Boletim de Geociências da Petrobras

Topographic lineaments extraction in digital elevation model: a comparison between human interpretation and a convolutional neural network output

Lineamentos topográficos em modelos digitais de elevação: uma comparação entre extração manual e computacional usando redes neurais convolucionais com arquitetura U-Net

[1]Pedro Rossi, [2]Julio Cesar Horta de Almeida, [3]Antônio de Pádua Pires, [3]Mario Neto Araujo
[1]Petrobras, Rio de Janeiro, Rio de Janeiro, Brazil, pedrorossi@petrobras.com.br (Autor correspondente)
[2]Rio de Janeiro State University, Rio de Janeiro, Rio de Janeiro, Brazil
[3]Petrobras, Rio de Janeiro, Rio de Janeiro, Brazil

ABSTRACT

Topographic lineaments seen on digital elevation models (DEM) provide good insights into geological structures and lithotypes. To extract useful information from these images, interpretation is a crucial step. Interpreting DEM images is conventionally done by trained scientists drawing line segments over the image, in an often repetitive, time-consuming, and undoubtedly subjective process. Recently, machine learning techniques have been employed for recognizing patterns in DEM data and extracting lineaments in semi-automated mode. This technique has the potential to help scientists reduce subjectivity and reduce the life cycle of interpreting lineaments on remote sensed images. In this paper we take on the task of extracting lineaments from Shuttle Radar Topography Mission (SRTM) DEM data in the central Ribeira fold belt in SE Brazil, a deep eroded Neoproterozoic-Cambrian orogen. We analyze how human interpretation varied over five rounds of lineaments annotations in the same area and compare them with the output of a convolutional neural network (CNN) with U-net architecture trained to classify images at pixel level into lineaments or non-lineaments. Results show that the criteria for manual lineament extraction varied substantially throughout the interpretation rounds. Lineaments became fewer, and repeatability increased towards the later versions. The CNN hyperparameters had to be set through a series of trial-and-error tests but eventually yielded results that mimic the ones obtained manually. CNNs are powerful tools to process and analyze data, but require time and effort to build, label, parametrize, train and test. In this sense it is justifiable to apply CNNs over large datasets, when it is unfeasible to interpret lineaments manually, otherwise, it is more practical to have a human draw by hand with two or more interpretation versions to reduce the subjectivity.

Keywords: machine learning, manual lineament interpretation, Convolutional Neural Networks, U-Net, topographic lineaments.

RESUMO

Lineamentos topográficos presentes em modelos digitais de elevação (MDE) oferecem boas indicações sobre estruturas e litotipos. Para extrair informação útil desses dados a interpretação é uma etapa crucial, tradicionalmente feita manualmente por um geocientista, em um processo repetitivo, demorado e sem dúvida subjetivo. Recentemente, técnicas de aprendizado de máquina têm sido empregadas para reconhecer padrões e extrair lineamentos de MDEs de maneira semi-automática. Esta técnica tem mostrado potencial de reduzir drasticamente o tempo necessário para a obtenção de lineamentos e reduzir a subjetividade da interpretação em dados adquiridos por sensores remotos. Neste trabalho assumimos a tarefa de extrair lineamentos de dados SRTM (Shutle Radar Topography Mission) na região central da Faixa de dobramentos Ribeira, SE do Brasil, um orógeno Neoproterozóico-cambriano profundamente erodido. Analisamos como a interpretação humana variou ao longo de cinco rodadas de interpretação de lineamentos na mesma área e comparamos os resultados com extrações de lineamentos feitas por uma Rede Neural Convolucional Profunda (RNCP) com arquitetura U-Net treinada para classificar imagens a nível de pixel como lineamento ou não-lineamento. Os resultados mostram que os critérios para extração manual de lineamentos variaram substancialmente ao longo das diferentes rodadas de interpretação. Lineamentos se tornaram menos numerosos e a repetibilidade aumentou em direção às últimas versões. Os hiperparâmetros da RNCP foram definidos em uma série de testes por tentativa e erro, mas entregaram resultados que mimetizam aqueles obtidos manualmente. RNCPs são uma poderosa ferramenta de análise e processamento de dados, mas requerem tempo para construir, anotar, parametrizar, treinar e testar. Neste sentido, se justifica utilizá-las em grandes conjuntos de dados quando não é viável a interpretação manual. Do contrário, é mais prático que um humano desenhe os lineamentos à mão com duas ou mais versões de interpretação para reduzir a subjetividade.

Palavras-chave: aprendizado de máquina, interpretação manual de lineamentos, Redes Neurais Convolucionais, U-Net, lineamentos topográficos.

Recebido: 18 de junho de 2025 | Recebido no formato de revisão: 13 de agosto de 2025 | Aceito: 19 de agosto de 2025 | Disponível online: 30 de outubro de 2025
CC BY 4.0 - Attribution 4.0 International
DOI: https://doi.org/10.70369/188dww73

1. INTRODUCTION

Large amounts of data available in geosciences are best displayed as images, for example seismic data, gravimetry, magnetometry, satellite images, outcrops, thin section photos, and digital elevation models. To give meaning to these images and extract useful data, one must go through the process of interpretation, in which one object or structure is similar, and therefore related to others, and can be classified in the same group.

Manual interpretation is subjective. The way that human visual perception works is still not fully understood. What people see is not just a simple case of translating the retinal stimuli. Vision is the result of unconscious inference, making assumptions based on visual clues and previous experience stored in memory (von Helmholtz, 1866 apud Froner, 2013). This means that the results of one's visual interpretation are conditioned by their former experience, cultural background, and knowledge.

For example, some operators tend to draw lineaments strictly where they are visible, whereas others may be inclined to extrapolate to areas with no clear linear structures to connect two or more lineaments to create a big one. If a structure is slightly curve, some operators will draw one lineament that will overall represent the feature, but others will choose to draw two or more line segments to represent the bending feature. These different interpretation styles are frequently related to the conceptual model adopted by the operator. For instance, if the operator believes that lineaments are expressions of normal faulting, he or she will likely draw lineaments in a fault linkage pattern. On the other hand, if the operator believes that lineaments are related to foliation in a fold belt, he or she will tend to capture folds and shear zones patterns. Hence, although the experience of the operator adds value to the interpretation product, it may also be an important source of bias. In an experiment conducted by Peacock (2019), it is suggested that one operator tended to measure more fractures from satellite images and aerial photographs as his experience increased, and instead of viewing just each individual fracture, it became easier to identify fractures as part of a larger structure. After some interpretation work the image will be full of lines that may cover some of the minor linear features, increasing the chances of letting them pass unnoticed. Or yet, will tend to distribute lineaments in a uniform way, so they will look harder in areas where naturally less structures occur.

There are biological issues that may influence the interpretation results and introduce an unconscious bias. People have different quality eyesight or may work under different psychological conditions that can alter their willingness to detect patterns (Peacock, 2019). Working for long periods of time may induce repetitive stress injury, that will affect the speed and style of interpretation.

The term "lineament" was introduced by Hobbs (1904) but had been used in rather loose manner until O'Leary et al. (1976) proposed the definition: "A lineament is a mappable, simple or composite linear feature of a surface, whose parts are aligned in a rectilinear or slightly curvilinear relationship, and which differs distinctly from the patterns of adjacent features and presumably reflects a subsurface phenomenon." Topographic lineaments are linear crests and valleys, ridgelines and slope breaks (Jordan et al., 2005), not to be mistaken by other linear features such as roads, agricultural areas borders, and other manmade high pixel contrast structures. They are superficial expressions of planar structures such as dykes, faults, rock beddings and shear zones truncated by the topographic surface, and therefore important in many fields of geoscience. Identifying linear features on maps or remotely sensed data is one of the keys to understanding groundwater occurrence (Sander, 2007; Akinluyi et al., 2018; Chandra et al., 2006) and incorporating structural information into landslide assessment (Ramli et al., 2010; Yusof et al., 2011; Lulseged et al., 2005). Lineaments can be representative of fracture zones, which act as secondary porosity and may control migration and accumulation of hydrocarbons (Bashliov, 1982; Enoh et al., 2021; Saha, 2022) and play a main role predicting the distribution of ore deposits, as they often represent discontinuities such as faults, where hydrothermal fluids circulate and concentrate certain elements. In Iran for example, more than 90% of metallic ore deposits lay within 15km of fault-related lineaments and 50% are closer than 5km (Meshkani at al., 2013). In northeast Brazil major lineaments control a large number of mineral-bearing pegmatites (Liu et al., 2000).

