Ph.D. Thesis
Doctor of Philosophy

DTU Physics
Department of Physics

Machine Learning in 4D Seismic Data Analysis

Deep Neural Networks in Geophysics

Jesper Sören Dramsch

Kongens Lyngby 2019

DTU Physics
Department of Physics
Technical University of Denmark

Fysikvej
Building 311
2800 Kgs. Lyngby, Lyngby
info@fysik.dtu.dk
Tel.: +45 4525 3344
https://www.fysik.dtu.dk

Summary

Preface

This xxx thesis was prepared at the department of Applied Mathematics and Computer Science at the Technical University of Denmark in fulfillment of the requirements for acquiring a yyy degree in zzz.

Kongens Lyngby, November 4, 2019

Jesper Sören Dramsch

Acknowledgements

Publication List

Journal Articles

Dramsch, Jesper Sören, A. N. Christensen, and M. Lüthje (2019a). "Deep Unsupervised 4D Seismic 3D Time-Shift Estimation with Convolutional Neural Networks". In: *IEEE Transactions in Geoscience and Remote Sensing*. In Preparation.

Dramsch, Jesper Sören, A. N. Christensen, and M. Lüthje (2019b). "Let's do the Time Warp again! – Revisiting Dynamic Time Warping – A practical tutorial in Python on North Sea field data". In: *Geophysics*. In Review.

Dramsch, Jesper Sören, M. Lüthje, and A. N. Christensen (2019f). "Complex-valued neural networks for machine learning on non-stationary physical data". In: *Computers & Geoscience*. In Preparation.

Aabø, T. M., **Dramsch, Jesper Sören**, C. L. Würtzen, S. Seyum, F. Amour, M. Welch, and M. Lüthje (2020). "An integrated workflow for fracture characterization in chalk reservoirs, applied to the Kraka Field". In: *Marine and Petroleum Geology* 112. Accepted, p. 104065. ISSN: 0264-8172. DOI: https://doi.org/10.1016/j.marpetgeo.2019.104065. URL: http://www.sciencedirect.com/science/article/pii/S026481721930501X.

Peer-Reviewed Conference Proceedings

Aabø, T. M., **Dramsch, Jesper Sören**, M. Welch, and M. Lüthje (2017a). "Correlation of Fractures From Core, Borehole Images and Seismic Data in a Chalk Reservoir in the Danish North Sea". In: *79th EAGE Conference and Exhibition 2017*. EAGE. DOI: 10.3997/2214-4609.201701283. URL: https://doi.org/10.3997/2214-4609.201701283.

Dramsch, Jesper Sören and M. Lüthje (2018b). "Deep-learning seismic facies on state-of-the-art CNN architectures". In: *SEG Technical Program Expanded Abstracts 2018*. Society of Exploration Geophysicists, pp. 2036–2040. DOI: 10.1190/segam2018-2996783.1. URL: https://doi.org/10.1190/segam2018-2996783.1.

Mosser, L., W. Kimman, **Dramsch, Jesper Sören**, S. Purves, A. De la Fuente Briceño, and G. Ganssle (2018d). "Rapid seismic domain transfer: Seismic velocity inversion and modeling using deep generative neural networks". In: *80th EAGE Conference and Exhibition 2018*. EAGE. DOI: 10.3997/2214-4609.201800734. URL: https://doi.org/10.3997/2214-4609.201800734.

Peer-Reviewed Workshop Proceedings

Dramsch, Jesper Sören, F. Amour, and M. Lüthje (2018). "Gaussian Mixture Models For Robust Unsupervised Scanning-Electron Microscopy Image Segmentation Of North Sea Chalk". In: *First EAGE/PESGB Workshop Machine Learning*. EAGE. DOI: 10.3997/2214-4609.201803014. URL: https://doi.org/10.3997/2214-4609.201803014.

Dramsch, Jesper Sören and M. Lüthje (2018d). "Information Theory Considerations In Patch-Based Training Of Deep Neural Networks On Seismic Time-Series". In: *First EAGE/PESGB Workshop Machine Learning*. EAGE. DOI: 10.3997/2214-4609.201803020. URL: https://doi.org/10.3997/2214-4609.201803020.

Dramsch, Jesper Sören, G. Corte, H. Amini, M. Lüthje, and C. MacBeth (2019c). "Deep Learning Application for 4D Pressure Saturation Inversion Compared to Bayesian Inversion on North Sea Data". In: *Second EAGE Workshop Practical Reservoir Monitoring 2019*. EAGE. DOI: 10.3997/2214-4609.201900028.

Dramsch, Jesper Sören, G. Corte, H. Amini, M. Lüthje, and C. MacBeth (2019e). "Including Physics in Deep Learning – An Example from 4D Seismic Pressure Saturation Inversion". In: *81st EAGE Conference and Exhibition 2019 Workshop Programme*. EAGE. DOI: 10.3997/2214-4609.201901967. URL: https://doi.org/10.3997/2214-4609.201901967.

Not Peer Reviewed

Aabø, T. M., M. Welch, **Dramsch, Jesper Sören**, M. Lüthje, S. Seyum, F. Amour, and C. L. Würtzen (2017b). *Fracture Characterization and Modelling in the Kraka Field*. Tech. rep. Danish Hydrocarbon Research and Technology Centre.

Dramsch, Jesper Sören and M. Lüthje (2018a). *Deep Learning: From Cats to 4D Seismic - Reducing cycle time and model training cost in asset management*. Tech. rep. Danish Hydrocarbon Research and Technology Centre. DOI: 10.6084/m9.figshare.7422629. URL: https://doi.org/10.6084/m9.figshare.7422629.

Dramsch, Jesper Sören and M. Lüthje (Dec. 15, 2018c). *Deep-learning seismic facies on state-of-the-art CNN architectures. figshare*. Open-Source Software. DOI: 10.6084/m9.figshare.7227545. URL: https://doi.org/10.6084/m9.figshare.7227545.

Contents

CHAPTER 1

Introduction

CHAPTER 2

Methods & Theory

2.1 4D seismic

4D seismic is the analysis of seismic data that was acquired over the same area after some calendar time has passed. This analysis usually includes matching two seismic cubes and analyzing the shifts as well as the amplitude difference after alignment. These can theoretically be inverted for physical properties that cause the change of the seismic image.

4D seismic data analysis suffers from the superposition of multiple effects on the seismic imaging. These effects include changes in the acquisition equipment due to technological advances, changes in acquisition geometry (source-receiver mismatch), as well as physical changes in the subsurface. These physical changes are in part due to fluid movement in the subsurface, as well as, changes in the geology due to compaction and expansion. These geomechanical effects change the position of the reflectors, the thickness of stratigraphy and the physical properties such as density and wave velocity.

Amplitude differences are the standard analysis tool in 4D sesimic interpretation. Once the seismic cubes have been aligned, the amplitude difference can be interpreted by experts. Interpreters will look for both the differences to see fluid movements. Additionally, by-passed zones can be identified by ”low difference zones” in generally mobile reflector packets.

Time shifts were a main tool to align seismic data for amplitude difference interpretation. The time shifts and time strains themselves can be interpreted on their own. Time shifts extraction is mostly done in z-direction by comparing traces. The most common methods implementing time shift extraction operate solely in 1D on traces. Prominently, the 1D windowed cross-correlation is used due to its computational speed and general robustness. The main drawback of this method is, however, that the result is highly dependent on the window-size.

More recently research into pre-stack time shift extraction and 3D-based methods is conducted. These methods take into account changes in the geology and seismic illumination.

The subsurface changes recorded by the seismic data can be related to subsurface changes. These changes include porosity, fluid saturation and pressure changes. This inversion process is non-unique and is often reliant on prior information. Many inversions rely on Bayesian processes to de-risk the inversion.

2.2 Machine Learning

Machine Learning (ML) is the discipline of defining a statistical or mathematical models based on data. These ML models are either trained in a supervised or unsupervised fashion, which usually results in them learning a decision boundary, or a representation or structure of the data respectively. Historically, ML has been an interest in geoscience but has not gained momentum due to sparse data, computational capability, and availability of algorithms. Geoscience data was often not available and still is often not available with a reliable ground truth. However, particularly Neural Networks (NNs) have found broad interest in geophysical applications, Bayesian methods are often used in inversion schemes and recent software developments have changed the research entirely.

Recently, the subfield Deep Learning (DL) has reignited interest in the wider field of ML by outperforming rule-based algorithms on computer vision tasks, such as image classification and segmentation (Bishop, 2016). These developments have propelled developments in other non-related fields such as biology (Ching et al., 2018), chemistry (Schütt et al., 2017), medicine (Shen et al., 2017) and pharmacology (Kadurin et al., 2017). DL utilizes many-layered artificial NN to approximate an objective function. In recent years the open source movement, democratization of access to computing power and developments in the field of DL have rekindled interest in applications of ML to geoscience. The availability of free open source libraries such as skikit-learn (Pedregosa et al., 2011) has made ML methods and several tools for the application of rigorous statistical evaluation of experiments without explicit expert knowledge widely available. Furthermore, Tensorflow (Martín Abadi et al., 2015), PyTorch (Paszke et al., 2017), and Keras (Chollet et al., 2015) have made NNs easily accessible and provide experimentation capabilities to transfer recent developments in ML research to other scientific fields.

Algorithms and methods in ML can be organized in different ways. Two ways to categorize algorithms are based on the training or based on the learned distribution. In training, these algorithms can be categorized into supervised and unsupervised methods, where supervised methods learn the functional mapping from x, being the data, to y, being the ground truth or label for the data. When the ground truth is not known, unsupervised methods can be applied to determine structures and relationships within the data. Semi-supervised, and weakly supervised try to propagate partial labels to similarly distributed data and then learn the supervised mapping f(x) = y. Alternatively, ML algorithms can be categorized into generative methods that learn the joint probability distribution or discriminative methods that learn a decision boundary to optimally separate data. Additionally, methods can be distinguished by application. Assigning labels to data is called classification. The general, continuous application to map data from the input to the output domain is called regression. Finding relationships and agglomerations of data is called clustering. Most algorithms can be applied to several of these categories, such as support vector machines that can function as classifier and regressor.

Applications in ML are quickly evolving and many are improved by mathematical insights, engineering features and increased availability of data. This thesis focuses on the application of NNs, which come in different implementation details and particularly NN architectures are often re-implemented with slight differences that deviate from the original published architecture. Particularly in NN we have to focus on the most practical building blocks, to be able to give a comprehensive overview.

2.2.1 History of Machine Learning

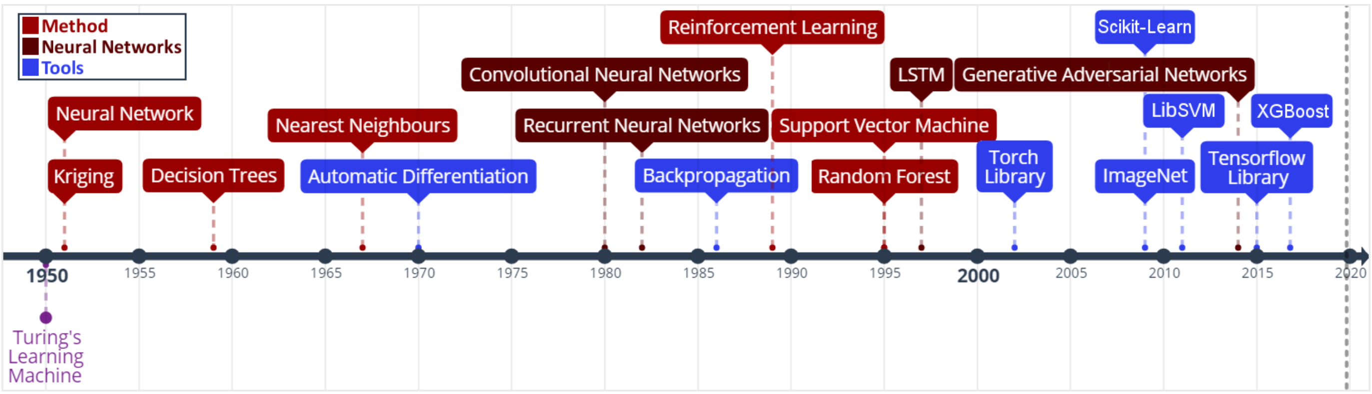


Figure 2.1: Selection of notable milestones in machine learning.

