A Small-Scale Network for Seismic Patterns Classification

Monacér da Silva[1,2], Jean Charléty[1], Aurélia Fraysse[2] and Jean-Christophe Pesquet[3]
[1] *IFP Energies Nouvelles*
Rueil-Malmaison, France
[2] *Laboratoire des Signaux et Systèmes (L2S)*
CNRS, Univ. Paris-Saclay, Gif-Sur-Yvette, France
[3] *Centre de Vision Numérique (CVN),*
INRIA Saclay, Univ. Paris-Saclay, *Gif-Sur-Yvette, France*
Contact: monacer-ericson.da-silva@ifpen.fr, jean.charlety@ifpen.fr,
aurelia.fraysse@l2s.centralesupelec.fr, jean-christophe.pesquet@centralesupelec.fr

***Abstract*—Deep Convolutional Neural Networks (DCNNs) correspond to the state-of-art for image classification. However to train such systems it is necessary to have access to a large number of samples and powerful computational resources, given the huge number of involved parameters. In the field of seismic images, large and freely available databases are scarce due to their strategic interest. In this situation, large architectures lead to hardly tractable problems in terms of overfitting. In this paper, we propose a reduced-size CNN with low computational cost that allows high accuracy performance on two small seismic datasets. The results are compared with KNN, SVM and LeNet.**

***Index Terms*—Seismic Data Analysis, Machine Learning, Deep Neural Network, Classification, Small Datasets**

I. Introduction

Over the last years a wide variety of techniques have been developed to extract information from seismic images in order to improve their analysis for tasks ranging from lithology, mining, oil and gas exploration, civil construction to nuclear waste deposition [1]. The nature of such information can yield a direct geological interpretation when its extraction is based on a multi-attribute analysis. In this case, an attribute is a quantity derived from seismic data that is suitable to enhance relevant information of different seismic characteristics.

A large number of attributes were defined in this context, some of them useless or redundant, and their choice became an issue to be addressed [2]. Consequently, developments have been carried out to characterize seismic patterns with only a few descriptors as those based on texture similarity [3]–[5]. For the analysis of a seismic section and the labeling of its patterns, attribute selection is problematic as it depends on different choices for the most important aspects that characterize a pattern. An abstract analysis - a black box approach - can better capture nonlinear relationships (e.g. between seismic waves and anisotropic media) and be more robust to noise and processing errors [6], [7]. Therefore, the useful quantities provided by such powerful methods allow the detection or separation of different patterns and simplify the automation of complex tasks [1].

Such technologies, among them Deep Convolutional Neural Networks (DCNNs) [8], allow the questionable task of attribute/feature selection to be avoided. Given the recent developments provided by DCNNs, some successful results have been obtained in the field of seismic analysis as well [9].

In Section II a development guideline for the task of classification and segmentation of seismic patterns using DCNNs is provided. Then, we present the problem associated with the use of DCNNs, the seismic data constraints and the motivation for a reduced-size CNN able to learn with a small labeled dataset. Section III details the proposed CNN architecture. A preliminary comparison between the proposed CNN and other algorithms is presented in Section IV and the last section draws conclusions and perspectives.

II. Description of the problem

In the seismic domain, structures of interest depend strongly on the final application. Structures as faults, salt-domes, channels, and gas chimneys can be associated with gas and oil reservoirs discovery [10]; recognizing facies can assist in understanding the depositional environment of a region [11]. Each task (segmentation or classification) involves a different type of DCNN architecture. For segmentation, the choice will commonly be based on a U-Net [12] with its encoder-decoder structure for delineation of faults [13], salt-domes [14], horizons [15], facies [16]. For classification only the encoder part is considered. Some works show that different approaches (multi-attribute analysis or black box) using different algorithms have the possibility to classify seismic patterns with high-accuracy. A different set of attributes have been chosen in [17], [18] to feed various algorithms. In [19], a texture descriptor is given as a feature extractor to feed a DCNN in an unsupervised context. In [20] authors follow the conventional way, by providing directly the images as inputs for the DCNN and in [21] different architectures are compared.

