Li et al. (2022), Nasim et al. (2022), Wang et al. (2023)	9
Frequency weighted intersection over union	Alaudah et al. (2019), Guazzelli et al. (2020), Trinidad et al. (2022), Chen et al. (2022b), Tolstaya and Egorov (2022), Li et al. (2022), Nasim et al. (2022)	7
Precision	Shi et al. (2019), Wang and Chen (2021)	2
Recall	Shi et al. (2019), Wang and Chen (2021)	2
F1 score (Dice)	Shi et al. (2019), Wang et al. (2021)	2
Qualitative evaluation	Waldeland et al. (2018), Li et al. (2023)	2

models, tuning hyperparameters, and determining whether a model meets the desired performance criteria.

Note from Table 8 that numerous metrics have been utilized for evaluating seismic segmentation, with pixel accuracy being the most frequently employed. A detailed description of the metrics is presented in Section 5.4.

4. Public Interpreted Datasets

In this section, a brief description of the datasets encountered during the review process is provided, focusing on the datasets for which the interpretation is available. Unlike computer vision and other fields of study, seismic segmentation lacks large-scale publicly available annotated datasets suitable for training and evaluating deep learning models (Alaudah et al., 2019). Although there are many seismic volumes available for download through institutions, e.g., Open data from SegWiki[2], New Zealand Petroleum & Minerals[3] and TerraNubis[4], these data are not annotated. Hence, there is no means to validate the interpretation of these datasets, and as such, they are not the focus of this study.

[2] https://wiki.seg.org/wiki/Open_data
[3] https://www.nzpam.govt.nz/
[4] https://terranubis.com/

In an effort to offer openly labeled data for subsequent studies, Baroni et al. (2019), Alaudah et al. (2019), and Silva et al. (2019) released labeled datasets that are accessible for researchers to download and utilize under a Creative Commons Attribution license. Baroni et al. (2019) released an interpretation of the Penobscot 3D Dataset and discussed the labels and the use of these data. Both Alaudah et al. (2019) and Silva et al. (2019) released interpreted versions of the F3 Netherlands dataset. Baroni et al. (2019) and Silva et al. (2019) focused on discussing the data itself and the results of third-party studies using their dataset. As for Alaudah et al. (2019), one of their main goals was to establish a benchmark for future comparison of results, providing train and test splits, train set size, and metrics for comparison. In addition, a labeled version of the Parihaka Seismic Data was released (Inc), which was interpreted by a team from Chevron U.S.A. as the ground truth for the 2020 SEG Annual Meeting Machine Learning Interpretation Workshop.

Table 9 summarizes open-interpreted datasets, including their size, number of classes, and citations in related works.

Table 9
Summary of the labeled datasets available, comparing their size, number of classes, and related citations.

Dataset	Size (I,X,T)	Classes	Citations
F3 (Alaudah *et al.*)	601×901×255	6 LF	Alaudah et al. (2019), Guazzelli et al. (2020), Trinidad et al. (2022), Abid et al. (2022), Li et al. (2022), Chen et al. (2022b), Tolstaya and Egorov (2022), Wang et al. (2021), Nasim et al. (2022)
F3 (Silva *et al.*)	651×951×462	10 LF	Silva et al. (2019), Wang and Chen (2021), Monteiro et al. (2022), Wang et al. (2023)
F3 (Conoco-Phillips)	651×951×462	9 SF	Zhang et al. (2019),Liu et al. (2020a), Zhang et al. (2021)
Parihaka (Chevron)	590x782x1006	6 LF	Li et al. (2022), Tolstaya and Egorov (2022), Monteiro et al. (2022), Su-Mei et al. (2022), Li et al. (2023)
Stanford VI-E (Lee and Mukerji)	150×200×200	3 LF	Guazzelli et al. (2020), Chen et al. (2022a)
Penobscot (Baroni *et al.*)	601×482×1501	7 LF	Chevitarese et al. (2018), Nasim et al. (2022)
TGS Salt (TGS Salt Identification Challenge)	22k*(101×101)	2 (S/NS)	Mukhopadhyay and Mallick (2019)

I: Number of inlines; LF: Lithofacies; SF: Seismofacies; S/NS: Salt/Not salt; T: Number of timeslices; X: Number of crosslines.

Notice that the F3 Netherlands is the most explored among the ones with a public interpretation of facies. Its most-used interpreted version is the one proposed by Alaudah et al. (2019), followed by the ones proposed by Silva et al. (2019) and Conoco-Phillips (F3). Since it was made public in 2020 with a detailed geological segmentation for a competition, the Parihaka New Zealand (interpretation by Chevron (Inc)) has had increased usage. The Penobscot dataset (with Baroni et al. (2018) interpretation) and other Synthetic data are also sometimes used. The studies quoted in Table 9 are further discussed in Section 5.

The main characteristics of the open-interpreted datasets with available descriptions are summarized as follows:

Parihaka - Chevron (Inc) : the survey is offshore Taranaki, North Island, New Zealand. It is part of the Taranaki basin, which is made up mainly of terrestrial, marginal marine, and shallow marine sediments. It was made publicly available by New Zealand Petroleum and Minerals (NZPM), and the labels were provided by Chevron U.S.A. Inc. The 3D volume contains six classes: Basement/Other, Slope Mudstone A, Mass Transport Deposit, Slope Mudstone B, Slope Valley, and Submarine Canyon System. It has a 3D volume with 590 inlines, 782 crosslines, and 1006 timeslices.

F3 Netherlands - Alaudah et al. (2019) : this model is based on the F3 Block, located on the North Sea continental shelf offshore of the Netherlands. This block is related to the Step Graben and the Dutch Central Graben, two tectonic structures marked by distinct lithostratigraphic units of varied thickness. Also, the area is deformed by strong halokinetic. The authors delimited seven lithostratigraphic units: the Upper North Sea group, the Middle North Sea group, the Lower North Sea group, the Chalk group, the Rijnland group, the Scruff group, and the Zechstein group. They made available a volume consisting of 700 inline sections and 1200 crossline sections and suggested one train and two test splits.

F3 Netherlands - Silva et al. (2019) : this segmentation was also based on the F3 Block. For the purposes of machine learning, the authors reinterpreted 9 horizons separating the 10 classes: North Sea Supergroup, Chalk Group, Rijnland Group, Schieland, Scruff and Niedersachsen Groups, Altena Group, Germanic Trias Group, Zechstein Group, Rotliegend Group, and Carboniferous Group. All inline and crosslines were interpreted, with 651 and 951 slices. Also, approximately 190,000 labeled patches were generated for the inlines and crosslines.

Penobscot - Baroni et al. (2019) : this dataset was obtained in the Scotian shelf in Nova Scotia, Canada. The horizons were reinterpreted to generate 7 seismofacies characterized by their reflection patterns. For machine learning tasks, they opened 1083 labeled seismic lines and also the patches cropped from them.

TGS Salt Identification Challenge (Kaggle): The dataset comprises randomly selected images from different subsurface locations. These images have dimensions of 101 x 101 pixels, with each pixel labeled as either salt or sediment. Furthermore, the depth of each imaged location is included alongside the seismic images. The objective of this competition was to identify and segment the areas containing salt within a non-salt background.

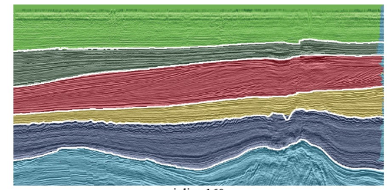

Figure 2: Example of semantic segmentation applied to inline 160 the F3 Netherlands dataset. White lines represent manually interpreted horizons (ground truth), while colored regions represent the class (label) assigned to each pixel after the task of segmentation. This particular example utilizes six interpreted horizons as its classes, but different interpretations may include a different number of horizons. Image source: (Silva et al., 2019)

5. Deep Learning for Seismic Segmentation

As previously stated, conventional machine learning techniques are limited in their ability to process natural data in their raw form (LeCun et al., 2015). This would not, however, hinder the development of robust artificial intelligence methods, as deep learning algorithms have gained significant popularity over the last decades, mainly due to the availability of data as well as the advancements in computational power. These algorithms brought numerous advantages over traditional methods, such as the capacity to automatically learn features from data, reducing the need for manual feature engineering, and the ability to recognize patterns in large amounts of complex, unstructured data.

The fundamental building block of a deep learning algorithms is the neural network. It consists of layers of interconnected nodes, each performing a mathematical operation on the input data. The nodes in each layer transform the data and pass them to the next layer. The initial layers usually capture simple features, while deeper layers learn increasingly complex and abstract representations (Zeiler and Fergus, 2014). These sets of layers allow the model to learn intricate patterns in the data, making it well-suited for complex tasks such as image classification, speech recognition and natural language processing.

One particular area in which deep learning has managed to successfully discover intricate structures in high-dimensional data is computer vision (Krizhevsky et al., 2017; Farabet et al., 2013; Szegedy et al., 2015; Tompson et al., 2014). CNNs are a specialized type of deep neural network that is commonly used in image and video processing applications. They are particularly well-suited for tasks involving visual recognition, classification, and segmentation, and have significantly improved the state of the art in computer vision.

As the name states, CNNs are composed of convolutional layers, which perform a convolution operation by applying a set of learnable filters, also called kernels, to the input image. Each filter is trained to detect specific features, such as edges, corners, and textures. By sliding the filters over the input image in small steps, this process computes feature maps, which capture local patterns in the input. After each convolution operation, an activation function is element-wisely applied to the feature map to introduce non-linearity to the model. Then, the spatial dimensions of the feature maps are reduced through pooling layers or stride/padding strategies, decreasing the computational complexity and the risk of overfitting. These layers aggregate information from local regions, preserving the most important features while discarding less relevant details. What is left is a dense representation of the input in a low-dimension space. The final layers produce the model's predictions, which can vary depending on the desired task.

Since seismic volumes can be regarded as a type of image (Wang et al., 2019), CNNs have been applied to numerous geophysical applications to segment and classify facies in seismic cubes (Salles Civitarese et al., 2018; Zhao; Zhang et al., 2021). Differently from image classification, which assigns a single label to the entire image, image segmentation assigns labels to individual pixels or groups of pixels. This task automates the process of understanding the spatial distribution of objects and their boundaries within an image and it is therefore of great interest for the analysis of seismic facies. Figure 2 shows an example of segmentation applied to a slice of a seismic volume, where the label of each pixel is denoted by a different color.

Dedicated approaches have emerged since traditional CNNs have shown promising results in semantic segmentation tasks. In this section, we elucidate specialized methods developed for seismic segmentation, delve into their foundational components, critically assess their outcomes, and highlight potential avenues for future research. More precisely, first we delineate three distinct tasks associated with seismic segmentation in Section 5.1. Second, in Section 5.2, we describe the primary deep learning architectures serving as the backbone for these tasks. Third, in Section 5.3, we explore potential loss functions that have been and can be employed. Fourth, in Section 5.4, we provide a critical review on evaluation metrics. Fifth, Sections 5.5 and 5.6 provide a comprehensive review of both supervised and unsupervised solutions, respectively. Finally, a concise summary of experimental results achieved in the literature is presented in Section 5.7.

5.1. Seismic Segmentation Tasks

The concept of "facies" is widely used in geology, particularly in sedimentology, in which the term refers to the sum of the characteristics of a sedimentary unit (Middleton and Hampton, 1973). These characteristics include dimensions, sedimentary structures, grain sizes and types, color and biogenic content of the sedimentary rock. Not all aspects of the rock are necessarily indicated in the facies name, and in some circumstances, it may be important to emphasize different characteristics. The full range of the characteristics of a rock would be given in the facies description that form part of any study of sedimentary rocks (Nichols, 2009).