Creativity, learning, and intelligence with regard to computers have been discussed as early as of the first programmer Ada Lovelace (Taylor, 1843).

> ”The Anlytical Engine has no pretensions whatever to *originate* any thing. It can do whatever we *know how to order it* to perform. It can *follow* analysis; but it has no power of *anticipating* any analytical relations or truths. Its province is to assist us in making *available* what we are already acquainted with. This it is calculated too effect primarily and chiefly of course, through its executive faculties; but it is likely to exert an *indirect* and reciprocal influence on science itself in another manner.” – Note G, Page 689, Ada A. Lovelace. (Taylor, 1843); Emphasis taken from source text.

This notion was challenged by Alan Turing (Turing, 1950) who proposed the ”Learning Machine”, which specifically predict genetic algorithms, a metaheuristic that finds application in optimization and search problems. Evolutionary computing and genetic algorithms specifically can perform some machine learning tasks (Goldberg and Holland, 1988). This is generally considered the commencement of Artificial Intelligence (AI) and ML, however, they rely heavily on earlier developments in statistics such as the Bayesian theorem (Bayes, 1763) and Markov processes (Markov, 1971). The first method, we include on the timeline in Figure 2.1 is ”kriging” Krige (1951), which is based on Gaussian

Processes, these form an important category of non-parametric machine learning these days. Gaussian processes are often also attributed to work of Kolmogorov (1939) on time series. Another method was developed to mimic the human brain, namely Neural Networks (NNs). The construction of the first NN machine by Minsky (Russell and Norvig, 2010) was soon followed by the "Perceptron", a binary decision boundary learner (Rosenblatt, 1958). The decision is made according to

$$\begin{aligned} o_j &= \sigma\left(\textstyle\sum_i w_{ij} x_i + b\right) \\ &= \begin{cases} 1 & \sum_i w_{ij} x_i + b > 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{2.1}$$

which describes a linear system of the input data x, the weights w and bias b and a binary activation funtion σ. The linear system is still used in modern neurons, however, the activation σ is usually a Rectifier function. Shortly after, Belson (1959) describe the first Decision Tree (DT), which learns hierarchical decision systems. The next method, Nearest Neighbour (KNN) search, was introduced by Cover and Hart (1967) to solve the traveling salesman problem. Two decades later Q-learning (Watkins, 1989) introduces a method to reinforcement learning that is still used to this day. The final two methods in the timeline were introduced in 1995. Random Forests (RFs) (Ho, 1995) introduce ensemble learning of weak learning Decision Trees (DTs). Support Vector Machines (SVM) (Cortes and Vapnik, 1995) introduce a strong learner that aims to maximize the margin between classes.

These methods have been improved upon over the decades. Specific milestones that accelerated further developments in NN are automatic differentiation (Linnainmaa, 1970) and consequently applying this to backpropagate errors in Deep Neural Networks (DNNs) (Rumelhart et al., 1988). Backpropagation itself as a concept existed earlier (Kelley, 1960; Bryson, 1961), followed by a simplification by using the chain rule (Dreyfus, 1962). These enable effective implementation of NNs today. Moreover, open sourcing the Torch library (Collobert et al., 2002) made and assembling the ImageNet database (Deng et al., 2009) has accelerated developments in computer vision and enabled modern developments in deep learning. In the same year the library Scikit-Learn (Pedregosa et al., 2011) was established, which introduced a common open source API (Buitinck et al., 2013) for a diverse and growing set of shallow machine learning models (e.g. SVMs, RFs, KNNs, shallow NNs). Scikit-learn has had a profound impact on machine learning applications across the sciences and the API is modelled in other open source libraries. Chang and Lin (2011) introduced a widely used implementation for Support Vector Machines (SVM), which is also used in Scikit-Learn. Recently, the Tensorflow library (Martín Abadi et al., 2015) was introduced for open source deep learning models, with some different design choices than Pytorch. In this open environment fueled by competitions (e.g. ImageNet (Russakovsky et al., 2013), Netflix Prize (Bennett, Lanning, et al., 2007), Kaggle (Goodfellow et al., 2013)) XGBoost (Chen and Guestrin, 2016), a library for extreme gradient tree boosting was developed.

Recent developments in deep learning are based in Neural Networks (NNs), hence, we highlight some key developments in Figure 2.1. Convolutional Neural Networks

(CNNs) are essential in the modern computational vision systems, they were inspired by (LeCun et al., 2015) the concept of Neocognitron (Fukushima, 1980). In the same decade Recurrent Neural Networks (RNNs) were introduced implemented as Hopfield Networks (Hopfield, 1982). While Hopfield networks are not a general RNN, they provide content-adressable memory with the internal state memory. Hochreiter and Schmidhuber (1997) implement the Long Short-Term Memory (LSTM), which contain internal states (i.e. memory) that can process temporal sequences, still used and performing to the state-of-the-art in sequence analysis and Natural Language Processing (NLP) to this day. Recently, Generative Adversarial Network (GAN) (Goodfellow et al., 2014b) introduced a system of NNs that can create new samples from a distribution. The GAN consists of two NNs, a generator and a discriminator, which generate samples from a noise distribution and judge the validity of the sample respectively. We discuss NNs in more detail in section 2.2.2

2.2.2 Neural Networks (NNs)

Neural Network (NN) as a class of ML algorithms are very diverse and versatile. NNs have persisted for decades and their nomenclature has changed in this time. NNs were long called Artificial Neural Network (ANN), which has changed to simply NN, usually prepended with the class of Neural Network, namely Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Deep Neural Network (DNN), which I will discuss in more detail.

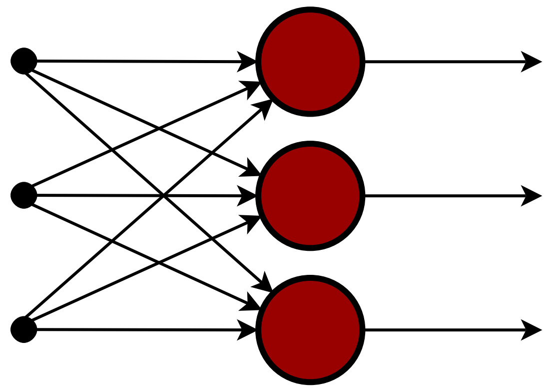
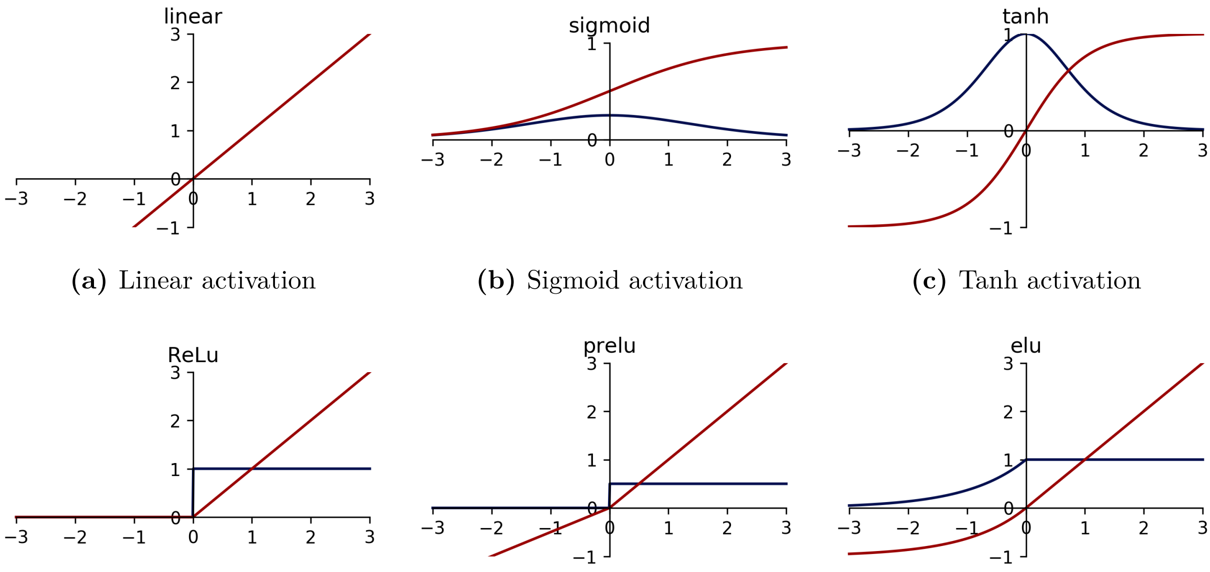


Figure 2.2: A simple NN.

Neural Networks (NNs) can be approached from several theoretical bases. Mathematically, NNs are directed acyclical graphs with edges and nodes. In neural computation, these are generally referred to as weights and nodes or neurons. In Figure 2.2, we present a simple densely connected Multi-Layer Perceptron (MLP) with three inputs and three outputs. This configuration is equivalent to a linear regression model. The inputs are distributed across the nodes, and each weight is multiplied with a weight inherent to that graph edge. During the trianing of this machine learning model, these weights get adjusted to obtain a generalizable result. Each node sums the contributions of these weights and possibly a bias, which is trainable but does not take input data. This amounts to each node performing

$$a_j = \sigma\left(\sum_i w_{ij} x_i + b\right), \tag{2.2}$$

with a signifying the activation at a node, i, j being the index of the source and target node respectively, w being the trainable weight, and b being the trainable bias, and σ representing an activation function. Activation functions are an active topic of research, but the generally perform a non-linear transformation of the activation at the node.



(a) Linear activation **(b)** Sigmoid activation **(c)** Tanh activation

(d) ReLu activation **(e)** PReLu activation ($\alpha = .5$) **(f)** ELU activation ($\alpha = 1$)

Figure 2.3: Common Activation functions (red) and derivatives (blue).

In Figure 2.3 I present common activation functions used in NNs. The activation functions introduce non-linearities into the network to transform the linearly scaled input to arbitrary non-linear outputs. The mathematical function in Figure 2.3(b) and Figure 2.3(c) are used less, because of the vanishing gradient problem (Hochreiter, 1991). These occur in the extrema of both functions, where the function saturates and the gradient is close to zero for large values of x. Rectifiers presented in Figure 2.3(d), 2.3(e), and 2.3(f), circumvent this problem by one-sided saturation.

Training the Model Before training, each weight and bias is assigned an initial number that is drawn from a distribution appropriate to the network architecture and data (LeCun et al., 2012; Glorot and Bengio, 2010; He et al., 2015). These strategies collectively initialize weights in a pseudo-random way within limits. The data is then passed through the network, which calculates a result. This result is then compared to the ground truth, using a loss function (e.g. Mean Absolute Error (MAE), Mean Squared Error (MSE)). The resulting error Δt is then used to correct the weights and biases in the network, calculating the correction per weight Δw_{ij} recursively (for many-layered networks).

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} = -\eta \delta_j a_i, \tag{2.3}$$

with η being the learning rate and δ being

$$\delta_j = \begin{cases} \sigma'(\mathrm{net}_j)\Delta t & \text{if } j \text{ is output node,} \\ \sigma'(\mathrm{net}_j) \sum_{j-1} \delta_{j-1} w_{j(j-1)} & \text{if } j \text{ is hidden node.} \end{cases} \tag{2.4}$$

Therefore, hidden nodes are reliant on the result δ_{j-1} of the node at index $j-1$ (Goodfellow et al., 2016). The training of the model can be done on a per-sample basis, which is Stochastic Gradient Descent (SGD) or in the case of noisy inputs, the mean error of several samples can be calculated to perform mini-batch gradient descent. Iteration over forward and backward passes adjusts the weights to predict the correct result.

The optimization of the backpropagation is performed using SGD or other gradient-based optimizers such as the Adam optimizer (Kingma and Ba, 2014). However, during training of the NN, it is important to ensure that the network learns a general relationship instead of memorizing the input data. This memorization is called overtraining, or overfitting. Overfitting can be avoided by regularizations like weight decay (Krogh and Hertz, 1992) and Nesterov momentum or (Sutskever et al., 2013), which modify the optimization loop. Alternatively, methods like Dropout (Hinton et al., 2012) and Batch-Normalization (Ioffe and Szegedy, 2015) modify the network at training time. Moreover, a diverse training set and train-val-test split help avoid overfitting and ensure generalization of the trained model.