All the aforementioned references for segmentation or classification tasks provide good results. But they have as a common characteristic the availability of a huge amount of data.

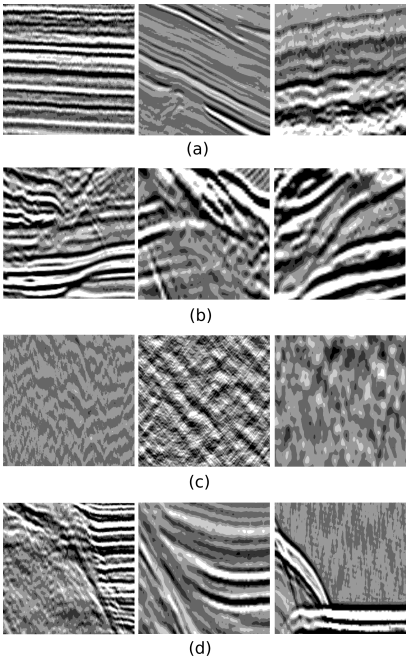

Fig. 1: Three samples for each class: (a) *Flat*, (b) *Fold*, (c) *Low Interest*, (d) *Sigmoid*.

In particular, annotated databases associated with the use of big architectures are available. This requirement is mandatory when using DCNNs. The most common benchmark datasets to compare different architectures contains from 10^4 to more than 10^6 samples as inputs to networks with millions of parameters. In the seismic image domain, large tagged image databases are scarce: it is hampered by strategic interests and labeling is an expensive and time consuming process. Due to this fact, [10] emphasized the use of weak labels and unsupervised learning. Despite that, the most common approach continues to be that of supervised learning with the help of data augmentation [22]. However there are some fields that are sensitive to arbitrary transformations applied to images. To contextualize this last point, let us consider the dataset that we have at our disposal: the IFPEN's dataset. It has 580 grayscale images containing 512×512 pixels and four patterns/classes (*flat*, *fold*, *low interest* and *sigmoid*). Each class represents a units with a geological interpretation and the classes are respectively composed of $221, 223, 100$ and 36 samples. Figure 1 shows 3 samples for each class of the dataset.

Consider the 90° rotation of a horizontal profile. The seismic meaning is different. In Figure 2 all four images presented are in the same class, fold (Figure 1(b)). Even without any transformation (a) and (b) are close to the flat class (Figure 1(a)) and in turn (c) and (d) to the sigmoid class (Figure 1(d)). In this domain, smooth transitions from one pattern to another is common and transformations as data augmentation must be done carefully.

To deal with such problems, [23] used a scattering network (SN) [24], with promising results in textural multiclass classification with small datasets [25]. SN is analogous to the

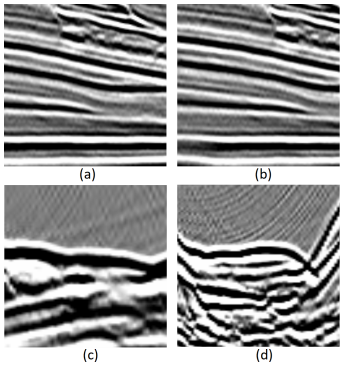

Fig. 2: Four examples of the fold class.

feature extractor block of a DCNN, and needs to be associated with a classifier or a classifier block. In between a feature reduction can be performed as the output of a SN has a high-dimensionality.

In the following, we propose a simple CNN with reduced computational cost that performs well for classification of two seismic labeled datasets even when a small quantity of samples is available.

