Data Mining and Knowledge Discovery (2025) 39:0
https://doi.org/10.1007/s10618-024-01079-y

A systematic review of deep learning for structural geological interpretation

Gustavo Lúcius Fernandes · Flavio Figueiredo · Raphael Siston Hatushika · Maria Luiza Leão · Breno Augusto Mariano · Bruno Augusto Alemão Monteiro, et al. ***[full author details at the end of the article]***

Received: 6 March 2024 / Accepted: 6 November 2024

Abstract
It is well known that seismic data (or seismic volumes/images) are one of the primary work materials of the oil and gas industry. Nevertheless, the manual interpretation of such data has become increasingly time-consuming and prone to errors. This problem arises due to the massive amount of data and the complexity and variability of geological patterns. Over the last few years, image processing methods from the Artificial Intelligence (AI) sub-field Deep Learning (DL) have become valuable tools for seismic volume interpretation. DL techniques are already incorporated in commercial seismic interpretation software such as GeoTeric and Earth Science Analytics. Overall, DL accelerates the seismic interpretation process and enhances the results' precision and consistency. This allows for a better understanding of geological structures and reduces the risks associated with exploring and producing natural resources. Despite these facts, a systematic and broad review of DL techniques for different structural geological interpretation tasks has not been made available to researchers and practitioners. This article comprehensively reviews current techniques, models, and practices in DL-based seismic volume interpretation based on three different structural interpretation tasks: Fault interpretation, Horizon estimation, and Relative Geological Time (RGT) estimation, three complementary pillars of the geological framework. Our review is based on 85 relevant research articles, provides a foundational view of DL concepts, and introduces taxonomies for applying DL to structural geological interpretation. Our review also paves the way for further research by presenting the challenges encountered when applying deep learning in geological tasks and additional directions for future endeavors. The major applications of the techniques presented here are identifying architectural depositional elements and petrophysical properties of seismic images for natural oil and gas exploration.

Keywords Deep learning · Machine learning · Fault detection · Horizon estimation · Relative geological time

Responsible editor: Jianlin Cheng.

Published online: 25 November 2024

1 Introduction

Seismic interpretation plays a crucial role for a wide range of geological and geophysical applications. Some examples are:stratigraphic analyses (Mitchum Jr et al. 1977a, b; Vail et al. 1977), geomorphologic analyses (Posamentier 2000, 2004), structural analyses (Misra and Mukherjee 2018), rock physical properties (Havelia et al. 2021; Alabbad et al. 2021, 2023), bathymetry and seafloor characterization (Mosher et al. 2006; Power and Clarke 2019; Riera et al. 2019; Cox et al. 2020, 2021; Lebrec et al. 2022), mineral exploration and ore evaluation (Malehmir et al. 2012; Manzi et al. 2012), and, for hydrocarbon exploration (Payton 1977).

For the above tasks, seismic data is the critical component of interest, and seismic reflection is one of the most used techniques for imaging the subsurface (Ashcroft 2011; Sheriff and Geldart 1995) and acquiring such data. IN layman's terms, seismic reflection involves the transmission of seismic waves into the subsurface and the recording of the reflections of these waves at the interface between rocks of different physical properties according to the acoustic impedance contrast and incidence angle. By analyzing the travel times and amplitudes of reflected waves, geoscientists can create detailed 2D/3D subsurface images that represent changes in the physical properties of rocks in the subsurface.

Since the introduction of seismic data, seismic interpretation has been traditionally conducted by manually identifying key seismic reflections, primarily in 2D seismic images. That is, over the years, the use of 2D data for interpreting depositional systems and predicting lithology became more widespread, especially after the publication of AAPG Memoir 26 in 1977 (Payton 1977). This marked a significant development in the field of seismic interpretation as a whole. With time, 3D seismic data became more readily available, interpreters began transitioning from 2D to 3D interpretation. The advantage of 3D data is its ability to provide a more comprehensive view of subsurface structures and improve the accuracy of interpretation. Regardless of the number of dimensions, manual interpretations is laborious, and geoscientists commonly rely on software tools to aid them on this task. Something we now detail.

To complement traditional manual reflection picking, new software tools enabled semi-automatic and full-volume interpretation of 3D seismic data. They offer advantages such as generating 3D models, attribute analysis, and easy access to various display options (Alcalde et al. 2019). These methods have improved the speed and accuracy of interpretation and made it possible to analyze complex seismic datasets more effectively. However, manual seismic interpretation (even if aided by software tools) remains susceptible to the influence of uncertainties, biases, and subjectivity (Alcalde and Bond 2022). To mitigate these issues, Deep Learning (DL) techniques have recently been applied to seismic interpretation. DL algorithms can automatically extract features, recognize patterns, and assist in horizon and fault detection. Overall, DL seismic interpretation offers several advantages. That is, with DL we are able to: (1) process large volumes of seismic data quickly and consistently and this efficiency is especially valuable

when working with extensive 3D seismic datasets; (2) analyze seismic data at high speed enabling rapid interpretation and decision-making; (3) not suffer variations in interpretation that can affect human interpreters; (4) recognizing subtle patterns and features in seismic data that may be challenging interpreters to detect; (5) reduce the subjective nature of interpretation; (6) lead to cost savings in exploration and production activities by streamlining the interpretation process, reducing the need for extensive manual labor, and enabling more efficient reservoir characterization.

Given the numerous benefits of employing DL in seismic interpretation, there has been a substantial increase in articles related to this task over recent years. In this survey, our goal is to provide an overview of such publications as they cover a wide range of topics, from identifying architectural elements in depositional environments, such as channels (Pham et al. 2019; Gao et al. 2020), salt bodies (Shi et al. 2019; Sen et al. 2020), and clinoforms (Gao et al. 2023), to addressing geophysical challenges like waveform inversion (Wu and McMechan 2019; He and Wang 2021), and noise attenuation (Yu et al. 2019). Nonetheless, the predominant publications have been on fault interpretation, horizon segmentation, and Relative Geological Time (RGT) estimation in seismic data. These tasks are the focus of our survey, and they are fundamental to constructing a comprehensive geological framework that includes structural and stratigraphic interpretation, something we now detail.

In the context of structural interpretation, the identification and analysis of faults play a central role, significantly influencing reservoir exploration and the strategic positioning of wells (Ottesen Ellevset et al. 1998; Rivenæs et al. 2005). The detection of faults is crucial for building reliable structural models (Caumon et al. 2009) and reducing risks associated with drilling and production. Furthermore, the process of fault detection is intricately linked to horizon interpretation. Accurately interpreting seismic horizons across fault zones often necessitates accurately identifying faults, as they serve as critical boundary controls. In the realm of stratigraphic interpretation, the interpretation of seismic horizons is critical. The horizon segmentation is fundamental for understanding the subsurface geology and identifying geological features, including stratigraphic sequences and depositional environments. Like the fault interpretation, it serves as the basis for subsequent interpretation tasks. Equally important for stratigraphic interpretation is using RGT volumes (Stark 2003, 2004, 2005). These volumes assign a relative geologic time value to each point within the seismic volume, thereby facilitating the extraction of stratigraphic information and providing a chronostratigraphic framework for understanding the temporal aspects of subsurface geology, making them essential for accurate stratigraphic interpretations.

Recognizing the critical role of fault interpretation, horizon segmentation, and RGT estimation in constructing an extensive geological framework, this survey offers an exhaustive review of the latest techniques, models, and practices in DL. It analyzes 85 pertinent articles focusing on these three tasks within seismic data interpretation. The selection of these tasks is informed not only by their integral importance in seismic interpretation–encompassing both structural and stratigraphic aspects, but also by their potential to cover a diverse range of methodologies from

the machine learning (ML) domain. This includes approaches in classification, regression, and segmentation.

Furthermore, our work complements other surveys, as we only identified literature reviews or systematic reviews related to fault detection among the three investigated tasks. An analysis conducted by Wang et al. (2018) addressed fault and salt dome detection, exploring the application of machine learning to enhance and automate these activities. Although a section of this review focuses on deep learning models, we observed that the number of presented works is limited, possibly due to the infrequent use of DL algorithms in these tasks at that time. Another study conducted (Tariq et al. 2021) offered a systematic review on the use of data science and machine learning in the oil and gas industry, dedicating a section to the application of machine learning in fault detection. More recently, An et al. (2023) presented a comprehensive systematic review on using deep learning for automated fault interpretation. Our research diverges from these works by concentrating on a broader spectrum of seismic interpretation methods, which are vital for developing a detailed geological framework model, including structural interpretation (i.e., fault detection) and stratigraphic interpretation (i.e., horizon extraction, RGT estimation). Furthermore, our systematic review brings to light newer methods not covered by An et al. (2023).

This work is organized as follows. The next section provides a review of the Traditional Interpretation Procedure (Sect. 2) and is followed by a Background Section (Sect. 3) covering essential explanations and definitions related to the mathematical notation used throughout this study. Additionally, we provide an overview of standard machine learning tasks, common deep learning architectures, computer vision, and the image processing techniques used in our systematic review papers. This section also delves into the concepts of evaluation metrics and loss functions used to train DL models. After presenting the required background, our systematic review process is described in Sect. 4. Subsequently, Sect. 5 presents a comprehensive literature review on deep learning applications in fault interpretation, horizon segmentation, and RGT analysis and summarizes the main datasets currently utilized in the literature. Finally, Sect. 6 presents current challenges and future opportunities for research, while Sect. 7 concludes our survey.

2 Traditional interpretation methods

In geophysical exploration methods, seismic surveying takes precedence due to its unparalleled capability to detect subsurface features, ranging from large-scale to small-scale (Bjorlykke 2015). Geoscientists extensively analyze seismic data to deduce information about subsurface rock formation composition, fluid content, extent, and geometry. This analysis hinges on integrating seismic inlines, crosslines, time slices, and horizon attributes (Hesthammer et al. 2001) and the interpretation process can be categorized into several types based on the geological features or information they aim to extract from seismic data, encompassing lithologic interpretation, stratigraphic interpretation, and structural interpretation.

The efficacy of these various categories of seismic interpretation can be significantly augmented through the utilization of seismic attributes. A seismic attribute is a measurement based on seismic data that helps us visually enhance features of interpretation interest (Herron 2011). An adequate seismic attribute can extract information from seismic data that would otherwise be lost and is either directly responsive to some desired geologic feature or property of interest (Anstey et al. 2007). One goal of seismic attributes is to capture the interpreter's pattern recognition expertise by quantifying the amplitude, continuity, and morphological features in the seismic data. For example, amplitude thresholding channels or hydrocarbon reservoirs and coherence attribute captures discontinuities, such as faults (Chen and Sidney 1997; Herron 2011). However, despite their improvements, these seismic attributes also have limitations due to the manual selection of attributes and threshold values, which can make the process subjective and potentially time-consuming (Wu et al. 2019). In the following subsection, we will discuss the traditional methodologies employed in seismic interpretation, along with the limitations of these methods for both structural and stratigraphic analyses. These conventional approaches cover manual interpretation techniques, the application of seismic attributes, and the utilization of non-deep learning algorithms. Our discussion will particularly emphasize the tasks of fault interpretation, horizon delineation, and RGT estimation, which constitute the primary objectives of our study.

2.1 Overview of strucutural interpretation

Structural seismic interpretation is the process of analyzing seismic data and characterize the structural features of the subsurface, including faults, folds, and other geological structures. This interpretation involves identifying and delineate geological structures, which helps geoscientists understand the deformation and structural history of an area.

2.1.1 Faults interpretation methods

Faults play a central role in structural seismic interpretation, enabling geoscientists to build accurate 3D structural models of the subsurface. These models are used in various applications: hydrocarbon reservoir characterization, faults can act as barriers or conduits for fluid flow within reservoirs, impacting production and reservoir management; mitigating hazards in drilling and production operations; tectonic analysis, to understand the geological history and tectonic processes that have shaped an area.

Traditional methods of interpreting faults from seismic data involve manual interpretation, which can be labor-intensive and time-consuming. These conventional methods rely on the interpretation of various geophysical surveys and the visual identification of seismic reflections and discontinuities associated with faulted geological structure. Therefore, this method involves dealing with multiple challenges, including ambiguity, data limitations, subjectivity, and the need for geological context (Frodeman 1995; Rankey and Mitchell 2003; Bond et al. 2007; Saltus and

Blakely 2011). Some studies have attempted to quantify these ambiguities, notable among them are the works of Bond et al. (2007), Bond (2015) and Macrae et al. (2016).

Before the DL 'boom', several methods were developed to enhance the detection and interpretation of faults, utilizing an array of additional attributes or mathematics operations. Semblance analysis involves computing the semblance attribute, which measures the coherency of seismic data along different directions. Faults often disrupt the coherency of seismic reflections, making them detectable through semblance analysis (Marfurt et al. 1998). Coherence is another attribute used to detect faults. It quantifies the similarity between seismic traces and is sensitive to discontinuities caused by faults (Bahorich and Farmer 1995; Marfurt et al. 1999; Wu 2017). Variance analysis focuses on the variance of seismic data, which can highlight variations in the seismic signal caused by faulting (Van Bemmel and Pepper 2000; Randen et al. 2001). Curvature attributes are calculated to identify changes in the curvature of seismic reflections. Faults often exhibit distinctive curvature patterns (Roberts 2001; Al-Dossary and Marfurt 2006; Di and Gao 2016). Edge detection techniques are used to identify abrupt changes in seismic signal intensity, which can be indicative of fault locations (Di and Gao 2014). The fault likelihood attribute, proposed by Hale (2013), quantifies the likelihood of faults within seismic images and can be used to extract fault surfaces.

2.2 Overview of stratigraphic interpretation

Stratigraphic seismic interpretation concerns understanding the arrangement and characteristics of sedimentary rock layers or strata over time. Its analysis involves deciphering the geological history recorded in seismic data by recognizing reflection patterns, identifying unconformities or discontinuities, and establishing a temporal framework for the subsurface strata. This understanding is essential for various geological applications, including hydrocarbon exploration and reservoir characterization.

2.2.1 Horizon interpretation methods

One of the critical elements for stratigraphic analysis is the horizon. Interpreted horizons are fundamental for building structural and stratigraphic models of the subsurface: (1) provide the basis for mapping subsurface geological structures, identifying potential reservoirs, and characterizing rock properties within those reservoirs; (2) define sequences and sequence boundaries, which are essential for understanding the depositional history and stratigraphic relationships within sedimentary basins; and (3) provide geologically reasonable control for extending lithology interpretations away from well data.

The traditional approach to horizon interpretation is based on identifying critical seismic reflections within the data. Their continuity and termination patterns typically characterize these reflections. Interpreters would manually pick stratigraphic surfaces or horizons of interest from seismic data. This process involves visually

identifying and tracing these horizons along seismic lines or within 3D seismic volumes, which can be labor-intensive and time-consuming.

The methods for interpreting seismic horizons have evolved significantly over the years, primarily driven by advancements in computer technology and the increasing workload of seismic data interpretation. These methods leverage various data attributes and computational techniques to automate horizon tracking and improve the accuracy and speed of seismic interpretation. Gersztenkorn and Marfurt (1999) introduced an eigenstructure-based coherence algorithm to automate horizon detection that detected discontinuities like horizons or faults in seismic data by comparing different coherence algorithms. Pauget et al. (2009) introduced a global geological method that auto-tracked horizons by considering links between seismic samples. It iteratively adjusted these links until reaching a global minimum, automating horizon interpretation. Another popular method for automated horizon interpretation is Slope-based. They involved estimating local slopes and multigrid slopes of seismic traces to identify horizons (Lomask et al. 2006; Fomel 2010; Wu and Hale 2013, 2015; Monniron et al. 2016). These methods leveraged gradient information in seismic data to detect horizons. Stark (2003) and Wu and Zhong (2012) introduced phase-unwrapping techniques to assist in seismic horizon interpretation. These methods focused on phase information in seismic data to identify horizons. Figueiredo et al. (2007), Figueiredo et al. (2014), Figueiredo et al. (2015) proposed waveform classification methods for horizon interpretation. They used waveform characteristics to automate the mapping of horizons. Wu and Fomel (2018) developed a horizon interpretation method based on multigrid correlation. This approach considered the correlation between seismic traces to identify horizons.

Interpreting the horizon is just the beginning of stratigraphic analysis; it's also essential to understand the geometric relationships between seismic reflections observed in the data. Seismic reflection terminations, which are the endpoints or boundaries of seismic reflections, play a crucial role in identifying seismic-scale stratigraphic discontinuities or unconformities. The architecture of seismic reflections can be described using terms such as concordant, onlapping, downlapping, toplapping, or erosionally truncating. These terminologies aid geoscientists in understanding how different strata relate to each other and erosional surfaces or unconformities.

2.2.2 RGT estimation methods

Another way to extract horizons and understand geologic structures is using the concept of the RGT volume first presented by Stark (2003), Stark (2004), Stark (2005). In this volume, each sample is assigned a relative geologic time value that corresponds to the reflection amplitude of that sample (Stark 2003, 2005). Unlike a seismic image, where sample values represent seismic amplitude, the RGT volume assigns values representing geologic time. These values typically increase with depth in the subsurface, and the iso-values extracted from an RGT volume correspond to seismic horizons. The RGT volume is the most comprehensive product for stratigraphic interpretation, as it provides a global view of all seismic horizons while simultaneously conveying the concept of geological time.

The RGT volume has proven valuable in various applications, such as to improve horizon extraction (Stark 2004; Fomel 2010; Wu and Zhong 2012), sedimentologic interpretation (Hongliu et al. 2012), stratigraphic interpretation (Pauget et al. 2009), geologic body detection (Prazuck et al. 2015), and structural implicit modeling (Irakarama et al. 2021; Laurent et al. 2016; Renaudeau et al. 2019).

Constructing the RGT volume can be challenging due to several difficulties, especially when using a simple manual or automatic tracking approach. The accuracy and reliability of the RGT volume depend heavily on the precision of horizon tracking. If horizons are not tracked accurately, it can lead to errors in assigning relative geologic time values to seismic samples (Wu and Zhong 2012). When using a tracking-based approach, the RGT volume may have limited resolution. This limitation arises because the RGT values are assigned based on discrete geologic time surfaces corresponding to tracked horizons. Information between these horizons is often ignored, resulting in a less detailed RGT volume. In practice, it may be challenging to track all seismic horizons comprehensively. Incomplete horizon tracking can leave gaps in the RGT volume, making it less useful for interpreting the subsurface geology (Wu and Zhong 2012).