The train-val-test split separates the data into three parts. The training and validation set are available during training and hyperparameter tuning, the test set, however, should only be used once to ensure generalization of the model. The train test is used during the optimization loop, the actual training of the model, with the validation set ensuring generalization of the model to unseen data within the loop. In and of itself, the train and validation data would be sufficient, if no other changes to the model were made based on the results of the validation data. Since hyperparameter tuning and model selection are a common procedure today, these present an implicit source of information leakage from the validation set into the data. The hyperparameter tuning will often pose an optimization loop in itself that optimizes based on the results on the "unseen" validation set, essentially implicitly fitting the model to the validation data, therefore, a separate test set is necessary to ensure true generalization.

2.2.2.1 Feed Forward Networks

Feed forward Neural Networks (NNs) or MLPs are the simplest for of NN. In its simplest form it uses a set of linear equations to approximate a function. The network can be described as a graph with edges and nodes. In the neural information community the nodes are often named neurons. These neurons are arranged into layers in Figure 2.4. The first layer in a NN is the input layer with a number of nodes corresponding to the number of input data points. The input nodes are connected to the next layer by the graph's edge. The next node can be the output layer. The weights between subsequent are floating point numbers that scale each input point and determine the value at the output nodes.

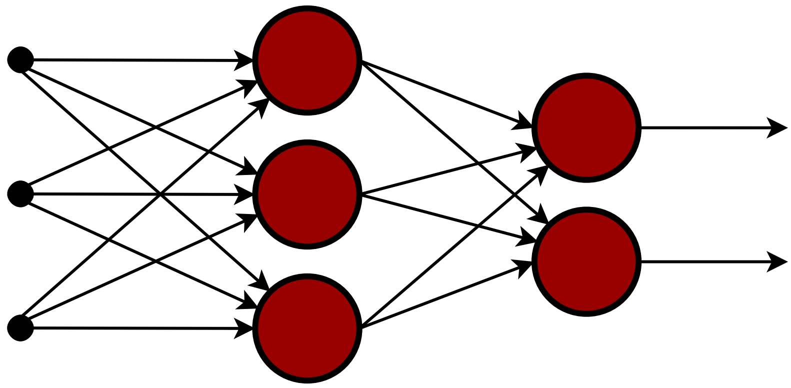


Figure 2.4: Feed forward NN.

NNs gain their powerful learning capabilities from adding layers (see Figure 2.4) in between the input and output node and applying a non-linear activation function. Non-linear activations scale the input from the edge at each neuron. Historically, these have been straight-forward mathematical functions such as $tanh()$ and $sig()$ (cf. Figure 2.3). These suffer from some short-comings that were overcome to leverage multi-layered DNNs.

2.2.2.2 Self-Organizing Maps

Self-Organizing Maps (SOM), also named Kohonen-networks (Kohonen, 1982) are a special case of networks that do not modify the flow of data from the input to the output nodes. They treat each data point as a node and adjust the weights between each node in on a similarity metric. These tend to perform well on spatially correlated data and find good adoption in geoscience.

2.2.2.3 Recurrent Networks

A special configuration of NN is the Recurrent Neural Network (RNN). These networks use edges that feed back into the network. RNNs are used in two applications in ML. They can preserve hidden states, which gives them temporal context sensitivity. Application two is time series analysis similar to feed-forward NNs, where the input is a time step that can be analyzed within the context of surrounding time steps. These RNN represent cyclic directed graphs of computation, as opposed to the other types of NN we discuss, which are acyclic directed graphs. In Figure 2.5 we show the changes of a simple RNN graph compared to a feed forward NN in Figure 2.4. The RNN loops back into itself, which is often regarded as the internal state or feedback. This internal state enables content addressable memory and good performance on

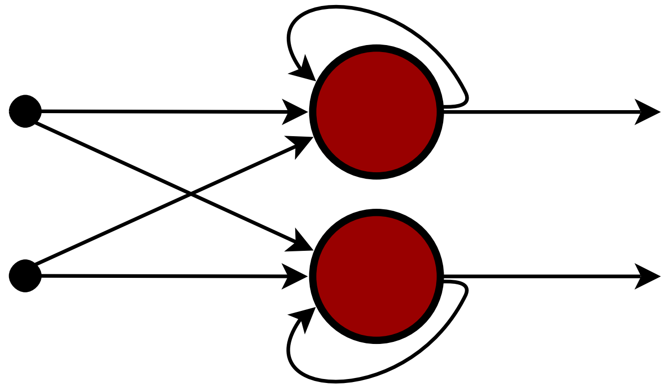


Figure 2.5: Recurrent NN.

Hopfield Networks are one type of recurrent networks that model the human memory. Hopfield networks and their subclasses can be used for pattern recognition. They are guaranteed to find a pattern, however, they are known to converge to local minima. Boltzman machines are configured like Hopfield networks, in contrast to deterministic Hopfield networks, their response to an input is stochastic. Boltzman machines draw from a joint distribution, making them a generative model.

Long Short-Term Memory (LSTM) is a type of RNN that models memory. Details differ in implementations of Long Short-Term Memory (LSTM), however the main criteria are three gates and an inner cell.

- Input Gate
- Forget Gate
- Output Gate

The input gate regulates the contribution of input values to the internal cell. The forget gate regulates the persistence of values in the cell. Finally, the output gate regulates the contribution of the input value to the output value convolved with the cell state.

2.2.2.4 Deep Networks

Improvements in computational power made it possible to train many-layered NNs (see Figure 2.6). These Deep Neural Networks (DNNs) are at the core of recent developments in Deep Learning (DL), leading to the re-implementation of many algorithms into openly available libraries, which has led to further innovative uses of these building blocks. These networks leverage the combinatorial power of NN layers. In deep NNs gradient propagation led to exploding or vanishing gradients before. New non-saturating activation functions lead to stabilization of training DNN (cf. Figure 2.3).

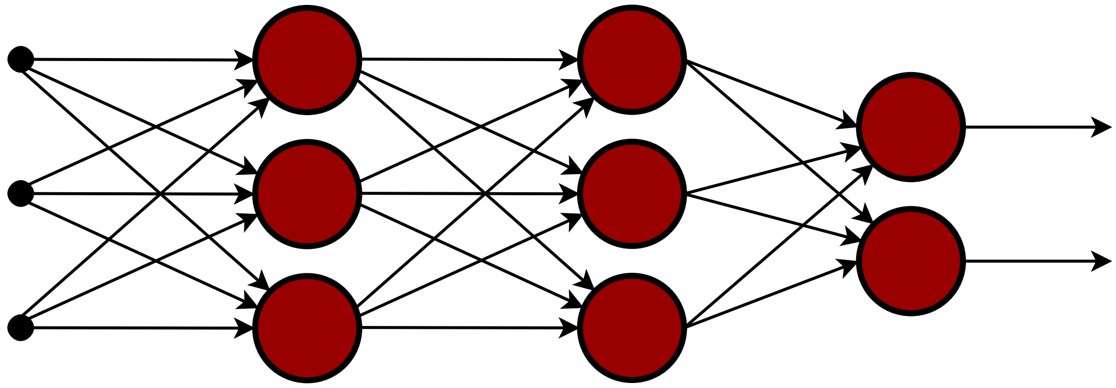


Figure 2.6: Deep Feed forward NN.

2.2.2.5 Convolutional Networks

Convolutional Neural Network (CNN) were developed in computer vision to automatically learn a filter that spatially correlates data. The convolutional kernels are computationally efficient due to weight sharing, making them feasible for very deep networks (cf. section 2.2.2.4). CNNs have had the biggest influence on the renaissance of modern ML. These building blocks for NNs are very good for image data and data where spatially correlated information provides valuable context. It has therefore quickly gained attention in seismic interpretation and seismic data analysis. CNNs like other NNs are optimized by stochastic gradient descent, optimizing a defined loss over the chosen task.

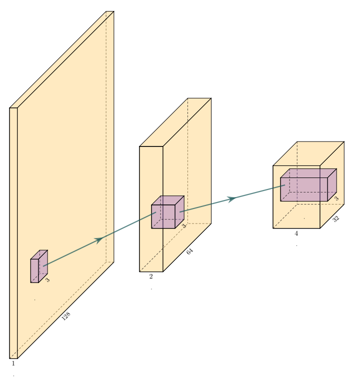


Figure 2.7: Schematic of a CNN filter (purple) in the image data (orange).

For a two-dimensional CNN, the convolution of the $m \times n$-dimensional image G with a filter matrix f can be expressed as:

$$G^*(x,y) = \sum_{i=1}^{n}\sum_{j=1}^{m} f(i,j) \cdot G(x-i+c,\ y-j+c), \tag{2.5}$$

resulting in the central result G^* around the coordinate c. Realistically, the calculation is done in the Fourier domain, due to the Convolution theorem reducing the computational complexity from $\mathcal{O}(n^2)$ to $\mathcal{O}(n \log n)$ with

$$\mathcal{F}\{f * g\} = k \cdot \mathcal{F}\{f\} \cdot \mathcal{F}\{g\}, \tag{2.6}$$

with $\mathcal{F}\{f\}$ denoting the Fourier transform of f and k being a normalization constant. This reduces the convolution to a simple multiplication in the Fourier domain.

Figure 2.7 shows the schematic of connected convolutional layers in a CNN. The network learns a specified number of 3×3 filters from the initial image. Strided convolutions with a step-size larger than 1 or Pooling layers are used to reduce the spatial extent of the image. The repeated downsampling of the image and extraction of convolutional filters has been shown to work for computer vision tasks. Historically, the CNN architecture AlexNet (Krizhevsky et al., 2012) was the first CNN to enter the ImageNet challenge and improved the classification error rate from 25.8 % to 16.4 % (top-5 accuracy). This has propelled research in CNNs, resulting in error rates on ImageNet of 2.25 % on top-5 accuracy in 2017 (Russakovsky et al., 2015).

2.2.2.6 Generative Adversarial Networks

Goodfellow et al. (2014a) introduced Generative Adversarial Network (GAN) as a combination of two CNNs. Different modifications exist that draw from the original GAN, these modifications add more regularization and other feedback loops, as GANs are notoriously difficult to train without careful fine-tuning. These modifications include Wasserstein losses (Arjovsky et al., 2017), and gradient penalization (Gulrajani et al., 2017) for regularization, or cycle-consistent loss for unsupervised training (Zhu et al., 2017).

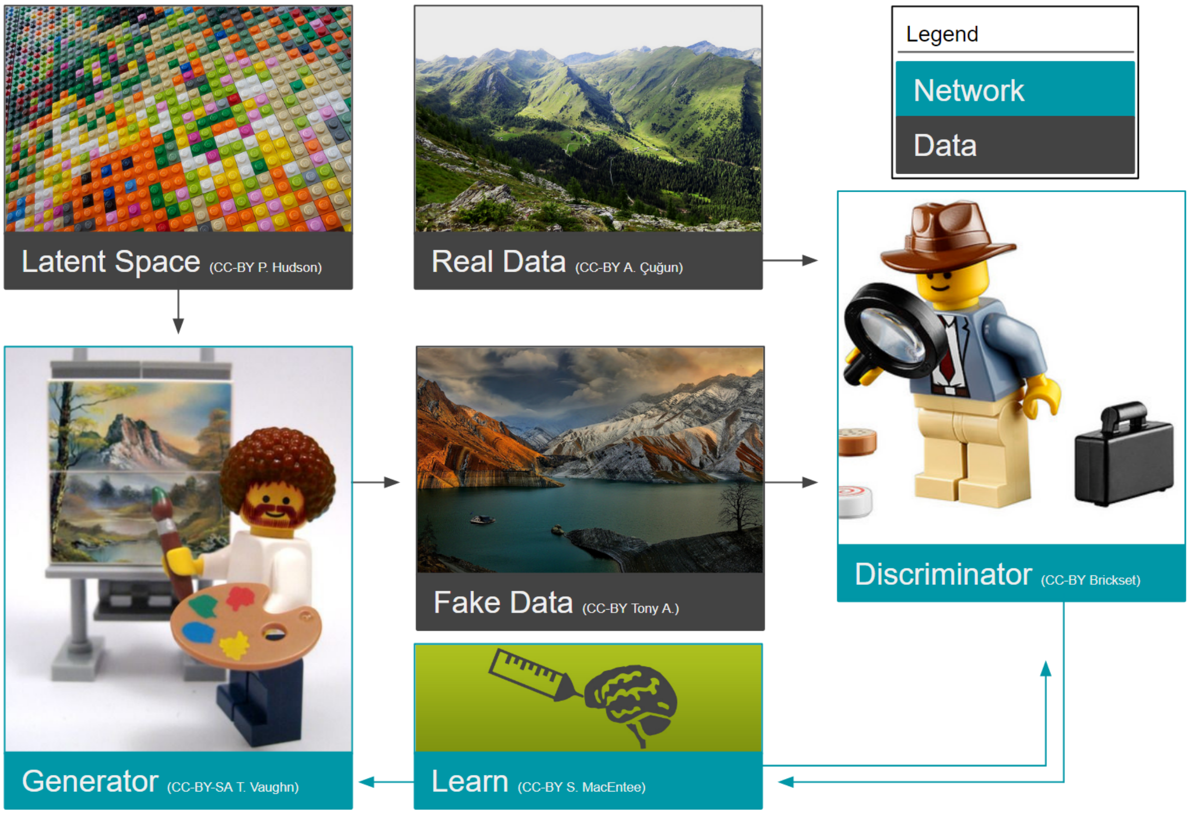


Figure 2.8: Schematic of a Generative Adversarial Network.