Deep Learning (DL) has been increasingly applied to image segmentation problems. Previous works show that, in the field of geosciences, it has the potential to drastically reduce the data interpretation life cycle. An et al. (2021) applied deep convolutional networks (DCNN) to extract fault planes from seismic data, and achieved performances comparable to expert human interpretation, besides reducing time requirements from months to hours. Haslebacher et al. (2023) aimed to identify and classify linear features on the Jupiter´s icy moon Europa. The authors stated that one expert took 4 hours of focused time to map 260 features on a dataset that covers less than 0.1% of Europa's surface. At that rate it would take 500 days of 8-hour focused workdays to map the whole surface. A network composed of convolutional and fully connected layers was employed,

trained in 5.2 hours, and yielded satisfactory results in minutes, even though it is difficult to assess the time involved in building, testing, and evaluating the network. Zhang et al. (2018) tested several deep learning models that classifies geological structures in different categories, such as anticline, ptygmatic folds, gneissose structure, ripple marks, fault, boudin, and others. It was trained over 2206 labeled images, and the best model reached 90% accuracy. Daynac et al. (2024) employed a U-Net CNN to map dune outlines and crests with 90% precision, a promising prospect for monitoring dune dynamics.

This paper describes the use of a CNN trained to find topographic lineaments in DEM data. Its results are compared to a series of manual classic interpretations, and they prove to be, although different in nature, satisfactory for analyzing big data sets, and probably adequate for most lineaments analysis in geosciences.

1.1. Geological setting

The studied area is 52 km wide and 140 km across and extends from the city of Andrelândia, in the north, to the coast (Ilha Grande) in the south. It is inserted in the tectonic context of the Ribeira belt (Hasui et al., 1975). The Ribeira belt is an NE-SW trending orogen that extends for more than 1400 km in the costal southeast of Brazil. It is the result of processes that culminated in the amalgamation of the Gondwana supercontinent between 600 and 540 Ma (Trouw et al., 2000; Schmitt et al., 2004; Heilbron et al., 2020; Freitas et al., 2021) during the Panafrican-Brasiliano orogeny. Driven by lithospheric plate movement, the Archean/Palaeoproterozoic São Francisco-Congo and Angola paleocontinents collided during the last phases of the Neoproterozoic and early Cambrian, forming the supercontinent West Gondwana (Brito Neves et al., 1999). Each tectonic compartment that crops out in the present time is representative of this collision, such as deformed and metamorphized magmatic arcs and continental margin sediments (Figure 1). What is visible at the surface today are the roots of a deeply eroded orogen.

The last major tectonic event recorded was the development of the Continental Rift of Southeastern Brazil (CRSB), as defined by Riccomini et al. (2004) in the Paleogene. It consists of a series of ENE-WSW grabens, approximately 900 km long, roughly along the Brazilian SE cost. In the target area the CRSB developed the Resende and Volta Redonda basins in *en échelon* pattern (Cogné et al., 2013).

In the DEM data it is possible to recognize the overall NE SW fabric trend of the Ribeira belt's foliation, suture lines, shear zones and major faults. The Resende and Volta Redonda basins are seen as smooth terrain at the central north region, and a conspicuous corridor of lineaments of N-S/NNW direction is present in the northern portion. The southern coastal region and Ilha Grande island is a mountainous region dominated by plutonic rocks (Figure 1).

Previous works in the studied area relate the topographic lineaments to normal and transcurrent faulting and Riedel associated structures (Negrão et al., 2020). Guedes et al. (2016) points out that Precambrian structural lineaments are parallel to the Resende-Ilha Grande Dyke swarm (RIGDS) and should control the emplacement of this series of NNW-SSE oriented tholeiitic dykes. The NNE lineaments in the central Ribeira foldbelt contrasts with the main NE direction and coincide with a series of Neoproterozoic-Ordovician foliations, shear zones, lithological boundaries and fractures filled by pegmatitic veins. According to Giro et al. (2021) the NNE lineaments represent a crustal weakness zone that was recurrently reactivated.

2. MATERIALS AND METHODS

2.1. Data

The Shuttle Radar Topography Mission (SRTM) digital elevation model data was acquired by Space Shuttle Endeavour from US space agency NASA in 2000. The raw data was processed by INPE, the Brazilian institute for space research to the resolution of 30 m per pixel, and 1 m of vertical resolution. The data was downloaded from the Topodata.

2.2. Manual interpretation

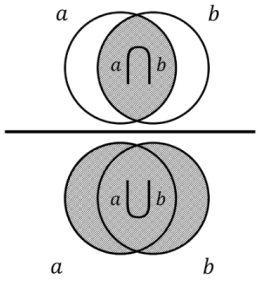
For classical, human interpretation, DEM data was loaded into Esri® ArcMap 10.5, and Hillshade effect was applied, as in Figure 2. Hillshade enhances the topography and makes the image more natural to the eyes by emulating the patterns of light and shade produced by sunlight. This, however, introduces a bias component, as it highlights structures that are perpendicular to the pseudo illumination, but it dims the ones that are parallel (Cardoso-Fernandes et al., 2022). Because of that it is a common practice to interpret DEM images under different pseudo lighting directions. In this case two different pseudo illumination directions were used: 315 and 265 azimuths, both had 45 degrees elevation. These directions were chosen for been nearly orthogonal to the most prominent linear structures. Lineaments were interpreted in two scales: 1:500.000 and 1:250.000 and defined by two points as a line segment, no smaller than 50mm at each scale, so minimum lineament length is 1,250 m in the field. To avoid fatigue bias interpretation

sections lasted one hour, maximum. Lineaments extractions were made by the author, and then repeated three times under the same conditions, with a time spam of at least two weeks between them, allowing a repeatability measurement. Another version of manual extraction was performed by a first-year undergraduate student, to assess the conceptual model bias.

Manual lineament interpretation may vary significantly, even when done by the same operator, under the same conditions. Mabee et al. (1994) suggests a method of manually interpreting lineaments based on repeated mapping sessions by the same observer or more than one observer superimposed, aiming the reduction the inherent subjectivity associated with interpretation. Lineaments were visually assessed and considered coincident when azimuths were within ± 5° and the separation distance between them was within 2 mm at the scale of drawing.

To quantify the repeatability among the various versions of mapping, we apply the Jaccard index of similarity, a well-accepted metric for comparing datasets (Fletcher et al., 2018). For that we take the ratio between the superimposed lineaments and the total amount of lineaments. Equation 1 a) defines the Jaccard index for two datasets a and b and in the case of lineaments similarity, the coinciding ones must be removed from the denominator, as shown in Equation 1 b) and since $J(a,b)$ is different from $J(b,a)$, the mean value was taken.

a) $$J(a,b) = \frac{a \cap b}{a \cup b}$$ or



b) $$J(a,b) = \frac{|\, a \cap b \,|}{|\, a \cup b \,|} = \frac{|\, a \cap b \,|}{|\, a \,| + |\, b \,| - |\, a \cap b \,|}$$

Equation 1. A. Jaccard index of similarity, where a and b are different interpretation datasets. **B.** Jaccard index for overlaying data.