Several non-Deep Learning methods have been proposed for estimating RGT or generating volumes of horizons. The stratal slicing method, for instance, employs reference horizons to interpolate a stratal time model or horizon volume (Zeng et al. 1998a, b). Other methods utilize seismic reflector dips or seismic normal vectors, which are computed for every image sample to be perpendicular to seismic amplitude reflectors (Lomask et al. 2006; Fomel 2010). The phase-unwrapping techniques can be applied to generate RGT volumes by unwrapping the seismic instantaneous phase. Additionally, some methods iteratively fit horizon slopes with seismic reflection slopes to automatically extract horizons and estimate RGT maps (Stark 2003, 2004, 2005; Wu and Zhong 2012).

Finally, seismic data interpretation plays an essential role in the geophysical investigation process involving seismic reflection. The manual interpretation of stratigraphic facies, faults, and horizons involves collaboration among geologists, geophysicists, and petrophysicists. However, given the substantial volume of seismic data, this process is highly time-consuming. Consequently, there is growing interest in leveraging machine learning algorithms to augment its interpretation. While experts offer valuable insights through this method, it is labor-intensive, prone to subjectivity, and limited in efficiently handling large datasets. Nonetheless, manual interpretation remains crucial for calibrating and validating automated methods.

3 Required background

In this section, we present some background in Image Processing and Machine Learning (ML) overall. The goal of this discussion is to present the reader with the required background on the major topics of ML required to understand the literature raised in our systematic review.

3.1 Image processing and computer vision

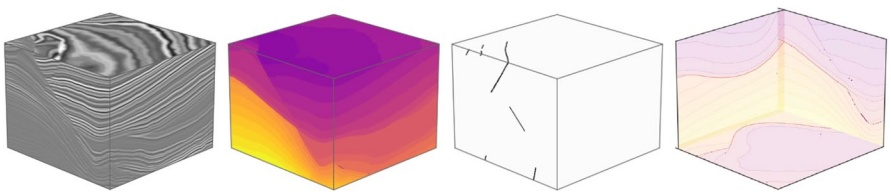
Computer Vision (CV) is an interdisciplinary domain that enables computers to extract high-level insights from images or frames. In contrast, image processing involves various techniques to process visual data (Gonzalez and Woods 2008). The data used is structured as matrices or tensors for both CV and image processing. A two-dimensional matrix represents a grayscale image with height and width coordinates, whereas a color image is a 3d Tensor. Thus, Black and White (BW) images become matrices, whereas color images become tensors with one matrix per red, green, and blue color (referred to as channels). Cells in such matrices/tensors are intensity values.

Seismic data are similar to BW images with a depth dimension. Thus, they are also tensors but with different intensity ranges. That is, these tensors may have values that do not represent colors. Authentic seismic images will not have intensity values in this [0, 255] range; however, a simple scaling represents the seismic data as an image. Figure 1a shows an example of a seismic image.

Over the years, it was possible to notice an overlap between Computer Vision, Image Processing (Gonzalez and Woods 2008), and Machine Learning. This combination contributed to the emergence of methods and algorithms that may be employed to understand and extract meaningful information from images. ML currently provides the most successful tools in image processing and is revised next.

3.2 Common machine learning tasks

ML may be summarized as a set of computational algorithms and statistical methods devised to train supervised (or predictive), clustering, or generative functions from finite datasets (Sarker 2021; Harshvardhan et al. 2020). The tasks analyzed in this work seek to train predictive functions. For descriptive purposes, let $\mathcal{D} = \{(x_i, y_i)\}$ be some finite dataset where (x_i, y_i) are tuples of data-points, with x_i coming from some domain $\mathcal{X}$ and, similarly, y_i coming from some other, but not necessarily different, domain $\mathcal{Y}$, i.e., $y_i \in \mathcal{Y}$. Here, i is used to capture the example, that is, y_0 is the response for example 0, and $i \in [0, n)$, where n is the number of examples ($n = |\mathcal{D}|$). The goal of supervised ML is to have the computer learn some parameterized function $h_\theta : \mathcal{X} \mapsto \mathcal{Y}$, with $\hat{y} = h_\theta(x)$ being a prediction. To understand our notation, one



(a) A Seismic Image is a 3d-Tensor (b) RGT Values for a Seismic Image (c) Fault Labels in a Seismic Volume (d) Horizons and RGT Overlays Per Slices

Fig. 1 Example of input data and different expected outputs

may consider a simple linear regression task. Here, an ML algorithm will train some model of the form: $\hat{y} = h_{\boldsymbol{\theta}}(x) = \boldsymbol{\theta}_0 + \boldsymbol{\theta}_1 x$. With a finite collection of data points $x_i \in \mathbb{R}$ and $y_i \in \mathbb{R}$ (where $\mathcal{X} = \mathcal{Y} = \mathbb{R}$). The goal of the ML algorithm is to find the set parameters $\boldsymbol{\theta} = [\theta_0, \theta_1]^T$ ($[\theta_0, \theta_1]$ is a row vector of two elements, and $[\theta_0, \theta_1]^T$ is it's transpose) that generates "good" predictions. In summary, the objective of an ML algorithm is to learn a function. In our notation, $\boldsymbol{\theta}$ is the column vector (we employ bold for vectors) of model parameters.

To find the best parameters, it is common for ML methods to employ an optimization approach where a suitable loss function is differentiated and minimized regarding $\boldsymbol{\theta}$. We deter our discussion loss functions to Sect. 3.6. For now, our goal is to summarize the main ML tasks (next) and how they relate to seismic interpretation. Before doing so, we further point out that these definitions will change slightly for other tasks, such as clustering and generative modeling.

A standard categorization of supervised machine learning tasks involves distinguishing between regression and classification tasks. The task is typically called regression when the target responses are real numbers. Conversely, classification tasks establish relationships between features and discrete outcome values. Another type of machine learning task is segmentation. The primary objective of a segmentation task is to partition an image into distinct segments based on predefined criteria. Finally, another necessary type of ML task is object detection. In summary, in this task, models need to identify instances of objects in images.

3.2.1 Regression task

As stated, when working with regression tasks in machine learning, the main focus is to investigate the relationship between features or variables, $\mathbf{x}_i \in \mathbb{R}^d$ (where $\mathbf{x}_i$ represents a vector of features and d is the number of features) and a continuous outcome value, $y_i \in \mathbb{R}$. For instance, if we wished to predict the rent value of an apartment given its neighborhood location, size in square meters, and the number of bedrooms (Maulud and Abdulazeez 2020).

One possible regression task related to seismic volume interpretation is the estimation of the RGT (Bi et al. 2021; Geng et al. 2020; Abubakar et al. 2022). As the name implies, we are interested in building a regression model that can predict the relative geological time of an input seismic volume. For this task, the outcome is not a single value, but a tensor or values indicating the relative time. Through a simple study of the output's domain, one can realize that the geological time of a seismic volume increases monotonically through the vertical axis, except in some areas of "discontinuity" caused by faults, for example. By doing so, one can use certain types of models (classical machine learning or deep learning based), if desired, that output a monotonically increasing prediction (Wehenkel and Louppe 2019; Runje and Shankaranarayana 2022). Figure 1b shows a simple example of a seismic volume RGT prediction inline slice. We can see that the values along the vertical axis monotonically increase as time increases.

3.2.2 Classification task

Classification tasks can be categorized into two main types: binary classification and multiclass classification. In binary classification, each example $\mathbf{x}_i$ is assigned a label y_i, where y_i belongs to the set $\{0, 1\}$. A $y_i = 1$ value indicates that the example $\mathbf{x}_i$ belongs to a specific class, while $y_i = 0$ signifies that the example does not belong to the class. In multiclass classification, the label vector $\mathbf{y}_i$ is defined as $\mathbf{y}_i \in \{0, 1\}^k$, where k represents the number of classes and the vector size. It is common for classes to be mutually exclusive, meaning an example belongs to only one class. In other words, the sum of elements in $\mathbf{y}_i$ equals 1, indicating the presence of a single class label. It's worth noting that while exclusivity of classes is a typical scenario, there are settings where examples may have multiple labels, deviating from this rule.

During the process of seismic volume interpretation, one common task done by interpreters is related to fault inference, which can be translated to a classification task (Wei et al. 2022; Wu et al. 2021; Yu et al. 2022). One may be interested in knowing if a specific seismic image contains or not one or more faults, as shown in Fig. 1c (Yu et al. 2022; Egorov 2019; An et al. 2021). This scenario is a binary classification task. Another scenario involves determining the type of each fault in an image (Wu et al. 2019). This case is a multiclass classification. The latter also permeates tasks related to segmentation problems, which will be discussed in the next, where we also want to detect the location of a particular fault in the image or seismic volume.

3.2.3 Segmentation and object detection tasks

Beyond classifying what an image represents as a whole, there is a need to perform pixel-level classification, which corresponds to a specific type of task: image segmentation. Image segmentation is crucial in numerous computer vision applications, images, or videos.

Image segmentation tasks can be broadly classified into two categories: (i) semantic segmentation, where machine learning models are tasked with categorizing each pixel based on the specific type of object it represents, essentially discerning its semantics (e.g., identifying cars or dogs); (ii) instance segmentation, which involves machine learning models classifying pixels individually within each instance in the image, without considering their overall class membership (Hafiz and Bhat 2020).

We can draw a parallel between segmentation-based tasks and seismic volume interpretation when dealing with seismic horizon inference (Dodda et al. 2022; Dyer and Manral 2022; Guo et al. 2020; Gupta et al. 2019). As the name implies, we are interested in building a segmentation model that estimates where are the horizons in a given seismic volume or a seismic image. Figure 1d illustrates an example of an instance segmentation task applied to horizon interpretation. This figure also shows the RGT values overlayed with the horizons. Different from the other figures, we focus on one slice for each coordinate. Notice how RGT values highlight horizons, indicating that one task is not unrelated to the other.

The purpose of the object detection task is to identify and localize specific instances of objects from one or many classes, such as people, animals, or geological faults in an image (Ikeuchi 2021). Thus, object detection aims to draw bounding boxes around each instance of the object of interest and classify them into predefined categories or classes. The object detection task provides information about the presence and location of objects in an image.

Image segmentation is frequently mistaken for object detection tasks. However, while object detection provides a coarser analysis of an image by detecting only objects in a particular scene, image segmentation provides a fine-grained understanding of an image, classifying each instance or semantic presented in the input (Hafiz and Bhat 2020).

3.3 Unsupervised and semi-supervised machine learning

So far, the tasks we have mentioned fall into *supervised* machine learning. That is, to learn models, we employ both features ($\mathbf{x}_i$) and responses (y_i). However, many ML tasks exist to extract patterns from datasets with no labels. These are unsupervised learning methods, and data clustering is the most common class of such methods.

Similarly, semi-supervised learning aims to extract patterns from datasets with *few* labels. In other words, datasets where we have only a handful of labeled examples and a wide range of unlabeled examples. These commonly exploit the distance between supervised and unsupervised examples to extract patterns.

Given that most of the papers that comprise our review focus on unsupervised ML tasks, we presented more details on such tasks. However, some papers do explore supervised and semi-supervised methods.

3.4 Common Deep Learning architectures

Deep Learning (DL), a subset of Machine Learning, comprises models constructed with multiple processing layers, which aim to learn the representation of a data set (LeCun et al. 2015). With the increase in computational resources and the amount of data available, DL models tend to perform better when compared to conventional ML (Sarker 2021). Next, we will present the most common architectures of models that use DL.

3.4.1 Multilayer perceptron

The Perceptron (McCulloch and Pitts 1943) is a simple structure based on a single artificial neuron that receives multiple weighted inputs and produces a binary output. The motivation behind the creation of the Perceptron is related to the idea of developing a computational model inspired by the functioning of the human brain and capable of performing classification tasks. After training, the Perceptron represents the hyperplane that splits the data respecting the labels of the data. It's

important to note that the training converges if and only if the data is linearly separable. When faced with a more complex problem in which data separation cannot be efficiently achieved by a single line, the limitations of the Perceptron become apparent, compromising its ability to perform the task adequately.

Motivated by this problem, the Multilayer Perceptron (MLP) was created with the objective of separating datasets that are not linearly separable. In order to achieve this goal, the Multilayer Perceptron (or MLP) stacks several Perceptron functions. Each Perceptron in an MLP is a neuron, and each connection between neurons is a parameter of the MLP. That way, each layer feeds the next layer with the results of its operations. This process ripples through all the layers until it gets to the final output layer. The MLP is the basis of Convolutional Neural Networks (CNNs) explored next.

3.4.2 Convolutional neural networks (CNNs)

Convolutional Neural Networks, or CNNs, are a kind of neural network architecture that became a popular choice when dealing with tasks related to image and time-series-based data. A CNN basic structure has a part destined to feature learning, i.e. the convolutional and pooling steps; and a part destined to the task itself, e.g. classification, which is done usually through the usage of a fully-connected neural network, in other words a multilayer perceptron, whose input is a flattening of the output of the previous layer.

Convolutions and poolings compose the basic operations made during the feature learning phase, extracting meaningful features. Inspired by the mathematical definition of a convolution between two continuous functions (Weisstein 2003), the convolution made by CNNs convolves what is called a kernel, in other words a filter (an n-dimensional tensor), with an input image, in order to extract key components of the input, therefore learning different features from the input data (Albawi et al. 2017). The kernel's weights used during this phase are learned and modified during the network training step, in a way that the convolutions are able to extract meaningful information about the input data. Pooling, on the other hand, is used to reduce the output's dimension, creating a more dense representation of the output, and also helping to make the representation approximately invariant to small translations of the input (Goodfellow et al. 2016). This concept of convolutional neural networks was first introduced by LeCun et al. (1998) through a network called LeNet-5, to solve the problem proposed by the MNIST handwritten digits dataset, which consists of classifying which handwritten digit a certain image represents.

Krizhevsky et al. (2012) proposed a deeper convolutional neural network architecture called AlexNet, achieving state-of-art recognition accuracy and winning the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), the most difficult ImageNet challenge regarding visual object recognition task. Throughout the years, deeper convolutional neural networks were proposed with millions of parameters, like the VGG "family" (Sengupta et al. 2019), and with different concepts like residual networks, for instance with ResNet architectures (Wu et al. 2019). Although we may think that having a deeper architecture, with more parameters, leads to better

results, recent studies show that we may have positive and negative impacts due to the depth of a neural network (Sun et al. 2016).

3.4.3 Generative adversarial network (GANs)

Generative Adversarial Network (GAN) is a machine learning framework for estimating generative models (Goodfellow et al. 2020). Data generation may aid in geological interpretation by augmenting the training sets of data. GANs are typically based on deep learning methods and is composed of two models, a generator and a discriminator. The generator learns to generate data considering the distribution of real data, in other words, the generator learns how to create fake data that is similar to real data. Then, the discriminator learns to differentiate the generator's fake data from the real data. During the training, the discriminator penalizes the generator if the artificial data produced is not good enough. This adversarial interplay between the generator and discriminator eventually refines the generator's capacity to generate data closer to real data (Goodfellow et al. 2020).

The first GAN was proposed by Goodfellow et al. (2020) and it uses MLP networks as models for the generator and the discriminator. Since then, several papers on GAN have been published presenting improvements and applications Wang et al. (2021). These works can be grouped according to their objectives of improving generated image diversity (helps avoid mode collapse) (Arjovsky et al. 2017; Dai et al. 2018; Kossaifi et al. 2018; Li et al. 2017), increase in generated image quality (Karras et al. 2017; Brock et al. 2018; Zhang et al. 2019), stabilizing training (Kodali et al. 2017; Chu et al. 2020), and applications (Ledig et al. 2017; Lutz et al. 2018; Zhu et al. 2017).

Some advances in the field of GANs, which were built upon the initially proposed GAN model, are described in the following works: (Arjovsky et al. 2017) helps to avoid vanishing gradient and mode collapse. Radford et al. (2015) published the first work to use a deconvolutional neural network as the generator architecture, and Mirza and Osindero (2014) feeds the generator and the discriminator with additional information to condition them and has some control over the generated output.

An approach like GAN offers a fitting solution for addressing challenges related to seismic data. As mentioned previously, obtaining seismic data is costly, that way the data may not be available sufficiently for some applications. In addition, real seismic data may contain noise and may not have suitable characteristics for training a deep learning model satisfactorily. For denoising and data generation tasks, such as in the situations mentioned above, the implementation of generative adversarial networks has proved to be a well-suited choice.

3.4.4 Transformers and self-attention mechanism

Transformer is a deep learning model that has been widely used in various fields, such as Computer Vision (CV), Natural Language Processing (NLP), and audio processing. The Transform architectures are based on a mechanism called self-attention. The key insight behind this mechanism is its ability to mimic human cognitive

attention, emphasizing essential parts of the input and de-emphasizing less important features.

Wang et al. (2018) introduced self-attention to computer vision and demonstrated significant advancements in tasks such as video understanding and object detection. In image classification, attention mechanisms help the model focus on relevant features and regions that contribute the most to the decision. Assigning different weights to different spatial locations or image channels, the model can emphasize informative regions and disregard irrelevant and/or noisy regions. Similarly, attention mechanisms have proven to be valuable in image segmentation tasks. The attention module assists the model in capturing the spatial relationship between pixels, enabling more precise label assignment to each pixel based on its local and global context (Guo et al. 2022).

Attention mechanisms have emerged as a promising approach in the field of seismic volume interpretation. These mechanisms have shown their value in multiple tasks, such as seismic horizon picking, fault inference, and segmentation. In the context of horizon picking, attention is helpful to highlight the relationships between seismic traces and attend to relevant parts of the seismic volume. When it comes to fault inference, attention is employed to emphasize the planar discontinuities within the seismic volume. Additionally, in segmentation tasks attention enables the model to capture spatial dependencies, crucial to accurate segmentation.

3.5 Evaluation metrics

Evaluation metrics serve as tools to measure the effectiveness of machine learning models after training and facilitate comparisons between different models. Following, we will introduce the key evaluation metrics applicable to different ML tasks.

3.5.1 Classification

The output of a classification model can be categorized by four possible scenarios: True Positive, True Negative, False Positive, and False Negative. A true positive output indicates the model's precise identification of a positive instance as positive, while a true negative output reflects the model's accurate classification of a negative instance as negative. False positives arise as the model's errors, where negatives are erroneously interpreted as positives. This has the potential to result in inappropriate actions driven by flawed predictions. Conversely, false negatives arise when positives are mistakenly classified as negatives. In such scenarios, important instances can be overlooked, leading to significant consequences.

