CATSEYES: CATEGORIZING SEISMIC STRUCTURES WITH TESSELLATED SCATTERING WAVELET NETWORKS

Yash Bhalgat[1,2], *Jean Charléty*[1], *Laurent Duval*[1]

[1] IFP Energies nouvelles
92852 Rueil-Malmaison, France
{firstname.lastname}@ifpen.fr

[2] Department of Electrical Engineering, IIT Bombay
Mumbai, Maharashtra 400076, India
yashsb@umich.edu (now with University of Michigan)

ABSTRACT

As field seismic data sizes are dramatically increasing toward exabytes, automating the labeling of "structural monads" — corresponding to geological patterns and yielding subsurface interpretation — in a huge amount of available information would drastically reduce interpretation time. Since customary designed features may not account for gradual deformations observable in seismic data, we propose to adapt the wavelet-based scattering network methodology with a tessellation of geophysical images. Its invariances are expected to be able to thwart the effect of the tectonics. The sparse structure of extracted feature vectors suggest to resort to dimension reduction methods before classification. The most promising one is based on a tessellated version of scattering decompositions, combined with a standard affine PCA classifier. Extensive comparative results on a four-class seismic database show the effectiveness of the proposed method in terms of seismic data labeling and object retrieval, in affordable computational time.

Index Terms— Classification, Scattering networks, Seismic processing, Machine learning, Dimension reduction

1. INTRODUCTION

Seismic data analysis has long been a topic of choice for advanced signal and image processing [1]. The inherent complexity and size of geophysical data has for instance nurtured widely-used techniques, from wavelet transforms [2, 3, 4, 5] to sparse deconvolution [6]. To produce sound images of underground structures, geophysics strive to acquire, model and process huge sets of seismic traces, ending up in stacked or migrated datasets (Fig. 1). The latter represent (distorted because indirect) geological formations in the shape of various seismic patterns, readable by geophysicists, within the wiggling bandpass nature of seismic signals. They for instance reveal depositional sequences, structural bodies (like salt domes [7]), geological incidents (like faults and fractures [8]), or poorly imaged zones. Their analysis is of primary importance to understand the tectonic and sedimentary history of regions, and their potential in finding hydrocarbon traps. Such an analysis can be seen as dimension reduction from highly redundant seismic data [9], albeit preserving essential geological features [10]. Automating the labeling of so-called rock types [11], seismic facies, associated lithofacies or structural elements (hereafter called *structural monads*) would drastically reduce interpretation time and cost. The term *structural monad* (abbreviated as SM) is proposed to singularize units with putative geophysical or geological interpretation, resistant to a family of admissible deformations, such as shift, shear or scaling for seismic data. However, the huge variability in seismic acquisition/processing and geological mechanisms, together with the notable data volume, hinder an accurate human-based expert interpretation.

In Section 2 we review related endeavors on seismic structure classification. With a shortage of widely-shared tagged seismic databases and purpose, we detail our evaluation methodology on a 2D seismic database structuration in Section 3. In Section 4, we motivate a novel approach based on scattering wavelet networks. Reminders are first provided, with an analysis of the sparse structure of feature vectors. A benchmark of different dimension reduction algorithms is then performed. A tessellated version of scattering transforms is then provided. Its performance is evaluated with different training/testing proportions, and sensitive improvements are obtained in terms of accuracy, at an affordable cost (Section 5). Conclusions and perspectives are finally drawn.