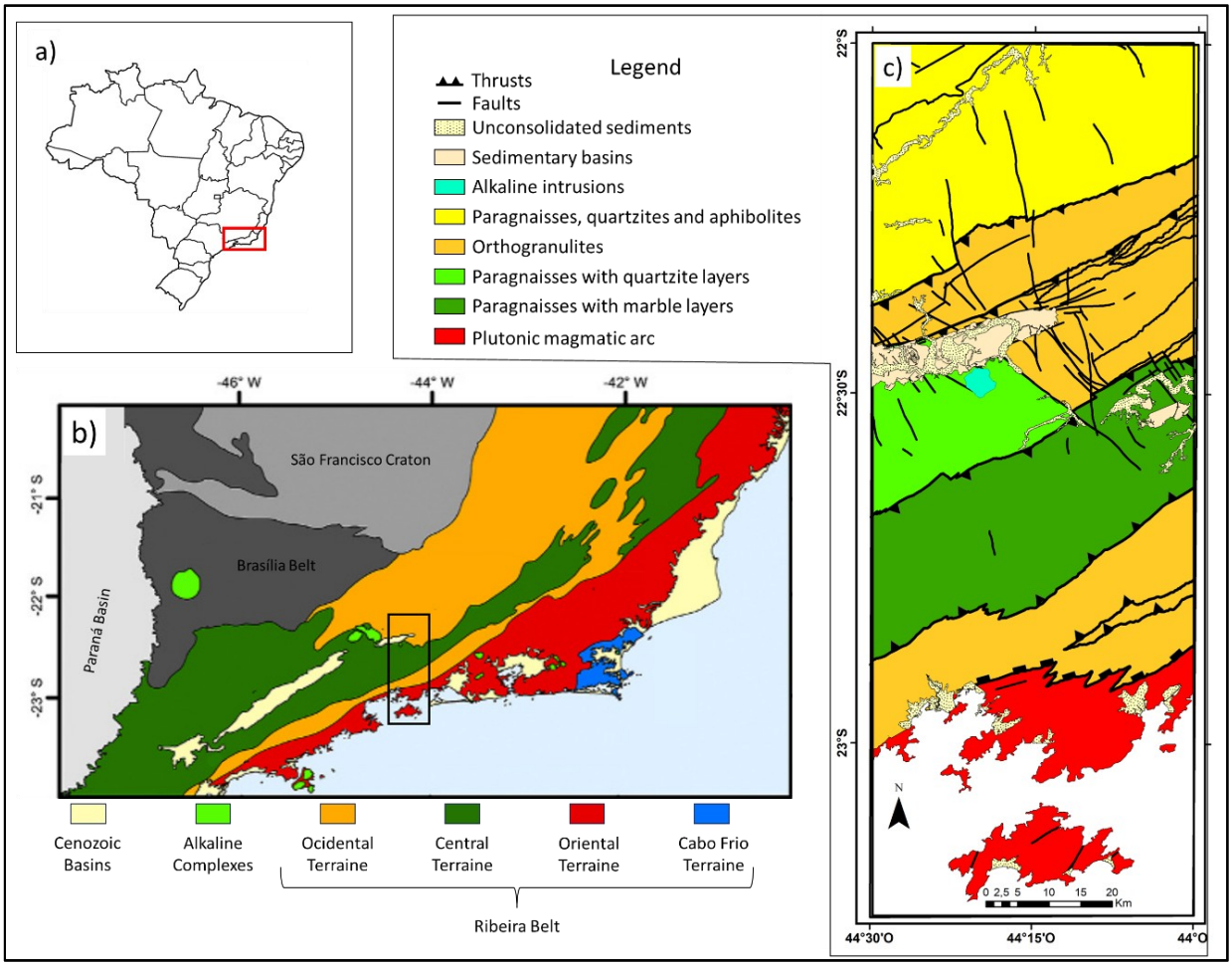


Figure 1. Tectonic configuration of the studied area. **A.** Ribeira Belt location. **B.** Major geological features of southeastern Brazil (modified from Heilbron et al., 2016 and Giro et al., 2021). **C.** main lithological occurrences in Ribeira belt (Modified from Heilbron et al., 2016).

2.3. CNN lineaments extraction methodology

With the objective of predicting topographic lineaments in DEM data we adopted a deep-learning semantic segmentation approach, by training a CNN in a supervised workflow. We chose a U-Net architecture (Ronnenberger et al., 2015) for the network with PyTorch as the development framework, a widely applied method for image segmentation because of its versatility and due to the availability of a large expertly annotated dataset used for training. Our implementation easily allows for variations in architecture parameters such as depth and width of the network. Conventional methods, as the one proposed by Farahbakhsh et al. (2018), use edge detection, adaptive filters and principal components analysis to satisfactorily extract lineaments from DEMs. This approach, however, requires extensive data pre-processing and attribute calculations to highlight the desired features. U-Nets, on the other hand, updates its parameters during training to maximize the association between input and output and finds by its own which relations work best.

The network's parameters were trained not on the studied area, but on a control area, about 50 km east of the studied area (Figure 2), using lineaments manually interpreted and published by Giro et al. (2021), under conditions so similar that criteria variation is neglected. This ensured that the results in the studied area did not simply replicate the training data and show that the CNN is properly generalizing on unseen areas.

2.4. CNN workflow

First the DEM data and mask containing the lineaments were cropped to the same size, normalized between zero and one, and saved as TIFF format. They were then divided into a subset of images called tiles, made of 128 by 128 pixels each, and subsequently loaded into the CNN. In the data loader step a stride is defined for training and validation, and it controls the number of tiles loaded into the CNN (Equation 2). The bigger the stride, the fewer tiles are loaded, and less information is used to train the model, resulting in faster training. For example, a stride equal to 1 would load all possible tiles, approximately 5×10^6, whereas setting a stride to 7 would load approximately 10^5 tiles. Attention is required to load enough tiles for adequate training and avoid overfitting.

$$n = \left[floor\left(\frac{C-T}{s}\right) + 1 \right] \times \left[floor\left(\frac{R-T}{s}\right) + 1 \right]$$

Equation 2. Describes the number of tiles of size (T) that can be cut from an image with (C) columns and (R) rows of pixels and loaded for training and validation, given the stride (s). Floor is the function that rounds down the number.

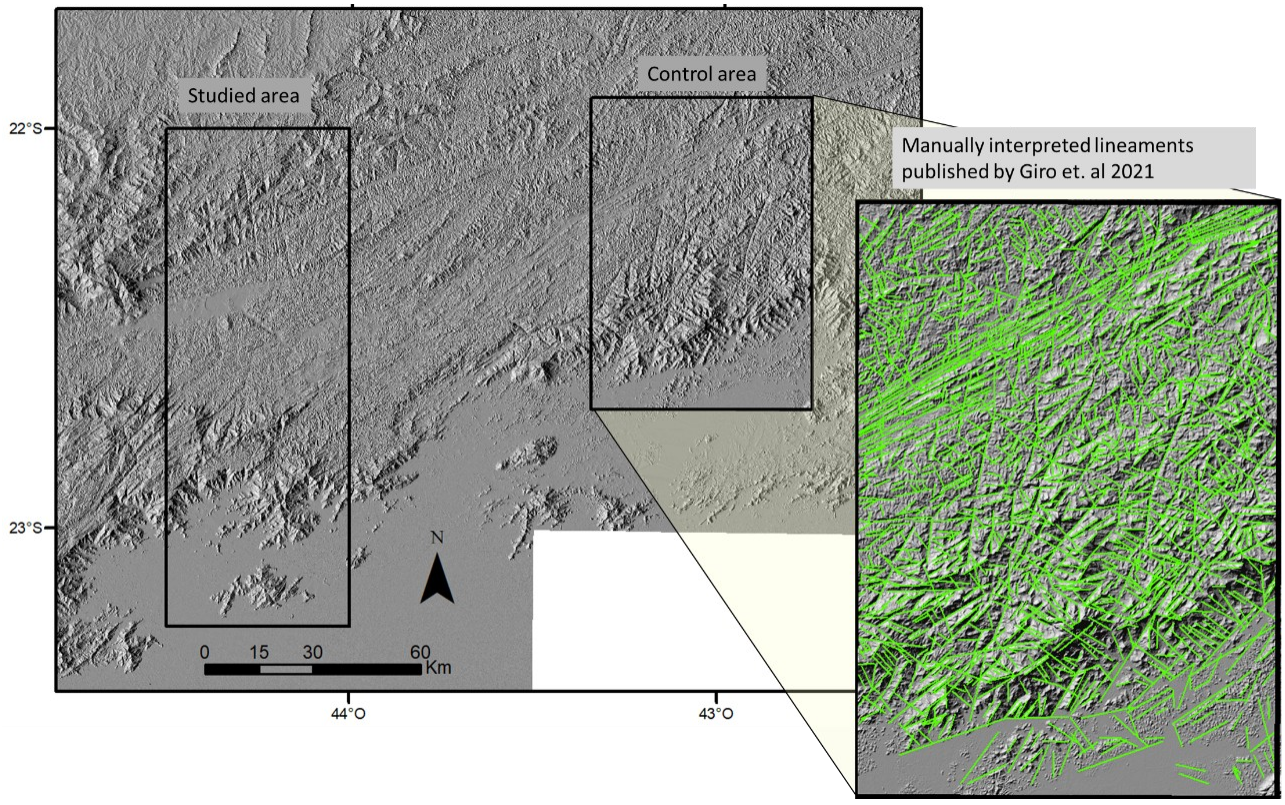


Figure 2. Control area, where 1437 manually interpreted lineaments by Giro et al. (2021) were used to train the CNN, later applied in the studied area. Pseudo illumination azimuth 265°, 45° inclination.

Next, the actual training takes place, and the chosen loss function was the Dice loss function (Equation 3). The Dice loss function is based on the intersection of sets, and it yields robust results in image segmentation when applied to unbalanced datasets, that is, when one of the classes occurs much more frequently than others. In this case, where we aim to classify images into two classes (lineament and non-lineament), the Dice loss function is defined as:

Where:

r = reference value, training data

p= Probability predicted by the neural network

ε = stabilization term that avoids division by zero problems

$$DL_2 = 1 - \frac{\sum_{n=1}^{N} p_n r_n + \epsilon}{\sum_{n=1}^{N} p_n + r_n + \epsilon} - \frac{\sum_{n=1}^{N} (1-p_n)(1-r_n) + \epsilon}{\sum_{n=1}^{N} 2 - p_n - r_n + \epsilon}$$

Equation 3. Dice loss function for two classes. Appropriate for image classification tasks in unbalanced datasets.

