



# Automatic fault mapping in remote optical images and topographic data with deep learning

Lionel Mattéo, Isabelle Manighetti, Yuliya Tarabalka, Jean-michel Gaucel, Martijn van den Ende, Antoine Mercier, Onur Tasar, Nicolas Girard, Frédérique Leclerc, Tiziano Giampetro, et al.

## ► To cite this version:

Lionel Mattéo, Isabelle Manighetti, Yuliya Tarabalka, Jean-michel Gaucel, Martijn van den Ende, et al.. Automatic fault mapping in remote optical images and topographic data with deep learning. Journal of Geophysical Research: Solid Earth, 2021, 10.1029/2020JB021269 . hal-03191101

**HAL Id: hal-03191101**

**<https://inria.hal.science/hal-03191101>**

Submitted on 11 May 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# JGR Solid Earth

## RESEARCH ARTICLE

10.1029/2020JB021269

### Special Section:

Machine Learning for Solid Earth Observation, Modeling, and Understanding

### Key Points:

- We adapt a U-Net Convolutional Neural Network to automate fracture and fault mapping in optical images and topographic data
- We provide a trained model  $M_{\text{Ref}}$  able to identify and map fractures and faults accurately in image data of various types and resolutions
- We use  $M_{\text{Ref}}$  to analyze fault organization, patterns, densities, orientations and lengths in six fault sites in western USA

### Supporting Information:

Supporting Information may be found in the online version of this article.

### Correspondence to:

L. Mattéo,  
[matteo@geoazur.unice.fr](mailto:matteo@geoazur.unice.fr)

### Citation:








Mattéo, L., Manighetti, I., Tarabalka, Y., Gaucel, J.-M., van den Ende, M., Mercier, A., et al. (2021). Automatic fault mapping in remote optical images and topographic data with deep learning. *Journal of Geophysical Research: Solid Earth*, 126, e2020JB021269. <https://doi.org/10.1029/2020JB021269>

Received 28 OCT 2020  
Accepted 25 MAR 2021

© 2021. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs License](#), which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

## Automatic Fault Mapping in Remote Optical Images and Topographic Data With Deep Learning

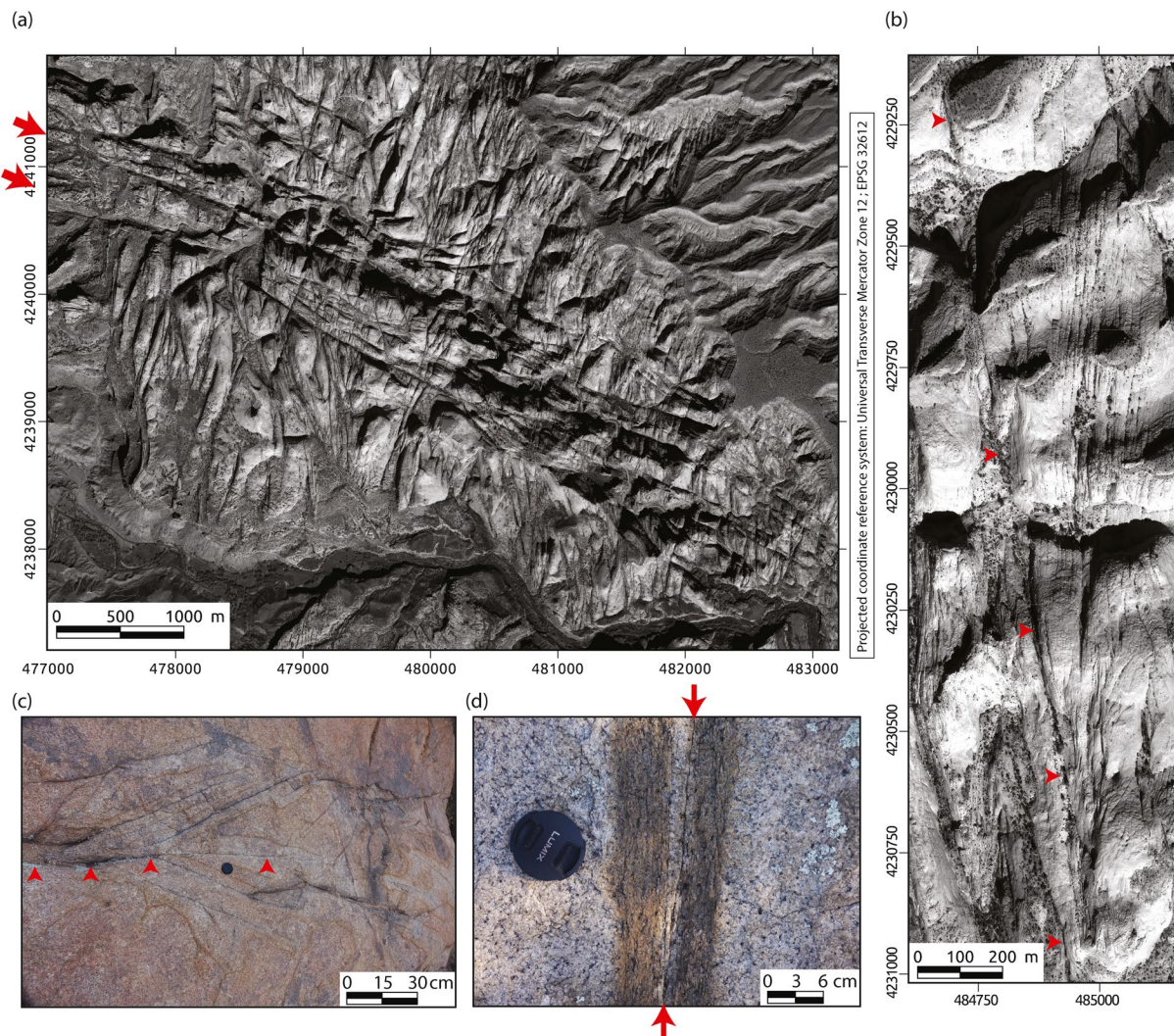
Lionel Mattéo<sup>1</sup> , Isabelle Manighetti<sup>1</sup> , Yuliya Tarabalka<sup>2</sup>, Jean-Michel Gaucel<sup>3</sup>, Martijn van den Ende<sup>1</sup> , Antoine Mercier<sup>1</sup> , Onur Tasar<sup>4</sup>, Nicolas Girard<sup>4</sup>, Frédérique Leclerc<sup>1</sup> , Tiziano Giampetro<sup>1</sup>, Stéphane Dominguez<sup>5</sup> , and Jacques Malavieille<sup>5</sup> 

<sup>1</sup>Université Côte d'Azur, Observatoire de la Côte d'Azur, IRD, CNRS, Géoazur, Sophia Antipolis, France, <sup>2</sup>LuxCarta, Sophia Antipolis, France, <sup>3</sup>Thales Alenia Space, Cannes, France, <sup>4</sup>Université Côte d'Azur, Inria, Sophia Antipolis, France, <sup>5</sup>Université de Montpellier, Géosciences Montpellier, Montpellier, France

**Abstract** Faults form dense, complex multi-scale networks generally featuring a master fault and myriads of smaller-scale faults and fractures off its trace, often referred to as damage. Quantification of the architecture of these complex networks is critical to understanding fault and earthquake mechanics. Commonly, faults are mapped manually in the field or from optical images and topographic data through the recognition of the specific curvilinear traces they form at the ground surface. However, manual mapping is time-consuming, which limits our capacity to produce complete representations and measurements of the fault networks. To overcome this problem, we have adopted a machine learning approach, namely a U-Net Convolutional Neural Network (CNN), to automate the identification and mapping of fractures and faults in optical images and topographic data. Intentionally, we trained the CNN with a moderate amount of manually created fracture and fault maps of low resolution and basic quality, extracted from one type of optical images (standard camera photographs of the ground surface). Based on a number of performance tests, we select the best performing model,  $M_{\text{Ref}}$ , and demonstrate its capacity to predict fractures and faults accurately in image data of various types and resolutions (ground photographs, drone and satellite images and topographic data).  $M_{\text{Ref}}$  exhibits good generalization capacities, making it a viable tool for fast and accurate mapping of fracture and fault networks in image and topographic data. The  $M_{\text{Ref}}$  model can thus be used to analyze fault organization, geometry, and statistics at various scales, key information to understand fault and earthquake mechanics.

## 1. Introduction

Fractures and faults are widespread in the Earth's crust, and associated with telluric hazards, including tectonic earthquakes, induced seismicity, landslides, rock reservoir fracturing and seepage, among others (e.g., Scholz, 2019). While fractures are generally small-scale, shallow, planar cracks in rocks (Barton & Zoback, 1992; Bonnet et al., 2001; Segall & Pollard, 1983), faults span a broad range of length scales ( $10^{-6}$ – $10^3$  km) and surface-to-depth widths ( $1$ – $10^2$  km), and have a complex 3D architecture (e.g., Taponnier & Molnar, 1977; Wesnousky, 1988). At all scales faults form dense networks (“fault zones”) including a master fault and myriads of secondary fractures and faults that intensely dissect the host rock embedding the master fault, sometimes up to large distances away from the master fault (Figure 1) (Chester & Logan, 1986; Granier, 1985; Manighetti, King, & Sammis, 2004; Mitchell & Faulkner, 2009; Perrin, Manighetti, & Gaude-mer, 2016; Smith et al., 2011). This intense secondary fracturing/faulting off the master fault is referred to as “damage” and has been extensively studied in the last decades (Cowie & Scholz, 1992; Faulkner et al., 2011; Manighetti, King, & Sammis, 2004; Perrin, Manighetti, Ampuero, et al., 2016; Savage & Brodsky, 2011; Zang et al., 2000) because it provides clues to understand fault mechanics. In particular, intense fracturing makes damaged rocks more compliant than the intact host material, which in turn impacts the master fault behavior (Bürgmann et al., 1994; Perrin, Manighetti, Ampuero, et al., 2016; Schlagenhauf et al., 2008). This issue is of particular importance when the master fault is seismogenic, that is, has the capacity to produce earthquakes. The architecture and the mechanical properties of the fault zone have indeed a strong impact on the earthquake behavior on the master fault: they partly control the rupture initiation and arrest and thus the rupture extent, but also the amplitude of the ground displacements and accelerations and thus the



**Figure 1.** Optical images of fault traces at ground surface, at different scales. Fault traces appear as dark lineaments in images, forming dense networks. At all scales, more pronounced, master fault traces (red arrows, not exhaustive) are “surrounded” (on one or both sides) by dense networks of closely spaced secondary faults and fractures of various lengths, generally oblique to and splaying off the master fault trace. (a and b) Pléiades satellite images of Waterpocket region, USA (ID in data statement). Two fault families are observed on (a), trending  $\sim$ NNE-SSW and  $\sim$ WNW-ESE. In both images, some of the faults have a vertical component of slip and thus form topographic escarpments that are seen through the shadows they form. (c and d) hand-held camera ground images in Granite Dells, USA; in (c) secondary faults splay from the underlined master trace, forming a fan at its tip; in (d), the underlined master fault trace is flanked on either sides by a narrow zone of intense fracturing, generally referred to as inner damage. Note the color change across the densely fractured inner damage zone.

earthquake magnitude and harmful potential (Hutchison et al., 2020; Manighetti, Campillo, Bouley, & Cotton, 2007; Oral et al., 2020; Radiguet et al., 2009; Stirling et al., 1996; Thakur et al., 2020; Wesnousky, 2006).

Given the prominent role of damage zones in fault and earthquake mechanics, accurate quantification of the 3D architecture of faults is of primary importance. At depth, this determination is still challenging as most geophysical imaging techniques rely on a number of assumptions and lack the resolution to discriminate the multiple closely spaced fractures and faults off the master fault “plane” (Yang, 2015; Zigone et al., 2019), while borehole provide valuable yet local observations (e.g., Bradbury et al., 2007; Morrow et al., 2015).

Alternatively, most faults intersect the ground surface, where they form clear traces generally with a footprint in the topography (Figure 1). These traces offer a great opportunity to observe most of the fault zone architecture at the ground surface. Therefore, most fault observations over the last century have been made at the ground surface, and rendered into 2D maps reporting the surface fault traces. Historically, these maps



were produced directly in the field from visual observations of the surface fault traces and the measurement of their characteristics (strike, dip, slip mode, and net displacements), and were reported as trace maps with specific attributes (e.g., Watterson et al., 1996). Over the last few decades however, the rapidly expanding volume of satellite and other remote sensing data has greatly assisted the fault trace mapping. Under appropriate conditions (e.g., no cloud overcast for optical images), surface fault traces (and fracture traces, depending on data resolution) are clearly seen in remote images and topographic data, and can thus be analyzed remotely. Optical images are most commonly used for fault trace analysis, as the faults appear as clear and sharp sub-linear or curvilinear “lines” of pronounced color gradients (e.g., Chorowicz et al., 1994; Manighetti, Tapponnier, Gillot, et al., 1998; Tapponnier & Molnar, 1977).

Accurately mapping fault and fracture traces requires a specific expertise as these traces have complex, curvilinear shapes and assemble, connect or intersect in an even more complex manner, yet partly deterministic (Figure 1). Because this complexity arises from the master fault growth and mechanics, it is of prime importance that it is properly recovered in the fault trace maps. Fault mapping is thus an expert task that requires a rich experience in fault observation and a fair understanding of fault mechanics. Would oversimplified or overlooked mapping be done, the signature of the mechanical processes would be lost and the faults left misunderstood. However, even when the fault mapping is done by an expert, various sources of uncertainties affect the mapping (e.g., Bond, 2015; Godefroy et al., 2020, and see below). The expert mapping is commonly done manually: the expert recognizes the fracture and fault traces visually in the remote images (and possibly other data), and reproduces these traces as hand-drawn lines in a Geographic Information System (GIS) environment. These environments allow fault attributes such as trace hierarchical importance, thickness, interruptions, connections, and slip mode to be labeled in various ways (various line thicknesses, colors, symbols, etc.) (e.g., Flodin & Aydin, 2004; Manighetti, Tapponnier, Gillot, et al., 1998). Some of the uncertainties on fault traces can also be informed.

Manual mapping of fault and fracture traces by an expert is arguably the most accurate processing technique. However, it is extremely time consuming, while the expertise might not be always available, which prohibits the analysis of large areas at high resolution, and dramatically limits the amount of available accurate fault maps. Edge detection algorithms have been proposed (e.g., Canny, 1986; Grompone von Gioi et al., 2012; Sobel, 2014) and used to automate fault mapping in image data, as these algorithms can rapidly extract all the objects expressed as a strong gradient in the images. However, these edge detectors generally suffer from poor accuracy and do not always discriminate fractures and faults from other linear features, especially in textured images (Drouyer, 2020). Another semi-automated approach named Ant-tracking (Monsen et al., 2011; Yan et al., 2013) has been developed to detect linear discontinuities in optical images but it needs heavy pre-processing of the images (for instance, to remove vegetation) and a priori constraints on the faults that are searched for (orientations, curvature, etc.). Other sophisticated approaches, based on stochastic modeling and ranking learning system, have demonstrated high efficiency to detect curvilinear structures in images (Jeong, Tarabalka, Nisse, & Zerubia, 2016; Jeong, Tarabalka, & Zerubia, 2015), but these approaches require to set many a priori parameters to define the structures that are searched for, and were never applied to faults.

Therefore, an automatic solution for rapid and accurate fracture and fault mapping in optical images would be greatly beneficial to better characterize the structural, geometric, and hence mechanical properties of fault zones. Here, we propose to address this challenge using deep machine learning, and more specifically, a Convolutional Neural Network (CNN) model (LeCun, Haffner, et al., 1999). Machine learning comprises a plethora of numerical methods that can learn from available data and make predictions in unseen data (process called generalization, e.g., Abu-Mostafa et al., 2012; Goodfellow, Bengio, et al., 2016). Machine learning is not a new domain of research, with the first landmark works having emerged in the 50–60's (see Figure 1 in Drams (2020) and references therein). However, its demands for large quantities of data and computational resources required by the learning process have only been met over the last two decades or so (e.g., Bishop, 1995; Carbonell et al., 1983; Devilee et al., 1999; Ermini et al., 2005; Goodfellow, Bengio, et al., 2016; LeCun, Bengio, & Hinton, 2015; MacKay, 2003; Meier et al., 2007; Mjolsness & DeCoste, 2001; Van der Baan & Jutten, 2000). This has spawned numerous applications in computer vision, natural language processing, and behavioral analysis (Sturman et al., 2020; Voulodimos et al., 2018; Young et al., 2018). Among the various algorithms, CNNs have shown great potential to tackle the complex task of image data



analysis (medical, satellite, etc.) and the capacity to rapidly process vast volumes of data (e.g., Krizhevsky et al., 2012; Russakovsky et al., 2015; Simonyan & Zisserman, 2014). In most computer vision tasks, CNNs indeed achieve state-of-the-art performance.

Currently, there is a rapidly expanding number of scientific works developing and/or applying CNNs and other machine learning techniques to assist researchers in detecting, modeling, or predicting specific features in a broad range of domains, including medical sciences (e.g., Ronneberger et al., 2015; Shen et al., 2017), remote sensing (e.g., Lary et al., 2016; Maggiori et al., 2017; Tasar, Tarabalka, & Alliez, 2019), chemistry (e.g., Schütt et al., 2017), climate sciences (e.g., Rolnick et al., 2019), engineering (e.g., Drouyer, 2020), and geosciences (e.g., Bergen et al., 2019). In the domain of geosciences, machine learning methods have especially flourished in earthquake seismology (e.g., Hulbert, Rouet-Leduc, Jolivet, & Johnson, 2020; Kong et al., 2019; Mousavi & Beroza, 2020; Perol et al., 2018; Ross, Meier, & Hauksson, 2018; Ross, Meier, Hauksson, & Heaton, 2018; Rouet-Leduc, Hulbert, & Johnson, 2019; Rouet-Leduc, Hulbert, McBrearty, & Johnson, 2020; Zhu & Beroza, 2019), and seismic and geophysical imaging of the Earth's interior with a major focus on sub-surface image interpretation for resource exploration (e.g., Babakhin et al., 2019; Dramsch & Lüthje, 2018; Gramstad & Nickel, 2018; Haber et al., 2019; Meier et al., 2007; Waldeland et al., 2018; Wu & Zhang, 2018). A few other works have been conducted in oceanography (Bézenac et al., 2019), geodesy (Rouet-Leduc, Hulbert, McBrearty, & Johnson, 2020), volcanology (Ren, Peltier, et al., 2020), and rock physics (Hulbert, Rouet-Leduc, Johnson, et al., 2019; Ren, Dorostkar, et al., 2019; Rouet-Leduc, Hulbert, Lubbers, et al., 2017; Srinivasan et al., 2018; You et al., 2020). However, few works have been conducted so far to address fault issues with machine learning, and all of them have targeted seismic images, that is, transformed vertical images of the Earth interior where actual faults cannot be observed directly (Araya-Polo et al., 2017; Guitton, 2018; Tingdahl & De Rooij, 2005; Wu and Fomel, 2018; Zhang et al., 2019). As far as we know, there has been no study addressing the identification and mapping of tectonic fractures and faults directly observable in optical images of the Earth ground surface.

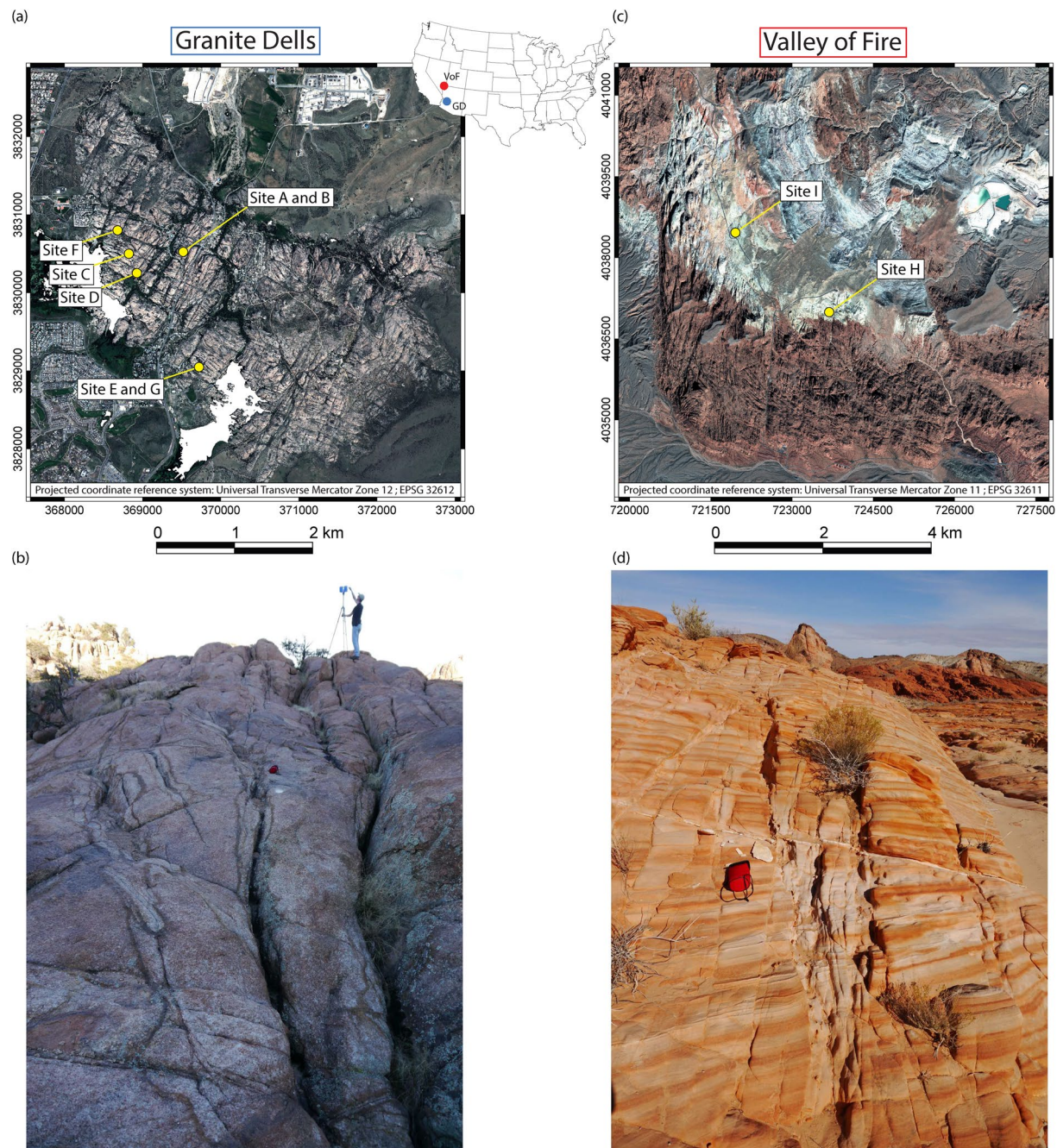
In the present study, we leverage recent deep learning developments for this task. We have adapted a CNN with U-Net architecture (Ronneberger et al., 2015) from Tasar, Tarabalka, and Alliez (2019), which was originally dedicated to identify buildings, roads, and vegetation in optical satellite images, to the challenge of fault and fracture detection. We develop a reference model,  $M_{\text{Ref}}$ , that we train with a small amount of image data and basic manual mapping, so as to simulate the common situation where only rough manual fault maps and few images are available. We apply  $M_{\text{Ref}}$  to different types of images (hand-held camera, drone, satellite) at distant sites, and demonstrate that it predicts the location of faults and fractures in images better than existing edge detection algorithms, and at a high level of accuracy similar to, and possibly greater than, expert fault mapping capacities. Using a standard vectorization algorithm, we analyze the  $M_{\text{Ref}}$  predictions to measure the geometrical properties of the fault zones. We provide the trained reference model  $M_{\text{Ref}}$  as a new powerful tool to assist Earth scientists to rapidly identify and map complex fault zones in ground images and quantify some of their geometrical characteristics.