Figure 2.8 shows the basic working of GANs. The arrows are colored in blue and grey, where the blue paths show network feedback and grey shows the progression of data. These networks learn from each other, where the generator draws from latent space (a noise vector) to create a fake version of a target. The discriminator tries to discern whether the presented data is real or generated from the adversarial generator. These networks leverage game theory to outperform each other and comparative networks. They reach a Nash equilibrium during training, which describes the concept on a non-cooperative game reaching steady state (Nash, 1951).

2.2.3 Neural Architectures

Neural Networks can generally be assembled in different architectures. In Figure 2.10 we present reported performances of neural architectures on the classification task of the ImageNet challenge. The colors in this figure express different classes of architectures. Early networks that broke ground as the new state-of-the-arts in image classification are the AlexNet, VGG-16, and VGG-19. These networks clearly do not leverage some tricks that modern CNNs implement, the VGG-16 with a relatively high amount of parameters is known to generalize well on transfer learning tasks however (**Dramsch, Jesper Sören** and Lüthje, 2018).

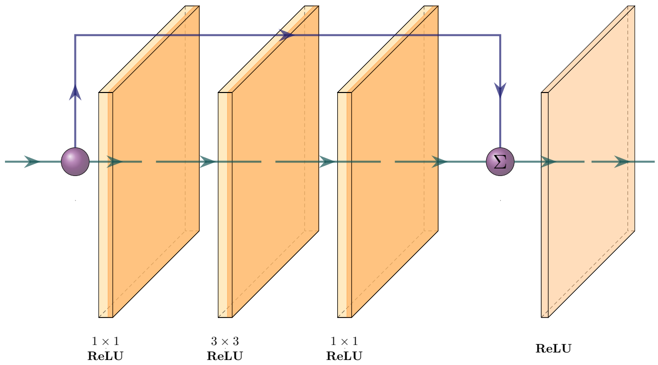


Figure 2.9: Resnet Block with two 1×1 convolutional layers that frame a 3×3 convolutional layer with ReLU activation each. The result being added with the original data, also known as identity..

Research into deep convolutional networks showed that the data in the network would lose signal with increasing depth. Hence, the limitation of VGG at 19 layers. Residual blocks introduced a solution to this problem by implementing a shortcut between the original data and the output from the block. Figure 2.9 presents the original ResNet block architecture, which was used in ResNet-50 and ResNet-101 in Figure 2.10 (He et al., 2016). Details on ResNet blocks differ, the main take-away being the sum or concatenation of the original data with the block output. DenseNets (Huang et al., 2017a) and Inception-style networks (Szegedy et al., 2015) are other approaches to build deeper NNs.

The categories of AmoebaNet, NASNet, and EfficientNet are a more recent development in neural architecture research, based on Neural Architecture Search (NAS). The AmoebaNet is based on Evolutionary Computing and hand-tuning the solution to search for an ideal neural architecture to solve the task (Real et al., 2019). The NASNet fixes the overall architecture, but uses a controler RNN to modify the blocks within the architecture (Zoph et al., 2018). The EfficientNet architecture was also acquired by NAS, by optimizing for both accuracy and FLOPS to reduce the computational cost (Tan and Le, 2019). Moreover, Tan and Le (2019) derives a method of compound scaling for deep neural networks. While ResNet-50 and ResNet-101 differ only in depth, the authors derive a relationship between depth, width and resolution-scaling of deep neural networks.

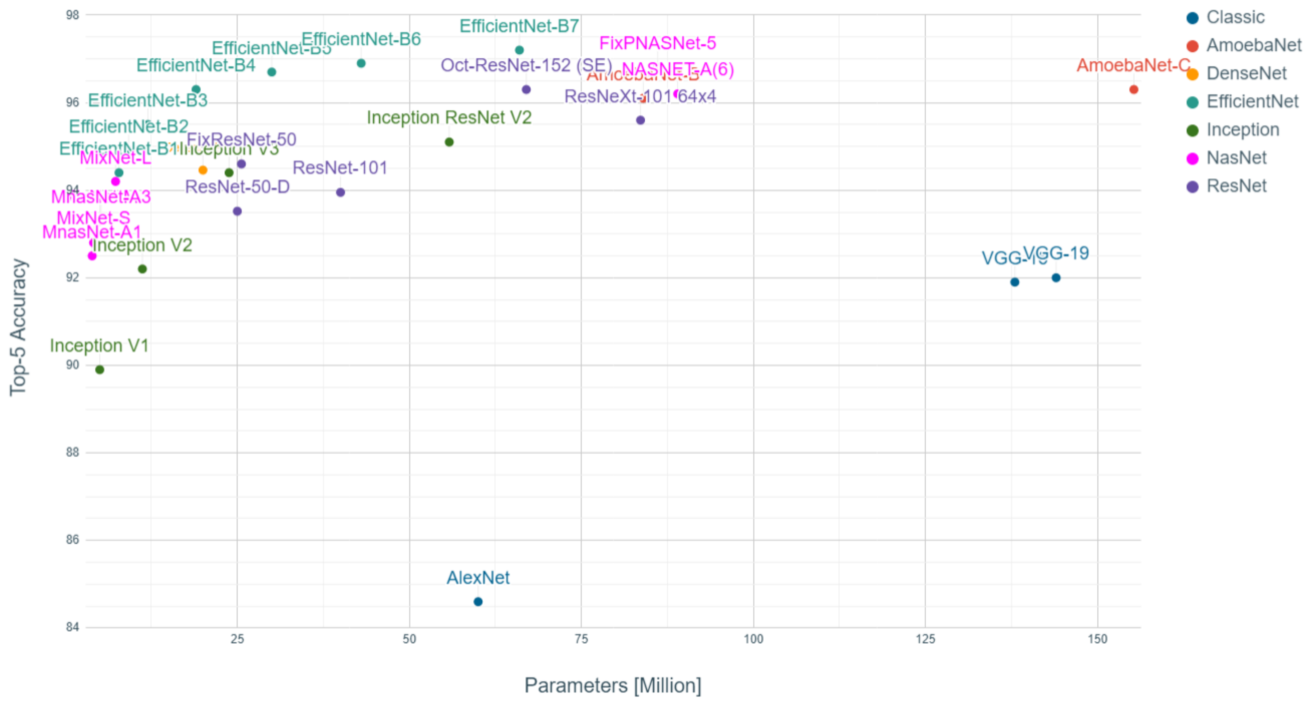


Figure 2.10: Top-5 Accuracies of Neural Architectures on ImageNet plotted against Million Parameters, color-coded to similar network type.

Apart from building deeper networks for image classification, the neural architectures can serve as a forcing function to the task the network is built for. Encoder-Decoder networks will compress the data with a combination of downsampling layers, which in the case of a computer vision could either be strided convolutions or pooling layers after convolutional layers. During these operations, the number of filters increases, while the spatial extent is diminished significantly. This encoding operation is equivalent to a lossy compression, with the low-dimensional layer called "code" or "bottleneck". The bottleneck is then upsampled by either strided transpose Convolutions or upsampling layers that perform a specified interpolation. This is the Decoder of the Encoder-Decoder pair. These networks can be used for data compresssion in AutoEncoders (AEs), where the decoder restores the original data as good as possible (Hinton and Salakhutdinov, 2006). Alternatively, the Decoder can learn a dense classification task like semantic segmentation or seismic interpretation.

Unets present a special type of encoder-decoder networks, that learn semantic segmentation on from small datasets (Ronneberger et al., 2015). Originally developed on biomedical images, the network found wide acceptance in data sparse disciplines. The Unet implements shortcut connections between convolutional layers of equal extent in the Encoder and Decoder networks. This alleviates the pressure of the network learning and reconstructing everything from the bottleneck in isolation.

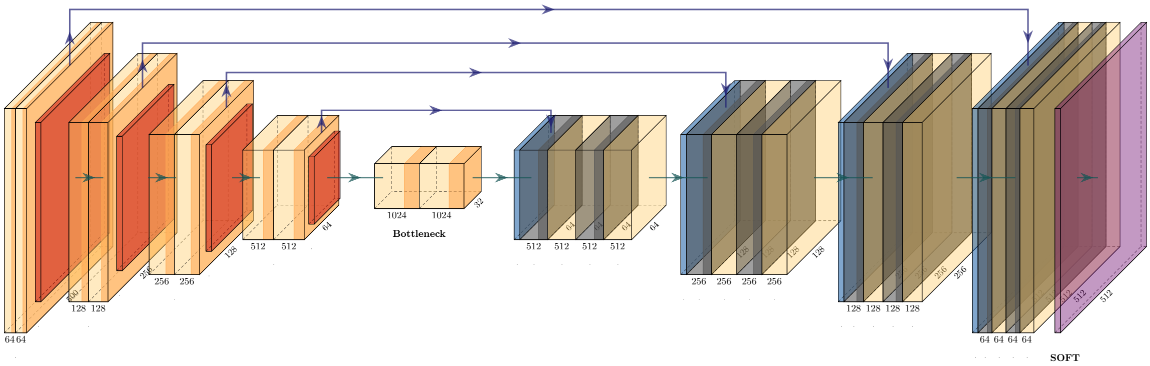


Figure 2.11: Unet after Ronneberger et al. (2015) using 2D convolutional layers (yellow) with ReLU activation (orange) and skip connection between equal-dimensional layers. The Encoder uses pooling (red), while the Decoder uses Upsampling layers (blue), witha final SoftMax layer (purple) for classification / semantic segmentation.

2.3 Machine Learning in Geoscience

The development of the subfield of deep learning has lead to advances in many scientific fields that are not directly related to the larger field of artificial intelligence. This section focuses on historic use-cases of machine learning models in geoscience and evaluate these in the context of recent advances in deep learning. I provide an overview of supervised and unsupervised methods that have persevered. Furthermore, I distinguish implementations of deep neural network topologies and advanced machine learning methods in geoscientific applications. I go on to investigate where these methods differ from previously unsuccessful attempts at application.

Early on Machine Learning (ML) has been reviewed in a geophysical context. Early publications of ML in geoscience apply NNs to geophysical problems. Particularly seismic processing lends itself to explore NNs as general functional approximator (Hornik et al., 1989). McCormack (1991) review of the emerging tool of neural networks in 1991. He highlights the application of pattern recognition and is very succinct in describing basic math associated with neural computing. The wording of most parts has changed, as compared to today. Generally this gives a good baseline and McCormack gives a good illustration and overview with examples in well log classification and trace editing. As the paper appeared in The Leading Edge, it is not peer reviewed, yet it does give a good historic overview. The author summarizes NN applications over the 30 year prior to the review and hightlights automated well-log analysis and seismic trace editing. The review comes to a conclusion that these methods show promise as general approximators.

Baan and Jutten (2000) review the most recent advancements in Neural Networks (NNs) in geophysical applications. It goes into much detail on the neural networks employed in 2000 and the difficulties in building these models and training them. They identify the following subsurface geoscience applications through history: First-break picking, electromagnetics, magnetotellurics, seismic inversion, shear-wave splitting, well log analysis, trace editing, seismic deconvolution, and event classification. The authors evaluate the application of NNs as subpar to physics-based approaches. The paper concludes that neural networks are too expensive and complex to be of real value in geoscience. Generally, this review focuses very much on exploration geoscienc.