The loss function guides the training, according to the hyperparameters, as the model iteratively acquires its parameters in a way to maximize the relationship between the DEM data and the interpreted lineaments. The model output is assessed visually, as it predicts the lineaments in the control area, and by its metrics. The training and validation loss values are expected to be as low as possible and validation accuracy and intersection over union, that measure how much of the classified pixels are correct, are expected to be as high as possible. These four metrics are normalized between zero and one. The time spent during training is expected to be as little as possible. Finally, the approved model receives as input the DEM data from the studied area, processes it, and classifies each pixel as lineament or non-lineament. After that a Hough transform is applied so that only line segments are output as predicted lineaments. Figure 4 shows a schematic workflow and hyperparameters for model 22, considered the best performer. The code is available in a repository at GitHub (link at Section Computer code availability).

2.4.1. Neural network architecture

The technique for lineaments extraction is based on semantic segmentation, which is the ability to segment an unknown image into different parts and objects at a pixel level (Guo et al., 2018). In this case we aim to classify the image's pixels in two classes: lineament or no-lineament. To achieve this, we apply a CNN deep learning technique with a U-net architecture. The U-net architecture for CNN's was developed for image analysis and classification in medical science (Du et al., 2020), but recently it's been used in many different fields, including geoscience. The name refers to the characteristic U-shape, and it consists of two paths. The first is the encoder, or contractive path. This is where the input image is deconstructed, as it gains more "depth" at the expense of its size. In other words, through a series of convolutions and pooling operations the image's representation decreases in height and width but gets more channels (Figure 3). After reaching the innermost segment the representation enters the decoder, or expansive path. Here the data is sampled through a series of transposed convolutions, that gradually restores the data to its original image size. Also, at the decoder path some information from the encoder path is directly transferred in an operation called "skip connection". This information is concatenated to the representation so the CNN can best learn to classify and localize features (Siddique et al., 2021).

The CNN's output is an image where white pixels represent the non-lineament class and black the lineament class. The results are noisy, with small clusters of black pixels scattered across the image (Figure 4), as opposed to the desired aligned disposition. Hence, a Hough transform was applied to the images produced by the CNN, which is a method for isolating a particular shape within an image, in this case, lines. It works by transforming spatially extended patterns to produce features in a parameter space (Hassanein et al., 2015). In practice, it converts the aligned classified pixels output by the CNN into straight line segments, eligible for comparison with the manually interpreted lineaments.

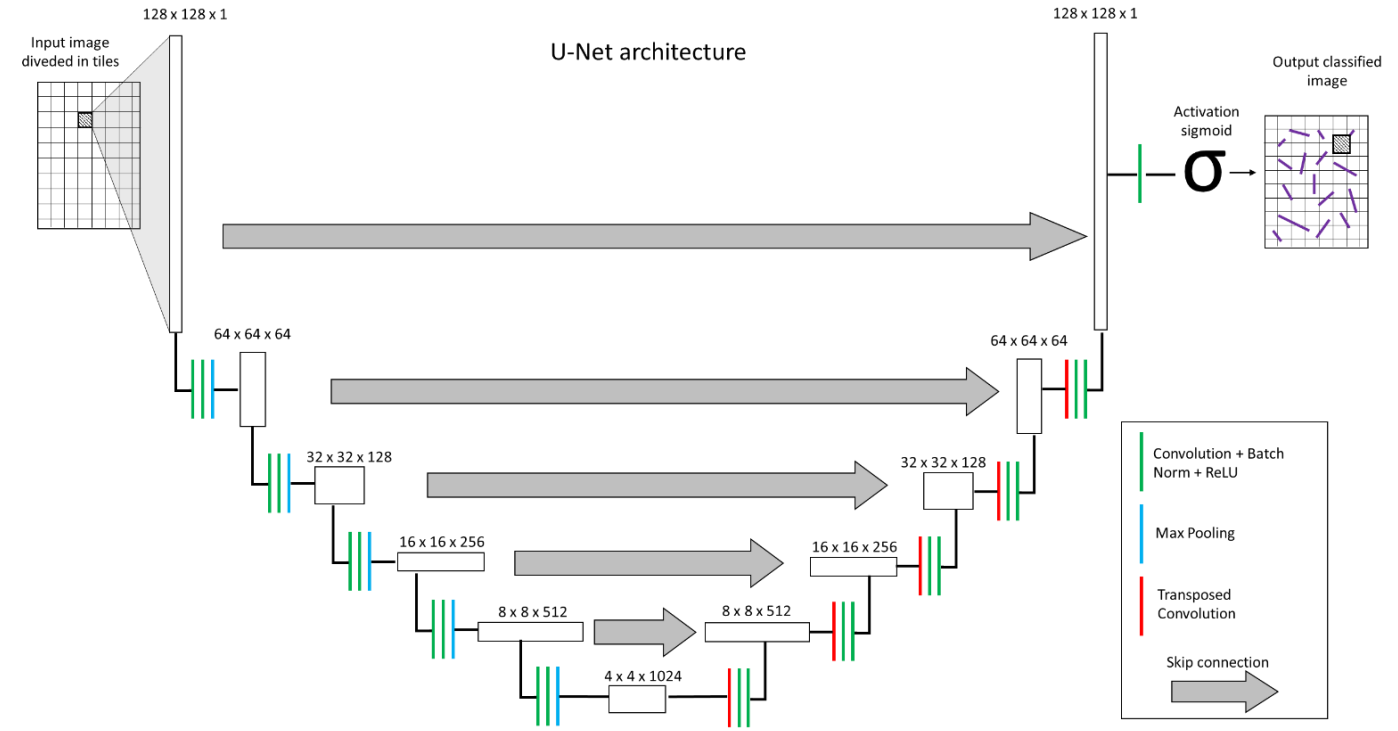


Figure 3. Schematic representation of the U-Net architecture. The input image is divided into tiles, and feed to CNN. In the first half of the way the data is rearranged through a series of convolutions and maxpooling. It loses height and width but gains channels (or depth). In the second half it is upscaled back to its original dimensions and outputs the classified image.

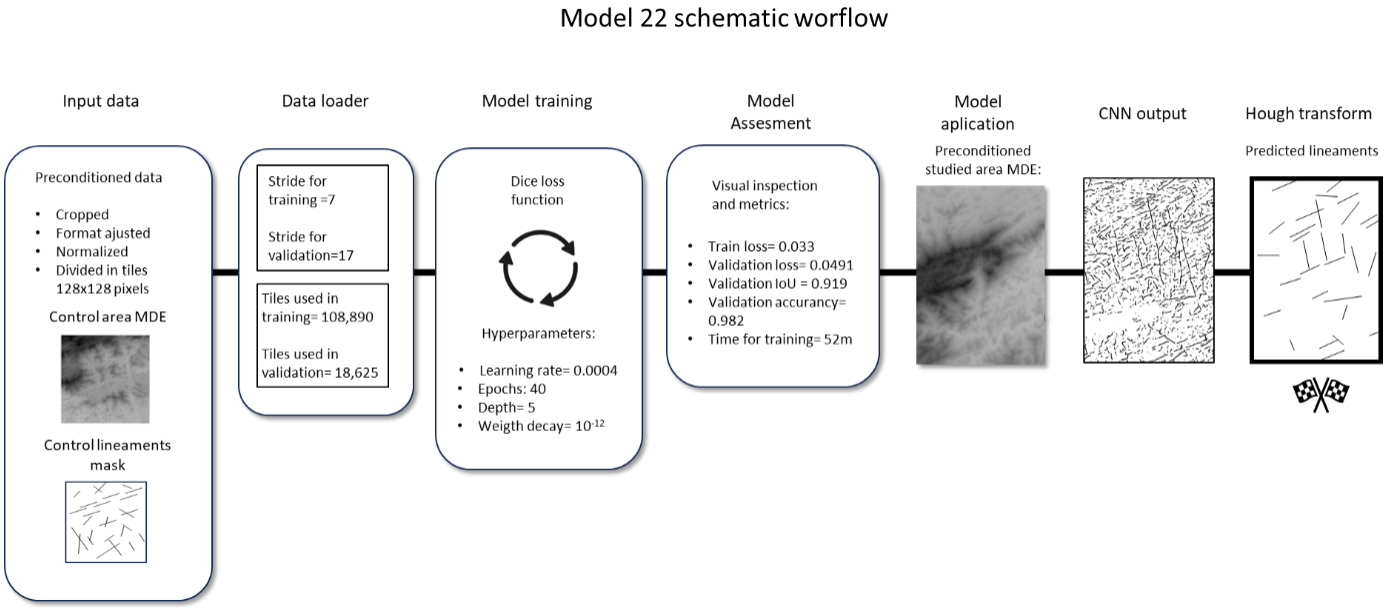


Figure 4. Model 22 schematic workflow and characteristics

2.4.2. Hyperparameters definition

A total of 30 models were tested, over which the CNN hyperparameters were extensively varied and tested, in a trial-and-error fashion that shaped the output to the expected results, a collection of lineaments.