## 2. Image, Topographic, and Fault Data

### 2.1. Fault Sites

We analyze faults in two distant sites differing in context, geology, fault slip modes and fault patterns, as well as in ground morphology and texture (Figure 2). In both sites, faults are inactive (i.e., currently not accommodating deformation) and were exhumed from several km depth (de Joussineau et al., 2007; DeWitt et al., 2008).

The first site, “Granite Dells,” is a large granite rock outcrop ( $\sim 15 \text{ km}^2$ ) with sparse vegetation, in central Arizona, USA (Haddad et al., 2012) (Figures 2a and 2b). The granite, of Proterozoic age, is dissected by a dense network of fractures and faults spanning a broad range of lengths from millimeters to kilometers (Figures 3a–3g). While the fractures are short open cracks with no relative motion, the faults are longer features with clear evidence of lateral slip (en echelon segmentation, pull-apart connections, slickensides). They have no or little vertical slip preserved and therefore they hardly imprint the topography (i.e., no or small escarpment), but where they have sustained differential erosion (Figures 2b and 3a–3g). Overall, fractures and faults organize in two families, trending about NE-SW and NW-SE and cross-cutting each other,

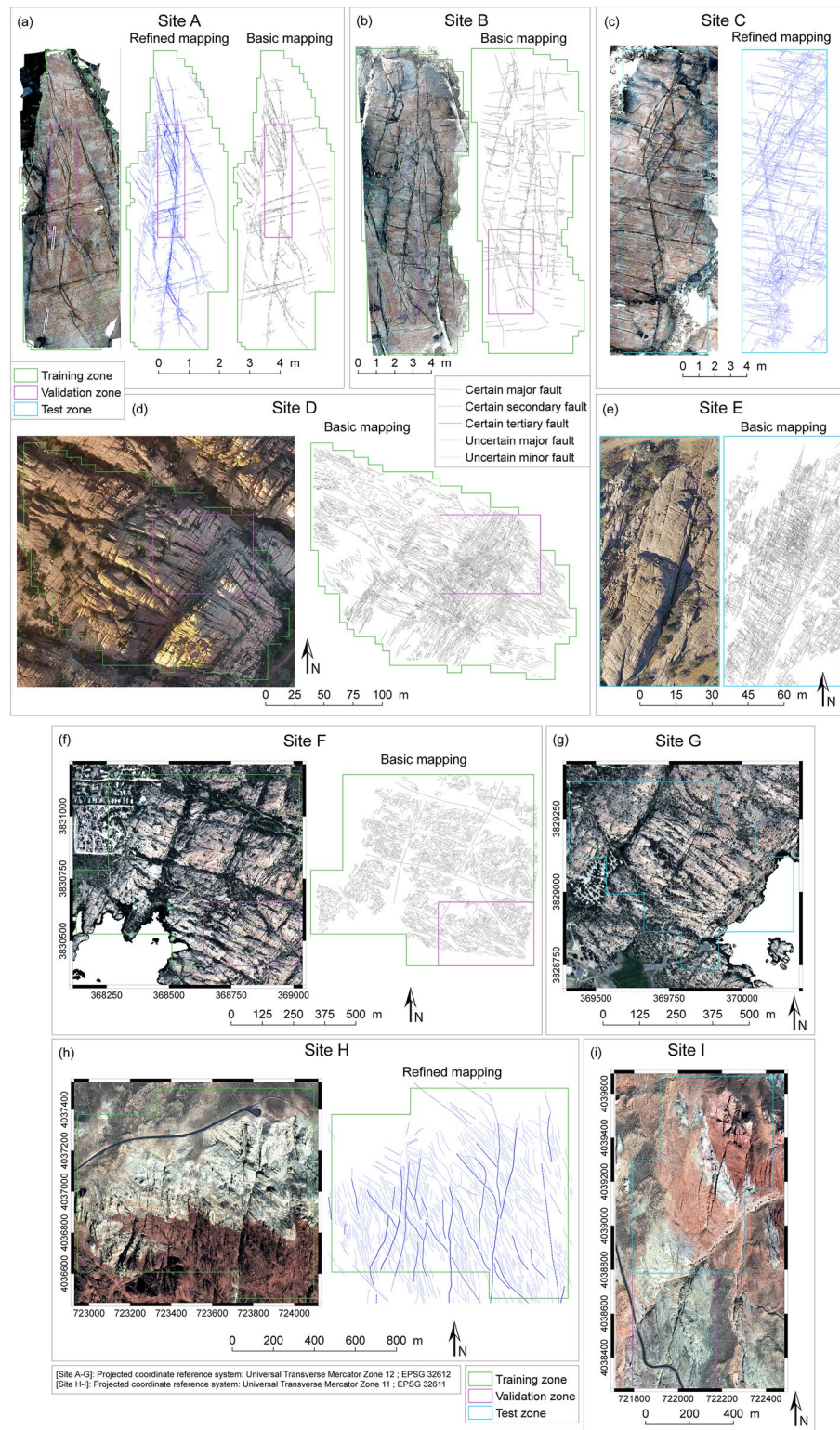


**Figure 2.** (a and b) Granite Dells, Arizona and (c and d) Valley of Fire, Nevada sites. (a and c) show Pléiades satellite images of the sites (ID in data statement), with location of the sub-sites analyzed in present study. (b and d) show field views of some of the faults and fractures (photographs courtesy of I. Manighetti). Note the vertical component of slip on some of them, forming topographic escarpments. In Granite Dells (b), most fracture and fault traces are surrounded by an alteration front (darker color).

yet with no clear systematic chronological relationship. The fracture and fault density is extremely high throughout the site, making their manual mapping difficult and time consuming.

A second site, “Valley of Fire,” is a large outcrop ( $\sim 30 \text{ km}^2$ ) of the Jurassic Aztec Sandstones formation, in southern Nevada, USA (de Joussineau & Aydin, 2007; Myers & Aydin, 2004) (Figure 2c). The site is desertic and thus has no vegetation. It is densely dissected by faults of meters to kilometers of length, showing clear evidence of both lateral (left- and right-lateral) and vertical (normal) thus oblique slip (Figures 2c, 2d, 3h,





**Figure 3.** Sites, images, and ground truth fault maps used in present study. (a–g) from Granite Dells site, (h–i) from Valley of Fire site. (a–c): hand-held camera ground images, (d and e): drone images, (f–i): Pléiades satellite images (ID in data statement). Training, validation and test zones are indicated in green, purple and blue, respectively. Sites C and G are only used as test zones. Basic (black) and refined (blue) mapping were done manually (not shown on Site I for figure clarity). Fault hierarchy is indicated only in refined maps in (a, c, and h) through different line thicknesses, but is most visible when zooming the figures out. Dotted lines are for uncertain fractures and faults. Note, in most sites, the very high density of fractures and faults, the existence of different fault families with various orientations, and the presence of non-tectonic features. Also note that, in all sites, the manual mapping is incomplete (as exhaustive manual mapping is practically unachievable).



**Table 1**  
*Characteristics of Image and Topographic Data at Each Site*

Site	Number of images in training data set (green)	Number of images in validation data set (pink)	Number of images in test data set (blue)	Number of images in total data set	Total area of analyzed zone (m <sup>2</sup> )	Spatial resolution of image & DSM (m)	Fault length range (m)	Approximate vertical resolution of DSM (m)	Type of mapping
A	1,365	203	–	1,568	25.7	$0.5 \times 10^{-3}$	0.001–10	0.005	Refined (HR) & basic
B	809	120	–	929	87.7	$1.2 \times 10^{-3}$	0.001–10	0.01	Basic
C	–	–	912	912	101.0	$1.3 \times 10^{-3}$	0.001–10	0.01	Refined (VHR)
D	678	154	–	832	$31.4 \times 10^3$	$24 \times 10^{-3}$	0.01–100	0.5	Basic
E	–	–	325	325	$6.9 \times 10^3$	$10 \times 10^{-3}$	0.01–100	0.5	Basic
F	30	6	–	36	$590 \times 10^3$	0.5	10–700	1	Basic
G	–	–	19	19	$310 \times 10^3$	0.5	10–400	1	Basic
H	63	–	–	63	$1,032 \times 10^3$	0.5	10–1,000	1	Refined (HR)
I	–	12	29	41	$670 \times 10^3$	0.5	10–1,000	1	Basic

*Note.* Each image refers here to a  $256 \times 256$  pixels tile.

Abbreviation: HR, high resolution; VHR, very high resolution.

and 3i) (de Joussineau & Aydin, 2007). Most faults have thus formed topographic escarpments from meters to tens of meters high. Faults organize in two principal families trending about N-S and NW-SE (Ahmadov et al., 2007; Myers & Aydin, 2004).

## 2.2. Optical Image and Topographic Data

To document fractures and faults over a broad range of lengths from mms to kms, we have acquired different types of optical images on six subregions within the two sites (Figure 2), amounting to a total of ~4,600 individual images over the two sites. We have then used Structure from Motion (SfM) photogrammetric technique (e.g., Bemis et al., 2014; Bonilla-Sierra et al., 2015; Harwin & Lucieer, 2012; Turner et al., 2012) to calculate, from the individual images, both the topography of each subregion (Digital Surface Model or DSM) and its ortho-image (rectified mosaic-like global image). The six subregions being almost free of vegetation, the DSM are equivalent to the DTM (Digital Terrain Model) most commonly used. The resolutions of the DSM and the ortho-images are here the same.

We have acquired four types of images (Figures 2 and 3, and Table 1):

- In three subregions of the Granite Dells area later referred to as Sites A, B, and C (Figures 2a and 3a–3c), we have acquired a total of 165 RGB photos “on the ground” with a standard optical camera (Panasonic Lumix GX80) elevated by 3 m above the ground. Sites A, B, and C are about 3, 6, and 7 m wide and 10, 15, and 20 m long, respectively. The ortho-images and DSMs (calculated with Agisoft Metashape software, AgiSoft Metashape Professional (Version 1.4.4, 2018; <http://agisoft.com>) of Sites A, B, and C have an ultra-high resolution of 0.5, 1.2, and 1.3 mm, respectively. They thus allow recovery of fractures and faults down to a few millimeters in length, while the extent of the surveys limits the longest recovered faults to 10–15 m (Table 1). As we could not acquire GPS ground control points in the sites, the ortho-images and DSMs are not geo-referenced. However, to ease their description, we will refer to as “North,” “South,” “West,” and “East” for the Top, Bottom, Left and Right sides of the images, respectively
- In two other subregions of the Granite Dells site referred to as Sites D and E (Figures 2a, 3d, and 3e), we have acquired ~40 RGB photos from an optical camera onboard a drone (DJI Phantom 4) flying less than 100 m above ground level. Sites D and E are about  $230 \times 150 \text{ m}^2$  and  $50 \times 100 \text{ m}^2$ , respectively. Their ortho-images and DSMs, here georeferenced, have a high resolution of 2.4 and 1 cm, respectively, allowing recovery of fractures and faults down to a few centimeters of length, while the longest recovered faults are a few 10–100 m long (Table 1)

- (c) Over the Granite Dells site, we have also acquired three Pléiades satellite optical images with different view angles (ID in data citations), which we use to analyze two subregions referred to as Sites F and G (Figures 2a, 3f, and 3g). Each Pléiades image is delivered as two files, one panchromatic image with one band at a resolution of 50 cm, and one multispectral image with four bands at a resolution of 2 m. Using the panchromatic tri-stereo images, we calculated a DSM of the site at 50 cm resolution (Micmac software, free open-source solution for photogrammetry, see Rupnik, Daakir, & Deseilligny, 2017; Rupnik, Deseilligny, et al., 2016). Then, using the multispectral Pléiades image with the view angle closest to the Nadir, we derived a color ortho-image of the site. To calculate the color ortho-image at 50 cm resolution, we used pansharpening techniques from GDAL/OGR (2018) to fuse the highest resolution panchromatic image with the lower resolution multispectral image. The F and G sites are  $\sim 900 \times 790 \text{ m}^2$  and  $760 \times 640 \text{ m}^2$ , respectively, and therefore the longest recovered faults are 400–700 m long. Although the resolution of the images is high, the smallest faults that are well resolved are  $\sim 10 \text{ m}$  long
- (d) In the Valley of Fire site, we have acquired three Pléiades satellite optical images (ID in data citations) with different view angles to analyze two subregions referred to as Sites H and I (Figures 2c, 3h, and 3i). As above, we calculated a DSM and an ortho-image of Sites H and I at 50 cm resolution. The H and I sites are  $\sim 1.1 \times 1 \text{ km}^2$  and  $1.4 \times 0.6 \text{ km}^2$ , respectively, so that the longest recovered faults are about 1 km long. Although the resolution of the images is high, the smallest faults that are well resolved are  $\sim 10 \text{ m}$  long

### 2.3. Fault Ground Truth Derived From Manual Mapping

Manual fault mapping of parts of the sites is used as ground truth to train the CNN models. Fractures and faults were hand-mapped in the images and topographic data following common procedures in active tectonics (e.g., Manighetti, Tapponnier, Courtillot, et al., 1997; Manighetti, Tapponnier, Gillot, et al., 1998; Tapponnier & Molnar, 1977; Watterson et al., 1996). Mapping was done with Erdas Ermapper ([www.geosystems.fr](http://www.geosystems.fr)) or QGIS (QGIS Development Team, 2020). Faults are represented by curvilinear lines of different thicknesses made to inform on the fault hierarchy (major vs. minor faults, see below). It has to be noted that this information on fault hierarchy is provided to assist the tectonic analysis, but is not used in the deep learning training (it is a further scope). Fault maps are eventually stored as shapefiles, that is, vector files containing geographic information.

To explore which level of expertise in fault mapping is needed to properly train the Deep Learning models, we produced three qualities of mapping (Table 1):

- Sites A, B, D, E, and F were mapped by a student with limited expertise in fault mapping (Figure 3). This produces a comparatively low resolution ground truth, where fault traces are simplified while many minor fractures and faults are missing and the fault hierarchy is not indicated. These mappings are referred to as “basic” in the following
- Site A and H were mapped by an expert geologist, yet at a moderate resolution (Figures 3a and 3h). This means that while most faults are mapped accurately and their hierarchy informed, some details in their trace are simplified while the smallest fractures and faults are not mapped. The expert mapping is referred to as “refined” in the following
- Site C was mapped by an expert geologist, at the highest possible resolution (Figure 3c). This means that all fractures and faults visible up to the highest resolution of the data were mapped as accurately as possible, with their hierarchy informed. The C fault map is thus the most exhaustive ground truth

In the refined fault maps, faults traces are discriminated into 3 levels of hierarchy during the mapping, using lines of different thickness (Figures 3a–3h): from thicker to thinner lines, (1) the major faults which have the most pronounced and generally the longest traces, and hence likely accommodate the greatest displacements, (2) the secondary faults and fractures, which have less pronounced but still clear traces and are generally of shorter length, and (3) the tertiary faults and fractures which have subtle traces, commonly looking like very thin cracks in the rock. It has to be noted that, because displacement typically varies along a fault (e.g., Manighetti, King, Gaudemer, et al., 2001), the thickness of the line made to represent a fault trace can vary along its length. Again, the information on fault hierarchy is provided to assist the tectonic analysis, but is not used in the following deep learning training.

Regardless of the degree of expertise to produce the manual fault mapping, the drawn fault traces have four uncertainties: (1) a fault trace generally looks as a dark, thick lineament of several pixels width in optical images. This width results from different causes, including the real thickness of the fault trace (from mm to m depending on fault size, e.g., Sibson, 2003), the existence of intense fracturing juxtaposed to the fault trace (see Figure 1d), the apparent widening of the fault trace due to a vertical component (see shadows in Figures 1a–1c), and the weathering of the fault trace at the ground surface. Generally, the exact position of the fault (i.e., the exact intersection “line” between the fault plane and the ground surface) cannot be assessed accurately within this several-pixels wide trace, even by the best expert; (2) Even in cases where the fault trace is clear, its drawing by the user cannot have a pixel-wise resolution: for a matter of time, mapping is commonly done at a lower resolution than the image pixel. The mapping thus conveys uncertainties, especially where the fault trace is tenuous, and when the mapping is done at low resolution (as in the basic mapping here); (3) some fault traces are uncertain; while they look like fault traces, they might represent other features, like alteration zones or erosion patterns. Where such an ambiguity arises, the uncertain fault traces are represented by dotted lines; (4) depending on the scale and resolution at which the data is analyzed, some faults may not have been mapped while they do exist in the outcrop; exhaustive manual mapping is practically unachievable. Therefore, the fault ground truth is inevitably incomplete and holding uncertainties.

The fault shapefiles are converted into raster maps where each pixel is assigned a probability to be a fault. If actual fault traces would be one-pixel lines, the fault pixels would be assigned a probability of 1 (i.e., being a fracture or a fault), while all other pixels in the image would be assigned a probability of 0 (i.e., not being a fracture or a fault). To account for the uncertainties described above as items 1 and 2, we applied to the fault map raster a 2D Gaussian filter (Deng & Cahill, 1993) with a standard deviation of 0.8 pixels. The 2D Gaussian filter acts as a point-spread function: it assigns a probability of 1 (i.e., being a fracture or a fault) to the pixels along the drawn fault lines, and a probability decreasing in a Gaussian manner from 1 to 0 to the 3 pixels-wide zone either side (probabilities normalized after Gaussian spreading). Furthermore, along uncertain faults (dotted lines, and uncertainty described above as item 3), the largest probability is set to 0.5. Together these make the ground truth fault lines having a 7 pixels width and a heterogeneous probability distribution ranging from 0 (i.e., 100% certainty that it is not a fracture or a fault) to 1 (i.e., 100% certainty that it is a fracture or a fault).

In summary, the data used as input in the CNN algorithm described below include aligned optical images, DSMs and ground truth fault probability maps, organized in two files: (1) the optical images (3 bands) and the DSM (1 band) are concatenated into a 3D array (4 bands) of pixel values, and (2) the ground truth fault probability map is provided as a 2D array of probability pixel values.

### 3. Deep Learning Methodology

We use a CNN of U-Net architecture that we have adapted from Tasar, Tarabalka, and Alliez (2019). The algorithm is written in Python, and employs the TensorFlow library (<https://www.tensorflow.org/about/bib>; Abadi et al., 2015).

#### 3.1. Principles of Deep Learning and Convolutional Neural Networks

Here we present a mostly qualitative description of the Deep Learning methods employed in this study, aiming to provide a general overview of our approach. For a more quantitative exposition of the methodologies, we refer to Text S1.

In many situations, relatively simple relationships can be established between two quantities. While these relationships are not necessarily linear or physically interpretable, they can be easily visualized and identified owing to their low-dimensional character (i.e., they can be captured in a two-dimensional graph). On the other hand, some relationships are more abstract and much harder to characterize; one prime example is that of individual pixels in an image and their relationship to the contents of the image (e.g., a human face, an animal, a tectonic fault, etc.). While a single pixel does not uniquely define the contents of the image, it is the arrangement and ordering of pixels of particular color and intensity that are interpreta-



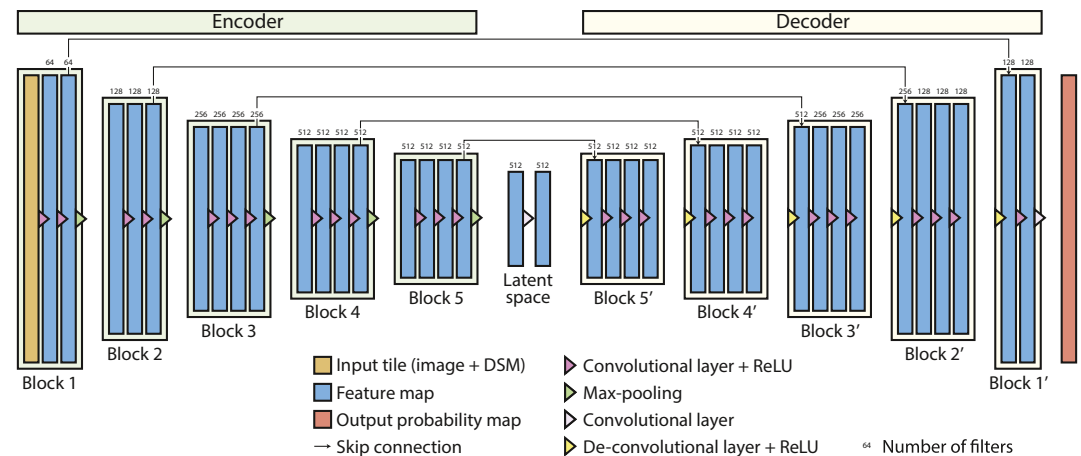
ble by an observer. This makes the mapping from the input (image pixels) to the output (image contents) high-dimensional. The human brain excels at creating such abstract, high-dimensional relationships, but reproducing this flexibility with computational systems (e.g., in computer vision tasks) has traditionally been challenging.

As a subclass of Machine Learning algorithms, Deep Learning draws inspiration from the human brain, allowing Artificial Neural Networks (ANN) to “learn” from the input data through incremental updates of computational units that are colloquially referred to as “neurons” (see LeCun, Bengio, & Hinton, 2015 for a review). The input data can be already interpreted (these interpreted data are called “ground truth,” and the learning is said “supervised”), or not (unsupervised learning), and there exist combinations. ANNs attempt to find an optimal relationship between the input  $y$  and output  $\hat{y}$  defined by a set of parameters  $\theta$ , often called “weights” and “biases.” During “training” of the model, the parameters  $\theta$  are incrementally adjusted by the network to optimize a pre-defined cost (or loss) function. The relationship between  $y$  and  $\hat{y}$  may not be obvious to describe in terms of interpretable criteria, but it may be empirically approximated by a large number of non-linear functions and their parameters. While the analogy between biological and artificial neural networks is mostly symbolic, ANNs nonetheless achieve state-of-the-art performance in abstract tasks such as computer vision and text interpretation.

Practically speaking, ANNs are constructed in a hierarchical fashion in which the parameters are grouped in layers. A given input vector  $y$  is passed onto the first layer, defined by a set of parametrized functions, to produce an intermediate output  $y_1$ . This intermediate output is passed onto a subsequent layer to produce the next output  $y_2$ , and this procedure is repeated several times until the terminal layer in the ANN produces the final model output  $\hat{y}$ . While the functions inside each layer are simple (see Text S1) and typically characterized by only two parameters (a “weight” and a “bias,” the former being a matrix and the latter a vector), the non-linear combination of a multitude of such functions permits an abstract and extremely high-dimensional transformation from the input to the output. In a supervised setting (as adopted in this study), the output of the model  $\hat{y}$  is compared with a pre-defined ground truth  $y$ , and during the training phase the parameters  $\theta$  within each layer are incrementally modified to minimize the difference between the model output and the ground truth. The details of the types and number of functions used in each layer, as well as the number of layers and the cost function that measures the difference between the model output and the ground truth, all fall under the header of the model “architecture.” This architecture is chosen by the user depending on the task to be performed, the type and volume of input data, and the available computational capacities. The architecture parameters (number of filters, blocks, layers, loss function, etc.) are set manually and called “hyper-parameters,” by opposition to the model “parameters” (such as filter weights) which are calculated by the network.

One of the most common architectures employed in numerous scientific domains is the CNN (Fukushima, 1975; LeCun, Bengio, & Hinton, 2015) and its variations. By design, CNNs leverage the notion that spatial and temporal data (e.g., time series, images, and video or sound clips) are mostly locally correlated: 2 pixels that are close together in an image are more likely to be part of the same object than 2 pixels that are far away. Each layer in a CNN consists of kernels (also called “filters,” represented by small matrices) that process a small part of the input data at a given time, and that are reused at every location within the data set (Text S1). Intuitively, this reusing of the kernels can be envisioned as sliding each of them through an image or a time series so as to examine a specific feature in the image or data (e.g., an edge with a particular orientation, another type of motif, a combination of motifs, etc.), and then constructing an output holding information on this specific feature. This “scanning” of the input data by the kernels occurs through the multiplication of the input image pixel values with the weights of the filter matrix and can be represented mathematically as a convolution operation (hence the term “Convolutional Neural Network”) (Figure A1 in Text S1).