III. A SMALL-SCALE CNN

The indication that DCNNs can perform well on small datasets is found in [26], interpreting a CNN as an ensemble of sub-networks with small correlation, being robust to overfitting. Considering that seismic data have an intrinsic textural characteristic [4] and that DCNNs present bias over textural information [27], a small CNN may emphasize such characteristics and minimize overfitting issues. In [23] and [28] is showed the possibility to reduce drastically the amount of representative information for seismic images. In [23], a suitable number of features provided by the SN was selected to feed a classifier, reducing its number from 2×10^6 features to 256. While in [28], the inner flux of a CNN made this reduction possible. So exploring the downsampling rate and filter hyperparameters could allow to reduce the dimensionality of the feature maps and at the same time to reduce the size of the network while keeping a good performance.

The choice of an architecture follows the "general belief that for flexible learning methods with finite samples, the best prediction performance is provided by a model of optimum complexity" [29]. We want to keep the network as shallow and simple as possible. We choose to build our network starting with LeNet [30] since seismic and handwriting data show a simpler aspect when compared with other benchmark datasets, at least visually. However, even a small network as LeNet with only 2 convolutional layers (CL) and 3 fully connected layers (FCL) suffers from overfitting and instabilities. To deal with such situation, in a first step our intention is to explore the redundancy of the seismic images. Even if there are approaches based on the compressed learning framework to find a sparse representation of the data for classification [31], here the intention is to reduce the dimensionality of the feature

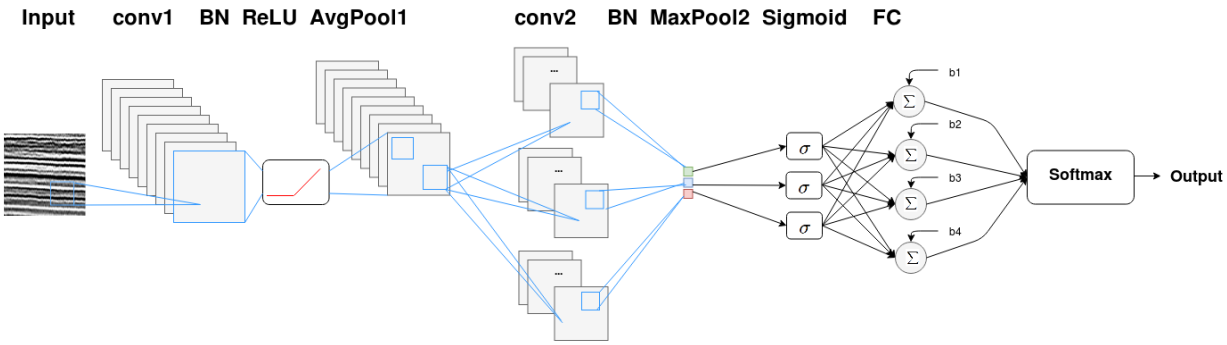

Fig. 3: Complete diagram of the reduced CNN.

maps by tuning the hyperparameters of the network. A more aggressive downsampling rate is considered. The stride value for the convolutional operators can be explored in the same way. Stride controls how the filters convolve over the input image. When considering a canonical 2-D convolution, the stride value is equal to one, meaning that it shifts at a pace of one pixel at a time. The choice for bigger filter sizes is based on the suitability with the dimensions of the images. Bigger values for these hyperparameters allow to reduce drastically the dimensionality of the feature maps from a layer to another, making possible to reduce the number of layers. In the first layer the number of filters was chosen empirically, considering the smallest number that did not affect the performance. For the last layer, a small number of filters was selected in order to have a more compact representation and the possibility to visualize how the clusters are spatially organized. While keeping good performance, reducing the original architecture end up with a 2 CLs and 1 FCL network. Average pooling was preferred because it makes simpler the analysis of the architecture when compared with max pooling [32]. However to have a good performance, it is still necessary to have a max pooling layer after the second CL. The sigmoid operator allows a more stable convergence.