When evaluating a model's performance for classification tasks, accuracy is one of the most commonly employed metrics. This metric measures the ratio of correct predictions (true positives and true negatives) to the total instances, as shown in the Eq. 1. However, its apparent simplicity can be deceiving, especially in scenarios characterized by imbalanced data distributions. To illustrate, consider a binary classification setting where 90% of the instances belong to class 0 and the remaining 10% to class 1. Even if the model consistently predicts class 0, resulting in a seemingly impressive 90%

accuracy, it is important to note that it misclassified 10% of the data. This becomes especially critical when one or both of the classes hold vital significance. In such cases, the model's utility comes into question.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total}} \tag{1}$$

To address this issue, precision and recall serve as two complementary metrics. Precision measures the ratio of predicted positives to actual positives (equation 2). In other words, when the model predicts a positive, how probable is it to be accurate. In the context of binary classification, with 90% of instances belonging to class 0 and the remaining 10% to class 1, the precision score would be 0. This is because it never predicts class 1, which constitutes a true positive.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{2}$$

Conversely, recall assesses the ratio of actual positives to accurate predictions (Eq. 3). In essence, it quantifies how improbable it is for the model to predict a true case as false. Similar to the rationale behind the precision score being 0 in the provided example, the recall would also result in a value of 0.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{3}$$

The F1 score represents the weighted harmonic mean of precision and recall, with values ranging between 0 and 1 (Eq. 4). Achieving an equilibrium between precision and recall is essential for attaining robust performance. The goal is to elevate both precision and recall. The F1 score addresses false positives and false negatives, thereby enhancing comprehension of the model's performance.

$$\text{F1 score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{4}$$

3.5.2 Regression

As mentioned in Sect. 3.2, regression analysis constitutes an approach that seeks to understand the relationship between the characteristics of a data set and a resulting continuous value. This characteristic underlies the difference between regression analysis and classification, with the latter focused on discretely predicting the category to which an instance belongs. It is essential, therefore, to use specific evaluation metrics for each type of task. Metrics traditionally employed in evaluating classification models cannot be directly applied to regression analysis. This is due to the continuous nature of the predictive output in regression tasks. As a consequence, the evaluation of the performance of a regression model is expressed through the quantification of the prediction error.

The Mean Squared Error (MSE) stands as a metric that measures the average of the squared difference between the predicted and real values. Within Eq. 5, y_i denotes the 'i-th' real value within the dataset, $\hat{y}_i$ symbolizes the corresponding predicted value, and n is the number of data points. This difference is squared, nullifying the directional aspect and resulting in a positive error value. Furthermore, squaring accentuates substantial errors, effectively magnifying their impact. In essence, this process inflates disparities, meaning that larger discrepancies between predicted and real values lead to correspondingly amplified squared errors. Consequently, a higher metric value directly correlates with inferior model performance.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{5}$$

Root Mean Squared Error (RMSE) is an extension of the Mean Squared Error. One limitation of the MSE metric is that it yields squared units, meaning that if the model tries to predict attributes like the depth of a seismic fault, the resulting MSE value would be expressed in squared units. RMSE rectifies this by restoring the metric's dimensionality to its original scale, achieved through the square root of MSE (Eq. 6). Despite this adjustment, RMSE maintains the other advantageous properties inherent to the MSE metric.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{6}$$

Another well-known metric is the The Mean Absolute Error (MAE). It measures the absolute gap between real values and predictions, averaging across all observations (Eq. 7).

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{7}$$

3.5.3 Segmentation

The need to compare predicted and ground truth images or volumes introduces numerous challenges when assessing a model's performance. Ideally, one aims to quantify the degree of similarity between a prediction and its corresponding real value.

Pixel Accuracy (PA) serves as a segmentation metric, representing the proportion of pixels in an image that are correctly classified. This metric computes the ratio between the number of accurately classified pixels (true positives and true negatives) and the total pixel count in the image (Eq. 8). It is the accuracy equivalent for image segmentation.

$$\text{PA} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total}} \tag{8}$$

Intersection Over Union (IoU), also known as Jaccard-Index, is a metric that measures the area of the intersection between the predicted segmentation mask ($\hat{\mathcal{Y}}$) and the real segmentation mask ($\mathcal{Y}$) over union of the two classes (Eq. 9). It exhibits slightly improved performance in handling unbalanced classes when compared to PA. In binary or multi-class segmentation scenarios, IoU is individually computed for each class, followed by taking the mean. Similar to accuracy providing a favorable outcome despite less-than-ideal model performance, IoU can also exhibit instability in the presence of unbalanced classes. For instance, considering the same example, if class 1 occupies approximately 10% of the image and the model predicts an image composed solely of class 0, the IoU might yield a substantial score of 45% (when a simple mean is employed). This value is considerable due to the intersection of the images amounting to around 90%.

$$\text{IoU} = \frac{|\hat{\mathcal{Y}} \cap \mathcal{Y}|}{|\hat{\mathcal{Y}} \cup \mathcal{Y}|} \tag{9}$$

The Dice coefficient (DC) is an evaluation metric widely used in the computer vision community that assesses the alignment between the predicted segmentation mask ($\hat{\mathcal{Y}}$) generated by the segmentation method and real mask ($\mathcal{Y}$). To this, it quantifies the degree of agreement by measuring the double overlap area relative to the total number of pixels in both segmentations (equation 10). Consequently, this metric assigns greater significance to successful matches compared to IoU.

$$\text{DC} = \frac{2 * |\hat{\mathcal{Y}} \cap \mathcal{Y}|}{|\hat{\mathcal{Y}}| + |\mathcal{Y}|} \tag{10}$$

3.5.4 Object detection

When dealing with object detection, we can also apply the concepts of True Positives, True Negatives, False Positives, and False Negatives, already discussed in the evaluation metrics of classification tasks. However, in the case of object detection, it is necessary to set a threshold to decide what instance counts as a correct detection and what does not. One commonly employed method for setting this threshold is by employing the Intersection Over Union (IoU) value, a metric previously explained. Consequently, we define an object as detected when the overlap between the model's predicted boundary and the actual object's boundary surpasses the defined threshold. Conversely, if this IoU falls short of the threshold, we consider the object as undetected.

This approach enables the computation of two crucial metrics: precision, indicating a model's ability to correctly identify pertinent objects, and recall, reflecting its capability to identify all relevant instances. From these metrics, it is possible to construct a precision-recall curve, which essentially represents the trade-off between precision and recall across various threshold values related to the bounding boxes generated by the model. Generally, a lower threshold results in a higher number of

items being retrieved, thereby boosting recall but potentially reducing precision. Conversely, a higher threshold yields enhanced precision but may decrease recall.

Average Precision (AP) is a summary metric derived from the precision-recall curve. It is computed by summing the product of the change in recall (R_i - R_{i-1}) and precision (P_i) at each threshold point denoted as i, among n thresholds. This approach ensures that AP considers precision at different recall levels, with the shift in recall values serving as a weight, emphasizing the importance of high-recall regions within the Precision-Recall curve. High recall regions are valuable because they indicate that the model is capturing a large portion of the relevant items. By assigning greater importance to these high-recall regions through the use of the shift in recall values as a weight, AP acknowledges the model's ability to achieve both comprehensiveness and accuracy in its predictions.

$$\mathrm{AP} = \sum_{i=1}^{n} (R_i - R_{i-1}) * P_i \tag{11}$$

The Mean Average Precision (mAP) is the overall accuracy of object detectors across all categories within a given dataset. It is essentially the arithmetic mean of the individual AP scores computed for each class k, among m classes.

$$\mathrm{mAP} = \frac{1}{m} \sum_{j=1}^{m} AP_k \tag{12}$$

3.6 Loss functions

In ML algorithms, choosing a suitable loss function is crucial for the quality of the obtained results. That's because the loss function error gradient guides the choice of the model's parameters, and the minimization of inadequate loss function error can lead to a model being unable to approximate the desired function. As such, a quick discussion on the most common cost functions employed, including cross-entropy, balanced cross-entropy, focal loss, dice, MSE, and SSIM, is important before the more in-depth analysis of the state-of-the-art ML models to come.

Loss functions are commonly defined as the average of individual losses per example. In this sense, they may be written as:

$$L_\theta(\mathcal{D}) = \frac{1}{|\mathcal{D}|} \sum_{i=1}^{|\mathcal{D}|} l(h_\theta(\mathbf{x}[i]), \mathbf{y}[i]) = l(\hat{y}_i, y_i)$$

Where y_i is the correct label, and $\hat{y}$ is the prediction. Observe that $\hat{y}_i = h_\theta(\mathbf{x}[i])$ is the output of the network. In this sense, the last layer of a network is commonly tasked with scaling the output to the task at hand. For instance, in binary classification, the last layer is commonly activated via a sigmoid, whereas for regression, it is a linear activation. We now discuss the particular forms of l, the individual loss.

Binary Cross-Entropy (BiCE) is among the most commonly used loss functions and is a specific case of two classes Cross-Entropy (CE) loss. Formally, BiCE is by Eq. 13.

However, as is often the case on fault identification tasks (where the number of pixels corresponding to a fault is immensely smaller than no-fault pixels) BiCE may lead the model to guess the dominant label on imbalanced data sets.

$$l(\hat{y}_i, y_i) = BiCE(\hat{y}, y) = -(y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)) \tag{13}$$

For this reason, balanced cross-entropy is frequently utilized in seismic analysis (Zhang et al. 2022; Wu et al. 2019). Balanced cross-entropy is defined as shown in Eq. 14, where β is a weighting factor that determines the impact of each class on the overall cost. Commonly, β is defined as the ratio between the number of true examples and the total number of entries (Wu et al. 2019).

$$BaCE(\hat{y}_i, y_i) = -(\beta * y_i \log(\hat{y}_i) + (1 - \beta) * (1 - y_i) \log(1 - \hat{y}_i)), \tag{14}$$

Another regular approach for addressing the imbalance issue is Focal Loss (Lin et al. 2018), which uses a modulation factor to emphasize hard examples as opposed to easy ones, as shown in Eq. 15. Where p_t is the estimated probability by the model for class $y = 1$, and for $y = -1$ class, the probability is $1 - p_t$. The γ parameter adjusts the rate at which examples easily classified are down-weighted. Similarly to balanced cross-entropy, $\alpha_t \in [0, 1]$ is a class weighting factor.

$$Focal(\hat{y}_i, y_i) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t), \tag{15}$$

Dice loss is a function defined as one minus Dice Coefficient (mentioned in Sect. 3.5), this adaptation is necessary to allow the function minimization during the neural network training process. It can be given by the Eq. 16, where $\hat{\mathcal{Y}}$ and $\mathcal{Y}$ denote the prediction and ground-truth pixels set for a class.

$$Dice(\hat{\mathcal{Y}}, \mathcal{Y}) = 1 - \frac{2|\mathcal{Y} \cap \hat{\mathcal{Y}}|}{|\mathcal{Y}| + |\hat{\mathcal{Y}}|}, \tag{16}$$

Seismic interpretation subtasks, as RGT estimation, can be modeled as a regression problem. In this case, the most common metric employed is Mean Square Error (MSE). This metric serves as an evaluation measure and holds utility as a loss function.

The MSE, however, is a point-wise metric, i.e., it evaluates the output only concerning its corresponding pixel on the label. To capture the structural similarity between seismic images, the mean structural similarity index (MSSIM) is often adopted (Bi et al. 2021; Yang et al. 2023).

4 Our systematic review process

In this section, we provide a detailed description of the execution of our Systematic Review process, aimed at selecting the relevant literature for the survey. A systematic review is a data extraction process that consists of a comprehensive procedure that (i) selects and analyzes relevant studies from the literature, (ii) extracts relevant data/meta-data from them, and (iii) synthesizes the findings.

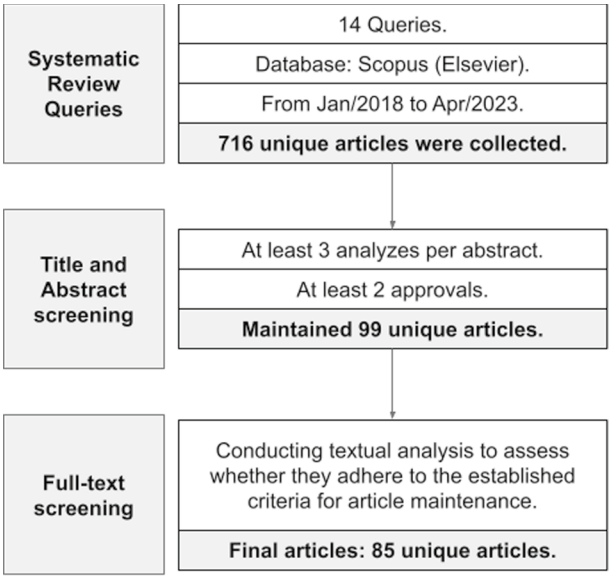


Fig. 2 The systematic review process

It organizes essential components of any research effort, including examining and synthesizing existing material on the state of knowledge in a study area. This review process allows us to grasp a larger picture of a given research field. In other words, we can answer questions that individual research papers could not answer and spot flaws in primary research papers that should be addressed in future studies (Page et al. 2021). Moreover, by employing a systematic literature review and data collection method, we aim to reduce biases in data collection and improve the validity of our presented findings.

Before presenting our review, the methods and steps used to conduct a systematic review and extract data from the relevant literature will be described throughout this section (Fig. 2). These steps include database selection, query keywords, inclusion and exclusion criteria, and data extraction methods. We'll also discuss strategies for evaluating the included research quality and approaches for data synthesis and presentation. We adopt and adapt parts of the PRISMA2020 statement (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) (Page et al. 2021) to structure our findings, which includes a checklist of items to guide the report. We also followed the five steps suggested by Khan et al. (2003) and by [137],[1] a popular systematic review software.

As stated, our survey focuses on the process of automated seismic data interpretation of geological features using machine learning methods, more precisely, deep neural networks. From a geological point of view, our systematic review will inquire

[1] https://www.covidence.org.

about the tasks of fault detection, horizon extraction, and Relative Geologic Time (RGT) estimation from seismic data. We shall focus on computer vision applications from a computational perspective, referring to the problems given, mainly classification, segmentation, and regression tasks.

4.1 Systematic review queries

The first stage in conducting a systematic review is to frame the questions that the study will answer. Such questions should be relevant in light of the review's intended use and the knowledge gap that has to be filled. The question-framing process aids researchers in focusing on the presented problem and keeping track of the literature review scope.

The second component of the systematic review consisted of searching for relevant publications. Search engines are employed as search tools for this task, returning publications according to a provided query and defined restrictions. In our setting, Elsevier's Scopus[2] database was used to search and analyze academic papers. Both engines can query multiple repositories using specific keyword searches and date constraints.

All queries were restricted by date, as we were interested in papers made public between 2018 and April of 2023. We built a series of queries using geological and DL keywords to provide in-scope results. These queries are presented in Table 1. The table also presents the number of papers found by each query. In summary, our search resulted in 716 unique entries. We then used Rayyan[3] (Ouzzani et al. 2016) on the next steps.

4.2 Title and abstract screening

Title and abstract screening is an important stage in systematic reviews, meta-analyses, and other forms of research synthesis. This procedure entails evaluating the significance of possibly relevant research based on their titles and abstracts. It efficiently identifies research likely to match the project'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 716 unique entries, providing a flag (included, maybe, or excluded) signaling individual concerns about whether the study should stand to the next stage. Our procedure involved registering at least three reviewers per paper, with no maximum amount of opinions. If at least two reviewers approved a paper, it was kept for the next stage. Some of the excluding reasons in this stage are stated below. By the end of this procedure, 99 articles were selected to move forward in the process.

Excluding Reasons: The article does not use seismic data or deep learning techniques, the approach solely incorporates shallow machine learning methods, the work is a survey, an introductory paper, or a dissertation, and the predominance content of the article lies beyond the intended scope.

[2] https://www.elsevier.com/solutions/scopus.

[3] https://www.rayyan.ai.

Table 1 Searched queries and number of papers returned

Query	#
("deep learning" OR "convolutional neural network" OR "recurrent neural network" OR "graph neural network" OR "deep neural network") AND ("stratigraphy" OR "chronostratigraphy" OR "eustasy" OR "accommodation space")	138
("deep learning" OR "convolutional neural network" OR "recurrent neural network" OR "graph neural network" OR "deep neural network") AND ("tectonic" OR "subsidence")	200
("deep learning" OR "convolutional neural network" OR "recurrent neural network" OR "graph neural network" OR "deep neural network") AND (("fault" AND "seismic") OR ("fold" AND "seismic"))	437
("deep learning" OR "convolutional neural network" OR "recurrent neural network" OR "graph neural network" OR "deep neural network") AND ("relative geologic time" OR "sequence stratigraphy" OR "stacking pattern" OR "stratum pattern")	29
("deep learning" OR "convolutional neural network" OR "recurrent neural network" OR "graph neural network" OR "deep neural network") AND ("systems tracts" OR "aggradational" OR "progradational" OR "retrogradational")	3
("deep learning" OR "convolutional neural network" OR "recurrent neural network" OR "graph neural network" OR "deep neural network") AND ("onlap" OR "downlap" OR "baselap" OR "toplap")	0
("deep learning" OR "convolutional neural network" OR "recurrent neural network" OR "graph neural network" OR "deep neural network") AND ("erosional truncation" OR "discordance" OR " structural restoration" OR " structural balancing")	54
("deep learning" OR "convolutional neural network" OR "recurrent neural network" OR "graph neural network" OR "deep neural network") AND ("paleobathymetry" OR "depositional environment")	22
("deep learning" OR "convolutional neural network" OR "recurrent neural network" OR "graph neural network" OR "deep neural network") AND (("base level" AND "basin") OR ("climate" AND "geology"))	59
("generative adversarial networks") AND (("fault" AND "seismic") OR ("fold" AND "seismic") OR "stratigraphy" OR "chronostratigraphy" OR "eustasy" OR "accommodation space")	19
("generative adversarial networks") AND ("tectonic" OR "subsidence" OR "relative geologic time" OR "sequence stratigraphy" OR "stacking pattern" OR "stratum pattern")	6
("generative adversarial networks") AND ("systems tracts" OR "aggradational" OR "progradational" OR "retrogradational" OR "onlap" OR "downlap" OR "baselap" OR "toplap")	4
("generative adversarial networks") AND ("erosional truncation" OR "discordance" OR " structural restoration" OR " structural balancing" OR "paleobathymetry" OR "depositional environment")	7
("generative adversarial networks") AND (("base level" AND "basin") OR ("climate" AND "geology"))	5
Total entries	**983**
Total unique entries	**716**

4.3 Full-text screening

Since only 99 studies were accepted in the previous phase, the full-text screening step was conducted parallel with data extraction. By doing this, we could look

deeper into the eligible papers and better understand the approaches used for automated seismic interpretation. This allowed us to extract relevant metadata and key elements regarding these articles, further discussed in Sect. 4.4. At this point, we also present eligibility criteria for including and excluding papers during the full-text screening stage. Some studies had both reasons for inclusion and exclusion, and some had more than one.