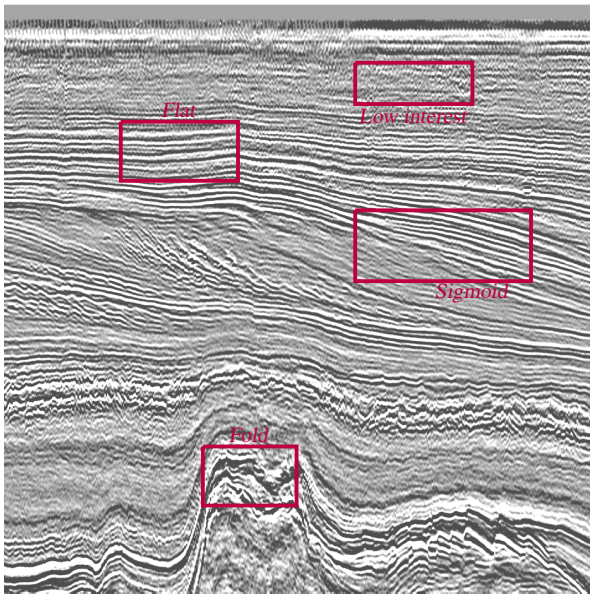


Fig. 1: Migrated seismic sections, with exemplars of four instances of structural monads: *Flat*, *Sigmoid*, *Fold*, *Low interest*.

2. PRIOR WORK

As exemplified in Figure 1, the nature of seismic images is mostly textural, with fuzzily-defined regions, and limited edge-based delineation. Hence, their characterization has been pursued for decades with various combinations of features (often termed attributes in the geophysics community) and classifiers. Interestingly, seismic has soon resorted to the computation of complex attributes [12] derived

from the Hilbert transform. Instantaneous frequency, phase and envelope have long been favored attributes for data labeling and classification. Since then, many works have aimed at simplifying and classifying seismic data, through a wealth of signal/image processing techniques: compression, normalization and morphological operations [9], different breeds of directional filter banks [13, 14] or multi-scale feature extraction or wavelets [15]. Due to the complexity of feature combination and learning, recent studies focus on specific SMs (structural monads) such as salt domes, with a combination of gray-level co-occurrence matrices and Gabor filters in [16], gradients of textures in [7] or curvelet coefficients in [17]. Even techniques from speech recognition (like mel-frequency cepstrum coefficients or MFCC) have been used [18].

When a predetermined discrete collection of tailored features is used, their resistance to different sampling patterns, wave propagation or seismic processing deformations is usually not evident. The appropriateness of an union of such feature sets for different SMs is not fully guaranteed. In other words, no current seismic classification scheme is available to cover all kinds of features. We chose to address the categorization [19, p. 13 sq.] of different kinds of SM, in the line of [17]. Instead of relying on custom features, we chose a more attribute-agnostic framework, named Scattering Networks or Scattering Transform. It stems from nearly-invariant complex wavelet representations, and aims at providing a mathematical rationale behind the Convolutional Neural Networks (CNN) instance of deep learning [20]. It roots on frames of oriented complex/Hilbert pair [21] of wavelet transform levels, intertwined with non-linear operators mimicking pooling operators. The proposed **CatsEyes** framework alludes to the **Cat**egorization of **Seis**mic monads, with sector-wise scattering coefficients (Figures 2(b-c-f-g-j-k-n-o)) reminiscent of dark-glowing cat retina, colored-insert glass marbles. Scattering transforms have been used scarcely in geophysics so far [22, 23], on types of geophysical data different from the ones addressed here. Inspired by morphological and textural analogies between fingerprints and seismic data, exploited for instance in denoising with wave atoms [24], our proposal builds upon recent results on fingerprint classification [25].

3. SEISMIC DATABASE EVALUATION METHODOLOGY

3.1. Seismic data sources

Unlike the mainstream image field, due to the shortage of widely-shared tagged seismic image databases to design and evaluate classification algorithms, a first priority resides in building a database which can also serve as a reference for testing other algorithms. The data come from subparts of 2D seismic sections that were acquired in different tectonic zones at the border of the American continent: from the New-Jersey to the Amazonian mouth. From four wide sections, subparts were extracted to feed the database with four categories named: *Flat*, *Fold*, *Sigmoid* and the rest, denoted by *Low interest*. *Sigmoid* may refer to geological sequences or strata called toplap, offlap, onlap or downlap. The last *Low interest* class encompasses natural noise present in data, seismic processing artifacts (migration) and spurious signal that come before the sea bottom and thus is not considered since it does not carry geological information. These 38 subparts were then cropped to a common size of 1178×1932 pixels.

