

2024-03

# Leveraging Spatial Metadata in Machine Learning for Improved Objective Quantification of Geological Drill Core

Grant, LJC

<https://pearl.plymouth.ac.uk/handle/10026.1/22405>

---

10.1029/2023ea003220

Earth and Space Science

American Geophysical Union (AGU)

---

*All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.*

1     **Leveraging Spatial Metadata in Machine Learning for**  
2     **Improved Objective Quantification of Geological Drill**  
3     **Core**

4     **Lewis, J. C. Grant<sup>1</sup>. Miquel Massot-Campos<sup>2</sup>. Rosalind, M. Coggon<sup>1</sup>. Blair**  
5     **Thornton<sup>2</sup>. Francesca Rotondo<sup>1</sup>. Michelle Harris<sup>3</sup>. Aled, D. Evans<sup>1</sup>. Damon,**  
6     **A. H. Teagle<sup>1</sup>**

7     <sup>1</sup>School of Ocean and Earth Science, National Oceanography Centre Southampton, University of  
8     Southampton, Southampton, SO14 3ZH, UK.

9     <sup>2</sup>Centre for In Situ and Remote Intelligent Sensing, University of Southampton, Southampton SO16 7QF,  
10     U.K.

11    <sup>3</sup>School of Geography, Earth and Environmental Sciences, Plymouth University, Plymouth PL4 8AA, UK

12     **Key Points:**

- 13     • Use of spatial metadata improves state-of-the-art unsupervised feature extraction.  
14     • Semi-supervised methods using spatial metadata outperform supervised methods  
15     for the same expert labelling effort.  
16     • New methods described enable the standardisation of core-logging processes and  
17     improve cross-dataset comparisons.

---

Corresponding author: Lewis, J. C. Grant, [L.Grant@soton.ac.uk](mailto:L.Grant@soton.ac.uk)

## 18 **Abstract**

19 Here we present a method for using the spatial x-y coordinate of an image cropped from  
 20 the cylindrical surface of digital 3D drill core images and demonstrate how this spatial  
 21 metadata can be used to improve unsupervised machine learning performance. This ap-  
 22 proach is applicable to any dataset with known spatial context, however, here it is used  
 23 to classify 400 m of drillcore imagery into 12 distinct classes reflecting the dominant rock  
 24 types and alteration features in the core. We modified two unsupervised learning mod-  
 25 els to incorporate spatial metadata and an average improvement of 25 % was achieved  
 26 over equivalent models that did not utilize metadata. Our semi-supervised workflow in-  
 27 volves unsupervised network training followed by semi-supervised clustering where a sup-  
 28 port vector machine uses a subset of  $M$  expert labelled images to assign a pseudolabel  
 29 to the entire dataset. Fine-tuning of the best performing model showed an  $f_1$  (macro av-  
 30 erage) of 90 %, and its classifications were used to estimate bulk fresh and altered rock  
 31 abundance downhole. Validation against the same information gathered manually by ex-  
 32 perts when the core was recovered during the Oman Drilling Project revealed that our  
 33 automatically generated datasets have a significant positive correlation (Pearson's  $r$  of  
 34 0.65-0.72) to the expert generated equivalent, demonstrating that valuable geological in-  
 35 formation can be generated automatically for 400 m of core with only  $\sim 24$  hrs of domain  
 36 expert effort.

## 37 **Plain Language Summary**

38 This work presents a novel method for using the spatial context of digital core im-  
 39 ages to improve the descriptive accuracy of unsupervised machine learning algorithms.  
 40 The addition of spatial metadata improves model performance by an average of 25 %,   
 41 with the best performing model in this study achieving an accuracy score of 90 %. The  
 42 output of this model was then used to estimate the amount of fresh and altered rock within  
 43 a 400 m long drill core, which was shown to be of comparable quality to the same es-  
 44 timations made by geologists on the cores themselves.

## 45 **1 Introduction**

46 Drilling into the Earth to recover cores for geological analysis is an essential tool  
 47 that provides valuable insight into otherwise inaccessible environments, yielding datasets  
 48 utilized for mining, infrastructure planning and reconstructing the history of the planet.  
 49 The task of describing these cores falls to specialists who systematically work through  
 50 the recovered material to produce a series of descriptive and quantitative logs (core-logging)  
 51 as well as visual core descriptions (VCDs). The features documented may include, but  
 52 are not limited to, changes in rock type, veins and alteration features, structural mea-  
 53 surements, and variations in relative mineral abundance downhole. These tasks are time  
 54 consuming and rely on subjective estimates of the abundance of key features within a  
 55 core. Furthermore, human interpretation tends to overestimate the abundance of a given  
 56 feature in a scene, causing estimates to vary widely between individuals (Olmstead et  
 57 al., 2004; Finn et al., 2010) and objective automated methods could resolve this under-  
 58 lying bias. In addition to VCDs, cores are digitally imaged, and in the case of scientific  
 59 drilling, their physical properties are measured prior to detailed petrographic and geo-  
 60 chemical analyses (Jarrard et al., 2003; Kelemen et al., 2020), but additional process-  
 61 ing is needed to make these datasets machine readable, limiting their use in emerging  
 62 machine learning applications. During drilling campaigns downhole wireline geophys-  
 63 ical logs of the borehole wall may also be collected, providing useful continuous datasets  
 64 for comparing borehole features with recovered core material to compensate for incom-  
 65 plete core recovery (Tominaga et al., 2009; Tominaga & Umino, 2010). Most attempts  
 66 to automate the classification of rock-types downhole initially focused on applying ar-  
 67 tificial neural networks (ANN) to one-dimensional borehole data (Tominaga et al., 2009;

68 Ma, 2011; Al-Mudhafar, 2017; J. He et al., 2019). However, using only numerical data  
69 has the limitation of providing less direct information about the rock when compared  
70 to core images (Chai et al., 2009; Thomas et al., 2011).

71 Most recent efforts to automatically classify rock-types using images of drill core have  
72 utilized convolutional neural networks (CNN) as they are more suited to image analy-  
73 sis (LeCun et al., 1995). When training a CNN to classify images, there are three main  
74 types of machine learning; supervised, unsupervised and semi-supervised, which involves  
75 a combination of unsupervised learning followed by a less intensive supervised step (Camps-  
76 Valls et al., 2007). The initial 'learning' stage of training is where a CNN determines which  
77 images it considers similar and dissimilar, however, additional steps are required to as-  
78 sign classifications or labels to the images. In supervised learning, each training image  
79 has been labelled to give the model a target output to work towards, however this re-  
80 quires significant effort on the part of the annotator. In contrast, unsupervised learn-  
81 ing does not involve any labelling effort as the network extracts salient information from  
82 each image, referred to as a latent representation, and clustering techniques allow group-  
83 ing of images based on these simplified representations. An expert then inspects these  
84 clusters and provides a label to each. When taking a semi-supervised approach, a sub-  
85 set of expert labelled images can be provided to an unsupervised model to allow it to  
86 both cluster and assign a label to all images. Images are labelled based on where their  
87 latent representations plot in the hyper-dimensional feature space relative to the expert  
88 labelled subset. To date, there have been numerous attempts to use neural networks to  
89 classify images of drill core, all of which have taken slightly different approaches.

90 Zhang et al. (2017) used a supervised approach to train a CNN to classify a dataset of  
91 1500 2D grayscale borehole wall resistivity images into three texturally distinct sedimen-  
92 tary rock types (sandstone, shale and conglomerate). Their number of training images  
93 was class imbalanced with an order of magnitude more sandstone images used in an at-  
94 tempt to improve their model's ability to identify potential hydrocarbon reservoirs. Sim-  
95 ilarly, Alzubaidi et al. (2021) used a supervised workflow to compare the performance  
96 of several CNN model architectures in identifying three sedimentary rock types in pho-  
97 tos of boxed core sections (box photos) with the ResNeXt-50 CNN architecture out-performing  
98 other networks. Their training dataset consisted of 76,500 (25,500 per class) 2 cm<sup>2</sup> patches  
99 cropped from the box photos and all models were trained to identify non-core artifacts  
100 in the images to avoid them being labelled as classes of geological interest. Although this  
101 work showed promising results, such models are only capable of classifying a few distinct  
102 classes of rock and consequently have only limited applicability to more complex image  
103 datasets that display greater variability of geological features. Most recently, Fu et al.  
104 (2022) demonstrated a supervised workflow based on fine-tuning CNNs to identify 10  
105 rock types commonly encountered during subsurface engineering projects. Their work  
106 showed ResNeSt-50 produced the best prediction accuracy of 99.6 %. Supervised train-  
107 ing of models requires careful preparation of the input data by an expert to ensure each  
108 desired class is well represented. For this reason, Fu et al. (2022) trained their models  
109 using 15,000 3 cm<sup>2</sup> labelled images of best-case examples of each rock type having first  
110 discarded images not of interest, such as crushing structures and crayon marks. Images  
111 removed from the training dataset were also defined based on what the authors believed  
112 would confuse the CNNs and cause them to mis-classify features of interest.

113 A concerted effort to label a large database of images of all known rock types would pro-  
114 vide a widely applicable training dataset, however, unlike in satellite imagery and ob-  
115 ject recognition research, there are no publicly available training datasets for classify-  
116 ing common rock types in drill core (Deng et al., 2009; Van Etten et al., 2018). This is  
117 partly because resources are rarely put toward labelling such datasets, but also because  
118 it is difficult to combine individual datasets with variable resolution and quality, often  
119 stored in different media and file formats, into a single database. In response to these  
120 limitations, this study is intended to provide researchers with a means of analysing large  
121 numbers of images on a per-dataset basis with minimal effort in the hope that widely  
122 applicable training datasets of rock images can begin to emerge. Furthermore, use of spa-



CNN framework	Feature extraction	Spatial metadata	Reference
Autoencoder (AE)	unsupervised (autoencoder)	N	Yamada et al. (2021)
Location Guided Autoencoder (LGA)	unsupervised (autoencoder)	Y	Yamada et al. (2021)
SimCLR	unsupervised (contrastive learning)	N	Chen et al. (2020)
GeoCLR	unsupervised (contrastive learning)	Y	Yamada, Prügel-Bennett, et al. (2022)
ResNet18	supervised	N	He et al. (2016)

Table 1: List of the machine learning models used in this study. The feature extraction column identifies whether the model learns with (supervised) or without (unsupervised) domain expert input and the spatial metadata column identifies models which utilize spatial information accompanying images during training (Y = yes, N = no). The references provided are those that outline the original development of each model.

123 tial information alongside numerical datasets have been shown to improve the automatic  
124 classification of geological information stored in the data (Yamada et al., 2021; Hill et  
125 al., 2015, 2021), and here we make a first attempt at leveraging spatial information when  
126 classifying digital geological core imagery.

127 In this study we modify two unsupervised learning frameworks originally designed to use  
128 3D geolocational metadata for improved semantic interpretation of seafloor imagery (Yamada  
129 et al., 2021; Yamada, Prügel-Bennett, et al., 2022; Yamada, Massot-Campos, et al., 2022)  
130 to instead use the x-y coordinate of where an image lies on the surface of a 3D drill core  
131 image. The first framework uses an autoencoder that was trained both with and with-  
132 out the addition of this spatial metadata, whereas the second uses two contrastive learn-  
133 ing methods, one that makes use of metadata, and another that does not (Table 1). The  
134 performance of each framework is reviewed to determine which is most accurate and we  
135 present a novel semi-supervised workflow for training CNNs using images accompanied  
136 by spatial metadata. The output of the best performing model is then used to automat-  
137 ically generate a downhole log of hydrothermal alteration extent, which is bench-marked  
138 against expert generated alteration logs.