Reasons for inclusion: The article has innovative methods or architectures and presents exceptional results or multi-scale approaches. It addresses the tasks of fault detection, horizons estimation, or RGT estimation. The work discusses data augmentation, loss function analysis, metrics, and transfer learning.

Reasons for exclusion: The article lacks a relevant contribution or presents redundant content.

Due to the large number of papers, only one reviewer gave his opinion for approval or rejection of each study to compose the final report. After trialing the publications by full-text screening, 85 publications were selected to compose the main content of this report. The main aspects of each paper can be evaluated in Sect. 4.4.

4.4 Data extraction

We now turn to the final step of the review. Here, the obtained data must be organized and synthesized. Given that we did not have access to the full text of all articles from the previous step, in this section, we focus on 85 of the 99 previously selected papers (those with full text available).

Year: The first organizational aspect is the year of publication. In the search methods, we filtered studies published since 2018. Figure 3 displays the number of relevant publications available in the selected years, considering those chosen for the final data analysis. As we can see, the periods spanning 2018 to 2019 and 2021 to 2022 experienced the biggest spikes in publication numbers, with each interval witnessing an increase of 7 articles. The decrease in publications in 2023 is expected given that this is the current year.

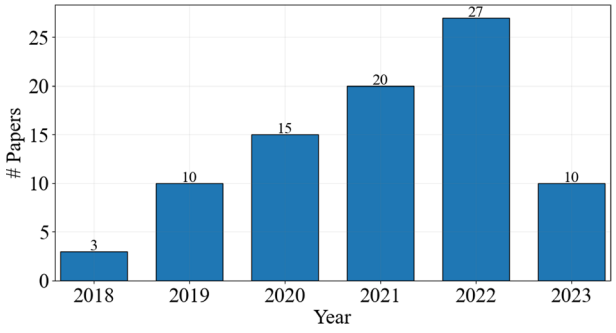


Fig. 3 Papers per year

Table 2 Per geological task

Task	Count
Fault inference	56
Horizon inference	9
RGT inference	6
Data generation	3
Denoising	3
Resolution improvement	3
Facies classification	2
Data interpolation/reconstruction	1
Metrics	1
Structural modeling	1
Total	85

Table 3 Papers per paradigm

Learning paradigm	Count
Supervised	74
Unsupervised	7
N/A	3
Self-supervised	1
Total	85

Task: Secondly, we organized the studies according to their main tasks to seismic interpretation. The resulting count of paradigm usage is shown in Table 2.

We have split the publications into three objectives: fault inference, horizons inference, and RGT inference. Other tasks can contribute to improving the results of the main ones or use some technique that can be applied in the context of the main ones. The main tasks and their relationship with machine learning problems are addressed throughout this work.

The last two studied dimensions (Year and Task) are not specific to machine learning. For this reason, we now break down papers per ML concepts (Tables 3, 4, and 5).

Machine Learning Paradigm (Task): We begin by characterizing the ML paradigm (or ML task) employed. We use the work ML paradigm in contrast to the ML task to avoid confusion with the previously described geological task. Thus, we tried to identify the ML paradigms used in the articles. This result is shown in Table 3.

Looking at the table, it can be noticed that the supervised approach is the most employed (87.05% of the articles). This result is due to the nature of the main objectives, fault inference, horizons inference, and RGT inference, all requiring labeled data. Moreover, from a machine learning perspective, these geological interpretations can be classified as classification (fault and horizon inference) and regression (RGT inference) tasks.

Table 4 Per cost function

Cost function	Count
Not specified	28
Bal. cross-entropy	19
Cross-entropy (CE)	15
Binary CE	14
MSE	8
Focal Loss	4
Dice	3
Horizon Loss	3
Structural Similarity Loss	3
Mean Squared Error + KL divergence Loss	2
MSE + KL	2
MS-SSIM	2
Smoothed Dice Loss	2
Multi-Scale Structural Similarity	2
Mean Squared Error + Discriminator Loss	2
Mean Squared Error + Cosine Similarity	1
Squared Hinge Loss	1
Tanh Cross-Entropy	1
L1	1
Multi-Scale Residual Loss	1
Supervised Loss	1
Cross-Entropy (CE) + L1	1
Adversarial Loss	1
KL divergence Loss	1
Voxel Loss	1
Sparse Binary Cross-Entropy	1
Cycle Consistency Loss	1
KL divergence Loss + Bernoulli Likelihood	1
Mean Squared Error + L2	1
Identity Loss	1

Table 5 Papers per input

Input style	Count
3D data	41
2D data	31
3D and 2D data	7
1D data	2
2.5D data	2
3D and 1D data	1
N/A	1
Total	85

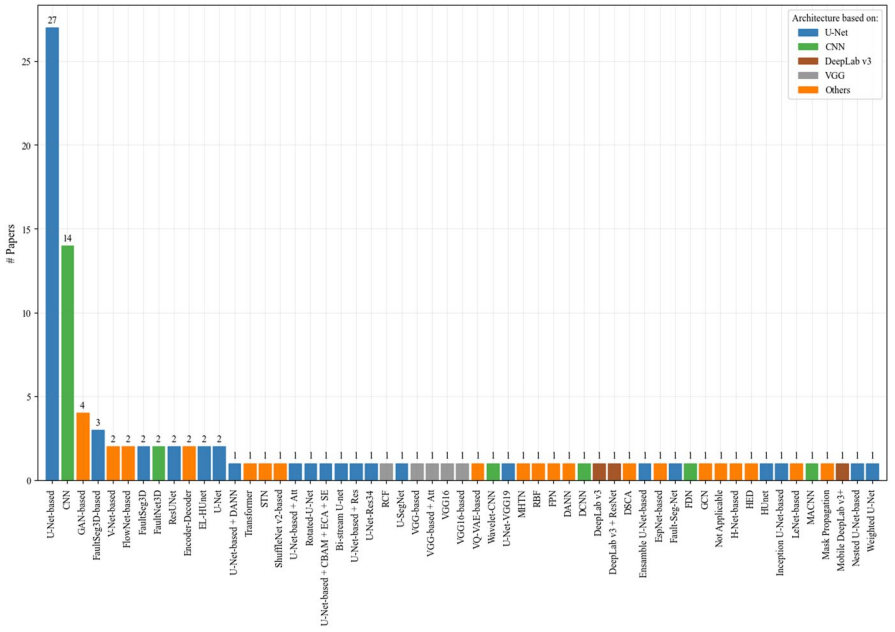


Fig. 4 Deep architectures per paper

Cost functions: As we have mentioned previously, it is common for ML algorithms to optimize a cost function to train models. Thus, we provide the cost functions used in the accepted studies in this topic. Table 4 exhibits the frequency of main cost functions employed.

Altogether 30 different cost functions were used. The most frequently used are Balanced Cross-Entropy (22.35%), Cross-Entropy (17.65%), Binary Cross-Entropy (16.47%), and MSE (9.41%). The four most used cost functions are also the most known in the literature, this fact may justify their wide use in the analyzed works. However, the focal loss and dice loss cost functions are theoretically more adequate to deal with the type of data used and the tasks of fault inference, horizons inference, and RGT inference than the Cross-Entropy loss, for example.

Machine Learning Input Type: The input type refers to the data format used to train the network. The types of data used in the selected works are shown in Table 5.

The input data type choice can be guided by several factors, such as available computational capacity, work objective, and task type. Despite the higher computational cost, 57,64% of the studies used 3D data (tensors). 2D data (matrices) were also widely used in some studies (44.70% of the articles), in which the output was interpolated to have a volume and extract fault planes. In some studies, the objective was to detect a pattern locally using small data sizes. These cited examples help to justify the lack of a direct relationship between the type of data used and the tasks.

Deep Learning Architectures: Finally, we turn to the Deep Learning (DL) architecture used. The frequency (number of papers) of usage of each architecture is shown in Fig. 4.

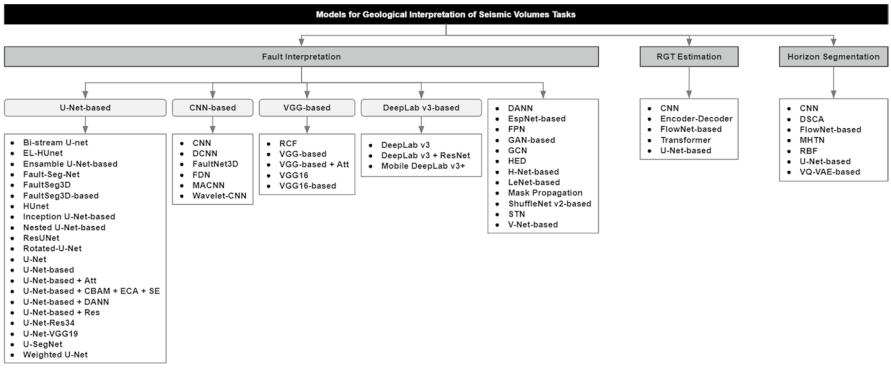


Fig. 5 Taxonomy

For this figure, we use the name of the architecture when the authors of the paper give one. This does not necessarily mean that architectures are inherently different across papers. For instance, most papers use some variation of the U-Net network (Ronneberger et al. 2015). The U-Net is a network initially proposed for medical image segmentation and later became known for general segmentation tasks (mainly due to its popularity with seismic volume). Consequently, several architectures are variations of the U-Net in the figure. When an "U"-shaped (also called U-Net-based) architecture was used but not named, we also added it to the figure.

Furthermore, we also point out that whenever we were able to identify the architectural family used (such as the VGG family (Sengupta et al. 2019)) we also added it to the figure. It is also important to point out that several papers mention generic Convolutional Neural Networks (CNNs) as the central architecture.

In our systematic review, we identified that most of the works, specifically 53 (62.35%), utilized architectures based on U-Net. In second place were CNN-based architectures, which featured in 20 (23.52%) of the works. Altogether, we classified the models into a diverse array of 54 distinct types.

4.5 Taxonomy

Based on the models identified through the systematic review, we developed a taxonomy that categorizes the Deep Learning architectures used for each of the three tasks defined in this survey (Fig. 5). In all applicable cases, we group models that share the same architecture.

5 From ML tasks to geological tasks

This section presents the main articles related to the three geological tasks that are the focus of this survey, namely: fault inference (Sect. 5.2), Relative Geologic Time (RGT) inference (Sect. 5.3), and horizons inference (Sect. 5.4). As was discussed in Sect. 4, several papers share similar DL architectures. Thus, to simplify our

discussion, the three subsections above present a general overview of how different papers tackle each one of the geological tasks. It is essential to point out that each task is inherently complex. Therefore, this is the primary reason we grouped articles based on geological tasks, not DL model.

Due to the large number of research papers (85 as discussed in Sect. 4), which are the focus of study in this review, summarizing this vast literature is challenging. Thus, before going into our discussions for the particular geological task, we present the reader with a table of abbreviations that are used throughout this section.

Finally, before presenting our discussion, we point out that Table 6 presents the abbreviations used throughout this section. This table was created to help the reader grasp the large literature we summarize here.

5.1 Datasets

When using deep learning methods, the data used for training play a critical role in the quality of the predictions (Cortes et al. 1994). Therefore, understanding the seismic volume is critical to design effective models. This data can be of two types: real and synthetic. Real data captures the exact geological attributes in a particular basin, while synthetic data try to mimic these attributes through geological and geophysical models or modifying real seismic data.

The acquisition and processing of real seismic data involve a series of steps aimed at representing the subsurface geological reality, aiming to enhance the quality and interpretability of the acquired seismic data, such as noise removal, velocity correction, filtering, migration, and advanced techniques like data inversion. For tasks such as seismic horizon inference, fault inference, and relative geological time estimation (RGT), geologists and geophysicists still need to label a massive portion of the seismic volume, which is tedious and challenging. The quality of labeling plays an essential role in the model's performance. However, due to interpretation ambiguities and the large volume of data, it is practically impossible to guarantee labels that represent all desired features throughout the real seismic volume without ambiguity and within a short timeframe by the geoscientist.

Synthetic data, along with its corresponding labels, is artificially produced through the utilization of numerical and geophysical models and the transformation of real field data. The generation of synthetic seismic data typically involves four distinct stages: firstly, the generation of a random reflectivity field; secondly, the creation of folding and faulting patterns; thirdly, the application of smoothing convolutions to minimize sharp edges; and lastly, the introduction of random noise to enhance realism. This process allows for the creation of extensive volumes of synthetic seismic data, which can be easily customized to suit specific scenarios, thereby enhancing the ease of conducting experiments and tests. Consequently, the cost-effectiveness and adaptability of synthetic data render it a choice for training deep learning models.

Generalizing a model trained on synthetic data to real data can be tricky, and it is still an open problem in machine learning in geosciences and all machine learning fields. The artificially generated seismic can be placed under several conditions,

Table 6 Table of abbreviations

Column	Abbreviation	Meaning	Column	Abbreviation	Meaning
Ref	ASPP	Atrous spatial pyramid pooling	Metrics	ODS	Optimal dataset scale
	Att	Attention		OIS	Optimal image scale
	CBAM	Convolutional block attention module		PearCor	Pearson correlation
	DANN	Domain adversarial neural network		Pr	Precision
	DCNN	Deep denoising CNN		Pr+	Precision modified
	DSCA	Deep sparse convolutional autoencoders		PSNR	Peak Signal-to-noise ratio
	ECA	Efficient channel attention		R2	Coefficient of determination
	FDN	Fault denoising network		Rc	Recall
	FPN	Feature pyramid network		Rc+	Recall modified
	GCN	Graph convolutional network		R. Jacard	Robust Jaccard
	HED	Holistically-nested edge detection		RMSE	Root mean squared error
	MHTN	Multitask horizon tracking network		ROC	Receiver operating characteristic
	MACNN	Multiscale attention CNN		Sen	Sensitivity
	OPM	Overlapped patch merging		SNR	Signal-to-noise ratio
	RBF	Radial basis function		Spe	Specificity
	RCF	Richer convolutional features		SSIM	Structural similarity
	Res	Residual block		T-SNE	t-Distributed stochastic neighbor embedding
	SE	Squeeze-and-excitation block		WCA	Weighted class average accuracy
	STN	Spatial transformer network	Loss	Adv	Adversarial loss
DType	R	Real		BaCE	Balanced cross-entropy
	S	Synthetic		Ber	Bernoulli likelihood
Metrics	Acc	Accuracy		BiCE	Binary cross-entropy
	AP	Average precision		CE	Cross-entropy
	AUC	Area under the ROC curve		CosSim	Cosine similarity
	Brier	Brier score		Cyc	Cycle consistency loss
	ECE	Expected calibration error		Dice	Dice loss

Table 6 (continued)

Column	Abbreviation	Meaning	Column	Abbreviation	Meaning
	F1	F1-Score		Discr	Discriminator loss
	FreqS	Frequency spectrum		Focal	Focal loss
	FDA	Fault detection accuracy index		H.Loss	Horizon loss
	FSMI	Fault similarity index		Id	Identity loss
	HEERV	Horizon extraction error by RGT value		L1	L1-norm
	IoU	Intersection over Union		L2	L2-norm
	IoU+	Intersection over Union modified		MS-SSIM	Multi-scale structural similarity
	Jacc	Jaccard index		MSR	Multi-scale residual loss
	KL	Kullback?leibler divergence		S.Dice	Smoothed dice loss
	MAE	Mean absolute error		S.Hinge	Squared hinge loss
	MIoU	Mean intersection over union		S.Loss	Supervised loss
	MRPD	Mean relative percentage difference		SBiCE	Sparse binary cross-entropy
	MSE	Mean squared error		TCE	Tanh cross-entropy
	NLL	Negative log-likelihood		Voxel	Voxel loss

Table 7 Most common public datasets used on articles of this survey

Dataset	Location	Data type	Attributes	Used
Wu et al	N/A	Synthetic	Fault	25/75
F3	Dutch North Sea	Synthetic, Real	Fault	9/75
Thebe Gas Fied	Exmouth Plateau, Australia	Real	Fault	4/75

yielding a more generalized model on paper. However, the systematic differences between field and synthetic data are significant. Although the synthetic data mimic the characteristic of the real seismic volume, many times the model used to generate the data cannot capture the complexity and variability existing in the interested basin. Thus, regularization techniques may help model generalization (Zhang et al. 2021).

One common approach to address this challenge is combining synthetic and real data. The real data will contain essential and unique geological features about the seismic, while the synthetic data contributes to different scenarios, allowing for a more robust learning (Seib et al. 2020). This technique acts as a data augmentation process, preventing overfitting. It is also possible to train a model on synthetic data and transfer its learning to real data by fine-tuning its parameters.

Ensemble learning combines different models' predictions to minimize generalization errors. Since data variance, noise, and bias are the significant sources of error, it is possible to train several models with several different seismic volumes and assemble them together (Li et al. 2022).

Likewise, standardization aims to reduce the domain gap of the two distributions, i.e, to minimize the difference in the statistical distribution of different data. One of the most common methods of standardization, apart from min-max normalization, is the z-score, that is a simple mathematical function which aims to measure how distant a certain point is from the population's mean. Other standardization methods have been proposed, each using an empirical or theoretical function to reduce the domain gap (Tang et al. 2023).

To further reduce the gap between real and synthetic data, recent efforts have been made use of more advanced deep learning techniques to enhance generated data. Jun Park et al. (2022), used neural style transfer to help incorporate features of real seismic surveys into artificial data. Durall et al. (2021) used cycleGAN to create an image-2-image translation from real data, to synthetic data. The proposed model is able to learn common features in real seismic data, and replicate it on synthetic ones, increasing the similarity between the two domains. Taking a different approach, Zhao et al. (2023) employ a GAN to generate seismic images and their respective labels in an end-to-end process. On their proposed architecture, a fault mask is used as a constraining input, and the generator must learn to create seismic images which respect the indicated fault locations.

The three most commonly used public datasets are listed in the Table 7. Used in about 33% of the articles, the data artificially generated by Wu et al. (2019) is

present in a wide variety of works regarding the creation of synthetic seismic volumes, or model's performance comparison for example.

