Complex Neural Networks for Estimating Epicentral Distance, Depth, and Magnitude of Seismic Waves

Nicolae-Cătălin Ristea and Anamaria Radoi, *Member, IEEE*

***Abstract*— Taking advantage of the latest advances in deep learning for seismology, we address earthquake characterization from a data-driven perspective. Many of the usual procedures for extracting information from seismograms require processing a large volume of data using empirical and physics rule-based techniques. In this letter, we propose a novel approach for estimating epicentral distance, depth, and magnitude directly from individual raw three-component seismograms of 1-min length observed by single stations. Our convolutional neural network-based method is able to handle complex-valued representations of the seismic data in the time–frequency domain by using dedicated convolutional and activation functions. In this way, our method benefits both from extracting relevant information through time–frequency domain analysis and from designing a single architecture that deals with complex information. The proposed method achieves a mean absolute error of 4.51 km for epicentral distance, 6.15 km for depth, and 0.26 for magnitude estimation. The experiments were conducted over a publicly available and large database, STanford EArthquake data set (STEAD), and the comparisons with current state-of-the-art approaches show the effectiveness of the proposed approach. Source code and best model are available at https://github.com/ristea/stead-earthquake-cnn.**

***Index Terms*— Complex neural networks, convolutional neural networks (CNNs), seismic source characterization, seismic waves.**

I. Introduction

THE past few years have witnessed an increased interest toward the interpretation of seismic data by means of signal processing, machine learning, and, particularly, deep learning techniques, e.g., convolutional neural networks (CNNs) [1]–[3], recurrent neural networks [4], and graph neural networks [5]. Advances in monitoring volcano-seismic events [6], seismic data compression [7], seismic signal denoising [8], phase classification [9], seismic structure representation [10], or earthquake early warning systems [3] are just several areas that have shown promising results in this direction. One of the primary tasks in earthquake seismology is seismic source characterization. In case of emergency, the requirement for timely dissemination of information leads to providing accurate and rapid estimates of the seismic source characteristics, preferably without the intervention of an expert. However, the majority of the techniques dealing with seismic source characterization involve data acquired by multiple stations, for which nontrivial data fusion techniques and geometry of the seismic networks have to be specified or learned [5].

Lately, source location and characterization of seismic waves from single-station observations has gained attention [1], [2], [4], [11]. Starting from ConvNetQuake [1], the first CNN-based architecture used for earthquake detection and location from a single-station waveform, Lomax *et al.* [11] extended the model to characterize earthquakes over a broad range of distances and magnitudes by classifying the waveforms in a larger number of classes (i.e., discriminate between seismic event and noise, 50 distance, 20 magnitude, 20 depth, and 36 azimuth categories). The advantage of this approach is that it can be directly applied on the seismogram waveforms, without additional pre-processing or feature extraction, which leads to a decreased running time for the monitoring and analysis of seismic signals. However, the characterization errors are sometimes large due to possible overfitting of the CNN model.

A combined convolutional-recurrent neural network was employed to estimate the earthquake's magnitude directly from raw single-station waveforms [4]. The regression is performed using long short-term memory (LSTM)-based networks, whose input is connected to the output of a series of two cascaded CNNs with the role of performing automatic feature extraction and dimensionality reduction. Comparing with the case of using multiple stations (up to four) to estimate the magnitude of the earthquakes, the results shown in [4] demonstrate that a low error rate can still be achieved when using a single-station raw signal.

A Bayesian-based approach toward earthquake location from single-station observations is described in [2]. The task of predicting various parameters that characterize the seismic source is regarded as a regression problem for which two separate Bayesian neural network architectures are proposed. A temporal convolutional network, composed of causal dilated convolutions and residual connections, is designed to estimate epicentral distance and P travel time from one 1-min seismogram, whereas the back-azimuth angle is estimated, with a mean error rate of only 1°, using a network of standard 1-D convolutional layers applied over 1.5 s of the waveforms. Although achieving low mean estimation errors (i.e., 7.3- and 6.7-km mean errors for predicting the epicentral distance and depth), the approach proposed in [2] does not provide a method for predicting the magnitude of an earthquake and it uses additional information regarding the arrival times of the P and S waves.

In this letter, we propose a new method for estimating epicentral distance, depth, and magnitude from raw

Manuscript received November 30, 2020; revised February 3, 2021; accepted February 9, 2021. Date of publication February 25, 2021; date of current version December 31, 2021. *(Corresponding author: Anamaria Radoi.)*

The authors are with the CAMPUS Research Institute, University Politehnica of Bucharest, 060042 Bucharest, Romania, and also with the Department of Applied Electronics and Information Engineering, University Politehnica of Bucharest, 060042 Bucharest, Romania (e-mail: r.catalin196@yahoo.ro; anamaria.radoi@upb.ro).