The technique used to understand the subsurface via seismic data is named seismic facies analysis, which describes and interprets the seismic reflection parameters, such as configuration, continuity, amplitude, and frequency, within rock layers (strata) of a depositional sequence (Vail, 1987). Such analysis is of great significance in the oil and gas industry since it provides information regarding the possible distribution of rock layers and geobodies. In conjunction with additional information, it may indicate lithology and possible spots of hydrocarbon accumulation (Chevitarese et al., 2018). Interpretation of seismic facies is conducive to analyzing subsurface geologic environments and further predicting oil and gas reservoirs.

As detailed in Section 3.4, this survey centers on tasks associated with identifying lithofacies, seismofacies and saline bodies. In computational terms, all these tasks can be formulated as image segmentation problems applied to seismic data (Wang et al., 2019), regardless of which specific task is being addressed. The difference is that, in general, the identification of salt bodies comprises a binary segmentation task, in which there are "salt" and "not salt", whereas lithofacies and seismofacies commonly relate to multi-class segmentation problems, with the main difference stemming from the class labeling process.

Lithofacies When the facies description is associated with the physical and chemical characteristics of a rock, such as grain size, mineralogy, porosity, and permeability, this is referred to as lithofacies. In other words, the lithofacies identification is based directly on geological observations (Nichols, 2009). In the realm of deep learning solutions, our literature review found that supervised learning is the predominant approach for tackling this problem, as evidenced in several works (Chevitarese et al., 2018; Silva et al., 2019; Alaudah et al., 2019; Guazzelli et al., 2020; Wang and Chen, 2021; Trinidad et al., 2022; Abid et al., 2022; Tolstaya and Egorov, 2022; Wang et al., 2021). Nevertheless, unsupervised learning methods have also been explored (Chen et al., 2022b; Nasim et al., 2022), while semi-supervised approaches are discussed in (Su-Mei et al., 2022; Li et al., 2022; Monteiro et al., 2022; Wang et al., 2023). As for the publicly available datasets, F3 Netherlands, Penobscot, and Parihaka NZ all come equipped with lithofacies annotations. It is important to note that, in the case of F3 Netherlands, there are distinct public interpretations. For instance, Alaudah et al. (2019) defined six classes of lithofacies, whereas Silva et al. (2019) annotated ten classes.

Seismofacies Seismofacies consist of seismic reflections whose patterns, such as amplitude, frequency, and geometry, are different from those of adjacent groups (West et al., 2002). In summary, seismofacies analysis aims to interpret the depositional environment and facies distribution directly from seismic data (Dumay and Fournier, 1988). From our literature review, we found that, although supervised techniques are the most common (Zhang et al., 2019; Wang et al., 2019; Li et al., 2020; Zhang et al., 2021), there were also papers addressing the problem with unsupervised (Li et al., 2023) and semi-supervised (Liu et al., 2020a) learning. The interpretation of the F3 Netherlands with 9 seismofacies provided by Conoco-Phillips Norge (F3) is the only one available.

Salt bodies Besides lithofacies and seismofacies classification, salt bodies identification constitutes another task of interest. Due to the low permeability associated with salt bodies, they may form seals for reservoirs. Furthermore, they have a relatively high sound velocity, which makes them important to obtain an accurate velocity model in the vicinity of the salt bodies (Waldeland et al., 2018). This task can be modeled as a binary segmentation problem, and when the salt/not salt labels are available, then supervised learning naturally applies. Our survey found the following datasets related to this task: Barents Sea (Waldeland et al., 2018), SEAM Phase I (Mukhopadhyay and Mallick, 2019), and TGS Salt Identification Challenge (Shi et al., 2019). The first dataset is private; the second is available upon request; and the third can be publicly downloaded.

5.2. Deep Learning Architectures

Among the many architectures that have been proposed to tackle the image segmentation problem, the encoder-decoder is widely used deep learning architecture for processing unstructured data, such as images, videos, and natural language texts. Its fundamental design pattern has proven effective in various deep learning tasks. The architecture consists of two primary components: an encoder and a decoder, each playing a distinct role in transforming the input data. The encoder resembles the convolutional layers of a CNN, in which the input image is processed to extract high-level features and is encoded into a lower-dimensional representation. What distinguishes this architecture from a traditional CNN is the fact that instead of being fed to a fully connected network, the output of the convolutional layers is then used as input to the next component, the decoder, which takes the encoded representation from the encoder and reconstruct the desired output by transforming the data into its original format, as illustrated Figure 2.

Encoder-decoder networks are also commonly deployed to image generation tasks, in which the decoder is responsible for *upsampling* the encoded data to generate a full-resolution output. The upsampling process typically involves the use of transposed convolutions or interpolation techniques. Throughout the remainder of this section, we will elucidate some of the frequently employed architectures that leverage the encoder-decoder design for seismic image segmentation tasks.

UNet. Introduced in 2015 (Ronneberger et al., 2015), this architecture stands out as one of the most extensively employed in the semantic segmentation field (Table 6). Initially conceived for medical image segmentation, it swiftly found utility in various other domains (Iglovikov and Shvets, 2018; Yao et al., 2018; Çiçek et al., 2016; Kandel et al., 2020; Nazem et al., 2021). The name UNet comes from the fact that its shape resembles a "U", as displayed in Figure 3, where the descending path (left side) is the encoder and the ascending path (right side) is the decoder. The network's architecture comprises a connection between the contracting path that extracts contextual information and a symmetrical expanding path that facilitates accurate localization (Ronneberger et al., 2015). These connections allow the encoder to pass low-level features to the decoder since the downsampling operations result in the loss of information as the data gets encoded. More precisely, a "connection" between layers is represented by copying the activation map of one encoding layer, which results from a convolution followed by activation, and concatenating it to the corresponding layer on the decoder. Hence, the UNet is capable of preserving both low-level features from data as well as high-level ones, making it especially well suited for tasks that require fine details, such as biomedical and seismic image analysis.

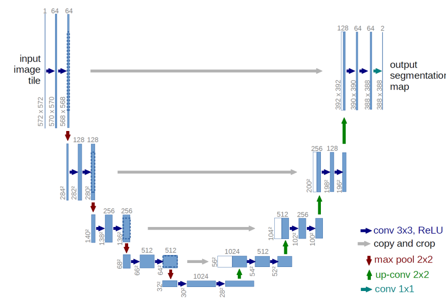

Figure 3: The UNet architecture (with a 572x572 2D input image as an example). Blue rectangles represent multi-channel feature maps, with the number of channels at the top of each rectangle and the corresponding resolution at the lower left corner. Arrows represent operations, and the skip connections step is represented by the copy and crop operation (gray arrows). Image source: (Ronneberger et al., 2015)

DeconvNet. This architecture was proposed in 2015 (Noh et al., 2015) and incorporates the convolutional layers of a popular CNN called VGG16 (Simonyan and Zisserman, 2014) as its encoder. As illustrated in Figure 4, it comprises thirteen convolutional layers and five max-pooling layers, making it a robust yet simple network for feature extraction. In contrast with the VGG16, however, the DeconvNet drops the fully connected layers and replaces them with a mirrored version of the convolutional layers to form its decoder, where the convolutions become deconvolutions and perform upsampling instead of downsampling. Because

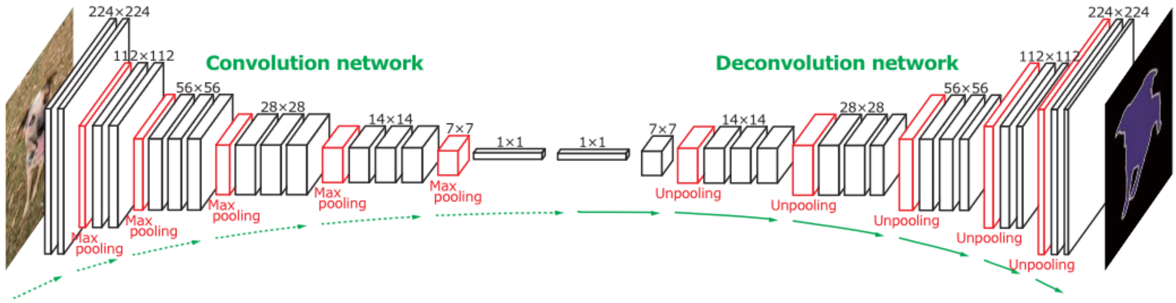

Figure 4: Architecture of the DeconvNet. The convolutional network based on the VGG16 is followed by a multi-layer deconvolution network to generate a segmentation map of the input image. Image source: (Noh et al., 2015)

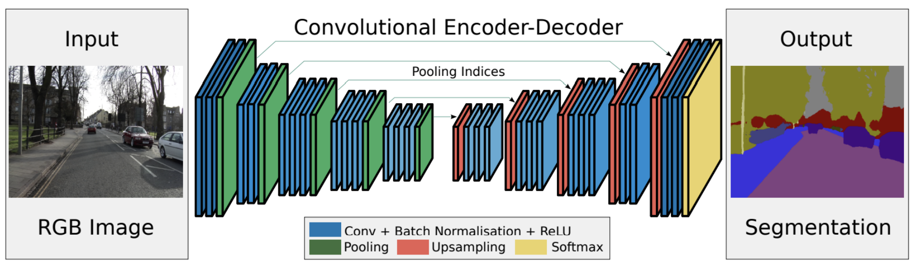

Figure 5: The SegNet architecture. There are no fully connected layers, only convolutional ones. The encoder (left half) generates pool indices, which are then fed to the decoder (right half) to produce sparse feature maps. Image source: (Badrinarayanan et al., 2017)

of this robustness, DeconvNet is particularly useful for tasks where precise pixel-level segmentation is crucial, such as seismic and medical image analysis. In comparison with UNet's original applicability on image segmentation tasks, DeconvNet was primarily designed to tackle image reconstruction and image generation problems. Due to the lack of shortcut connections in its architecture, DeconvNet models usually struggle to preserve spatial information during the upsampling in the deconvolution size. As a result, DeconvNet is easier to train at the cost of being less accurate in segmenting small objects than UNet-based models.

SegNet. Proposed in 2017 (Badrinarayanan et al., 2017) and shown in Figure 5, the SegNet architecture focuses on accurate pixel-wise segmentation. However, SegNet uses a specific approach to manage the complexity of the model and to make it more efficient. Similar to the DeconvNet, it relies on the design of the convolutional layers of the VGG16 for its encoder, but with some key differences. In contrast to conventional max pooling, SegNet not only retains the maximum values during the pooling operation but also stores their corresponding indices, preserving the spatial locations of the maximum activations. The indices are then used later in the decoder for upsampling. While this approach has the same goal as the UNet, which is to prevent the loss of information in the pooling layers, SegNet manages to be more efficient by choosing not to store entire activation maps at each convolutional layer, enabling accurate segmentation while using less memory and fewer parameters compared to other architectures.