## 139 2 Methods

### 140 2.1 Background

#### 141 2.1.1 Artificial Neural Networks

142 An artificial neural network (ANN) is a computer model inspired by the structure  
143 of the human brain and consists of multiple layers of stacked artificial neurons, also re-  
144 ferred to as perceptrons or nodes (Rosenblatt, 1962). Each artificial neuron is a math-  
145 ematical model that takes multiple binary inputs ( $x$ ) and gives a binary output deter-  
146 mined by whether the weighted sum of the inputs meet some threshold value ( $t$ ). The  
147 weight ( $w$ ) assigned to a given input expresses its importance to the output, and the weight  
148 and threshold parameters can be adjusted to customize a model to a particular task. To  
149 exert control on how easily a neuron will give a 1, the threshold is often replaced by a  
150 bias ( $b \equiv -t$ ) and the neuron’s activation function is expressed using the following dot  
151 product:

$$\text{output} = \begin{cases} 0, & \text{if } w \cdot x + b < 0, \\ 1, & \text{if } w \cdot x + b > 0. \end{cases} \quad (1)$$

152 The layers of stacked neurons between the input and output layers of an ANN are called  
153 hidden layers and each neuron in a hidden layer receives its input from every neuron of  
154 the previous layer. Therefore, each neuron in an ANN is fully connected to each neu-  
155 ron in the adjacent layers (Fig. 1a) (Krogh, 2008). By using weights, biases and activa-  
156 tion functions, each hidden layer extracts features within its input, and multiple hidden  
157 layers make a flexible model capable of identifying complex patterns within a dataset.  
158 The final output layer of an ANN provides a prediction for the information passed through  
159 the hidden layers, and the number of neurons in this layer depends on the application.  
160 In the case of a binary classification model the last layer would contain only two nodes,  
161 but for more complex cases the number of nodes will be equal to the number of poten-  
162 tial classes in the input data. One drawback of using ANNs for image processing is that  
163 each neuron possesses a unique weight and bias, requiring great processing power due  
164 to the large number of parameters handled by the model. This, as well as the fact that  
165 ANNs do not achieve spatial invariance, that is they are unable to recognise features re-  
166 gardless of their specific location in an image, limit their use in computer vision appli-  
167 cations.

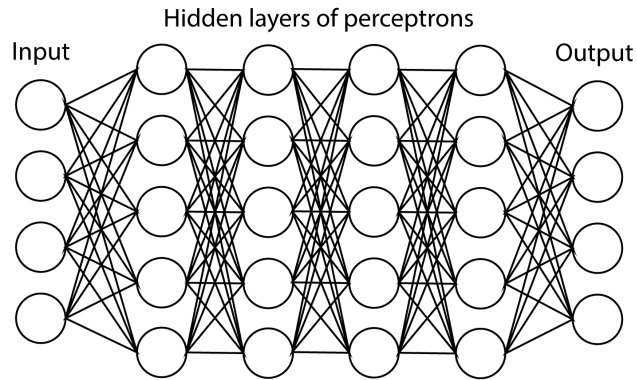
### 168 **2.1.2 Convolutional Neural Networks**

169 A Convolutional neural network (CNN) is a type of deep ANN developed in the  
170 early 1990s that can account for the spatial structure of input data (LeCun et al., 1995).  
171 Innovations over the last decade have made CNNs increasingly popular for computer vi-  
172 sion tasks, as their architecture is particularly suited for image analysis (Krizhevsky et  
173 al., 2012; Russakovsky et al., 2015). Unlike the fully connected layers of an ANN, each  
174 neuron in a CNN’s first hidden layer corresponds to a rectangular region of a defined size  
175 and location in the input image (Fig. 1b). This rectangular region is processed by a con-  
176 volutional filter or kernel with a weight and bias and is known as a ‘local receptive field’.  
177 The local receptive field then moves across the input neurons while keeping the same weight  
178 and bias when mapping information to its corresponding neuron in the hidden layer. It-  
179 erating this process across an image (convolution) creates a hidden layer (feature map)  
180 of neurons capable of detecting the same feature anywhere in the image. Each hidden  
181 layer of a CNN can contain multiple feature maps, and at shallow levels they can detect  
182 simple features, such as lines and shapes, whereas at deeper layers increasingly more com-  
183 plex features become identifiable. Convolution layers are then followed by pooling lay-  
184 ers that simplify the output of each feature map by summarising a specified sub-region  
185 into a condensed feature map (Fig. 1b). Pooling in a CNN is a downsampling operation  
186 that reduces spatial dimensions while retaining important features, aiding in computa-  
187 tional efficiency, and promoting robustness and generalization of the network. The ma-  
188 jor benefit of using local receptive fields and pooling layers, is that they make CNNs well  
189 adapted to handle spatial invariance within an image. The feature maps created by convolu-  
190 tion-pooling layers are multi-dimensional arrays (tensors), which make them suitable for iden-  
191 tifying complex features, but not for assigning class scores or probabilities. Therefore,  
192 the tensors produced by the last convolution-pooling layer are flattened into a one-dimensional  
193 vector that is fed into a fully connected layer of neurons. This fully connected layer trans-  
194 forms its input into high-level features that can be used for classification and regression.  
195 The output layer of a CNN is also a fully connected layer consisting of as many neurons  
196 as possible classes, and the input image is classified depending on which of these neu-  
197 rons is triggered by its activation function (Fig. 1b).

### 198 **2.1.3 Unsupervised Machine Learning**

199 In computer vision, there are many publicly available databases of labelled images,  
200 such as ImageNet, MS COCO and CIFAR-100, that can be used to train CNNs to clas-  
201 sify common objects. However, a supervised approach cannot be used when the classes  
202 within these datasets have no relevance to the application domain. In fields such as ge-

a)



b)

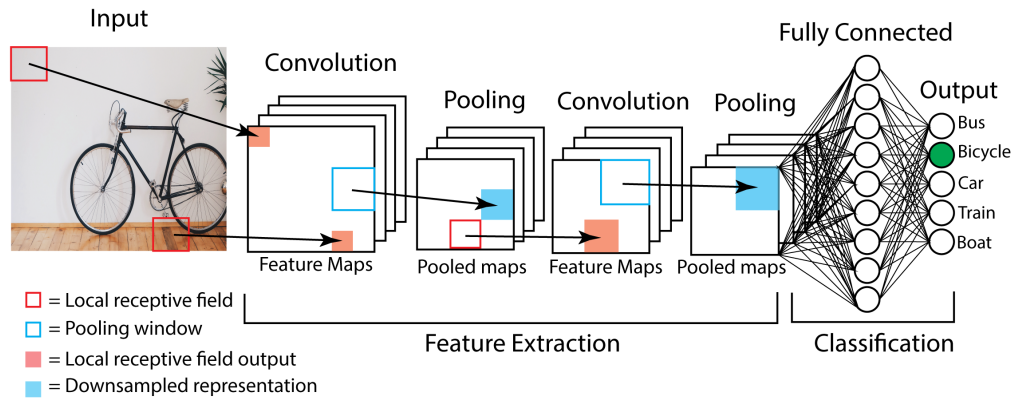


Fig. 1: Diagrams showing (a) the basic structure of an artificial neural network and (b) a convoluted neural network.

203 ology, there are no large labelled datasets of rock images available to pre-train a model  
 204 and the labelling effort required to generate enough training images for supervised ap-  
 205 proaches would be too time consuming, particularly in the context of a drilling campaign.  
 206 This is because a given campaign often involves drilling numerous holes that may yield  
 207 hundreds or thousands of meters of complex drill core, all of which needs describing by  
 208 an expert. A solution to this is to utilize unsupervised CNN frameworks capable of ex-  
 209 tracting salient information from geological images without any prior labelling effort, and  
 210 two such frameworks include autoencoders and contrastive learning.

#### 211 **2.1.4 Autoencoders**

212 An autoencoder (AE) is a form of neural network architecture used for unsuper-  
 213 vised learning and dimensionality reduction that consists of two elements. Firstly, an en-  
 214 coder ( $f$ ) that takes an input ( $x$ ) and compresses it into a lower dimensional represen-  
 215 tation, or latent representation ( $h = f_\phi(x)$ ) (Fig. 2Bii). Secondly, a decoder ( $g$ ) which  
 216 uses the latent representation to re-construct the input to give  $x_r = g_\theta(h)$  (Fig. 2Biii),  
 217 where  $\phi$  and  $\theta$  are the parameters of the encoder and decoder, respectively. Where the  
 218 input data is continuous ( $\{x\}_{i=1}^n$ ), the difference between  $x$  and  $x_r$  (reconstruction loss)  
 219 can be calculated using the mean square error, making the optimizing objective (loss func-  
 220 tion) of the AE:

$$\min_{\phi, \theta} L_{rec} = \min \frac{1}{n} \sum_{i=1}^n \|x_i - x_{ri}\|^2 \quad (2)$$

221 The major objective of network training in machine learning is to find the minimum loss.  
 222 Clustering techniques are often used to improve the grouping of similar datapoints in la-  
 223 tent space by using both the reconstruction loss ( $L_{rec}$ ) and clustering loss ( $L_{clust}$ ) (Aljalbout  
 224 et al., 2018; Min et al., 2018). The purpose of  $L_{rec}$  is to learn realistic features, whereas  
 225  $L_{clust}$  promotes discrimination and grouping of feature points within the latent space (Min  
 226 et al., 2018). When using deep clustering, the loss function becomes:

$$L_{all} = (1 - \lambda)L_{rec} + \lambda L_{clust} \quad (3)$$

227 where  $\lambda \in \{0,1\}$  is a hyperparameter that balances  $L_{rec}$  and  $L_{clust}$  and should be set  
 228 to prevent over/under fitting of the model for a given dataset. If set too low, over-fitting  
 229 will occur as the model has learnt too much about the noise in the data, limiting its abil-  
 230 ity to identify characteristic features of each class. In contrast, if set too high under-fitting  
 231 occurs as the model becomes too simplistic and overlooks key patterns in the data.  $L_{clust}$   
 232 can be obtained by calculating the Kullback-Leibler (KL) divergence loss between the  
 233 soft assignment probability of sample  $i$  belonging to cluster  $j$  with an auxiliary target  
 234 distribution using the following equation (Xie et al., 2016):

$$L_{clust} = KL(P||Q) = \sum_i \sum_k p_{ik} \log \frac{p_{ik}}{q_{ik}} \quad (4)$$

235 where  $p_{ik}$  and  $q_{ik}$  are the  $i_{th}$  sample of the  $k_{th}$  cluster of the target ( $P$ ) and soft ( $Q$ ) prob-  
 236 ability distributions (Van der Maaten & Hinton, 2008). Calculating all soft assignments  
 237 for a sample produces probability distribution  $Q$ , whereas the target probabilistic dis-  
 238 tributions ( $P$ ) are derived by squaring  $q_{ik}$  and normalizing by the sum of its soft cluster  
 239 frequencies:

$$q_{ik} = \frac{(1 + \|h_i - \mu_k\|^2)^{-1}}{\sum_{k'} (1 + \|h_i - \mu_{k'}\|^2)^{-1}} \quad (5)$$

$$p_{ik} = \frac{q_{ik}^2 / f_k}{\sum_{k'} q_{ik'}^2 / f_{k'}} \quad (6)$$

240 where  $h_i = f_\phi(x_i)$ ,  $\mu_k$  is the centroid of cluster  $k$ , and  $f_k = \sum_i q_{ik}$  is the soft cluster  
 241 frequency. Making use of  $h$ , which is a compact version of the original input, allows auto-  
 242 encoders to pick out only the most salient features in the data.

### 243 **2.1.5 Location Guided Autoencoder**

244 Spatial information is important in many applications, and while CNNs can find  
 245 patterns within an image, many spatial patterns are larger than the footprint of a sin-  
 246 gle image cropped from a larger scene and CNNs cannot correlate these patterns. In re-  
 247 sponse, Yamada et al. (2021) developed a novel location guided autoencoder (LGA) for  
 248 automated semantic interpretation of seafloor images that utilizes 3D geolocational meta-  
 249 data. Their base autoencoder for feature extraction uses AlexNet (Krizhevsky et al., 2012),  
 250 where the encoder is AlexNet’s original architecture and the decoder is an inverted ver-  
 251 sion of the encoder (Fig. 2Bii). The LGA was designed with the assumption that “two  
 252 images captured close together look more similar than those far apart”. Using this as-  
 253 sumption, the position of data in the latent space ( $h_i$  and  $h_j$ ) is modified by account-  
 254 ing for the distance between the locations ( $y_i$  and  $y_j$ ) of the original images ( $x_i$  and  $x_j$ ) (Fig. 2Aii).  
 255 The assumption can then be applied by using a Gaussian distribution as a kernel to quan-  
 256 tify the affinity between  $h$  and geographical space ( $y$ ) (Fig. 2Biv):

$$q'_{ij} = \frac{(1 + \|h_i - h_j\|)^{-1}}{\sum_{i'} \sum_{j'} (1 + \|h_{i'} - h_{j'}\|^2)^{-1}} \quad (7)$$

$$p'_{ij} = \frac{(1 + d(y_i y_j))^{-1}}{\sum_{i'} \sum_{j'} (1 + d(y_{i'} y_{j'}))^{-1}} \quad (8)$$

257 where  $q'_{ij}$  and  $p'_{ij}$  are the values of the affinity matrices at index  $(i, j)$  in the latent space  
 258 ( $Q'$ ) and physical space ( $P'$ ) respectively, and  $d(y_i, y_j) = \min \|y_i, y_j\|^2 d_{max}^2$ . In this con-  
 259 text  $d_{max}$  is the user-defined maximum distance between two locations that will be cor-  
 260 rected and will vary on the application domain and scale of the image scene. The LGA  
 261 is trained to minimize the KL divergence between  $Q'$  and  $P'$  using the following loss func-  
 262 tion:

$$L_{all} = L_{rec} + \lambda L_{geo} = L_{rec} + KL(P' \| Q') \quad (9)$$

263 This approach results in  $h_i$  and  $h_j$  being moved closer together in feature space if they  
 264 are close in physical space.