Digital Object Identifier 10.1109/LGRS.2021.3059422

1-min seismograms observed by single stations, without relying on additional information regarding the arrival times of P and S waves. The proposed approach is based on the recent concept of complex CNN, which takes full advantage of the complex information retrieved from the short-time Fourier transform (STFT) representation of the seismic signals. In order to avoid convergence issues and to accelerate the training process, we define a novel complex residual layer, which benefits from the advantages of skip connections, but allows the processing of complex information at once. The method can be applied directly on raw signals, leveraging the preprocessing pipeline that leads to a shorter response time. The proposed approach is validated on a large-scale global data set of seismic signals, that is, STanford EArthquake data set (STEAD) [12].

The rest of this letter is structured as follows. Section II presents the proposed method for estimating epicentral distance, depth, and magnitude of seismic waves. Section III presents a set of experimental results and comparisons with existing approaches for estimating earthquake parameters. Finally, Section IV concludes this letter.

II. Proposed Method

A. Time–Frequency Representations of Seismic Signals

The majority of the state-of-the-art results in computer vision involve CNNs. Tasks regarding image recognition involve stacked convolutional layers, which are able to extract meaningful features at each processing level. In order to apply CNNs to our task, we first transform the time-domain seismic wave signal into the corresponding 2-D time–frequency representation. One of the most commonly used time–frequency representations is the discrete STFT

$$\mathrm{STFT}\{x[n]\}[k,r] = \sum_{m=0}^{L-1} x[rR+m]\cdot w[m]e^{-j\frac{2\pi}{N_x}km} \tag{1}$$

where $x[n]$ is the discrete input signal, $w[n]$ is a window function of length L, N_x is the STFT length, and R is the step size [13]. For simplicity, we chose the Hamming window as the window function. In order to perform the time–frequency analysis of the seismic signals, other time–frequency representations can be considered, e.g., the continuous wavelet transform with different base functions [14], S-transform [15], [16]. Nevertheless, considering the lower computational complexity of the STFT compared to other time–frequency methods, we restrict ourselves to STFT.

Typically, seismic stations record signals in three directions: east (E), north (N), and perpendicular (Z). Therefore, a seismic signal is a three-component vector $s[n]$ of dimensions $(3, N)$, where n is the discrete-time variable and N is the length of the signal. Consequently, the STFT is computed for each direction separately and concatenated into a multidimensional tensor.

B. Complex Neural Networks

Time–frequency transforms, e.g., STFT or S-transform, are usually performed in order to obtain meaningful seismic data representations. However, the output of such transforms is complex-valued representations. In most of the deep learning approaches based on time–frequency representations, the amplitude of these representations is considered, whereas the phase is neglected (see [17]). This approach is less computationally expensive but involves information loss and may, eventually, lead to poor results.

Currently, the vast majority of the neural blocks and activation functions for deep learning are based on real-valued operations and representations [18], [19]. However, recent work on the theoretical analysis of neural networks suggests that complex numbers could have a richer representational capacity [20]. Thus, the output of time–frequency transforms is more natural to be feedforward into neural blocks, which are intrinsically able to process complex-valued numbers. Our proposed solution toward estimating parameters of the seismic waves is based on complex neural layers, first introduced in [20] as a solution for various tasks in computer vision, music transcription, or speech spectrum prediction.

In order to perform the equivalent of a traditional real-valued convolution in the complex domain, we convolve a complex filter matrix $\mathbf{W} = \mathbf{A} + i \cdot \mathbf{B}$ by a complex vector $\mathbf{h} = \mathbf{x} + i \cdot \mathbf{y}$, where $\mathbf{A}$ and $\mathbf{B}$ are real-valued matrices and $\mathbf{x}$ and $\mathbf{y}$ are real-valued vectors. As the convolution operator is distributive, convolving the vector $\mathbf{h}$ by the filter $\mathbf{W}$, we obtain the core approach for complex convolutional layers

$$\mathbf{W} * \mathbf{h} = (\mathbf{A} * \mathbf{x} - \mathbf{B} * \mathbf{y}) + i \cdot (\mathbf{B} * \mathbf{x} + \mathbf{A} * \mathbf{y}). \tag{2}$$

Moreover, for the pooling operations and activation functions, we define complex max-pooling layers and complex rectified linear unit ($\mathbb{C}$ReLU) activations, as described by the following equations:

$$\mathrm{MaxPool}(\mathbf{W}) = \mathrm{MaxPool}(\mathbf{A}) + i \cdot \mathrm{MaxPool}(\mathbf{B}) \tag{3}$$

$$\mathbb{C}\mathrm{ReLU}(\mathbf{W}) = \mathrm{ReLU}(\mathbf{A}) + i \cdot \mathrm{ReLU}(\mathbf{B}) \tag{4}$$

where MaxPool and ReLU are the standard max-pooling operation and rectified linear activation function, respectively.

As shown in [20], this approach outperforms the classical method of individually processing only the magnitude or the phase of the time–frequency representations because the complex layers process data in accordance with the complex arithmetic, which leads to extracting more relevant features for complex data representations.