Mjolsness and DeCoste (2001) review ML in a broader context outside of exploration geoscience. They illustrate recent successes of ML in analyzing sattelite data and computer robotic geology. The authors include graphical models, Random Markov Models (RMMs), Hidden Markov Models (HMMs), and SVMs. They further highlight limitations to vector data, therefore failing richer data such as graphs and text data. Moreover, the authors from NASA JPL go into detail on pattern recognition in automated rovers to identify geological prospects on Mars. They state:

> "The scientific need for geological feature catalogs has led to multiyear human surveys of Mars orbital imagery yielding tens of thousands of cataloged, characterized features including impact craters, faults, and ridges." - (Mjolsness and DeCoste, 2001)

The authors evaluate how especially the introduction of SVM have allowed the identification of geomorphological features without modeling the processes behind. Further they mention recurrent neural networks in gene expression data, a method that has experienced a renaissance in deep learning. The paper is very short and succinct in evaluating prospects without going into detail on the algorithms itself. In contrast, we expect our review to go more into depth and explore the applications in geoscience further.

2.3.1 History of Machine Learning in Geoscience

Machine learning, statistical, and mathematical models have a long history in geoscience. Markov models have been used to describe sedimentology as early as the 1970s (Schwarzacher, 1972) and the use of k-means in geoscience as early as 1964 (Preston and Henderson, 1964). In geophysics applications of NNs to perform seismic devonvolution were published in the 1980s Zhao and Mendel (1988). Early tree-based methods were chiefly used in economic geology and exploration geophysics for prospectivity mapping with Decision Trees (DTs) (Newendorp, 1976; Reddy and Bonham-Carter, 1991). SVM has early on been applied to AVO classification Li and Castagna, 2004 and geological facies delineation for hydrological analysis (Tartakovsky, 2004). Due to some changes in nomenclature of methods through time, it has been difficult to identify all publications. Moreover, this thesis mostly focuses on the application of NNs, however, we give an additional overview of geoscientific applications of shallow ML.

2.3.1.1 Neural Networks in Geoscience

Early applications of neural networks were prominent in seismic data processing and analysis. Zhao and Mendel (1988) use a NN to perform seismic deconvolution early on. An application of seismic inversion with NNs was published by Röth and Tarantola (1994). Early ML-based electromagnetic geophysics performs subsurface localization (Poulton et al., 1992) and magnetotelluric inversion via Hopfield NNs (Zhang and Paulson, 1997). Feng and Seto (1998) applied NN to model geomechanical microfractures in triaxial compression tests. Interestingly, Legget et al. (1996) used a combination of Self-Organizing Maps (SOM) and back-propagation NNs that function similar to modern day Convolutional Neural Networks (CNNs) to perform 3D horizon tracking (Leggett et al., 2003). With the recent DL explosion, papers on seismic interpretation have gotten very popular, given the similarity to 2D segmentation tasks (cf. Table 2.1).

2.3.2 Challenges of ML in geoscience

2.3.3 Partial Solutions to Challenges

Topic	Publications
First Break Picking	Murat and Rudman (1992), McCormack et al. (1993), Dai and MacBeth (1997), and Ross et al. (2018)
Ground Penetrating Radar	Al-Nuaimy et al. (2000), Gamba and Lossani (2000), Shihab et al. (2002a), Shihab et al. (2002b), Youn and Chen (2002), Birkenfeld (2010), Cui et al. (2010), Maas and Schmalzl (2013), Núñez-Nieto et al. (2014), Mertens et al. (2016), Hansen and Cordua (2017), and Kilic and Eren (2018)
Seismic Deconvolution	Zhao and Mendel (1988), Wang and Teng (1997), Calderón-Macías et al. (1997), and Harrigan et al. (1991)
Seismic Horizon Picking	Huang et al. (1990), Legget et al. (1996), Zhang et al. (2001), and Leggett et al. (2003)
Seismic Interpretation	Meldahl et al. (2001), Strecker and Uden (2002), Klose (2006), Zheng et al. (2014), Marroquín (2014), Qi et al. (2016), Zhao et al. (2016), Roden et al. (2015), Huang et al. (2017b), Lewis and Vigh (2017), Waldeland and Solberg (2017), Guo et al. (2017), Zhao et al. (2017), Veillard et al. (2018), Araya-Polo et al. (2017), **Dramsch, Jesper Sören** and Lüthje (2018), Chevitarese et al. (2018), Gramstad and Nickel (2018), Guitton (2018), Purves et al. (2018), Shafiq et al. (2018a), Shafiq et al. (2018b), Waldeland et al. (2018), AlRegib et al. (2018), Le Bouteiller et al. (2018), Li et al. (2018), Sacrey and Roden (2018), Shafiq, Prabhushankar, et al. (2018c), and Wu and Zhang (2018)
Seismic Inversion	Röth and Tarantola (1994), Langer et al. (1996), Iturrarán-Viveros (2012), Ansari (2014), Verma et al. (2014), Golsanami et al. (2015), Schuster (2018), Araya-Polo et al. (2018), Mosser et al. (2018b), Mosser et al. (2018a), and Richardson (2018)
Seismic Tomography	Bauer et al. (2008) and Braeuer and Bauer (2015)
Seismic Well-Tie	Chaki et al. (2018)
Well-Log analysis	Huang et al. (1996), Fung et al. (1997), Bhatt and Helle (2002), Helle and Bhatt (2002), Asoodeh and Bagheripour (2014), Anifowose et al. (2017), Saporetti et al. (2018), Maiti and Tiwari (2010), Chang et al. (2002), Bauer et al. (2015), Emelyanova et al. (2017), and Carreira et al. (2018)

Table 2.1: Neural Networks in Geophysics.

CHAPTER 3

Synopsis

CHAPTER 4

Data Preparation and Analysis

CHAPTER 5

Foundations of Deep Learning for Seismic Data Analysis

CHAPTER 6

Deep Neural Networks for 4D Seismic Inversion

CHAPTER 7
Deep Convolutional Networks for 4D Time Shift Extraction

Acronyms

AE AutoEncoder

AGI Artificial General Intelligence

AI Artificial Intelligence

AUC Area Under Curve

ANN Artificial Neural Network

BN Batch Normalization

CNN Convolutional Neural Network

CPU Central Processing Unit

DCGAN Deep Convolutional Generative Adversarial Network

DL Deep Learning

DNN Deep Neural Network

DT Decision Tree

ELU Expenantial Linear Unit

EM Expectation-maximization

FCN Fully Convolutional Network

FFT Fast Fourier transform

GAN Generative Adversarial Network

GMM Gaussian mixture model

GPU Graphical Processing Unit

HMM Hidden Markov Model

iid independent and identically distributed

KL Kullback-Leibler divergence

KNN Nearest Neighbour

LReLu Leaky Rectified Linear Unit

LSTM Long Short-Term Memory

MAE Mean Absolute Error

MCMC Markov Chain Monte Carlo

ML Machine Learning

MLP Multi-Layer Perceptron

MSE Mean Squared Error

NAS Neural Architecture Search

NLP Natural Language Processing

NN Neural Network

NRMS Normalized Root Mean Sqared Error

PReLU Parameterized Rectified Linear Unit

ReLU Rectified Linear Unit

RF Random Forest

RNN Recurrent Neural Network

RMM Random Markov Model

RMS Root Mean Sqared Error

ROC Receiver Operating Characteristic

SGD Stochastic Gradient Descent

SOM Self-Organizing Maps

SOTA state-of-the-art

SVM Support Vector Machines

TF Tensorflow

TPU Tensor Processing Unit

VAE Variational AutoEncoder

Bibliography

Bayes, T. (1763). "LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S". In: *Philosophical transactions of the Royal Society of London* 53, pp. 370–418.

Taylor, R. (1843). *Scientific memoirs, selected from the transactions of foreign academies of science and learned societies, and from foreign journals*. Vol. 3. R. and JE Taylor.

Markov, A. A. (1906). "Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga". In: *Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete* 15.135-156, p. 18.

Kolmogorov, A. N. (1939). "Sur l'interpolation et extrapolation des suites stationnaires". In: *CR Acad Sci* 208, pp. 2043–2045.

Turing, A. M. (1950). "I.—Computing Machinery and Intelligence". In: *Mind* LIX.236, pp. 433–460. ISSN: 0026-4423. DOI: `10.1093/mind/LIX.236.433`. eprint: `http://oup.prod.sis.lan/mind/article-pdf/LIX/236/433/30123314/lix-236-433.pdf`. URL: `https://doi.org/10.1093/mind/LIX.236.433`.

Krige, D. G. (1951). "A statistical approach to some mine valuation and allied problems on the Witwatersrand". English. PhD thesis. Johannesburg.

Nash, J. (1951). "Non-cooperative games". In: *Annals of mathematics*, pp. 286–295.

Rosenblatt, F. (1958). "The perceptron: a probabilistic model for information storage and organization in the brain." In: *Psychological review* 65.6, p. 386.

Belson, W. A. (1959). "Matching and prediction on the principle of biological classification". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 8.2, pp. 65–75.

Kelley, H. J. (1960). "Gradient theory of optimal flight paths". In: *Ars Journal* 30.10, pp. 947–954.

Bryson, A. E. (1961). "A gradient method for optimizing multi-stage allocation processes". In: *Proc. Harvard Univ. Symposium on digital computers and their applications*. Vol. 72.

Dreyfus, S. (1962). "The numerical solution of variational problems". In: *Journal of Mathematical Analysis and Applications* 5.1, pp. 30–45.

Preston, F. W. and J. Henderson (1964). *Fourier series characterization of cyclic sediments for stratigraphic correlation*. Kansas Geological Survey.

Cover, T. and P. Hart (1967). "Nearest neighbor pattern classification". In: *IEEE transactions on information theory* 13.1, pp. 21–27.

Linnainmaa, S. (1970). “The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors”. In: *Master’s Thesis (in Finnish), Univ. Helsinki*, pp. 6–7.

Markov, A. A. (1971). “Extension of the limit theorems of probability theory to a sum of variables connected in a chain”. In: *Dynamic probabilistic systems* 1. Reprint in English of (Markov, 1906), pp. 552–577.

Schwarzacher, W. (1972). “The semi-Markov process as a general sedimentation model”. In: *Mathematical Models of Sedimentary Processes.* Springer, pp. 247–268.

Newendorp, P. D. (1976). “Decision analysis for petroleum exploration”. In:

Fukushima, K. (1980). “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position”. In: *Biological cybernetics* 36.4, pp. 193–202.

Hopfield, J. J. (1982). “Neural networks and physical systems with emergent collective computational abilities”. In: *Proceedings of the national academy of sciences* 79.8, pp. 2554–2558.

Kohonen, T. (1982). “Self-organized formation of topologically correct feature maps”. In: *Biological cybernetics* 43.1, pp. 59–69.

Goldberg, D. E. and J. H. Holland (1988). “Genetic Algorithms and Machine Learning”. en. In: *Mach. Learn.* 3.2-3, pp. 95–99. ISSN: 0885-6125, 1573-0565. DOI: `10.1023/A:1022602019183`. URL: `https://link.springer.com/article/10.1023/A:1022602019183`.

Rumelhart, D. E., G. E. Hinton, R. J. Williams, et al. (1988). “Learning representations by back-propagating errors”. In: *Cognitive modeling* 5.3, p. 1.

Zhao, X. and J. M. Mendel (1988). “Minimum-variance deconvolution using artificial neural networks”. In: *SEG Technical Program Expanded Abstracts.* URL: `https://library.seg.org/doi/pdf/10.1190/1.1892433`.

Hornik, K., M. Stinchcombe, and H. White (1989). “Multilayer feedforward networks are universal approximators”. In: *Neural Netw.* 2.5, pp. 359–366. ISSN: 0893-6080. DOI: `10.1016/0893-6080(89)90020-8`. URL: `http://www.sciencedirect.com/science/article/pii/0893608089900208`.

Watkins, C. J. C. H. (1989). “Learning from delayed rewards”. In:

Huang, K. Y., W. R. I. Chang, and H. T. Yen (1990). “Self-organizing neural network for picking seismic horizons”. In: *SEG Technical Program Expanded.* URL: `https://library.seg.org/doi/pdf/10.1190/1.1890183`.