5.2 Fault inference

Fault Inference was the task most commonly found in the literature. Overall, 66 papers were selected in this review, with a quick summary of all chosen works available in Table 8. We found that:

- Among the evaluated works, U-Net-based architectures dominated the landscape for fault interpretation tasks, with variations of U-Net featuring prominently in 45 (68.18%) of the studies. Following behind, CNN architectures emerged as the second most prevalent choice, being employed in 15 (22.72%) of the analyzed works. It's worth noting that one paper did not assess any architecture for this task.
- Among the works, 41 (62.12%) employed 3D data formats, while we identified one work where no data was utilized.
- Out of all the works, 54 (81.81%) employed synthetic data, with only one work not utilizing any data. A more in-depth discussion about datasets and data types will be held in Sect. 5.1.
- The most frequently employed evaluation metric in the studies was accuracy, with 21 of them (31.82%) utilizing it as their primary metric. Notably, in 22 (33.33%) of the works, the specific evaluation metric used could not be identified.
- The loss function most commonly encountered in the works was Balanced Cross Entropy, found in 18 of them (27.27%). However, in 17 studies (25.76%), the specific loss function used could not be discerned.

Beyond the task of Fault Inference, we also found some papers that performed tasks that were considered essential for the good results of other methods, though they should have focused on the inference itself. Amongst these, some papers focused on the generation of synthetic data and will be discussed in Sect. 5.1. With a similar objective of improving data quality, other works centered on removing noise from authentic seismic images (denoising) or increasing their resolution (super-resolution). Finally, one paper focused on the selection of seismic attributes, which could serve as additional input to aid the process of Fault Inference.

Though the task of Fault Inference is most commonly modeled as a semantic segmentation problem, a lot of the architectures we encountered during the review are based on simple CNN networks, more typically applied to classification tasks. Generally, in these works (Zhao and Mukhopadhyay 2018; Zhao 2020; Wei et al. 2022; Wang et al. 2022; Cunha et al. 2020; Pochet et al. 2019; Mosser and Naeini 2021; Liu et al. 2018), the model is trained to output the class of the center pixel of the input image, meaning the network has to be independently applied to several

Table 8 A summary of the Fault papers

Ref	Input	DType	MS	TL	OA	Others T.	Metrics	Loss
Alfarhan et al. (2022): Resnet/VGG	2D	R		✓			Pr/Rc/F1/IoU	BaCE
Alohali et al. (2022): FaultSeg3D	3D	R			✓		Acc/IoU/Jacc	BiCE
An et al. (2021): HED/RCF/DeepLab v3+	2D	R			✓		AP/OIS/ODS	BaCE/BiCE
An and Dong (2023): DeepLab v3	2D	R/S		✓			AP/OIS/ODS	-
Cheng et al. (2022): U-shaped	3D	S			✓		-	-
Cunha et al. (2020): CNN	2D	S		✓			F1/AUC	S. Hinge
Di et al. (2020): V-Net/H-Net	3D	R			✓		-	CE
Dinh and Nguyen (2022): Inception U-Net	3D	S					Acc	BiCE
Dou et al. (2023): GAN	3D	R	✓			Ger	PSNR/SSIM	TCE
Du et al. (2022): CNN	2D	R				Den	PSNR/DB	MSE
Durall et al. (2021): U-Shaped/GAN	2D	R/S				Ger	–	BiCE/Adv/Cyc/Id
Egorov (2019): V-Net+SE	–	S	✓				Dice	BaCE
Feng et al. (2021): FaultSeg3D		S	✓		✓		Acc	
Gao et al. (2021): U-Shaped+Att	2D/3D	S	✓				Acc/Pr/Rc	S. Dice
Gao et al. (2022): Nested U-Net	2D/3D	S					–	S. Dice
Guillon et al. (2020): N/A	3D	R	–		✓		Pr+/Rc+/IoU+/MSE/MAE	–
Hu et al. (2022): VGG16+ASPP	2D	R	✓		✓		Acc/MIoU	BiCE
Jiang and Norlund (2020): GAN	3D	S				Resol	–	–
Jiang et al. (2022): U-Shaped	–	S	✓		✓		–	–
Jiang and Norlund (2022): N/A	–	–	–	–	–	AttrSel	–	–
Jing et al. (2023): U-Shaped	3D	S			✓		Acc	BaCE
Li et al. (2021): U-Shaped	2D	S				Den/Resol	MIoU	MSE/MS-SSIM
Li et al. (2019): U-Shaped	2D	S					FDA	BiCE
Li et al. (2022): FaultSeg3D/EL-HUnet	3D	S	✓				Pr/Rc/F1/AUC	BaCE
Li et al. (2023): Fault-Seg-Net	2D	S	✓				Pr/Rc/F1/MIoU	MSR/Dice/Focal
Lin et al. (2022): U-Shaped+Att	2.5D	R/S	✓				F1/WCA/IoU	Focal

Table 8 (continued)

Ref	Input	DType	MS	TL	OA	Others T.	Metrics	Loss
Liu et al. (2018): CNN	2D	S			✓		Acc	CE
Lu et al. (2018): GAN	2D	R				Resol	FreqS	MSE+Discr
Lyu et al. (2022): U-SegNet	3D	R/S	✓				Pr/Rc/ROC	Focal
Ma and Li (2021): Rotated-U-Net	3D	R/S	–					CE
Ma and Duan (2022): Weighted U-Net	3D	S	✓				–	BiCE
Mosser and Naeini (2021): CNN	2D	R/S			✓		IoU	CE/Brier/ECE
Mosser et al. (2020): U-Shaped	3D		✓		✓			CE
Mosser and Zabihi Naeini (2022): Ensamble U-Net	3D	S			✓		NLL/IoU/Brier/ECE	BaCE/BiCE/KL+Ber
Norlund and Jiang (2022): U-Net	3D	R/S					–	BaCE
Otchere et al. (2022): ResUnet	3D	R/S			✓		–	–
Palo et al. (2023): GCN	2D	R/S			✓		Acc	S.Loss
Jun Park et al. (2022): DeepLab v3/VGG	2D	R/S				Ger	KL/T-SNE	BiCE
Pochet et al. (2019): CNN	2D	S					Acc/Sen/Spe/F1/AUC	CE
Sang et al. (2020): CNN	2D	S				Den/Ger	–	MSE
Sarajaervi et al. (2020): U-Shaped	3D	S			✓		R.Jaccard	–
Shen et al. (2022): FaultSeg3D	3D	S	✓		✓		–	BiCE
Tang et al. (2023): U-Net+Att+OPM	2.5D	S	✓				–	BaCE
Van-Ha and Thanh-An (2022): U-Shaped	3D	S	✓				Acc	CE
Wang and Ma (2020): VGG + Att	3D	R	✓				Acc/Pr/Rc/F1	BaCE
Wang et al. (2021a): ShuffleNet v2+ASPP	3D	R/S	✓				–	–
Wang et al. (2021b): ResUNet	3D	S	✓				Acc/Rc/F1/IoU	BiCE/KL
Wang et al. (2022): CNN	3D	R/S			✓		–	CE
Wei et al. (2022): CNN	2D	S		✓	✓		Acc/Sen/Spe/F1/AUC	Focal/CE/BaCE
Wu et al. (2021): U-Shaped	3D	S					–	BaCE

Table 8 (continued)

Ref	Input	DType	MS	TL	OA	Others T.	Metrics	Loss
Wu et al. (2022): U-Shaped	3D	S	✓				Acc	BiCE/Dice
Wu et al. (2019): FaultNet3D	2D/3D	S			✓		Pr/Rc/ROC	–
Wu et al. (2019): FaultSeg3D	3D	S					Pr/Rc/ROC	BaCE
Wu et al. (2020): N/A	–	–	–	–	–	Ger	–	–
Yan et al. (2021): U-Shaped	2D	R/S		✓			–	BaCE
Yu et al. (2022): UNet+CBAM+ECA+SE	3D	S	✓				Pr/Rc/ROC/F1	BaCE
Zhang et al. (2022): HUnet/EL-HUnet/U-Net	2D/3D	R/S	✓				Pr/Rc/F1	BaCE
Zhao et al. (2023): U-Shaped/RCF/FPN	2D	S	✓			Ger	Wass	CE+L1/BaCE
Zhao and Mukhopadhyay (2018): CNN	3D	S			✓		–	–
Zhao (2020): CNN	3D	S			✓		–	–
Zhou et al. (2020): Mask Propagation	3D	S	–				–	–
Zhou et al. (2021a): U-Shaped	3D	R/S		✓			Acc	BaCE
Zhou et al. (2021b): Domain Adversarial	2D	R/S		✓	✓		Acc	–
Zhou et al. (2023): U-Shaped	3D	S				Resol	Acc/Rc/PSNR/SSIM	SSIM/Voxel
Zhou et al. (2022): U-Net – wavelet pooling	3D	R					Acc	–
Zhu et al. (2022): EspNet/STN	2D/3D	R/S	✓				Acc	BaCE/SBiCE/MSE+L2

Input indicates the type of data used for training, inform Data Type (DType) shows the data used for training, real or synthetic, Multi-Scale (MS) indicates multi-scale technique uses, Transfer Learning (TL) indicates whether transfer learning uses Other Approach (OA) indicates that it approached the problem differently, Other Tasks (Others T.) informs the complementary task used

overlapping patches until it is able to predict the value of every pixel and, thus, finalizing the segmentation.

In the exact center-pixel-based modeling, (Palo et al. 2023) proposes using a Graph Convolutional Network instead of a simple CNN. The authors treat each of the individual 2D input patches as a node in a graph; two nodes are connected if they belong to the same class, i.e., fault or non-fault. The job of the GCN is to classify each of the nodes, given the extracted amplitude and attribute patches. Compared to a regular CNN, the GCN was orders of magnitude faster and achieved competitive, if not better, results, especially when combined with a successive difference. This measures the difference between seismic layers.

More frequently, however, a segmentation-specific network was used to classify every input pixel simultaneously. For this, many authors have used Fully Convolutional Networks, which were already commonplace in the more extensive deep-learning literature. Namely, DeepLabV3 and DeepLabv3+ were the most popular architectures, being employed with slight modification in some works (An and Dong 2023; Jun Park et al. 2022; An et al. 2021; Li et al. 2023). Wu et al. (2021) modified another traditional DL model, the VGG16, to create a novel FCN architecture, where the deconvolutional strides in the up-sampling stage of the network are gradually reduced to maintain a high level of detail.

As mentioned previously, U-Net-based architectures were the most commonly utilized method for the Fault Inference task. The original U-Net is composed of four down-sampling convolutional and four up-sampling blocks convolutional. Many authors use the pioneer raw architecture, only adapting the networks to the inputs (Cheng et al. 2022; Li et al. 2019; Sarajaervi et al. 2020). Given its high computational cost, many authors modify this architecture to lower the down and up-sampling stages, but the symmetric shape of the network remains the same.

Most notoriously, Wu et al. (2019) propose to train a simplified 3D U-Net with synthetic data. The created network, named FaultSeg-3D, was largely successful and became a staple in the literature, being largely used as both a baseline or initial model in several other works (Alohali et al. 2022; An et al. 2021; Dinh and Nguyen 2022). Further simplifiyng the model proposed by Wu et al. (2019), Dinh and Nguyen (2022) introduced the Inception U-Net. This architecture was able to achieve higher accuracies than FaultSeg3D (Wu et al. 2019) with 12x fewer overall parameters, though its convergence did require four times as many epochs.

To enrich the learnable features of the model, Ma and Li (2021) places an inverse-sample block (IS block) based on an inverse residual structure, aiming to reduce the loss of information at the bottom of the network. This block is placed right before the encoding part of the U-shaped architecture. The block works similarly to a small U-Net, making use of downsampling, upsampling, and skip-connection. Their result indicates that the IS block, alongside dropouts, provides a more efficient model.

Gao et al. (2022) propose a simplified version of U^2-Net (Qin et al. 2020) which more aptly fit the small amount of data generally used for the task of Fault Inference. Their Nested U-Net is trained using a smoothed dice loss, which the authors note, after tests, was the most suitable cost function for the problem. The model was built for both 3D and 2D inputs and generated fault maps which, visually, were more clear and coherent than their baseline U-Net.

A frequent problem in the above networks is their general inability of accurately capturing minor faults. Though several factors contribute to this outcome, a large part is due to the model's detected features being mainly of lower resolution and, therefore, capable only of highlighting large-scale faults. To addres this issue, Li et al. (2023) proposed a promising novel architecture, the Fault-Seg-Net. The network was trained with a compound loss that combined focal loss, dice loss, and the sum of absolute errors to achieve excellent results on synthetic and field examples.

Also hoping to detect minor faults better, An et al. (2021) adopts a network originally built for multi-scale edge detection in images, the Holistically Nested Edge Detection Network (HED) Xie and Tu (2015), as well as a variation of it, Richer Convolutional Features for Edge Detection (RCF) Liu et al. (2017). These two new approaches were then compared with FaultSeg3D and DeepLabV3+ using standard edge detection evaluation metrics and yielded great results, though the new networks required more time to be aptly trained.

Li et al. (2022) additionally propose a U-Net variant of HED, the HU-Net, and compare it with a vanilla U-shaped network, achieving results similar to An et al. (2021). By combining several light-weight HU-Nets, the authors then propose El-HU-Net, an ensemble learning version of HU-Net that outperformed all other baselines. Lastly, they dive into the training data itself, verifying that a mixture of synthetic and real data on the training was beneficial, and a larger size of input was also connected to better results.

Zhang et al. (2022) conduct another series of experiments confirming all the findings of Li et al. (2022). Furthermore, the authors also compare the use of 3D and 2D data, concluding that, though 3D data did lead to better continuity in detected faults, 2D models generally outperformed 3D ones, essentially, they hypothesize, due to the larger volume of 2D data.

Also attacking the systematic difference between 2D and 3D data, Tang et al. (2023) proposed a U-shaped architecture to treat 2.5D data, which outperforms several 3D models, yielding less noisy predictions than the traditional 2D models. The remarkable results are due to two critical factors: an empirical-based standardization of the seismic, a significative decrease in the domain gap between real and synthetic data, and an innovative architecture. In this approach, the model receives a series of adjacent 2D seismic images in order to predict one 2D seismic in fault or non-fault. Moreover, the architecture is not a simple UNet: The model uses a Swin Transformer, a block to compute self-attention, dividing the input in a series of windows and shifted windows. In addition, the pooling, crucial to the autoencoder architecture, is different from the average or max pooling. The proposed architecture uses Overlapped Patch Merging (OPM), which divides the input into overlapping windows.

Similarly, Lin et al. (2022) also use 2.5D data on a U-shaped architecture. However, instead of using Swin Transformers, the attention mechanism used by the authors is the Squeeze-and-Excitation module, which computes the channel-wise attention of the input. The authors showed the efficacy of a 2.5D approach and conducted a series of experiments in cropping the seismic volume into different size 2D patches. Yu et al. (2022) built different 3D U-shaped networks with different attention mechanisms, such as Squeeze-and-Excitation, CBAM, and ECA. Their results

indicate that the addition of the attention modules contributes to more robust learning, outperforming several literature models, including the standard 3D UNet.

Other works aim to improve small-fault predictions by pre-processing the input image rather than changing the network itself. Using a simple yet effective approach, Ma and Duan (2022) enhances faults by combining Principal Component Analysis and edge-preserving fault enhancement techniques, proposed by Yang et al. (2010). The improved images are fed at different scales to different 3D U-Nets, with the final output being the weighted average between predictions of all networks. This weighted U-Net, similar to the models described above, creates a multi-scale model. Shen et al. (2022) highlight the faults of the input seismic images via a 3D scattered wavelet transform, which visually improved the result of their proposed U-Net. We note, however, that this processing took nearly 10 h to be completed on the 200 seismic volumes.

Some papers attempt to improve the seismic volumes rather than simply making faults more evident. Commonly, a big culprit of a model's poor outputs is the lousy quality of the seismic images, which are often riddled by noise and low resolution. Addressing this issue, many authors have used deep-learning models to increase the resolution of the model's input in a process called image super-resolution (Zhou et al. 2023; Lu et al. 2018; Jiang and Norlund 2020; Otchere et al. 2022; Li et al. 2021) or to denoise the original image (Li et al. 2021; Sang et al. 2020; Du et al. 2022; Otchere et al. 2022).

In Zhou et al. (2023), prior fault information is used to train a teacher network, which then, by knowledge distilation (Gou et al. 2021), aids in the training of a student network. This process allows the model to accurately learn to generate super-resolution images, though the network is more expensive than simple CNN-based approaches.

Li et al. (2021) add a subpixel convolution layer (Shi et al. 2016) to a traditional U-shaped architecture to reduce the network training time and memory utilization while still increasing the receptive field. The new model is able to quickly and efficiently extract a super-resolution counterpart for the original image whilst simultaneously denoising it. A standard Fault inference model is then tested on the improved images, leading to better results. Overall, though, it should be noted that, during the denoising process, artifacts may arise in low-quality regions.

Instead of attempting to minimize the semantic gap between real and synthetic data, other works aim to mitigate this issue by applying transfer learning methods to their methods. An and Dong (2023) conducts a systematic review of transfer learning applied to Fault Inference alongside several experiments to determine the best ways of pre-training a network. The authors suggest that networks should be pre-trained on ImageNet. However, synthetic data does help to mitigate the need for large, labeled, real seismic datasets.

Cunha et al. (2020) compare three transfer learning methods applied to a network trained initially with synthetic data: in the first, named full fine tuning (FFT), all the weights of the model are modified in the fine-tuning process; in the second, the convolutional layers are treated as a fixed feature extractor, and solely the classifier, i.e., the dense layers are trained; the third is similar to the second, except an SVM is used to perform the classification. Unsurprisingly, FFT outperformed the other

approaches, but the reduced number of training weights of other methods allowed the model to be rapidly tuned using only a CPU.

Other works have shown that, even with little real training data, quality results could still be achieved with the aid of transfer learning. In Alfarhan et al. (2022), the authors used a ResNet model already pre-trained on ImageNet (Deng et al. 2009) as the backbone for his U-Net architecture. These pre-loaded weights allowed for the training to be conducted with a small number of data points whilst maintaining good results.

Yan et al. (2021) trained a raw 2D U-Net on synthetic data and subsequently fine-tuned the model using real data. To automatically label the real data, fault enhancement techniques were employed. This approach yielded performance improvements, requiring only a minimal amount of actual seismic data. Similarly, Zhou et al. (2021a) trained a 3D U-Net on synthetic data. Nevertheless, instead of fine-tuning their model to real data, the authors treat the difference between synthetic and real seismic with a domain adversarial network.

A different method sometimes used to improve the quality of fault predictions is to do additional processing on the outputs of the inference model (Hu et al. 2022; Zhao and Mukhopadhyay 2018; Wei et al. 2022; Qi et al. 2022; Alohali et al. 2022). Typically post-processing stacks are based around morphological operations, that is, set of processing techniques that alter the shape of the image's features (Goyal 2011). These operands can be applied to remove artifacts or unwanted imperfections commonly created by other processing tasks. Alohali et al. (2022) uses a probabilistic variation of the Hough Transform Galamhos et al. (1999), a line detection algorithm, which accelerates it's computation. Similarly, Li et al. (2019) uses a different line extraction method, the Line Segment Detection (LSD) Von Gioi et al. (2012), which performed better than the Hough transforms on their experiments.