DeepLab. This term refers to a series of architectures that employ spatial pyramid pooling (ASPP) (Liang-Chieh et al., 2015; Chen et al., 2017a,b, 2018) developed by researchers at Google, being initially introduced in 2014 with DeepLabV1 and having its most recent version with DeepLabV3+. This network currently holds state-of-the-art performance in several semantic segmentation tasks, including seismic image segmentation (Du et al., 2021; Polat, 2022). The original architecture introduced the concept of dilated convolutions to capture multi-scale contextual information. These convolutions have a configurable dilation rate that allows the network to have a larger receptive field without increasing the number of parameters. As shown in Figure 6, the three versions use ASPP modules, which use parallel dilated convolutional layers with different dilation rates. This captures context at various scales and improves segmentation performance. Future versions of this architecture would aim to improve the model's efficiency and ability

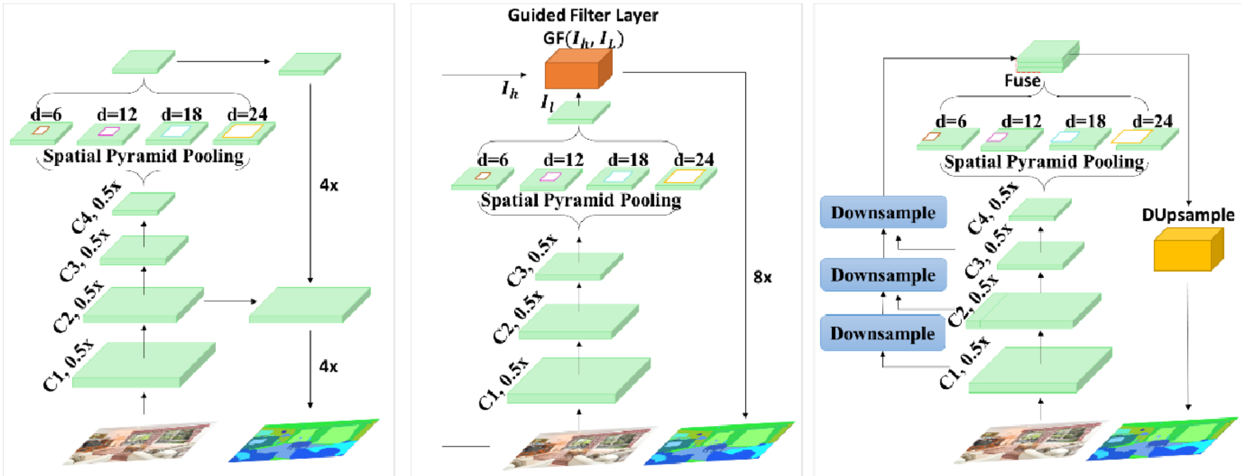

Figure 6: Overview of DeepLabv1 network architecture (*left*), implementation of guided filter layer over DeepLabv2 (*middle*), and illustration of *DUpsampling* module employed after the last layer of DeepLabv3 (*right*). Image source: (Sediqi and Lee, 2021).

to capture features at different levels of abstraction, thus yielding better segmentation accuracy.

5.3. Loss Functions

As detailed in Section 3.4, when addressing Machine Learning or Deep Learning challenges, the primary goal is to optimize a model, typically achieved by minimizing a loss function. Table 7 provides an overview of the cost functions commonly employed in the literature under review. Noteworthy selections include Cross-Entropy, Adversarial Loss, and Contrastive Loss, each accompanied by their respective variations and adaptations.

Cross-Entropy Loss (CEL). The cross-entropy cost function is widely employed as a primary error metric in many deep learning applications, including seismic segmentation tasks. The cross-entropy formulation can be represented as

$$\mathcal{L}_{\text{CEL}}(p,t) = -\sum_{i=1}^{n} t_i \log(p_i), \tag{1}$$

where n represents the number of classes t and p are one-dimensional vectors. $t_i \in \{0,1\}$ holds the true target label and p_i represents the probability of the predicted class. Note that this function gauges the probability of the predicted class against the intended class output. It penalizes deviations based on their magnitude from the expected outcome. With a logarithmic nature, substantial disparities yield higher penalties, while minor differences result in lesser penalties.

The cross-entropy loss offers a notable advantage due to its straightforward implementation and optimization. Its smooth and convex shape also supports the efficient convergence of gradient-based optimization methods to reach the global minimum. Conversely, a significant drawback of this loss function is its sensitivity to outliers and imbalanced data. In cases where one class vastly outnumbers the others within the dataset, this loss function may prioritize the dominant class, resulting in sub-optimal performance for the minority classes. To tackle this challenge, various adaptations of the cross-entropy loss have been introduced, such as weighted cross-entropy, focal loss, and class-balanced loss.

Adversarial Loss (ADV). The adversarial cost function arises within the framework of Generative Adversarial Networks (GANs), as introduced by Goodfellow et al. (2014). In summary, GANs comprise a generative model denoted as G and a discriminative model referred to as D, which are trained in cooperation. The generator aims to minimize the penalty whereas the discriminator endeavors to maximize it. This oppositional dynamic between the two models gives rise to the term "adversarial". It can be defined as

$$\mathcal{L}_{\text{ADV}}(p,t) = \sum_{i=1}^{n} \mathbb{E}_t[\log D(t_i)] + \mathbb{E}_p[\log\left(1 - D(G(p_i))\right)], \tag{2}$$

where $\mathbb{E}_t$ is the expected value over all real data instances, $D(t)$ is the discriminator's estimate of the probability that real data instance x is real, $\mathbb{E}_p$ is the expected value over all generated synthetic instances, $G(p)$ is the generator's output for a given noise p, and $D(G(p))$ is the discriminator's estimate of the probability that a counterfeit instance is real.

Adversarial training and GANs are both important for generating realistic images. In the field of seismic segmentation, some researchers have employed this method successfully (Liu et al., 2020a), where a GAN framework is used to train a generator network to create synthetic seismic data and a discriminator network to distinguish between real and synthetic data. This adversarial training process helps refining the synthetic data generation over time, making it increasingly indistinguishable from real seismic data. The synthetic data can then be used to augment training datasets, which can improve the performance and robustness of seismic segmentation models.

Contrastive Loss (CON). Contrastive loss is a type of loss function that is used to train machine learning models to learn representations of data where similar examples are close together and dissimilar examples are far apart. Basically, it quantifies the vector's separation from another sample of the same category and juxtaposes it with the separation from negative instances. This function ensures that the penalty remains minimal when positive examples result in more comparable (close) representations, while negative instances yield less comparable (distant) ones. Typically, cosine distances are employed to gauge these similarities, subsequently serving as prediction probabilities. For a pair of samples (p_1, p_2), the contrastive loss is defined as

$$\mathcal{L}_{\text{CON}}(p_1, p_2, t) = \\ tE(p_1, p_2)^2 + (1 - t)\max(\alpha - E(p_1, p_2), 0), \tag{3}$$

where $t = 0$ indicates that samples are similar by annulling the Euclidean distance term $E(p_1, p_2) = ||p_1 - p_2||^2$ on the left side of the equation. Differently, $t = 1$ signalizes that samples are dissimilar as it minimizes the second term $\max(\alpha - E(p_1, p_2), 0)$, which is equivalent to maximizing the Euclidean distance until some threshold α.

Some researchers have employed the contrastive loss in pre-training methodologies tailored to 3D seismic data. In this approach, slices within the same data block are considered positive samples, while those originating from different blocks are treated as negatives. In the CONSS method proposed by Li et al. (2022), for instance, the contrastive loss optimization encourages the neural network to reduce intra-class distances while increasing inter-class separation. It leads to the creation of sharper decision boundaries that ultimately enhance classification accuracy and robustness.

Boundary Awareness. Regardless of the recent approaches designed to tackle semantic segmentation of seismic data, the vast majority have failed to address the lack of intricate boundary details along the outputs produced by neural network models. This issue is commonly found in architectures employing downsampling operations to encompass broader receptive fields.

Few approaches have been proposed to tackle boundary misclassification. For instance, Bertasius et al. (2015) and Takikawa et al. (2019) introduced boundary-aware information flow and multitask training techniques. Although these operations aid in encoding contextual information around each pixel, they tend to propagate feature information across the image, resulting in feature smoothing along object boundaries (Wang et al., 2022). In a similar fashion, Yuan et al. (2020) developed a technique to model the relationship between boundary and interior pixels. However, this method often shows significant errors, especially for small and slender objects, particularly at object boundaries.

Inspired by the aforementioned drawbacks, Wang et al. (2022) introduced a novel approach known as Active Boundary Loss (ABL), which progressively attempts to align predicted boundaries with ground-truth boundaries during training. In contrast to the cross-entropy loss, which solely supervises pixel-level classification, ABL focuses on the correspondence between predicted and actual boundaries. This approach prompts the network to focus more on boundary pixels, consequently enhancing segmentation outcomes. Unfortunately, ABL is computationally expensive to train as it requires the calculation of the boundary gradient for each training image. In addition, it is also sensitive to the hyperparameters, such as the learning rate and ABL's weight factor in the overall loss function.

5.4. Evaluation Metrics

As stated in Section 3.4, a variety of evaluation metrics have been used on the seismic segmentation task, but only four of them have been repeatedly employed. According to Table 8, pixel accuracy (PA), intersection over union (IoU), mean class accuracy (MCA), frequency weighted intersection over union (FWIoU) are present in a large number of papers. For this reason, we channel our exploration efforts toward these four evaluation metrics in this survey. Please refer to the survey conducted by Minaee et al. (2022)

for detailed definitions of precision, recall and F1 score (Dice) metrics in the context of image segmentation using deep learning. For the following definitions, consider the following notation: TP is the number of true positives; TN is the number of true negatives; FP is the number of false positives; FN is the number of false negatives; n is the number of pixels in the image; m is the number of classes. While in the multi-class context, positive elements refer to elements of the class being considered, whereas negative elements are the ones belonging to all other classes.

Pixel Accuracy. PA is a simple evaluation metric that measures the percentage of correctly classified pixels in an image. It computes the ratio of correctly classified pixels to the total number of pixels in an image. While it provides a primary measure of exactness, it does not consider the specific predicted classes and may not be suitable for imbalanced datasets. It is defined as

$$PA = (TP + TN)/n. \tag{4}$$

Mean Class Accuracy. MCA is a metric that computes the average accuracy for each class in a dataset. It computes the percentage of correctly classified pixels for each class and then takes the average across all classes. MCA provides a more fine-grained evaluation compared to PA by considering individual class performances, which can be helpful in scenarios where certain classes are more important than others. It is defined as

$$MCA = \frac{1}{m}\sum_{i=1}^{m} TP_i/(TP_i + FN_i). \tag{5}$$

Mean Intersection over Union. The mIoU is the average of the Intersection over Union (IoU) for all classes and provides a comprehensive measure of the overall segmentation performance. It measures the overlap between the predicted and ground truth segmentation masks for each class and is defined as

$$mIoU = \frac{1}{m}\sum_{i=1}^{m} TP_i/(TP_i + FP_i + FN_i). \tag{6}$$

Frequency Weighted Intersection over Union. The FWIoU also measures the overlap between the predicted and ground truth segmentation masks for each class and calculates the intersection ratio to the union of the predicted and ground truth regions. $FWIoU$ extends this concept by weighting the IoU score of each class by the frequency of that class in the dataset. It gives more importance to more prevalent classes, allowing a more representative evaluation of the overall segmentation performance. It is defined as

$$FWIoU = \sum_{i=1}^{m} \frac{n_i}{n} TP_i/(TP_i + FP_i + FN_i). \tag{7}$$

Boundary F1 Score. The BF1 metric (Csurka et al., 2013) was not found in the reviewed papers related to seismic data segmentation tasks. Nevertheless, this metric holds promise for application within this specific problem domain. The BF1 metric quantifies the proximity between the predicted boundary of an object and the corresponding ground-truth boundary. Its potential utility extends to the segmentation of seismic data, primarily due to its ability to enhance the precision of boundary delimitation between distinct classes.