### 265 **2.1.6 Contrastive Learning**

266 Contrastive learning is an unsupervised machine learning technique that attempts  
 267 to learn features in an image by comparing similar pairs of images close together in  $h$   
 268 to a random dissimilar pair embedded far apart in  $h$ . The aim of this comparison is to  
 269 maximize the similarity between positive pairs (images that look similar) and minimize  
 270 the similarity between negative pairs (images that look dissimilar). An issue with con-  
 271 trastive learning is that you must confirm that the positive pair of images are indeed sim-  
 272 ilar. In response, Chen et al. (2020) developed a framework for self-supervised contrastive  
 273 learning of visual representations (SimCLR) that attempts to improve agreement between  
 274 variably augmented images ( $x_i$  and  $x_j$ ) derived from the same original image ( $x$ ). At each

275 training iteration, a minibatch (i.e. a small subset) of  $N$  images is taken for augmenta-  
 276 tion. During augmentation, random cropping, colour distortion and Gaussian blur are  
 277 applied before a CNN is used as a base encoder ( $f(\cdot)$ ) that extracts representations, known  
 278 as feature vectors ( $h_i$ ), from the augmented images ( $h_i = f(x_i)$ ) (Fig. 2Cii). These vec-  
 279 tors then act as the input for a projection head ( $g(\cdot)$ ) consisting of a two-layer multi-layer  
 280 perceptron (MLP), which produces an embedding ( $z_i = g(h_i)$ ) that is mapped to a la-  
 281 tent space (Fig. 2Ciii) where the following loss function is applied to compute the con-  
 282 trastive loss ( $\ell$ ):

$$\ell_{i,j} = -\log \left( \frac{\exp \left( \frac{\text{sim}(z_i, z_j)}{\tau} \right)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp \left( \frac{\text{sim}(z_i, z_j)}{\tau} \right)} \right) \quad (10)$$

283 where  $\text{sim}()$  is the cosine similarity;  $\tau$  is a temperature parameter that controls the penalty  
 284 given to hard negative samples, which controls the smoothness of the probability distri-  
 285 bution (Wang & Liu, 2021; Kumar & Chauhan, 2022); and  $\mathbb{1}_{[k \neq 1]} \in \{0,1\}$  is the indi-  
 286 cator function, which is set to 1 when  $k \neq 1$ . The total loss ( $L$ ) for the minibatch can  
 287 then be calculated as:

$$L = \frac{1}{2N} \sum_{k=1}^N [\ell(2k-1, 2k) + \ell(2k, 2k-1)] \quad (11)$$

288 At each training iteration a stochastic gradient descent (SGD) optimizer with linear rate  
 289 scaling is used to update the base encoder and projection head parameters toward the  
 290 fastest training loss (Goyal et al., 2017). Fine-tuning of a CNN trained using SimCLR  
 291 also showed improved accuracy even with two orders of magnitude fewer hand labelled  
 292 images provided (Chen et al., 2020).

### 293 **2.1.7 GeoCLR**

294 Although the method proposed in SimCLR works well to present individual sim-  
 295 ilar and dissimilar images, it does not account for spatial patterns with footprints larger  
 296 than a single image. To overcome this limitation, Yamada, Prügel-Bennett, et al. (2022)  
 297 developed ‘georeference contrastive learning of visual representation’ (GeoCLR) to ef-  
 298 ficiently train CNNs by leveraging georeferenced metadata. Their dataset consisted of  
 299 86,772 seafloor images collected by an autonomous underwater vehicle (AUV) from a sin-  
 300 gle locality, and each image had an associated depth, northing and easting. In summary,  
 301 GeoCLR generates a similar image pair ( $\tilde{x}_i$  and  $\tilde{x}'_j$ ) from two different images ( $x$  and  $x'$ )  
 302 that are close together in physical 3D space (Fig. 2Ci). Image  $x$  possesses a unique ge-  
 303 olocation ( $g_{east}, g_{north}, g_{depth}$ ) and image  $x'$  is then selected from a batch of images with  
 304 a 3D geolocation ( $g'_{east}, g'_{north}, g'_{depth}$ ) within a given distance ( $r$ ) of image  $x$  provided  
 305 it meets the following criteria:

$$\sqrt{(g'_{east} - g_{east})^2 + (g'_{north} - g_{north})^2 + \lambda(g'_{depth} - g_{depth})^2} \leq r \quad (12)$$

306 A scaling factor ( $\lambda$ ) is used to include or exclude images that are close but at different  
 307 depth. Once image pairs are selected the same augmentations are applied as SimCLR  
 308 to generate the similar image pair ( $x_i$  and  $\tilde{x}'_j$ ) (Yamada, Prügel-Bennett, et al., 2022).  
 309 Using a semi-supervised framework, the average classification accuracy of GeoCLR was  
 310 10.2 % higher than an identical CNN trained using SimCLR alone, highlighting the value  
 311 of utilizing geolocational metadata when using a latent space for feature extraction.

312

## 2.2 Adapting Spatial Machine Learning for Drill Core Imagery

313

314

315

316

317

318

319

320

321

322

323

324

325

Here we present a modification of LGA and GeoCLR that involves calculating 2D cylindrical (x-y) coordinates, instead of 3D Cartesian coordinates, to guide semantic interpretation of a 2D core image (Fig. 2Aii). Typically, images taken during scientific coring operations include: 2D scans of a cut surface of a core section half, 2D images of core sections (either cut or uncut) in a core box, or 3D line scans taken on a 360 degree core scanner that images the outer surface of the uncut core. As 2D images are more common, and when a 3D image is unwrapped it is also 2D (Fig. 2Ai), spatial metadata accompanying a given cropped patch from a core image is a 2D x-y coordinate. All these image formats capture visual information about the rocks in the form of a three-channel (RGB) 2D array where the top and bottom of the image have an associated depth down hole. Cores also have different diameters depending on the drill bit used to collect them and this information can be used to calculate the horizontal position of a given patch ( $s_i$ ) as a function of the minimum ( $m_i$ ) and maximum ( $M_i$ ) width of the original image:

$$s_i = m_i + \frac{f_i}{\left(\frac{M_i}{n}\right)}(M_i - m_i) \quad (13)$$

326

327

328

329

Where  $n$  is the number of adjacent patches that fit horizontally into  $M_i$  and depends on the image resolution and user defined patch size, and  $f_i \in \{0, \dots, \left(\frac{M_i}{n}\right)\}$  is the horizontal patch index. Similarly, the vertical position ( $s_j$ ) of each patch can be calculated in the same fashion:

$$s_j = m_j + \frac{f_j}{\left(\frac{M_j}{n}\right)}(M_j - m_j) \quad (14)$$

330

331

332

333

334

335

336

Where  $m_j$  and  $M_j$  are the minimum and maximum depth of the original image and  $f_j \in \{0, \dots, \left(\frac{M_j}{n}\right)\}$  is the vertical patch index. Our proposed workflow calculates a horizontal 2D spatial location, or polar coordinate, for a given patch and combines this with the depth downhole the patch is from to give an x-y coordinate ( $s_i, s_j$ ) which is used to determine how close patches are in physical space. Following the methods described above for GeoCLR, our polar coordinate system is used to select  $\tilde{x}'$  from a batch of images with a spatial location ( $s'_i, s'_j$ ) that meets the following criteria:

$$\sqrt{(s'_i - s_i)^2 + (s'_j - s_j)^2} \leq r \quad (15)$$

337

338

339

340

Patch pairs ( $x_i$  and  $\tilde{x}'_j$ ) then go through the same augmentations used by SimCLR and GeoCLR to extract features from the input data. In contrast, the LGA was modified to use ( $s_i, s_j$ ) when quantifying the affinity between  $h$  and  $y$  using a Gaussian distribution as a kernel, where sigma is set to  $d_{max}$ .



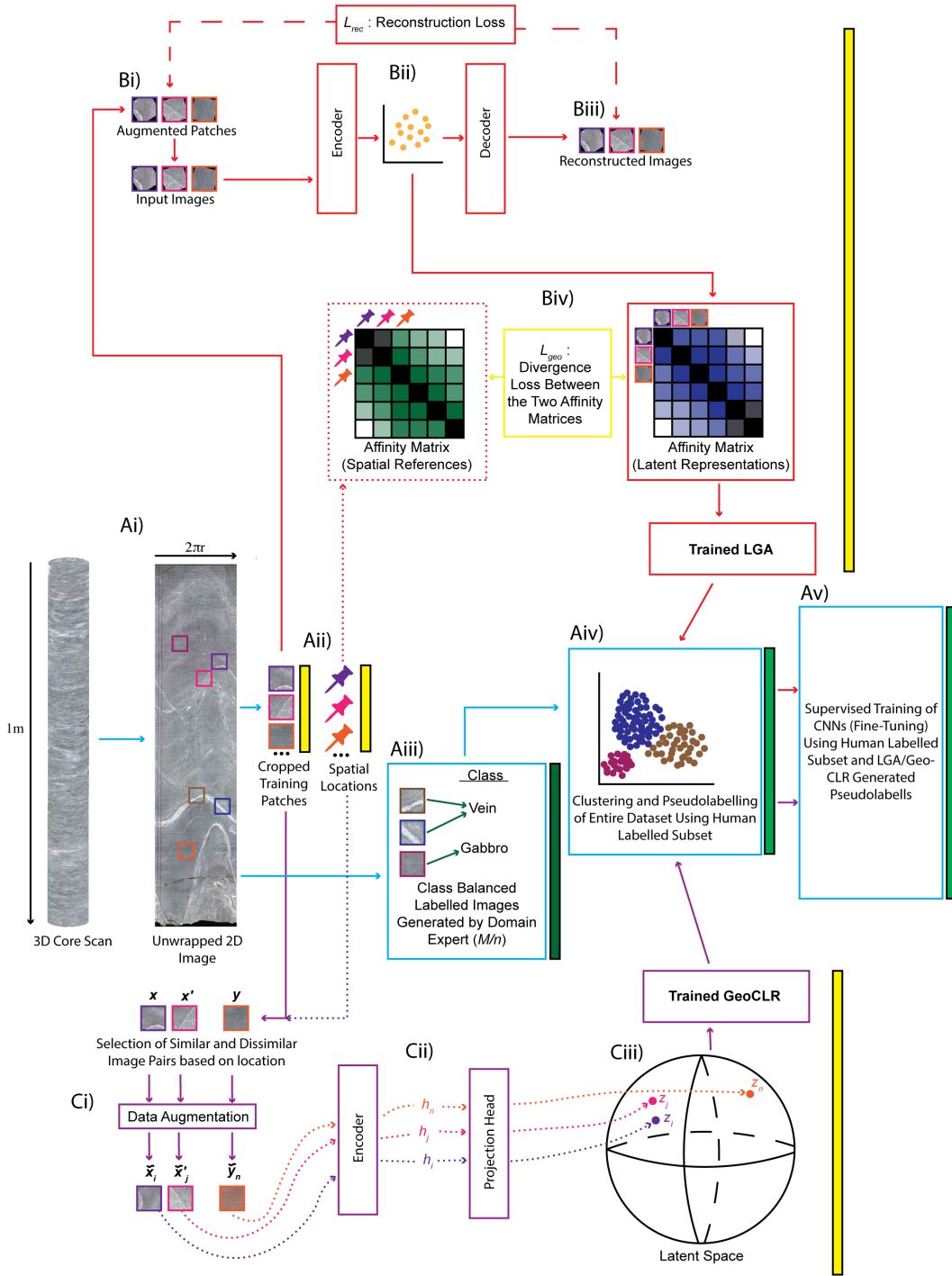


Fig. 2: Diagram of the semi-supervised workflow used in this study. Unsupervised feature extraction used both the LGA (red path) and GeoCLR (purple path) to create latent representations of the dataset, and the workflow involves image processing, sampling, labelling and clustering prior to fine tuning of the pre-trained CNNs generated by LGA and GeoCLR (blue path). Yellow bars indicate completely automated steps, whereas dark and light green bars indicate supervised and semi-supervised steps, respectively. Modified after Yamada *et al.* (2021, 2022).