Though works often focus on small faults, large ones may also be difficult to capture since their long extension blends itself into the overall seismic shape. Attacking this issue, Cheng et al. (2022) uses a standard U-shaped network to delineate major faults and a seismic gradient disorder to highlight minor faults. Jiang et al. (2022) uses as an input for a U-Net, a frequency decomposed seismic data in order to highlight bigger faults.

Aiming to build a probabilistic network for fault prediction and, additionally, hoping to measure uncertainty in the model's outputs, Mosser and Zabihi Naeini (2022) creates a U-shaped architecture and compare different network approaches: fully deterministic (balanced and unbalanced cross-entropy), concrete dropout, SWAG and ensemble-based (three models). The probabilistic ensemble-based network outperforms the other approaches in most metrics, though its computational cost is notably higher.

Sarajaervi et al. (2020) uses a standard U-Net model trained on synthetic data. However, the metric used to evaluate the performance is rather different: the robust Jaccard metric. In a similar effort to improve the evaluation of Fault Inference methods, Guillon et al. (2020) adds tolerance to traditional classification metrics, ensuring they are less sensitive to the pixel-wise discrepancies between detected faults and labels. To achieve this, the proposed modified precision, recall, and IoU, rather than simply comparing a pixel in the output with its counterpart in the label, create a

Gaussian distance function. This measures how close a predicted fault is to the real one.

5.3 RGT inference

In our systematic literature review, we selected eight articles using deep learning methods related to the RGT inference task. These articles are summarized in Table 9. The main points found were:

- The prevailing architectural choice in works related to the RGT inference task was the U-Net variation. These U-Net-based architectures were referenced in 4 (50%) of the selected works.
- 62.5% of the studies employed data in a 3D format.
- 62.5% of the studies utilize synthetic data.
- The metrics MAE, Accuracy, RMSE, Precision, MRPD, Rc, and F1 were each employed in 2 (25%) of the papers. In 3 papers (37.50%), the evaluation metric used was not disclosed.
- MSE emerges as the most prevalent loss function, featured in 50% of the examined works. One paper (7.14%) does not specify the loss function it employed.

Next, we highlight some of the works found. Bi et al. (2020) used a U-Net based network (a modification of the U-Net network) to obtain a seismic volume and its

Table 9 A summary of the RGT papers. Input indicates the type of data used for training, inform Data Type (DType) shows the data used for training, real or synthetic, Transfer Learning (TL) indicates whether transfer learning was used, Metrics shows the metrics used, and Loss tells the cost function employed

Ref	Input	DType	TL	Metrics	Loss
Bi et al. (2020): U-Net-based	3D	S			MSE/SSIM
Bi et al. (2021): U-Net-based + SE	3D	S		MAE/RMSE/MRPD/HEE	SSIM/BaCE
Di et al. (2021): Encoder-Decoder	3D	R		Acc/Pr/Rc/F1	CE/MSE
Di et al. (2022): Encoder-Decoder	3D/1D	R		Acc/Pr/Rc/F1	CE/MSE
Geng et al. (2019): U-Net-based	2D	S			CosSim + MSE
Geng et al. (2020): U-Net-based + Res	2D	S		MAE/RMSE/MRPD	MSE
Li and Abubakar (2020): FlowNet-based	3D	R			
Yang et al. (2023): Transformer/CNN	2D	S		HEERV	MS-SSIM/Dice/H.Loss

corresponding RGT. They used synthetic data and a loss function combination of MSE and SSIM loss, as follows: $L = L_{MSE} + \beta(1 - |L_{SSIM}|)$.

Bi et al. (2021) used a U-shaped network (a modification of the U-Net network) with Squeeze and Extraction (SE) residual blocks and channel-wise attention mechanism to obtain an RGT volume that allows the interpretation of faults and horizons. They used synthetic data generated without any manual constrain and based part of the process on the method proposed by Wu et al. (2020). As loss functions, they used $1 - L_{SSIM}$ for the RGT inference task and Balanced Cross-Entropy as a loss function for the fault inference task.

Li and Abubakar (2020) proposed obtaining a volume RGT from the optical flow between two 2D-seismic image pairs. Optical flow fields in video motion estimation tasks represent a dense pixel-to-pixel correspondence between two video frames. The developed workflow estimates the optical flow between each pair of consecutive seismic slices and extracts any horizon or segmentation. Additionally, to obtain the RGT, vertical order constraints were applied during the training and prediction phases. The architecture used here was based on FlowNet (Dosovitskiy et al. 2015). The authors used an encoder-decoder type network with multi-resolution shared weight feature layers and cross-correlation layers. A flow field was randomly generated and applied to a previously elastically deformed seismic image to generate the data that, together with the original seismic image, comprise the network input. The randomly generated field flow serves as the ground truth for the network. In the work, no comparison was established between the proposed form and other RGT methods.

Yang et al. (2023) proposed a multi-task learning network for RGT estimation, extracting horizons and detecting faults simultaneously, which used transformer architecture to capture the global structure patterns in a seismic image. Additionally, it used prior horizon constrain to maintain lateral consistency and improve network performance in complex cases and also decomposed the 3D problem into an independently parallel 2D estimation. The architecture RGT estimation branch is modified from a network named Global–Local Path. It is constructed with an encoder using a transformer backbone to capture global structural information and a decoder with CNN backbone and Selective Feature Fusion (SFF) modules to merge global (from the encoder) and local (from CNN) information for RGT estimation. A CNN was used for the fault detection task with shared features from the RGT branch.

In training the RGT branch, the sum of two loss functions was used. One is based on known horizon segments, and the other is on RGT labels. The first one penalizes if the RGT estimated value along a known horizon segment is not the same and is given by the equation: $\sum_{i=0}^{N} \sum_{j=0}^{M_i} (y_{i,j} - \bar{y}_i)^2$, where N is the number of horizons segments, M is the number of points at each horizon, $y_{i,j}$ is the RGT estimated value of j-th point of i-th horizon, and $\bar{y}_i$ is the RGT value average estimated at the i-th horizon. The second loss function is known as Multi-Scale Structural Similarity (MS-SSIM).

In the works developed by Di et al. (2021) and Di et al. (2022), the RGT was used as input to improve the performance of the proposed model for the facies

classification and identification tasks. In Di et al. (2022), the authors used the seismic and RGT as input to the facies identification task and seismic, RGT, and well data to estimate the porosity properties of the facies. It uses actual data and data from three wells. It uses the difference between the input and reconstructed (output) RGT as part of the loss. Three networks are trained: one to extract features, identify facies, and reconstruct the RGT.

5.4 Seismic horizon inference

In our systematic literature review, we selected eleven articles using deep learning methods related to the horizon inference task. These articles are summarized in the Table 10. Overall, the models for Horizon Inference found in the literature were diverse and mainly were simple regarding the deep learning techniques used. More traditional losses, namely least squares (Wu and Fomel 2018), are still expected. However, novel machine-learning-based efforts have surpassed it, and we now discuss some of these efforts.

Our main findings were:

- The predominant architectural choices in studies related to horizon inference were CNN-based and U-Net-based architectures, with each option being cited in 4 works, accounting for 36.36%.
- A majority of the works, specifically 6 (54.55%), utilize data in a 2D format.
- The overwhelming majority of articles, precisely 9 (81.82%), employed real data in their research.
- The most frequently encountered evaluation metrics in the studies were Precision, R2, and F1 score, each appearing in 2 works, constituting 16.67% usage for

Table 10 A summary of the Horizons articles

Ref	Input	DType	TL	Metrics	Loss
Dodda et al. (2022): DSCA	2D	R		MSE/PearCor/R2	MSE + KL
Dyer and Manral (2022): U-Net-based/RBF	2D	R			
Guo et al. (2020): CNN	3D	S		Pr/F1	CE
Gupta et al. (2019): CNN	2D	R		Pr/Rc/F1	
Li and Abubakar (2021): FlowNet-based	3D	R			
Luo et al. (2023): MHTN/U-Net-based	2D	R		MAE	H.Loss
Siahkoohi et al. (2020): CNN	2D	R			
Syed et al. (2022): U-Net-based	3D	R/S			
Wu et al. (2019): U-Net-based	1D	R		Acc	CE
Yang and Sun (2020): CNN	2D	R		R2/RMSE	CE
Yuan et al. (2022) VQ-VAE-based	1D	S		Entropy	MSE

Input indicates the type of data used for training, inform Data Type (DType) shows the data used for training, real or synthetic, Transfer Learning (TL) indicates whether transfer learning was used, Metrics shows the metrics used, and Loss tells the cost function employed

each metric. Additionally, in 4 studies (36.36%), no metrics were employed for evaluation.

- The Cross-Entropy loss function was the most commonly utilized among the works, with 3 (27.27)% of them employing this function. Furthermore, in 5 studies (45.45%), loss functions were not explicitly mentioned.

Next, we've featured a few works. Dodda et al. (2022) formluates the problem of Horizon Inference as a classification task, in which horizons are nubmered, and the network must classify the center pixel of a seismic patch to it's correct horizon. For this, the authors use an autoencoder to track horizons in a seismic images. A sparsity constrain is additionally added on the proposed network, allowing for the model to more easily find relevant features which, ultimetly, leads to better overall results. Similarly, a simple CNN is employed by Yang and Sun (2020) to determine if the center pixel of a 2D patch is part of a pre-determined, partly labeled horizon.

Though it is expected of a good Horizon inference method to keep any two horizons in a strictly constant order, where their lines never cross each other, that is, often, not the case with deep-learning based methods. To solve this so called cross-horizon phonema (Luo et al. 2023) presents a multitask network focused on horizon tracking called MHTN. It's architecture is made of a shared layer, an auxiliary task, a main task "steps" and a horizon correction part. The first aims to generate multi-scale feature maps for an input seismic data. These feature maps are fed to the auxiliar and main tasks, that aim to obtain the region of the target horizon based on an object detection algorithm and horizon tracking based on a semantic segmentation algorithm, respectively. The last part, as the name implies, aims to obtain a more precise horizon tracking result, removing the cross-horizon phenomenon.

A frequent limitation present in the previous methods is the small amount of horizons concurrently mapped by the model. Addressing this issue, Li and Abubakar (2021) treats seismic images similarly to videos, to create a model based on optical flow. The core concept proposed in this novel method is to consider a seismic an either the inline, crossline or time directions as a sequence of frames, and, thus, estimate the motion of pixel between those frames. Then, two consecutive frames are fed into an encoder-decoder model, which detects a single horizon, specified by a seed-point in the image. By choosing multiple random seed-points, the model is albe to densely track multiple horizons in the same seismic. Additionally, to avoid the cross-horizon phonomena, a vertical constrain is added to the model, which prevents a younger horizon from deeping below a new one.

Aiming to quantify uncertainty in Horizon Tracking algorithms, Siahkoohi et al. (2020) uses the aforementioned least squares method to extract horizons from various samples of a previously obtained posterior distribution of seismic images. The authors then use a Monte-Carlo sampling method to calculate the confidence intervals of the tracking method and find that areas near faults, which have a great variability between samples, are associated with greater uncertainty.

Similarly, Yuan et al. (2022) calculates the entropy to estimate the uncertainty in horizon auto-picking. The authors use a modified version of a non-conventional architecture called VQ-VAE, which is an autoencoder. After turning the model

deeper, the authors introduce Gaussian noise to the input 1D vectors. Their technique takes advantage of the directional structure tensor and dilated horizon search strategy, aiming to correctly identify large fault displacements. The results indicate an accurate and less uncertain horizon picking within the VQ-VAE framework.

6 Challenges and opportunities

The previous information has highlighted the potential of applying DL techniques to fault interpretation, horizon segmentation, and relative geological time analysis. However, it is essential to acknowledge that seismic volume interpretation is inherently challenging, presenting numerous opportunities for further improvement. Throughout this section, we detail some challenges that still exist in current research and opportunities for future research efforts.

Lack of Labeled Data and Sparse Data: As highlighted in Sect. 5.1, the scarcity of labeled datasets poses a significant challenge in optimizing the performance of models in Deep Learning applied to the tasks described in this work. The increasing reliance on synthetic datasets, exemplified by the presence of the Wu et al. (2019) dataset in approximately 33% of the analyzed studies, presents an opportunity to explore innovative strategies to overcome this limitation. However, even when resorting to synthetic data, the complication of sparse labels becomes apparent, highlighting the need to develop enhanced labeling techniques or expand these datasets. This challenging aspect paves the way for the research and implementation of more effective methods for generating synthetic data that can more faithfully capture the nuances of real geological data. Bridging the gap between synthetic and real data is a promising opportunity to enhance deep learning models' generalization and practical applicability in geology.

Lack of Multi-Task Data: The three geological tasks we analyze in this study reveal an intrinsic interdependence. In the literature, we identified a significant gap related to the scarcity of datasets that comprehensively satisfy these three interconnected tasks. The establishment of a carefully annotated database encompassing these tasks could pave the way for the advancement of multi-task learning models. Such models can potentially improve efficiency since, during the learning process, they could create a structure of mutual improvement and control between tasks.

Inherent Geographical Bias of Datasets: Another relevant challenge arises in the generalization of models trained exclusively with geological data from a specific region since the particularities of the source location can significantly impact performance in different contexts. For example, models trained on geological data extracted from regions in Norway may not perform well when used in geological data from Brazil. Furthermore, applying models trained on synthetic data to real data presents complexities because, as mentioned previously, synthetic data cannot adequately capture the nuances of real-world scenarios in many cases.

Scalability: A significant challenge in model training lies in scalability. The scarcity of annotated datasets for the specified tasks and a limited number of available instances presents a notable difficulty. There is a concern regarding the model's effective generalization in smaller datasets, increasing the risk of overfitting,

where the model memorizes training data rather than learning patterns. Complex models like deep neural networks may encounter obstacles in extracting robust representations from small datasets. Training times can be significantly extended despite reduced data quantities, particularly in highly complex models with varied data sizes. Integrating more potent computational resources like Graphical Processing Units (GPUs) or Tensor Processing Units (TPUs) becomes crucial for optimizing training in such scenarios. It's worth noting that parallelism strategies can be explored with the adoption of these additional resources, thus accelerating the training process. However, it's essential to acknowledge that incorporating extra resources may incur substantial costs, necessitating a balanced approach considering data quality, model complexity, and computational efficiency.

7 Conclusions

The interpretation of seismic volumes in geosciences is a crucial task with applications in the exploration and production of natural resources, such as oil and gas. It is a time-consuming and extremely laborious process, which can result in high ambiguity of interpretation. Over the years, computer science has played a crucial role in assisting with this endeavor, leading to the development of specialized software and companies dedicated to visualizing and processing seismic data.

The incorporation of DL into the oil and gas industry, particularly in seismic and geological interpretation, has brought significant economic benefits and technological advancements. Utilizing DL algorithms, AI exhibits the ability to efficiently process extensive seismic and geological datasets, leading to the identification of intricate patterns and correlations. This enhanced comprehension of subsurface characteristics and improved identification of potential oil and gas reservoirs substantially diminish exploration uncertainties and elevate the efficiency of prospect selection. Furthermore, AI greatly contributes to project optimization within the oil and gas sector, covering critical areas like well-site selection, equipment sizing, resource allocation, and strategic decision-making. As a result, there is a substantial reduction in costs and an enhancement in project efficiency, translating into tangible economic advantages. Additionally, AI's impact extends to safety by analyzing real-time operational data such as pressure, temperature, and fluid flow. Through the identification of anomalies and behavioral patterns indicative of potential safety hazards, AI algorithms enable prompt responses and effective mitigation of operational risks, ensuring both personnel safety and environmental preservation. Furthermore, by achieving a better interpretation of seismic data, AI plays a pivotal role in the oil and gas industry, especially in areas like CCUS (carbon capture, utilization, and storage). In this context, AI is instrumental in optimizing carbon capture and storage processes by analyzing vast datasets. This ensures the safe storage of captured carbon, leading to a significant reduction in greenhouse gas emissions and mitigating the environmental impact associated with oil and gas activities.

In this survey, we have provided a general overview of DL methods as valuable tools for seismic interpretation. Specifically, our survey has focused on fault

detection, horizon segmentation, and RGT regression tasks. While our survey only offers a broad overview of current DL methods, it draws comparisons with classical machine learning techniques, provides insights into public and private datasets utilized for model evaluation, and introduces commonly used evaluation metrics. Furthermore, it highlights that the applications of deep learning (DL) in the geological tasks addressed in this work show promising results. This study emphasizes the use of both real and synthetic geological data to validate these techniques. For example, Alfarhan et al. (2022) achieved excellent results in fault detection using the U-Net architecture. Another interesting case is Luo et al. (2023), who implemented horizon estimation using a model they developed based on multitask network cascade (MNC), achieving good results. Additionally, Bi et al. (2020) performed the task of RGT estimation by training the model on synthetic data and obtained positive results in a qualitative validation. These examples highlight the effectiveness and versatility of DL techniques in seismic data interpretation, showcasing their potential to transform current geological practices, reduce exploration uncertainties, and optimize resource allocation in the oil and gas industry. Consequently, the findings of this survey can be of great benefit to geoscientists, computer scientists, and data scientists involved in the interpretation of seismic volumes and related tasks.

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

Computer code availability Not applicable; this is a systematic review without any original code.