The BF1 is based on the F1 measure and extends the Berkeley contour matching score (Martin et al., 2004), which computes the F1-measure from precision and recall values with a distance error tolerance of θ. Let $B^c gt$ be the boundary map of the binary ground truth segmentation map for class c, $S^c gt$, with $S^c gt(z) = [[Sgt(z) == c]]$ and [[z]] is the Iverson bracket notation, i.e. [[z]]=1 if z=true and 0, otherwise. Similarly, $B^c ps$ is the contour map for the binary predicted segmentation map $S^c ps$. If θ is the distance error tolerance, the adapted precision (P^c) and recall (R^c) for each class are, respectively,

$$P^c = \frac{1}{|Bps|}\sum_{z\in B^c ps} [[d(z, B^c gt) < \theta]] \tag{8}$$

and

$$R^c = \frac{1}{|Bgt|}\sum_{z\in B^c gt} [[d(z, B^c ps) < \theta]], \tag{9}$$

where d is the Euclidean distance. Given that, the F_1^c is defined as

$$F_1^c = 2P^c R^c/(P^c + R^c). \tag{10}$$

The F_1^c scores of all classes are then averaged to generate the F1 score per image ($BF1$), and the result in the full dataset is computed by averaging the $BF1$ of all images.

Multi-metric approach. In many segmentation tasks, class imbalance is common, where some classes have significantly more or fewer pixels than others. Using metrics such as $FWIoU$ and $mIoU$ that account for class frequencies helps mitigate the impact of imbalanced datasets and provides a fair evaluation of model performance across all classes. Moreover, for those specifically interested in models delivering precise delineation of class borders, metrics such as $BF1$ become crucial. Therefore, an optimal evaluation approach involves simultaneously considering multiple metrics. The integration of these metrics facilitates a comprehensive evaluation of image segmentation models, addressing various facets including accuracy, class-specific performance, class frequencies, border delineation, and overall segmentation quality. Essentially, a multi-metric approach enhances our understanding of the model's performance and provides guidance for further improvements.

5.5. Supervised Approaches

Supervised learning is a machine learning paradigm where an algorithm learns from labeled training data to make predictions. It involves learning a mapping from input data to output labels or values based on a set of input-output pairs. In the case of image segmentation, an input-output pair consists of a standard RGB image and a mask with the same resolution, where each pixel is labeled according to its ground truth. The same logic can be applied to the domain of seismic volumes, in which inline and crossline sections from the cube form the input images, and each one has a manually annotated mask delimiting the horizons (or classes). Despite the difficulty inherent in labeling such data, interpretations of public datasets have been made available in recent years, allowing researchers to tackle fully supervised methods. Amongst these interpretations, some of the most popular ones, as shown in Table 9, are those of Alaudah et al. (2019), Silva et al. (2019) for the F3 Netherlands and Chevron U.S.A. Inc. (Inc) for the Parihaka dataset.

Since CNNs have achieved state-of-the-art results in many computer vision tasks, a natural initial approach for the task of facies segmentation involves the use of such techniques. One of the earlier methods found in our investigation employed CNNs for automated seismic interpretation (Waldeland et al., 2018). More precisely, the authors proposed a binary classification method to identify salt bodies, where the network would use an encoder to learn a dense representation and then perform a pixel-wise classification as salt or non-salt. Although this approach achieved good results, they highlighted that CNNs require large amounts of training data and must be carefully designed to perform well. Regardless, their work showed that deep learning could be successfully applied to seismic data, and thus it laid the foundation for future works to explore new architectures and improve the results.

Numerous well-established deep learning architectures, originally designed for standard supervised image classification tasks, have been employed for seismic segmentation, such as AlexNet (Krizhevsky et al., 2012b), VGG (Simonyan and Zisserman, 2014), DeconvNet (Noh et al., 2015), and ResNet (He et al., 2016), which have achieved excellent results (Zheng et al., 2019; Dramsch and Lüthje, 2018; Waldeland and Solberg, 2017; Sun et al., 2017). However, seismic facies are very different from traditional images, motivating many works to design architectures specifically for this domain. For instance, Chevitarese et al. (2018) proposed modifications of the network topology to reduce the number of parameters and operations while still improving the accuracy of test data. These modifications were based on different ideas from previous methods, which include small convolutional filters, similar to those on VGG, residual units, as in ResNet, as well as utilizing different receptive fields and regularization methods. After performing experiments with all of these variations, their best model was entitled Danet-3, and it would then achieve state-of-the-art performance in terms of published results for seismic facies classification on the Penobscot dataset.

Despite the great success of CNNs in image classification, it was not well established up to this point that end-to-end convolutional networks could perform well in semantic labeling tasks such as facies classification. One major problem in CNNs is that they have a trade-off between classification and localization accuracy, which was evidenced by Girshick et al. (2014). Their work highlighted the need for region-based approaches to improve both localization and classification accuracy. Furthermore, deeper networks with many convolution and pooling layers have proven to be the most successful models for image classification, but they have a high spatial invariance due to their large receptive field. In other words, the deeper we go into a network, the more we lose the location information of objects within the image, which is crucial for tasks that require high spatial detail such as seismic segmentation.

To overcome the aforementioned limitation, deconvolution networks have become an alternative over traditional

CNNs with fully-connected layers. As described in Section 5.2, these networks use a symmetric encoder-decoder architecture composed of convolution and pooling layers in the encoder, and deconvolution and unpooling layers in the decoder. Such architectures can achieve finer and more accurate results than those of a fully-connected network. As an example, Alaudah *et al.*, the same authors of a publicly available interpretation of the F3 Netherlands block, utilized two baseline deconvolution networks to evaluate the performance of such architectures on this dataset (Alaudah et al., 2019). By separately training the models with patches and sections of the cube, experiments showed that, indeed, these architectures could better incorporate spatial and contextual information, since section-based models obtained better results than patch-based ones. This is a consequence of the fact that patches are acquired at different depths in the data, and some classes typically exist in specific depths, while sections span across all depths.

As previously mentioned, a big limitation of deep learning and CNN techniques is the requirement for large, high-quality annotated datasets to perform well. Hence, coming up with a strategy to utilize the few annotated samples more efficiently became as important as developing robust networks for seismic segmentation. UNet (Ronneberger et al., 2015) was one of the first proposed encoder-decoder architectures to tackle this limitation. Originally targeting biomedical images, it is one of the most popular networks for the task of image segmentation in many fields (Falk et al., 2019; Zhang et al., 2018; Pan et al., 2020; Zhao et al., 2019). It relies on the strong use of data augmentation techniques such as elastic deformations to the training images, which is a very common variation in biological tissue and even in rock formations, thus creating a model that is invariant to such deformations. Moreover, UNet also addresses the problem of separating touching objects of the same class, e.g. by applying a weighted loss that assigns higher values to borders of such values. This technique, while originally envisioning cells that are very close to each other in medical images, is also useful when it comes to segmenting geological facies, since delimiting horizons with fine detail in seismic volumes remains a challenge even for experts (Zhao et al., 2015). Because of these innovations, a series of works have employed UNet for geological facies segmentation tasks with successful results, ranging from the detection of salt-bodies (Shi et al., 2019) to complex tasks such as identifying seismofacies (Wang et al., 2019; Li et al., 2020) and lithofacies (Wang and Chen, 2021; Trinidad et al., 2022).

After establishing UNet as a robust baseline architecture for image segmentation, subsequent efforts focused on enhancing its classification accuracy. One mechanism that grew in popularity in recent years and has significantly improved the performance of deep learning models is called *attention* (Niu et al., 2021). This mechanism was initially inspired by how humans pay attention to different aspects of information when processing data, and it allows networks to focus on relevant information adaptively. It works by assigning different attention scores to different parts of the input data, such as words from a sentence or regions of an image, and higher scores dictate the level of importance of each part according to the task that is being performed. Furthermore, another popular mechanism in deep learning is the use of dilated convolutions (Contreras et al., 2021). As briefly mentioned previously, the context information in networks such as UNet is acquired by the pooling layers to expand the receptive field, which leads to the global information and resolution being gradually lost through the layers of the network. To address this issue, dilated convolutions can expand the receptive field without losing global information, which is achieved by increasing the space between the kernel values to expand the receptive field without increasing too much computation and loss of context information. Both the attention mechanism and dilated convolutions are desirable when dealing with large, complex seismic data, and were successfully incorporated into UNet to achieve better performance (Wang et al., 2019; Li et al., 2020; Trinidad et al., 2022).

Our literature review revealed that recent works have predominantly focused on modifying well-known architectures rather than adhering to baseline models. Previously, we showed an example of how networks like UNet could be modified by introducing more recent and complex deep learning techniques such as attention and dilated convolutions, but other architectures, such as SegNet and DeepLabV3+, have been explored and modified as well. Some of the modifications on these networks include: (i) varying the number of convolutional layers (Chevitarese et al., 2018); (ii) experimenting with different input sizes (Zhang et al., 2021); (iii) applying dropout and regularization (Chevitarese et al., 2018; Mukhopadhyay and Mallick, 2019); and (iv) testing with different loss functions (Trinidad et al., 2022). The former modifications (i and ii) aim to enhance efficiency, as there is a trade-off between computational cost and accuracy when handling extensive seismic data volumes, as noted by Zhang et al. (2019). In contrast, the latter adjustments (iii and iv) primarily address overfitting concerns while also effectively capturing intricate details in the predicted segmentation.

Another strategy employed by the most recent works is to combine two or more networks in an ensemble method. As an example, Zhang et al. (2019) employed both SegNet and UNet to tackle a segmentation task, as well as a binary classification for each seismic facies. The final segmentation was derived by merging the predicted facies based on the results that achieved the highest accuracies. The work of Abid et al. (2022) also combined two baseline networks, DeepLabV3+ and SegNet-18, to create ensemble models. The predicted scores for each class are averaged, and the highest average probability is labeled for the class of each pixel. Both of these works take inspiration from a classic machine learning strategy, which consists of separately training multiple weak classifiers and combining the results to obtain a strong classifier. This strategy is often called ensemble learning (Dietterich, 2000), and it has shown to be a promising technique in deep learning as well, managing to improve on the classification accuracy of standard baseline models (Zhang et al., 2019; Abid et al., 2022).

It is important to note that most works using supervised methods rely on data augmentation to compensate for the lack of large, high-quality annotated data. This strategy, however, has its limitations, and thus, many works have turned their attention to unsupervised and semi-supervised solutions, as detailed in the following section.

5.6. Strategies for Data-Scarce Scenarios

The performance of supervised learning approaches is highly dependent on the amount and the quality of annotated data (Jing and Tian, 2019). Consequently, reducing the costs and time required for labeling large datasets is paramount, especially in highly specialized fields (Balestriero et al., 2023); unfortunately, this is not always achievable. Therefore, there is a growing interest in adopting deep learning models to scenarios where only few labeled data are available, relying on approaches such as few-shot, semi- and self-supervised learning (Sun et al., 2019; Su et al., 2020; Xian et al., 2020). In this section, we introduce several studies that offer solutions for seismic facies segmentation when either no annotated data or only a sparse amount is available.