Its flux (Figure 3): **1st CL**: 9 filters with size 6×6 and stride equal to 1×1; **batch normalization (BN)**: *momentum* = 0.9 (a convex combination between the older batches (0.9) and the current one (0.1)); **ReLU**; **average pooling (AvgPool1)**: 6×6; **2nd CL**: 3 filters with size 9×9 and stride equal to 2×2; **batch normalization (BN)**: *momentum* = 0.9; **max pooling (MaxPool2)**: 4×4; **sigmoid operator**; **FCL**: 4 nodes (with bias); **Output**: Log-Softmax. The resulting number of parameters is equal to $2,539$. The order of the layers are not the same as commonly explored in the literature, but empirically this choice leads to a better performance when using average pooling in the middle of the network.

IV. Results

The proposed CNN was tested with different ratios for the training set, the images being chosen randomly. For each ratio 50 simulations are performed. The hyperparameters values were selected so as to maximize performance: the batch size is equal to 20, the optimizer is the Nesterov Gradient method with learning rate equal to 0.007, *momentum* = 0.9, and the cost function is the log-likelihood.

The LANDMASS is a freely available database [33]. It contains 4 classes: chaotic-horizons, faults, horizons and salt-domes corresponding to 5.140, 1.251, 9.384 and 1.891 8-bits grayscale 99×99 images, respectively. For testing with a small dataset, 300 images from each class were chosen randomly.

A. Initialization

After the sigmoid operator, the output is embedded in the cube $[0,1]^3$, making possible its visualization. After convergence, the features are located close to the vertices $(1,0,1)$, $(0,1,1)$, $(1,1,0)$ and $(0,0,0)$ (but not always the same class at the same position what is dependend of the initialization).

Thus it is possible to initialize the FCL in order to have an affine map from the aforementioned vertices to the expected one-hot vector for each class in the CNN output. In this way, it is possible to achieve a more stable and faster convergence. The CLs are initialized randomly with a uniform distribution [34].

B. Evaluation

We perform comparisons with LeNet, Support Vector Machine (SVM), and k-nearest neighbors (knn) that were observed to outperform other algorithms such as random forest, binary trees and knn for other configurations. For LeNet the parameters that provide the best performance are: batch size = 20; learning rate = 0.05 and 0.1 for the equalized IFPEN data and for the reduced LANDMASS, respectively. The best configuration for SVM is a polynomial kernel with degree equal to 2 and a regularization parameter equal to 100. For the knn, we set k=1. The set of features used to feed the SVM and knn are the images histograms with 256 bins. This choice was based on the connection between the histograms and the different seismic patterns [35] and it was already explored in [5], [36]. The LeNet and SVM were using the same ratio between the training and validation sets as the small-scale CNN for each simulation. The results are shown in Tables I and II. For the CNN, we chose the number of epochs which, for a given simulation, provides the maximum accuracy for the training and test stages, the averages in the tables are computed for this epoch number.

TABLE I: IFPEN DATASET: Average accuracy of 50 simulations for each different ratios for the training set

	Average Accuracy (%)									
	Training					Validation				
Method	50%	60%	70%	80%	83%	50%	60%	70%	80%	83%
CNN	99.94	99.80	99.87	99.43	99.99	**92.64**	**90.49**	**92.96**	**81.56**	**96.9**
LeNet	85.52	86.01	85.96	82.79	80.75	84.1	84.06	82.12	79.76	75.88
SVM	**100**	**100**	**100**	**100**	**100**	73.3	73.3	73.9	74.1	74.3

TABLE II: LANDMASS DATASET: Average accuracy of 50 simulations for different percentages for the training set

		Average Accuracy (%)									
		Training					Validation				
Method		50%	60%	70%	80%	83%	50%	60%	70%	80%	83%
CNN		**99.88**	**99.88**	**99.9**	**99.96**	**99.95**	**98.46**	**94.41**	**94.92**	**98.43**	**99.27**
LeNet		71.15	83.29	87.44	85.48	87.04	70.98	81.37	87.91	86.33	83.76
KNN	k=1	-	-	-	-	-	87.2	88.3	89.1	89.3	89.6