For the task of pixel-wise classification (i.e., assigning a label to each pixel in the image), a CNN takes an image (either a single-channel greyscale or multi-channel color image as is the case here) as an input to the first layer, which produces a layer output that is proportional in size to the input image, possibly with a different number of channels. The layer output is then passed onto subsequent layers as described above. A given channel of the intermediate output is often referred to as “feature map” as it is constructed by a convolutional kernel that is said to represent a “feature.” In spite of this terminology, a single kernel does



**Figure 4.** Architecture of U-Net CNN used in present study. See text for detailed description. DSM, Digital Surface Model; ReLU, Rectified Linear Unit.

not necessarily represent a well-defined object in the input image, and the interpretation of CNNs (and ANNs in general) has thus proven to be a challenging topic of ongoing research. A typical CNN contains many layers with kernels that hierarchically interact to collectively extract the relevant objects in an image. In general, the more complex the image and the searched objects, the “deeper” the CNN needs to be (i.e., greater number of layers and kernels). A common interpretation is that early layers extract simple patterns such as edges, corners, etc., while deeper layers make combinations of these primitives to extract more complex patterns and shapes, with the final layers extracting complete objects (e.g., Lee et al., 2009). For deep CNNs, the number of filters that “scan” the input image is very high, often counting up to thousands to millions. It is important to keep in mind that these filters are not pre-defined by the user, but learned by the network (Text S1).

### 3.2. Architecture of the CNN Model Used in Present Study

We have adapted the CNN model developed by Tasar, Tarabalka, and Alliez (2019) to originally detect and map buildings, high vegetation, roads, railways and water in optical images (Figure 4). This CNN is a variant of a U-Net architecture (Ronneberger et al., 2015), which, as any U-Net, includes two sub-networks, called encoder and decoder, with skip connections between the encoder and decoder units.

The main role of the encoder is to extract the objects of concern (called “context”) that contribute to final predictions; conversely, the role of the decoder is to synthesize these features into a prediction and locate them accurately. The encoder is a “contracting path” because it down-samples the data and coarse-grains them through “pooling layers,” generally added after each or a short series of convolutional layers (Text S1). The pooling operation consists in a sliding window that scans the output of the previous layer, looking for the average (“mean pooling”) or the maximum (“max pooling”) value at each position of the scan (Boureau et al., 2010). The pooling operations thus reduce the size of the feature maps, and allow the next filters to have a larger field of view of the data, and thus the capacity to identify larger motifs. Here we use max pooling layers.

Because the encoder down-samples the data, the searched features can be accurately detected but their localization is not recovered accurately. Therefore, a decoder sub-network is added to perform the complementary operation: it up-scales the data (it is thus an “expanding path”) and enables the localization of the identified features. To facilitate the encoding-decoding process, the model is here designed to be symmetric, with the output of each block in the encoder being concatenated to the input of its symmetrical counterpart in the decoder. These concatenations, or “skip connections,” regularize the training process by smoothing the high-dimensional loss landscape defined by the loss function  $\mathcal{L}$  (see Section 3.3), leading to faster convergence to the global minimum (Li et al., 2018). Each block in the decoder performs the same number of convolution operations as its counterpart in the encoder, but which are followed by an up-scaling operation

in the form of a strided deconvolution layer. More details can be found in Tasar, Tarabalka, and Alliez (2019) and in Text S1.

Here, both the encoder and the decoder have 13 convolutional layers, and one central layer that forms the “latent space” (see below) (Figure 4). The 13 layers of the encoder are distributed over five blocks. Each of the first two blocks contains two convolutional layers while each of the three other blocks are composed of three convolutional layers. The convolutional layers of the first block each comprises  $64 \times 3 \times 3$  convolutional filters. The number of filters increases by a factor of 2 in the next blocks until reaching 512 filters in blocks 4 and 5. To ensure that the center of each  $3 \times 3$  filter corresponds to the first pixel of the input image, a padding of 1 is applied to the image (i.e., a border of 1 pixel of value zero is added to the input image, to match the outermost line of the matrix) (Figure A1 in Text S1). Each filter then slides by 1 pixel-step over the image (“stride” of 1 pixel). Each convolutional layer is followed with a standard rectified linear unit (ReLU) function, aimed to introduce non-linearity in the model, and the resulting feature maps are down-sampled with a  $2 \times 2$  max-pooling layer that reduces the resolution of the feature maps by a factor 2. During the encoding process, the characteristics of the input data are summarized in a compact set of numbers stored in the central latent space. The decoder is symmetric to the encoder and thus also consists of five blocks. The first four blocks are composed of one deconvolutional layer and three convolutional layers, and the last block includes one deconvolutional layer and one convolutional layer. The final layer in the network is a convolution layer with a single filter, followed by a sigmoid activation function to ensure that the output falls between 0 and 1 (so that the output can be interpreted as a confidence score). The weights and biases of the filters are updated at each iteration of the training stage using an Adam optimizer (Kingma & Ba, 2014) with a fast learning rate of 0.0001. The weights are randomly initialized following Glorot and Bengio (2010). No regularization (dropout, batch normalization, L2-regularization, etc.) was performed.

### 3.3. Training Procedure

The available data are 3 bands-optical images, topographic data and manual fault maps. For each site, we have merged the image and topographic data so as to obtain a four bands data set (red, green, blue, and topography). In the following, we refer to this data as “image” even though it includes topographic data. The total available data is split into a training set, a validation set, and a test set. To prevent overfitting, we split the data by assigning dedicated regions of the mapped areas (see Figure 3). Limited by the GPU on-board memory availability, to train the CNN models we have split each data set into small tiles called “patches” of  $256 \times 256$  pixels and we have grouped in a batch (a set of patches), 12 of such patches picked randomly in the training data.

Owing to the relatively small volume of ground truth data available for training, overfitting of the model might occur during the training. Over-fitting is when the model learns so much from a specific data set that it becomes unable to predict anything else than this data set. To prevent overfitting, we artificially augmented the data set by applying geometrical and numerical transformations, including random horizontal rotations (here we used rotations by  $90^\circ$ ,  $180^\circ$ , and  $270^\circ$ ), flip transformations, color adjustments (with changes in contrast and gamma correction), and the addition of Gaussian noise (Buslaev et al., 2020). The degree of overfitting is monitored during training by evaluating the model performance on an independent validation data set, which is not used to constrain the model parameters directly. Training of the model is arrested when the model performance on the validation set systematically decreases, while simultaneously the model performance increases on the training set (a primary indicator of overfitting) – see Figure S2 for the stopping criterion for  $M_{\text{Ref}}$ .

To evaluate the performance of the model during the training, we use the cross entropy loss function which is more sensitive to confident false classifications than a mean-squared error function (Nielsen, 2015). As a first step here, we adopt a binary classification approach, assigning to each pixel a class “fault” or “not a fault” (i.e., fault hierarchy is not taken into account). To account for the uneven distribution of fractures and faults in the image data, we use the proportion of “not-a-fault” pixels as a weight parameter ( $\beta$ ; calculated according to the training and validation sites that are used to train and validate the current model) in the cross entropy calculation. The weighted binary cross entropy loss ( $\mathcal{L}$ ) is the average calculated from each pixel ( $n$ ) of the ground truth ( $y$ ) and the prediction ( $\hat{y}$ ) expressed as follows (Equation 1; Sudre et al., 2017):



$$L = -\frac{1}{n} \times \sum_{i=1}^n y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \beta \quad (1)$$

### 3.4. Estimating the Performance of the Models

The performance of the models is first estimated in a qualitative manner, on the one hand through the visual inspections of the results (i.e., fracture and fault probability maps), and on the other hand through the comparison of the results with those obtained with the most common available algorithms, such as edge detectors. We specifically use an edge detection algorithm that computes color gradients in the images (from Canny, 1986; later referred to as “Canny edge filter”), and another one which selects linear gradients (from Grompone von Gioi et al., 2012; later referred to as “GVG detector”). We also compare the results with those obtained with CNN models having different hyper-parameters or a different architecture.

We then quantify the performance of the models. Among the various CNN models that we calculate, the performance is first measured on the validation data sets, in order to select the most appropriate architecture and hyper-parameters for the CNN model. We refer to this most appropriate model as the “reference” model,  $M_{\text{Ref}}$ . We then demonstrate the robustness of  $M_{\text{Ref}}$  through its application to independent test data sets (different sites and images, Table 1) that were never seen during the training or the validation.

Because neither the ground truth fault maps nor the model predictions are binary 0 and 1 values, we use a performance metric that handles continuous probabilities. We have chosen the Tversky index (TI, Equation 2, Tversky, 1977), a variant of the most common Intersection over Union (IoU) metrics. The latter measures the ratio between the summed number of pixels (expressed as an area) whose prediction is consistent with the ground truth, and the total area encompassing both the ground truth and the predictions, yet only applies to binary data. The Tversky index is more flexible as it can be applied to continuous probabilities, while it also takes into account (1) the small proportion of pixels being a fracture or a fault in an otherwise dominantly un-fractured area, and (2) the incomplete ground truth, in which some fractures or faults might have not been mapped while they actually exist and are well predicted by the model. The TI metric ( $0 \leq \text{TI} \leq 1$ ) is calculated per pixel, then summed over the pixel population  $n$ , as follows (Equation 2):

$$\text{TI} = \frac{\sum_{i=1}^n (y_i * \hat{y}_i)}{\sum_{i=1}^n (y_i * \hat{y}_i + \gamma(1 - y_i)\hat{y}_i + (1 - \gamma)y_i(1 - \hat{y}_i))} \quad (2)$$

The coefficient  $\gamma$  denotes the proportion of “certain” fractures and faults in  $y$ ; it is thus the number of pixels of probability greater than 0 normalized to the total number of pixels in the image. In the sites studied here  $\gamma$  is less than 1% on average. Therefore,  $\hat{y}\hat{y}$  evaluates the pixels detected as fractures or faults in both the ground truth and the prediction,  $(1 - \gamma)\hat{y}$  measures the fractures and faults that are predicted by the model but do not exist in the ground truth, and  $\gamma(1 - \hat{y})$  measures the fractures and faults that exist in the ground truth but are not predicted by the model. The two latter terms are weighted by  $\gamma$  to account for sparsity of fractures and faults in the site. A TI value of 1 indicates a perfect prediction, with a value of 0 indicating the converse. Note that, in this TI calculation, only certain fractures and faults are considered in the ground truth.

To compare our results with those of prior works, we also use common metrics such as the receiving operating characteristic (ROC) curve, which compares the “False positive rate” (predicted fault/fracture location not present in the ground truth) to the “True positive rate” (predicted fault/fracture location present in the ground truth) at increasing probability thresholds. With this metric good model performance is characterized by a low False positive rate and a high True positive rate.

To further measure the accuracy of the reference model ( $M_{\text{Ref}}$ ), we qualitatively compare the density, azimuth and length distributions of the ground truth and the predicted fractures and faults in the validation zone. Finally, we test the robustness of  $M_{\text{Ref}}$  through its application to unseen data sets of different types (different types of images, such as drone and Pléiades, Table 1).

## 4. Defining a “Reference Model” $M_{\text{Ref}}$

In this section and the following, we discuss the qualitative results in both the entire validation or test zone, and in a smaller “reference” area chosen for its tectonic complexity.

### 4.1. Selecting the Most Appropriate CNN Architecture

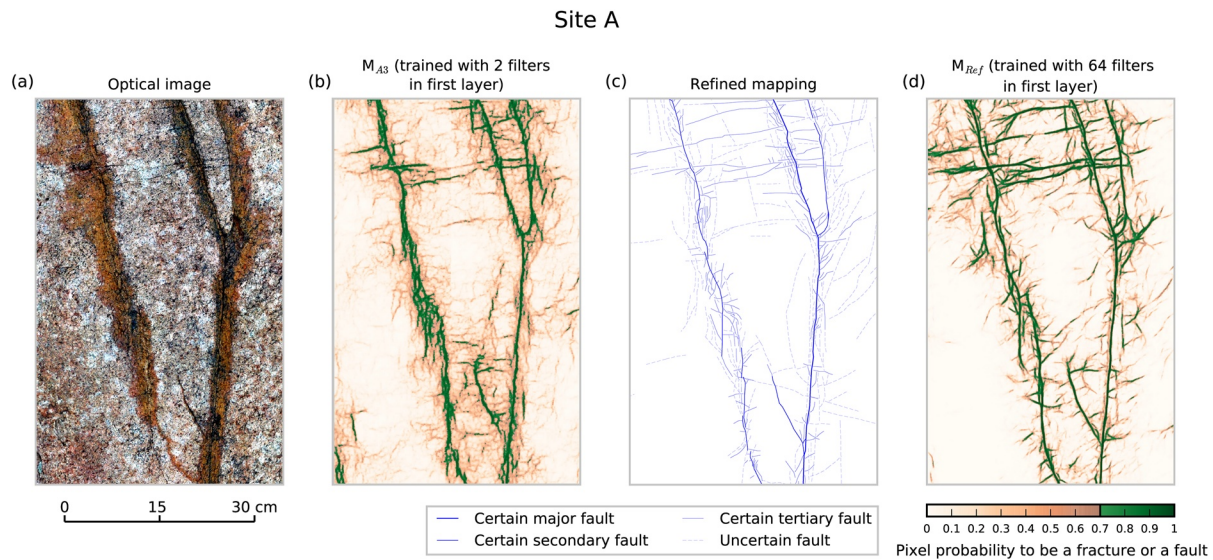
To explore which architecture and model hyper-parameters are the most appropriate for fracture and fault detection in the given image data, we set-up seven U-Net models with different numbers of filters, blocks, layers, and loss functions. While we could not explore the entire range of possibilities, we examined the CNN architectures most commonly used in prior works (see Ide & Kurita, 2017; Ronneberger et al., 2015; Tasar, Tarabalka, & Alliez, 2019).

We designed seven models differing from our reference architecture  $M_{\text{Ref}}$  described in Section 3:

- One model (referred to as  $M_{A1}$ ) with an architecture as described in Section 3, but using a Mean Squared Error (MSE) loss function
- One model (referred to as  $M_{A2}$ ) with an architecture as described in Section 3, yet lacking the last block 5 of the encoder and the corresponding block 5' in the decoder
- Five models (referred to as  $M_{A3}$  to  $M_{A7}$ ) with an architecture as described in Section 3, yet having 2, 4, 8, 16, and 32 filters in the first layer, respectively, instead of 64 in the reference architecture  $M_{\text{Ref}}$

All these models were trained on the same data. To have a large and homogeneous training data set, we used the basic mapping ground truth on Sites A and B (green zones in Figures 3a and 3b, including a total of 2,174 tiles of  $256 \times 256$  pixels). We then estimated the TI scores of the models in the validation zones (in pink in Figures 3a and 3b, including a total of 323 tiles of  $256 \times 256$  pixels). Using basic fault maps as training data simulates the common situation where only rough manual fault maps are available.

Because a refined ground truth exists in Site A but not in Site B, two TI values are calculated: (1) a  $TI_R$  calculated only in Site A with reference to the refined ground truth, and (2) a  $TI_B$  calculated in both Sites A and B with reference to the basic ground truth. In the latter case, we average the two TI values. Figure S3 shows the TI values for the seven models  $M_{A1}$  to  $M_{A7}$  and the reference architecture  $M_{\text{Ref}}$  used with the same data sets. First, for all models but  $M_{A1}$ , the  $TI_R$  calculated with reference to the refined mapping is systematically slightly lower than the  $TI_B$  calculated with reference to the basic mapping. The difference is likely due to the training done with a ground truth not containing as many details as the refined fault map, leading to a model unable to make predictions at sufficient detail. The difference is small however, and both TI values peak at  $0.67 \pm 0.01$  for the best  $M_{\text{Ref}}$  model. Looking at the models individually,  $M_{A1}$  delivers the lowest TI (0.35), almost twice lower than the TI of the reference model, demonstrating the importance of the loss function. Reducing the number of blocks in the architecture of the CNN from 5 ( $M_{\text{Ref}}$ ) to 4 ( $M_{A2}$ ) does not have a strong impact on the performance of the model ( $TI_B$  and  $TI_R$  of 0.63 and 0.59, respectively). In contrast, the number of filters in the first layer of the model has a significant impact on the results. While 16 filters ( $M_{A6}$ ) might be sufficient to get a reasonable result (TI of 0.62–0.64), a minimum of 32 filters ( $M_{A7}$ ) is suggested to obtain near-optimal performance. It is interesting to note that, based on the visual inspection of the results, the simplest model  $M_{A3}$  with only two filters in the first block predicts the major and secondary faults fairly well (Figure S4), as well as some of the more minor features (Figure 5b). The model also recognizes that the ruler in the left bottom corner of Figure S4 does not correspond to a fracture, in spite of its straight features and sharp contrast with its surroundings. Furthermore, most predictions at moderate probabilities (0.4–0.6) underline real cracks in the rock (compare Figures 5a and 5b). Some of these were labeled as “uncertain” faults in the ground truth (Figure S4c), but most were not mapped. Therefore, although having only a small number of parameters, the  $M_{A3}$  2-filter model already provides a realistic view of the cracks and faults distribution in the site. The more expressive (i.e., larger)  $M_{A7}$  model provides a prediction almost as good as  $M_{\text{Ref}}$  (Figure S3), yet with slightly larger uncertainties on fault locations (“thicker” probability “lines”), and more apparent “noise” in the form of short fractures with low probabilities (Figure S5). However, when looking at these short fractures in more details (Figure S5b), we can see that most of these do coincide with real cracks in the rock.



**Figure 5.** Impact of number of filters in first layer. (a) Optical image of part of Site A (from validation zone, see Figure 3a), used as a reference example zone. (b) Predictions from a model with architecture of  $M_{Ref}$ , however with two filters only in first layer. (c) Refined mapping, with fault hierarchy indicated (better seen when zooming). (d) Predictions from  $M_{Ref}$  (64 filters in first layer).

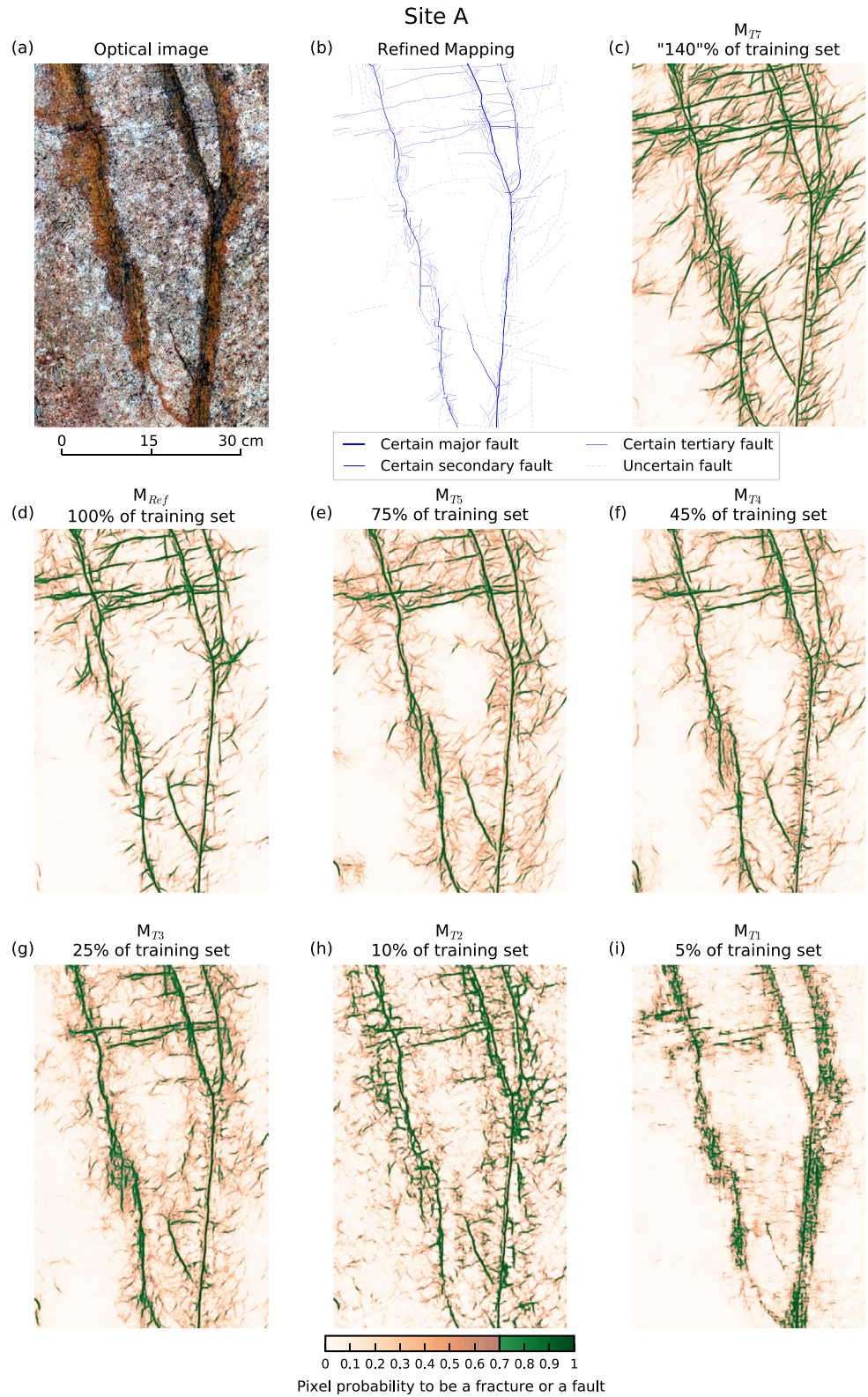
Therefore, while the larger model  $M_{Ref}$  is better performing from a metrics point of view and in its capacity to locate the major faults with high fidelity (narrow probability lines), the smaller models and especially the smallest 2-filter model provide realistic fracture and fault predictions down to the smallest scales, with an apparent greater capacity to predict the smallest cracks than has  $M_{Ref}$ . This likely results from the models having been trained with basic mapping ignoring small fractures. Because the adopted loss function penalizes any differences between the model output and the ground truth (regardless of whether or not the ground truth is accurate), any model trained on the basic mapping will be biased toward ignoring small-scale details that are absent in the basic ground truth. However, because these small-scale details are subtle, a more expressive model is more likely to learn this implicit bias than do the less expressive, smaller models.

#### 4.2. Sensitivity of Model Performance to Training Data Size

To assess the prediction capacities of the reference architecture, we explore the impact of the size of the training set on the performance of the models. To generate consistent comparisons, we use the same data sets as in the prior section. Within the total training zone in green (Sites A and B, Figures 3a and 3b), we randomly select subset zones for training, accounting for 5%–100% of the total training zone. This yields six models labeled  $M_{T1}$  to  $M_{T6}$ , with  $M_{T6}$  the case at 100% equivalent to  $M_{Ref}$ . For reference, the minimum training size of 5% covers an area of  $\sim 1.8 \text{ m}^2$  (109 tiles of  $256 \times 256$  pixels, i.e., 7 millions of pixels).