Such a database has a very limited number of exemplar images for providing sufficient training data for most algorithms. Hence, templates do not span all intra-class variations, and the base is not representative enough. A more consistent database is thus required.

3.2. Database structuration and curation

Resultingly, the following steps were followed in the creation of the evaluation database:

- **Cropping:** an extended set of images of size 512×512 is created by cropping overlapping patches from the original 38 images. For the underrepresented *Sigmoid* class, some Regions of Interest were manually selected. At the end of this step, 1300 exemplars were obtained, with a distribution of 400, 440, 360, 100 images in the *Flat*, *Fold*, *Low interest*, *Sigmoid* classes, respectively.
- **Curation:** when cropped to a smaller size, sometimes features for a specific class are lost. A handful of unfit exemplars are manually removed.
- **Repetition removal:** some images from the cropped database are almost replicas. To reduce the bias toward redundant templates, we choose to automatically remove images with Structural Similarity Index (SSIM, [26]) above 0.9, resulting in a better coverage among all classes.

Finally, we obtain a high-quality database with a non-uniform distribution of annotated templates as 221, 223, 100 and 36 images in *Flat*, *Fold*, *Low interest*, *Sigmoid* classes respectively. An exemplar for each template is depicted in Fig. 2(a-e-i-m).

4. TESSELLATED SCATTERING TRANSFORM

4.1. Motivations

Scattering wavelet networks (ScatNets or ScatterNets) have been introduced by S. Mallat [27] and his co-authors. They are being studied and expanded in several paths [28, 29]. They offer a solid framework for a mathematical formulation aiming at deep network understanding. Using cascades of complex wavelet transform frames and non-linear modulus operators to build feature vectors, they yield data representations stable to controlled deformations. Features obtained with ScatNets show promising results for texture classification [30, 31], inverse problems [32] or chemical structures [33]. Along with local translation and small-deformation invariance, energy preservation and higher-order information retention, ScatNets have predetermined weights and structure, unlike convolution neural networks [34]. Resultingly, almost no training is required in learning and setting network's hyper-parameters (layers, filters) for a specific application, often requiring significant expertise and a handful of experimentation.

4.2. Reminders on the ScatNet architecture

Let $I(x, y)$, $(x, y) \in \mathbb{R}^2$, represent a two-dimensional image. Let $\phi_J(x, y) = 2^{-2J}\phi(2^{-J}(x, y))$ denote a low-pass filter controlled by the scaling factor J. If ψ is a mother wavelet, the family of its rotated and dilated versions is denoted by $\{\psi_\lambda\}$, with the multi-parameter $\lambda = (\theta, j) \in \{\Lambda = \Theta \times [1, 2, \ldots, J]\}$, where θ indexes orientations and 2^j are dyadic scales.

For the wavelet family $\{\psi_\lambda\}_{\lambda\in\Lambda}$, wavelet coefficients obtained through $\{I(x, y) * \psi_\lambda\}_{\lambda\in\Lambda}$ mostly retain the high-frequency variations of the signal. A coarse approximation is obtained by a smoothing with the low-pass filter ϕ_J, resulting in the compound wavelet feature vector (FV):

$$|\mathcal{W}I(x, y)| = \{I(x, y) * \phi_J, |I(x, y) * \psi_\lambda|\}_{\lambda\in\Lambda}. \qquad (1)$$

In (1) however, wavelet coefficients preserve translation invariance only up to 2^j, and not up to the maximum scale 2^J reached by the

the first component (the average, stable to deformation [27]). This can be resolved by resorting to a modulus operator whose output is convolved with ϕ_J. Now $S_0 I(x,y) = I(x,y) * \phi_J$ represents the zeroth layer of the scattering network. The non-linear part is also averaged to get first-order scattering coefficients:

$$S_1 I(x,y,\lambda_1) = |I(x,y) * \psi_{\lambda_1}| * \phi_J \,. \tag{2}$$

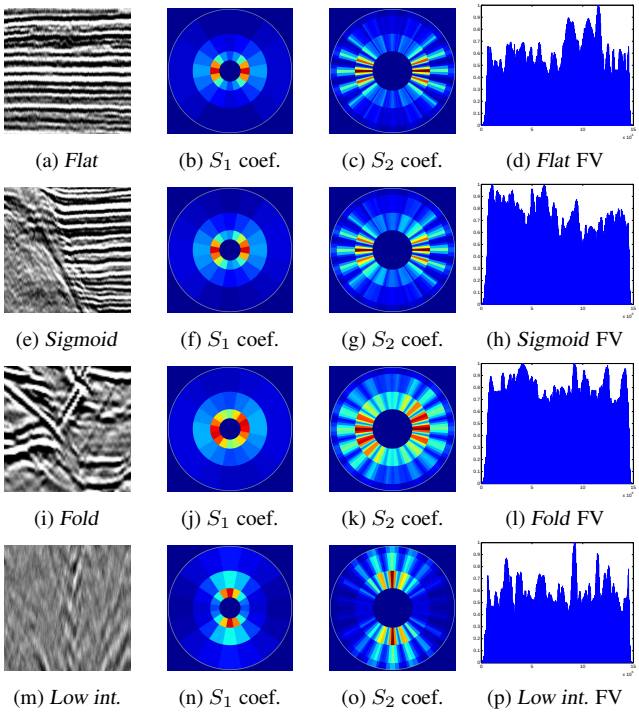
Similarly, second-order scattering coefficients are obtained as:

$$S_2 I(x,y,\lambda_1,\lambda_2) = ||I(x,y) * \psi_{\lambda_1}| * \psi_{\lambda_2}| * \phi_J \,.$$

Further, higher-order scattering coefficients can be computed recursively with:

$$S_m I(x,y,\lambda_1,\ldots,\lambda_m) = |||I(x,y) * \psi_{\lambda_1}| \ldots | \cdots * \psi_{\lambda_m}| * \phi_J \,.$$

As in [30], inter-class differences can be observed through concentric diagrams of scattering coefficients. Figures 2(b-f-j-n) and 2(c-g-k-o), respectively, result from computing $S_1 I(x,y,\lambda_1)$ and $S_2 I(x,y,\lambda_1,\lambda_2)$ with the left-hand images. Inside the rings, each angular sector represents one (θ, j) combination and is filled with the average value obtained from the cascaded convolution. Such multi-level pie charts, computed over entire images, illustrate the classification power of scattering networks, as well as potential overlaps between sought classes. The latter observation motivates a tessellated[1] scattering network approach described hereafter.



(a) *Flat* (b) S_1 coef. (c) S_2 coef. (d) *Flat* FV

(e) *Sigmoid* (f) S_1 coef. (g) S_2 coef. (h) *Sigmoid* FV

(i) *Fold* (j) S_1 coef. (k) S_2 coef. (l) *Fold* FV

(m) *Low int.* (n) S_1 coef. (o) S_2 coef. (p) *Low int.* FV

Fig. 2: Left: exemplars for each class. Center: multi-level, angular sector representation of a two-level scattering transform. Right: flattened feature vectors.

[1]From Latin *tessella*, small squares.

4.3. Feature vector representation and dimension reduction

The most natural manner to construct FVs (feature vectors) — to feed classifiers — reduces to using all scattering coefficients. Such a flattening vectorizes the entirety of the S_m scattering layers. As they result from many couples of orientations and scales, or convolutions with an oversampled filter bank, their size can become extremely large. Exemplar FVs depicted in Figures 2(d-h-l-p) have a total length of about 15×10^6 coefficients. Even if a down-sampling can reduce this length to 7×10^4 coefficient, the high-dimension of the feature space may adversely affects the classification accuracy, due to the modest size of the database (584 images). However, we observe that the FVs are highly compressible. Their normalized and sorted values for the different classes are represented in Fig. 3. As ScatNets are energy preserving, the steep decay in the upper-right indicates that most of the information is carried by very few important coefficients. Slightly differing decay [35] regimes can be observed across the different classes. This can be emphasized with the close-up graph in gray, within the highest $0.1\,\%$ of the coefficients in the steepest portion of the curve.