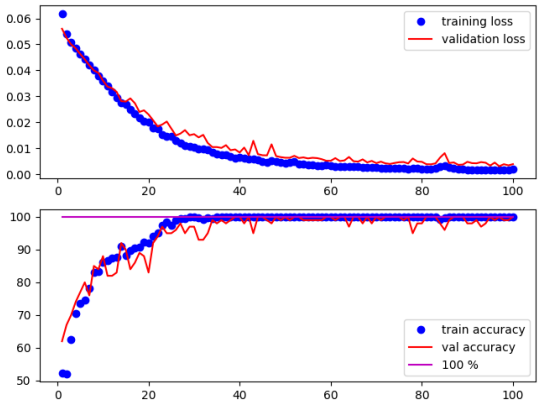

Fig. 4: One curve of the CNN's performance (IFPEN dataset): Above: training and validation cost curves for the reduced CNN; Below: training and validation accuracy curves.

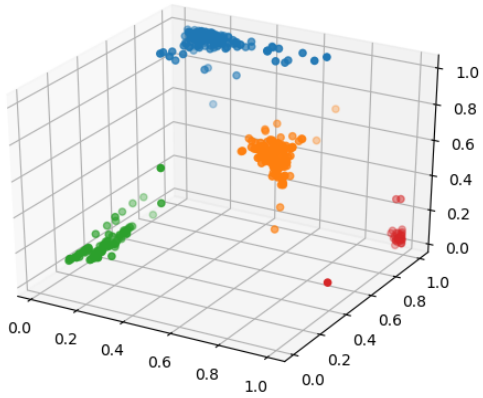

Fig. 5: Cluster after the sigmoid operator (IFPEN dataset).

For the IFPEN dataset, the small-scale CNN was observed to have a more unstable behavior. However, it is possible to obtain a smooth convergence as shown in Figure 4 but only with 83% of training data. It is possible to visualize the clusters for each class after the sigmoid operator in Figure 5, which shows that they are completely separated and concentrated close to the aforementioned vertices. For the best results with $50, 60, 70$ and 80% respectively, $4, 2, 1$ and 1 images are missclassified. For the LANDMASS database the convergence is usually smooth with some isolated spikes.

As expected the choice of the dataset influences the behavior of each classifier, making necessary to choose other values of the parameters and a different configuration.

For LeNet, a poorer performance is evidenced, mainly for the IFPEN dataset, possibly because it presents more variability among samples and the classes are unbalanced. Such behavior is expected when dealing with small datasets, even for networks much smaller than the state-of-the art architectures.

V. Conclusions and Perspectives

In this paper we presented a small-scale network that allows to classify two seismic databases with a small number of samples. A possible explanation for such a good behavior may be related to the complexity of the database. If LeNet provides a good performance for MNIST, it does not perform as well for CIFAR-10/100 [37], since MNIST is probably less complex. The same reason may apply to seismic images. However, it is difficult to argue in these terms since there is no universal technique to quantify the complexity of datasets, which may highly depend on the application [38]. While there exist approaches to measure inter-dataset similarity, these developments are recent and they were not tested with our dataset [39]. Another point to be taken into consideration is the instability that can arise with CNNs, in accordance with the discussion in [40] about the difficult of training a small neural network when compared with bigger ones.

Even if a complete theory for DCNNs is not well developed, it is known that a CNN approximates a function with few examples only if this function have some regularity properties [41]. A small network with a simpler information flow could provide a more qualitative interpretation with interest for a seismic interpreter [6] and since the mathematical analysis for CNNs is simplified for a two-layers network [42], [43], there is the possibility to carry out an analysis to better understand the links existing between its structure and the data of interest [44].

REFERENCES

[1] A. E. Barnes, "Seismic attributes in your facies," *CSEG recorder*, vol. 26, no. 7, pp. 41–47, 2001.

[2] ——, "Redundant and useless seismic attributes," *Geophysics*, vol. 72, no. 3, pp. P33–P38, 2007.

[3] D. Gao, "Application of three-dimensional seismic texture analysis with special reference to deep-marine facies discrimination and interpretation: Offshore angola, west africa," *AAPG bulletin*, vol. 91, no. 12, pp. 1665–1683, 2007.