## 2.3 Experiment and Workflow

In this study the performance of frameworks that utilize the spatial context of training images (GeoCLR and LGA) are compared to equivalent methods that do not use this context (SimCLR and AE) (Table 1). Additionally, a 4800 (400/class) image subset was used for supervised training of ResNet18 to benchmark against the performance of unsupervised learning results. A summary of the models tested in this study can be seen in Table 1.

### 2.3.1 Dataset

All images used in this study are of core recovered from Oman Drilling Project (OmanDP) Hole GT1A drilled into gabbroic rocks from the Semail ophiolite (Fig. 3), an ancient slab of ocean crust preserved on the Arabian margin (Kelemen et al., 2020). All cores were imaged using a DMT CoreScan3 digital line scanner which rotated them about their cylindrical axis as the DMT incrementally imaged the full length of the core exterior. Cores were imaged one section at a time, and each section was no longer than 1 m, as this was the maximum length the scanner could fit. Each section had a blue and red crayon line drawn along its length to indicate way up and as a guide for where it was to be cut into an archive (preserved for future reference) and working (for sampling) half. When orientated to its original vertical position, the blue line is to the left of the red. The total depth of Hole GT1A is 403.4 m; cores collected from the upper 254.2 m were drilled with an HQ diamond bit yielding core with a diameter of 63.5 mm (1995 pixels). Below this depth, coring used a narrower PQ bit and cores are 47.8 mm in diameter (1493 pixels) (Kelemen et al., 2020). All images were taken at a 10 pixel/mm resolution and stored as bitmap files.

Core exterior images collected during the OmanDP were an excellent candidate for this study due to the large amount of data accompanying them in the form of VCDs and detailed core logs generated by expert geologists. Therefore, all labelling of training and validation images in this study were cross referenced and groundtruthed to these data, as well as confirmed by the geologists involved in the description of these cores.

### 2.3.2 Training Image Preparation

Raw bitmap images were prepared for training by: 1) transposing to the correct vertical orientation, 2) cropping any valueless pixel columns from image borders, 3) ‘rotating’ the image horizontally until the blue cutting line was at 100 pixels from the left of the image (Fig. 2Ai). Many of the images had been rotated more than 360° during scanning, making the apparent resolution of 10 pixels/mm inaccurate. However, this only duplicates ~20 pixels either side of the vertically rotated raw image. In some cases, images were over-rolled (>540°), which was resolved by cropping them to the correct width of 1995 or 1493 pixels, depending on core diameter. Uneven surfaces appear as visual interference, particularly at either end of a section with angular contacts with the sections above or below it. Spurious reflections are also present where tape was used to hold fractured core together during scanning, or where foam was used as a spacer in some cases where material was too fragmented to scan. Once prepared, all section images were segmented to produce 722,157 100x100 pixel (1 cm<sup>2</sup>) patches that were used to train the machine learning models (Fig. 2Aii). Patch size was chosen to be small enough to avoid multiple classes occurring in a single image, but large enough to be labelled by an expert (Fig. 4).

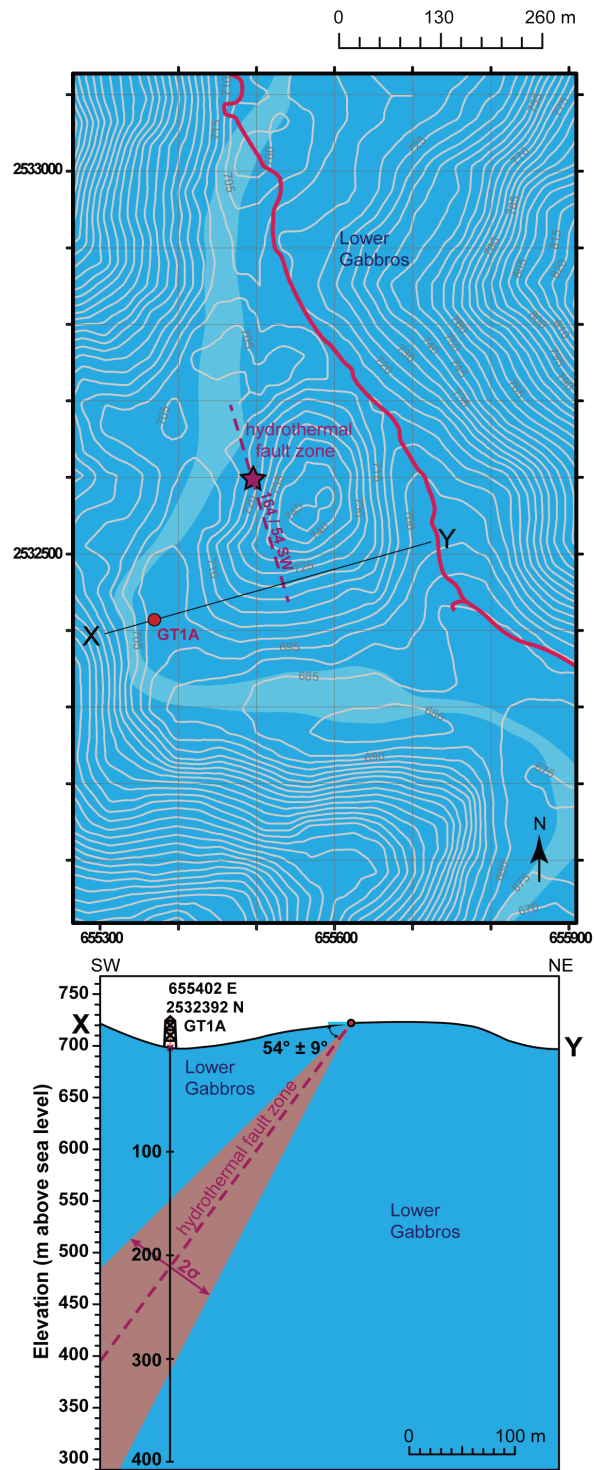


Fig. 3: Location and cross section of the Hole GT1A drill site, Oman. From Kelemen et al. (2020)

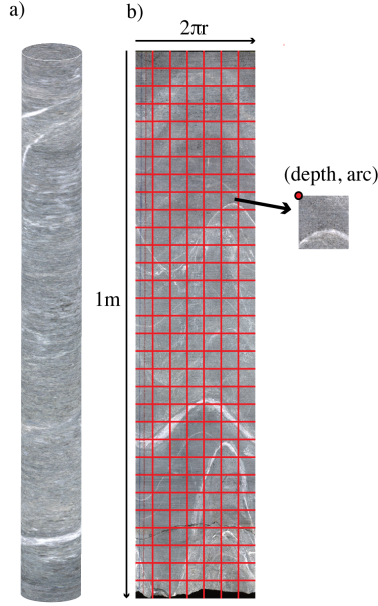


Fig. 4: a) Diagram of a 3D scan of a section of core, and b) the unrolled 2D version of (a) with an example of the segmentation style used to generate training patches used in this study (red grid not to scale). The top left corner of each training patch is the location of the patch's depth and arc position on the core surface (right).

388

### 2.3.3 Self-supervised Learning Configuration

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

Configurations for all models are set to those deemed optimal by Yamada et al. (2021) and Yamada, Prügel-Bennett, et al. (2022) during their development of the LGA and GeoCLR methods, except for threshold closeness ( $d_{max}$  and  $r$ ) and number of training cycles. All training patches were expanded to  $227 \times 227$  pixels during feature extraction for the AE and LGA, as this was the size required by the AlexNet-based autoencoder. In contrast, SimCLR and GeoCLR methods re-scale each patch to a resolution of 2 mm/pixel and randomly crop out a  $224 \times 224$  region for use during training (Yamada, Prügel-Bennett, et al., 2022; Yamada et al., 2021). The number of dimensions in latent space ( $h$ ) for the autoencoders is set to 16, whereas for SimCLR and GeoCLR it is set to 128. For all frameworks, the number of images fed into the model at each training iteration (mini-batch) was set to 256 and training ran for 200 iterations (epochs). Patches physically adjacent in all directions to  $x_i$  were deemed close enough spatially to assume they will look similar, therefore  $d_{max}$  and  $r$  were set to 1.5 cm. Hyperparameters such as learning rate and weight decay for all models were set to the optimal values determined during their development (see reference in Table 1).

404

### 2.3.4 Geologically Constrained Semi-supervised Clustering

405

406

407

408

409

410

411

412

A total of 12 classes were defined to be representative of the most common rock types and features that occur downhole within Hole GT1A (Kelemen et al., 2020) (Fig. 5). All 722,157 image patches were used during self-supervised learning, and two subsets of 100 and 300 per class were expert labelled for validation and training, respectively (Fig. 2Aiii). Several classes are not of geological interest so to avoid these features being incorrectly labelled, they were treated as distinct classes. These include spurious noise from tape and foam, as well as crayon lines and dark empty space. Gabbros in Hole GT1A were subdivided based on their colour, with light grey, more felsic, patches being termed sim-

413 ply ‘gabbro’. Gabbro with  $\sim 1\text{-}5\%$  darker minerals was termed ‘olivine-bearing gabbro’,  
 414 whereas patches with  $\sim 6\text{-}50\%$  dark minerals were referred to as ‘olivine gabbro’, and  
 415 patches containing  $\geq 50\%$  dark minerals were labelled as ‘mela-olivine gabbro’. Dark min-  
 416 erals in Hole GT1A are primarily a mix of olivine and clinopyroxene and distinguishing  
 417 between the two in the training images was not always possible. Therefore, all expert  
 418 labels given to patches were groundtruthed to the lithology and modal abundances recorded  
 419 for the appropriate interval in the OmanDP VCDs. Other classes considered of inter-  
 420 est for alteration logging included: veins composed of white minerals (vein type A), veins  
 421 that contain a mix of prehnite and chlorite (vein type B), ‘fracture’ and ‘alteration zone’,  
 422 which were also groundtruthed using the OmanDP vein and alteration logs (Kelemen  
 423 et al., 2020). Here alteration refers to parts of the core where primary igneous miner-  
 424 als have been replaced by secondary phases due to hydrothermal alteration and/or de-  
 425 formation, which occurs in Hole GT1A mostly as patches, halos and densely spaced vein  
 426 networks. Within Hole GT1A there is variability in the dominant secondary minerals  
 427 present in an alteration zone (Kelemen et al., 2020; Greenberger et al., 2021), however,  
 428 all were placed in a single class to capture zones of focused alteration. In many cases,  
 429 patches labelled as alteration zone could be confused as a type of vein if the annotator  
 430 only looks at the  $1\text{x}1\text{ cm}$  patch. However, when the spatial context of a patch revealed  
 431 that it sits within an altered interval, and is not part of a single linear vein, it was la-  
 432 belled as ‘alteration zone’.

433 For all experimental configurations a class-balanced approach was used where an equal  
 434 number of representative expert annotations per class ( $M/n$ ) were manually generated.  
 435 A class-balanced approach can be time consuming when compared to other selection meth-  
 436 ods (Yamada, Prügel-Bennett, et al., 2022). However, it ensures all labels provided are  
 437 representative of the high intra-class variation at the cm-scale in the rocks. Each model  
 438 was trained multiple times, varying  $M/n$  to find its optimal value. Labelling 100 images  
 439 for each of the 12 classes in this study took  $\sim 16\text{-}24$  hrs. Therefore, a maximum of  $M/n$   
 440  $= 300$  was chosen because the time taken to manually label more images would be in-  
 441 efficient in the context of real-time core analysis during a geological coring project. For  
 442 a given  $M/n$ , self-supervised training produced a latent representation of the dataset be-  
 443 fore a support vector machine with a radial basis function as a kernel (R-SVM) was used  
 444 to classify the data based on the expert annotated subset (Fig. 2Aiv). The outcome of  
 445 this classification is all images are assigned a computer-generated pseudolabel, which were  
 446 then compared to the expert labelled validation subset to quantify the accuracy of each  
 447 model. The best performing configuration for each model was then fine-tuned with the  
 448 pseudolabels generated by the R-SVM, and in all cases ResNet18 was used as the fine-  
 449 tuning classifier (Fig. 2Av).