Harrigan, E., J. R. Kroh, W. A. Sandham, and T. S. Durrani (1991). “Seismic wavelet extraction using artificial neural networks”. In: *1991 Second International Conference on Artificial Neural Networks.* ieeexplore.ieee.org, pp. 95–99. URL: `https://ieeexplore.ieee.org/abstract/document/140293/`.

Hochreiter, S. (1991). “Untersuchungen zu dynamischen neuronalen Netzen”. In: *Diploma, Technische Universität München* 91.1.

McCormack, M. (1991). “Neural computing in geophysics”. In: *Lead. Edge* 10.1, pp. 11–15. ISSN: 1070-485X. DOI: `10.1190/1.1436771`. URL: `https://doi.org/10.1190/1.1436771`.

Reddy, R. and G. Bonham-Carter (1991). "A Decision-Tree Approach to Mineral Potential Mapping in Snow Lake Area, Manitoba". In: *Canadian Journal of Remote Sensing* 17.2, pp. 191–200. DOI: 10.1080/07038992.1991.10855292. eprint: https://doi.org/10.1080/07038992.1991.10855292. URL: https://doi.org/10.1080/07038992.1991.10855292.

Krogh, A. and J. A. Hertz (1992). "A simple weight decay can improve generalization". In: *Advances in neural information processing systems*, pp. 950–957.

Murat, M. E. and A. J. Rudman (1992). "AUTOMATED FIRST ARRIVAL PICKING: A NEURAL NETWORK APPROACH1". In: *Geophys. Prospect.* 40.6, pp. 587–604. ISSN: 0016-8025, 1365-2478. DOI: 10.1111/j.1365-2478.1992.tb00543.x. URL: http://doi.wiley.com/10.1111/j.1365-2478.1992.tb00543.x.

Poulton, M., B. Sternberg, and C. Glass (1992). "Location of subsurface targets in geophysical data using neural networks". In: *Geophysics* 57.12, pp. 1534–1544. ISSN: 0016-8033. DOI: 10.1190/1.1443221. URL: https://doi.org/10.1190/1.1443221.

McCormack, M. D., D. E. Zaucha, and D. W. Dushek (1993). "First-break refraction event picking and seismic data trace editing using neural networks". In: *Geophysics*. ISSN: 0016-8033. URL: https://library.seg.org/doi/abs/10.1190/1.1443352.

Röth, G. and A. Tarantola (1994). "Neural networks and inversion of seismic data". In: *J. Geophys. Res.* 99.B4, p. 6753. ISSN: 0148-0227. DOI: 10.1029/93JB01563. URL: http://doi.wiley.com/10.1029/93JB01563.

Cortes, C. and V. Vapnik (1995). "Support-vector networks". In: *Machine learning* 20.3, pp. 273–297.

Ho, T. K. (1995). "Random decision forests". In: *Proceedings of 3rd international conference on document analysis and recognition*. Vol. 1. IEEE, pp. 278–282.

Huang, Z., J. Shimeld, M. Williamson, and J. Katsube (1996). "Permeability prediction with artificial neural network modeling in the Venture gas field, offshore eastern Canada". In: *Geophysics* 61.2, pp. 422–436. ISSN: 0016-8033. DOI: 10.1190/1.1443970. URL: https://doi.org/10.1190/1.1443970.

Langer, H., G. Nunnari, and L. Occhipinti (1996). "Estimation of seismic waveform governing parameters with neural networks". In: *J. Geophys. Res.* 101.B9, pp. 20109–20118. ISSN: 0148-0227. DOI: 10.1029/96JB00948. URL: http://doi.wiley.com/10.1029/96JB00948.

Legget, M., W. A. Sandham, and T. S. Durrani (1996). "3D horizon tracking using artificial neural networks". In: *First Break*. ISSN: 0263-5046. URL: http://www.earthdoc.org/publication/publicationdetails/?publication=28049.

Calderón-Macías, C., M. Sen, and P. Stoffa (1997). "Hopfield neural networks, and mean field annealing for seismic deconvolution and multiple attenuation". In: *Geophysics* 62.3, pp. 992–1002. ISSN: 0016-8033. DOI: 10.1190/1.1444205. URL: https://doi.org/10.1190/1.1444205.

Dai, H. and C. MacBeth (1997). "The application of back-propagation neural network to automatic picking seismic arrivals from single-component recordings". In: *J. Geophys. Res. [Solid Earth]*. URL: https://onlinelibrary.wiley.com/doi/abs/10.1029/97JB00625.

Fung, C. C., K. W. Wong, and H. Eren (1997). “Modular artificial neural network for prediction of petrophysical properties from well log data”. In: *IEEE Trans. Instrum. Meas.* 46.6, pp. 1295–1299. ISSN: 0018-9456. DOI: 10.1109/19.668276. URL: http://dx.doi.org/10.1109/19.668276.

Hochreiter, S. and J. Schmidhuber (1997). “Long short-term memory”. In: *Neural computation* 9.8, pp. 1735–1780.

Wang, J. and T.-L. Teng (1997). “Identification and picking of S phase using an artificial neural network”. In: *Bull. Seismol. Soc. Am.* 87.5, pp. 1140–1149. ISSN: 0037-1106. URL: https://pubs.geoscienceworld.org/ssa/bssa/article-abstract/87/5/1140/120211.

Zhang, Y. and K. V. Paulson (1997). “Magnetotelluric inversion using regularized Hopfield neural networks”. In: *Geophys. Prospect.* ISSN: 0016-8025. URL: http://www.earthdoc.org/publication/publicationdetails/?publication=33881.

Feng, X.-T. and M. Seto (1998). “Neural network dynamic modelling of rock microfracturing sequences under triaxial compressive stress conditions”. In: *Tectonophysics* 292.3, pp. 293–309. ISSN: 0040-1951. DOI: 10.1016/S0040-1951(98)00072-9. URL: http://www.sciencedirect.com/science/article/pii/S0040195198000729.

Baan, M. van der and C. Jutten (2000). “Neural networks in geophysical applications”. In: *Geophysics* 65.4, pp. 1032–1047. ISSN: 0016-8033. DOI: 10.1190/1.1444797. URL: https://doi.org/10.1190/1.1444797.

Gamba, P. and S. Lossani (2000). “Neural detection of pipe signatures in ground penetrating radar images”. In: *IEEE Trans. Geosci. Remote Sens.* 38.2, pp. 790–797. ISSN: 0196-2892. DOI: 10.1109/36.842008. URL: http://dx.doi.org/10.1109/36.842008.

Al-Nuaimy, W., Y. Huang, M. Nakhkash, M. T. C. Fang, V. T. Nguyen, and A. Eriksen (2000). “Automatic detection of buried utilities and solid objects with GPR using neural networks and pattern recognition”. In: *J. Appl. Geophys.* 43.2, pp. 157–165. ISSN: 0926-9851. DOI: 10.1016/S0926-9851(99)00055-5. URL: http://www.sciencedirect.com/science/article/pii/S0926985199000555.

Meldahl, P., R. Heggland, B. Bril, and P. de Groot (2001). “Identifying faults and gas chimneys using multiattributes and neural networks”. In: *Lead. Edge* 20.5, pp. 474–482. ISSN: 1070-485X. DOI: 10.1190/1.1438976. URL: https://doi.org/10.1190/1.1438976.

Mjolsness, E. and D. DeCoste (2001). “Machine learning for science: state of the art and future prospects”. en. In: *Science* 293.5537, pp. 2051–2055. ISSN: 0036-8075. DOI: 10.1126/science.293.5537.2051. URL: http://dx.doi.org/10.1126/science.293.5537.2051.

Zhang, L., J. Quieren, and J. Schuelke (2001). “Chapter 10 Self-Organizing Map (SOM) network for tracking horizons and classifying seismic traces”. In: *Handbook of Geophysical Exploration: Seismic Exploration*. Ed. by M. M. Poulton. Vol. 30. Pergamon, pp. 155–170. DOI: 10.1016/S0950-1401(01)80024-0. URL: http://www.sciencedirect.com/science/article/pii/S0950140101800240.

Bhatt, A. and H. B. Helle (2002). "Determination of facies from well logs using modular neural networks". In: *Pet. Geosci.* ISSN: 1354-0793. URL: http://www.earthdoc.org/publication/publicationdetails/?publication=37980.

Chang, H.-C., D. C. Kopaska-Merkel, and H.-C. Chen (2002). "Identification of lithofacies using Kohonen self-organizing maps". In: *Comput. Geosci.* 28.2, pp. 223–229. ISSN: 0098-3004. DOI: 10.1016/S0098-3004(01)00067-X. URL: http://www.sciencedirect.com/science/article/pii/S009830040100067X.

Collobert, R., S. Bengio, and J. Mariéthoz (2002). *Torch: a modular machine learning software library.* Tech. rep. Idiap.

Helle, H. B. and A. Bhatt (2002). "Fluid saturation from well logs using committee neural networks". In: *Pet. Geosci.* ISSN: 1354-0793. URL: http://pg.lyellcollection.org/content/8/2/109.short.

Shihab, S., W. Al-Nuaimy, and A. Eriksen (2002a). "Image processing and neural network techniques for automatic detection and interpretation of ground penetrating radar data". In: *Proceedings of the 6th*. URL: https://www.researchgate.net/profile/Asger_Eriksen/publication/255015233_Image_processing_and_neural_network_techniques_for_automatic_detection_and_interpretation_of_ground_penetrating_radar_data/links/54c664b60cf219bbe4f842ba/Image-processing-and-neural-network-techniques-for-automatic-detection-and-interpretation-of-ground-penetrating-radar-data.pdf.

Shihab, S., W. Al-Nuaimy, Y. Huang, and A. Eriksen (2002b). "Neural network target identifier based on statistical features of GPR signals". In: *Ninth International Conference on Ground Penetrating Radar*. Vol. 4758. International Society for Optics and Photonics, pp. 135–139. DOI: 10.1117/12.462228. URL: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/4758/0000/Neural-network-target-identifier-based-on-statistical-features-of-GPR/10.1117/12.462228.short.

Strecker, U. and R. Uden (2002). "Data mining of 3D poststack seismic attribute volumes using Kohonen self-organizing maps". In: *Lead. Edge* 21.10, pp. 1032–1037. ISSN: 1070-485X. DOI: 10.1190/1.1518442. URL: https://doi.org/10.1190/1.1518442.

Youn, H.-S. and C.-C. Chen (2002). "Automatic GPR target detection and clutter reduction using neural network". In: *Ninth International Conference on Ground Penetrating Radar*. Vol. 4758. International Society for Optics and Photonics, pp. 579–583. DOI: 10.1117/12.462229. URL: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/4758/0000/Automatic-GPR-target-detection-and-clutter-reduction-using-neural-network/10.1117/12.462229.short.

Leggett, M., W. A. Sandham, and T. S. Durrani (2003). "Automated 3-D Horizon Tracking and Seismic Classification Using Artificial Neural Networks". In: *Geophysical Applications of Artificial Neural Networks and Fuzzy Logic*. Ed. by W. A. Sandham and M. Leggett. Dordrecht: Springer Netherlands, pp. 31–44. ISBN: 9789401702713. DOI: 10.1007/978-94-017-0271-3_3. URL: https://doi.org/10.1007/978-94-017-0271-3_3.

Li, J. and J. Castagna (2004). "Support Vector Machine (SVM) pattern recognition to AVO classification". In: *Geophys. Res. Lett.* 31.2, p. 948. ISSN: 0094-8276. DOI: 10.1029/2003GL018299. URL: http://doi.wiley.com/10.1029/2003GL018299.

Tartakovsky, D. M. (2004). "Delineation of geologic facies with statistical learning theory". In: *Geophys. Res. Lett.* 31.18, p. 121. ISSN: 0094-8276. DOI: 10.1029/2004GL020864. URL: http://doi.wiley.com/10.1029/2004GL020864.

Hinton, G. E. and R. R. Salakhutdinov (2006). "Reducing the dimensionality of data with neural networks". In: *science* 313.5786, pp. 504–507.

Klose, C. D. (2006). "Self-organizing maps for geoscientific data analysis: geological interpretation of multidimensional geophysical data". In: *Computational Geosciences* 10.3, pp. 265–277. ISSN: 1573-1499. DOI: 10.1007/s10596-006-9022-x. URL: https://doi.org/10.1007/s10596-006-9022-x.