Figure S6 shows the resulting TI values, while Figures S7, S8, and 6 show the predicted faults, over the Site A validation zone and in selected regions (Figure 6). As expected, the performance of the models is greater for larger training sets (Figure S6). However, the performance saturates at an acceptable value ( $TI > 0.6$ ) onward from a moderate training size of 25% ( $M_{T3}$ ). This is also at that stage that non-tectonic features such as the ruler in the left bottom quadrant of the image are properly discriminated (Figure S7g). However, the performance is still reasonable with a very minimum training data size of 5%–10% ( $TI \sim 0.48$ – $0.58$ ,  $M_{T1}$  and  $M_{T2}$ ). This is even clearer on the fault images (Figures S7, S8, and 6); while fault lines are recovered in a sharper and most complete way and better localized with large training ( $\geq 45\%$ ), the major fault traces are predicted fairly well with a training set as small as 10% and even 5%. The major differences between the models are seen in the amount and organization of “noise” in the form of short fractures with low probabilities. While these supposedly fractures are disorganized for models trained with less than 25% of the training data, they progressively organize and simplify toward realistic patterns





**Figure 6.** Impact of training data size. (a and b) as in Figure 5. (d) shows predictions from  $M_{Ref}$  (i.e., 100% of the available training data), while (e–i) show predictions with decreasing amount of training data, and (c) predictions with additional training data (from Site C). See text for details.

with more training data (Figures S7d–S7f, 6d–6f, and S8c–S8e). Altogether these suggest that, while training the model with little data does not allow to make accurate predictions, it might be sufficient to predict major faults fairly well.

For reference, fracture and fault detection by the canny edge filter provides a TI of less than 0.05 while that from the GVG detector has a TI around 0.5 (Figure S6). Therefore, the CNN models presented here systematically outperform these edge detector algorithms even when a CNN model is trained with only a few images.

To further examine the impact of the training data size, we have trained another model ( $M_{T7}$ ) where the training data are expanded with the inclusion of the refined ground truth available at Site C (912 more tiles, thus equivalent to a size of  $\sim 140\%$  of that of A + B training size). The scores are then measured as before (Figure S6). The TI value reveals that, whereas the training data set has been significantly enlarged and enriched with higher resolution and accuracy expert mapping, the model scores do not improve. However, a greater number of actual short fractures are recovered (Figure S7c, 6c, and S8b), and those, being fairly well organized, seem to coincide with real cracks in the rock (most of which were not mapped). The capacity of the model  $M_{T7}$  to more accurately predict the smallest features is a result from the training including refined ground truth where many small cracks were labeled. Altogether these suggest that a moderate-size and moderate-resolution manual fault mapping might be used to train the model successfully to identify and locate the majority of fractures and faults but, possibly, the smallest ones. To predict the latter, the ground truth needs to include some of them.

#### 4.3. Sensitivity of Model Performance to “Quality” of Training Data

We here further examine the impact of the “quality” of the fault mapping training data set on the reference model ( $M_{\text{Ref}}$ ) results.

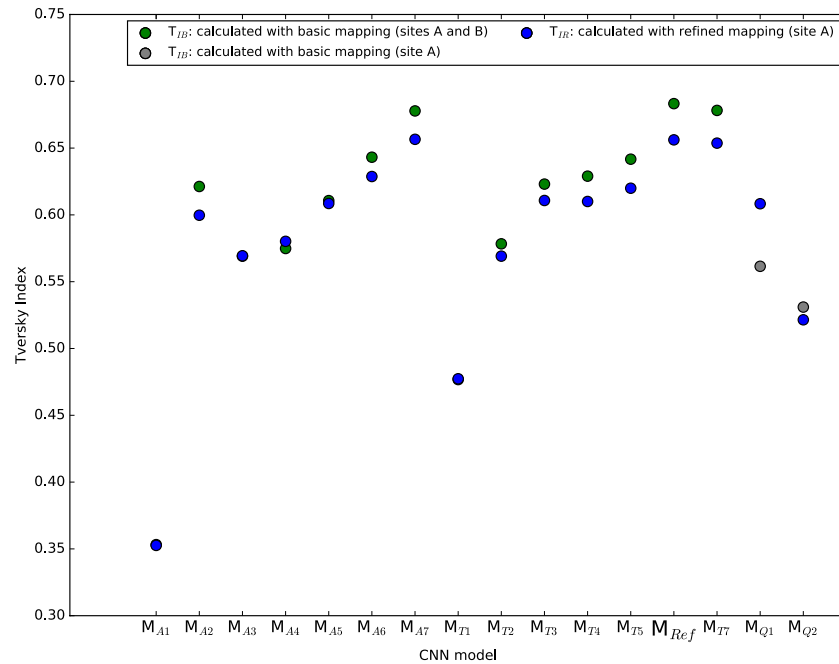
In Site A, we train two models  $M_{Q1}$  and  $M_{Q2}$  (training in the green zone in Figure 3a), the former with the refined ground truth map from the expert, the latter with the more basic ground truth map from the student. The scores of the models are then estimated in the validation zone of Site A, with respect to both the refined ( $TI_R$ ) and the basic mapping ( $TI_B$ ). The TI values (Figure 7) reveal that the model trained and validated with the refined mapping ( $M_{Q1}$ ) exhibits the best performance of the four realizations, with a TI of  $\sim 0.62$  against lower values (0.52–0.56) for the other models. Figures S9 and 8 visually confirm that, while both  $M_{Q1}$  and  $M_{Q2}$  predict the major faults well, their continuity is best revealed in  $M_{Q1}$ , which also recovers a greater number of short faults present in the refined ground truth.  $M_{Q1}$  actually provides a rich fault mapping that well represents the complexity of the dense fracturing and faulting observed in the raw image. The predictions are so detailed that the en echelon fault patterns, the small relay zones between faults and the intense fracturing in inner damage zones are well recovered (Figure S9b), while they are not or much less in the  $M_{Q2}$  predictions (Figure S9e).

Therefore, training the model with a “basic quality” mapping is sufficient to properly recover the principal fractures and faults. However, training the model with more accurate and higher resolution mapping is necessary to allow it to perform richer predictions.

#### 4.4. Reference Model

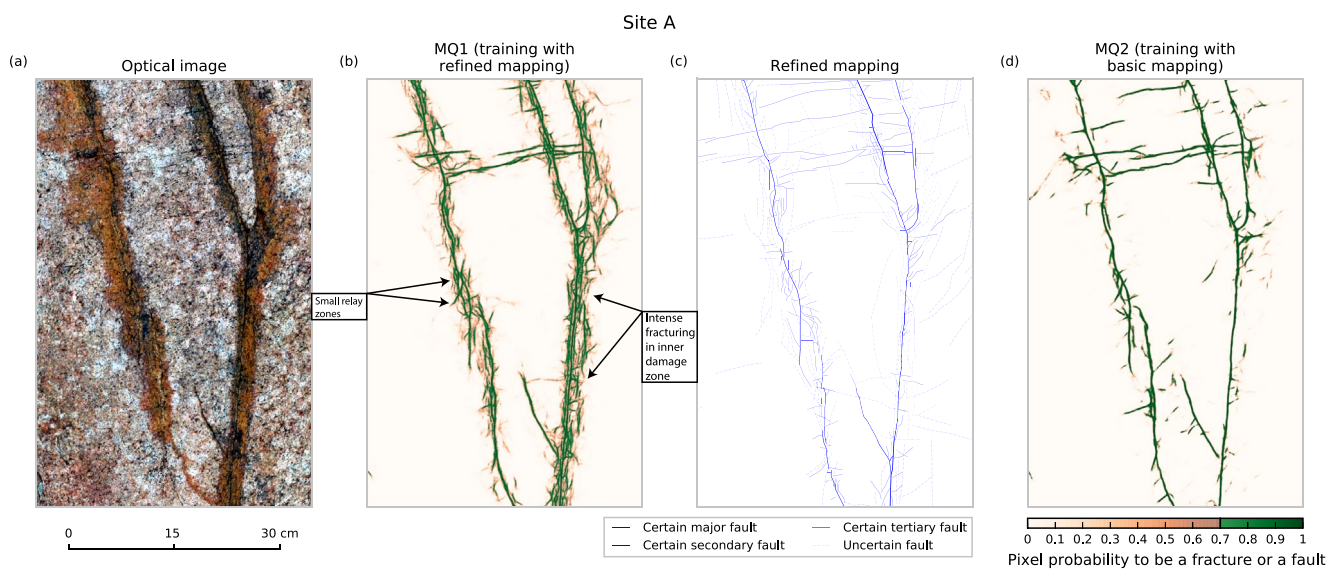
The scores of all the models calculated in the previous sections are presented in Table 2 and Figure 7. We remind the reader that the slightly slower  $TI_R$  values mainly result from the training being made on basic mapping, thus at a lower resolution than the refined map used for the  $TI_R$  calculation. The results demonstrate that, while most models provide reasonable results (most have a  $TI \geq 0.55$ ), the reference architecture  $M_{\text{Ref}}$  described in Section 3 provides the most accurate predictions (along with  $M_{A7}$ , yet see Figure S5 for its lower prediction capacities). Therefore, in the following, we use the reference model  $M_{\text{Ref}}$ .

Figure S10 shows the standard ROC curve for all the models. The ROC curves are evaluated with the basic mapping in Sites A and B (Figures S10a and S10b) and also with the refined mapping in Site A (Figure S10c).



**Figure 7.** Scores of all calculated models. See text for description of the different models. All models but  $M_{Q1}$  were trained with basic mapping (in both Sites A and B for all models but  $M_{Q2}$ ). Tversky index  $TI_B$  in green is calculated with reference to basic mapping in both Sites A and B (averaged values), while  $TI_R$  in blue is calculated with reference to refined mapping in Site A only.  $M_{Q1}$  and  $M_{Q2}$  models were trained with refined and basic mapping, respectively, in Site A only. The Tversky index in gray are calculated with respect to basic mapping.  $M_{Ref}$  provides the highest score, especially with respect to basic mapping.

In keeping with prior findings, the models trained with the smallest data sets ( $M_{T1}$  and  $M_{T2}$ ) have the lowest scores, while  $M_{Ref}$  and  $M_{T7}$  (largest training data set) have the highest True positive rates at all False positive rates, hence are the most performant models.



**Figure 8.** Impact of training data quality. (a and c) as in Figure 5. (b and d) show  $M_{Ref}$  predictions when  $M_{Ref}$  is trained with refined and with basic mapping, respectively (in Site A only). Clearly, training the model with more accurate and higher resolution ground truth allows it to predict more significant tectonic features and to provide greater details on these features. See for instance the finer details in the architecture of the inner damage zone of the fault branch to the right, and in the segmentation and arrangements of the fault traces.



**Table 2**

*Tversky Index of all Calculated Models Compared to Reference Model  $M_{Ref}$*

Model	$TI_B$ (reference to basic mapping in Sites A and B)	$TI_R$ (reference to refined mapping in Site A)
$M_{A1}$	0.35	0.35
$M_{A2}$	0.62	0.6
$M_{A3}$	0.57	0.57
$M_{A4}$	0.57	0.58
$M_{A5}$	0.61	0.61
$M_{A6}$	0.64	0.63
$M_{A7}$	0.68	0.66
$M_{T1}$	0.48	0.48
$M_{T2}$	0.58	0.57
$M_{T3}$	0.62	0.61
$M_{T4}$	0.63	0.61
$M_{T5}$	0.64	0.62
$M_{Ref}$	<b>0.68</b>	<b>0.66</b>
$M_{T7}$	0.68	0.65
$M_{Q1}$	0.56	0.61
$M_{Q2}$	0.53	0.52

*Note.* All models but  $M_{Q1}$  were trained with basic mapping.  $TIB$  is calculated with reference to basic mapping in both Sites A and B, while  $TIR$  is calculated with reference to refined mapping in Site A only.

The calculations were done using a NVIDIA GeForce GTX 1080 GPU with 8 GB available memory. With 64 filters, a model is trained over one epoch in about 9 min. As we run the model up to 44 epochs (Figure S2a),  $M_{Ref}$  was generally calculated in  $\sim 6$  h (though 10–12 h are needed to examine overfitting). Decreasing the number of filters to 32 decreases the training time by a factor of  $\sim 3$ , making the prediction calculated in  $\sim 2$  h.

## 5. Detailed Evaluation of Reference Model Fault Predictions

In the following sections, we evaluate the performance of the reference model  $M_{Ref}$  in more detail. Specifically, we examine its predictions for optical images from Sites A and B, and for previously unseen data of the same type and of different types (drone photography and optical satellite data).

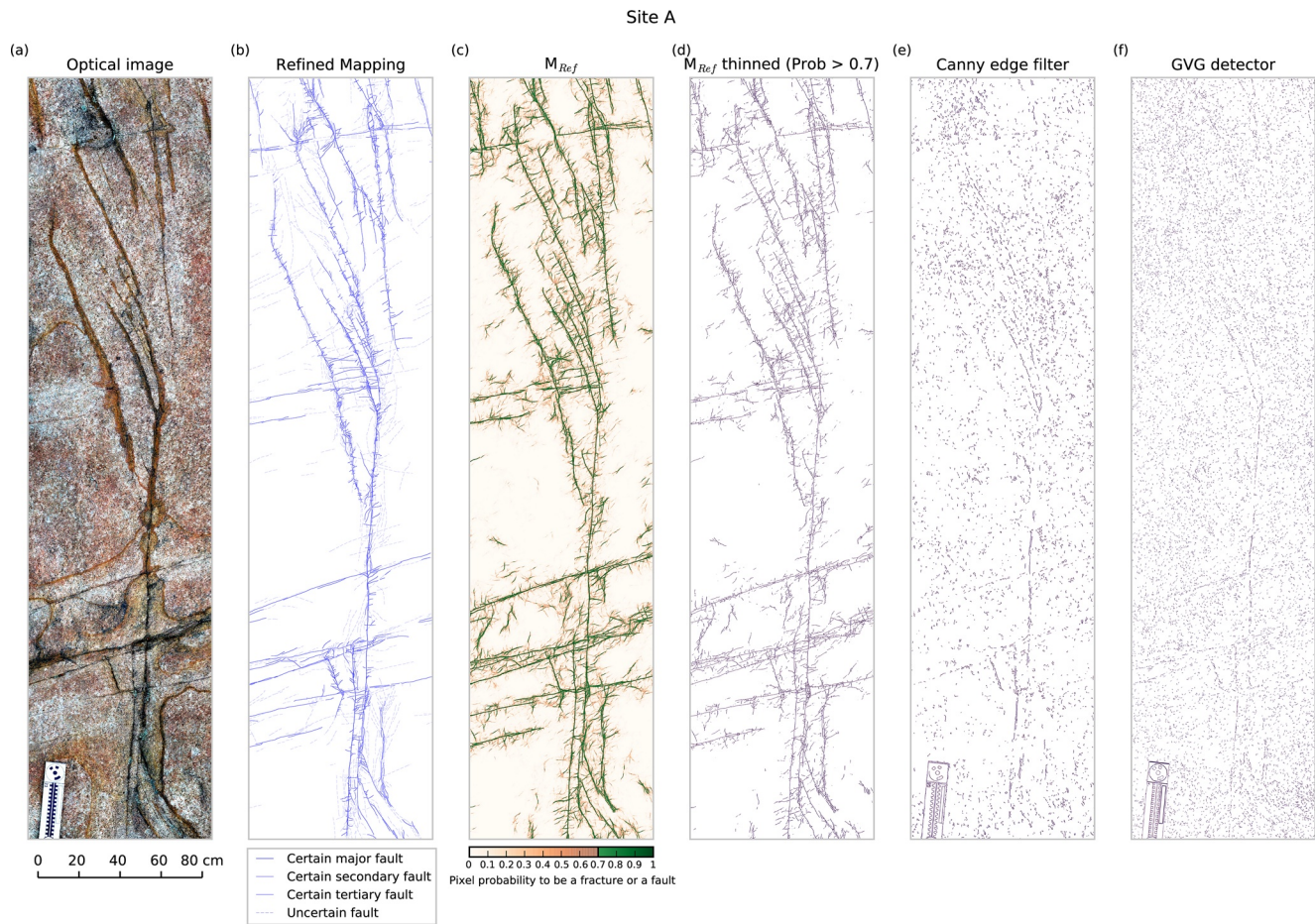
### 5.1. Results in Sites A and B

#### 5.1.1. Site A

Figure 9 shows the  $M_{Ref}$  model results in the validation zone of Site A, and compares them to the raw optical image and the refined ground truth fault map. A zoomed view was shown in Figure 5d. The range of probabilities of the predictions is shown as a color scale. In Figure 9, the predictions are also shown as “thinned lines” (see explanation below), and compared to the results obtained with the Canny edge filter and the GVG detector.

We observe that the predictions of the CNN  $M_{Ref}$  model compare favorably to the manual mapping, revealing both the main fault traces and the secondary faults. Furthermore, most fractures and faults labeled as uncertain by the expert are recognized by the model. By contrast, the two edge detectors fall short to detect the actual faults and fractures in the images. Even the main fault traces are not well recovered by these algorithms, and there is no continuity in the fault traces: they are split into numerous small, isolated segments. In the results given by the CNN  $M_{Ref}$  model the continuity of the faults and fractures is much better preserved.

The model results reveal the tectonics of the site with great richness (Figure 9; enlargements provided in Figure 10). While the model does not provide any interpretation, a fault expert can see that the general architecture of the central  $\sim$ “N-S” fault system is well recovered, showing the main fault trace splaying into long, oblique secondary faults at both tips. Another fault set, trending  $\sim$ “ENE,” is also well revealed. The architecture of smaller faults is likewise well depicted, with the model predicting remarkably well their tip splays (Figure 10a), their en echelon arrangement (Figure 10b), the relay zones between faults or segments (Figure 10c), the small bends in fault traces (Figure 10), the strongly oblique fracturing in between overlapping segments (Figure 9), and the dense sub-parallel fracturing forming inner damage either sides of major fault traces (Figure 10d). From the geometry of the en echelon arrangements and pull-apart type-relay zones along the  $\sim$ “N-S to NW-SE”-trending faults, we infer that the latter have a right-lateral component of slip. Although evidence is less numerous along the  $\sim$ “ENE”-trending faults, a few relay zones and en echelon dispositions suggest that these faults also have a right-lateral component of slip. That the two fault families have a similar slip mode suggest they are not coeval. Actually, although there is still ambiguity, several lines of evidence in both the refined map and the model predictions suggest that the  $\sim$ “N-S to NW-SE” faults cross-cut the  $\sim$ “ENE” faults, and thus post-date the latter (Figure 10f). Finally, the model predictions also recover fairly well the sub-perpendicular very short cracks identified by the expert along most significant fault traces (Figure 10e). Analyzing further the tectonic significance of the results is beyond the scope of the study. However, we demonstrate here the amazing capacity of the model  $M_{Ref}$  to map fractures and faults as accurately as the expert, and even in greater details in some places.



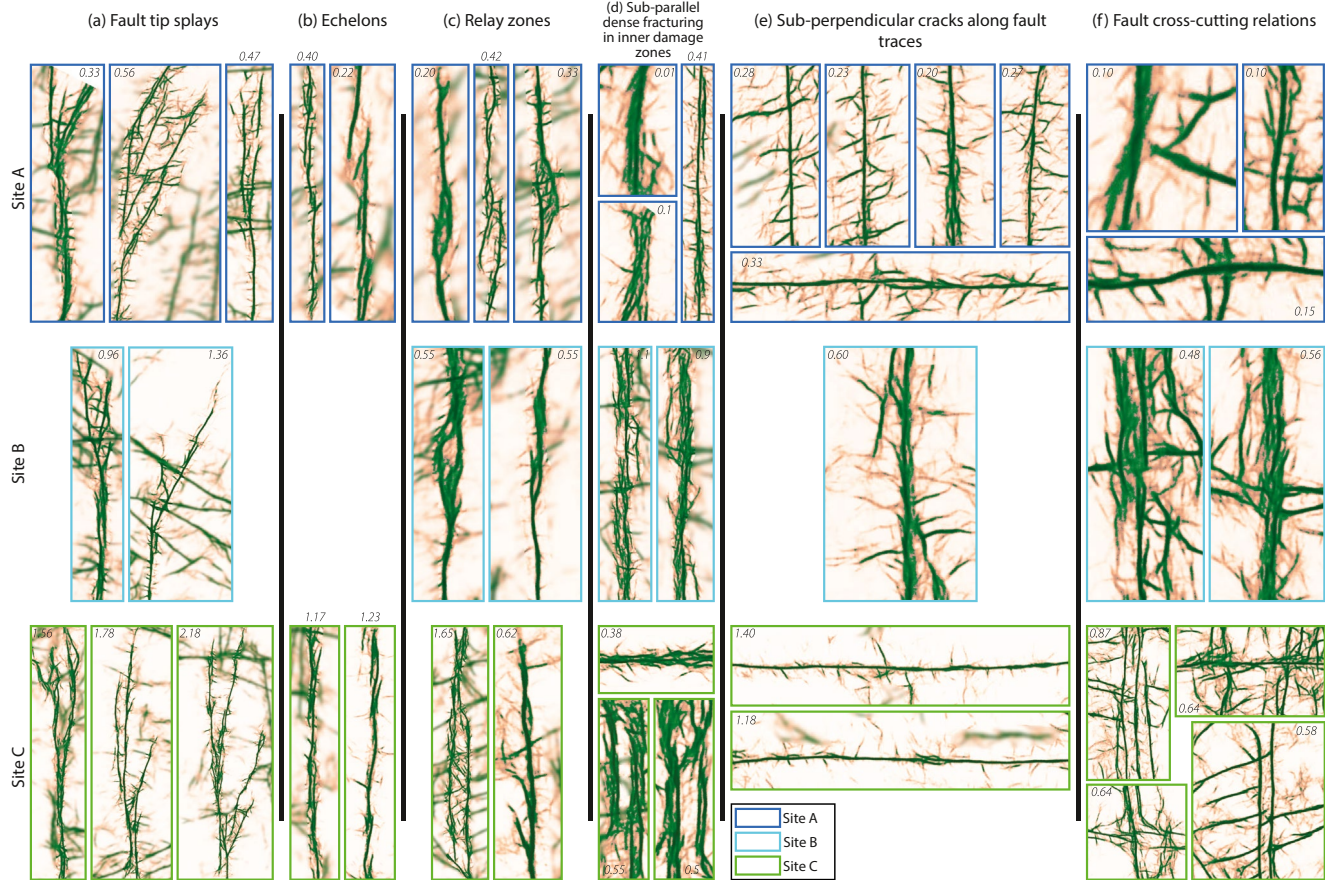
**Figure 9.**  $M_{ref}$  predictions on Site A. (a) Optical image of validation zone in Site A (see Figure 3a); (b) Refined mapping, with fault hierarchy indicated (better seen when zooming); (c) Predictions from  $M_{ref}$ ; (d)  $M_{ref}$  predictions with probability  $> 0.7$  thinned for vectorization (see text); (e) Predictions from Canny edge filter algorithm; (f) Predictions from GVG detector algorithm.

To further analyze the validity of the results, we compare the spatial density and the orientation of the fractures and faults in both the ground truth and the  $M_{ref}$  probability map, in the validation zone shown in Figure 9.

A prerequisite for this analysis is to transform the predicted probability maps into fault vectors to make them comparable with the ground truth fault lines (rasterized into 1-pixel wide lines). For this purpose, we process the CNN predictions as follows: (a) the model probability map at a given probability threshold, here  $> 0.7$ , is converted into a discretized binary map (i.e., a value of 1 is given to any pixel with a probability  $> 0.7$ , and a value of 0 is given to any other pixels); (b) this binary map is subsequently thinned down to faults traces of 1-pixel wide representations: the pixels from the borders of the fault traces are removed iteratively until no more pixels can be removed without disconnecting the pixels from each other (Guo & Hall, 1989; Lam et al., 1992) (Figure 9d); (c) finally, we vectorize the binary thinned map to obtain a separate line for each fault (GRASS Development Team, 2020). While this vectorization cuts the faults anywhere they meet another fault, which may be inadequate (see Discussion section), it provides a first-order view of the fault distribution and characteristics. To make this overview most relevant, we have removed the predicted fractures of less than 2 pixels length, as they are clearly noise.