References

Abubakar A, Di H, Kaul A, Li C, Li Z, Simoes V, Truelove L, Zhao T (2022) Deep learning for end-to-end subsurface modeling and interpretation: an example from the groningen gas field. Lead Edge 41(4):259–267

Alabbad A, Dvorkin J, Altowairqi Y, Duan ZF (2021) Rock physics based interpretation of seismically derived elastic volumes. Front Earth Sci 8:620276

Alabbad A, Humphrey JD, El-Husseiny A, Altowairqi Y, Dvorkin JP (2023) Rock physics modeling and quantitative seismic interpretation workflow for organic-rich mudrocks. Geoenergy Sci Eng 227:211824

Albawi S, Mohammed TA, Al-Zawi S (2017) Understanding of a convolutional neural network. In: 2017 international conference on engineering and technology (ICET), pp 1–6. IEEE

Alcalde J, Bond CE, Johnson G, Kloppenburg A, Ferrer O, Bell R, Ayarza P (2019) Fault interpretation in seismic reflection data: an experiment analysing the impact of conceptual model anchoring and vertical exaggeration. Solid Earth 10(5):1651–1662. https://doi.org/10.5194/se-10-1651-2019

Alcalde J, Bond CE (2022) Subjective uncertainty and biases: The impact on seismic data interpretation. In: Interpreting subsurface seismic data, pp 103–123. Elsevier

Al-Dossary S, Marfurt KJ (2006) 3d volumetric multispectral estimates of reflector curvature and rotation. Geophysics 71(5):41–51

Alfarhan M, Deriche M, Maalej A (2022) Robust concurrent detection of salt domes and faults in seismic surveys using an improved unet architecture. IEEE Access 10:39424–39435

Alohali R, Alzubaidi F, Van Kranendonk M, Clark S (2022) Automated fault detection in the Arabian basin. Geophysics 87(4):101–109

An Y, Dong R (2023) Understanding the effect of different prior knowledge on cnn fault interpreter. IEEE Access 11:15058–15068

An Y, Guo J, Ye Q, Childs C, Walsh J, Dong R (2021) Deep convolutional neural network for automatic fault recognition from 3d seismic datasets. Comput Geosci 153:104776

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 Sci Rev 243:104509. https://doi.org/10.1016/j.earscirev.2023.104509

Anstey N, Bahorich MS, Bridges SR, Bahorich MS, Farmer SL, et al (2007) Overview of seismic attributes. In: Chopra S, Marfurt KJ (eds.) Seismic Attributes for Prospect Identification and Reservoir Characterization. Chap. 1. https://doi.org/10.1190/1.9781560801900.ch1

Arjovsky M, Chintala S, Bottou L (2017) Wasserstein generative adversarial networks. In: International conference on machine learning, pp 214–223. PMLR

Ashcroft W (2011) A petroleum geologist's guide to seismic reflection. Wiley, London

Bahorich M, Farmer S (1995) 3-d seismic discontinuity for faults and stratigraphic features: the coherence cube. Lead Edge 14(10):1053–1058

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. J Geophys Res Solid Earth 126(9):2021–021882

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. OnePetro

Bjorlykke K (2015) Introduction to petroleum geology. Petroleum geoscience: from sedimentary environments to rock physics. Springer, Berlin, Heidelberg, pp 1–29. https://doi.org/10.1007/978-3-642-34132-8_1

Bond CE (2015) Uncertainty in structural interpretation: lessons to be learnt. J Struct Geol 74:185–200

Bond CE, Gibbs AD, Shipton ZK, Jones S et al (2007) What do you think this is?"Conceptual uncertainty"in geoscience interpretation. GSA Today 17(11):4

Brock A, Donahue J, Simonyan K (2018) Large scale gan training for high fidelity natural image synthesis. arXiv preprint arXiv:1809.11096

Caumon G, Collon-Drouaillet P, Veslud C, Viseur S, Sausse J (2009) Surface-based 3d modeling of geological structures. Math Geosci 41:927–945

Chen Q, Sidney S (1997) Seismic attribute technology for reservoir forecasting and monitoring. Lead Edge 16(5):445–448

Cheng Z, Bian L, Chen H, Wang X, Ye D, He L (2022) Multi-scale fracture prediction technique via deep learning, seismic gradient disorder and aberrance: applied to tight sandstone reservoirs in hutubi block, southern junggar basin. Interpretation (United Kingdom) 10(4):1–81

Chu C, Minami K, Fukumizu K (2020) Smoothness and stability in gans. arXiv preprint arXiv:2002.04185

Cortes C, Jackel LD, Chiang W-P (1994) Limits on learning machine accuracy imposed by data quality. Adv Neural Inf Process Syst 7

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 10 June 2023

Cox DR, Knutz PC, Campbell DC, Hopper JR, Newton AM, Huuse M, Gohl K (2020) Geohazard detection using 3d seismic data to enhance offshore scientific drilling site selection. Sci Drill 28:1–27

Cox DR, Huuse M, Newton AM, Sarkar AD, Knutz PC (2021) Shallow gas and gas hydrate occurrences on the northwest greenland shelf margin. Mar Geol 432:106382

Cunha A, Pochet A, Lopes H, Gattass M (2020) Seismic fault detection in real data using transfer learning from a convolutional neural network pre-trained with synthetic seismic data. Comput Geosci 135:104344

Dai Q, Li Q, Tang J, Wang D (2018) Adversarial network embedding. In: Proceedings of the AAAI conference on artificial intelligence, vol. 32

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, pp 248–255. IEEE

Di H, Gao D (2014) Gray-level transformation and canny edge detection for 3d seismic discontinuity enhancement. Comput Geosci 72:192–200

Di H, Gao D (2016) Efficient volumetric extraction of most positive/negative curvature and flexure for fracture characterization from 3d seismic data. Geophys Prospect 64(6):1454–1468

Di H, Truelove L, Li C, Abubakar A (2020) Accelerating seismic fault and stratigraphy interpretation with deep cnns: a case study of the taranaki basin, new zealand. Lead Edge 39(10):727–733

Di H, Li Z, Abubakar A (2022) Using relative geologic time to constrain convolutional neural network-based seismic interpretation and property estimation. Geophysics 87(2):25–35
Di H, Li Z, Abubakar A (2021) Using relative geologic time to constrain seismic facies classification using neural networks. In: SEG/AAPG/SEPM first international meeting for applied geoscience & energy. OnePetro
Dinh V-HT, Nguyen T-A (2022) 3d-inception-unet: a light-weight u-net variant with inception blocks for 3d fault segmentation in seismic data. In: Proceedings–2022 RIVF international conference on computing and communication technologies, RIVF 2022, 191–196
Dodda VC, Kuruguntla L, Elumalai K (2022) Seismic horizon estimation based on deep learning technique. ECS Trans 107(1):11449
Dosovitskiy A, Fischer P, Ilg E, Hausser P, Hazirbas C, Golkov V, Van Der Smagt P, Cremers D, Brox T (2015) Flownet: Learning optical flow with convolutional networks. In: Proceedings of the IEEE international conference on computer vision, pp 2758–2766
Dou Y, Li K, Duan H, Li T, Dong L, Huang Z (2023) Mda gan: Adversarial-learning-based 3-d seismic data interpolation and reconstruction for complex missing. IEEE Trans Geosci Remote Sens 61
Du R, Liu W, Fu X, Meng L, Liu Z (2022) Random noise attenuation via convolutional neural network in seismic datasets. Alex Eng J 61(12):9901–9909
Durall R, Tschannen V, Ettrich N, Keuper J (2021) Generative models for the transfer of knowledge in seismic interpretation with deep learning. Lead Edge 40(7):534–542
Dyer L, Manral S (2022) Machine learning assisted seismic interpretation: a case study of the loppa high area, barents sea. First Break 40(11):33–39
Egorov D (2019) Automatic fault interpretation from seismic data via convolutional neural networks. In: EAGE subsurface intelligence workshop, vol. 2019, pp 1–5. European Association of Geoscientists & Engineers
Feng R, Grana D, Balling N (2021) Uncertainty quantification in fault detection using convolutional neural networks. Geophysics 86(3):41–48
Figueiredo AM, Gattass M, Szenberg F (2007) Seismic horizon mapping across faults with growing neural gas. In: 10th international congress of the brazilian geophysical society, p 172. European Association of Geoscientists & Engineers
Figueiredo* A, Silva F, Silva P, O. Martins L, L. Milidiú R, Gattass M (2015) A clustering-based approach to map 3d seismic horizons. In: 14th international congress of the brazilian geophysical society & EXPOGEF, Rio de Janeiro, Brazil, 3-6 August 2015, pp 1166–1170. Brazilian Geophysical Society
Figueiredo A, Silva F, Silva P, Milidiú RL, Gattass M (2014) A seismic facies analysis approach to map 3d seismic horizons. In: SEG international exposition and annual meeting, p 2014. SEG
Fomel S (2010) Predictive painting of 3d seismic volumes. Geophysics 75(4):25–30
Frodeman R (1995) Geological reasoning: geology as an interpretive and historical science. Geol Soc Am Bull 107(8):960–968
Galamhos C, Matas J, Kittler J (1999) Progressive probabilistic hough transform for line detection. In: Proceedings. 1999 IEEE computer society conference on computer vision and pattern recognition (Cat. No PR00149), vol. 1, pp 554–560. IEEE
Gao K, Huang L, Zheng Y (2022) Fault detection on seismic structural images using a nested residual u-net. IEEE Trans Geosci Remote Sens 60:1–15
Gao H, Wu X, Zhang J, Sun X, Bi Z (2023) Clinoformnet-1.0: stratigraphic forward modeling and deep learning for seismic clinoform delineation. Geosci Model Dev 16(9):2495–2513
Gao Y, Han X, Wang X, Huang W, Scott M (2020) Channel interaction networks for fine-grained image categorization. In: Proceedings of the AAAI conference on artificial intelligence, vol. 34, pp. 10818–10825
Gao K, Huang L, Zheng Y, Lin R, Hu H, Cladouhos T (2021) Automatic fault detection on seismic images using a multiscale attention convolutional neural network. Geophysics 87(1)
Geng Z, Wu X, Shi Y, Fomel S (2020) Deep learning for relative geologic time and seismic horizons. Geophysics 85(4):87–100
Geng Z, Wu X, Shi Y, Fomel S (2019) Relative geologic time estimation using a deep convolutional neural network. In: SEG international exposition and annual meeting. OnePetro
Gersztenkorn A, Marfurt KJ (1999) Eigenstructure-based coherence computations as an aid to 3-d structural and stratigraphic mapping. Geophysics 64(5):1468–1479
Gonzalez RC, Woods RE (2008) Digital Image Processing. Prentice Hall, Upper Saddle River, N.J. http://www.amazon.com/Digital-Image-Processing-3rd-Edition/dp/013168728X

Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press, Cambridge
Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2020) Generative adversarial networks. Commun ACM 63(11):139–144
Gou J, Yu B, Maybank SJ, Tao D (2021) Knowledge distillation: a survey. Int J Comput Vision 129:1789–1819
Goyal M (2011) Morphological image processing. IJCST 2(4):59
Guillon S, Joncour F, Barrallon P-E, Castanié L (2020) Ground-truth uncertainty-aware metrics for machine learning applications on seismic image interpretation: application to faults and horizon extraction. Lead Edge 39(10):734–741
Guo Y, Peng S, Du W, Li D (2020) Fault and horizon automatic interpretation by cnn: a case study of coalfield. J Geophys Eng. https://doi.org/10.1093/jge/gxaa060
Guo M-H, Xu T-X, Liu J-J, Liu Z-N, Jiang P-T, Mu T-J, Zhang S-H, Martin RR, Cheng M-M, Hu S-M (2022) Attention mechanisms in computer vision: a survey. Comput Vis Media 8(3):331–368
Gupta H, Pradhan S, Gogia R, Srirangarajan S, Phirani J, Ranu S (2019) Deep learning-based automatic horizon identification from seismic data. In: SPE annual technical conference and exhibition. OnePetro
Hafiz AM, Bhat GM (2020) A survey on instance segmentation: state of the art. Int J Multimed Inf Retr 9(3):171–189
Hale D (2013) Methods to compute fault images, extract fault surfaces, and estimate fault throws from 3d seismic images. Geophysics 78(2):33–43
Harshvardhan G, Gourisaria MK, Pandey M, Rautaray SS (2020) A comprehensive survey and analysis of generative models in machine learning. Comput Sci Rev 38:100285
Havelia K, Manral S, Murineddu A (2021) Lithology, porosity and saturation joint prediction using stochastic rock physics modelling and litho-petro-elastic inversion. First Break 39(11):33–43
He Q, Wang Y (2021) Reparameterized full-waveform inversion using deep neural networks. Geophysics 86(1):1–13
Herron DA (2011) First steps in seismic interpretation. Society of Exploration Geophysicists, Texas
Hesthammer J, Landrø M, Fossen H (2001) Use and abuse of seismic data in reservoir characterisation. Mar Pet Geol 18(5):635–655. https://doi.org/10.1016/S0264-8172(01)00011-3
Hongliu Z, Xiaomin Z, Rukai Z, Zhang Q (2012) Guidelines for seismic sedimentologic study in non-marine postrift basins. Pet Explor Dev 39(3):295–304
Hu G, Hu Z, Liu J, Cheng F, Peng D (2022) Seismic fault interpretation using deep learning-based semantic segmentation method. IEEE Geosci Remote Sens Lett 19:3041301
Ikeuchi K (2021) Computer vision: a reference guide. Springer, Berlin
Irakarama M, Laurent G, Renaudeau J, Caumon G (2021) Finite difference implicit structural modeling of geological structures. Math Geosci 53(5):785–808
Jiang F, Jaramillo A, Angelovich S, Norlund P, Toms J (2022) Implementation of frequency-dependent fault identification by convolutional neural networks with uncertainty analysis. SEG Technical Program Expanded Abstracts 2022-August, 1709–1713
Jiang F, Norlund P (2020) Super resolution of fault plane prediction by a generative adversarial network. In: First EAGE digitalization conference and exhibition, vol. 2020, pp 1–5. EAGE Publications BV
Jiang F, Norlund P (2022) Machine Learning-Based Feature Importance Analysis of Seismic Attributes to Assist Fault Prediction. In: International Petroleum technology conference, IPTC 2022
Jing J, Yan Z, Zhang Z, Gu H, Han B (2023) Fault detection using a convolutional neural network trained with point-spread function-convolution-based samples. Geophysics 88(1):1–14
Jun Park M, Jennings J, Clapp B, Biondi B (2022) Realistic synthetic data generation using neural style transfer: Application to automatic fault interpretation. In: Second international meeting for applied geoscience & energy, pp 1714–1718. Society of Exploration Geophysicists and American Association of Petroleum .
Karras T, Aila T, Laine S, Lehtinen J (2017) Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196
Khan KS, Kunz R, Kleijnen J, Antes G (2003) Five steps to conducting a systematic review. J R Soc Med 96(3):118–121
Kodali N, Abernethy J, Hays J, Kira Z (2017) On convergence and stability of gans. arXiv preprint arXiv:1705.07215
Kossaifi J, Tran L, Panagakis Y, Pantic M (2018) Gagan: Geometry-aware generative adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 878–887

Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 25

Laurent G, Ailleres L, Grose L, Caumon G, Jessell M, Armit R (2016) Implicit modeling of folds and overprinting deformation. Earth Planet Sci Lett 456:26–38

Lebrec U, Riera R, Paumard V, O'Leary MJ, Lang SC (2022) Morphology and distribution of submerged palaeoshorelines: insights from the north west shelf of Australia. Earth Sci Rev 224:103864

LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324

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

Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z, et al (2017) Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4681–4690

Li S, Yang C, Sun H, Zhang H (2019) Seismic fault detection using an encoder-decoder convolutional neural network with a small training set. J Geophys Eng 16(1):175–189

Li J, Wu X, Hu Z (2021) Deep learning for simultaneous seismic image super-resolution and denoising. IEEE Trans Geosci Remote Sens 60:1–11

Li Z, Abubakar A (2020) Complete sequence stratigraphy from seismic optical flow without human labeling. In: SEG technical program expanded abstracts 2020, pp 1248–1252. Society of Exploration Geophysicists

Li Z, Abubakar A (2021) Seismic flownet: Using optical flow field for dense horizon interpretation. In: 82nd EAGE annual conference & exhibition, vol. 2021, pp 1–5. EAGE Publications BV

Li S, Liu N, Li F, Gao J, DIng J (2022) Automatic fault delineation in 3-d seismic images with deep learning: Data augmentation or ensemble learning. IEEE Trans Geosci Remote Sens 60

Li X, Li K, Xu Z, Huang Z, Dou Y (2023) Fault-seg-net: A method for seismic fault segmentation based on multi-scale feature fusion with imbalanced classification. Comput Geotech 158

Lin T-Y, Goyal P, Girshick R, He K, Dollár P (2018) Focal loss for dense object detection

Lin L, Zhong Z, Cai Z, Sun AY, Li C (2022) Automatic geological fault identification from seismic data using 2.5d channel attention u-net. Geophysics 87(4)

Li Y, Schwing A, Wang K-C, Zemel R (2017) Dualing GANs

Liu Y, Cheng M-M, Hu X, Wang K, Bai X (2017) Richer convolutional features for edge detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3000–3009

Liu Z, Song C, She B, Li K, Yao X, Hu G (2018) Visual explanations from convolutional neural networks for fault detection. SEG Tech Progr Expand Abstr 2226–2230

Lomask J, Guitton A, Fomel S, Claerbout J, Valenciano AA (2006) Flattening without picking. Geophysics 71(4):13–20

Lu P, Morris M, Brazell S, Comiskey C, Xiao Y (2018) Using generative adversarial networks to improve deep-learning fault interpretation networks. Leading Edge 37(8):578–583

Luo Y, Zhang G, Zhang J, Li Y, Lin Y, Li B, Liang C, Li L (2023) Sequence-constrained multitask horizon tracking. Geophysics 88(2):15–27

Lutz S, Amplianitis K, Smolic A (2018) Alphagan: Generative adversarial networks for natural image matting. arXiv preprint arXiv:1807.10088

Lyu F, Zhou H, Liu J, Zhou J, Tao B, Wang D (2022) A buried hill fault detection method based on 3d u-segnet and transfer learning. J Pet Sci Eng 218

Macrae EJ, Bond CE, Shipton ZK, Lunn RJ (2016) Increasing the quality of seismic interpretation. Interpretation 4(3):395–402

Ma Q, Duan T (2022) Multiscale fault and fracture characterization methods. In: SEG/AAPG international meeting for applied geoscience & energy. OnePetro

Malehmir A, Durrheim R, Bellefleur G, Urosevic M, Juhlin C, White DJ, Milkereit B, Campbell G (2012) Seismic methods in mineral exploration and mine planning: a general overview of past and present case histories and a look into the future. Geophysics 77(5):173–190

Ma Z, Li Y (2021) Rotated-unet: A seismic fault identification network based on inverse sampling block construction. In: 2021 3rd international academic exchange conference on science and technology innovation (IAECST), pp 740–744. IEEE

Manzi MS, Gibson MA, Hein KA, King N, Durrheim RJ (2012) Application of 3d seismic techniques to evaluate ore resources in the west wits line goldfield and portions of the west rand goldfield, south africa. Geophysics 77(5):163–171

Marfurt KJ, Kirlin RL, Farmer SL, Bahorich MS (1998) 3-d seismic attributes using a semblance-based coherency algorithm. Geophysics 63(4):1150–1165