The first models had a training and validation stride of 16 and 31 respectively (~21,000 tiles for training and ~5,600 for validation), and yielded highly varying amplitude validation metrics over training, which was resolved by lowering the learning rate, so that the validation values could work their way down the loss function smoothly. The tradeoff was that it took longer to complete the training, as more epochs were necessary to achieve convergence. Next, the strides decreased to 5 for training and 11 for validation (~213,000 tiles for training

and ~44,000 for validation), in a model that took 9 hours to train. The visual inspection of the result showed no improvement from models that required 1,5 hours to train with strides set to 10 for training and 23 for validation (training and validation strides were never multiples to avoid overfitting). One model trained in 15 seconds with 25 epochs was able to capture the overall lineaments trend. Up to this point CNNs were tested with 3 depth blocks, or 3 maxpooling operations at the encoder path. When a 5 depth blocks architecture was tried, the results showed substantial improvement, with sharper and less noisy lineaments. Models with even more depth blocks did not show better results. Finally, models were trained and validated with subsets of data, restricted in height or width. As digital images are matrices of pixels, we can think of height and width as lines and columns. Model M22, for example, was trained using all the control image lines, but only 1/3 of its columns. It was as if the training image was divided in 3 N-S stripes, and only one was used for training and validation, the west stripe. Its results were considered the best compromise between the time spent on training and results obtained. Figure 4 displays the CNN's schematic workflow with model 22's hyperparameters.

3. RESULTS AND DISCUSSION

Over five rounds of manually interpreting lineaments significant differences are observed. The number of lineaments manually acquired decreased throughout the interpretation sessions (Figure 5). The author states that at the beginning of the project (V1) his moral/willingness was high, and overexcitement might have led to the annotation of second order lineaments as seen in a set of NW lineaments in the black rectangles (Figure 5B). These second order lineaments contrast with the major ones formed by the rock´s foliation, main shear zones and the lineament corridor at the northern portion of the studied area. Over the following rounds of interpretation, when revisiting the data (V2, V3 and V4), the author decided that these lineaments make up a background, and do not differ from the pattern of adjacent features, as defined by O'Leary et al. (1976). These background lineaments exist, but are tightly spaced, and could not be represented in the chosen scales of mapping. That demonstrates that the criteria for choosing lineaments often fluctuate as mapping proceeds. During repeated mapping sessions the author focused more on the main structures, to best represent the lineaments in the studied area. This becomes clear when looking at the charts in Figure 6. As the number of lineaments decreases from the early mapping sessions to the latter, the Jaccard similarity index of the mapping sessions grouped two by two tends to increase among the later versions.

The highest similarity is found between versions 3 and 4 whereas the lowest is the one comparing CNN output Model 30 to the first-year student version -STD (Figure 6). The early versions also contain the largest number of non-repeating lineaments, indicating that they have lower consistency, or make up second order lineaments and probably should not be represented. Over the course of repeating the interpretation, lineaments became fewer, and more robust, as the criteria for recognizing lineaments consolidated.

The higher similarity between a model and a handmade version is between M22, considered to be the best, and version 3, suggesting that although there is a tendency to ramp similarity up towards later versions, version 4 might have been too minimalistic (Figure 6). In the process of selecting the most important lineaments, some relevant ones could have been left out.

The results obtained by the first-year student (STD) showed that even though not familiarized with the geological setting, she was undoubtedly able to capture the overall structures and trends. A major structure, the NEE suture zone at northern area, however, that is known to extend for hundreds of kilometers as a straight corridor (arrow 1), was poorly represented, and the footprint of merging images at the southern area was interpreted as a lineament (arrow 2), as she was unaware of this kind of noise (Figure 5F).

The distribution of lineament's lengths is represented by boxplots in Figure 7, where the box's bottom shows the first quartile, the top shows the third quartile, the midline is the distribution's median and the "x" the mean value. The whiskers are the two lines outside the box and represent the distribution's minimum and maximum.

Manually interpreted lineaments are seen from V1 to V4 and STD. They are slightly positively skewed, meaning that there are more long than short lineaments. The lower whiskers tend to be similar, as lengths were limited to 1250 m, or 0,5 cm at mapping scale. Results obtained by the CNN show behaviors significantly different from humans. For example, computers often represent a slightly curvilinear feature as 3 or 4 small line segments, whereas humans tend to draw a single lineament, or draw closely spaced parallel lineaments, representing associated crests and valleys, while humans tend to choose just one of them. Computers place lineaments where costal line shores are straight, while humans tend to recognize shorelines as features not related to topographic lineaments, although they can coincide. From Figure 7 it's possible to note that handmade lineaments are longer than the ones produced

by the computer (M22 and M30). Humans tend to have a broader view. We can extrapolate our interpretation and connect hill crests or valleys that line up together as a single lineament, even in areas where they are not visible. Computers, on the other hand, tend to annotate lineament where they are strictly apparent, drawing shorter lineaments.

The rose diagrams in Figure 7 show that directions did not change significantly over versions and kept mean direction around 32°.

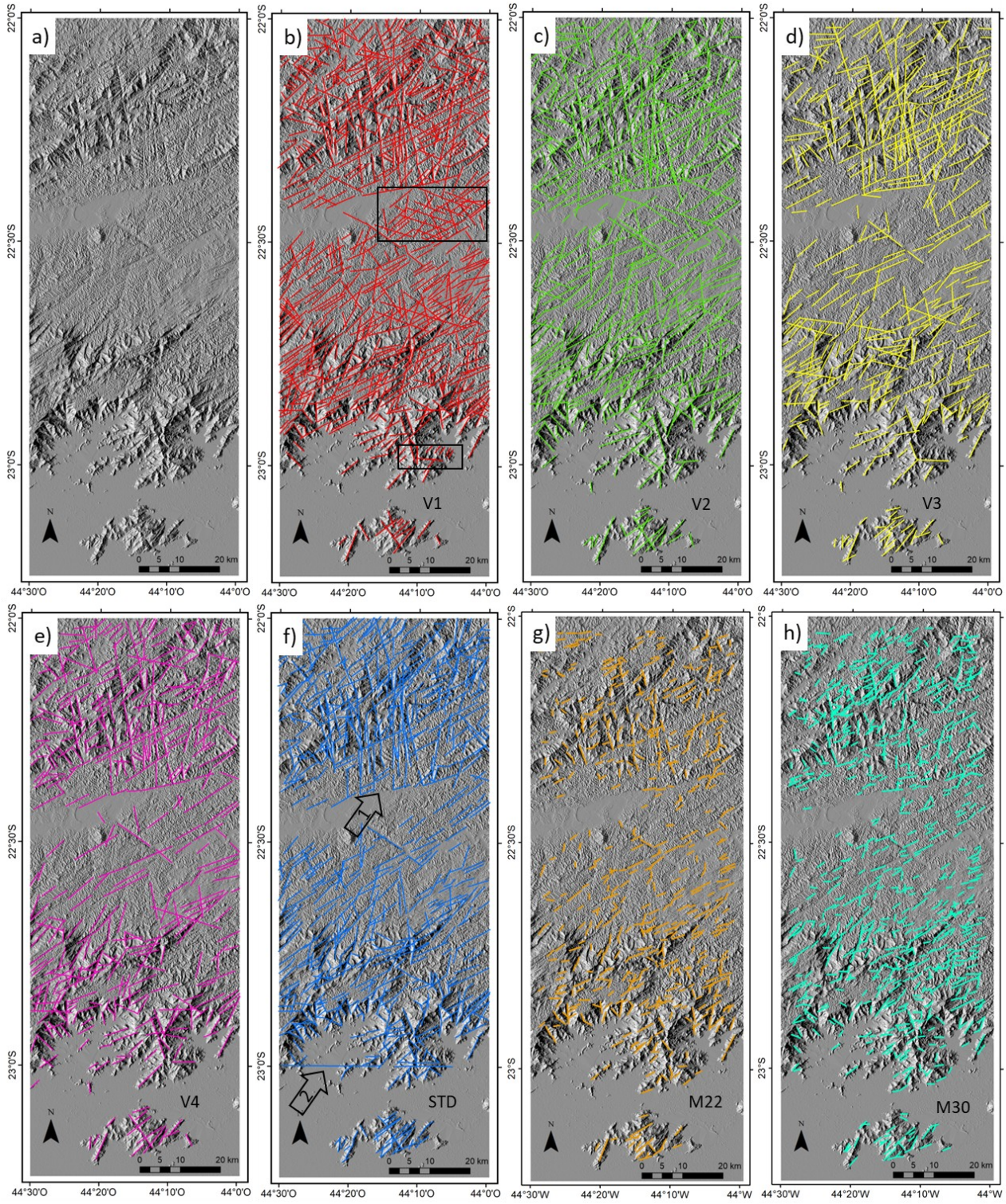


Figure 5. Human extraction of topographic lineaments. **A.** DEM, pseudo illumination azimuth 265, elevation 45°. **B.** First round of interpretation – 726 lineaments. Black rectangles show areas where second order lineaments were interpreted. **C.** Second round of interpretation – 493 lineaments. **D.** Third round of interpretation – 410 lineaments. **E.** Fourth round of interpretation – 393 lineaments. **F.** First-year undergrad student interpretation – 744 lineaments. Arrow 1 indicates a poorly represented major lineament, arrow 2 indicates a lineament drawn over an image merge footprint. **G.** Model 22's output – 782 lineaments. **H.** Model 30's output 552 - lineaments.