[4] T. Hegazy, Z. Wang, and G. AlRegib, "The role of perceptual texture dissimilarity in automating seismic data interpretation," in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2015, pp. 118–122.

[5] Z. Long, Y. Alaudah, M. A. Qureshi, Y. Hu, Z. Wang, M. Alfarraj, G. AlRegib, A. Amin, M. Deriche, S. Al-Dharrab *et al.*, "A comparative study of texture attributes for characterizing subsurface structures in seismic volumes," *Interpretation*, vol. 6, no. 4, pp. T1055–T1066, 2018.

[6] F. Aminzadeh and P. deGroot, "A neural networks based seismic object detection technique," in *Society of Exploration Geophysicists Technical Program Expanded Abstracts 2005*. Society of Exploration Geophysicists, 2005, pp. 775–778.

[7] H. Zheng, Z. Zhang, and E. Liu, "Non-linear seismic wave propagation in anisotropic media using the flux-corrected transport technique," *Geophysical Journal International*, vol. 165, no. 3, pp. 943–956, 2006.

[8] J. Fan, C. Ma, and Y. Zhong, "A selective overview of deep learning," *arXiv preprint arXiv:1904.05526*, 2019.

[9] K. Kislov and V. Gravirov, "Deep artificial neural networks as a tool for the analysis of seismic data," *Seismic Instruments*, vol. 54, no. 1, pp. 8–16, 2018.

[10] G. AlRegib, M. Deriche, Z. Long, H. Di, Z. Wang, Y. Alaudah, M. A. Shafiq, and M. Alfarraj, "Subsurface structure analysis using computational interpretation and learning: A visual signal processing perspective," *IEEE Signal Process. Mag.*, vol. 35, no. 2, pp. 82–98, 2018.

[11] F. Berton and F. F. Vesely, "Seismic expression of depositional elements associated with a strongly progradational shelf margin: Northern santos basin, southeastern brazil," *Brazilian Journal of Geology*, vol. 46, no. 4, pp. 585–603, 2016.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image society of exploration geophysicistsmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[13] X. Wu, L. Liang, Y. Shi, and S. Fomel, "Faultsociety of exploration geophysicists3d: Using synthetic data sets to train an end-to-end convolutional neural network for 3d seismic fault society of exploration geophysicistsmentation," *Geophysics*, vol. 84, no. 3, pp. IM35–IM45, 2019.

[14] Y. Shi, X. Wu, and S. Fomel, "Saltsociety of exploration geophysicists: Automatic 3d salt society of exploration geophysicistsmentation using a deep convolutional neural network," *Interpretation*, vol. 7, no. 3, pp. SE113–SE122, 2019.

[15] B. Peters, E. Haber, and J. Granek, "Does shallow geological knowledge help neural-networks to predict deep units?" in *Society of Exploration Geophysicists Technical Program Expanded Abstracts 2019*. Society of Exploration Geophysicists, 2019, pp. 2268–2272.

[16] H. Di, D. Gao, and G. AlRegib, "Developing a seismic texture analysis neural network for machine-aided seismic pattern recognition and classification," *Geophysical Journal International*, vol. 218, no. 2, pp. 1262–1275, 2019.

[17] T. Wrona, I. Pan, R. L. Gawthorpe, and H. Fossen, "Seismic facies analysis using machine learning," *Geophysics*, vol. 83, no. 5, pp. O83–O95, 2018.

[18] J. Halotel, V. Demyanov, and A. Gardiner, "Value of geologically derived features in machine learning facies classification," *Mathematical Geosciences*, pp. 1–25, 2019.

[19] A. J. Bugge, J. Lie, and S. Clark, "Automatic facies classification and horizon tracking in 3d seismic data," in *First EAGE/PESGB Workshop Machine Learning*, vol. 2018, no. 1. European Association of Geoscientists & Engineers, 2018, pp. 1–3.