450

### 451 **2.3.5 Supervised Training Configuration**

452 Supervised learning methods use labelled data that have corresponding target la-  
 453 bels or outputs, whereas unsupervised learning networks extract the underlying struc-  
 454 ture of the data with no target output. Unsupervised approaches are used in this study  
 455 to generate a latent space before  $M/n$  expert labelled images are provided for the au-  
 456 tomatic assignment of computer-generated pseudolabels to the entire dataset, which then  
 457 allow for fine tuning. Fine tuning of a neural network takes the initial pre-trained net-  
 458 work as a starting point before adjusting its parameters by re-training using a labelled  
 459 subset of the dataset in a supervised fashion. All semi-supervised frameworks trained  
 460 with  $M/n = 100$  and  $M/n = 300$  were fine-tuned by feeding the entire pseudolabelled  
 461 dataset into ResNet18 with a minibatch size of 128, learning rate and weight decay of  
 462  $1\text{x}10^{-5}$  and Adam optimizer (Kingma and Ba, 2014)(Fig. 2Aiv-v). Models trained with  
 463 other values of  $M/n$  were not fine-tuned as it would have been computationally burden-  
 464 some.



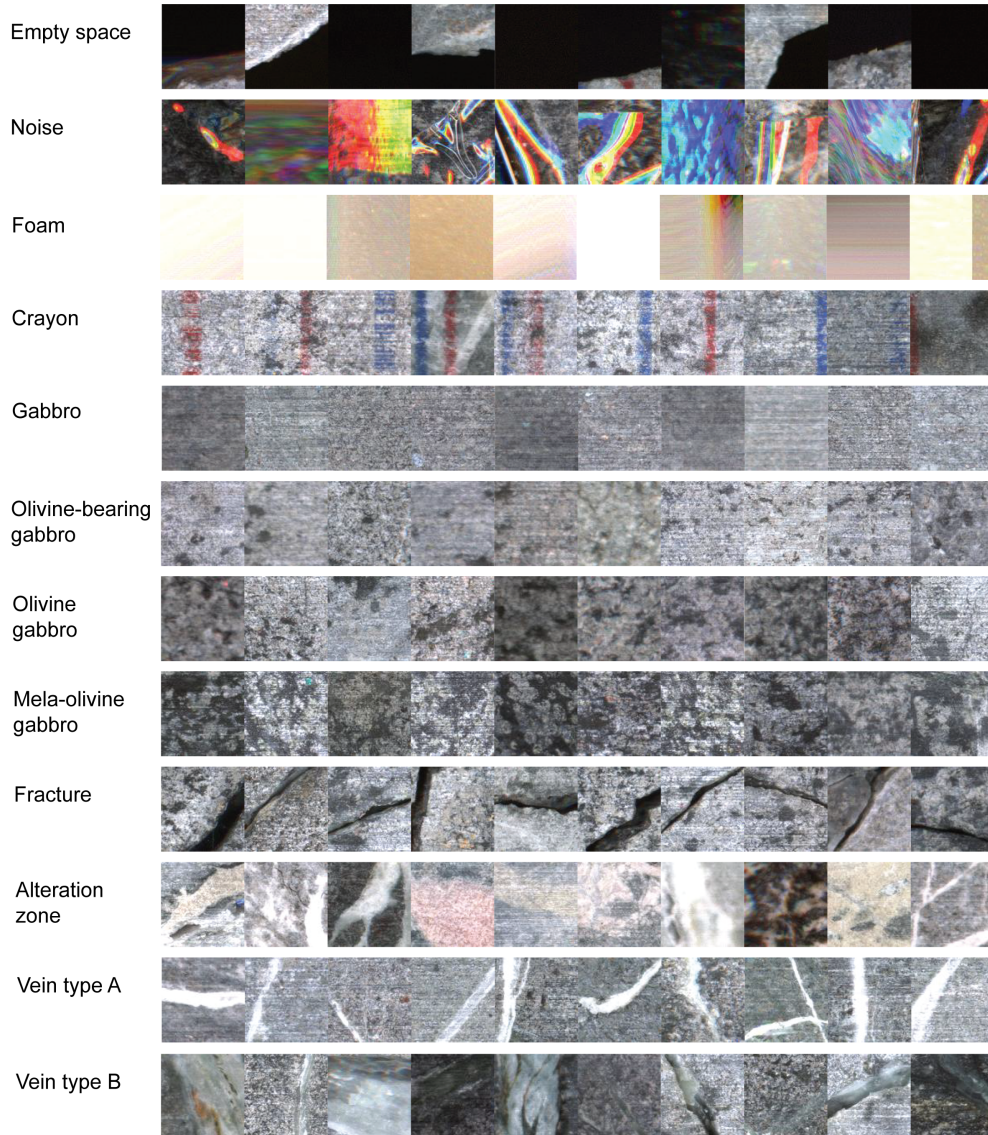


Fig. 5: Example images of expert defined classes that each model was trained to identify. These were chosen to represent the most common rock types in Hole GT1A as well as highlight areas of intensified hydrothermal alteration and fracturing. A total of 400 images were expert labelled per class, with 300 used for training and 100 for validation.

465 To quantify the improvement of using unsupervised feature extraction prior to fine tun-  
 466 ing over a simple supervised approach that would require the same expert labelling ef-  
 467 fort, the expert labelled training (300/class) and validation (100/class) images were also  
 468 used for supervised training of ResNet18. ResNet18 was pre-trained using ImageNet (Deng  
 469 et al., 2009), and its hyperparameters were set to the same as those used during fine-tuning  
 470 of the unsupervised frameworks. Training ran for 200 epochs with a batch size of 128  
 471 and the last layer of the network was set to the number of classes in the ImageNet database  
 472 (1000). This is because the last layer of the pre-trained ResNet18 model used for fine-  
 473 tuning is also 1000, due to the number of classes in the ImageNet database, which matches  
 474 the approach Yamada et al. (2021); Yamada, Prügel-Bennett, et al. (2022) took when  
 475 fine tuning CNNs trained using their LGA and GeoCLR methods.

## 476 2.4 Validation

477 When quantifying the performance of machine learning algorithms there are a num-  
 478 ber of commonly used performance metrics, such as accuracy, precision and recall. Pre-  
 479 vious attempts to use machine learning to classify core images have primarily reported  
 480 model performance using only accuracy. However, when the proportions of each class  
 481 within the training dataset are imbalanced accuracy can be inflated in cases where the  
 482 model does particularly well at classifying the most abundant classes. In the case of us-  
 483 ing unsupervised learning approaches, the relative abundance of each expected class in  
 484 the dataset is not known. Therefore, in this study we use the  $f_1$  score for each class to  
 485 quantify model performance as it accounts for both the model’s ability to correctly iden-  
 486 tify positive instances (precision) and to capture all positive instances (recall):

$$Precision = \frac{TP}{TP + FP} \quad (16)$$

$$Recall = \frac{TP}{TP + FN} \quad (17)$$

$$f_1 = 2 \frac{Precision \cdot Recall}{Precision + Recall} \quad (18)$$

487 Where true positive (TP), false positive (FP), true negative (TN) and false negative (FN)  
 488 results were generated by comparing the machine learning classifications given to the 1200  
 489 validation images labelled by domain experts. The overall performance of each model  
 490 is then presented in this paper as the class-averaged  $f_1$  score (macro average) where  $f_{1i}$   
 491 is the  $f_1$  score of class  $i$  and  $n$  is the total number of classes identified in the dataset:

$$f_{1(\text{macro average})} = 0.5 \sum_{i=1}^n f_{1i} \quad (19)$$

## 492 3 Results

### 493 3.1 Training Evaluation

494 In all cases GeoCLR showed best performance, and the  $f_1$  scores (macro average)  
 495  $\pm 1\sigma$  for all model configurations can be seen in Table 2. At all values of M/n, the AE  
 496 performed the worst, demonstrating that, without incorporation of spatial metadata, auto-  
 497 encoders are not suitable for classification of core images. Excluding the AE, increas-  
 498 ing M/n improved classification for all other models. The AE only showed minor improve-  
 499 ment up to M/n = 200 before accuracy began to decrease (Fig. 6). For the remaining  
 500 models, sigmoidal growth is seen where most of the improvement in accuracy occurs at

CNN framework	Training type	M/n					
		3	10	30	100	200	300
AE	semi-supervised	0.08 ± 0.04	0.08 ± 0.04	0.07 ± 0.04	0.09 ± 0.04	0.10 ± 0.05	0.10 ± 0.03
LGA	semi-supervised	0.33 ± 0.25	0.38 ± 0.27	0.48 ± 0.25	0.59 ± 0.21	0.60 ± 0.22	0.62 ± 0.21
SimCLR	semi-supervised	0.34 ± 0.23	0.49 ± 0.20	0.60 ± 0.19	0.69 ± 0.16	0.73 ± 0.15	0.74 ± 0.14
GeoCLR	semi-supervised	<b>0.40 ± 0.18</b>	<b>0.60 ± 0.14</b>	<b>0.74 ± 0.13</b>	<b>0.84 ± 0.07</b>	<b>0.86 ± 0.07</b>	<b>0.86 ± 0.07</b>
ResNet18	supervised	0.33 ± 0.34	0.34 ± 0.35	0.52 ± 0.31	0.67 ± 0.21	0.82 ± 0.13	0.84 ± 0.11

Table 2: Results of each model trained using the semi-supervised workflow presented in this paper, as well as supervised learning results for ResNet18. All models were trained using an increasing number of training images per class (M/n) and all results are  $f_1$  scores (macro average)  $\pm$  1SD.

501 the lower end between M/n = 3 and M/n = 100 (Fig. 6). Both the LGA and SimCLR  
502 show relatively large increases in accuracy at M/n >100 compared to GeoCLR, suggest-  
503 ing they would have further improved with M/n >300. At no point does the LGA out-  
504 perform contrastive learning or supervised methods, however it consistently outperforms  
505 the AE with a maximum of 52 % improvement. This indicates that introduction of spa-  
506 tial metadata when training auto-encoders drastically improves performance.  
507 Peak performance of GeoCLR is achieved with M/n =  $\sim$ 100, as performance only in-  
508 creases by  $\sim$ 2 % before plateauing with increased M/n. Furthermore, with only M/n =  
509 30, GeoCLR was able to outperform SimCLR and ResNet18 trained with an order of mag-  
510 nitude more annotations by 7 % and 5 %, respectively. Both contrastive learning meth-  
511 ods outperform ResNet18 at lower values of M/n, but at M/n >100 ResNet18 is more  
512 accurate than SimCLR and begins to achieve comparable performance to GeoCLR with  
513 increasing M/n. However, GeoCLR requires less domain expert effort to produce higher  
514 accuracy image classification than supervised (ResNet18) and black box (SimCLR) mod-  
515 els.

### 516 3.2 Class Identification

517 At lower values of M/n, the LGA outperforms the contrastive learning frameworks  
518 in correctly identifying non-geological classes, such as noise, foam and empty space. In  
519 contrast, both SimCLR and GeoCLR outperform the LGA in correctly distinguishing  
520 geological classes with fewer expert-generated labels (M/n <100). SimCLR correctly iden-  
521 tifies foam and empty space in almost all cases, however it fails to reliably distinguish  
522 crayon from the rock on which it was drawn. Regardless of increasing M/n, the LGA poorly  
523 distinguishes between classes containing single linear features, such as fractures and veins.  
524 The gabbroic rock classes share a lot of visual similarity, given they are defined by sub-  
525 divisions of a property that actually spans a spectrum of values (dark mineral abundance).  
526 This is particularly evident at the extreme ends of the colour index used to define them  
527 in this study. These shared characteristics cause both the LGA and SimCLR to mis-label  
528 5-8 % olivine-bearing gabbro as olivine gabbro, whereas GeoCLR only confuses 4 % and  
529 6 % of olivine-bearing gabbro for gabbro and olivine gabbro, respectively. For the more  
530 mafic-rich (higher proportion of dark mineral) classes, all models mis-label  $\geq$ 10 % of mel-  
531 olivine gabbro as olivine gabbro.