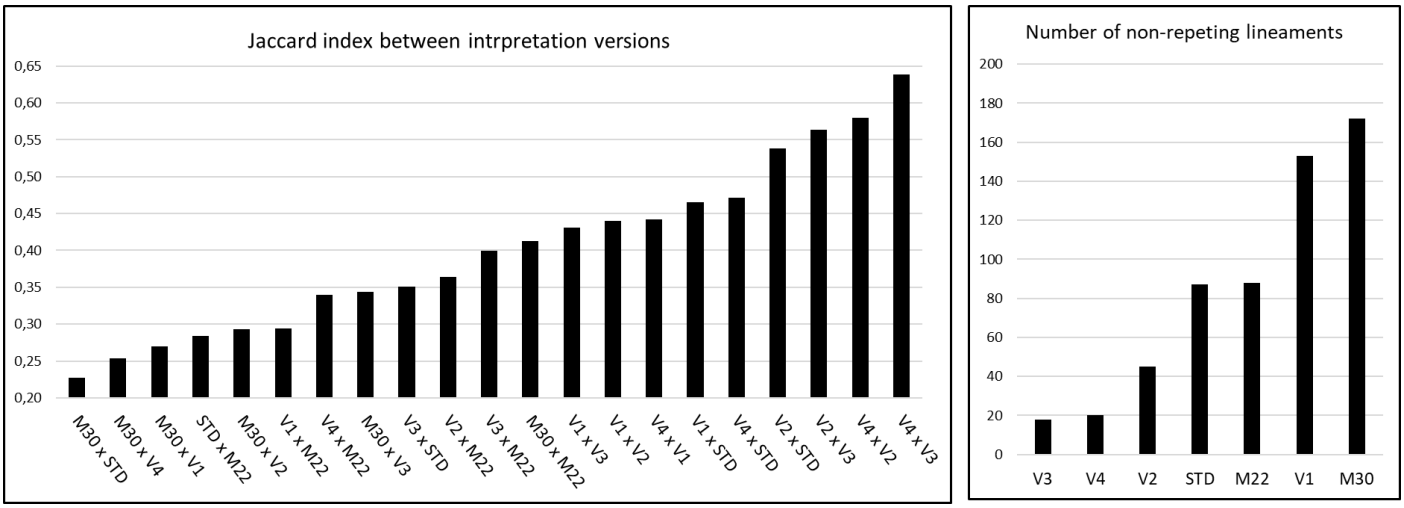


Figure 6. V1 to V4 represents the versions of mapping done by the author, STD represents the one from the first-year undergraduate student, M22 and M30 two CNN outputs. **Left:** Jaccard index of similarity between interpretation sessions. **Right:** Percentage of lineaments that appear in only once, indicating non-consistent interpretations.

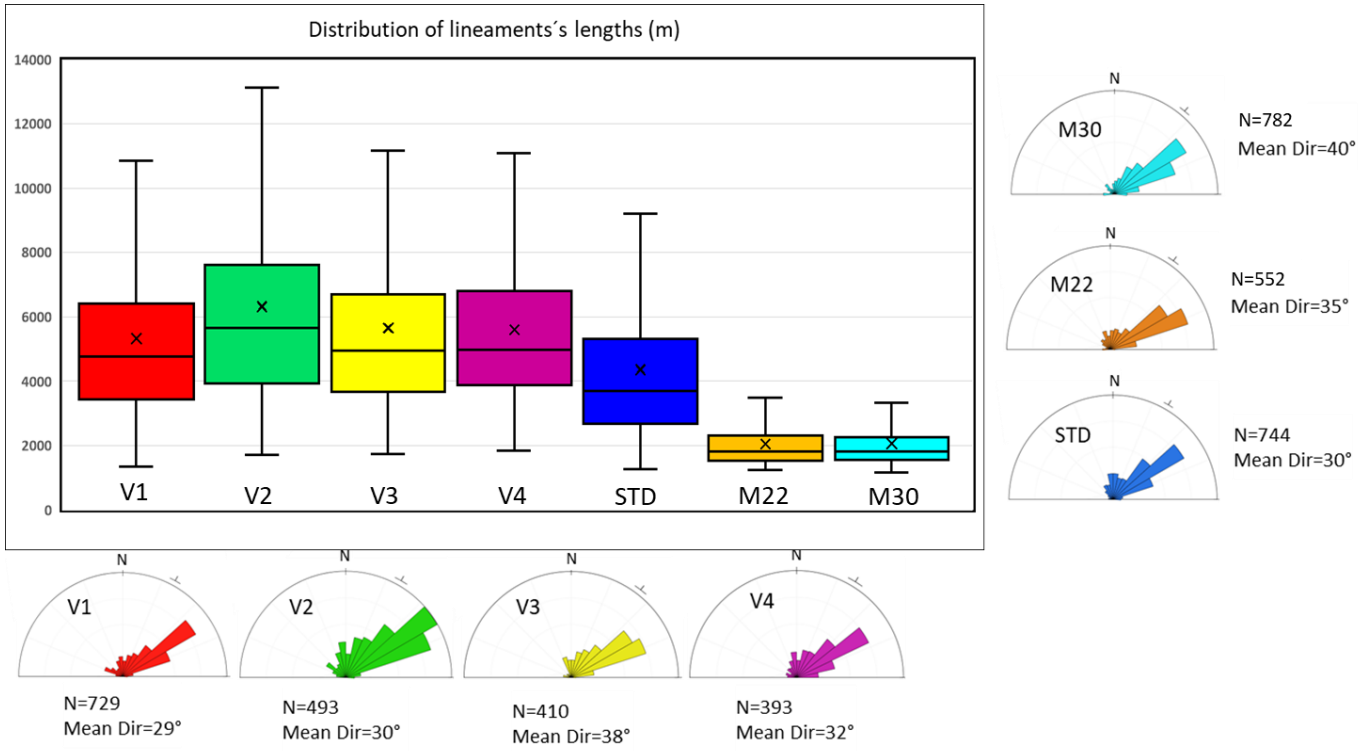


Figure 7. Boxplots representing the distribution of lineament lengths and rose diagrams showing directions.

4. CONCLUSIONS

In this paper we compare the results of lineaments extracted from DEM data by humans and by computers using a CNN with U-Net architecture. Interpretation, although essential to data incorporation, is subjective, and there is no ground truth from which our results can be compared. What we did was to analyze the various versions of lineaments extractions grouped two by two and assess their repeatability as a proxy for consistency.

Results from manual extraction show that the criteria for identifying and drawing lineaments changed over repeated interpretations session. Early versions tend to capture more lineaments that often are not repeated in following sessions, but as the studied area becomes familiar to the interpreter, the criteria for selecting lineaments improve and second order lineaments were left out of later versions. The best similarity between computer models and handmade interpretations was M22 and version 3, suggesting that version 4, the last one, could have become too minimalistic.

A CNN was written, trained, and then applied to the studied area. Results show that once hyperparameters have been determined and the CNN satisfactorily trained,

it can be a powerful tool to reduce the time and subjectivity involved in interpreting lineaments in DEM data. However, the trial-and-error process required to find satisfactory hyperparameters and post-processing applied to improve results can take a lot of time and can even neutralize the gains obtained by applying CNN technology. Moreover, U-Net's complex architecture has, in general, high computational cost and can underperform when the training data is unbalanced.

Humans are immensely versatile. We can rapidly adapt to perform complex tasks using our common sense. Computers, on the other hand, have no common sense, and even though they can "learn" to predict lineaments each data set and hyperparameters, they will probably yield poor results in different areas. Therefore, it is justifiable to build and apply CNNs to predict lineaments over large datasets with similar characteristics in such that it is feasible to label a training dataset that is representative of the whole data. Otherwise, it is far more practical to manually extract lineaments with two or more interpretation versions performed by one or more people, aiming at the reduction of the inevitable subjectivity.

ACKNOWLEDGMENTS

We would like to thank Rayane Pereira Guedes for creating the STD interpretation version, João Paulo Giro for providing the lineaments used for training the CNN, and Petrobras for the support during my master's studies. Also, thanks to visiblegeology.com for providing an excellent on-line free service for building stereograms and rose diagrams.