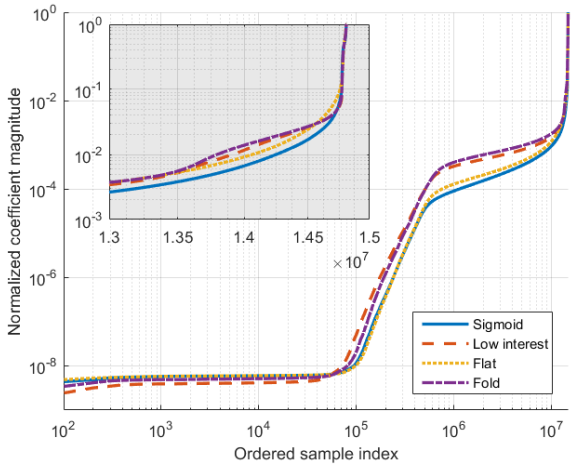


Fig. 3: Compressibility of ordered scattering transform feature vectors from Fig. 2(d-h-l-p) in log-log-scale.

To exploit redundancy in these highly sparse vectors, we first benchmark various feature selection methods [36] collected at Arizona State University[2]. The feature selection methods applied to our database are based on: Gini index [37], χ^2 statistics [38], Correlation-based Feature Selection (CFS) [39], Kruskal-Wallis (KW) [40], Minimum Redundancy and Maximum Relevance (mRMR) [41], sparse multinomial logistic regression algorithm with Bayesian regularisation (SBMLR) [42] and Fisher score [43]. Reduced feature sizes (#RFS) are globally shrunk again by an order of magnitude, toward thousands of coefficients, as indicated in Table 1-(column 2).

4.4. Tessellated scattering networks

Several additional ingredients can be melt into scattering networks for performance improvement. While excellent results are obtained on textures [30, 31], adaptations to seismic features could provide enhanced results. For instance, Figure 1 indicates that gradual

[2]`featureselection.asu.edu/old/software.php`

changes could be observed between structural monads: a *Sigmoid* or *Fold* behavior may slowly warp to a *Flat* morphology. Due to data acquisition and processing, regions of *Low interest* neighbor other regions with higher geological value.

Similarly, the *Sigmoid* class can be viewed as a combination of two classes: *Fold* and *Flat*, as can be seen from Fig. 2(a-e-i). Important classification features for "overlapped" classes can thus be located in different regions of an image. We thus extract and combine FVs from a tessellation of subparts of initial images. This operation can be thought as a diversity enhancement to account for gradual morphing between structural monads. Such an approach has also been successfully applied recently to fingerprints [25], whose oriented textures combined with localized information reveal important in classification tasks.

In this work, images (of size 512×512) are divided into $4 \times 4 = 16$ non-overlapping blocks. Scattering wavelet transform coefficient are extracted from each block. Computing the mean of each of the convolutions was shown [27, 31] to correspond to the energy of the convolutions. The complete "tessellated" ScatNet feature vector is formed by concatenating FVs of all the 16 blocks, formed by the pooling/averaging operations over the global scattering transform.

4.5. Scattering network specifications and category assignment

To evaluate the relative performance of the different dimension reduction methods, we select $|\Theta| = 8$ orientations and $J = 3$ scales, using a geophysics-inspired Morlet wavelet filter bank [2, 3]. A scattering transform consisting of three layers ($m = 2$) is either applied on the whole images or their tessellated blocks. The size of an FV for each block after the pooling step, as described above, is given by $\sum_{i=0}^{m} L^i \binom{J}{i}$. Since the final FV is obtained by concatenating all these partial FVs, the size of a complete FV is given by $16 \times \sum_{i=0}^{m} L^i \binom{J}{i}$. These FVs are finally passed to a generative Principal Component Analysis (PCA) affine classifier from [31].