Semi-supervised learning falls within the techniques that leverage labeled and unlabeled data for training. Typically, it involves a small quantity of labeled data paired with a substantial amount of unlabeled data. One prominent approach in this domain are the Generative Adversarial Networks, introduced by Goodfellow et al. (2016). These models have gained widespread acknowledgement for their versatility in generating realistic data. GANs differ from traditional neural networks because they comprise two competing entities: the generator and the discriminator. The generator's role is to create samples from an unknown target probability distribution using random noise in a low-dimensional space as input, while the discriminator functions as a judge, distinguishing generated samples from real ones.

In the context of seismic data analysis, a comparison of supervised and semi-supervised learning was conducted in a study by Liu et al. (2020a), who employed a supervised CNN and a semi-supervised GAN to tackle facies classification using seismic reflection data derived from a facies model. Two distinct deep learning frameworks were proposed: a conventional CNN that relies on a substantial quantity of labeled data and a semi-supervised GAN framework that only requires limited data expanded from well-log locations during the network training. The discriminator is structured as an eight-class classifier: seven facies classes plus a real or fake discriminant. During the inference phase, the generator is no longer used, and the discriminator produces the final predictions. In their comparison, the semi-supervised method surpassed the supervised approach in the scarce-label scenario for both synthetic and real field data.

To alleviate the costs associated with data annotation, one strategy is to choose a handful of examples (crosslines and/or inlines) for annotation and then utilize these annotated samples to train a model for segmenting the remainder of the seismic volume. In essence, this entails sampling representative and distinctive examples that encapsulate the overarching patterns within the seismic volume while excluding redundant and anomalous instances. However, this process is not trivial, as geological layers often display horizontal continuity as a result of the natural sediment deposition process, which leads to an abundance of redundant information in adjacent sections. Therefore, the distance between training and testing sections within the seismic volume has a significant impact, particularly in few-shot scenarios, as also observed by Su-Mei et al. (2022) and

Wang et al. (2023). With this in mind, when attempting segmentation with a limited number of labeled sections, it is essential to consider the sampling strategy to prevent data redundancy and enhance data diversity.

One straightforward approach was employed by Wang and Chen (2021), who trained a UNet (Ronneberger et al., 2015) using a reduced number of equally spaced samples from the F3 Dataset. Although their results were promising, it is worth noting that this method merged imbalanced classes in the dataset and was tested on a small testing set. A more ingenious strategy was introduced by Chen et al. (2022b), who presented a sampling method that leverages the Harris corner detector (Derpanis, 2004) to enhance the dataset with spatial features capable of more effectively discriminating seismic sections. With that, the method could select representative samples without any supervision. However, despite the notable achieved improvements compared to a baseline approach, their strategy still required the utilization of at least one-quarter of the entire seismic volume.

A similar approach to address the scarcity of labeled data involves propagating available labels while considering the lateral variation in seismic data. In their research, Su-Mei et al. (2022) introduced a method that relied on five seismic sections recognized as “well-labeled examples”. At the beginning of the training process, they assessed the similarity between the data in the selected sections and the remaining 3D volume using cosine similarity. They defined subsets of similar sub-datasets associated with each reference data by establishing a threshold. Within each subset, the labels were assigned following the reference labeled section, and the entire subset was employed for a supervised training phase. This complex approach aimed at simulating the process of iterative work done by interpreters.

Transfer learning is another approach to be considered in scenarios where data is limited. Transfer learning is a method that focuses on carrying knowledge from a source domain to a target domain (Zhuang et al., 2020). It aims at achieving better target learners without depending on a large amount of data in the target domain. The broader definition can be divided into three categories from a label-setting aspect: transductive, inductive, and unsupervised transfer learning (Zhuang et al., 2020; Pan and Yang, 2009). Domain adaptation is sometimes regarded as a synonym for transfer learning, but it can be seen as a specific case of transductive transfer learning (Pan and Yang, 2009). In this setting, the source and target tasks are alike, but the source and target domains are dissimilar, i.e., the data has the same feature space in both domains, but they have distinct probability distributions (Pan and Yang, 2009). Domain adaptation is also defined as the process of adapting source domains to a target domain, addressing the domain shift by bringing the distribution of both domains closer (Weiss et al., 2016).

In our literature review, we identified a few studies that employed transfer learning for seismic segmentation as a way to address the challenge of limited data availability. Wang et al. (2021) initially trained a UNet model for lithofacies segmentation using the Parihaka NZ dataset as the baseline model. Afterward, they harnessed the acquired model weights for fine-tuning on the F3 Netherlands dataset. Their study aimed to achieve optimal results with fewer training and validation examples during the fine-tuning phase while maximizing the utilization of frozen layers. Their method involved exploring various hyperparameters for fine-tuning in a setup where the training and validation sets comprised 10 and 100 labeled sections. Additionally, they investigated the sequence of unfreezing layers within the model when adapting it to the target domain. They identified two local optima for the hyperparameter: one involving minimal unfreezing of just the final layers and another where approximately 18 layers in the decoding section were unfrozen. These findings suggest that fine-tuning the model is most effective when only a few of the final layers are retrained.

Similarly, Nasim et al. (2022) employed deep domain adaptation, using the F3 dataset as the source domain and the Penobscot dataset as the target domain. Their primary focus was on classes with limited labeled data. They proposed a network called EarthAdaptNet, which combined elements of both UNet and DaNet, incorporating an ASPP module. They employed the correlation alignment (CORAL (Sun and Saenko, 2016)) method to facilitate unsupervised deep domain adaptation. Notably, the CORAL method minimizes domain shift by aligning the distributions of source and target domains without requiring any target labels. Their testing results revealed that the performance was quite limited without domain adaptation methods. However, a substantial enhancement in performance was observed following the application of their refinement method, surpassing traditional architectures like UNet, particularly for the smaller classes.

Using networks pre-trained within specific domains has become well-established but sometimes requires domain adaptation techniques. One approach to deal with the domain shift relies on pre-training models within the target domain without manually annotated labels, i.e., applying unsupervised or, more specifically, self-supervised learning (SSL) techniques (Jing et al., 2018; Chen et al., 2020b). Self-supervised learning lies within unsupervised learning techniques where networks are explicitly trained using automatically forged labels. In computer vision, SSL approaches have gained significant attention, often relying on pre-text tasks for pre-training within the target domain (Noroozi and Favaro, 2016; Pathak et al., 2016; Gidaris et al., 2018; He et al., 2019; Chen et al., 2020a; Chen and He, 2021; Monteiro et al., 2022).

Following this approach, Wang et al. (2023) employed a two-stage method, relying on input reconstruction pre-training. In the initial unsupervised stage, a UNet was pre-trained to reconstruct input data. Subsequently, the encoder weights were kept frozen in the supervised setting and a new decoder designed specifically for lithofacies segmentation was trained. To assess the effectiveness of their approach, the model performance was evaluated under few-shot scenarios and compared with a supervised setup in the same context. Additionally, they investigated the benefits of the semi-supervised approach by exploring feature maps after pre-training and estimating segmentation uncertainty using deep ensembles (Lakshminarayanan et al., 2017). Their findings demonstrated that not only did the semi-supervised training enhance performance with limited labels, but it also reduced uncertainty. Similarly, Monteiro et al. (2022) also embraced SSL techniques for pre-training, followed by fine-tuning for the target task. Their pre-training process relied on contextual-based pre-text tasks, including training models to recognize image rotation (Gidaris et al., 2018) and reconstructing jigsaw puzzles from shuffled tiles within the image (Noroozi and Favaro, 2016). Subsequently, the models were fine-tuned in few-shot scenarios and the results compared with training from scratch, demonstrating improved performance through SSL pre-training. Additionally, they explored the benefits of deep model ensembles, which further enhanced performance by combining activations from multiple models.

A recent self-supervised learning technique that has become popular for tasks where obtaining labeled data is expensive or impractical is *contrastive learning* (Chen et al., 2020a; He et al., 2019; Chen and He, 2021). Its primary objective is to learn meaningful representations of data by contrasting positive pairs (similar samples) and negative pairs (dissimilar samples) within the dataset. The training aims at bringing similar samples or classes closer together in the latent space while pushing apart dissimilar ones, thereby enhancing the model's understanding of underlying patterns. One major benefit of adopting these strategies is that the models can also benefit from large amounts of unlabeled data. Additionally, the pre-training stage can occur within the target domain, diminishing the data domain shift.

Li et al. (2022) conducted a study that combines contrastive learning with conventional supervised learning. They utilized a limited number of labeled samples, leveraging the supervised loss for segmentation. More specifically, they extracted features from regions characterized by high confidence for each class, considering these regions as positive pairs (exhibiting similar features) and negative pairs (displaying dissimilar features). The overall loss function consisted of the sum of the supervised and contrastive loss. This approach yielded compelling results, even when using only 1% of the available training data. In another study involving contrastive learning, Li et al. (2023) developed a framework for automatic seismic facies clustering, eliminating manual labeling. Their method is a one-stage, end-to-end process. Seismic cubes were used instead of seismic traces or their variants for constructing a training dataset, which likely enhanced lateral consistency and the stability of facies mapping. Additionally, they incorporated seismic attributes, a conventional segmentation method, as geological constraints into the network alongside the seismic data (Zhao et al., 2017). By replacing data augmentations with seismic attributes, the method enabled the contrastive learning framework to process both types of inputs. This allowed for the maximization of similarities between seismic and multi-attribute cubes at the same location while minimizing similarities between cubes from different positions.

5.7. Summary of the Results

In this section, we provide a synthesis of the current state of the art in the field of facies segmentation of seismic images. Table 10 provides a detailed breakdown of each reviewed study, including the analyzed dataset, the number of corresponding classes, the results achieved in the four most widely-used metrics, as well as the architecture and loss function employed in the segmentation models.

Table 10
Compilation of results, showing the dataset and interpretation, the commonly available metrics, the adopted loss and architectures.