Marfurt KJ, Sudhaker V, Gersztenkorn A, Crawford KD, Nissen SE (1999) Coherency calculations in the presence of structural dip. Geophysics 64(1):104–111
Maulud D, Abdulazeez AM (2020) A review on linear regression comprehensive in machine learning. J Appl Sci Technol Trends 1(4):140–147
McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 5:115–133
Mirza M, Osindero S (2014) Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784
Misra AA, Mukherjee S (2018) Seismic structural analysis. Atlas of Structural Geological Interpretation from Seismic Images, 15–26
Mitchum Jr RM, Vail PR, Sangree JB (1977) Seismic stratigraphy and global changes of sea level: Part 6. stratigraphic interpretation of seismic reflection patterns in depositional sequences: Section 2. application of seismic reflection configuration to stratigraphic interpretation
Mitchum Jr R, Vail PR, Thompson III S (1977) Seismic stratigraphy and global changes of sea level: Part 2. the depositional sequence as a basic unit for stratigraphic analysis: Section 2. application of seismic reflection configuration to stratigraphic interpretation
Monniron M, Frambati S, Quillón S, Berthoumieu Y, Donias M (2016) Seismic horizon and pseudo-geological time cube extraction based on a riemmanian geodesic search. In: 2016 IEEE 12th image, video, and multidimensional signal processing workshop (IVMSP), pp 1–5. IEEE
Mosher D, Bigg S, LaPierre A (2006) 3d seismic versus multibeam sonar seafloor surface renderings for geohazard assessment: case examples from the central scotian slope. Lead Edge 25(12):1484–1494
Mosser L, Naeini E (2021) Deep probabilistic neural networks for geoscience. In: 82nd EAGE annual conference & exhibition, vol. 2021, pp 1–5. EAGE Publications BV
Mosser L, Purves S, Naeini EZ (2020) Deep bayesian neural networks for fault identification and uncertainty quantification. In: First EAGE digitalization conference and exhibition, vol. 2020, pp 1–5. EAGE Publications BV
Mosser L, Zabihi Naeini E (2022) A comprehensive study of calibration and uncertainty quantification for bayesian convolutional neural networks-an application to seismic data. Geophysics 87(4)
Norlund P, Jiang F (2022) Improving machine learning approaches to seismic fault imaging through training augmentation. In: International petroleum technology conference. OnePetro
Otchere DA, Tackie-Otoo BN, Mohammad MAA, Ganat TOA, Kuvakin N, Miftakhov R, Efremov I, Bazanov A (2022) Improving seismic fault mapping through data conditioning using a pre-trained deep convolutional neural network: a case study on groningen field. J Pet Sci Eng 213
Ottesen Ellevset S, Knipe R, Svava Olsen T, Fisher Q, Jones G (1998) Fault controlled communication in the sleipner vest field, norwegian continental shelf; detailed, quantitative input for reservoir simulation and well planning. Geol Soc 147(1):283–297
Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A (2016) Rayyan-a web and mobile app for systematic reviews. Syst Rev 5:1–10
Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE et al (2021) The prisma 2020 statement: an updated guideline for reporting systematic reviews. Int J Surg 88:105906
Palo P, Routray A, Mahadik R, Singh S (2023) Fault detection in seismic data using graph convolutional network. J Supercomput
Pauget F, Lacaze S, Valding T (2009) A global approach in seismic interpretation based on cost function minimization. In: SEG international exposition and annual meeting, p 2009. SEG
Payton CE (1977) Seismic Stratigraphy: Applications to Hydrocarbon Exploration. AAPG memoir. American Association of Petroleum Geologists. https://books.google.com.br/books?id=z0kQvgAACAAJ
Pham N, Fomel S, Dunlap D (2019) Automatic channel detection using deep learning. Interpretation 7(3):43–50
Pochet A, Diniz PHB, Lopes H, Gattass M (2019) Seismic fault detection using convolutional neural networks trained on synthetic poststacked amplitude maps. IEEE Geosci Remote Sens Lett 16(3):352–356
Posamentier HW (2000) Seismic stratigraphy into the next millennium; a focus on 3d seismic data. In: American association of petroleum geologists annual conference, New Orleans, LA, vol. 118
Posamentier HW (2004) Seismic geomorphology: imaging elements of depositional systems from shelf to deep basin using 3d seismic data: implications for exploration and development
Power H, Clarke S (2019) 3d seismic-derived bathymetry: a quantitative comparison with multibeam data. Geo-Mar Lett 39(6):447–467

Prazuck C, Durot B, Savajol V, Lacaze S (2015) Interpretation of complex geo-bodies using a relative geological time model: exmouth sub-basin, australia. In: SEG international exposition and annual meeting, p 2015. SEG
Qi J, Laudon C, Marfurt K (2022) An integrated machine learning-based fault classification workflow. In: SEG/AAPG international meeting for applied geoscience & Energy. OnePetro
Qin X, Zhang Z, Huang C, Dehghan M, Zaiane OR, Jagersand M (2020) U2-net: Going deeper with nested u-structure for salient object detection. Pattern Recogn 106:107404
Radford A, Metz L, Chintala S (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434
Randen T, Pedersen SI, Sønneland L (2001) Automatic extraction of fault surfaces from three-dimensional seismic data. In: SEG International Exposition and Annual Meeting, p 2001. SEG
Rankey EC, Mitchell JC (2003) That's why it's called interpretation: impact of horizon uncertainty on seismic attribute analysis. Lead Edge 22(9):820–828. https://doi.org/10.1190/1.1614152
Renaudeau J, Malvesin E, Maerten F, Caumon G (2019) Implicit structural modeling by minimization of the bending energy with moving least squares functions. Math Geosci 51(6):693–724
Riera R, Bourget J, Paumard V, Wilson ME, Shragge J, George AD, Borgomano J, Wilson T (2019) Discovery of a 400 km2 honeycomb structure mimicking a regional unconformity on three-dimensional seismic data. Geology 47(12):1181–1184
Rivenæs JC, Otterlei C, Zachariassen E, Dart C, Sjøholm J (2005) A 3d stochastic model integrating depth, fault and property uncertainty for planning robust wells, njord field, offshore norway. Pet Geosci 11(1):57–65
Roberts A (2001) Curvature attributes and their application to 3 d interpreted horizons. First Break 19(2):85–100
Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pp 234–241. Springer
Runje D, Shankaranarayana SM (2022) Constrained monotonic neural networks. arXiv preprint arXiv:2205.11775
Saltus RW, Blakely RJ (2011) Unique geologic insights from" non-unique" gravity and magnetic interpretation. GSA Today 21(12):4–11
Sang W, Yuan S, Yong X, Jiao X, Wang S (2020) Dcnns-based denoising with a novel data generation for multidimensional geological structures learning. IEEE Geosci Remote Sens Lett 18(10):1861–1865
Sarajaervi M, Bo TH, Goledowski B, Nickel M (2020) Robust evaluation of fault prediction results: Machine learning using synthetic seismic. In: First EAGE digitalization conference and exhibition, vol. 2020, pp 1–5. EAGE Publications BV
Sarker IH (2021) Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. SN Comput Sci 2(6):420
Sarker IH (2021) Machine learning: algorithms, real-world applications and research directions. SN Comput Sci 2(3):160
Seib V, Lange B, Wirtz S (2020) Mixing real and synthetic data to enhance neural network training–a review of current approaches. arXiv preprint arXiv:2007.08781
Sen S, Kainkaryam S, Ong C, Sharma A (2020) Saltnet: a production-scale deep learning pipeline for automated salt model building. Lead Edge 39(3):195–203
Sengupta A, Ye Y, Wang R, Liu C, Roy K (2019) Going deeper in spiking neural networks: Vgg and residual architectures. Front Neurosci 13:95
Shen S, Li H, Chen W, Wang X, Huang B (2022) Seismic fault interpretation using 3-d scattering wavelet transform cnn. IEEE Geosci Remote Sens Lett 19
Sheriff RE, Geldart LP (1995) Exploration seismology. Cambridge University Press, Cambridge
Shi Y, Wu X, Fomel S (2019) Saltseg: automatic 3d salt segmentation using a deep convolutional neural network. Interpretation 7(3):113–122
Shi W, Caballero J, Huszár F, Totz J, Aitken AP, Bishop R, Rueckert D, Wang Z (2016) Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1874–1883
Siahkoohi A, Rizzuti G, Herrmann FJ (2020) Uncertainty quantification in imaging and automatic horizon tracking - a bayesian deep-prior based approach. In: SEG technical program expanded abstracts 2020, pp 1636–1640. Society of Exploration Geophysicists

Stark TJ (2003) Unwrapping instantaneous phase to generate a relative geologic time volume. In: SEG international exposition and annual meeting, p. 2003. SEG
Stark TJ (2005) Generation of a 3d seismic "wheeler diagram" from a high resolution age volume. In: SEG technical program expanded abstracts 2005, pp. 782–785. Society of Exploration Geophysicists
Stark TJ (2004) Relative geologic time (age) volumes-relating every seismic sample to a geologically reasonable horizon. Lead Edge 23(9):928–932
Sun S, Chen W, Wang L, Liu X, Liu T-Y (2016) On the depth of deep neural networks: a theoretical view. In: Proceedings of the AAAI conference on artificial intelligence, vol. 30
Syed BS, Mukherjee J, Ditia T, Seliem A, Al Kobaisi AS, Andrews B, Jaramillo A, Norlund P (2022) Interpreting subtle faults and multiple horizons layers using machine learning and data-driven approach. In: ADIPEC. OnePetro
Tang Z, Wu B, Wu W, Ma D (2023) Fault detection via 2.5 d transformer u-net with seismic data preprocessing. Remote Sens 15(4):1039
Tariq Z, Aljawad MS, Hasan A, Murtaza M, Mohammed E, El-Husseiny A, Alarifi SA, Mahmoud M, Abdulraheem A (2021) A systematic review of data science and machine learning applications to the oil and gas industry. J Pet Explor Prod Technol 11:1–36
Vail P, Mitchum Jr R, Thompson III S (1977) Seismic stratigraphy and global changes of sea level: Part 3. relative changes of sea level from coastal onlap: section 2. application of seismic reflection configuration to stratigrapic interpretation
Van Bemmel PP, Pepper RE (2000) Seismic signal processing method and apparatus for generating a cube of variance values. Google Patents. US Patent 6,151,555
Van-Ha TD, Thanh-An N (2022) 3d-faultseg-unet: 3d fault segmentation in seismic data using bi-stream u-net. Commun Comput Inf Sci 1688 CCIS, 477–488
Von Gioi RG, Jakubowicz J, Morel J-M, Randall G (2012) Lsd: a line segment detector. Image Processing OnLine 2:35–55
Wang Z, Di H, Shafiq MA, Alaudah Y, AlRegib G (2018) Successful leveraging of image processing and machine learning in seismic structural interpretation: a review. Lead Edge 37(6):451–461
Wang Z, She Q, Ward TE (2021) Generative adversarial networks in computer vision: a survey and taxonomy. ACM Comput Surv (CSUR) 54(2):1–38
Wang J, Zhang J-H, Zhang J-L, Lu F-M, Meng R-G, Wang Z (2021) Research on fault recognition method combining 3d res-unet and knowledge distillation. Appl Geophys 18(2):199–212
Wang E, Amaru M, Jayr S, Payne B (2021) Improved 3d neural network architecture for fault interpretation on field data. SEG Technical Program Expanded Abstracts 2021-September, 3189–3193
Wang X, Girshick R, Gupta A, He K (2018) Non-local neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7794–7803
Wang Z, Li B, Liu N, Wu B, Zhu X (2022) Distilling knowledge from an ensemble of convolutional neural networks for seismic fault detection. IEEE Geosci Remote Sens Lett 19
Wang B, Ma J (2020) Cnn based fault recognition with multi-scale fusion attention mechanism. In: IOP Conference series: earth and environmental science 474(4)
Wehenkel A, Louppe G (2019) Unconstrained monotonic neural networks. Adv Neural Inf Process Syst 32
Wei X-L, Zhang C-X, Kim S-W, Jing K-L, Wang Y-J, Xu S, Xie Z-Z (2022) Seismic fault detection using convolutional neural networks with focal loss. Comput Geosci 158
Weisstein EW (2003) Convolution. https://mathworld.wolfram.com/. Published by: Wolfram Research, Inc
Wu X (2017) Directional structure-tensor-based coherence to detect seismic faults and channels. Geophysics 82(2):13
Wu X, Fomel S (2018) Least-squares horizons with local slopes and multigrid correlations least-squares horizons. Geophysics 83(4):29–40
Wu X, Fomel S (2018) Least-squares horizons with local slopes and multigrid correlations. Geophysics 83(4):29–40
Wu X, Hale D (2015) Horizon volumes with interpreted constraints. Geophysics 80(2):21–33
Wu Y, McMechan GA (2019) Parametric convolutional neural network-domain full-waveform inversion. Geophysics 84(6):881–896
Wu X, Zhong G (2012) Generating a relative geologic time volume by 3d graph-cut phase unwrapping method with horizon and unconformity constraints. Geophysics 77(4):21–34

Wu Z, Shen C, Van Den Hengel A (2019) Wider or deeper: revisiting the resnet model for visual recognition. Pattern Recogn 90:119–133
Wu H, Zhang B, Lin T, Cao D, Lou Y (2019) Semiautomated seismic horizon interpretation using the encoder-decoder convolutional neural network horizon tracking using deep learning. Geophysics 84(6):403–417
Wu X, Liang L, Shi Y, Fomel S (2019) Faultseg3d: using synthetic data sets to train an end-to-end convolutional neural network for 3d seismic fault segmentation. Geophysics 84(3):35–45
Wu X, Shi Y, Fomel S, Liang L, Zhang Q, Yusifov AZ (2019) Faultnet3d: predicting fault probabilities, strikes, and dips with a single convolutional neural network. IEEE Trans Geosci Remote Sens 57(11):9138–9155
Wu X, Geng Z, Shi Y, Pham N, Fomel S, Caumon G (2020) Building realistic structure models to train convolutional neural networks for seismic structural interpretation. Geophysics 85(4):27–39
Wu J, Liu B, Zhang H, He S, Yang Q (2021) Fault detection based on fully convolutional networks (fcn). J Marine Sci Eng 9(3):1–13
Wu X, Hale D (2013) Extracting horizons and sequence boundaries from 3d seismic images. In: SEG international exposition and annual meeting, p 2013. SEG
Wu J, Shi Y, Wang W (2022) Fault imaging of seismic data based on a modified u-net with dilated convolution. Appl Sci (Switzerland) 12(5)
Xie S, Tu Z (2015) Holistically-nested edge detection. In: Proceedings of the IEEE international conference on computer vision, pp. 1395–1403
Yang L, Sun SZ (2020) Seismic horizon tracking using a deep convolutional neural network. J Petrol Sci Eng 187:106709
Yang P-J, Mu X, Zhang J-T (2010) Orientational edge preserving fault enhance. Chin J Geophys 53(12):2992–2997
Yang J, Wu X, Bi Z, Geng Z (2023) A multi-task learning method for relative geologic time, horizons, and faults with prior information and transformer. IEEE Trans Geosci Remote Sens 61:1–20. https://doi.org/10.1109/TGRS.2023.3264593
Yan Z, Zhang Z, Liu S (2021) Improving performance of seismic fault detection by fine-tuning the convolutional neural network pre-trained with synthetic samples. Energies 14(12)
Yu S, Ma J, Wang W (2019) Deep learning for denoising. Geophysics 84(6):333–350
Yuan C, Su M, Ni C, Liu X, Xu Y, Cui X (2022) Horizon auto-picking with quantitative uncertainty evaluation by using a modified vq-vae framework. J Geophys Eng 19(4):788–806
Yu T, Wang X, Chen TJ, Ding CW (2022) Fault recognition method based on attention mechanism and the 3d-unet. Computational Intelligence and Neuroscience 2022
Zeng H, Backus MM, Barrow KT, Tyler N (1998) Stratal slicing; part 1, realistic 3-d seismic model. Geophysics 63(2):502–513
Zeng H, Henry SC, Riola JP (1998) Stratal slicing, part ii: real 3-d seismic data. Geophysics 63(2):514–522
Zhang C, Bengio S, Hardt M, Recht B, Vinyals O (2021) Understanding deep learning (still) requires rethinking generalization. Commun ACM 64(3):107–115
Zhang B, Pu Y, Xu Z, Liu N, Li S, Li F (2022) Exploring factors affecting the performance of deep learning in seismic fault attribute computation. Interpretation 10(4):1–75
Zhang H, Goodfellow I, Metaxas D, Odena A (2019) Self-attention generative adversarial networks. In: International conference on machine learning, pp 7354–7363. PMLR
Zhao T (2020) 3D Convolutional Neural Networks for Efficient Fault Detection and Orientation Estimation. In: SEG international exposition and annual meeting 2019, pp 2418–2422. Society of Exploration Geophysicists
Zhao S, Ding R, Han T, Liu Y, Zhang J, Zhao L (2023) Fault2seisgan: a method for the expansion of fault datasets based on generative adversarial networks. Front Earth Sci 11:1091803
Zhao T, Mukhopadhyay P (2018) A fault detection workflow using deep learning and image processing. In: 2018 SEG international exposition and annual meeting. OnePetro
Zhou R, Cai Y, Zong J, Yao X, Yu F, Hu G (2020) Automatic fault instance segmentation based on mask propagation neural network. Artif Intell Geosci 1:31–35
Zhou R, Yao X, Hu G, Yu F (2021) Learning from unlabelled real seismic data: fault detection based on transfer learning. Geophys Prospect 69(6):1218–1234
Zhou R, Yao X, Wang Y, Hu G, Yu F (2021) Seismic fault detection with progressive transfer learning. Acta Geophys 69(6):2187–2203

Zhou R, Zhou C, Wang Y, Yao X, Hu G, Yu F (2023) Deep learning with fault prior for 3-d seismic data super-resolution. IEEE Trans Geosci Remote Sens 61:1–16

Zhou X, Yin Q, Wang B (2022) Automatic Fault Segmentation Using Wavelet Convolutional Neural Networks. In: Offshore technology conference Asia, OTCA 2022

Zhu J-Y, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision, pp 2223–2232

Zhu D, Li L, Guo R, Tao C, Zhan S (2022) 3d fault detection: using human reasoning to improve performance of convolutional neural networks. Geophysics 87(4)

Authors and Affiliations

Gustavo Lúcius Fernandes[1] · Flavio Figueiredo[1] · Raphael Siston Hatushika[2] · Maria Luiza Leão[1] · Breno Augusto Mariano[1] · Bruno Augusto Alemão Monteiro[1] · Fernando Tonucci de Cerqueira Oliveira[1] · Tales Panoutsos[1] · João Pedro Pires[1] · Thiago Martin Poppe[1] · Frederico Zavam[2]

✉ Flavio Figueiredo
flaviovdf@dcc.ufmg.br

[1] Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

[2] Petrobras, Rio de Janeiro, Brazil