532

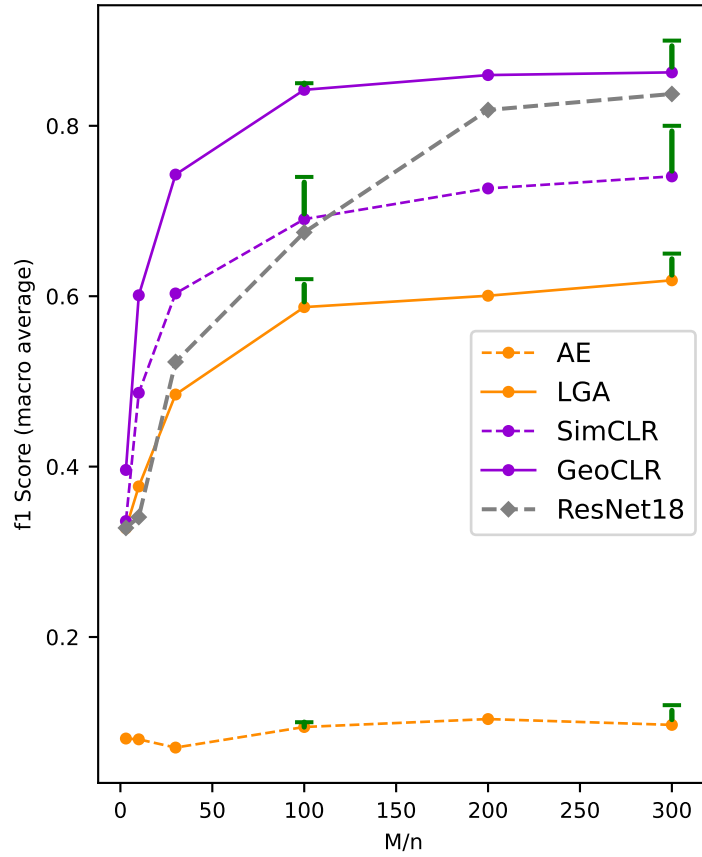


Fig. 6:  $f_1$  (macro average) scores of all models when 3, 10, 30, 100, 200, and 300 expert labelled images per class ( $M/n$ ) were used for training. Results of the contrastive learning methods are shown in purple, the results of the autoencoder methods are in orange, and the results of the supervised model are in grey. Solid lines indicate models that make use of spatial metadata and dashed lines are those that do not, whereas circles represent the unsupervised model results and diamonds supervised model results. Green lines indicate the performance increase gained by fine tuning.



M/n:	100	300
AE	0.09 ± 0.07	0.12 ± 0.06
LGA	0.62 ± 0.20	0.65 ± 0.19
SimCLR	0.74 ± 0.13	0.80 ± 0.10
GeoCLR	0.85 ± 0.08	<b>0.90 ± 0.05</b>

Table 3: Fine tuning results for each model given as  $f_1$  scores (macro average)  $\pm$  1SD.

533

### 3.3 Fine Tuning

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

Figure 7 compares fine-tuned networks pre-trained using the AE, LGA, SimCLR and GeoCLR frameworks to ResNet18. This comparison serves as an indicator of how well the semi-supervised methods outlined in this paper compare to commonly used supervised image classification techniques (Krizhevsky et al., 2012; K. He et al., 2016). Specific  $f_1$  scores were generated by averaging the scores of related groups of classes to highlight how well models classify geological ( $f_{geo}$ ), linear ( $f_{linear}$ ), bulk-rock ( $f_{geo}$ ) and noisy ( $f_{noise}$ ) classes (Fig. 7). All models except the AE are effective at filtering out noisy classes not of geological interest, whereas linear classes are those most often mis-classified. Yamada et al. (2021) demonstrated that their LGA improved the classification accuracy of linear classes to 53.7 %, as they had a characteristic spatial distribution. In this study linear features were the least well classified, even with spatial metadata, as the LGA gave an  $f_{linear}$  of  $0.43 \pm 0.07$ . Like the LGA, ResNet18 and SimCLR gave a relatively low  $f_{linear}$  when compared to  $f_{geo}$  and  $f_{bulk}$ , but are still more accurate than the LGA as they all have  $f_{linear} > 0.75$ . Unlike all other models, GeoCLR shows almost no variation between its ability to classify linear, bulk and geological features, and all have  $f_1$  scores of  $0.87 \pm 0.01-0.05$ . This consistent accuracy across class types, combined with its high  $f_{all} = 0.90 \pm 0.01$  and low error confirm that GeoCLR outperforms all other models evaluated in this study (Table 3). The classifications made by a fine-tuned model trained using our modified GeoCLR framework can therefore reliably be used to visualise and quantify geological features within a borehole.

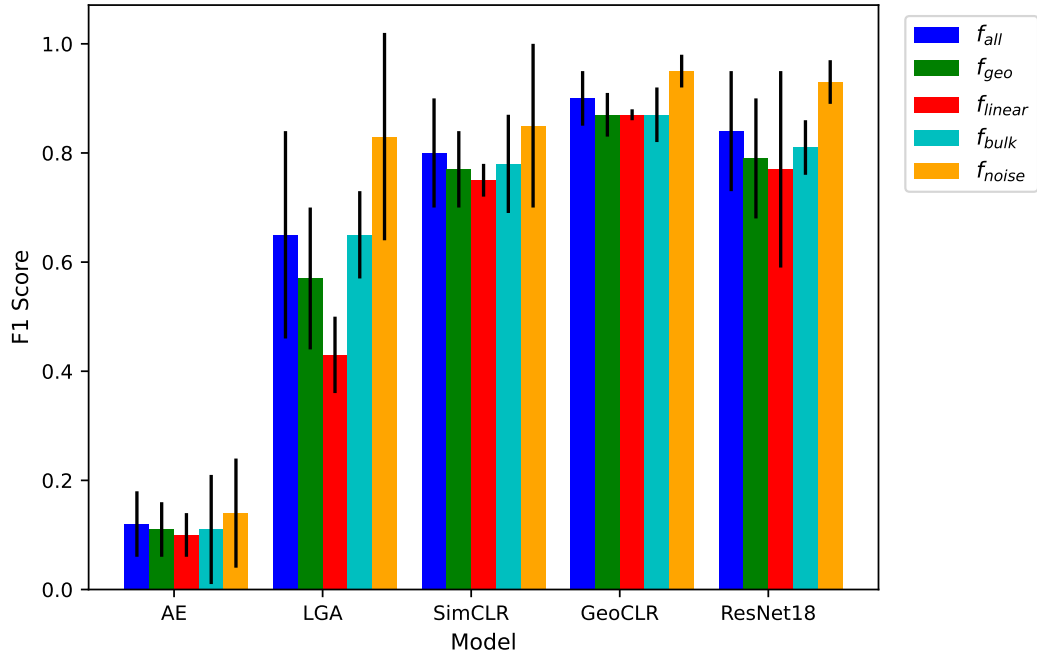


Fig. 7: Class averaged  $f_1$  scores for each model;  $f_{all}$  is the macro average across all classes;  $f_{geo}$  is the average score for geological classes only (gabbro, olive bearing gabbro, olivine gabbro, mela olivine gabbro, alteration zone, fracture, vein type A, vein type B);  $f_{linear}$  is the average score for linear classes (fracture, vein type A, vein type B);  $f_{bulk}$  is the average score for all bulk rock classes (gabbro, olivine bearing gabbro, olivine gabbro, mela olivine gabbro).  $f_{noise}$  is the average score for all non-geological classes. All errors are shown as 1SD.

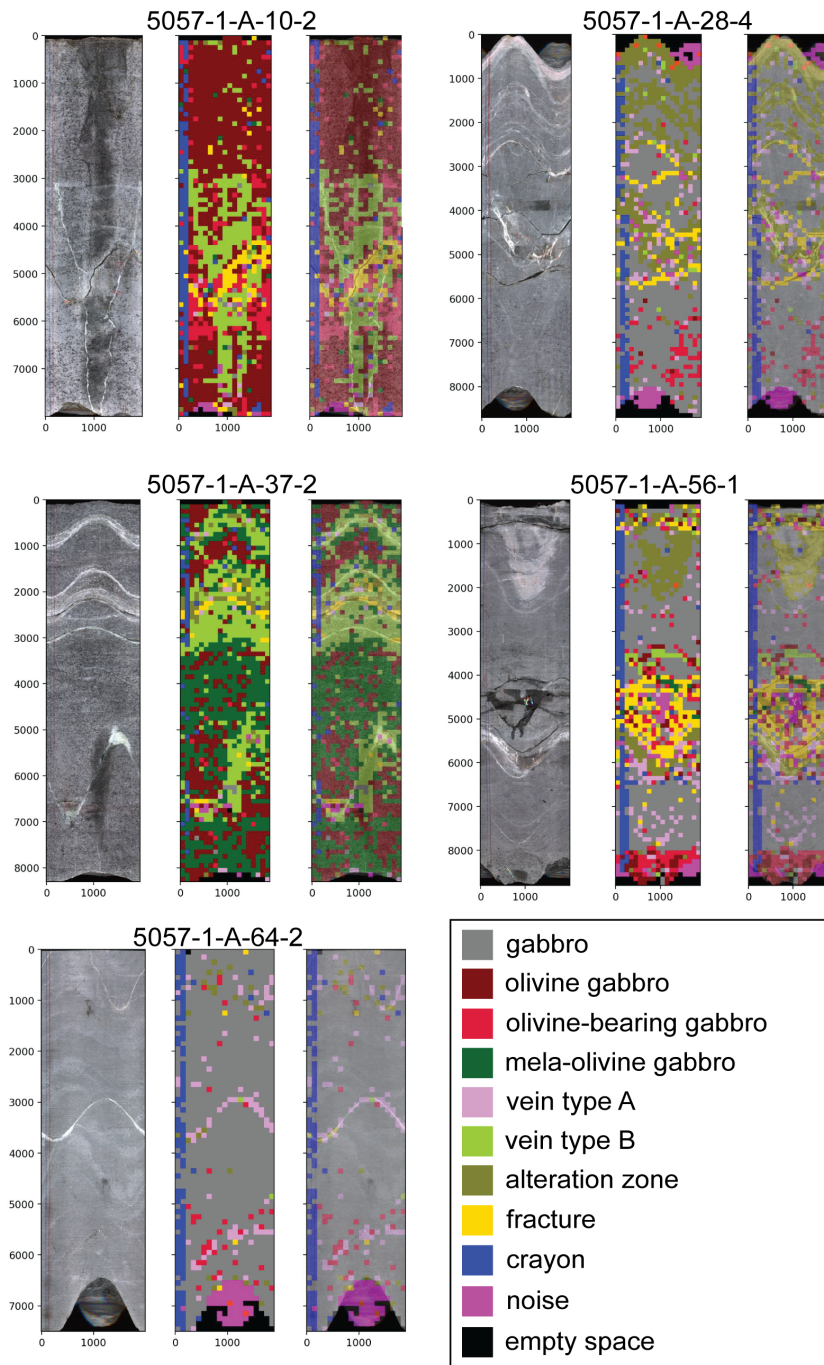


Fig. 8: Visual comparison of original Hole GT1A core-section images to the classifications given to patches taken from each section. For each section above, the left image is the original 360° DMT image, the middle image is constructed using the patches from the original section where colour corresponds to the class label generated using GeoCLR, the right image is the classified patch image overlaid above the original DMT image.

#### 4 Automated Alteration Logging

Plotting the relative abundance of fresh and altered rock downhole highlights regions of focused hydrothermal alteration within the ocean crust (Alt et al., 2010; Kelemen et al., 2020; Coggon et al., 2022; Teagle et al., 2023). During scientific drill core description alteration petrologists gather this data using visual estimations of alteration extent. The scale at which estimations are made often varies between expeditions and the quantification has an element of subjectiveness. Here we present a novel and automated approach to evaluating the spatial variations in the alteration extent downhole using the classifications generated by GeoCLR as a demonstration that AI-based approaches can standardise time-intensive geological tasks. Validation of our AI-based method is done by comparing it to an equivalent dataset generated by experts during the OmanDP. The expert-generated alteration data for Hole GT1A includes visual estimations of the average proportion of alteration features (halos, patches and deformation), as well as relatively fresh background rock within continuous downhole intervals. The depth and length of these intervals were defined by distinct changes in the nature/extent of alteration. To allow comparison with the cm-scale AI-based data through Hole GT1A, we assume that the proportions of alteration features in a given interval are representative of each centimeter of core in that interval. This assumption allowed a continuous downhole visual core description-based (VCD-based) estimate of the extent of alteration and background rock to be calculated by summing the proportions of all alteration types in an interval. A comparable depth-resolution dataset was then generated from the AI-based core logging data by calculating the percentage of patches labelled as 'alteration zone' by GeoCLR at each cm downhole (Fig. 8). Similarly, the proportion of images labelled as a class of gabbroic rock was used to infer the amount of relatively fresh background rock in each cm downhole. GeoCLR classified images of 'alteration zone' with an  $f_1 = 0.9$ , although 3 % and 5 % of the validation dataset were mis-labelled as foam and vein type A, respectively. Foam was inserted into regions too altered and fractured to be scanned on the DMT core scanner, and veins occur in conjunction with high levels of alteration in the core. Therefore, the presence of these classes are indicative of alteration, so their misclassification is not expected to significantly bias a plot of alteration extent downhole.