5. EXPERIMENTS AND RESULTS

Reduced feature size, classification accuracy and different training times, with various methods and training/testing ratios are reported in Table 1. Numbers were obtained from averages of 50 randomized trials from the seismic database. Standard deviations for classification accuracy spanned a $0.5\,\%$ to $1.5\,\%$ range. Among the standard dimension reduction methods detailed in [36], the Fisher score achieved the best overall classification with $81\,\%$ to $82\,\%$ and the smallest number of reduced features. However, for higher accuracies, we notice that the correct class retrieval saturates, in terms of statistical significance, within the randomized variations. However (bottom lines of Table 1), pure ScatNets exhibit a $5\,\%$ to $6\,\%$ improvement over Fisher score. Their tessellated heir is even augmented by $3\,\%$ to $6\,\%$. Interestingly, the 16-fold higher dimension of the tessellated feature size (3456) and times required for feature extraction and training (1814 *vs* 2214 and 54 *vs* 341 seconds respectively) are compensated by a faster feature classification ($0.47\,\mathrm{s}$ *vs* $0.14\,\mathrm{s}$). This observation could be related to the ability of ScatNets to jointly sparsify and linearize the original feature space. The overall $10\,\%$ to $11\,\%$ improvement in classification accuracy of the proposed tessellated scattering network over the Fisher score remains practically useful, as #RFS (reduced feature size), feature extraction and training times are comparable, with a six-fold faster final classification speed-up.

A typical confusion matrix is provided in Table 2. For instance, for $50\,\%$ training, from a total of 221 *Flat* exemplars, 111 test images are retained. Among them, 103 are correctly assigned, while 6 and 2 are assigned to *Fold* and *Sigmoid* classes respectively. The table confirms the accurate assignment of the tessellated scattering network to *Flat* and *Fold* structural monads, as well as a good classification of *Low interest* zones. As expected, *Sigmoid* structures can be wrongfully assigned to *Flat* or *Fold* more frequently.

Training		30%	50%	70%	Time (s.) with 50 % training		
Method	#RFS	Accuracy (%)			Feat. ext.	Train.	Class./FV
Gini	8725	62.4	64.9	65.2	7113	968	0.13
χ^2	7167	57.0	61.2	61.7	5623	797	0.11
CFS	5133	69.0	69.5	70.4	4784	557	0.17
KW	3133	69.1	71.6	71.8	2126	289	0.10
mRMR	4607	75.2	76.4	77.9	5275	782	0.31
SBMLR	3265	73.8	75.1	76.1	8982	1044	0.22
Fisher	2819	80.7	81.3	82.5	1931	376	0.87
ScatNet	216	86.6	87.1	87.6	1814	54	0.47
Tessellated	3456	90.9	91.6	93.5	2214	341	0.14

Table 1: Computational results. Left: comparisons for different dimension reduction/feature extraction methods with varying training percentages (methods referenced in Sec. 4.3-4.4) in #RFS (Reduced feature size) and accuracy/c. Right: times for feature extraction and selection, training, and PCA classification per feature vector.

	Flat	*Fold*	*Sigmoid*	*Low int.*
Flat	103	6	2	0
Fold	5	102	3	2
Sigmoid	1	3	14	0
Low int.	1	2	0	47

Table 2: Confusion matrix for 50 % training. Horizontal: true class; vertical: assigned class.

6. CONCLUSION AND PERSPECTIVES

We present an novel categorization of seismic patterns, defined as *structural monads* with a more generic acceptation. To account for geological and geophysical processes, we adopt scattering wavelet networks as deformation and translation invariant feature joint extractors and classifiers. A database with tagged *structural monads* is devised, drawn on public data, which could be shared for other publishable studies. An extensive comparison of feature vector dimension reduction methods for classification is performed, and the proposed tessellated scattering decomposition is shown the effective.

Addressing the robustness of the proposed methodology to seismic processing variability and noise, and to other related classifications such as rock types [11] is an immediate perspective. Improving results with a better consideration of smooth tectonic deformations, and visualizing the most important features [44] are trusted to further extend our evaluation database to controlled warping of original images.

7. ACKNOWLEDGMENTS

L. Duval would like to thank S. Mallat (ENS Ulm) for precious advice on classification and J.-M. Mengus (IFPEN) for seismic data simulation.