[20] D. S. Chevitarese, D. Szwarcman, E. V. Brazil, and B. Zadrozny, "Efficient classification of seismic textures," in *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2018, pp. 1–8.

[21] T. Zhao, "Seismic facies classification using different deep convolutional neural networks," in *Society of Exploration Geophysicists Technical Program Expanded Abstracts 2018*. Society of Exploration Geophysicists, 2018, pp. 2046–2050.

[22] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *2018 international interdisciplinary PhD workshop (IIPhDW)*. IEEE, 2018, pp. 117–122.

[23] Y. Bhalgat, J. Charlety, and L. Duval, "Catseyes: categorizing seismic structures with tessellated scattering wavelet networks," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 1618–1622.

[24] J. Bruna and S. Mallat, "Invariant scattering convolution networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1872–1886, 2013.

[25] L. Sifre and S. Mallat, "Rotation, scaling and deformation invariant scattering for texture discrimination," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 1233–1240.

[26] M. Olson, A. Wyner, and R. Berk, "Modern neural networks generalize on small data sets," in *Advances in Neural Information Processing Systems*, 2018, pp. 3619–3628.

[27] R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F. A. Wichmann, and W. Brendel, "Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness," *arXiv preprint arXiv:1811.12231*, 2018.

[28] A. Veillard, O. Morére, M. Grout, and J. Gruffeille, "Fast 3d seismic interpretation with unsupervised deep learning: Application to a potash network in the north sea," in *80th EAGE Conference and Exhibition 2018*, 2018.

[29] V. Cherkassky and F. M. Mulier, *Learning from data: concepts, theory, and methods*. John Wiley & Sons, 2007.

[30] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner *et al.*, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[31] A. Adler, M. Elad, and M. Zibulevsky, "Compressed learning: A deep neural network approach," *arXiv preprint arXiv:1610.09615*, 2016.

[32] P. L. Combettes and J.-C. Pesquet, "Deep neural network structures solving variational inequalities," *Set-Valued and Variational Analysis*, pp. 1–28, 2020.

[33] CeGP, "Landmass: large north-sea dataset of migrated aggregated seismic structures," 2015.

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

[35] T. M. Sheffield and B. A. Payne, "Geovolume visualization and interpretation: What makes a useful visualization seismic attribute?" in *Society of Exploration Geophysicists Technical Program Expanded Abstracts 2008*. Society of Exploration Geophysicists, 2008, pp. 849–853.

[36] J. Qi, T. Lin, T. Zhao, F. Li, and K. Marfurt, "Semisupervised multiattribute seismic facies analysis," *Interpretation*, vol. 4, no. 1, pp. SB91–SB106, 2016.

[37] F. Cotter and N. Kingsbury, "Deep learning in the wavelet domain," *arXiv preprint arXiv:1811.06115*, 2018.

[38] S. Singh and D. Shukla, "A review on various measures for finding image complexity," *International Journal of Scientific Research Engineering & Technology. ISSN*, pp. 2278–0882, 2017.

[39] I. Hwang, J. Lee, F. Liu, and M. Cho, "Simex: Express prediction of inter-dataset similarity by a fleet of autoencoders," *arXiv preprint arXiv:2001.04893v1*, 2018.

[40] S. Lawrence, C. L. Giles, and A. C. Tsoi, "What size neural network gives optimal generalization? convergence properties of backpropagation," Tech. Rep., 1998.

[41] S. Mallat, "Understanding deep convolutional networks," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150203, 2016.

[42] Y. Tian, "Symmetry-breaking convergence analysis of certain two-layered neural networks with relu nonlinearity," 2016.

[43] Y. Li and Y. Yuan, "Convergence analysis of two-layer neural networks with relu activation," in *Advances in neural information processing systems*, 2017, pp. 597–607.

[44] L. Pinto, D. Gopalan *et al.*, "Limiting network size within finite bounds for optimization," *arXiv preprint arXiv:1903.02809*, 2019.