Figure 11A (top row) compares the (a) fault spatial densities, (b) strikes, (c) strikes weighted to cumulative lengths, and (d) lengths among refined ground truth and model predictions (for probabilities  $> 0.7$ ). The fault density pattern is consistent overall among the two sets, with a tendency of the model, however, to identify more faults than the refined map when the latter only includes certain faults, and less faults than





**Figure 10.** Details of tectonic predictions. The figure shows zooms from  $M_{Ref}$  predictions in Sites A, B, and C accurately revealing relevant features in fault patterns such as splaying secondary (a) faults at master faults tips, (b) en echelon fault segments, (c) extensional relay zones between fault segments, (d) dense sub-parallel fracturing in inner damage zones, (e) sub-perpendicular cracks along fault traces, and (f) cross-cutting relations among different fault families. To best emphasize these tectonic features, we have blurred the adjacent parts of images. We have also rotated some of the views so that their orientation differs from their original one in Figure 9. The number in italic in each box is the longest length of the box frame, in meters. The observed fault patterns are basically scale-independent.

the refined map when the latter includes both certain and uncertain faults. The former case results from  $M_{Ref}$  producing predictions more detailed than the refined, yet moderate resolution ground truth, while the latter case results from  $M_{Ref}$  training done with basic ground truth lacking details on small features.

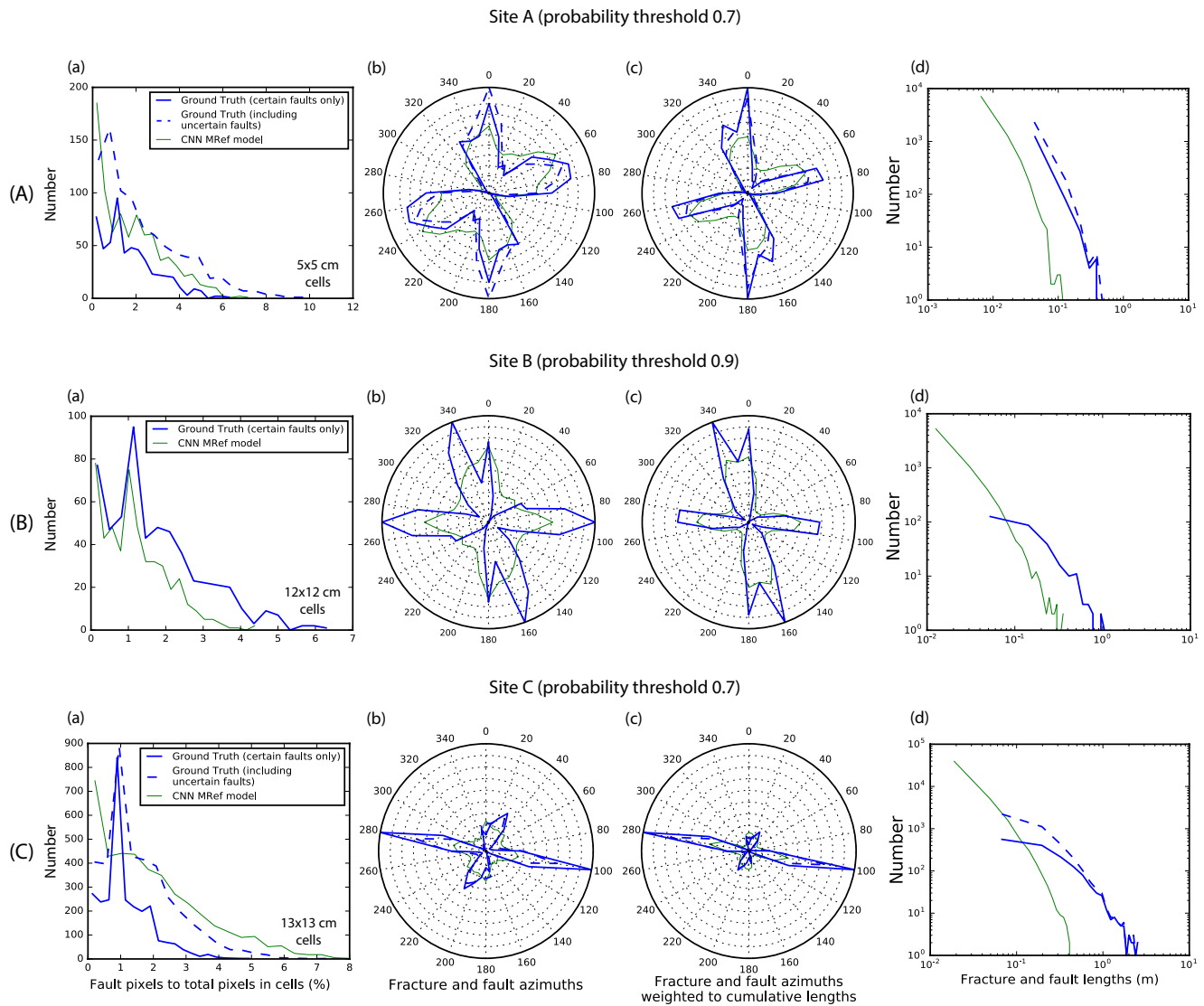
The distribution of the fault azimuths is also consistent among the ground truth and the predictions, especially when the azimuth distribution is weighted to the cumulative length of the faults. This shows that, while the model recovers a few unmapped fractures of various azimuths that are likely noise or fabric inside the granite (Figure 11A (b)), it primarily recovers actual fractures and faults organized in two dominant “N-S to NW” and “ENE” families (Figures 11A (c)).

By contrast, the model does not seem to recover the fault lengths properly (Figures 11A (d)): compared to the ground truth, it produces a larger number of very small faults, and a smaller number of larger faults. However, this is not necessarily inherent to the CNN model, but likely attributable to the vectorization process that improperly divides some of the fault traces and thus introduces artifacts in the recovered fault lengths. We come back to this issue in the Discussion section.

### 5.1.2. Site B

Figures S11 and 12 compare the raw optical image, the basic ground truth fault map and the CNN  $M_{Ref}$  prediction in the validation zone of Site B, and in a selected region, respectively. The prediction as thinned lines and the results of the canny edge filter and GVG detector are also shown on Figure S11.

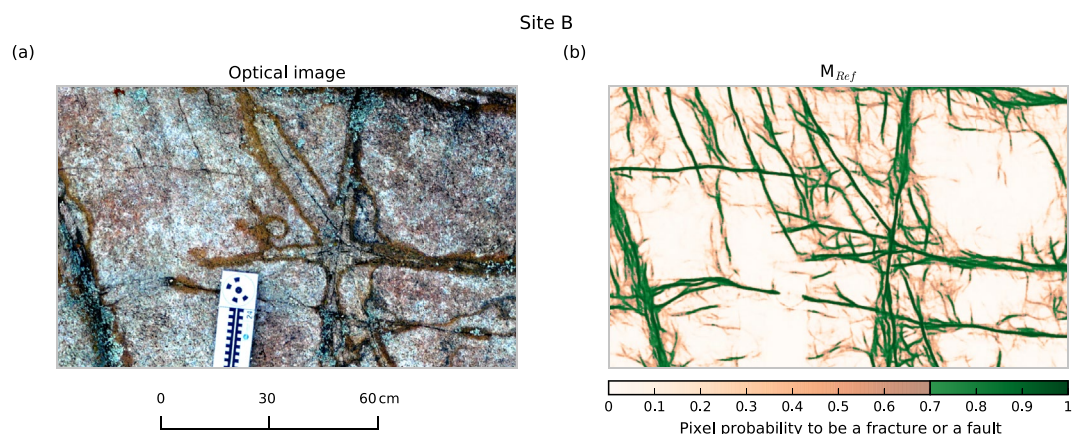




**Figure 11.** Fault statistics among ground truth and M<sub>Ref</sub> predictions. A, B, and C refer to as Sites A, B, and C, respectively. Ground truth data in blue, and M<sub>Ref</sub> predictions in green. In Sites A and C, the predictions considered are those with a probability >0.7, while in Site B, they are those with a probability >0.9. Thus, in the prediction, any pixel with a probability >0.7 (or 0.9 for Site B) is labeled as a fault pixel. (a) Overall fault density. It is expressed as the number of fault pixels normalized to the total number of pixels within cells of 100 × 100 pixels (corresponding to 5 × 5, 12 × 12, and 13 × 13 cm for Sites A, B, and C, respectively); (b) distribution of fault strikes. As faults are short, their azimuth is measured along the line connecting to two fault ends; (c) distribution of fault strikes weighted to cumulative lengths; (d) distribution of fault lengths. See text for discussion.

As for Site A, the M<sub>Ref</sub> model well predicts the major faults seen in the optical image, and provides a similar mapping as the ground truth, where both major and secondary faults and fractures are correctly predicted, albeit with much more details than in the basic ground truth. In contrast, the two edge detector algorithms do not recover the faults; they actually split the fractures and faults into a large number of small segments so that even the major faults are not identifiable as continuous features.

The predictions well highlight that three fault families co-exist in the site, trending ~“N-S”, ~“NW-SE” and ~“E-W” (remember that, as the data are not georeferenced, these orientations are arbitrary). As in Site A, the model recovers the incredibly fine details of the faults, such as their tip splays (Figure 10a), the relay zones between faults or segments (Figure 10c), the small bends in fault traces (Figure 12), the dense sub-parallel fracturing on either sides of major fault traces (Figure 10d), and sub-perpendicular short cracks along the principal fault traces (Figure 10e). There are very few geometric indicators of slip mode, and those suggest that the ~“E-W” faults might have a right-lateral component of slip. Although this remains



**Figure 12.**  $M_{\text{Ref}}$  predictions on Site B (partial view). (a) Optical image of part of validation zone in Site B (see Figure 3b). (b) Predictions from  $M_{\text{Ref}}$ . See text for discussion.

to be analyzed in detail, the model predictions suggest that the ~“N-S” faults cross-cut the ~“E-W” and the ~“NW-SE” faults, and might thus post-date them (Figures S11 and 12).

To allow the best comparison between the predictions and the ground truth, we analyze the faults predicted with a probability  $\geq 0.9$ . The fault density pattern is similar overall among the ground truth and the predictions (Figure 11B (a)), but the model predicts less “certain” faults than the ground truth. Like for Site A, while a small number of fractures with various azimuths are produced in the model (Figure 11B (b)), when weighted to the cumulative fault length, the azimuths of the predicted faults are in good agreement with the actual fault strikes (Figure 11B (c)). Here again, the vectorization procedure has divided some of the faults improperly, resulting in a too large number of small pieces compared to the ground truth (Figure 11B (d)).

In summary, these comparisons demonstrate that the CNN  $M_{\text{Ref}}$  predictions properly recover the distribution and architecture of the actual fracture and fault networks, down to their smallest features of  $\sim 1$  cm long. By contrast, the vectorization procedure we used to convert the predicted fault probability maps into vectors is not entirely appropriate as it does not recover the actual fracture and fault lengths (see Discussion section).

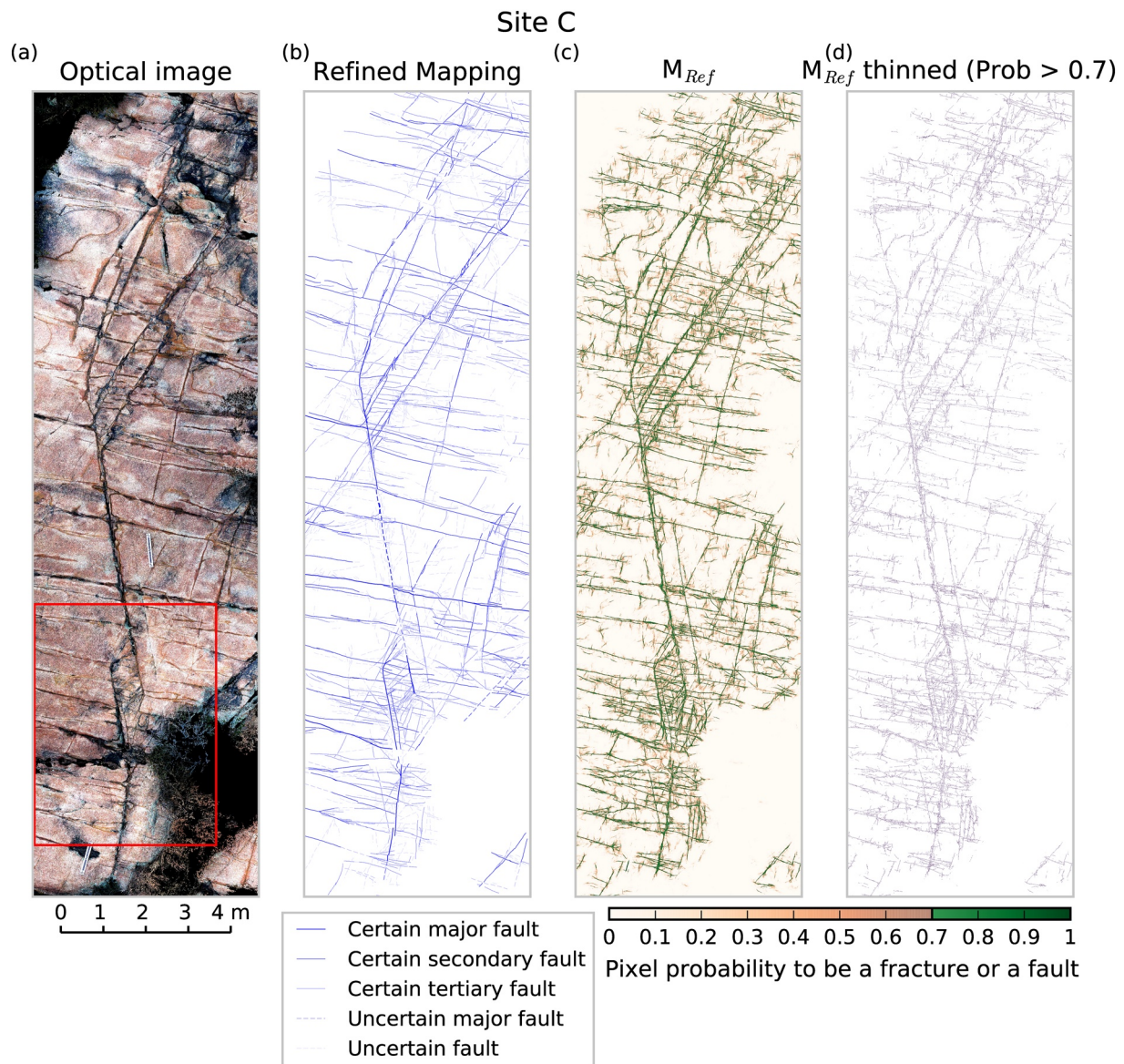
## 5.2. Predictions in Unseen Data of Similar Type

Here, we explore the accuracy of  $M_{\text{Ref}}$  to predict fractures and faults in another distant site, Site C, of which the data were never seen during the training or validation. Site C is located  $\sim 700$  m away from Sites A and B (Figure 2). The only commonalities between the data of Site C and those of Sites A and B is the nature of the optical images, which were also acquired “on the ground” with an hand-held camera (Table 1, “ground photogrammetry”), and the lithology (granite). The morphology and texture of Site C images are partly different however from those of Sites A and B (Figure 3).

Figures 13 and 14 show the predictions compared with the optical image and with the refined ground truth map, over the entire site and in a zoom, respectively. The thinned predictions are also shown, calculated at probabilities  $> 0.7$ .

We see that the  $M_{\text{Ref}}$  model trained on distant sites succeeds to detect and locate most of the fractures and faults visible in the image and mapped by the expert. The major fault traces are correctly predicted with high probabilities ( $\geq 0.9$ ) while the secondary and tertiary structures are predicted with lower probabilities, yet generally  $> 0.7$ .

The model actually recovers the tectonics of the site in great detail. Three fault families are revealed, trending  $\sim$ “NNW,” “NE,” and “WNW.” A major “NNW” fault system extends at the center of the site, and its architecture is well recovered: while the fault system has a sub-linear trace in its central part, it ends in the “North” by splaying into a few long “NE”-trending splays. In the “South,” the fault steps laterally to

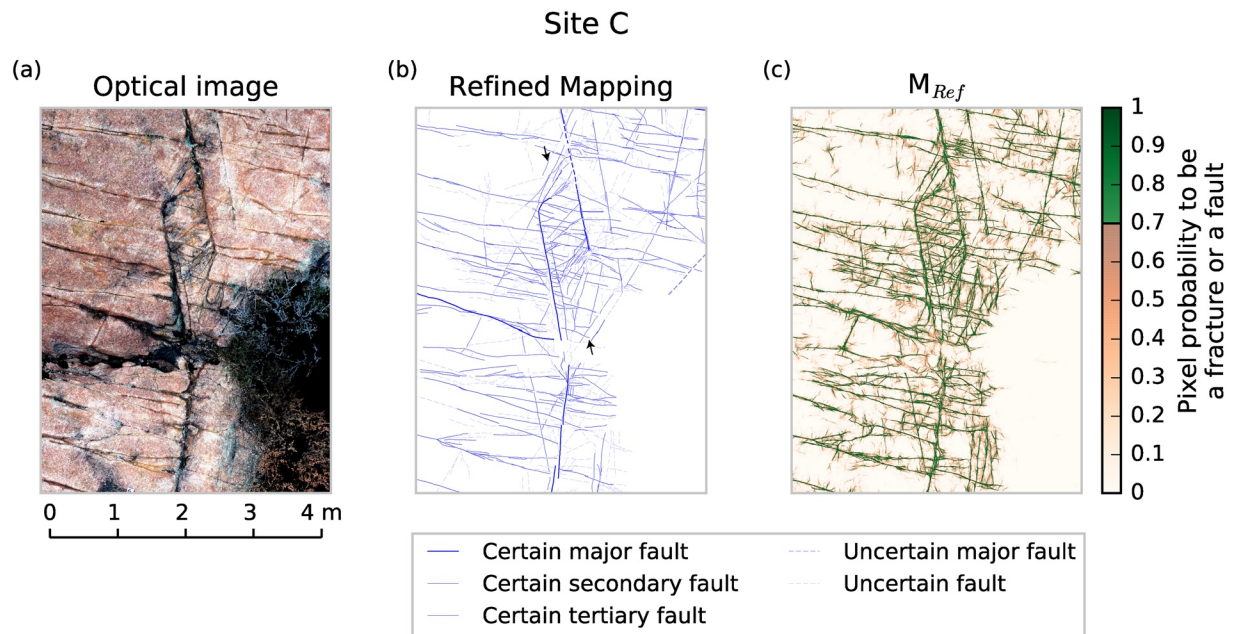


**Figure 13.**  $M_{ref}$  predictions on Site C (whole site). (a) Optical image of Site C (see Figure 3c); red box locates Figure 14; (b) Refined mapping, with fault hierarchy indicated (better seen when zooming); (c) Predictions from  $M_{Ref}$ ; (d)  $M_{Ref}$  predictions with probability > 0.7 thinned for vectorization. See text for discussion.

a next segment, and the extensional pull-apart-type relay zone in between the two segments (arrows in Figure 14b) is remarkably well recovered (Figure 14c), attesting to a right-lateral component of slip on the central “NNW”-trending main faults. As in Sites A and B, the model provides fine details that allow identifying secondary splay faults (Figure 10a), en echelon segments (Figure 10b) and relay zones (Figure 10c) between faults or segments. Altogether these geometries suggest a right-lateral component on the ~“NNW” faults, a right-lateral component on the ~“WNW” faults, and a left-lateral component on the ~“NE” faults. In terms of relative chronology, the ~“NNW” central fault system is revealed as cross-cutting the “WNW” fault set (Figures 13 and 14).

The fault density pattern is consistent overall among the ground truth and prediction (for probabilities > 0.7), even when uncertain faults are taken into account. The model actually predicts as many detailed well-constrained fractures as the expert but in the zones that are hardly fractured (Figure 11C (a)). The azimuth distributions are also consistent (Figure 11C (b)), especially when the azimuths are weighted to the





**Figure 14.** Mref predictions on Site C (partial view). (a) Optical image of part of Site C (location in Figure 13a); (b) Refined mapping, with fault hierarchy indicated (better seen when zooming). Black arrows indicate the pull-apart relay-zone between the two main ~"NNW"-trending fault branches; (c) Predictions from  $M_{Ref}$ . Note the great tectonic details predicted in the pull-apart relay zone.

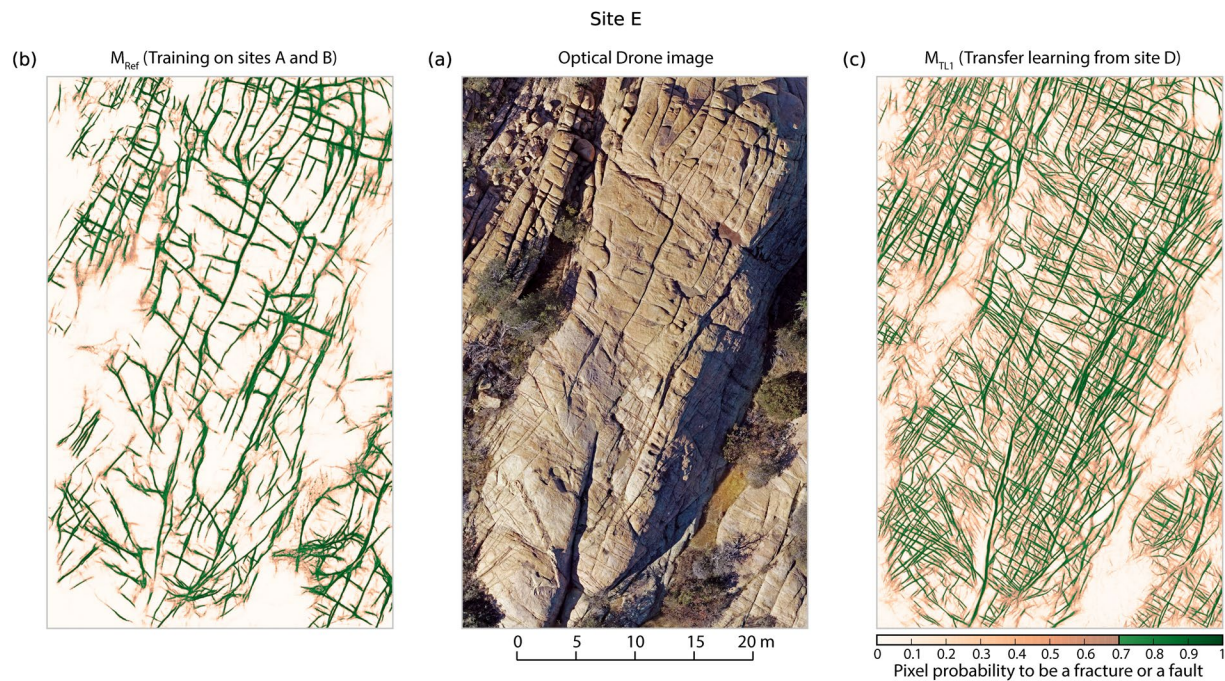
cumulative fault lengths (Figure 11C (c)). By contrast, as before, the vectorization procedure does not allow recovering the actual fault lengths properly (Figure 11C (d)).

### 5.3. Predictions in Unseen Data of Different Type

Here we examine the performance of the reference model  $M_{Ref}$  to predict fractures and faults in other images and DSM data, of different type and previously unseen by the model (Table 1). These new data are (1) optical images (and DSM) acquired from a drone at Sites D and E, in the same area as Sites A and B but distant by ~1 km (Figures 2 and 3), (2) optical satellite Pléiades images (and DSM) acquired at Sites F and G in the same area as Sites D and E (Figures 2, 3f, and 3g), and (3) Pléiades images (and DSM) acquired at remote and different Sites H and I in a far distant and different site (Table 1 and Figures 2c, 3h, and 3i). The ground truth in all these new sites but H is a basic, low resolution mapping.