8. REFERENCES

[1] L. C. Wood and S. Treitel, "Seismic signal processing," *Proc. IEEE*, vol. 63, no. 4, pp. 649–661, Apr. 1975.

[2] J. Morlet, G. Arens, E. Fourgeau, and D. Giard, "Wave propagation and sampling theory—part I: Complex signal and scattering in multilayered media," *Geophysics*, vol. 47, no. 2, pp. 203–221, 1982.

[3] J. Morlet, G. Arens, E. Fourgeau, and D. Giard, "Wave propagation and sampling theory—part II: Sampling theory and complex waves," *Geophysics*, vol. 47, no. 2, pp. 222–236, 1982.

[4] R. Klees and R. Haagmans, Eds., *Wavelets in the geosciences*, Number 90 in Lect. Notes Earth Sci. Springer, 2000.

[5] L. Jacques, L. Duval, C. Chaux, and G. Peyré, "A panorama on multiscale geometric representations, intertwining spatial, directional and frequency selectivity," *Signal Process.*, vol. 91, no. 12, pp. 2699–2730, Dec. 2011.

[6] A. Repetti, M. Q. Pham, L. Duval, E. Chouzenoux, and J.-C. Pesquet, "Euclid in a taxicab: Sparse blind deconvolution with smoothed ℓ_1/ℓ_2 regularization," *IEEE Signal Process. Lett.*, vol. 22, no. 5, pp. 539–543, May 2015.

[7] T. Hegazy, Z. Wang, and G. AlRegib, "The role of perceptual texture dissimilarity in automating seismic data interpretation," in *Proc. IEEE Global Conf. Signal Information Process.*, Orlando, FL, USA, Dec. 14-16, 2015.

[8] L. Schmidt, C. Hegde, P. Indyk, J. Kane, L. Lu, and D. Hohl, "Automatic fault localization using the generalized Earth mover's distance," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, May 4-9, 2014, pp. 8134–8138.

[9] R. A. Strelitz and Y. Keshet, "Integral transforms, data compression, and automatic analysis of seismic sections," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 6, pp. 982–991, Nov. 1990.

[10] T. Coléou, M. Poupon, and K. Azbel, "Unsupervised seismic facies classification: A review and comparison of techniques and implementation," *The Leading Edge*, vol. 22, no. 10, pp. 942–953, Oct. 2003.

[11] L. Shua, K. McIsaac, G. R. Osinski, and R. Francis, "Unsupervised feature learning for autonomous rock image classification," *Comput. Geosci.*, 2017.

[12] M. T. Taner, F. Koehler, and R. E. Sheriff, "Complex seismic trace analysis," *Geophysics*, vol. 44, no. 6, pp. 1041–1063, Jun. 1979.

[13] T. Randen and J. H. Husøy, "Filtering for texture classification: a comparative study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 4, pp. 291–310, Apr. 1999.

[14] J. Gauthier, L. Duval, and J.-C. Pesquet, "Optimization of synthesis oversampled complex filter banks," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3827–3843, Oct. 2009.

[15] B. Jafarpour, "Wavelet reconstruction of geologic facies from nonlinear dynamic flow measurements," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1520–1535, May 2011.

[16] A. Amin and M. Deriche, "Salt-dome detection using a codebook-based learning model," *IEEE Geosci. Rem. Sens. Lett.*, vol. 13, no. 11, pp. 1636–1640, Nov. 2016.

[17] Y. Alaudah and G. AlRegib, "Weakly-supervised labeling of seismic volumes using reference exemplars," in *Proc. Int. Conf. Image Process.*, Phoenix, AZ, USA, Sep. 25-28, 2016, pp. 4373–2377.

[18] T. Xie, X. Zheng, and Y. Zhang, "Seismic facies analysis based on speech recognition feature parameters," *Geophysics*, vol. 82, no. 3, pp. O23–O35, May 2017.

[19] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning. From Theory to Algorithms*, Cambridge University Press, 2014.

[20] C. Trabelsi, O. Bilaniuk, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. J. Pal, "Deep complex networks," *PREPRINT*, 2017.