COMPUTER CODE AVAILABILITY

Name of the code/library:
Lineament_DEM_extractor / Pytorch
Contact: Antônio Pires:
antonio.pires@petrobras.com.br; +5521 99783-9917
Hardware requirements: Intel Core i5 or i7 processor, Full HD resolution, ideally 1920×1080 and 8GB of RAM
Program language: Python
Software required: Python 3.9.13 and Pytorch 2.0
Program size: 4 KB
Year first available: 2024
Code available at github open repository:
https://github.com/PedroRossi99/Lineament_DEM_extractor.git

REFERENCES

Akinluyi, F.O., Olorunfemi, M.O., and Bayowa, O.G., 2018, Investigation of the influence of lineaments, lineament intersections and geology on groundwater yield in the basement complex terrain of Ondo State, Southwestern Nigeria: Applied Water Science, v. 8, n. 49, doi: https://doi.org/10.1007/s13201-018-0686-x.

Almeida F.F.M., Hasui Y., Brito Neves, B.B., and Fuck, R.A., 1977, As províncias estruturais do Brasil, *in* Simpósio de Geologia do Nordeste, 8th, Bol. Esp., 12p.

Almeida, F.F.M., 1967, Origem e evolução da plataforma brasileira: DNPM: Div. Geol. Min: Boletim 241, 36 p.

Almeida, F.F.M., 1976, The system of continental rifts bordering the Santos Basin, Brazil: Anais da Academia Brasileira de Ciências, n. 48, p.15-26.

Almeida, F.F.M.; Hasui, Y., Brito Neves, B.B., and Fuck, R.A., 1981, Brazilian structural provinces: An introduction: Earth-Science Reviews, v. 17, n. 1-2, p. 1-29, doi: https://doi.org/10.1016/0012-8252(81)90003-9.

Almeida, J., Dios, F., Mohriak, W.U., Valeriano, C.D.M., Heilbron, M., Eirado, L.G., and Tomazzoli, E., 2013, Pre-rift tectonic scenario of the eo-cretaceous Gondwana break-up along SE Brazil-SW Africa: insights from tholeiitic mafic dyke swarms: Geological Society, London, Special Publications, v. 369, p. 11-40, doi: https://doi.org/10.1144/SP369.24.

An, Y., Guo, J,, Ye, Q., Childs, C., Walsh, J., and Dong, R., 2021, Deep convolutional neural network for automatic fault recognition from 3D seismic datasets: Computers & Geosciences, v. 153, doi: https://doi.org/10.1016/j.cageo.2021.104776.

Ayalew, L., and Yamagishi, H., 2005, The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan: Geomorphology, v. 65, n. 1-2, p. 15-31, doi: https://doi.org/10.1016/j.geomorph.2004.06.010.

Bashilov, V.I., 1982, Lineaments and petroleum occurrences in the north of the East European Platform: International Geology Review, v. 24, n. 6, p. 715-718, doi: https://doi.org/10.1080/00206818209451013.

Baskin, C., Liss, N., Mendelson, A., and Zheltonozhskii, E, 2017, Streaming Architecture for Large-Scale Quantized Neural Networks on an FPGA-Based Dataflow Platform, in IEEE International Parallel and Distributed Processing Symposium Workshops, Canada, doi: https://doi.org/10.1109/IPDPSW.2018.00032.

Brito Neves, B.B.D., Campos Neto, M.D.C., and Fuck, R.A., 1999, From Rodinia to Western Gondwana: an approach to the Brasiliano-Pan African Cycle and orogenic collage: *Episodes*, v. *22,* n. 3, p. 155-166, doi: https://doi.org/10.18814/epiiugs/1999/v22i3/002.

Cardoso-Fernandes, J., Teodoro, A., Lima, A., Menuge, J.,Brönner, M., and Steiner, R., 2022, Sentinel-1 and ALOS data for lineament extraction: a comparative study, *in* SPIE Remote Sensing, doi: https://doi.org/10.1117/12.2636117.

Chandra, S., Rao, V.A., Krishnamurthy, N.S., Dutta, S., and Ahmed, S., 2006, Integrated studies for characterization of lineaments used to locate groundwater potential zones in a hard rock region of Karnataka, India: Hydrogeology Journal, n. 14, p. 1042-1051, doi: https://doi.org/10.1007/s10040-006-0097-1.

Clark Jr., S.F., Moore, R.B., Ferguson, E.W., and Picard, M.Z., 1996, Criteria and Methods for Fracture-Trace Analysis of the New Hampshire Bedrock Aquifer: U.S. Geological Survey Open-File Report, n. 96–479, 12 p., doi: https://doi.org/10.3133/ofr96479.

Cogné, N., Cobbold, P., Riccomini, C., and Gallagher, K., 2013, Tectonic setting of the Taubaté Basin (Southeastern Brazil): Insights from regional seismic profiles and outcrop data: Journal of South American Earth Sciences, v. 42, p. 94-204, doi: https://doi.org/10.1016/j.jsames.2012.09.011.

Daynac, J., Bessin, P., Pochat, S., Mourgues, R., and Shumack, S. 2024, A new workflow for mapping dune features (outline, crestline and defects) combining deep learning and skeletonization from DEM-derived data: Geomorphology, v. 463, doi: https://doi.org/10.1016/j.geomorph.2024.109369.

Du, G., Cao, X., Liang, J., Chen, X., and Zhan, Y., 2020, Medical Image Segmentation based on U-Net: A Review: Journal of Imaging Science and Technology, v. 64, p 020508-1 - 020508-12, doi: https://doi.org/10.2352/J.ImagingSci.Technol.2020.64.2.020508.

Enoh, M.A., Okeke, F.I., and Okeke, U.C. 2021, Automatic lineaments mapping and extraction in relationship to natural hydrocarbon seepage in Ugwueme, South-Eastern Nigeria: Geodesy and Cartography, v. 47, n. 1, p. 34-44, doi: https://doi.org/10.3846/gac.2021.12099.

Farahbakhsh, E., Chandra, R., Olierook, H.K.H., Scalzo, R., Clark, C., Reddy, S.M., and Müller, R.D., 2019, Computer vision-based framework for extracting tectonic lineaments from optical remote sensing data: International Journal of Remote Sensing, v. 41, p. 1760–1787, doi: https://doi.org/10.1080/01431161.2019.1674462.

Fletcher, S., and Islam, M.Z., 2018, Comparing sets of patterns with the Jaccard index: Australasian Journal of Information Systems, v. 22, doi: https://doi.org/10.3127/ajis.v22i0.1538.

Freitas, N.C., Almeida, J., Heilbron, M., Cutts, K. and Dussin, I., 2021, The Cabo Frio Thrust: a folded suture zone, Ribeira belt, SE Brazil: Journal of Structural Geology, v. 149, doi: https://doi.org/10.1016/j.jsg.2021.104379.

Froner B., Purves S.J., Lowell J., and Henderson J., 2013, Perception of visual information: the role of colour in seismic interpretation: First Break, v. 31, n. 4, doi: https://doi.org/10.3997/1365-2397.2013010.

Giro J.P., Almeida J., Guedes E., and Bruno H., 2021, Tectonic inheritances in rifts: The meaning of NNE lineaments in the continental rift of SE-Brazil: Journal of South American Earth Sciences, v. 108, doi: https://doi.org/10.1016/j.jsames.2021.103225.

Goldstein, A. and Marshak, S. 1988, Analysis of fracture array geometry (Ch. 12), *in* Marshak, S. and Mitra, G., eds., Basic Methods of Structural Geology: Prentice Hall, p. 249-268.

Guedes, E., Heilbron, M., Vasconcelos, P.M., Valeriano, C.M., Almeida, J. C. H., Teixeira, W. and Thomaz Filho, A. 2005, K-Ar and 40Ar/39Ar ages of dikes emplaced in the onshore basement of the Santos Basin (Resende Area, SE-Brazil): implications for South Atlantic opening and tertiary reactivation: South American Journal of Earth Sciences, v. 18, n. 3-4, p. 371-382, doi: https://doi.org/10.1016/j.jsames.2004.11.008.

Guedes E., Heilbron M., Valeriano C.M., Almeida J. C. H., and Szatmari P., 2016, Evidence of Gondwana early rifting process recorded by Resende-Ilha Grande Dike Swarm, southern Rio de Janeiro, Brazil: Journal of South American Earth Sciences, v. 67, p. 11-24, doi: https://doi.org/10.1016/j.jsames.2016.01.004.

Guo, Y., Liu, Y., Georgiou, T., and Lew, M.S., 2018, A review of semantic segmentation using deep neural networks: Int Journal Multimed Info Retr, v. 7, p. 87-93, doi: https://doi.org/10.1007/s13735-017-0141z.

Haslebacher, C., Thomas, N., and Bickel, V.T., 2023, LineaMapper: A deep learning-powered tool for mapping linear surface features on Europa: Icarus, v. 410, doi: https://doi.org/10.1016/j.icarus.2023.115722.