We first apply  $M_{Ref}$  to these new data, with no additional training. We then produce a set of upgraded models ( $M_{TL1}$  to  $M_{TL3}$ ), where  $M_{Ref}$  is enriched (through transfer learning or "fine-tuning," e.g., Pratt et al., 1991) with additional training using ground truth from the new sites. The ground truth mapping in Sites D to I being incomplete and of low resolution, we do not calculate the Tversky index of the models and only provide a visual analysis of the predictions.

Figures S12 and 15 show the predictions of  $M_{Ref}$  and  $M_{TL1}$  in Site E (in full and a selected region, respectively), never seen during the training or the validation, and compare them to the optical image. While the results of  $M_{Ref}$  are not perfect, the model predicts the major faults fairly well. In particular, it recovers the two dominant fault families, ~NNE-SSW and NW-SE, identifies the zones with greatest density of fracturing, and recognizes and ignores most of the vegetation. However, the fault traces are represented with several pixel-width lines with high probabilities, suggesting uncertainties in the fault trace location. With transfer learning and additional moderate amount of training on 678 images (along with 154 for the validation) and basic ground truth from Site D (training and validation zones in Figure 3d), the  $M_{TL1}$  model produces even better results: the great majority of the visible fault traces, major and minor, are now recovered, and mapped as sharp, narrow lines well expressing the details of the actual fault geometries and distributions. Figure 15 clearly shows some of these details. Among them, dense tip splay faults and cross-cutting networks are well



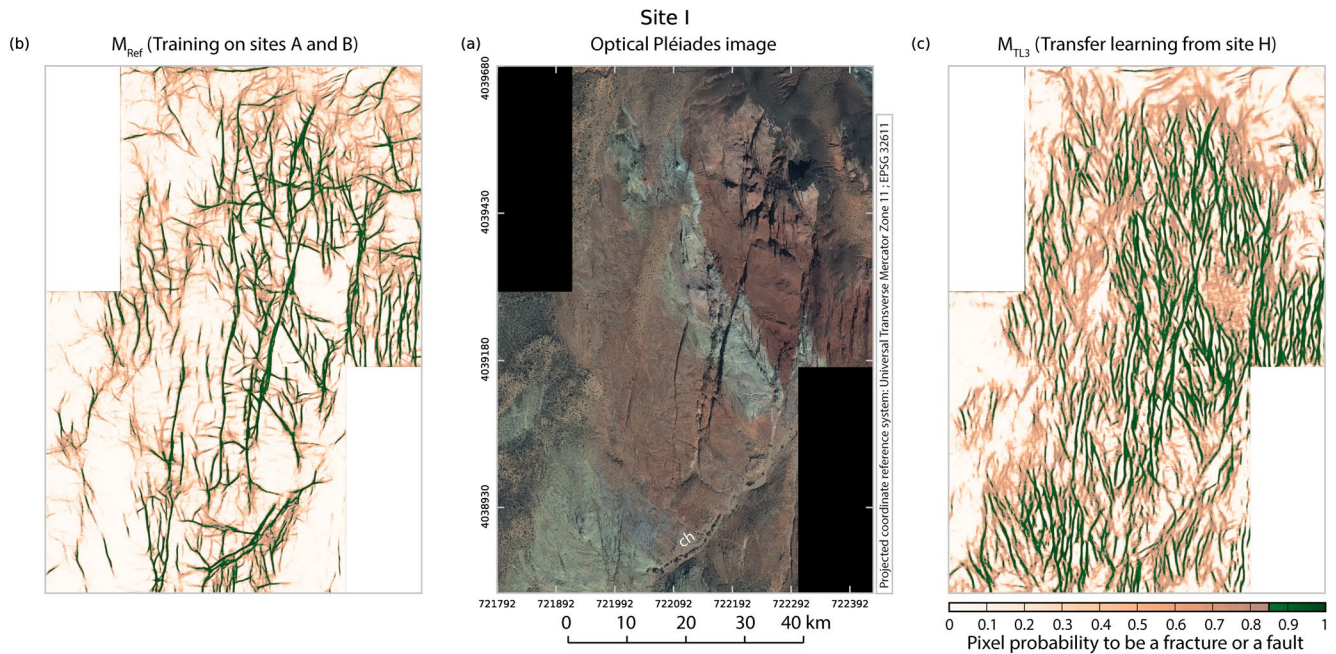
**Figure 15.** M<sub>ref</sub> predictions on Site E (partial view). (a) Drone optical image of part of Site E in Granite Dells (entire site shown in Figure S12); (b) Predictions from M<sub>Ref</sub> trained on ground images in distant Sites A and B in Granite Dells as before; (c) Predictions from M<sub>Ref</sub> trained with additional learning from Site D in Granite Dells. Ground truth in Site E was done on drone data, different from the data used to train M<sub>Ref</sub>. While the predictions of M<sub>Ref</sub> alone well recover the major faults in the drone data at Site E, the additional training of M<sub>Ref</sub> with a few drone data and fault traces at Site D allows the predictions to recover most existing faults at Site E.

identified, while major fault traces are well discriminated. Although this remains to be analyzed in greater details, several indicators (relay zones, horsetails) suggest that the NNE faults have a left-lateral component of slip. The slip kinematics is more uncertain on the NW faults. The NNE faults generally offset the NW features, but opposite relations are observed at some places, suggesting the possible coeval slip of the two fault sets. The fault distribution is uneven, with greater secondary fracture and fault densities adjacent to faults with longest and most pronounced traces.

The application of M<sub>Ref</sub> to the Pléiades image at Site G (Granite Dells) (never seen during the training and the validation) is less satisfactory (Figure S13); the model only succeeds predicting the major faults with probabilities  $\geq 0.7$  and a few secondary faults with lower probabilities (in range of 0.5–0.7), but fails to predict most minor faults. This is likely due to the significant change in resolution among the images used for training and those used here for testing. With transfer learning and moderate additional training on Site F (training and validation zones in Figure 3f, including 30 and 6 images, respectively) the predictions are more accurate (model M<sub>TL2</sub>, Figure S13c), with the two dominant fault families well detected, and vegetation, roads and houses discriminated. By contrast, as in Site E before, the fault traces are represented as several pixel-width lines of high probabilities, indicating that the model lacks some confidence on actual fault details. The model inability of capturing details is also seen in the widespread low probabilities in the whole site: while the model infers that many small faults are present in the site, it is not certain of their existence or precise location. This suggests that the additional training that was done to transfer to the new data domain was based on too few samples (too few new Pléiades images).

Finally, in far distant Site I (Figure 16) (never seen during training), M<sub>Ref</sub> predicts most major faults fairly well with high probabilities ( $\geq 0.85$ ), as well as a number of smaller faults with lower probabilities (0.3–0.85). However, the model fails discriminating non-tectonic features such as stream channels (Ch, Figure 16b). The transfer learning (model M<sub>TL3</sub>, with additional training on 63 images and refined ground truth from Site H, Figure 3h) greatly improves the identification of both the major and minor faults, and makes the model discriminating non-tectonic features (Figure 16c). The architecture of the fault systems is well recovered,





**Figure 16.**  $M_{\text{ref}}$  predictions on Site I (test zone). (a) Pléiades satellite optical image of Site I, in Valley of Fire (location in Figure 3i). Ch for stream channel; (b) Predictions from  $M_{\text{Ref}}$  trained on ground images in far distant Sites A and B in Granite Dells as before; (c) Predictions from  $M_{\text{Ref}}$  trained with additional learning from Site H (Pléiades data in Valley of Fire). As the images, rocks and faults are here very different from those seen by  $M_{\text{Ref}}$  over its training, the prediction capacities of  $M_{\text{Ref}}$  alone are fairly low (b). However, with modest additional training from Site H (Pléiades images in Valley of Fire), the  $M_{\text{Ref}}$  predictions improve and allow recovering most existing faults at Site I (c).

highlighting the dense networks of NW-trending secondary faults splaying off a few NNE master faults. This architecture might suggest a left-lateral component of slip on the NNE master faults, as recognized in earlier works (de Joussineau et al., 2007).

Concluding this section, we observe that even with no additional training, the model identifies the main tectonic features in completely new and differing data sets. Fine-tuning the model with additional training on a small amount of new data greatly improves the quality of the predictions, making the deep learning model  $M_{\text{Ref}}$  a portable tool that can be applied to new data with little effort.

## 6. Discussion

### 6.1. U-Net Appropriate for Fracture and Fault Detection in Optical Images

Our work demonstrates that the proposed U-Net architecture ( $M_{\text{Ref}}$ ) is appropriate to automatically detect and map tectonic fractures and faults accurately in optical images and topographic data. As expected, the model performs better when applied to data of similar type as those used for the training, yet is satisfactory when transferred to other data types. Clearly, our model outperforms conventional edge detectors, and when properly trained, the model recognizes non-tectonic features such as vegetation, roads, buildings, stream channels, shadow contours, erosion traces, rock boundaries and the ruler present in some images, classifying them accordingly. The model predictions are in agreement with both the manual fault maps and the lineaments directly observable in the raw optical images. The agreement is first attested by the coincidence of the fracture and fault density and azimuth distributions among the ground truths and the predictions. It is also attested by the great coincidence between the expert fault mapping and the predicted fault traces: all major faults identified by the expert are well predicted by the model, with the complex geometry of their traces properly recovered, down to their smallest-scale details (Figure 10). Most minor faults mapped by the expert are also well recovered despite of their tenuous traces, while the model results verify that minor cracks manually mapped as “uncertain faults” are actual fractures. Furthermore, most of the small features predicted with low probabilities in unmapped zones reveal to coincide with observable yet



unmapped cracks in the rocks. Therefore, the fault maps produced by the model are tectonically relevant, and actually of high tectonic quality. They outperform the basic mapping accuracy and completeness, and provide more details than the expert fault maps. Furthermore, they outperform the manual mapping in timeliness since they can be generated in less than a few hours (for training) to a few seconds (for application) using a standard computer.

It has to be noted that, recently, Jafrasteh et al. (2020) used a Generative Adversarial Network to extract fractures and faults in our optical image data at Site A (no topography included). Although training was done with our refined mapping (contrary to  $M_{\text{Ref}}$  that was trained with basic mapping), the predictions are not as accurate as the U-Net  $M_{\text{Ref}}$  predictions (see their Figure 4c). While major faults are identified, they are divided into small pieces; minor faults are not well predicted; the predictions are noisy; and the model does not discriminate the pebble in the lower part of the image. This confirms the superiority of U-Net to automate fracture and fault mapping in optical image data.

## 6.2. Interpreting Learned Characteristics of Faults and Fractures

The model predicts faults as well as fractures, although the trace complexity of the former is much higher than that of the latter. Indeed, while fractures have linear and short traces, faults have more complex, curvilinear and long traces, including continuous and discontinuous sections, commonly stepping from one another or connecting to adjacent faults (Figure 1). The U-Net model employed here is well adapted to capture such complexities because it includes thousands to millions of parameters among thousands of convolution filters that, combined, “scan” every detail of the fault trace expressions in the images. While a moderate number of filters is sufficient to capture the principal characteristics of the fault traces (see Figure 5 with small number of filters), a greater amount of filters allows recovering greater details. Although it is not practical at present to determine the nature of the filters and their combinations, their success to identify fractures and faults demonstrates that some of them contain physical information on fractures and faults; that is, a measurable information that characterizes some of the actual fracture and fault properties (geometrical, mechanical, kinematical, etc.). Therefore, future works that visualize the CNN filters will undoubtedly contribute to learning new information on the physical characteristics of fractures and faults.

In an attempt to learn more on the fault physical characteristics encapsulated within  $M_{\text{Ref}}$ , we have performed an additional test. We examine the impact of the topography information on  $M_{\text{Ref}}$  predictions. Figure S14 shows  $M_{\text{Ref}}$  predictions on Site C when the training data includes (Figure S14c), and does not include (Figure S14d) the topography of the site. We remind the reader that fractures have no topographic imprint at the ground surface, while faults in Site C have a small but measurable topographic expression (Figure S14a). However, even though this expression is subtle, the model results differ significantly: while the model trained with no topography excels in recognizing the sub-linear and continuous fracture/fault traces in the image, it does not penalize any trace and thus does not recover the variable expression of the fault traces along their length, and it populates the site with a multitude of short lineaments that are likely noise (Figure S14d). In contrast, when the topographic property of faults is learned by the model, the predictions become more cautious in assessing that a sub-linear and continuous trace is a fracture or a fault. This demonstrates that the model learned to recognize long-range correlations that define faults and include topographic information in its assessment. When the topographic expression of faults is more pronounced as in Site E (Figure S15), the results differ even more significantly. The model trained with no topography (Figure S15c) marks any sub-linear trace as a fault/fracture, and eventually provides a uniform mapping where all sub-linear traces, including non-tectonic features, have a similarly high probability to be a fault. In contrast, the model trained with topography (Figure S15b) assigns a high probability to the major faults only (for they have vertical slip and thus a topographic imprint), allowing their clear emphasis.

Deciphering which physical properties of fractures and faults are recovered by CNNs is clearly a challenge for future research. All of the relevant information required for the predictions is retained in the low-dimensional latent space of the model. In the present  $M_{\text{Ref}}$ , when the spatial information is decomposed ( $8 \times 8$  pixels), there are 512 numbers comprising the latent space that precisely encode the input data, and hence the fractures and faults. This is a reasonable data size that could be investigated with clustering algorithms like t-SNE (Maaten & Hinton, 2008), which are capable to extract and group structures in the latent space

associated with a particular expression of the fault zone. In this way, a fault expert could identify fault zones sharing similarities and other fault zones appearing markedly different, and visually compare them for a qualitative interpretation of which characteristics define one type of fault zone or another. Moreover, this could help to visually evaluate whether these characteristics are physical or artificial (e.g., lighting conditions). In the former case, we might gain insights on the fault physical properties, while in the latter case we might learn how the U-Net model proceeds to identify and discriminate the faults.

### 6.3. Conditions for Model Generalization

Our results show that  $M_{\text{Ref}}$  has good generalization capacities, in spite of the fact that the training data set was modest in size and of moderate quality (basic mapping at low resolution and accuracy). This shows that training the model with a few rough fault maps accessible to most users is sufficient to obtain good predictions on a large zone. The capacity of the model to provide fair results with minimum training likely arises from the present algorithm more needing to get homogeneous than numerous training data: there is little need to repeat the same information to the algorithm as it learns very quickly, but it is necessary to provide it the information at the various scales that might be relevant (here faults and fractures from largest to smallest scales). It is also needed to learn the model with the differences (brightness, colors, contrast, etc.) that the various images might exhibit. Indeed, while the model has strong prediction capacities, it cannot predict features it has not learned. Although minimal, the expert mapping, including diverse samples, is thus mandatory for training. Even though the model cannot “discover” new features that were never learned, it can recognize many more features of the types it has learned than would an expert or any other user. This large prediction capacity is a form of generalization.

We have shown that the model  $M_{\text{Ref}}$  retains good prediction capacities in different images never seen during the training. In all cases,  $M_{\text{Ref}}$  identifies most major faults fairly well, despite of the significant differences among the various types and resolutions of optical images and topographic data. U-Net skip connections contribute to the generalization capacity of the model by eliminating local minima in the loss landscape (Li et al., 2018). However, we suggest that the broad generalization capacity of  $M_{\text{Ref}}$  is not related to its architecture and loss function only, but also to the geometric character of faults. Faults are indeed hierarchical features: independent of its size, a fault is a 3D system comprising a longer master fault and a myriad of smaller faults and fractures of different lengths that generally dissect the master fault embedding rocks (Chester & Logan, 1986; de Joussineau & Aydin, 2007; Manighetti, King, Gaudemer, et al., 2001; Manighetti, King, & Sammis, 2004; Perrin, Manighetti, & Gaudemer, 2016). U-Net is especially appropriate to analyze hierarchical features made of lower-level ones. Furthermore, faults have certain generic properties, that is, properties that are similar regardless of their scale, context, slip mode, etc (see Figure 10). Among these properties, their 3D architecture seems to be similar at all scales with secondary faults forming a fan-shaped halo around the master fault (de Joussineau & Aydin, 2007; Flodin & Aydin, 2004; Manighetti, King, & Sammis, 2004; Perrin, Manighetti, Ampuero, et al., 2016; Perrin, Manighetti, & Gaudemer, 2016), fault traces are segmented laterally in a generic fashion (Manighetti, Caulet, et al., 2015; Manighetti, Zigone, et al., 2009; Otsuki & Dilov, 2005), and the surface slip distribution along the faults have a generic envelope shape (Manighetti, Campillo, Sammis, et al., 2005; Manighetti, King, Gaudemer, et al., 2001). Therefore, faults form (self-)similar patterns over a broad range of scales, and this scale-independent similarity likely facilitates efficient identification of the fault patterns at every possible scale in the image data.

Training the model on more diverse data sets is evidently the most straightforward step to further improve the model generalization capacities, that is, its capacities to be applied successfully to other data types/regions. We suggest that this be done with transfer learning designed so as to keep the “memory” of the model  $M_{\text{Ref}}$  and enrich it progressively with more information. Not only will the model increase its “knowledge,” but the combination of different data types/sources will dilute the impact of irrelevant features such as vegetation, stream channels, shadows, non-tectonic steps, etc. Another way to improve the model generalization might be to perform “style transfer” (also called, image-to-image translation (Tasar et al., 2020b), with GANs (Generative Adversarial Networks, Goodfellow, Pouget-Abadie, et al., 2014; Son et al., 2019; Tasar et al., 2020a, 2020c). Using GANs, one can modify the training images, for instance by changing their color distribution, so that they look like the test images. Then the model can be trained on the modified

images, and it is expected that its predictions on the unseen different types of images will be significantly improved compared to those of the original model.

Manual fault maps are georeferenced vector files. As such, they can be collocated exactly with any georeferenced image or topographic data of the zone they cover. Therefore, a ground truth done on one type of image/data can be collocated and thus “attached” to any other image/data of different type and resolution. Coarse-graining or fine-graining the ground truth might be necessary to make it matching better the new image/data resolution. Following this approach might be a way to augment the image data with available ground truth, allowing for the additional training of the model and hence improvement of its generalization capacities.

#### **6.4. Uncertainties and Model Robustness**

In natural data such as faults, the ground truth used for training is inherently uncertain: (1) manual fault maps are never exhaustive; (2) actual fault traces cannot be located exactly at the pixel resolution (see Section 2.3), and the accuracy of the manual fault tracing depends on the scale and resolution at which mapping is done; (3) biases are introduced by the analyst who determines what constitutes a fault or fracture, and where it is located, based on personal experience (e.g., Bond, 2015). This is clearly seen when comparing the basic mapping with the refined mapping.

Therefore, a major question is how these different uncertainties in the ground truth, of both aleatory (items 1 and 2 above) and epistemic (item 3 above) nature, can be taken into account in the CNN algorithm. In this study, we have used ad hoc criteria, such as representing each manual fault line as a fixed-width Gaussian probability density function, and uncertain traces as lower probability Gaussians. This rendered every fault trace 7 pixels wide, and its probabilities varying from 1 (or 0.5 for uncertain faults) to 0 from the drawn line to either sides. However, tectonic observations show that most major and thus most certain faults have generally wider traces at the ground surface, while minor faults, which can be more ambiguous, have narrower surface traces. Therefore, a fixed-width Gaussian function is not fully appropriate to represent the natural fault traces. Clearly, more accurate procedures should be developed to account for the various sources of uncertainties in the fault ground truth.

That manual fault maps are inherently incomplete also implies that zones free of fault traces in the ground truth can actually contain fractures and faults that were simply omitted in the manual mapping. This calls for the need to enforce the model to recognize that any part of the rock site might contain some faults even when they are absent in the ground truth. This may be achieved by including an adversarial loss term. In the present study, we have not addressed this issue. In future work, one may consider to modify the loss function, treating the ground truth labels as the noisy observations of true labels (Mnih & Hinton, 2012).

Like any model, the CNN predictions have uncertainties. Yet their quantification is not straightforward. As shown in the different model realizations that we have calculated, for a given architecture, any variation in the number of filters, training data volume, data type, epoch number, etc., yields predictions that vary slightly. However, estimating how much and why the predictions vary is challenging. The common indirect approach to estimate uncertainties on the predictions is to assess their capacities to reproduce the ground truth; this is the “model score.” However, there are two major limitations when faults are concerned. First, as said before, fault ground truths are by essence incomplete and uncertain, so that any comparison to a ground truth conveys a bias related to it. In particular, the model might predict faults that do exist but were not mapped. In that case, the score is low while the model is correct. Second, the model score is estimated using a mathematical loss function, and depending on the function that is used, the score differs. Moreover, the pixel-wise model output does not represent a calibrated probability estimation that a fault is present at that location. In other words, a model output of 0.8 does not necessarily imply that there is an 80% probability that the pixel be a part of a fault. Rather, it means that the model is more confident that the pixel is a fault than it is for the other pixels with lower probabilities, and ANNs tend to overestimate this level of confidence (Guo et al., 2017). Furthermore, it is not certain that the confidence score is linear. Therefore, more robust methods for quality control of the model predictions need to be considered.



One approach might be to employ dropout regularization both during training and at inference time. Most commonly, dropout regularization (Srivastava et al., 2014) is used to reduce overfitting and to improve generalization during training by randomly setting the output of neurons in a neural network to zero. The model can therefore not rely on any single neuron or pathway to produce an accurate prediction, and is forced to factor in redundancy. For a given realization of the dropout connections, a random subset of the model is used to produce  $\hat{y}_i$ . By maintaining dropout at inference time, each model evaluation essentially utilizes a different model. This technique has been coined “Bayesian dropout,” as it has been shown that the application of dropout at inference time produces a sample from the model posterior distribution (Gal & Ghahramani, 2016). Thus, through repeated inference with dropout enabled, the model posterior distribution can be efficiently sampled, giving for each pixel a distribution of the probability that a fault exists at that location. By doing so, the validity of the predictions can be better estimated. For instance, if  $\hat{y}_i = 0.9(0.4 - 0.95)$ , where the numbers in parentheses indicate a suitable interquartile range, would indicate that, even though the mean of the prediction is high (0.9), the confidence interval extends down to low values (0.4), hinting at low precision. On the other hand a prediction  $\hat{y}_i = 0.2(0.18 - 0.23)$  expresses great confidence that no fault is present at that location. Following this interpretation, a prediction  $\hat{y}_i = 0.5(0.499 - 0.501)$  would imply that the model is extremely confident about being uncertain whether or not a given pixel belongs to a fault.

### 6.5. Recovering Fault Hierarchy and Connectivity

Any fault is part of a multi-scale fault network comprising multiple faults that have different lengths and levels of “importance” in terms of strain accommodation: major faults are those to have accommodated more slip over the long-term (compared to more minor faults) (e.g., Ben-Zion & Sammis, 2003; de Joussineau & Aydin, 2007; Manighetti, King, Gaudemer, et al., 2001; Manighetti, King, & Sammis, 2004; Tchalenko, 1970). They are thus generally the longest faults (as faults propagate laterally as they grow over time, e.g., Manighetti, King, Gaudemer, et al., 2001; Schlagenhauf et al., 2008), and those expected to produce the largest earthquakes in case of rupture. Recovering fault hierarchy, that is, which faults are major and which faults are more minor, is thus critical to understand fault growth, organization, mechanics, and seismic potential. The fault hierarchy is generally indicated in the ground truth with different labels (different line thicknesses). In the present study, we have not taken into account the hierarchical information that was available in the expert ground truth, and have instead followed, as a first step, a binary approach “fault – not a fault.” In future work, we aim to integrate this information from the ground truth and learn the model to predict the fault hierarchy. This might be done with a multi-label approach. Instead of performing the training of the model with a binary class distribution (0 for no fault, 1 for fault), the labels for each pixel would span all classes describing the fault hierarchy (here four in total: the class “no fault” plus the three hierarchical levels defined in the ground truth). Practically, this means that the last layer in the model architecture should feature 4 channels and a sigmoid activation to ensure that the output of the model is appropriately normalized between 0 and 1, such that the sum of the probabilities over all classes equals 1. All other components of the proposed model do not require modification. Subsequent post-processing on pixels belonging to one particular class may aid in obtaining better associations than when no hierarchical distinction is made.