Bennett, J., S. Lanning, et al. (2007). "The netflix prize". In: *Proceedings of KDD cup and workshop*. Vol. 2007. New York, NY, USA., p. 35.

Bauer, K., R. G. Pratt, C. Haberland, and M. Weber (2008). "Neural network analysis of crosshole tomographic images: The seismic signature of gas hydrate bearing sediments in the Mackenzie Delta (NW Canada)". In: *Geophys. Res. Lett.* 35.19, p. 340. ISSN: 0094-8276. DOI: 10.1029/2008GL035263. URL: http://doi.wiley.com/10.1029/2008GL035263.

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

Birkenfeld, S. (2010). "Automatic detection of reflexion hyperbolas in GPR data with neural networks". In: *World Automation Congress*. researchgate.net, pp. 1–6. URL: https://www.researchgate.net/profile/Sven_Birkenfeld/publication/221671913_Automatic_detection_of_reflexion_hyperbolas_in_GPR_data_with_neural_networks/links/00b7d5303634cb6e1a000000.pdf.

Cui, Y.-A., L. Wang, and J.-P. Xiao (2010). "Automatic feature recognition for GPR image processing". In: *Proc. World Acad. of Sci. Eng. Technol.* 61, pp. 176–179. ISSN: 1307-6884. URL: https://pdfs.semanticscholar.org/edd6/3447f33c032fe26dfb970e92f6194e98d pdf.

Glorot, X. and Y. Bengio (2010). "Understanding the difficulty of training deep feedforward neural networks". In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 249–256.

Maiti, S. and R. K. Tiwari (2010). "Neural network modeling and an uncertainty analysis in Bayesian framework: A case study from the KTB borehole site". In: *J. Geophys. Res.* 115.B10, E67. ISSN: 0148-0227. DOI: 10.1029/2010JB000864. URL: http://doi.wiley.com/10.1029/2010JB000864.

Russell, S. J. and P. Norvig (2010). *Artificial Intelligence - A Modern Approach, Third International Edition*. Pearson Education. ISBN: 978-0-13-207148-2. URL: http://vig.pearsoned.com/store/product/1,1207,store-12521%5C_isbn-0136042597,00.html.

Chang, C.-C. and C.-J. Lin (2011). "LIBSVM: A Library for Support Vector Machines". In: *ACM Trans. Intell. Syst. Technol.* 2.3, 27:1–27:27. ISSN: 2157-6904. DOI: 10.1145/1961189.1961199. URL: http://doi.acm.org/10.1145/1961189.1961199.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay (2011). "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

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

Iturrarán-Viveros, U. (2012). "Smooth regression to estimate effective porosity using seismic attributes". In: *J. Appl. Geophys.* 76, pp. 1–12. ISSN: 0926-9851. DOI: 10.1016/j.jappgeo.2011.10.012. URL: http://www.sciencedirect.com/science/article/pii/S0926985111002485.

Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012). "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*, pp. 1097–1105.

LeCun, Y. A., L. Bottou, G. B. Orr, and K.-R. Müller (2012). "Efficient backprop". In: *Neural networks: Tricks of the trade*. Springer, pp. 9–48.

Buitinck, L., G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux (2013). "API design for machine learning software: experiences from the scikit-learn project". In: *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pp. 108–122.

Goodfellow, I. J., D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, et al. (2013). "Challenges in representation learning: A report on three machine learning contests". In: *International Conference on Neural Information Processing*. Springer, pp. 117–124.

Maas, C. and J. Schmalzl (2013). "Using pattern recognition to automatically localize reflection hyperbolas in data from ground penetrating radar". In: *Comput. Geosci.* 58, pp. 116–125. ISSN: 0098-3004. DOI: 10.1016/j.cageo.2013.04.012. URL: http://www.sciencedirect.com/science/article/pii/S009830041300112X.

Russakovsky, O., J. Deng, Z. Huang, A. C. Berg, and L. Fei-Fei (2013). "Detecting avocados to zucchinis: what have we done, and where are we going?" In: *International Conference on Computer Vision (ICCV)*.

Sutskever, I., J. Martens, G. Dahl, and G. Hinton (2013). "On the importance of initialization and momentum in deep learning". In: *Proceedings of the 30th International Conference on Machine Learning*. Ed. by S. Dasgupta and D. McAllester. Vol. 28. Proceedings of Machine Learning Research 3. Atlanta, Georgia, USA: PMLR, pp. 1139–1147. URL: http://proceedings.mlr.press/v28/sutskever13.html.

Ansari, H. R. (2014). "Use seismic colored inversion and power law committee machine based on imperial competitive algorithm for improving porosity prediction in a heterogeneous reservoir". In: *J. Appl. Geophys.* 108, pp. 61–68. ISSN: 0926-9851. DOI:

10.1016/j.jappgeo.2014.06.016. URL: http://www.sciencedirect.com/science/article/pii/S092698511400192X.

Asoodeh, M. and P. Bagheripour (2014). "ACE stimulated neural network for shear wave velocity determination from well logs". In: *J. Appl. Geophys.* 107, pp. 102–107. ISSN: 0926-9851. DOI: 10.1016/j.jappgeo.2014.05.014. URL: http://www.sciencedirect.com/science/article/pii/S0926985114001487.

Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio (2014a). "Generative Adversarial Nets". In: *Advances in Neural Information Processing Systems 27*. Ed. by Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger. Curran Associates, Inc., pp. 2672–2680. URL: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf.

Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio (2014b). "Generative adversarial nets". In: *Advances in neural information processing systems*, pp. 2672–2680.

Kingma, D. P. and J. Ba (2014). "Adam: A method for stochastic optimization". In: *arXiv*. eprint: arXiv:1412.6980.

Marroquín, I. (2014). "A knowledge-integration framework for interpreting seismic facies". In: *Interpretation* 2.1, SA1–SA9. ISSN: 2324-8858. DOI: 10.1190/INT-2013-0057.1. URL: https://doi.org/10.1190/INT-2013-0057.1.

Núñez-Nieto, X., M. Solla, P. Gómez-Pérez, and H. Lorenzo (2014). "GPR Signal Characterization for Automated Landmine and UXO Detection Based on Machine Learning Techniques". en. In: *Remote Sensing* 6.10, pp. 9729–9748. DOI: 10.3390/rs6109729. URL: https://www.mdpi.com/2072-4292/6/10/9729/htm.

Verma, A. K., S. Chaki, A. Routray, W. K. Mohanty, and M. Jenamani (2014). "Quantification of sand fraction from seismic attributes using Neuro-Fuzzy approach". In: *J. Appl. Geophys.* 111, pp. 141–155. ISSN: 0926-9851. DOI: 10.1016/j.jappgeo.2014.10.005. URL: http://www.sciencedirect.com/science/article/pii/S0926985114002912.

Zheng, Z. H., P. Kavousi, and H. B. Di (2014). "Multi-attributes and neural network-based fault detection in 3D seismic interpretation". In: *Adv. Mat. Res.* ISSN: 1022-6680. URL: https://www.scientific.net/AMR.838-841.1497.

Bauer, K., J. Kulenkampff, J. Henninges, and E. Spangenberg (2015). "Lithological control on gas hydrate saturation as revealed by signal classification of NMR logging data". In: *J. Geophys. Res. [Solid Earth]* 120.9, pp. 6001–6017. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/2015JB012150.

Braeuer, B. and K. Bauer (2015). "A new interpretation of seismic tomography in the southern Dead Sea basin using neural network clustering techniques: INTERPRETATION OF TOMOGRAPHY IN THE SDSB". In: *Geophys. Res. Lett.* Lecture Notes Comput. Sci 42.22, pp. 9772–9780. ISSN: 0094-8276. DOI: 10.1002/2015GL066559. URL: http://doi.wiley.com/10.1002/2015GL066559.

Chollet, F. et al. (2015). *Keras*. https://keras.io.

Golsanami, N., A. Kadkhodaie-Ilkhchi, and A. Erfani (2015). "Synthesis of capillary pressure curves from post-stack seismic data with the use of intelligent estimators: a case study from the Iranian part of the South Pars ...". In: *Journal of Applied*. URL: https://www.sciencedirect.com/science/article/pii/S0926985114003413.

He, K., X. Zhang, S. Ren, and J. Sun (2015). "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification". In: *Proceedings of the IEEE international conference on computer vision*, pp. 1026–1034.

Ioffe, S. and C. Szegedy (2015). "Batch normalization: Accelerating deep network training by reducing internal covariate shift". In: *arXiv preprint arXiv:1502.03167*.

LeCun, Y., Y. Bengio, and G. Hinton (2015). "Deep learning". In: *nature* 521.7553, pp. 436–444.

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Y. Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. URL: http://tensorflow.org/.

Roden, R., T. Smith, and D. Sacrey (2015). "Geologic pattern recognition from seismic attributes: Principal component analysis and self-organizing maps". In: *Interpretation* 3.4, SAE59–SAE83. ISSN: 2324-8858. DOI: 10.1190/INT-2015-0037.1. URL: https://pubs.geoscienceworld.org/interpretation/article-abstract/3/4/SAE59/75892.

Ronneberger, O., P. Fischer, and T. Brox (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer, pp. 234–241.

Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei (2015). "ImageNet Large Scale Visual Recognition Challenge". In: *International Journal of Computer Vision (IJCV)* 115.3, pp. 211–252. DOI: 10.1007/s11263-015-0816-y.

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

Bishop, C. M. (2016). *Pattern Recognition and Machine Learning*. en. Springer New York. ISBN: 9781493938438. URL: https://market.android.com/details?id=book-kOXDtAEACAAJ.

Chen, T. and C. Guestrin (2016). "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. San Francisco, California, USA: ACM, pp. 785–794. ISBN: 978-1-4503-4232-2. DOI: 10.1145/2939672.2939785. URL: http://doi.acm.org/10.1145/2939672.2939785.

Goodfellow, I., Y. Bengio, and A. Courville (2016). *Deep Learning*. http://www.deeplearningbook.org. MIT Press.

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

Mertens, L., R. Persico, L. Matera, et al. (2016). "Automated Detection of Reflection Hyperbolas in Complex GPR Images With No A Priori Knowledge on the Medium". In: *IEEE Transactions on.* URL: https://ieeexplore.ieee.org/abstract/document/7230274/.

Qi, J., T. Lin, T. Zhao, F. Li, and K. Marfurt (2016). "Semisupervised multiattribute seismic facies analysis". In: *Interpretation* 4.1, SB91–SB106. ISSN: 2324-8858. DOI: 10.1190/INT-2015-0098.1. URL: https://doi.org/10.1190/INT-2015-0098.1.

Zhao, T., J. Zhang, F. Li, and K. Marfurt (2016). "Characterizing a turbidite system in Canterbury Basin, New Zealand, using seismic attributes and distance-preserving self-organizing maps". In: *Interpretation* 4.1, SB79–SB89. ISSN: 2324-8858. DOI: 10.1190/INT-2015-0094.1. URL: https://doi.org/10.1190/INT-2015-0094.1.

Anifowose, F., C. Ayadiuno, and F. Rashedian (2017). "Carbonate Reservoir Cementation Factor Modeling Using Wireline Logs and Artificial Intelligence Methodology". In: *79th EAGE Conference and Exhibition 2017-Workshops*. earthdoc.org. URL: http://www.earthdoc.org/publication/publicationdetails/?publication=89285.

Araya-Polo, M., T. Dahlke, C. Frogner, C. Zhang, T. Poggio, and D. Hohl (2017). "Automated fault detection without seismic processing". In: *Lead. Edge* 36.3, pp. 208–214. ISSN: 1070-485X. DOI: 10.1190/tle36030208.1. URL: https://doi.org/10.1190/tle36030208.1.

Arjovsky, M., S. Chintala, and L. Bottou (2017). "Wasserstein gan". In: *arXiv preprint arXiv:1701.07875*.

Emelyanova, I., M. Pervukhina, M. Clennell, et al. (2017). "Unsupervised identification of electrofacies employing machine learning". In: *79th EAGE Conference*. URL: http://www.earthdoc.org/publication/publicationdetails/?publication=89274.

Gulrajani, I., F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville (2017). "Improved training of wasserstein gans". In: *Advances in neural information processing systems*, pp. 5767–5777.

Guo, R., Y. S. Zhang, H. Lin, and W. Liu (2017). "Sweet Spot Interpretation from Multiple Attributes: Machine Learning and Neural Networks Technologies". In: *First EAGE/AMGP/AMGE Latin*. URL: http://www.earthdoc.org/publication/publicationdetails/?publication=90731.