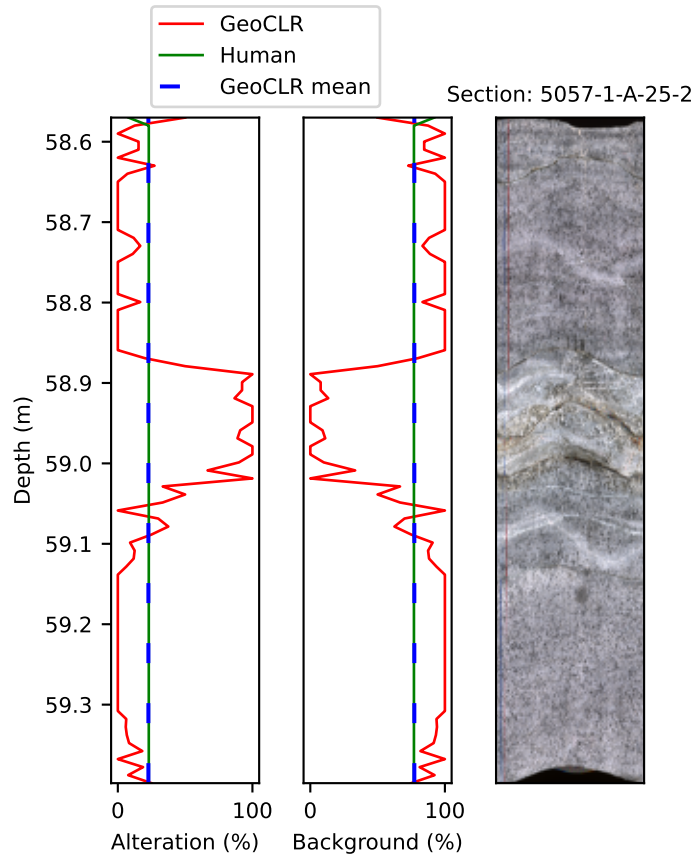


Fig. 9: Proportions of alteration and relatively fresh background rock through an example Hole GT1A core-section calculated using classifications generated using GeoCLR (red line) are compared to the equivalent data generated by alteration petrologists during the Oman Drilling Project (green line). The mean GeoCLR values through this interval (blue dashed lines) show excellent agreement with the visual core description-based estimates made by human experts. The vertical and horizontal scales of the core section image are equal.

586

587 The 1 cm depth resolution of the AI-generated data reveals high frequency shifts in alteration and background extent, whereas the lower-resolution VCD-based data displays sharp step-wise shifts between alteration intervals, which results in only a moderately positive correlation between the datasets (Table 4). Close inspection of a given 1 m section reveals that the AI-based data is capable of picking out small localised spatial variations in alteration that would be impractical for an expert to log (Fig. 9). However, the mean AI-based estimates of alteration extent and proportion of background rock through a given section show good agreement with the VCD-based estimates through the same interval (Fig. 9), further confirming that the AI-based approach is capable of identifying and quantifying geologically significant features identified by the experts - albeit at higher-resolution. To better visualise the broad variations in alteration extent within Hole GT1A, downhole-running averages for every 1 m (length of a core section) and 4 m (length of a full core) were also calculated to smooth both the AI and VCD-based data (Fig. 10). On average the AI-based data at a given depth is 1-2% higher than the VCD-based data,

600

601 however, the Pearson's Coefficient for the running averages of 1 m and 4 m shows sta-  
602 tistically significant positive correlations between the AI and VCD-based datasets (Ta-  
603 ble 4).

604 Overall, the large-scale variations in the smoothed VCD-based data are also captured  
605 by the AI-based data, and only two major discrepancies are observed at 146 m and 365  
606 m (Fig. 10). The first of these discrepancies occurs where a highly fractured interval has  
607 been visually identified as 40.5 % altered, whereas GeoCLR defined most of the inter-  
608 val as 'fracture'. The second occurs where GeoCLR underestimates the amount of back-  
609 ground rock by classifying patches of gabbro at this depth as 'crayon', highlighting the  
610 importance of minimising the markings made to the core surface prior to imaging.

611 The comparable performance of our AI-based approach using images alone to tradition-  
612 ally labour-intensive on-site core description demonstrates that AI methods have the po-  
613 tential to revolutionise current practices in the field. Specifically, rather than dedicat-  
614 ing time to visually quantifying features experts could dedicate more time to discrete sam-  
615 ple analysis or carrying out more detailed analysis of important intervals. Also, experts  
616 could dedicate time to labelling training images on-site while core is on display, as this  
617 would further ground classifications to the actual recovered material. One limitation, how-  
618 ever, is that cores are imaged one section at a time, so model training could not com-  
619 mence until drilling operations at a site are complete. Also, depending on the amount  
620 of recovered material, training time may take too long to be done on-site forcing AI-based  
621 approaches to be postponed until post-expedition. Regardless, it is clear that modify-  
622 ing on-site workflow with approaches such as that outlined in this work in mind would  
623 save significant amounts of time during a given coring campaign.

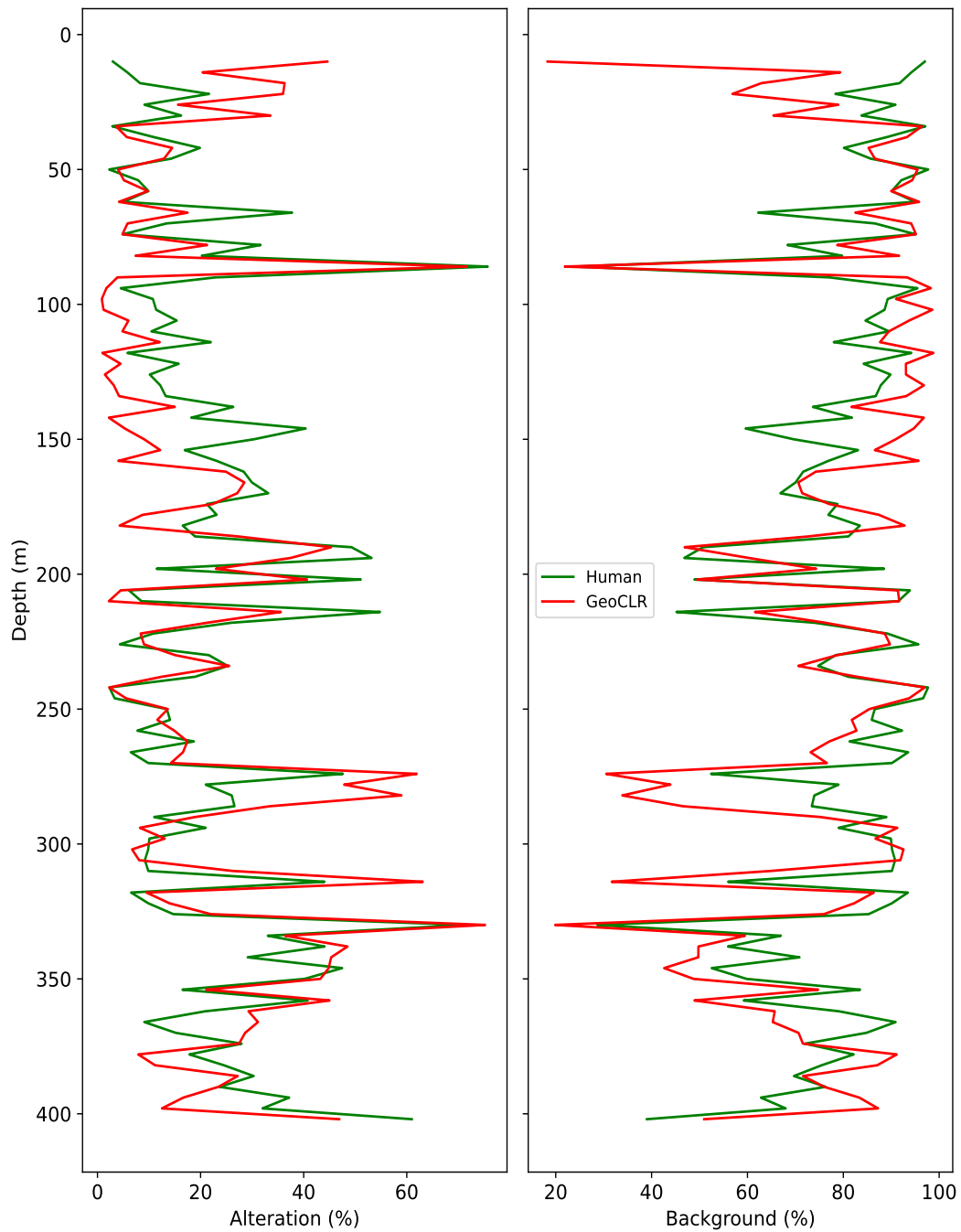


Fig. 10: Running average (window size = 4 m) for the abundance (%) of altered rock (left) and fresher protolith (right) downhole within Hole GT1A for both VCD-based estimations (green line) and AI-based estimations made using GeoCLR classifications (red line). As the VCD-based data sums to 100 % of the core surface, the AI-based data was normalized to also sum to 100 %.

Dataset		Window Size (cm)		
		1	100	400
Alteration%	<i>r</i>	0.50	0.71	0.72
	<i>p</i> value	0.00	$6.05 \times 10^{-61}$	$3.61 \times 10^{-17}$
	<i>n</i>	39784	392	99
Background%	<i>r</i>	0.48	0.68	0.65
	<i>p</i> value	0.00	$1.79 \times 10^{-54}$	$2.74 \times 10^{-13}$
	<i>n</i>	39784	392	99

Table 4: *Pearson's Coefficient (r)* and *p values* calculated by comparing the VCD-based and AI-based alteration log data. Analysis was performed for three different depth resolutions: 1 cm; and for running averages calculated using 1 m and 4 m window sizes. *n* indicates the number of data points compared for each iteration

## 5 Summary

This study presents a novel semi-supervised machine learning approach for the analysis and classification of geological images that utilizes spatial metadata for improved machine learning accuracy that can be implemented into existing CNN architectures. This method can be applied to any Earth or space image data sets that have accompanying spatial metadata, and implementing this workflow into several state-of-the-art machine learning frameworks has demonstrated that:

1. When only 30 labeled images per class are used for training, incorporating spatial metadata improves the classification accuracy of unsupervised auto-encoder and contrastive learning frameworks by 30 % and 11 %, respectively. Increasing this to >100 images per class further improves performance over non-spatially guided auto-encoders and contrastive learning by 50.7 % and 13.3 %.
2. Of the unsupervised learning models tested, spatially guided contrastive learning (GeoCLR) had the best classification accuracy, regardless of the number of expert-generated annotated images used for training. GeoCLR outperforms both non-spatially guided and supervised methods with an order of magnitude fewer expert-generated annotations and reaches maximum accuracy with  $\sim 100$  annotated images per class (1200 images).
3. Fine tuning of unsupervised models improves classification accuracy by an average of 2.25 %, and GeoCLR trained with 300 expert-generated annotations per class showed the best performance in this study with a classification accuracy of  $90 \pm 0.05$  % after fine tuning. Classes containing linear features, such as veins and fractures, with spatial context extending beyond the frame of a single patch are the least well classified class type for all models except GeoCLR, which labels all types of class with comparable accuracy.
4. Classifications generated using methods described here allow for the automated generation of downhole datasets traditionally created by experts over the course of days to weeks. Comparing downhole estimates of the amount of altered and relatively fresh rock based on both GeoCLR classifications and visual expert estimations indicate a statistically significant positive relationship (Pearson's Coefficient = 0.7). Therefore, our automated method provides a reliable and efficient means of analysing geological images at higher resolutions than would be feasible using current manual approaches.



## Open Research Section

All images and geological log data used in this study are available from the Oman Drilling Project website ([publications.iodp.org/other/Oman/OmanDP.html](http://publications.iodp.org/other/Oman/OmanDP.html)).