Faults are also connected systems. Any fault is connected to more minor faults generally splaying off its trace (e.g., de Joussineau & Aydin, 2007; Granier, 1985; Perrin, Manighetti, Ampuero, et al., 2016; Perrin, Manighetti, & Gaudemer, 2016). Recovering these connections is critical for they impact the mechanical properties of the whole fault zone, with consequences, for instance, on stress transfers among faults and earthquake behavior (Romanet et al., 2018; Ulrich et al., 2019). Although CNN computes the value of one pixel based on its neighbors, topology is not included so that pixels connected in specific patterns or meaningful pixel gaps might not be identified. One way to recover the topology of the faults could be to include a topological contribution to the training loss, as proposed by Hu et al. (2019). This does require, however, that all faults and fractures in the ground truth be disconnected (i.e., no overlap or crossing of traces), what is not the case at many places. Moreover, ambiguities may arise during the construction and comparison of the persistence diagrams, especially for areas of high fracture density. Another way might be to predict for each pixel, not only the probability to be a fault, but also the distance to the nearest fault, and even possibly the direction and distance toward the nearest fault (Ding & Bruzzzone, 2020). The output would thus be a 2D vector for each pixel, containing information on the fault continuity and connectivity.

Alternatively, one might use a different model architecture. When the fault mapping is performed manually, each fault trace is approximated by a set of straight-line segments, connected at anchor points. Thus, faults and fractures are representable by a finite collection of nodes connected by edges, or in other words, a graph. Recent advances in geometric deep learning (Bronstein et al., 2017) and graph neural networks (Gori et al., 2005; Zhou et al., 2019) have enabled learning strategies on non-Euclidean objects such as point cloud data (Wang et al., 2019), social networks (Hamilton et al., 2017), and molecules (Duvenaud et al., 2015). Valid operations on graphs are required to be invariant to the number of elements (nodes or edges) present in the graph, therefore allowing flexibility with respect to the number of nodes that represent the various faults in the system. In this way, a deep learning model can learn to map fractures in a similar way as human analysts do, namely as individually connected features, rather than as a fixed-size collection of unrelated pixels. The main challenge for graph-based learning in this context is to find an appropriate loss function that penalizes discrepancies between the model output and the ground truth. While in principle the model output and ground truth can always be rasterized and compared through a conventional cross-entropy or mean-squared error loss function, one needs to keep in mind that the loss function needs to be differentiable with respect to the model parameters in order to utilize the back-propagation technique (Rumelhart et al., 1986) that underlies the deep learning approach. Since rasterization is a non-differentiable operation, a different strategy to compare the output of a graph-based model with the vectorized ground truth needs to be considered. Addressing this challenge will be the topic of future research.

Developing one or other of the above approaches in the future should help to transform probability maps into more accurate fault vectors than those we derived here using common vectorization tools.

## 7. Conclusions

In the present study, we have developed a deep learning U-Net-type  $M_{\text{Ref}}$  model to automate the identification and mapping of tectonic fractures and faults in optical images and topographic data. We have demonstrated that, while trained on modest data volumes and qualities, our reference model  $M_{\text{Ref}}$  can be applied successfully to types of image data other than those used for training, and hence has good generalization capacities. This makes  $M_{\text{Ref}}$  a useful tool to assist geologists and geophysicists to rapidly identify and map fractures and faults in various types of images, including outcrop pictures acquired during field campaigns, images of earthquake surface ruptures, or images of rock experiments. In a few seconds to minutes depending on the data volume and available computer capacities, the  $M_{\text{Ref}}$  predictions provide an accurate knowledge of the fault/fracture existence, localization, and distribution in the zone of analysis, and highlight their overall architecture and organization: major faults are commonly revealed with the greatest probabilities, while more minor faults are emphasized with broader ranges of probabilities. The connections and cross-cutting relations among faults are well recovered, as well as their tectonic characteristics down to the smallest scales (lateral segmentation, en echelon arrangements, splaying faults, relay zones, intense fracturing such as in fault core and inner damage zone, etc.). The fault mapping produced by  $M_{\text{Ref}}$  is thus of great tectonic significance. While the vectorization approach we used here is not fully satisfactory, it allowed us to properly recover the densities and orientations of the faults and fractures, enabling their statistical analysis. However, refined vectorization approaches are needed to recover the individual fault lengths accurately. With these additional refinements, the  $M_{\text{Ref}}$  predictions will provide among the key information necessary to quantify the mechanical properties of the faulted rocks (e.g., Bonnet et al., 2001; Davy et al., 2018), an issue of critical importance in tectonic damage zones embedding master faults. The quantification of the elastic properties of the faulted/fractured rocks embedding a master fault is indeed critical to assess the behavior of an earthquake rupture on that fault (initiation, propagation, slip and acceleration amount and distribution, e.g., Cappa et al., 2014; Huang & Ampuero, 2011; Thakur et al., 2020). We thus anticipate that  $M_{\text{Ref}}$  will be helpful to make a step forward in these quantifications and related fault and earthquake modelings.

While  $M_{\text{Ref}}$  and any other supervised deep learning models are not expected to be able to predict anything else than what they have learned from humans, they are able to increase the range of scales analyzed. Therefore, they might be able to identify features undetectable by an analyst, here fractures and faults of very small scales undetected by the user. As such,  $M_{\text{Ref}}$  and its further refinements offer the opportunity

to analyze fault images down to very small scales never touched before. This should provide fracture and fault data over rarely explored ranges, a critical input to feed empirical fault scaling relations (or earthquake scaling laws if  $M_{\text{Ref}}$  is applied to rupture zones), such as those describing fault length statistics (e.g., Bonnet et al., 2001; de Joussineau & Aydin, 2007), or relating damage zone width and fault length (e.g., Perrin, Manighetti, & Gaudemer, 2016; Savage & Brodsky, 2011).

If, in the future, we become more able to decipher the physical fault properties that are identified by the CNN models, we might be able to discriminate which properties and hence which model parameters are most efficient to identify fractures and faults in optical images. This might enable us to prune the present models, keep only the kernels with the highest activations, and use them as intrinsic ingredients of any new CNN or other network designed for fracture and fault extraction in optical image data. Meanwhile, unsupervised learning might be explored to examine whether deep learning can recognize fractures and faults in images from their intrinsic properties and patterns alone. This would offer an opportunity to learn these properties and patterns without any interpretation bias, and to verify whether some of these fault properties and patterns are generic as it has been suggested (e.g., de Joussineau et al., 2007; Manighetti, Caulet, et al., 2015; Manighetti, King, Gaudemer, et al., 2001; Manighetti, King, & Sammis, 2004; Otsuki & Dilov, 2005; Perrin, Manighetti, & Gaudemer, 2016; Tchalenko, 1970).

Whatever the method used, supervised or unsupervised, relevant fault predictions could be stored in a database so as to form an increasing volume of available fault data worldwide. These predicted fault data could in turn form a specific ground truth usable to train new models in subsequent studies. With such an increasing volume of fault data and ground truth, fault predictions might evolve to 3D, especially through the combined modeling of surface images and seismic data. 3D predictions would increase our knowledge of fault zones architectures and fault networks connectivities, with critical implications for the understanding of fault and earthquake mechanics, stress transfers in the crust, and fluid pathways in geological reservoirs, among others.

## Data Availability Statement

The source code used in the study is an intellectual property of a consortium between Inria, CNES, and ACRI-ST (France) and cannot be freely distributed. But we provide the parameters of the trained model  $M_{\text{Ref}}$ , along with the image and topographic data and the ground truths used for training, validation and test. They are accessible on the repository <http://doi.org/10.5281/zenodo.4611494>, along with the supplementary figures produced in this study. The raw Pléiades satellite images used to calculate the DSMs are DS\_PHR1A\_201708171827015\_FR1\_PX\_W113N34\_0715\_00913, DS\_PHR1A\_201708171827174\_FR1\_PX\_W113N34\_0715\_00928\*, DS\_PHR1A\_201708171827293\_FR1\_PX\_W113N34\_0715\_00936 for the Granite Dells site, and DS\_PHR1B\_201701271830113\_FR1\_PX\_W115N36\_0612\_01482, DS\_PHR1B\_201701271830021\_FR1\_PX\_W115N36\_0612\_01464, DS\_PHR1B\_201701271830316\_FR1\_PX\_W115N36\_0612\_01450, DS\_PHR1A\_201708101830535\_FR1\_PX\_W115N36\_0612\_01425, DS\_PHR1B\_201702151833475\_FR1\_PX\_W115N36\_0612\_01462, DS\_PHR1B\_201702151834169\_FR1\_PX\_W115N36\_0612\_01461 DS\_PHR1B\_201702151833569\_FR1\_PX\_W115N36\_0612\_01485\* for the Valley of Fire site and DS\_PHR1B\_201709251826295\_FR1\_PX\_W112N38\_0908\_01849\*, DS\_PHR1A\_201710081826214\_FR1\_PX\_W112N38\_1006\_03007\* for WaterPocket region (Figure 1). IDs with a star refer to the orthorectified images used in the study. All Pléiades images can be obtained at <https://www.intelligence-airbusds.com/optical-and-radar-data/>.

## References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2015). Tensorflow: Large-scale machine learning on heterogeneous distributed systems (Vol. 19). arXiv preprint arXiv:1603.04467.
- Abu-Mostafa, Y. S., Magdon-Ismael, M., & Lin, H.-T. (2012). *Learning from data* (Vol. 4). AMLBook.
- Ahmadv, R., Aydin, A., Karimi-Fard, M., & Durllofsky, L. J. (2007). Permeability upscaling of fault zones in the Aztec Sandstone, Valley of Fire State Park, Nevada, with a focus on slip surfaces and slip bands. *Hydrogeology Journal*, 15(7), 1239–1250. <https://doi.org/10.1007/s10040-007-0180-2>
- Araya-Polo, M., Dahlke, T., Frogner, C., Zhang, C., Poggio, T., & Hohl, D. (2017). Automated fault detection without seismic processing. *The Leading Edge*, 36(3), 208–214. <https://doi.org/10.1190/le36030208.1>
- Babakhin, Y., Sanakoyeu, A., & Kitamura, H. (2019). Semi-supervised segmentation of salt bodies in seismic images using an ensemble of convolutional neural networks. In *German Conference on Pattern Recognition* (pp. 218–231). Springer.

## Acknowledgments

The Pléiades satellite image acquisition was supported by public funds received in the framework of GEOSUD, a project (ANR-10-EQPX-20) of the program “Investissements d’Avenir” managed by the French National Research Agency. The drone and ground images, and the work, have been funded by the French National Research Agency (ANR Grant FAULTS\_R\_GEMS # ANR-17-CE31-0008), the UCA<sup>IEDI</sup> Investments in the Future project managed by the French National Research Agency (grant ANR-15-IDEX-01, Academy of excellence #3), and the French CNRS (program Imag’In). The drone equipment was provided by the Arizona State University, and we greatly thank Chelsea Scott, Ramon Arrowsmith, and Tyler Scott for their help in the drone data acquisition, and for the great discussions and moments we shared with them on the field. Lionel Mattéo PhD has been funded by Thales Alenia Space and Observatoire de la Côte d’Azur, Université Côte d’Azur, France. The authors also thank Bilel Kanoun, Josiane Zerubia, and Sandrine Mathieu for insightful discussions. The authors are very grateful to Ramon Arrowsmith and an anonymous reviewer for their detailed, constructive comments that helped us to improve the manuscript.



- Barton, C. A., & Zoback, M. D. (1992). Self-similar distribution and properties of macroscopic fractures at depth in crystalline rock in the Cajon Pass Scientific Drill Hole. *Journal of Geophysical Research*, 97(B4), 5181–5200. <https://doi.org/10.1029/91jb01674>
- Bemis, S. P., Mickelthwaite, S., Turner, D., James, M. R., Akciz, S., Thiele, S. T., & Bangash, H. A. (2014). Ground-based and UAV-Based photogrammetry: A multi-scale, high-resolution mapping tool for structural geology and paleoseismology. *Journal of Structural Geology*, 69, 163–178. <https://doi.org/10.1016/j.jsg.2014.10.007>
- Ben-Zion, Y., & Sammis, C. G. (2003). Characterization of fault zones. *Pure and Applied Geophysics*, 160(3–4), 677–715. <https://doi.org/10.1007/pl00012554>
- Bergen, K. J., Johnson, P. A., Maarten, V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. *Science*, 363(6433). <https://doi.org/10.1126/science.aau0323>
- Bézenac, E. de, Pajot, A., & Gallinari, P. (2019). Deep learning for physical processes: incorporating prior scientific knowledge. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(12), 124009. <https://doi.org/10.1088/1742-5468/ab3195>
- Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford University Press.
- Bond, C. E. (2015). Uncertainty in structural interpretation: Lessons to be learnt. *Journal of Structural Geology*, 74, 185–200. <https://doi.org/10.1016/j.jsg.2015.03.003>
- Bonilla-Sierra, V., Scholtès, L., Donzé, F. V., & Elmoutie, M. K. (2015). Rock slope stability analysis using photogrammetric data and DFN-DEM modelling. *Acta Geotechnica*, 10(4), 497–511. <https://doi.org/10.1007/s11440-015-0374-z>
- Bonnet, E., Bour, O., Odling, N. E., Davy, P., Main, I., Cowie, P., & Berkowitz, B. (2001). Scaling of fracture systems in geological media. *Reviews of Geophysics*, 39(3), 347–383. <https://doi.org/10.1029/1999rg000074>
- Boureau, Y.-L., Ponce, J., & LeCun, Y. (2010). A theoretical analysis of feature pooling in visual recognition. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)* (pp. 111–118).
- Bradbury, K. K., Barton, D. C., Solum, J. G., Draper, S. D., & Evans, J. P. (2007). Mineralogic and textural analyses of drill cuttings from the San Andreas Fault Observatory at Depth (SAFOD) boreholes: Initial interpretations of fault zone composition and constraints on geologic models. *Geosphere*, 3(5), 299–318. <https://doi.org/10.1130/ges00076.1>
- Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., & Vandergheynst, P. (2017). Geometric deep learning: Going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4), 18–42. <https://doi.org/10.1109/MSP.2017.2693418>
- Bürgmann, R., Pollard, D. D., & Martel, S. J. (1994). Slip distributions on faults: effects of stress gradients, inelastic deformation, heterogeneous host-rock stiffness, and fault interaction. *Journal of Structural Geology*, 16(12), 1675–1690. [https://doi.org/10.1016/0191-8141\(94\)90134-1](https://doi.org/10.1016/0191-8141(94)90134-1)
- Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., & Kalinin, A. A. (2020). Albumentations: Fast and flexible image augmentations. *Information*, 11(2), 125. <https://doi.org/10.3390/info11020125>
- Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 679–698. <https://doi.org/10.1109/tpami.1986.4767851>
- Cappa, F., Perrin, C., Manighetti, I., & Delor, E. (2014). Off-fault long-term damage: A condition to account for generic, triangular earthquake slip profiles. *Geochemistry, Geophysics, Geosystems*, 15(4), 1476–1493. <https://doi.org/10.1002/2013gc005182>
- Carbonell, J. G., Michalski, R. S., & Mitchell, T. M. (1983). An overview of machine learning. In *Machine learning* (pp. 3–23). Elsevier.
- Chester, F. M., & Logan, J. M. (1986). Implications for mechanical properties of brittle faults from observations of the Punchbowl fault zone, California. *Pure and Applied Geophysics*, 124(1–2), 79–106. <https://doi.org/10.1007/bf00875720>
- Chorowicz, J., Luxey, P., Lyberis, N., Carvalho, J., Parrot, J.-F., Yürür, T., & Gündoğdu, N. (1994). The Maras Triple Junction (southern Turkey) based on digital elevation model and satellite imagery interpretation. *Journal of Geophysical Research*, 99(B10), 20225–20242. <https://doi.org/10.1029/94jb00321>
- Cowie, P. A., & Scholz, C. H. (1992). Physical explanation for the displacement-length relationship of faults using a post-yield fracture mechanics model. *Journal of Structural Geology*, 14(10), 1133–1148. [https://doi.org/10.1016/0191-8141\(92\)90065-5](https://doi.org/10.1016/0191-8141(92)90065-5)
- Davy, P., Darcel, C., Le Goc, R., & Mas Ivars, D. (2018). Elastic properties of fractured rock masses with frictional properties and power law fracture size distributions. *Journal of Geophysical Research: Solid Earth*, 123(8), 6521–6539. <https://doi.org/10.1029/2017JB015329>
- de Joussineau, G., & Aydin, A. (2007). The evolution of the damage zone with fault growth in sandstone and its multiscale characteristics. *Journal of Geophysical Research: Solid Earth*, 112(B12), B12401. <https://doi.org/10.1029/2006JB004711>
- de Joussineau, G., Mutlu, O., Aydin, A., & Pollard, D. D. (2007). Characterization of strike-slip fault-splay relationships in sandstone. *Journal of Structural Geology*, 29(11), 1831–1842. <https://doi.org/10.1016/j.jsg.2007.08.006>
- Deng, G., & Cahill, L. W. (1993). An adaptive Gaussian filter for noise reduction and edge detection. In *1993 IEEE conference record nuclear science symposium and medical imaging conference* (pp. 1615–1619). IEEE.
- Devilee, R. J. R., Curtis, A., & Roy-Chowdhury, K. (1999). An efficient, probabilistic neural network approach to solving inverse problems: Inverting surface wave velocities for Eurasian crustal thickness. *Journal of Geophysical Research*, 104(B12), 28841–28857. <https://doi.org/10.1029/1999jb000273>
- DeWitt, E., Langenheim, V., Force, E., Vance, R. K., Lindberg, P. A., & Driscoll, R. L. (2008). *Geologic Map of Prescott National Forest and the Headwater of the Verde River, Yavapai and Coconino Counties, Arizona*. US Department of the Interior, US Geological Survey.
- Ding, L., & Bruzzone, L. (2020). *Direction-aware Residual Network for Road Extraction in VHR Remote Sensing Images*. ArXiv Preprint ArXiv:2005.07232.
- Dramsch, J. S. (2020). 70 years of machine learning in geoscience in review. *Advances in Geophysics*, 61, 1–55. <https://doi.org/10.1016/bs.agph.2020.08.002>
- Dramsch, J. S., & Lüthje, M. (2018). Deep-learning seismic facies on state-of-the-art CNN architectures. In *SEG Technical Program Expanded Abstracts 2018* (pp. 2036–2040). Society of Exploration Geophysicists.
- Drouyer, S. (2020). An 'All Terrain' Crack Detector Obtained by Deep Learning on Available Databases. *Image Processing On Line*, 10, 105–123. <https://doi.org/10.5201/ipol.2020.282>
- Duvenaud, D., Maclaurin, D., Aguilera-Iparraguirre, J., Gomez-Bombarelli, R., Hirzel, T., Aspuru-Guzik, A., & Adams, R. P. (2015). Convolutional networks on graphs for learning molecular fingerprints. *Advances in Neural Information Processing Systems*, 9.
- Ermini, L., Catani, F., & Casagli, N. (2005). Artificial neural networks applied to landslide susceptibility assessment. *Geomorphology*, 66(1–4), 327–343. <https://doi.org/10.1016/j.geomorph.2004.09.025>
- Faulkner, D. R., Mitchell, T. M., Jensen, E., & Cembrano, J. (2011). Scaling of fault damage zones with displacement and the implications for fault growth processes. *Journal of Geophysical Research: Solid Earth*, 116(B5). <https://doi.org/10.1029/2010JB007788>
- Flodin, E. A., & Aydin, A. (2004). Evolution of a strike-slip fault network, Valley of Fire State Park, southern Nevada. *Geological Society of America Bulletin*, 116(1–2), 42–59. <https://doi.org/10.1130/b25282.1>
- Fukushima, K. (1975). Cognitron: A self-organizing multilayered neural network. *Biological Cybernetics*, 20(3–4), 121–136. <https://doi.org/10.1007/bf00342633>

- Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *International Conference on Machine Learning* (pp. 1050–1059). PMLR.
- GDAL/OGR contributors (2018). *GDAL/OGR geospatial data abstraction software library*. Open Source Geospatial Foundation.
- Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics* (pp. 249–256). PMLR.
- Godefroy, G., Caumon, G., Laurent, G., & Bonneau, F. (2020). Multi-scenario interpretations from sparse fault evidence using graph theory and geological rules. *Journal of Geophysical Research: Solid Earth*, 126(2), e2020JB020022. <https://doi.org/10.1029/2020JB020022>
- Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1). MIT Press.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672–2680).
- Gori, M., Monfardini, G., & Scarselli, F. (2005). A new model for learning in graph domains. In *Proceedings. 2005 IEEE International Joint Conference on Neural Networks* (Vol. 2, pp. 729–734). <https://doi.org/10.1109/IJCNN.2005.1555942>
- Gramstad, O., & Nickel, M. (2018). Automated interpretation of top and base salt using deep convolutional networks. In *SEG Technical Program Expanded Abstracts 2018* (pp. 1956–1960). Society of Exploration Geophysicists.
- Granier, T. (1985). Origin, damping, and pattern of development of faults in granite. *Tectonics*, 4(7), 721–737. <https://doi.org/10.1029/tc004i007p00721>
- GRASS Development Team (2020). *Geographic Resources Analysis Support System (GRASS GIS) Software*. Open Source Geospatial Foundation. Retrieved from <https://grass.osgeo.org>
- Grompone von Gioi, R., Jakubowicz, J., Morel, J.-M., & Randall, G. (2012). LSD: A line segment detector. *Image Processing On Line*, 2, 35–55. <https://doi.org/10.5201/ipol.2012.gjmr-lsd>
- Guillon, A. (2018). 3D convolutional neural networks for fault interpretation. In *80th EAGE Conference and Exhibition 2018* (Vol. 2018, pp. 1–5). European Association of Geoscientists & Engineers.
- Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). *On calibration of modern neural networks*. ArXiv:1706.04599 [Cs]. Retrieved from <http://arxiv.org/abs/1706.04599>
- Guo, Z., & Hall, R. W. (1989). Parallel thinning with two-subiteration algorithms. *Communications of the ACM*, 32(3), 359–373. <https://doi.org/10.1145/62065.62074>
- Haber, E., Fohring, J., McMillan, M., & Granek, J. (2019). Using machine learning to interpret 3D airborne electromagnetic inversions. *ASEG Extended Abstracts, 2019*(1), 1–4. <https://doi.org/10.1080/22020586.2019.12072978>
- Haddad, D., Zielke, O., Arrowsmith, J. R., Purvance, M. D., Haddad, A. G., & Landgraf, A. (2012). Estimating two-dimensional static stabilities and geomorphic settings of precariously balanced rocks from unconstrained digital photographs. *Geosphere*, 8(5), 1042–1053. <https://doi.org/10.1130/ges00788.1>
- Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems*.
- Harwin, S., & Lucieer, A. (2012). Assessing the accuracy of georeferenced point clouds produced via multi-view stereopsis from unmanned aerial vehicle (UAV) imagery. *Remote Sensing*, 4(6), 1573–1599. <https://doi.org/10.3390/rs4061573>
- Huang, Y., & Ampuero, J.-P. (2011). Pulse-like ruptures induced by low-velocity fault zones. *Journal of Geophysical Research*, 116(B12). <https://doi.org/10.1029/2011jb008684>
- Hulbert, C., Rouet-Leduc, B., Johnson, P. A., Ren, C. X., Rivière, J., Bolton, D. C., & Marone, C. (2019). Similarity of fast and slow earthquakes illuminated by machine learning. *Nature Geoscience*, 12(1), 69–74. <https://doi.org/10.1038/s41561-018-0272-8>
- Hulbert, C., Rouet-Leduc, B., Jolivet, R., & Johnson, P. A. (2020). An exponential build-up in seismic energy suggests a months-long nucleation of slow slip in Cascadia. *Nature Communications*, 11(1), 1–8. <https://doi.org/10.1038/s41467-020-17754-9>
- Hutchison, A. A., Böse, M., & Manighetti, I. (2020). Improving early estimates of large earthquake's final fault lengths and magnitudes leveraging source fault structural maturity information. *Geophysical Research Letters*, 47(14), e2020GL087539. <https://doi.org/10.1029/2020GL087539>
- Hu, X., Fuxin, L., Samaras, D., & Chen, C. (2019). Topology-preserving deep image segmentation. *Advances in Neural Information Processing Systems*, 5657–5668.
- Ide, H., & Kurita, T. (2017). Improvement of learning for CNN with ReLU activation by sparse regularization. In *2017 International Joint Conference on Neural Networks (IJCNN)* (pp. 2684–2691). IEEE.
- Jafrasteh, B., Manighetti, I., & Zerubia, J. (2020). Generative adversarial networks as a novel approach for tectonic fault and fracture extraction in high-resolution satellite and airborne optical images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 1219–1227. <https://doi.org/10.5194/isprs-archives-xliii-b3-2020-1219-2020>
- Jeong, S.-G., Tarabalka, Y., Nisse, N., & Zerubia, J. (2016). *Progressive tree-like curvilinear structure reconstruction with structured ranking learning and graph algorithm*. ArXiv Preprint ArXiv:1612.02631.
- Jeong, S.-G., Tarabalka, Y., & Zerubia, J. (2015). Stochastic model for curvilinear structure reconstruction using morphological profiles. In *2015 IEEE International Conference on Image Processing (ICIP)* (pp. 3575–3579). IEEE.
- Kingma, D. P., & Ba, J. (2014). *Adam: A method for stochastic optimization*. ArXiv:1412.6980 [Cs]. Retrieved from <http://arxiv.org/abs/1412.6980>
- Kong, Q., Trugman, D. T., Ross, Z. E., Bianco, M. J., Meade, B. J., & Gerstoft, P. (2019). Machine learning in seismology: Turning data into insights. *Seismological Research Letters*, 90(1), 3–14. <https://doi.org/10.1785/0220180259>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25(2), 1097–1105. <https://doi.org/10.1145/3065386>
- Lam, L., Lee, S.-W., & Suen, C. Y. (1992). Thinning methodologies—a comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(9), 869–885. <https://doi.org/10.1109/34.161346>
- Lary, D. J., Alavi, A. H., Gandomi, A. H., & Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7(1), 3–10. <https://doi.org/10.1016/j.gsf.2015.07.003>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- LeCun, Y., Haffner, P., Bottou, L., & Bengio, Y. (1999). Object recognition with gradient-based learning. In *Shape, contour and grouping in computer vision* (pp. 319–345). Springer.
- Lee, H., Grosse, R., Ranganath, R., & Ng, A. Y. (2009). Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *Proceedings of the 26th Annual International Conference on Machine Learning* (pp. 609–616). Association for Computing Machinery. <https://doi.org/10.1145/1553374.1553453>

- Li, H., Xu, Z., Taylor, G., Studer, C., & Goldstein, T. (2018). Visualizing the loss landscape of neural nets. *Advances in Neural Information Processing Systems*, 6389–6399.
- Maaten, L. van der, & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(Nov), 2579–2605.
- MacKay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge University Press.
- Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P. (2017). High-resolution aerial image labeling with convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 55(12), 7092–7103. <https://doi.org/10.1109/TGRS.2017.2740362>
- Manighetti, I., Campillo, M., Bouley, S., & Cotton, F. (2007). Earthquake scaling, fault segmentation, and structural maturity. *Earth and Planetary Science Letters*, 253(3–4), 429–438. <https://doi.org/10.1016/j.epsl.2006.11.004>
- Manighetti, I., Campillo, M., Sammis, C., Mai, P. M., & King, G. (2005). Evidence for self-similar, triangular slip distributions on earthquakes: Implications for earthquake and fault mechanics. *Journal of Geophysical Research: Solid Earth*, 110(B5). <https://doi.org/10.1029/2004jb003174>
- Manighetti, I., Caulet, C., De Barros, L., Perrin, C., Cappa, F., & Gaudemer, Y. (2015). Generic along-strike segmentation of Afar normal faults, East Africa: Implications on fault growth and stress heterogeneity on seismogenic fault planes. *Geochemistry, Geophysics, Geosystems*, 16(2), 443–467. <https://doi.org/10.1002/2014GC005691>
- Manighetti, I., King, G. C. P., Gaudemer, Y., Scholz, C. H., & Doubre, C. (2001). Slip accumulation and lateral propagation of active normal faults in Afar. *Journal of Geophysical Research*, 106(B7), 13667–13696. <https://doi.org/10.1029/2000jb900471>
- Manighetti, I., King, G., & Sammis, C. G. (2004). The role of off-fault damage in the evolution of normal faults. *Earth and Planetary Science Letters*, 217(3–4), 399–408. [https://doi.org/10.1016/s0012-821x\(03\)00601-0](https://doi.org/10.1016/s0012-821x(03)00601-0)
- Manighetti, I., Tapponnier, P., Courtillot, V., Gruszow, S., & Gillot, P.-Y. (1997). Propagation of rifting along the Arabia-Somalia plate boundary: The Gulfs of Aden and Tadjoura. *Journal of Geophysical Research*, 102(B2), 2681–2710. <https://doi.org/10.1029/96jb01185>
- Manighetti, I., Tapponnier, P., Gillot, P. Y., Jacques, E., Courtillot, V., Armijo, R., et al. (1998). Propagation of rifting along the Arabia-Somalia plate boundary: Into Afar. *Journal of Geophysical Research*, 103(B3), 4947–4974. <https://doi.org/10.1029/97jb02758>
- Manighetti, I., Zigone, D., Campillo, M., & Cotton, F. (2009). Self-similarity of the largest-scale segmentation of the faults: Implications for earthquake behavior. *Earth and Planetary Science Letters*, 288(3), 370–381. <https://doi.org/10.1016/j.epsl.2009.09.040>
- Meier, U., Curtis, A., & Trampert, J. (2007). Fully nonlinear inversion of fundamental mode surface waves for a global crustal model. *Geophysical Research Letters*, 34(16). <https://doi.org/10.1029/2007gl030989>
- Mitchell, T. M., & Faulkner, D. R. (2009). The nature and origin of off-fault damage surrounding strike-slip fault zones with a wide range of displacements: A field study from the Atacama fault system, northern Chile. *Journal of Structural Geology*, 31(8), 802–816. <https://doi.org/10.1016/j.jsg.2009.05.002>
- Mjolsness, E., & DeCoste, D. (2001). Machine learning for science: state of the art and future prospects. *Science*, 293(5537), 2051–2055. <https://doi.org/10.1126/science.293.5537.2051>
- Mnih, V., & Hinton, G. E. (2012). Learning to label aerial images from noisy data. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)* (pp. 567–574).
- Monsen, E., Hunt, D. W., Bounaim, A., Savary-Sismondini, B., Brenna, T., Nickel, M., et al. (2011). The automated interpretation of photo-realistic outcrop models. *Outcrops Revitalized—Tools, Techniques and Applications: Tulsa, Oklahoma, SEPM Concepts in Sedimentology and Paleontology*, 10, 107–136.
- Morrow, C., Lockner, D. A., & Hickman, S. (2015). Low resistivity and permeability in actively deforming shear zones on the San Andreas Fault at SAFOD. *Journal of Geophysical Research: Solid Earth*, 120(12), 8240–8258. <https://doi.org/10.1002/2015jb012214>
- Mousavi, S. M., & Beroza, G. C. (2020). A machine-learning approach for earthquake magnitude estimation. *Geophysical Research Letters*, 47(1), e2019GL085976. <https://doi.org/10.1029/2019gl085976>
- Myers, R., & Aydin, A. (2004). The evolution of faults formed by shearing across joint zones in sandstone. *Journal of Structural Geology*, 26(5), 947–966. <https://doi.org/10.1016/j.jsg.2003.07.008>
- Nielsen, M. A. (2015). *Neural networks and deep learning* (Vol. 2018). Determination Press.
- Oral, E., Weng, H., & Ampuero, J. P. (2020). Does a damaged-fault zone mitigate the near-field impact of supershear earthquakes?—Application to the 2018 7.5 Palu, Indonesia, earthquake. *Geophysical Research Letters*, 47(1), e2019GL085649. <https://doi.org/10.1029/2019gl085649>
- Otsuki, K., & Dilov, T. (2005). Evolution of hierarchical self-similar geometry of experimental fault zones: Implications for seismic nucleation and earthquake size. *Journal of Geophysical Research: Solid Earth*, 110(B3). <https://doi.org/10.1029/2004jb003359>
- Perol, T., Gharbi, M., & Denolle, M. (2018). Convolutional neural network for earthquake detection and location. *Science Advances*, 4(2), e1700578. <https://doi.org/10.1126/sciadv.1700578>
- Perrin, C., Manighetti, I., Ampuero, J.-P., Cappa, F., & Gaudemer, Y. (2016a). Location of largest earthquake slip and fast rupture controlled by along-strike change in fault structural maturity due to fault growth. *Journal of Geophysical Research: Solid Earth*, 121(5), 3666–3685. <https://doi.org/10.1002/2015JB012671>
- Perrin, C., Manighetti, I., & Gaudemer, Y. (2016b). Off-fault tip splay networks: A genetic and generic property of faults indicative of their long-term propagation. *Comptes Rendus Geoscience*, 348(1), 52–60. <https://doi.org/10.1016/j.crte.2015.05.002>
- Pratt, L. Y., Mostow, J., Kamm, C. A., & Kamm, A. A. (1991). Direct transfer of learned information among neural networks. In *AAAI* (Vol. 91, pp. 584–589).
- QGIS Development Team (2020). *QGIS Geographic Information System*. Open Source Geospatial Foundation. Retrieved from <http://qgis.osgeo.org>
- Radiguet, M., Cotton, F., Manighetti, I., Campillo, M., & Douglas, J. (2009). Dependency of near-field ground motions on the structural maturity of the ruptured faults. *Bulletin of the Seismological Society of America*, 99(4), 2572–2581. <https://doi.org/10.1785/0120080340>
- Ren, C. X., Dorostkar, O., Rouet-Leduc, B., Hulbert, C., Strebel, D., Guyer, R. A., et al. (2019). Machine learning reveals the state of intermittent frictional dynamics in a sheared granular fault. *Geophysical Research Letters*, 46(13), 7395–7403. <https://doi.org/10.1029/2019gl082706>
- Ren, C. X., Peltier, A., Ferrazzini, V., Rouet-Leduc, B., Johnson, P. A., & Brenguier, F. (2020). Machine learning reveals the seismic signature of eruptive behavior at Piton de la Fournaise volcano. *Geophysical Research Letters*, 47(3), e2019GL085523. <https://doi.org/10.1029/2019gl085523>
- Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., et al. (2019). *Tackling climate change with machine learning*. ArXiv Preprint ArXiv:1906.05433.
- Romanet, P., Bhat, H. S., Jolivet, R., & Madariaga, R. (2018). Fast and slow slip events emerge due to fault geometrical complexity. *Geophysical Research Letters*, 45(10), 4809–4819. <https://doi.org/10.1029/2018gl077579>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234–241). Springer.



- Ross, Z. E., Meier, M.-A., & Hauksson, E. (2018a). P wave arrival picking and first-motion polarity determination with deep learning. *Journal of Geophysical Research: Solid Earth*, 123(6), 5120–5129. <https://doi.org/10.1029/2017jb015251>
- Ross, Z. E., Meier, M.-A., Hauksson, E., & Heaton, T. H. (2018b). Generalized seismic phase detection with deep learning. *Bulletin of the Seismological Society of America*, 108(5A), 2894–2901. <https://doi.org/10.1785/0120180080>
- Rouet-Leduc, B., Hulbert, C., & Johnson, P. A. (2019). Continuous chatter of the Cascadia subduction zone revealed by machine learning. *Nature Geoscience*, 12(1), 75–79. <https://doi.org/10.1038/s41561-018-0274-6>
- Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., & Johnson, P. A. (2017). Machine learning predicts laboratory earthquakes. *Geophysical Research Letters*, 44(18), 9276–9282. <https://doi.org/10.1002/2017gl074677>
- Rouet-Leduc, B., Hulbert, C., McBrearty, I. W., & Johnson, P. A. (2020). Probing slow earthquakes with deep learning. *Geophysical Research Letters*, 47(4), e2019GL085870. <https://doi.org/10.1029/2019GL085870>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536. <https://doi.org/10.1038/323533a0>
- Rupnik, E., Daakir, M., & Deseilligny, M. P. (2017). MicMac—a free, open-source solution for photogrammetry. *Open Geospatial Data, Software and Standards*, 2(1), 14. <https://doi.org/10.1186/s40965-017-0027-2>
- Rupnik, E., Deseilligny, M. P., Delorme, A., & Klinger, Y. (2016). Refined satellite image orientation in the free open-source photogrammetric tools apero/micmac. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 3(1).
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3), 211–252. <https://doi.org/10.1007/s11263-015-0816-y>
- Savage, H. M., & Brodsky, E. E. (2011). Collateral damage: Evolution with displacement of fracture distribution and secondary fault strands in fault damage zones. *Journal of Geophysical Research: Solid Earth*, 116(B3), B03405. <https://doi.org/10.1029/2010JB007665>
- Schütt, K. T., Arbabzadah, F., Chmiela, S., Müller, K. R., & Tkatchenko, A. (2017). Quantum-chemical insights from deep tensor neural networks. *Nature Communications*, 8(1), 1–8. <https://doi.org/10.1038/ncomms13890>
- Schlagenhauf, A., Manighetti, I., Malavielle, J., & Dominguez, S. (2008). Incremental growth of normal faults: Insights from a laser-equipped analog experiment. *Earth and Planetary Science Letters*, 273(3–4), 299–311. <https://doi.org/10.1016/j.epsl.2008.06.042>
- Scholz, C. H. (2019). *The mechanics of earthquakes and faulting*. Cambridge University Press.
- Segall, P., & Pollard, D. D. (1983). Nucleation and growth of strike slip faults in granite. *Journal of Geophysical Research*, 88(B1), 555–568. <https://doi.org/10.1029/jb088ib01p00555>
- Shen, D., Wu, G., & Suk, H.-I. (2017). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, 19, 221–248. <https://doi.org/10.1146/annurev-bioeng-071516-044442>
- Sibson, R. H. (2003). Thickness of the seismic slip zone. *Bulletin of the Seismological Society of America*, 93(3), 1169–1178. <https://doi.org/10.1785/0120020061>
- Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*. ArXiv Preprint ArXiv:1409.1556.
- Smith, S. A. F., Billi, A., Toro, G. D., & Spiess, R. (2011). Principal slip zones in limestone: Microstructural characterization and implications for the seismic cycle (Tre Monti Fault, Central Apennines, Italy). *Pure and Applied Geophysics*, 168(12), 2365–2393. <https://doi.org/10.1007/s00024-011-0267-5>
- Sobel, I. (2014). *History and definition of the sobel operator*. Retrieved from the World Wide Web, 1505.
- Son, J., Park, S. J., & Jung, K.-H. (2019). Towards accurate segmentation of retinal vessels and the optic disc in fundoscopic images with generative adversarial networks. *Journal of Digital Imaging*, 32(3), 499–512. <https://doi.org/10.1007/s10278-018-0126-3>
- Srinivasan, G., Hyman, J. D., Osthus, D. A., Moore, B. A., O'Malley, D., Karra, S., et al. (2018). Quantifying topological uncertainty in fractured systems using graph theory and machine learning. *Scientific Reports*, 8(1), 1–11. <https://doi.org/10.1038/s41598-018-30117-1>
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56), 1929–1958.
- Stirling, M. W., Wesnousky, S. G., & Shimazaki, K. (1996). Fault trace complexity, cumulative slip, and the shape of the magnitude-frequency distribution for strike-slip faults: A global survey. *Geophysical Journal International*, 124(3), 833–868. <https://doi.org/10.1111/j.1365-246x.1996.tb05641.x>
- Sturman, O., von Ziegler, L., Schläppli, C., Akyol, F., Privitera, M., Slominski, D., et al. (2020). Deep learning-based behavioral analysis reaches human accuracy and is capable of outperforming commercial solutions. *Neuropsychopharmacology*, 45, 1942–1952.
- Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., & Cardoso, M. J. (2017). Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep learning in medical image analysis and multimodal learning for clinical decision support* (pp. 240–248). Springer.
- Tapponnier, P., & Molnar, P. (1977). Active faulting and tectonics in China. *Journal of Geophysical Research*, 82(20), 2905–2930. <https://doi.org/10.1029/jb082i020p02905>
- Tasar, O., Happy, S. L., Tarabalka, Y., & Alliez, P. (2020a). ColorMapGAN: Unsupervised domain adaptation for semantic segmentation using color mapping generative adversarial networks. *IEEE Transactions on Geoscience and Remote Sensing*, 58(10), 7178–7193. <https://doi.org/10.1109/TGRS.2020.2980417>
- Tasar, O., Happy, S. L., Tarabalka, Y., & Alliez, P. (2020b). *SemI2I: Semantically consistent image-to-image translation for domain adaptation of remote sensing data*. ArXiv:2002.05925 [Cs, Eess]. Retrieved from <http://arxiv.org/abs/2002.05925>
- Tasar, O., Tarabalka, Y., & Alliez, P. (2019). Incremental learning for semantic segmentation of large-scale remote sensing data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(9), 3524–3537. <https://doi.org/10.1109/JSTARS.2019.2925416>
- Tasar, O., Tarabalka, Y., Giros, A., Alliez, P., & Clerc, S. (2020c). StandardGAN: Multi-source domain adaptation for semantic segmentation of very high resolution satellite images by data standardization. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (192–193).
- Tchalenko, J. S. (1970). Similarities between shear zones of different magnitudes. *Geological Society of America Bulletin*, 81(6), 1625–1640. [https://doi.org/10.1130/0016-7606\(1970\)81\[1625:sbszod\]2.0.co;2](https://doi.org/10.1130/0016-7606(1970)81[1625:sbszod]2.0.co;2)
- Thakur, P., Huang, Y., & Kaneko, Y. (2020). Effects of low-velocity fault damage zones on long-term earthquake behaviors on mature strike-slip faults. *Journal of Geophysical Research: Solid Earth*, 125(8), e2020JB019587. <https://doi.org/10.1029/2020jb019587>
- Tingdahl, K. M., & De Rooij, M. (2005). Semi-automatic detection of faults in 3D seismic data. *Geophysical Prospecting*, 53(4), 533–542. <https://doi.org/10.1111/j.1365-2478.2005.00489.x>
- Turner, D., Lucieer, A., & Watson, C. (2012). An automated technique for generating georectified mosaics from ultra-high resolution unmanned aerial vehicle (UAV) imagery, based on structure from motion (SfM) point clouds. *Remote Sensing*, 4(5), 1392–1410. <https://doi.org/10.3390/rs4051392>
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84(4), 327. <https://doi.org/10.1037/0033-295x.84.4.327>

- Ulrich, T., Gabriel, A.-A., Ampuero, J.-P., & Xu, W. (2019). Dynamic viability of the 2016 Mw 7.8 Kaikōura earthquake cascade on weak crustal faults. *Nature Communications*, 10(1), 1–16. <https://doi.org/10.1038/s41467-019-09125-w>
- Van der Baan, M., & Jutten, C. (2000). Neural networks in geophysical applications. *Geophysics*, 65(4), 1032–1047. <https://doi.org/10.1190/1.1444797>
- Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience*, 2018. <https://doi.org/10.1155/2018/7068349>
- Waldeland, A. U., Jensen, A. C., Gelius, L.-J., & Solberg, A. H. S. (2018). Convolutional neural networks for automated seismic interpretation. *The Leading Edge*, 37(7), 529–537. <https://doi.org/10.1190/tle37070529.1>
- Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., & Solomon, J. M. (2019). Dynamic graph CNN for learning on point clouds. *ACM Transactions on Graphics*, 38(5), 1–12. <https://doi.org/10.1145/3326362>
- Watterson, J., Walsh, J. J., Gillespie, P. A., & Easton, S. (1996). Scaling systematics of fault sizes on a large-scale range fault map. *Journal of Structural Geology*, 18(2–3), 199–214. [https://doi.org/10.1016/s0191-8141\(96\)80045-9](https://doi.org/10.1016/s0191-8141(96)80045-9)
- Wesnousky, S. G. (1988). Seismological and structural evolution of strike-slip faults. *Nature*, 335(6188), 340–343. <https://doi.org/10.1038/335340a0>
- Wesnousky, S. G. (2006). Predicting the endpoints of earthquake ruptures. *Nature*, 444(7117), 358–360. <https://doi.org/10.1038/nature05275>
- Wu, H., & Zhang, B. (2018). *A deep convolutional encoder-decoder neural network in assisting seismic horizon tracking*. ArXiv Preprint ArXiv:1804.06814.
- Wu, X., & Fomel, S. (2018). Automatic fault interpretation with optimal surface voting. *Geophysics*, 83(5), O67–O82. <https://doi.org/10.1190/geo2018-0115.1>
- Yang, H. (2015). Recent advances in imaging crustal fault zones: A review. *Earthquake Science*, 28(2), 151–162. <https://doi.org/10.1007/s11589-015-0114-3>
- Yan, Z., Gu, H., & Cai, C. (2013). Automatic fault tracking based on ant colony algorithms. *Computers & Geosciences*, 51, 269–281. <https://doi.org/10.1016/j.cageo.2012.08.003>
- Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing [review article]. *IEEE Computational Intelligence Magazine*, 13(3), 55–75. <https://doi.org/10.1109/mci.2018.2840738>
- You, N., Li, Y. E., & Cheng, A. (2020). Shale anisotropy model building based on deep neural networks. *Journal of Geophysical Research: Solid Earth*, 125(2), e2019JB019042. <https://doi.org/10.1029/2019jb019042>
- Zang, A., Wagner, F. C., Stanchits, S., Janssen, C., & Dresen, G. (2000). Fracture process zone in granite. *Journal of Geophysical Research*, 105(B10), 23651–23661. <https://doi.org/10.1029/2000jb900239>
- Zhang, Q., Yusifov, A., Joy, C., Shi, Y., & Wu, X. (2019). FaultNet: A deep CNN model for 3D automated fault picking. In *SEG Technical Program Expanded Abstracts 2019* (pp. 2413–2417). Society of Exploration Geophysicists. <https://doi.org/10.1190/segam2019-3215930.1>
- Zhou, J., Cui, G., Zhang, Z., Yang, C., Liu, Z., Wang, L., et al. (2019). *Graph neural networks: A review of methods and applications*. ArXiv:1812.08434.
- Zhu, W., & Beroza, G. C. (2019). PhaseNet: A deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, 216(1), 261–273. <https://doi.org/10.1093/gji/ggy423>
- Zigone, D., Ben-Zion, Y., Lehujeur, M., Campillo, M., Hillers, G., & Vernon, F. L. (2019). Imaging subsurface structures in the San Jacinto fault zone with high-frequency noise recorded by dense linear arrays. *Geophysical Journal International*, 217(2), 879–893. <https://doi.org/10.1093/gji/ggz069>