Reference	Dataset	Cls	Architecture	Loss	PA	MCA	FWIoU	mIoU
Abid et al. (2022)	F3 (Alaudah)	6	DeepLabv3+, SegNet	CEL	98	97	NA	94
Li et al. (2022)	F3 (Alaudah)	6	DeepLabv3+	CON, CEL	98	95	95	91
Chen et al. (2022b)	F3 (Alaudah)	6	Hrnetv2-W32	CEL	93	87	77	88
Tolstaya and Egorov (2022)	F3 (Alaudah)	6	UNet, EfficientNet B1	CEL, Dice, TVL	94	95	91	85
Trinidad et al. (2022)	F3 (Alaudah)	6	ABUNet, ConvLSTM	Focal	94	85	88	NA
Alaudah et al. (2019)	F3 (Alaudah)	6	Encoder-decoder CNN	NA	90	82	83	NA
Guazzelli et al. (2020)	F3 (Alaudah)	6	CNN	NA	88	64	79	NA
Wang et al. (2021)	F3 (Alaudah)	6	UNet	Dice	82	NA	NA	NA
Wang and Chen (2021)	F3 (Silva)	8	UNet, BNN	CEL	97	NA	NA	94
Monteiro et al. (2022)	F3 (Silva)	10	ResNet50	CEL	NA	NA	NA	83
Wang et al. (2023)	F3 (Silva)	10	UNet	CEL, MSE	99	99	NA	NA
Zhang et al. (2021)	F3 (Conoco-Phillips)	9	3D CNN, SegNet, DeepLabv3+	SGD	NA	NA	NA	92
Zhang et al. (2019)	F3 (Conoco-Phillips)	9	SegNet, UNet-based	SGD	97	NA	NA	NA
Wang et al. (2019)	F3 (Private)	9	UNet-based	NA	NA	NA	NA	88
Liu et al. (2020a)	F3 (Private)	9	VGGNet, GAN	ADV	86	NA	NA	NA
Li et al. (2020)	F3 (Private)	4	ADDCNN	CEL	87	NA	NA	NA
Li et al. (2022)	Parihaka (Chevron)	6	DeepLabv3+	CON, CEL	97	92	94	88
Tolstaya and Egorov (2022)	Parihaka (Chevron)	6	UNet, EfficienteNet B1	CE, Dice, TVL	94	96	90	77
Monteiro et al. (2022)	Parihaka (Chevron)	6	ResNet50	CEL	NA	NA	NA	55
Su-Mei et al. (2022)	Parihaka (Chevron)	6	UNet	CEL, CS	95	NA	NA	NA
Nasim et al. (2022)	Penobscot (Baroni)	6	EAN	CORAL	85	78	77	62
Chevitarese et al. (2018)	Penobscot (Private)	7	Danet-FCN	NA	97	NA	NA	NA
Shi et al. (2019)	Synthetic (Salt/Not-salt)	2	UNet	CEL	96	NA	NA	NA
Mukhopadhyay and Mallick (2019)	TGS Salt (Salt/Sediment)	2	Bayesian SegNet	NA	91	NA	NA	NA

ABUnet: Atrous Bidirectional UNet; BNN: Bayesian neural network; CEL: Cross-entropy loss; CNN: Convolutional Network; CON: Contrastive loss; CS: Cosine similarity; EAN: EarthAdaptNet; FWIoU: Frequency weighted intersection over union; GAN: Generative Adversarial Network; LSTM: Long short-term memory; MCA: Mean class accuracy; mIoU: Mean Intersection over Union; MSE: Mean squared error; PA: Pixel accuracy; SGD: Stochastic Gradient Descent; TVL: Total Variation Loss.

First, as already mentioned in Section 5.4, recall that the most commonly reported metrics are PA, MCA, FWIoU, and mIoU. While PA is widely used in this context, it's worth noting that it's not the standard metric for segmentation tasks. Since segmentation tasks often involve class imbalance, relying solely on PA may introduce bias in performance assessment. Hence, an important direction for the field of seismic image segmentation is to embrace and incorporate evaluation strategies more commonly employed in general semantic segmentation research, such as the mIoU metric (Lateef and Ruichek, 2019), which was used in several studies (Abid et al., 2022; Li et al., 2022; Chen et al., 2022b; Tolstaya and Egorov, 2022; Wang and Chen, 2021; Monteiro et al., 2022; Zhang et al., 2021; Wang et al., 2019; Nasim et al., 2022).

The results reported in the reviewed works and displayed in Table 10 may seem to be enough to determine the optimal approach for each dataset. Unfortunately, this is not true, as the absence of a concrete benchmark leads to significant differences in the experimental setup of each study and make their results incomparable. For instance, consider the results reported for Wang and Chen (2021) and Wang et al. (2023). Although they employ the same dataset (F3 Netherlands interpreted by Silva et al. (2019)), a comparison between their results is infeasible due to differences in the way the data is used: While the former study merges classes and use only 30 sections for testing, the latter omits test set size and composition details.

Finally, considering the challenges posed by the limited availability of annotated data, some of the studies listed in Table 10 have proposed unsupervised or semi-supervised approaches for seismic segmentation (as indicated in Table 4). To maintain brevity and fairness, Table 10 showcases only the best results achieved by each study, which are typically achieved through fully supervised methods or with the maximum available training data. As such, it is essential to recognize that results reported for unsupervised or few-shot scenarios warrant a more in-depth analysis and should not be blindly compared to those obtained through fully supervised approaches.

6. Challenges and Opportunities

As discussed throughout this document, there is an increasing interest in employing deep learning techniques to address various seismic segmentation tasks. However, because there are still many difficulties for automating seismic segmentation, this field presents many opportunities for improvement and open challenges.

6.1. Challenges

Our systematic review and data analysis provided some insights into the difficulty of using consolidated methods of deep models when working with seismic data. Several challenges regarding the application of deep learning have been discussed throughout this document. Here, we debate and briefly synthesize them.

Poor and scarce labeled data. Working with seismic data involves several issues that influence deep learning algorithms. The complexity of geological features shown in seismic images is a significant problem. These structures are frequently subject to alternative interpretations, resulting in disparities among specialists. Unlike other domains of computer vision, there are few publicly available annotated datasets for seismic interpretation (Alaudah et al., 2019). Also, the confidentiality of seismic interpretations, driven by commercial and strategic considerations, exacerbates the problem.

Lack of standard protocols for performance comparison. Due to the scarcity of labeled data, training and assessing deep learning models for seismic data processing becomes problematic. While many studies may use the same dataset (See Table 9), authors frequently combine classes from the available datasets, leading to discrepancies in the segmentation method. This absence of consistency makes establishing a baseline for comparison difficult and limits reproducibility. Furthermore, the lack of a common benchmark divides results into a broad range of testing sets with varying sizes and assessment techniques, confounding performance comparisons across different methodologies even further.

6.2. Opportunities

With challenges come opportunities to improve the current state of research in the area of deep learning for seismic segmentation. First and foremost, there is a pressing need to establish a robust benchmarking methodology that can accommodate multiple datasets and their respective interpretations. Disregarding the notion of one interpretation being superior to another, building an effective procedure for assessing the quality of the proposed methods is essential. Among the available benchmarks, Alaudah et al. (2019) stands out as the most comprehensive and well-defined. An analog approach could (and should) be developed for the other available datasets (Silva et al., 2019; Baroni et al., 2019; Inc; F3).

A noteworthy observation is the increasing interest in addressing segmentation with limited labels, particularly in few-shot scenarios. Some attempts to improve performance with fewer data were described in Section 5 (Wang and Chen, 2021; Su-Mei et al., 2022; Li et al., 2022; Monteiro et al., 2022; Chen et al., 2022b). Approaches involving semi-supervised learning (Su-Mei et al., 2022), self-supervised learning (Li et al., 2022), transfer learning (Monteiro et al., 2022), domain adaptation (Nasim et al., 2022), and ensemble techniques (Abid et al., 2022) have shown promise in handling this challenge. These methodologies offer potential venues for further exploration in effectively tackling the issue of limited labeled data for seismic segmentation.

Considering the mentioned points, it is evident that there is room for improvement in the field of deep learning for seismic segmentation. Future research efforts should prioritize the development of a robust and reliable benchmark that can be employed consistently. Additionally, it is crucial to compare the accomplishments of various approaches when working with limited labeled data. This comparative analysis will help to bridge the gap between current methodologies and real-world scenarios, enabling their implementation in routine activities conducted by geoscientists.

7. Conclusion

This work presented a review of deep learning approaches designed for seismic data segmentation. It focused on facies delimitation but also addressed salt and channeled structure segmentation. Through a comprehensive and reproducible literature review, we identified and analyzed various approaches, architectures, and methodologies utilized in this field. Also, we critically assessed the merits and flaws of the existing studies. By analyzing the methodologies, experimental setups, and reported results, we identified areas of improvement and potential biases or limitations. This evaluation provide recommendations for future studies, including opportunities for refining existing approaches and addressing specific challenges in seismic data segmentation.

Regarding the challenges, the complexity of geological features in seismic images and the subjectivity of interpretations among specialists pose significant difficulties in establishing baselines and performance comparisons. In addition, the scarcity of publicly available annotated datasets further complicates the training and assessment of deep learning models.

Nevertheless, this work also identified opportunities for research in this field. One crucial and natural opportunity lies in establishing a robust benchmark methodology that accommodates multiple datasets and interpretations. Furthermore, exploring approaches such as few-shot learning, semi- and self-supervised learning, transfer learning, domain adaptation, and ensemble techniques can help address the challenge of limited labeled data in segmentation problems.

Acknowledgments

The authors would like to thank Petróleo Brasileiro S.A. for the technical and financial support through its cooperation agreement with the UFMG - Universidade Federal de Minas Gerais.

CRediT authorship contribution statement

Bruno A. A. Monteiro: Writing - Original Draft; writing review & editing; Investigation; Methodology; Formal analysis; Visualization; Validation; Project administration. **Leonardo M. S. Jorge:** Methodology (equal); writing - original draft (equal). **Rafael H. Vareto:** Methodology; writing - original draft; writing review & editing. **Bryan S. Oliveira:** Methodology (equal); writing - original draft (equal). **Luiz Alberto Lima:** Conceptualization (equal); resources (equal); validation (equal); writing review & editing (equal). **Alexei M. C. Machado:** Conceptualization (equal); project administration (equal); supervision (equal); validation (equal); writing review & editing (equal). **William Robson Schwartz:** Conceptualization (equal); project administration (equal); supervision (equal); validation (equal); writing review & editing (equal). **Pedro O. S. Vaz-de-Melo:** Conceptualization (equal); project administration (equal); supervision (equal); validation (equal); writing review & editing (equal).

References

Abid, B., Khan, B.M., Memon, R.A., 2022. Seismic facies segmentation using ensemble of convolutional neural networks. Wireless Communications and Mobile Computing 2022.

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2012. Slic superpixels compared to state-of-the-art superpixel methods. IEEE transactions on pattern analysis and machine intelligence 34, 2274–2282.

Alaudah, Y., Michałowicz, P., Alfarraj, M., AlRegib, G., 2019. A machine-learning benchmark for facies classification. Interpretation 7, SE175–SE187.

An, Y., Du, H., Ma, S., Niu, Y., Liu, D., Wang, J., Du, Y., Childs, C., Walsh, J., Dong, R., 2023. Current state and future directions for deep learning based automatic seismic fault interpretation: A systematic review. Earth-Science Reviews 243, 104509.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 39, 2481–2495.

Balestriero, R., Ibrahim, M., Sobal, V., Morcos, A., Shekhar, S., Goldstein, T., Bordes, F., Bardes, A., Mialon, G., Tian, Y., et al., 2023. A cookbook of self-supervised learning. arXiv preprint arXiv:2304.12210 .

Baroni, L., Silva, R.M., Ferreira, R.S., Civitarese, D., Szwarcman, D., Brazil, E.V., 2019. Penobscot dataset: Fostering machine learning development for seismic interpretation. arXiv:1903.12060.

Baroni, L., Silva, R.M., S. Ferreira, R., Chevitarese, D., Szwarcman, D., Vital Brazil, E., 2018. Penobscot interpretation dataset. Zenodo .

Bertasius, G., Shi, J., Torresani, L., 2015. Semantic segmentation with boundary neural fields, in: 2017 CVPR, IEEE Computer Society. pp. 3602–3610.

Bi, Z., Geng, Z., Gao, H., Wu, X., Li, H., 2020. 3d relative geologic time estimation with deep learning, in: SEG International Exposition and Annual Meeting, SEG. p. D021S008R007.

Bi, Z., Wu, X., Geng, Z., Li, H., 2021. Deep relative geologic time: A deep learning method for simultaneously interpreting 3-d seismic horizons and faults. Journal of Geophysical Research: Solid Earth 126, e2021JB021882.

Bjørlykke, K., 2015. Introduction to petroleum geology. Petroleum geoscience: From sedimentary environments to rock physics , 1–29.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2017a. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence 40, 834–848.

Chen, L.C., Papandreou, G., Schroff, F., Adam, H., 2017b. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587 .

Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European conference on computer vision (ECCV), pp. 801–818.

Chen, M., Wu, S., Bedle, H., Xie, P., Zhang, J., Wang, Y., 2022a. Modeling of subsurface sedimentary facies using self-attention generative adversarial networks (sagans). Journal of Petroleum Science and Engineering 214, 110470.

Chen, T., Kornblith, S., Norouzi, M., Hinton, G., 2020a. A simple framework for contrastive learning of visual representations, in: ICML, PMLR. pp. 1597–1607.

Chen, T., Kornblith, S., Swersky, K., Norouzi, M., Hinton, G., 2020b. Big self-supervised models are strong semi-supervised learners. arXiv:2006.10029.

Chen, X., He, K., 2021. Exploring simple siamese representation learning, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 15750–15758.

Chen, X., Liang, C., Huang, D., Real, E., Wang, K., Liu, Y., Pham, H., Dong, X., Luong, T., Hsieh, C.J., et al., 2023. Symbolic discovery of optimization algorithms. arXiv preprint arXiv:2302.06675 .

Chen, X., Zou, Q., Xu, X., Wang, N., 2022b. A stronger baseline for seismic facies classification with less data. IEEE TGRS 60, 1–10.

Chevitarese, D.S., Szwarcman, D., Brazil, E.V., Zadrozny, B., 2018. Efficient classification of seismic textures, in: 2018 International Joint Conference on Neural Networks (IJCNN), IEEE. pp. 1–8.

Chopra, S., Marfurt, K.J., 2005. Seismic attributes—a historical perspective. Geophysics 70, 3SO–28SO.

Chowdhary, K., Chowdhary, K., 2020. Natural language processing. Fundamentals of artificial intelligence , 603–649.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3d u-net: Learning dense volumetric segmentation from sparse annotation, in: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (Eds.), MICCAI 2016, Springer. Springer International Publishing, Cham. pp. 424–432.

Contreras, J., Ceberio, M., Kreinovich, V., 2021. Why dilated convolutional neural networks: A proof of their optimality. Entropy 23.

Covidence, . How to conduct a systematic review from beginning to end. https://www.covidence.org/blog/how-to-conduct-a-systematic-review-from-beginning-to-end/. Accessed: 2023-05-10.

Csurka, G., Larlus, D., Perronnin, F., Meylan, F., 2013. What is a good evaluation measure for semantic segmentation?., in: Bmvc, Bristol. pp. 10–5244.

Delfiner, P., Peyret, O., Serra, O., 1987. Automatic determination of lithology from well logs. SPE formation evaluation 2, 303–310.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database, in: 2009 IEEE conference on computer vision and pattern recognition, Ieee. pp. 248–255.

Deng, L., Hinton, G., Kingsbury, B., 2013. New types of deep neural network learning for speech recognition and related applications: An overview, in: 2013 IEEE international conference on acoustics, speech and signal processing, IEEE. pp. 8599–8603.

Derpanis, K.G., 2004. The harris corner detector. York University 2, 1–2.

Dietterich, T.G., 2000. Ensemble methods in machine learning, in: Multiple Classifier Systems, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 1–15.

Dos Santos, J.A., Gosselin, P.H., Philipp-Foliguet, S., Torres, R.d.S., Falao, A.X., 2012. Multiscale classification of remote sensing images. IEEE TGRS 50, 3764–3775.

Dramsch, J.S., Lüthje, M., 2018. Deep-learning seismic facies on state-of-the-art cnn architectures, in: 2018 SEG International Exposition and Annual Meeting, OnePetro. pp. SEG–2018.

Du, S., Du, S., Liu, B., Zhang, X., 2021. Incorporating deeplabv3+ and object-based image analysis for semantic segmentation of very high resolution remote sensing images. International Journal of Digital Earth 14, 357–378.

Dumay, J., Fournier, F., 1988. Multivariate statistical analyses applied to seismic facies recognition. Geophysics 53, 1151–1159.

Eli-Chukwu, N.C., 2019. Applications of artificial intelligence in agriculture: A review. Engineering, Technology & Applied Science Research 9, 4377–4383.

EnergyGlossary, 2023. Diagram of crosslines, inlines, and a time slice.

F3, M., . Data from the malenov machine learning seismic interpretation project from conocophillips norge. Accessed: 2023-05-19.

Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jaeckel, Z., Seiwald, K., Dovzhenko, O., Tietz, O., Dal Bosco, C., Walsh, S., Saltukoglu, D., Tay, T., Prinz, M., Palme, K., Simons, M., Ronneberger, O., 2019. U-net: deep learning for cell counting, detection, and morphometry. Nature Methods 16.

Farabet, C., Couprie, C., Najman, L., Lecun, Y., 2013. Learning hierarchical features for scene labeling. IEEE transactions on pattern analysis and machine intelligence 35, 1915–1929.

Felzenszwalb, P.F., Huttenlocher, D.P., 2004. Efficient graph-based image segmentation. International journal of computer vision 59, 167–181.

Gao, D., 2007. Application of three-dimensional seismic texture analysis with special reference to deep-marine facies discrimination and interpretation: Offshore angola, west africa. AAPG bulletin 91, 1665–1683.

Gidaris, S., Singh, P., Komodakis, N., 2018. Unsupervised representation learning by predicting image rotations. arXiv:1803.07728.

Girshick, R., 2015. Fast r-cnn, in: Proceedings of the IEEE international conference on computer vision, pp. 1440–1448.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. arXiv:1311.2524.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial networks. Advances in Neural Information Processing Systems 3.

Guazzelli, A.B., Roisenberg, M., Rodrigues, B.B., 2020. Efficient 3d semantic segmentation of seismic images using orthogonal planes 2d convolutional neural networks, in: 2020 International Joint Conference on Neural Networks (IJCNN), IEEE. pp. 1–8.

Harzing, A.W., . Publish or perish.

He, K., Fan, H., Wu, Y., Xie, S., Girshick, R., 2019. Momentum contrast for unsupervised visual representation learning. arXiv preprint arXiv:1911.05722 .

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn, in: Proceedings of the IEEE international conference on computer vision, pp. 2961–2969.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

He, Y., Carass, A., Liu, Y., Jedynak, B.M., Solomon, S.D., Saidha, S., Calabresi, P.A., Prince, J.L., 2021. Structured layer surface segmentation for retina oct using fully convolutional regression networks. Medical Image Analysis 68, 101856.

Herron, D.A., 2011. First steps in seismic interpretation. Society of Exploration Geophysicists.

Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N., et al., 2012a. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal processing magazine 29, 82–97.

Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R., 2012b. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580 .

Iglovikov, V., Shvets, A., 2018. Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation.

Inc, C.U., . 2020 seg annual meeting machine learning interpretation workshop. Accessed: 2023-05-19.

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., Wang, Y., 2017. Artificial intelligence in healthcare: past, present and future. Stroke and vascular neurology 2.

Jing, L., Tian, Y., 2019. Self-supervised visual feature learning with deep neural networks: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 43.

Jing, L., Yang, X., Liu, J., Tian, Y., 2018. Self-supervised spatiotemporal feature learning via video rotation prediction. arXiv preprint arXiv:1811.11387 .

Kaggle, . Tgs salt identification challenge. https://www.kaggle.com/competitions/tgs-salt-identification-challenge/data. Accessed: 2023-07-28.

Kandel, M., He, Y., Lee, Y.J., Chen, T., Sullivan, K., Aydın, O., Saif, M., Kong, H., Sobh, N., Popescu, G., 2020. Phase imaging with computational specificity (pics) for measuring dry mass changes in sub-cellular compartments. Nature Communications 11.

Khan, K.S., Kunz, R., Kleijnen, J., Antes, G., 2003. Five steps to conducting a systematic review. Journal of the royal society of medicine 96, 118–121.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al., 2023. Segment anything. arXiv preprint arXiv:2304.02643 .

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012a. Imagenet classification with deep convolutional neural networks, in: Pereira, F., Burges, C., Bottou, L., Weinberger, K. (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc.. pp. 84–90.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012b. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems 25.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. Imagenet classification with deep convolutional neural networks. Communications of the ACM 60, 84–90.

Lakshminarayanan, B., Pritzel, A., Blundell, C., 2017. Simple and scalable predictive uncertainty estimation using deep ensembles. Advances in neural information processing systems 30.

Lateef, F., Ruichek, Y., 2019. Survey on semantic segmentation using deep learning techniques. Neurocomputing 338, 321–348.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–44.

LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D., 1989. Backpropagation applied to handwritten zip code recognition. Neural computation 1, 541–551.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE 86, 2278–2324.

Lee, J., Mukerji, T., . The stanford vi-e reservoir: A synthetic data set for joint seismic-em time-lapse monitoring algorithms, in: 25th Annual Report, Stanford Center for Reservoir Forecasting, pp. 1–53.

Li, B., Liu, S., Xu, W., Qiu, W., 2018. Real-time object detection and semantic segmentation for autonomous driving, in: MIPPR 2017: Automatic Target Recognition and Navigation, SPIE. pp. 167–174.

Li, F., Zhou, H., Wang, Z., Wu, X., 2020. Addcnn: An attention-based deep dilated convolutional neural network for seismic facies analysis with interpretable spatial–spectral maps. IEEE TGRS 59, 1733–1744.

Li, J., Wu, X., Ye, Y., Yang, C., Hu, Z., Sun, X., Zhao, T., 2023. Unsupervised contrastive learning for seismic facies characterization. Geophysics 88, WA81–WA89.

Li, K., Liu, W., Dou, Y., Xu, Z., Duan, H., Jing, R., 2022. Contrastive learning approach for semi-supervised seismic facies identification using high-confidence representations. arXiv preprint arXiv:2210.04776 .

Li, W., Chen, H., Shi, Z., 2021. Semantic segmentation of remote sensing images with self-supervised multitask representation learning. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14, 6438–6450.

Liang-Chieh, C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A., 2015. Semantic image segmentation with deep convolutional nets and fully connected crfs, in: International Conference on Learning Representations, pp. arXiv–1412.

Liu, M., Jervis, M., Li, W., Nivlet, P., 2020a. Seismic facies classification using supervised convolutional neural networks and semisupervised generative adversarial networks. Geophysics 85, O47–O58.

Liu, X., Chen, X., Li, J., Zhou, X., Chen, Y., 2020b. Facies identification based on multikernel relevance vector machine. IEEE TGRS 58, 7269–7282.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431–3440.

Ma, Z., Mei, G., 2021. Deep learning for geological hazards analysis: Data, models, applications, and opportunities. Earth-Science Reviews 223, 103858.

Martin, D.R., Fowlkes, C.C., Malik, J., 2004. Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE transactions on pattern analysis and machine intelligence 26, 530–549.

de Matos, M.C., Osorio, P.L., Johann, P.R., 2007. Unsupervised seismic facies analysis using wavelet transform and self-organizing maps. Geophysics 72, P9–P21.

Middleton, G.V., Hampton, M.A., 1973. Part i. sediment gravity flows: mechanics of flow and deposition. Soc. Econ. Paleont. Miner. Short course .

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., Terzopoulos, D., 2022. Image segmentation using deep learning: A survey. IEEE transactions on pattern analysis and machine intelligence 44, 3523–3542.

Mondol, N.H., 2010. Seismic Exploration. Springer Berlin Heidelberg, Berlin, Heidelberg. chapter 17. pp. 375–402.