## Acknowledgments

Authors would like to thank Dr Jude Coggon (University of Southampton) for her help compiling all OmanDP data used in the study, as well as Dr. Michelle Harris (University of Plymouth) for her input on data presentation and insight into the expert generated datasets created by herself and others during the OmanDP. Finally we would like to thank Prof. Timothy Henstock (University of Southampton) for his input during the write-up. RMC was funded by a Royal Society University Research Fellowship (URF\R1\180320) and LG by a Royal Society award (RGF\EA\181072) to RMC. DAHT was funded by and NERC-NSF grant (NSFGEO-NEC: NE/W007517/1 “Data mining the deep”). This research used samples and data provided by the Oman Drilling Project. The Oman Drilling Project (OmanDP) has been possible through co-mingled funds from the International Continental Scientific Drilling Project (ICDP; Kelemen, Matter, Teagle Lead PIs), the Sloan Foundation – Deep Carbon Observatory (Grant 2014-3-01, Kelemen PI), the National Science Foundation (NSF-EAR-1516300, Kelemen lead PI), NASA – Astrobiology Institute (NNA15BB02A, Templeton PI), the German Research Foundation (DFG: KO 1723/21-1, Koepke PI), the Japanese Society for the Promotion of Science (JSPS no:16H06347, Michibayashi PI; and KAKENHI 16H02742, Takazawa PI), the European Research Council (Adv: no.669972; Jamveit PI), the Swiss National Science Foundation (SNF:20FI21163073, Früh-Green PI), JAMSTEC, the TAMU-JR Science Operator, and contributions from the Sultanate of Oman Ministry of Regional Municipalities and Water Resources, the Oman Public Authority of Mining, Sultan Qaboos University, CNRS-Univ. Montpellier, Columbia University of New York, and the University of Southampton.

## References

- Aljalbout, E., Golkov, V., Siddiqui, Y., Strobel, M., & Cremers, D. (2018). Clustering with deep learning: Taxonomy and new methods. *arXiv Preprint arXiv:1801.07648*.
- Al-Mudhafar, W. J. (2017). Integrating well log interpretations for lithofacies classification and permeability modeling through advanced machine learning algorithms. *Journal of Petroleum Exploration and Production Technology*, 7(4), 1023–1033. doi: 10.1007/s13202-017-0360-0
- Alt, J. C., Laverne, C., Coggon, R. M., Teagle, D. A., Banerjee, N. R., Morgan, S., ... Galli, L. (2010). Subsurface structure of a submarine hydrothermal system in ocean crust formed at the east pacific rise, odp/iodp site 1256. *Geochemistry, Geophysics, Geosystems*, 11(10). doi: 10.1029/2010GC003144
- Alzubaidi, F., Mostaghimi, P., Swietojanski, P., Clark, S. R., & Armstrong, R. T. (2021). Automated lithology classification from drill core images using convolutional neural networks. *Journal of Petroleum Science and Engineering*, 197, 107933. doi: 10.1016/j.petrol.2020.107933
- Camps-Valls, G., Marsheva, T. V. B., & Zhou, D. (2007). Semi-supervised graph-based hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10), 3044–3054. doi: 10.1109/TGRS.2007.895416
- Chai, H., Li, N., Xiao, C., Liu, X., Li, D., Wang, C., & Wu, D. (2009). Automatic discrimination of sedimentary facies and lithologies in reef-bank reservoirs using borehole image logs. *Applied Geophysics*, 6, 17–29. doi: 10.1007/s11770-009-0011-4
- Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In *International conference*

- 709        *on machine learning* (pp. 1597–1607). doi: 10.48550/arXiv.2002.05709
- 710    Coggon, R., Sylvan, J. B., Teagle, D. A., Reece, J., Chrsteson, G. L., Estes, E. R.,  
711        ... the Expedition 390 Scientists (2022). Expedition 390 preliminary report:  
712        South Atlantic Transect 1. *International Ocean Discovery Program*. doi:  
713        10.14379/iodp.pr.390.2022
- 714    Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A  
715        large-scale hierarchical image database. *Proceedings of the IEEE Conference on*  
716        *Computer Vision and Pattern Recognition*, 248–255. doi: 10.1109/CVPR.2009  
717        .5206848
- 718    Finn, P. G., Udy, N. S., Baltais, S. J., Price, K., & Coles, L. (2010). Assessing  
719        the quality of seagrass data collected by community volunteers in moreton  
720        bay marine park, australia. *Environmental Conservation*, 37(1), 83–89. doi:  
721        10.1017/S0376892910000251
- 722    Fu, D., Su, C., Wang, W., & Yuan, R. (2022). Deep learning based lithology classi-  
723        fication of drill core images. *Plos One*, 17(7), e0270826. doi: 10.1371/journal  
724        .pone.0270826
- 725    Goyal, P., Dollár, P., Girshick, R., Noordhuis, P., Wesolowski, L., Kyrola, A., ... He,  
726        K. (2017). Accurate, large minibatch sgd: Training imagenet in 1 hour. *arXiv*  
727        *Preprint arXiv:1706.02677*. doi: 10.48550/arXiv.1706.02677
- 728    Greenberger, R. N., Harris, M., Ehlmann, B. L., Crotteau, M. A., Kelemen, P. B.,  
729        Manning, C. E., ... Team, O. D. P. S. (2021). Hydrothermal alteration of the  
730        ocean crust and patterns in mineralization with depth as measured by micro-  
731        imaging infrared spectroscopy. *Journal of Geophysical Research: Solid Earth*,  
732        126(8), e2021JB021976.
- 733    He, J., La Croix, A. D., Wang, J., Ding, W., & Underschultz, J. (2019). Using  
734        neural networks and the markov chain approach for facies analysis and pre-  
735        diction from well logs in the precipice sandstone and evergreen formation,  
736        surat basin, australia. *Marine and Petroleum Geology*, 101, 410–427. doi:  
737        10.1016/j.marpetgeo.2018.12.022
- 738    He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image  
739        recognition. *Proceedings of the IEEE Conference on Computer Vision and Pat-*  
740        *tern Recognition*, 770–778.
- 741    Hill, E. J., Pearce, M. A., & Stromberg, J. M. (2021). Improving automated geolog-  
742        ical logging of drill holes by incorporating multiscale spatial methods. *Mathe-*  
743        *matical Geosciences*, 53, 21–53.
- 744    Hill, E. J., Robertson, J., & Uvarova, Y. (2015). Multiscale hierarchical domaining  
745        and compression of drill hole data. *Computers & Geosciences*, 79, 47–57.
- 746    Jarrard, R. D., Abrams, L. J., Pockalny, R., Larson, R. L., & Hirono, T. (2003).  
747        Physical properties of upper oceanic crust: Ocean drilling program hole 801c  
748        and the waning of hydrothermal circulation. *Journal of Geophysical Research:*  
749        *Solid Earth*, 108(B4). doi: 10.1029/2001JB001727
- 750    Kelemen, P. B., Matter, J. M., Teagle, D. A., Coggon, J. A., Team, O. D. P. S., et  
751        al. (2020). Oman drilling project: Scientific drilling in the samail ophiolite, sul-  
752        tanate of oman. *Proceedings of the Oman Drilling Project: College Station, Tx*  
753        *(International Ocean Discovery Program)*. doi: 10.14379/OmanDP.proc.2020
- 754    Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with  
755        deep convolutional neural networks. *Advances in Neural Information Process-*  
756        *ing Systems*, 25. doi: 10.1145/3065386
- 757    Krogh, A. (2008). What are artificial neural networks? *Nature Biotechnology*, 26(2),  
758        195–197. doi: doi.org/10.1038/nbt1386
- 759    Kumar, P., & Chauhan, S. (2022). Study on temperature ( $\tau$ ) variation for simclr-  
760        based activity recognition. *Signal, Image and Video Processing*, 16(6), 1667–  
761        1672. doi: 10.1007/s11760-021-02122-x
- 762    LeCun, Y., Bengio, Y., et al. (1995). Convolutional networks for images, speech, and  
763        time series. *The Handbook of Brain Theory and Neural Networks*, 3361(10).

- 764 Ma, Y. Z. (2011). Lithofacies clustering using principal component analysis and neural  
765 network: applications to wireline logs. *Mathematical Geosciences*, 43(4),  
766 401–419. doi: 10.1007/s11004-011-9335-8
- 767 Min, E., Guo, X., Liu, Q., Zhang, G., Cui, J., & Long, J. (2018). A survey of cluster-  
768 ing with deep learning: From the perspective of network architecture. *IEEE*  
769 *Access*, 6, 39501–39514. doi: 10.1109/ACCESS.2018.2855437
- 770 Olmstead, M. A., Wample, R., Greene, S., & Tarara, J. (2004). Nondestructive mea-  
771 surement of vegetative cover using digital image analysis. *HortScience*, 39(1),  
772 55–59. doi: 10.21273/HORTSCI.39.1.55
- 773 Rosenblatt, F. (1962). *Principles of neurodynamics: Perceptrons and the theory of*  
774 *brain mechanisms* (Vol. 55). Spartan Books Washington, DC.
- 775 Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., . . . others  
776 (2015). Imagenet large scale visual recognition challenge. *International Journal*  
777 *of Computer Vision*, 115, 211–252. doi: doi.org/10.1007/s11263-015-0816-y
- 778 Teagle, D., Reece, J., Coggon, R., Sylvan, J. B., Christeson, G. L., Williams, T. J.,  
779 & Estes, E. R. (2023). International ocean discovery program expedition 393  
780 preliminary report: South Atlantic Transect 2. *International Ocean Discovery*  
781 *Program Expedition Preliminary Report*, 393. doi: 10.14379/iodp.pr.393.2023
- 782 Thomas, A., Rider, M., Curtis, A., & MacArthur, A. (2011). Automated lithology  
783 extraction from core photographs. *First Break*, 29(6). doi: 10.3997/1365-2397  
784 .29.6.51281
- 785 Tominaga, M., Teagle, D. A., Alt, J. C., & Umino, S. (2009). Determination of  
786 the volcanostratigraphy of oceanic crust formed at superfast spreading ridge:  
787 Electrofacies analyses of odp/iodp hole 1256d. *Geochemistry, Geophysics,*  
788 *Geosystems*, 10(1). doi: 10.1029/2008GC002143
- 789 Tominaga, M., & Umino, S. (2010). Lava deposition history in odp hole 1256d: In-  
790 sights from log-based volcanostratigraphy. *Geochemistry, Geophysics, Geosys-*  
791 *tems*, 11(5). doi: 10.1029/2009GC002933
- 792 Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-sne. *Journal of*  
793 *machine learning research*, 9(11).
- 794 Van Etten, A., Lindenbaum, D., & Bacastow, T. M. (2018). Spacenet: A remote  
795 sensing dataset and challenge series. *arXiv Preprint arXiv:1807.01232*. doi: 10  
796 .48550/arXiv.1807.01232
- 797 Wang, F., & Liu, H. (2021). Understanding the behaviour of contrastive loss.  
798 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recogni-*  
799 *tion*, 2495–2504. doi: 10.48550/arXiv.2012.09740
- 800 Xie, J., Girshick, R., & Farhadi, A. (2016). Unsupervised deep embedding for cluster-  
801 ing analysis. *International Conference on Machine Learning*, 478–487. doi:  
802 10.48550/arXiv.1511.06335
- 803 Yamada, T., Massot-Campos, M., Prügel-Bennett, A., Pizarro, O., Williams, S. B.,  
804 & Thornton, B. (2022). Guiding labelling effort for efficient learning with  
805 georeferenced images. *IEEE Transactions on Pattern Analysis and Machine*  
806 *Intelligence*, 45(1), 593–607. doi: 10.1109/TPAMI.2021.3140060
- 807 Yamada, T., Prügel-Bennett, A., & Thornton, B. (2021). Learning features from  
808 georeferenced seafloor imagery with location guided autoencoders. *Journal of*  
809 *Field Robotics*, 38(1), 52–67. doi: 10.1002/rob.21961
- 810 Yamada, T., Prügel-Bennett, A., Williams, S. B., Pizarro, O., & Thornton, B.  
811 (2022). Geocl: Georeference contrastive learning for efficient seafloor im-  
812 age interpretation. *arXiv Preprint arXiv:2108.06421*. doi: doi.org/10.55417/  
813 fr.2022037
- 814 Zhang, P., Sun, J., Jiang, Y., & Gao, J. (2017). Deep learning method for lithol-  
815 ogy identification from borehole images. *79th EAGE Conference and Exhibi-*  
816 *tion 2017, 2017(1)*, 1–5. doi: 10.3997/2214-4609.201700945