Hassanein, A., S., Mohammad, S., Sameer, M., and Ragab, M.E., 2015, A Survey on Hough Transform, Theory, Techniques and Applications: International Journal of Computer Science Issues, v. 12, n. 1, doi: https://doi.org/10.48550/arXiv.1502.02160.

Hasui, Y., Carneiro, C.D.R., and Coimbra, A.M., 1975, The Ribeira folded belt: Revista Brasileira de Geociências, v. 5, n. 4, p. 257-266, doi: https://doi.org/10.25249/0375-7536.1975257266.

Heilbron M., Eirado L.G., and Almeida J.C.H., 2016, Mapa geológico e de recursos minerais do estado do Rio de Janeiro. CPRM: Rio de Janeiro.

Heilbron M., Mohriak, W., Valeriano C.M., Milani E., Almeida J.C.H., and Tupinambá, M., 2000, From collision to extension: the roots of the southeastern continental margin of Brazil, *in* Talwani and Mohriak., eds., Atlantic Rifts and Continental Margins: America Geophysical Union: Geophysical Monograph Series, v. 115, p. 1-34.

Heilbron, M., Silva, L.G.E., Almeida, J.C.H., et al., 2020, Proterozoic to Ordovician geology and tectonic evolution of Rio de Janeiro State, SE-Brazil: insights on the central Ribeira Orogen from the new 1:400,000 scale geologic map: Brazilian Journal of Geology, v. 50, p. 1-25, doi: https://doi.org/10.1590/2317-4889202020190099.

Heilbron, M., Valeriano, C.M. Valladares, C.S., and Machado, N., 1995, A Orogênese Brasiliana no segmento central da Faixa Ribeira, Brasil: Revista Brasileira de Geociências, v. 25, n. 4, p. 249-266, doi: https://doi.org/10.25249/0375-7536.1996249266.

Hobbs, W.H., 1904, Lineaments of the Atlantic Border region: GSA Bulletin, v. 15, n.1, p. 483-506, doi: https://doi.org/10.1130/GSAB-15-483.

Instituto Nacional de Pesquisas Espaciais, Brasil, 2008, Topodata: Banco de Dados Geomorfométricos do Brasil: http://www.dsr.inpe.br/topodata (accessed Agosto 2025).

Jordan, G., Meijninger, B.M.L., van Hinsbergen, D.J.J., Meulenkamp, J.E., and van Dijk, P.M., 2005, Extraction of morphotectonic features from DEMs: Development and applications for study areas in Hungary and NW Greece: International Journal of Applied Earth Observation and Geoinformation, v. 7, n. 3, p. 163-182, https://doi.org/10.1016/j.jag.2005.03.003.

Liu, C.C., Sousa, M.A., and Gopinath, T.R., 2000, Regional Structural Analysis by Remote Sensing for Mineral Exploration, Paraiba State, Northeast Brazil: Geocarto International, v. 15, n. 1, p. 70-77, doi: https://doi.org/10.1080/10106040008542142.

Lulseged A., and Hiromitsu Y., 2005, The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan: Geomorphology, v. 65, n. 1-2, p. 15-31, doi: https://doi.org/10.1016/j.geomorph.2004.06.010.

Mabee, S.B., Kenneth C.H., Donald U., and Wise, D.U., 1994, A method of collecting and analyzing lineaments for regional-scale fractured-bedrock aquifer studies: Groundwater, v. 32, n. 6, p. 884-894, doi: https://doi.org/10.1111/j.1745-6584.1994.tb00928.x.

McCulloch, W.S., and Pitts, W., 1943, A logical calculus of the ideas immanent in nervous activity: Bulletin of Mathematical Biophysics, v. 5, p. 115-133, doi: https://doi.org/10.1007/BF02478259.

Meshkani S.A., Mehrabi B., Yaghubpur A., and Sadeghi M., 2013, Recognition of the regional lineaments of Iran: Using geospatial data and their implications for exploration of metallic ore deposits: Ore Geology Reviews, v. 55, p. 48-63, doi: https://doi.org/10.1016/j.oregeorev.2013.04.007.

Negrão A.P., Mello C.L., Ramos R.R.C., Sanson M.S.R., Louro V.H.A., and Bauli P.G., 2020, Tectonosedimentary evolution of the Resende and Volta Redonda basins (Cenozoic, Central Segment of the Continental Rift of Southeastern Brazil): Journal of South American Earth Sciences, v. 104, doi: https://doi.org/10.1016/j.jsames.2020.102789.

O'Leary, D.W., Friedman J.D., and Pohn, H.A., 1976, Lineament, linear, lineation: some proposed new standards for old terms: GSA Bulletin, v. 87, n. 10, p. 1463-1469, doi: https://doi.org/10.1130/0016-7606(1976)87%3C1463:LLLSPN%3E2.0.CO;2.

Peacock, D.C., Sanderson, D.J., Bastesen, E., Rotevatn, A. and Storstein, T.H., 2019, Causes of bias and uncertainty in fracture network analysis: Norwegian Journal of Geology, v. 99, n. 1, p. 113-128, doi: https://doi.org/10.17850/NJG99-1-06.

Ramli, M.F., Yusof, N., Yusoff, M.K., Juahir, H., and Shafri H.Z.M., 2010, Lineament mapping and its application in landslide hazard assessment: a review: Bull Eng Geol Environ, v. 69, p. 215-233, doi: https://doi.org/10.1007/s10064-009-0255-5.

Riccomini, C., Sant'Anna, L.G., and Ferrari, A.L., 2004, Evolução geológica do rift continental do sudeste do Brasil, *in* Mantesso-Neto, V., Bartorelli, A., Dal Ré Carneiro, C., Brito Neves, orgs., B.B., eds., Geologia do continente sul-americano – evolução da obra de Fernando Flávio Marques de Almeida. São Paulo, Brasil: Beca.

Ronneberger, O., Fischer, P., and Brox, T., 2015, U-Net: Convolutional Networks for Biomedical Image Segmentation, in Navab, N., Hornegger, J., Wells, W., Frangi, A., eds., Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: Lecture Notes in Computer Science, v. 9351, doi: https://doi.org/10.1007/978-3-319-24574-4_28.

Saha, S.K., 2022, Remote Sensing and Geographic Information System Applications in Hydrocarbon Exploration: a Review: Journal Indian Soc Remote Sens, v. 50, p. 1457-1475, doi: https://doi.org/10.1007/s12524-022-01540-9.

Sander, P., 2007, Lineaments in groundwater exploration: a review of applications and limitations: Hydrogeology Journal, v. 15, p. 71-74, doi: https://doi.org/10.1007/s10040-006-0138-9.

Schmitt R.S., Trouw R.A.J., Schmus W.R.V.S., and Pimentel M.M., 2004, Late amalgamation in the central part of West Gondwana: new geochronological data and the characterization of a Cambrian collisional orogeny in the Ribeira Belt (SE Brazil): Precambrian Research, v. 133, n. 1-2, p. 29-61, doi: https://doi.org/10.1016/j.precamres.2004.03.010.

Siddique, N., Sidike, P., Elkin, C., and Devabhaktuni, V. 2021, U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications: IEEE Access, v. 9, p. 82031-82057, doi: https://doi.org/10.1109/ACCESS.2021.3086020.

Trouw, R., Heilbron, M., Ribeiro, A., Paciullo, F., Valeriano, C.; Almeida, J.C.H., Tupinambá, M., and Andreis, R., 2000, The Central Segment of The Ribeira Belt, *in* Cordani, U.G., Milani, E.J., Thomás Filho, A. and Campos, D.A., eds., Tectonic Evolution of South America, International Geological Congress, 31ed., Rio de Janeiro: Academia Brasileira de Ciências, p. 287-310.

Yusof, N., Ramli, M.F., Pirasteh S., and Shafri, H.Z.M., 2011, Landslides and lineament mapping along the Simpang Pulai to Kg Raja highway, Malaysia: International Journal of Remote Sensing, v. 32, n. 14, p. 4089-4105, dois: https://doi.org/10.1080/01431161.2010.484434.

Zhang, Y., Gang W., Mingchao L., and Shuai H., 2018, Automated Classification Analysis of Geological Structures Based on Images Data and Deep Learning Model: Applied Sciences, v. 8, n. 12, p. 2493, doi: https://doi.org/10.3390/app8122493.

AUTHORSHIP CONTRIBUTION STATEMENT

Pedro Rossi: Conceptualization, data curation, data interpretation, writing – original draft. Julio Almeida: Formal analysis, writing – review and editing. Antônio Pires: Software development, validation, writing- review and editing. Mário Neto: Writing review and editing.

FUNDING SOURCES

This research did not receive any funding from specific public agencies, commercial or non-profit organizations.