Monteiro, B.A.A., Oliveira, H., Santos, J.A.d., 2022. Self-supervised learning for seismic image segmentation from few-labeled samples. IEEE GRSL 19, 1–5.

Mukhopadhyay, P., Mallick, S., 2019. Bayesian deep learning for seismic facies classification and its uncertainty estimation, in: SEG Technical Program Expanded Abstracts 2019. Society of Exploration Geophysicists, pp. 2488–2492.

Najman, L., Schmitt, M., 1994. Watershed of a continuous function. Signal Processing 38, 99–112.

Nanda, N.C., 2021. Seismic data interpretation and evaluation for hydrocarbon exploration and production. Springer.

Nasim, M.Q., Maiti, T., Srivastava, A., Singh, T., Mei, J., 2022. Seismic facies analysis: a deep domain adaptation approach. IEEE TGRS 60, 1–16.

Nazem, F., Ghasemi, F., Fassihi, A., Dehnavi, A., 2021. 3d u-net: A voxel-based method in binding site prediction of protein structure. Journal of Bioinformatics and Computational Biology 19, 2150006.

Neubert, P., Protzel, P., 2014. Compact watershed and preemptive slic: On improving trade-offs of superpixel segmentation algorithms, in: 2014 22nd international conference on pattern recognition, IEEE. pp. 996–1001.

Nichols, G., 2009. Sedimentology and stratigraphy. John Wiley & Sons.

Niu, Z., Zhong, G., Yu, H., 2021. A review on the attention mechanism of deep learning. Neurocomputing 452, 48–62.

Nock, R., Nielsen, F., 2004. Statistical region merging. IEEE Transactions on pattern analysis and machine intelligence 26, 1452–1458.

Noh, H., Hong, S., Han, B., 2015. Learning deconvolution network for semantic segmentation, in: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 1520–1528.

Noroozi, M., Favaro, P., 2016. Unsupervised learning of visual representations by solving jigsaw puzzles. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 9910 LNCS, 69–84.

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics 9, 62–66.

Ouma, Y.O., Josaphat, S., Tateishi, R., 2008. Multiscale remote sensing data segmentation and post-segmentation change detection based on logical modeling: Theoretical exposition and experimental results for forestland cover change analysis. Computers & geosciences 34, 715–737.

Ouzzani, M., Hammady, H., Fedorowicz, Z., Elmagarmid, A., 2016. Rayyan — a web and mobile app for systematic reviews.

Page, M.J.e., 2021. The prisma 2020 statement: an updated guideline for reporting systematic reviews. BMJ .

Pan, S.J., Yang, Q., 2009. A survey on transfer learning. IEEE Transactions on knowledge and data engineering 22, 1345–1359.

Pan, Z., Xu, J., Guo, Y., Hu, Y., Wang, G., 2020. Deep learning segmentation and classification for urban village using a worldview satellite image based on u-net. Remote Sensing 12.

Parker, S.P., 1984. McGraw-Hill concise encyclopedia of science & technology. McGraw-Hill.

Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., Efros, A.A., 2016. Context encoders: Feature learning by inpainting. arXiv:1604.07379.

Polat, H., 2022. A modified deeplabv3+ based semantic segmentation of chest computed tomography images for covid-19 lung infections. International Journal of Imaging Systems and Technology 32, 1481–1495.

Qian, F., Yin, M., Liu, X.Y., Wang, Y.J., Lu, C., Hu, G.M., 2018. Unsupervised seismic facies analysis via deep convolutional autoencoders. Geophysics 83, A39–A43.

Reading, B., 1978. Sedimentary environments and facies. Progress in Physical Geography 4, 624–625.

Roksandić, M., 1978. Seismic facies analysis concepts. Geophysical prospecting 26, 383–398.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: MICCAI 2015, Springer. pp. 234–241.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. nature 323, 533–536.

Salles Civitarese, D., Szwarcman, D., Silva, R., Vital Brazil, E., 2018. Seismic facies segmentation using deep learning, in: Online Journal for E&P Geoscientists, pp. 1–16.

Sediqi, K.M., Lee, H.J., 2021. A novel upsampling and context convolution for image semantic segmentation. Sensors 21, 2170.

Shen, K., Quan, H., Han, J., Wu, M., 2022. Uro-gan: An untrustworthy region optimization approach for adipose tissue segmentation based on adversarial learning. Applied Intelligence 52, 10247–10269.

Shi, Y., Wu, X., Fomel, S., 2019. Saltseg: Automatic 3d salt segmentation using a deep convolutional neural network. Interpretation 7, SE113–SE122.

Silva, R.M., Baroni, L., Ferreira, R.S., Civitarese, D., Szwarcman, D., Brazil, E.V., 2019. Netherlands dataset: A new public dataset for machine learning in seismic interpretation.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 arXiv:1409.1556.

Song, C., Liu, Z., Wang, Y., Li, X., Hu, G., 2017. Multi-waveform classification for seismic facies analysis. Computers & Geosciences 101, 1–9.

Su, J.C., Maji, S., Hariharan, B., 2020. When Does Self-supervision Improve Few-Shot Learning? Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 12352 LNCS, 645–666.

Su-Mei, H., Zhao-Hui, S., Meng-Ke, Z., San-Yi, Y., Shang-Xu, W., 2022. Incremental semi-supervised learning for intelligent seismic facies identification. Applied Geophysics 19, 41–52.

Sun, B., Saenko, K., 2016. Deep coral: Correlation alignment for deep domain adaptation, in: Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14, Springer. pp. 443–450.

Sun, G., Hao, Y., Rong, J., Shi, S., Ren, J., 2017. Combined deep learning and multiscale segmentation for rapid high resolution damage mapping, in: 2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), pp. 1101–1105.

Sun, Q., Liu, Y., Chua, T.S., Schiele, B., 2019. Meta-transfer learning for few-shot learning, in: CVPR, pp. 403–412.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9.

Takikawa, T., Acuna, D., Jampani, V., Fidler, S., 2019. Gated-scnn: Gated shape cnns for semantic segmentation, in: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 5228–5237.

Tolstaya, E., Egorov, A., 2022. Deep learning for automated seismic facies classification. Interpretation 10, SC31–SC40.

Tompson, J., Jain, A., Lecun, Y., Bregler, C., 2014. Joint training of a convolutional network and a graphical model for human pose estimation, in: Advances in Neural Information Processing Systems, Curran Associates, Inc.. pp. 1–9.

Trinidad, M.J.C., Canchumuni, S.W.A., Feitosa, R.Q., et al., 2022. Seismic facies segmentation using atrous convolutional-lstm network, in: ABMEC-IACM, pp. 1–7.

Tzotsos, A., Karantzalos, K., Argialas, D., 2011. Object-based image analysis through nonlinear scale-space filtering. ISPRS Journal of Photogrammetry and Remote Sensing 66, 2–16.

Vail, P.R., 1987. Seismic stratigraphy interpretation using sequence stratigraphy: Part 1: Seismic stratigraphy interpretation procedure. Atlas of Seismic Stratigraphy .

Vareto, R.H., Günther, M., Schwartz, W.R., 2023. Open-set face recognition with neural ensemble, maximal entropy loss and feature augmentation, in: 2023 36th Conference on Graphics, Patterns and Images (SIBGRAPI), IEEE. pp. 1–6.

Waldeland, A., Solberg, A., 2017. Salt classification using deep learning, in: 79th eage conference and exhibition 2017, EAGE Publications BV. pp. 1–5.

Waldeland, A.U., Jensen, A.C., Gelius, L.J., Solberg, A.H.S., 2018. Convolutional neural networks for automated seismic interpretation. The Leading Edge 37, 529–537.

Wang, C., Zhang, Y., Cui, M., Ren, P., Yang, Y., Xie, X., Hua, X., Bao, H., Xu, W., 2022. Active boundary loss for semantic segmentation. arXiv:2102.02696.

Wang, D., Chen, G., 2021. Seismic stratum segmentation using an encoder–decoder convolutional neural network. Mathematical Geosciences 53, 1355–1374.

Wang, L., Joncour, F., Barrallon, P.E., Harribey, T., Castanie, L., Yousfi, S., Guillion, S., 2023. Semi-supervised semantic segmentation for seismic interpretation. Geophysics 88, 1–57.

Wang, R., Stitt, J., Shugar, A., 2021. Identifying Geologic Facies Through Seismic Dataset-to-Dataset Transfer Learning using Convolutional Neural Networks. Technical Report. Stanford University.

Wang, Z., Di, H., Shafiq, M.A., Alaudah, Y., AlRegib, G., 2018. Successful leveraging of image processing and machine learning in seismic structural interpretation: A review. The Leading Edge 37, 451–461.

Wang, Z., Li, F., Taha, T.R., Arabnia, H.R., 2019. Improved automating seismic facies analysis using deep dilated attention autoencoders, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 0–0.

Weiss, K., Khoshgoftaar, T.M., Wang, D., 2016. A survey of transfer learning. Journal of Big data 3, 1–40.

West, B.P., May, S.R., Eastwood, J.E., Rossen, C., 2002. Interactive seismic facies classification using textural attributes and neural networks. The Leading Edge 21, 1042–1049.

Wolf, M., Pelissier-Combescure, J., 1982. Faciolog-automatic electrofacies determination, in: SPWLA 23rd Annual Logging Symposium, OnePetro. pp. SPWLA–1982.

Wrona, T., Pan, I., Bell, R.E., Gawthorpe, R.L., Fossen, H., Brune, S., 2021. 3d seismic interpretation with deep learning: A brief introduction. The Leading Edge 40, 524–532.

Xian, Y., Lampert, C.H., Schiele, B., Akata, Z., 2020. Zero-shot learning – a comprehensive evaluation of the good, the bad and the ugly. arXiv:1707.00600.

Yao, W., Zeng, Z., Lian, C., Tang, H., 2018. Pixel-wise regression using u-net and its application on pansharpening. Neurocomputing 312.

Yuan, Y., Xie, J., Chen, X., Wang, J., 2020. Segfix: Model-agnostic boundary refinement for segmentation.

Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks, in: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), Computer Vision – ECCV 2014, Springer International Publishing, Cham. pp. 818–833.

Zhang, H., Chen, T., Liu, Y., Zhang, Y., Liu, J., 2021. Automatic seismic facies interpretation using supervised deep learning. Geophysics 86, IM15–IM33.

Zhang, Y., Liu, Y., Zhang, H., Xue, H., 2019. Seismic facies analysis based on deep learning. IEEE GRSL 17, 1119–1123.

Zhang, Z., Liu, Q., Wang, Y., 2018. Road extraction by deep residual u-net. IEEE GRSL 15, 749–753.

Zhao, T., . Seismic facies classification using different deep convolutional neural networks, in: 2018 SEG International Exposition and Annual Meeting, pp. 1–5.

Zhao, T., Jayaram, V., Roy, A., Marfurt, K.J., 2015. A comparison of classification techniques for seismic facies recognition. Interpretation 3, SAE29–SAE58.

Zhao, T., Li, F., Marfurt, K.J., 2017. Constraining self-organizing map facies analysis with stratigraphy: An approach to increase the credibility in automatic seismic facies classification. Interpretation 5, T163–T171.

Zhao, X., Yuan, Y., Song, M., Ding, Y., Lin, F., Liang, D., Zhang, D., 2019. Use of unmanned aerial vehicle imagery and deep learning unet to extract rice lodging. Sensors 19.

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

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., He, Q., 2020. A comprehensive survey on transfer learning. Proceedings of the IEEE 109, 43–76.