[21] C. Chaux, J.-C. Pesquet, and L. Duval, "Noise covariance properties in dual-tree wavelet decompositions," *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4680–4700, Dec. 2007.

[22] M. Glinsky, "Stratigraphic facies from the physics perspective of emergent phases of self organization," in *Proc. Geoflows conference*, 2013.

[23] B. Bougher and F. J. Herrmann, "Using the scattering transform to predict stratigraphic units from well logs," *CSEG Recorder*, pp. 22–25, Jan. 2016.

[24] L. Demanet and L. Ying, "Wave atoms and sparsity of oscillatory patterns," *Appl. Comput. Harmon. Analysis*, vol. 23, no. 3, pp. 368–387, 2007.

[25] P. Birajadar, Y. Bhalgat, U. Sharma, M. Haria, B. Singh, and V. Gadre, "A scattering wavelet network based approach to fingerprint classification," *Patt. Rec. Lett. (in review)*, Oct. 2017.

[26] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[27] S. Mallat, "Group invariant scattering," *Comm. Pure Appl. Math.*, vol. 65, no. 10, pp. 1331–1398, Oct. 2012.

[28] W. Czaja and W. Li, "Analysis of time-frequency scattering transforms," *Appl. Comput. Harmon. Analysis*, 2017.

[29] P. Grohs and T. Wiatowski, "Energy decay and conservation in deep convolutional neural networks," in *Proc. Int. Symp. Information Theory*, 2017.

[30] J. Bruna and S. Mallat, "Invariant scattering convolution networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1872–1886, Aug. 2013.

[31] L. Sifre and S. Mallat, "Rotation, scaling and deformation invariant scattering for texture discrimination," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, Portland, OR, USA, 2013.

[32] I. Dokmanić, J. Bruna, S. Mallat, and M. de Hoop, "Invariant multiscale statistics for inverse problems," in *Proc. Signal Processing with Adaptive Sparse Structured Representations (SPARS) workshop*, Lisbon, Portugal, Jun. 5-8, 2017.

[33] M. Hirn, S. Mallat, and N. Poilvert, "Wavelet scattering regression of quantum chemical energies," *PREPRINT*, 2017.

[34] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Ann. Conf. Neur. Inform. Proc. Syst.*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., pp. 1097–1105. 2012.

[35] I. Waldspurger, "Exponential decay of scattering coefficients," in *Proc. Int. Conf. Sampling Theory Appl.*, Tallin, Estonia, Jul. 3-7, 2017.

[36] Z. Zhao, F. Morstatter, S. Sharma, S. Alelyani, A. Anand, and H. Liu, "Advancing feature selection research — ASU feature selection repository," Tech. Rep., 2010.

[37] C. Gini, *Variabilità e mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche*, 1912.

[38] H. Liu and R. Setiono, "Chi2: feature selection and discretization of numeric attributes," in *Proc. 7th IEEE Int. Conf. Tools with Artificial Intelligence*, Herndon, VA, USA, Nov. 1995.

[39] M. A. Hall and L. A. Smith, "Feature selection for machine learning: Comparing a correlation-based filter approach to the wrapper," in *Proc. Int. Florida Artificial Intelligence Research Soc. Conf.*, Orlando, FL, USA, May 1-5, 1999.

[40] L. J. Wei, "Asymptotic conservativeness and efficiency of Kruskal-Wallis test for K dependent samples," *J. Am. Stat. Assoc.*, vol. 76, no. 376, pp. 1006–1009, Dec. 1981.

[41] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[42] G. C. Cawley, N. L. C. Talbot, and M. Girolami, "Sparse multinomial logistic regression via Bayesian L1 regularisation," in *Proc. Ann. Conf. Neur. Inform. Proc. Syst.*, Dec. 4-7, 2006, pp. 209–216.

[43] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, 2nd edition, 2001.

[44] F. Cotter and N. Kingsbury, "Visualizing and improving scattering networks," in *IEEE Int. Workshop Mach. Learn. Signal Process.*, Tokyo, Japan, Sep. 25-28, 2017.