Hansen, T. M. and K. S. Cordua (2017). "Efficient Monte Carlo sampling of inverse problems using a neural network-based forward—applied to GPR crosshole traveltime inversion". In: *Geophys. J. Int.* 211.3, pp. 1524–1533. ISSN: 0956-540X. DOI: 10.1093/gji/ggx380. URL: https://academic.oup.com/gji/article-abstract/211/3/1524/4157792.

Huang, G., Z. Liu, L. Van Der Maaten, and K. Q. Weinberger (2017a). "Densely connected convolutional networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708.

Huang, L., X. Dong, and T. Clee (2017b). "A scalable deep learning platform for identifying geologic features from seismic attributes". In: *Lead. Edge* 36.3, pp. 249–256.

ISSN: 1070-485X. DOI: 10.1190/tle36030249.1. URL: https://doi.org/10.1190/tle36030249.1.

Kadurin, A., S. Nikolenko, K. Khrabrov, A. Aliper, and A. Zhavoronkov (2017). "druGAN: An Advanced Generative Adversarial Autoencoder Model for de Novo Generation of New Molecules with Desired Molecular Properties in Silico". en. In: *Mol. Pharm.* 14.9, pp. 3098–3104. ISSN: 1543-8384, 1543-8392. DOI: 10.1021/acs.molpharmaceut.7b00346. URL: http://dx.doi.org/10.1021/acs.molpharmaceut.7b00346.

Lewis, W. and D. Vigh (2017). "Deep learning prior models from seismic images for full-waveform inversion". In: *SEG Technical Program Expanded Abstracts 2017*. SEG Technical Program Expanded Abstracts. Society of Exploration Geophysicists, pp. 1512–1517. DOI: 10.1190/segam2017-17627643.1. URL: https://doi.org/10.1190/segam2017-17627643.1.

Paszke, A., S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer (2017). "Automatic Differentiation in PyTorch". In: *NIPS Autodiff Workshop*.

Schütt, K. T., F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko (2017). "Quantum-chemical insights from deep tensor neural networks". en. In: *Nat. Commun.* 8, p. 13890. ISSN: 2041-1723. DOI: 10.1038/ncomms13890. URL: http://dx.doi.org/10.1038/ncomms13890.

Shen, D., G. Wu, and H.-I. Suk (2017). "Deep Learning in Medical Image Analysis". en. In: *Annu. Rev. Biomed. Eng.* 19, pp. 221–248. ISSN: 1523-9829, 1545-4274. DOI: 10.1146/annurev-bioeng-071516-044442. URL: http://dx.doi.org/10.1146/annurev-bioeng-071516-044442.

Waldeland, A. U. and A. Solberg (2017). "Salt classification using deep learning". In: *79th EAGE Conference and Exhibition*. URL: http://www.earthdoc.org/publication/publicationdetails/?publication=88635.

Zhao, T., F. Li, and K. Marfurt (2017). "Constraining self-organizing map facies analysis with stratigraphy: An approach to increase the credibility in automatic seismic facies classification". In: *Interpretation* 5.2, T163–T171. ISSN: 2324-8858. DOI: 10.1190/INT-2016-0132.1. URL: https://doi.org/10.1190/INT-2016-0132.1.

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

AlRegib, G., M. Deriche, Z. Long, H. Di, Z. Wang, Y. Alaudah, M. A. Shafiq, and M. Alfarraj (2018). "Subsurface Structure Analysis Using Computational Interpretation and Learning: A Visual Signal Processing Perspective". In: *IEEE Signal Process. Mag.* 35.2, pp. 82–98. ISSN: 1053-5888. DOI: 10.1109/MSP.2017.2785979. URL: http://dx.doi.org/10.1109/MSP.2017.2785979.

Araya-Polo, M., J. Jennings, A. Adler, and T. Dahlke (2018). "Deep-learning tomography". In: *Lead. Edge* 37.1, pp. 58–66. ISSN: 1070-485X. DOI: 10.1190/tle37010058.1. URL: https://doi.org/10.1190/tle37010058.1.

Carreira, V., C. P. Neto, and R. Bijani (2018). "A Comparison of Machine Learning Processes for Classification of Rock Units Using Well Log Data". In: *80th EAGE*

Conference and Exhibition 2018. earthdoc.org. URL: http://www.earthdoc.org/publication/publicationdetails/?publication=92908.

Chaki, S., A. Routray, and W. K. Mohanty (2018). "Well-Log and Seismic Data Integration for Reservoir Characterization: A Signal Processing and Machine-Learning Perspective". In: *IEEE Signal Process. Mag.* 35.2, pp. 72–81. ISSN: 1053-5888. DOI: 10.1109/MSP.2017.2776602. URL: http://dx.doi.org/10.1109/MSP.2017.2776602.

Chevitarese, D., D. Szwarcman, R. M. D. Silva, and E. V. Brazil (2018). "Seismic facies segmentation using deep learning". In: *ACE 2018 Annual Convention & Exhibition*. searchanddiscovery.com. URL: http://www.searchanddiscovery.com/documents/2018/42286chevitarese/ndx_chevitarese.pdf.

Ching, T., D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin, B. T. Do, G. P. Way, E. Ferrero, P.-M. Agapow, M. Zietz, M. M. Hoffman, W. Xie, G. L. Rosen, B. J. Lengerich, J. Israeli, J. Lanchantin, S. Woloszynek, A. E. Carpenter, A. Shrikumar, J. Xu, E. M. Cofer, C. A. Lavender, S. C. Turaga, A. M. Alexandari, Z. Lu, D. J. Harris, D. DeCaprio, Y. Qi, A. Kundaje, Y. Peng, L. K. Wiley, M. H. S. Segler, S. M. Boca, S. J. Swamidass, A. Huang, A. Gitter, and C. S. Greene (2018). "Opportunities and obstacles for deep learning in biology and medicine". en. In: *J. R. Soc. Interface* 15.141. ISSN: 1742-5689, 1742-5662. DOI: 10.1098/rsif.2017.0387. URL: http://dx.doi.org/10.1098/rsif.2017.0387.

Dramsch, Jesper Sören and M. Lüthje (2018). "Deep-learning seismic facies on state-of-the-art CNN architectures". In: *SEG Technical Program Expanded Abstracts 2018*. Society of Exploration Geophysicists, pp. 2036–2040. DOI: 10.1190/segam2018-2996783.1. URL: https://doi.org/10.1190/segam2018-2996783.1.

Gramstad, O. and M. Nickel (2018). "Automated Top Salt Interpretation Using a Deep Convolutional Net". In: *80th EAGE Conference and Exhibition 2018*. URL: http://www.earthdoc.org/publication/publicationdetails/?publication=92117.

Guitton, A. (2018). "3D Convolutional Neural Networks for Fault Interpretation". In: *80th EAGE Conference and Exhibition 2018*. URL: http://www.earthdoc.org/publication/publicationdetails/?publication=92118.

Kilic, G. and L. Eren (2018). "Neural network based inspection of voids and karst conduits in hydro–electric power station tunnels using GPR". In: *J. Appl. Geophys.* 151, pp. 194–204. ISSN: 0926-9851. DOI: 10.1016/j.jappgeo.2018.02.026. URL: http://www.sciencedirect.com/science/article/pii/S0926985118301484.

Le Bouteiller, P., J. Charléty, F. Delprat-Jannaud, D. Granjeon, and C. Gorini (2018). "Mixing Unsupervised and Knowledge-Based Analysis for Heterogeneous Object Delineation in Seismic Data". In: *80th EAGE Conference and Exhibition 2018*. earthdoc.org. URL: http://www.earthdoc.org/publication/publicationdetails/?publication=92121.

Li, L., X. W. Li, Z. Wan, Y. Liu, and L. Zhang (2018). "Multiscale Pre-Stack Seismic Attribute Enhancement Using Radial Basis Function Network". In: *80th EAGE Conference and*. URL: http://www.earthdoc.org/publication/publicationdetails/?publication=92122.

Mosser, L., O. Dubrule, and M. J. Blunt (2018a). “Stochastic seismic waveform inversion using generative adversarial networks as a geological prior”. In: arXiv: 1806.03720 [physics.geo-ph]. URL: http://arxiv.org/abs/1806.03720.

Mosser, L., W. Kimman, J. Dramsch, S. Purves, et al. (2018b). “Rapid seismic domain transfer: Seismic velocity inversion and modeling using deep generative neural networks”. In: *EAGE Conference and ...*. URL: http://www.earthdoc.org/publication/publicationdetails/?publication=92120.

Purves, S., B. Alaei, and E. Larsen (2018). “Bootstrapping Machine-Learning Based Seismic Fault Interpretation”. In: *ACE 2018 Annual Convention &*. URL: http://www.searchanddiscovery.com/abstracts/html/2018/ace2018/abstracts/2856016.html.

Richardson, A. (2018). “Seismic Full-Waveform Inversion Using Deep Learning Tools and Techniques”. In: arXiv: 1801.07232 [physics.geo-ph]. URL: http://arxiv.org/abs/1801.07232.

Ross, Z. E., M. A. Meier, and E. Hauksson (2018). “P-wave arrival picking and first-motion polarity determination with deep learning”. In: *J. Geophys. Res.* ISSN: 0148-0227. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2017JB015251.

Sacrey, D. and R. Roden (2018). “Solving exploration problems with machine learning”. In: *First Break*. ISSN: 0263-5046. URL: http://www.earthdoc.org/publication/publicationdetails/?publication=92017.

Saporetti, C. M., L. G. da Fonseca, E. Pereira, and L. C. de Oliveira (2018). “Machine learning approaches for petrographic classification of carbonate-siliciclastic rocks using well logs and textural information”. In: *J. Appl. Geophys.* 155, pp. 217–225. ISSN: 0926-9851. DOI: 10.1016/j.jappgeo.2018.06.012. URL: http://www.sciencedirect.com/science/article/pii/S092698511630667X.

Schuster, G. T. S. (2018). “Machine learning and wave equation inversion of skeletonized data”. In: *80th EAGE Conference & Exhibition 2018 Workshop*. URL: http://www.earthdoc.org/publication/publicationdetails/?publication=93370.

Shafiq, M. A., M. Prabhushankar, and G. AlRegib (2018a). “Leveraging Sparse Features Learned from Natural Images for Seismic Understanding”. In: *80th EAGE Conference and*. URL: http://www.earthdoc.org/publication/publicationdetails/?publication=92123.

Shafiq, M. A., M. Prabhushankar, and G. AlRegib (2018b). “Leveraging Sparse Features Learned from Natural Images for Seismic Understanding”. In: *80th EAGE Conference and*. URL: http://www.earthdoc.org/publication/publicationdetails/?publication=92123.

Shafiq, M. A., M. Prabhushankar, et al. (2018c). “Attention models based on sparse autoencoders for seismic interpretation”. In: *ACE 2018 Annual*. URL: http://www.searchanddiscovery.com/abstracts/html/2018/ace2018/abstracts/2856356.html.

Veillard, A., O. Morère, M. Grout, et al. (2018). “Fast 3D Seismic Interpretation with Unsupervised Deep Learning: Application to a Potash Network in the North Sea”. In: *80th EAGE Conference*. URL: http://www.earthdoc.org/publication/publicationdetails/?publication=92124.

Waldeland, A., A. Jensen, L. Gelius, and A. Solberg (2018). “Convolutional neural networks for automated seismic interpretation”. In: *Lead. Edge* 37.7, pp. 529–537. ISSN: 1070-485X. DOI: 10.1190/tle37070529.1. URL: https://doi.org/10.1190/tle37070529.1.

Wu, H. and B. Zhang (2018). “A deep convolutional encoder-decoder neural network in assisting seismic horizon tracking”. In: arXiv: 1804.06814 [physics.geo-ph]. URL: http://arxiv.org/abs/1804.06814.

Zoph, B., V. Vasudevan, J. Shlens, and Q. V. Le (2018). “Learning transferable architectures for scalable image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8697–8710.

Real, E., A. Aggarwal, Y. Huang, and Q. V. Le (2019). “Regularized evolution for image classifier architecture search”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33, pp. 4780–4789.

Tan, M. and Q. V. Le (2019). “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”. In: *arXiv preprint arXiv:1905.11946*.