C. Neural Network Architecture

Earthquake sensors register the seismic activity in a 3-D space. Therefore, we consider the STFT of all three components of a seismic signal (E, N, Z), which leads to an input tensor of size $(3, K, M)$, where each element is a complex number. The first dimension of the tensor refers to the sensor component and represents the number of channels for the input data. The last two dimensions correspond to the STFT parameters with $K = N_x/2$ (i.e., the seismic signals are real-valued) and M representing the number of short-time overlapping frames.

We designed a complex neural network architecture that can process the seismic data and provide as output real-valued estimates for the distance, depth, and magnitude of the earthquake. The proposed model is comprised of six convolutional

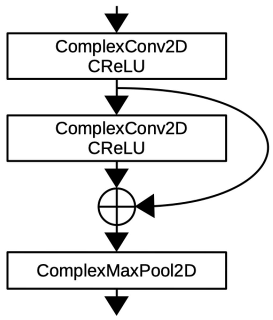

Fig. 1. Complex residual building block used in the network backbone.

blocks (conv) and a fully connected layer. The first two convolutional blocks are composed of a complex convolutional layer and a complex max-pooling layer, whereas the last convolutional block is composed of two layers of complex convolution and a complex max-pooling layer. As mentioned above, the complex convolutional layers and complex activation functions are specifically designed to work directly with complex values (i.e., real information and imaginary information are treated in parallel). This is a major asset when processing complex-valued time-frequency representations. In order to avoid convergence issues caused by vanishing gradients, we inserted skip connections from the third to the fifth convolutional block. First, introduced for residual neural networks [21], skip connections have the potential to improve generalization by reducing the impact that the nonconvex shape of the loss function has over the optimization process [22]. Our novel complex residual block, which is the basic building block of the proposed complex CNN-based architecture, is shown in Fig. 1.

The general architecture of the proposed network is presented in Fig. 2, whereas the detailed structure is shown in Table I. The layers are defined following (2)–(4). Considering that the dimensions of the feature maps are reduced after each convolutional block, the network is designed to be progressively wider in order to obtain a more complex data representation. Each convolutional layer has stride equal to 1 and corresponding padding such that the dimensions of the input and output feature maps of the convolutional layers remain constant. Moreover, each convolutional layer, except for the last one, is followed by the $\mathbb{C}$ReLU activation. The pooling layers downsample the input by a factor of 2. The last convolutional layer is followed by a fully connected layer with three output neurons that estimate continuous values, which corresponds to the estimations for distance, depth, and magnitude of the earthquake.

III. Experiments

We used the STEAD [12] for training and testing our complex network approach. STEAD is a large-scale global data set containing more than 1 million seismograms associated with roughly 450 000 earthquakes that occurred between January 1984 and August 2018. In our experiments, we used all earthquake events recorded by seismic stations at local distances. The data set was split into nonoverlapping training (70%), validation (15%), and test (15%) subsets randomly.

TABLE I

Detailed Structure of the Complex CNN-Based Architecture. Each Output Is Composed of Real and Imaginary Parts

Block	Layers	Output size
Conv & Pool 1	ComplexConv2D $(7 \times 7, 16)$ $\mathbb{C}$ReLU ComplexMaxPool2D (2×2)	$16 \times 256 \times 128$
Conv & Pool 2	ComplexConv2D $(7 \times 7, 16)$ $\mathbb{C}$ReLU ComplexMaxPool2D (2×2)	$16 \times 128 \times 64$
Complex Residual 3	ComplexConv2D $(5 \times 5, 32)$ $\mathbb{C}$ReLU ComplexConv2D $(5 \times 5, 32)$ $\mathbb{C}$ReLU ComplexMaxPool2D (2×2)	$32 \times 64 \times 32$
Complex Residual 4	ComplexConv2D $(3 \times 3, 64)$ $\mathbb{C}$ReLU ComplexConv2D $(3 \times 3, 64)$ $\mathbb{C}$ReLU ComplexMaxPool2D (2×2)	$64 \times 32 \times 16$
Complex Residual 5	ComplexConv2D $(3 \times 3, 96)$ $\mathbb{C}$ReLU ComplexConv2D $(3 \times 3, 96)$ $\mathbb{C}$ReLU ComplexMaxPool2D (2×2)	$96 \times 16 \times 8$
Conv & Pool 6	ComplexConv2D $(3 \times 3, 128)$ $\mathbb{C}$ReLU ComplexConv2D $(3 \times 3, 128)$ ComplexMaxPool2D (2×2)	$128 \times 8 \times 4$
FC	Fully Connected	1×3

All waveforms from the data set have three components, each of 1-min length with a sampling rate of 100 Hz (i.e., 6000 samples).

A. Experimental Setup

In order to have a smaller dynamic range for the predicted values, we scaled each label value by dividing it with an $\alpha = 10$ factor. We performed grid search to find the optimal STFT parameters for our setup, $N_x = 1024$ and $R = 22$, and kept only one side of the spectrum, considering that the spectrum is symmetric (i.e., the raw input is a real-valued signal).

We tuned the hyperparameters of our complex CNN model on the validation set in order to minimize the chance of overfitting in the hyperparameter space. We trained the model for 50 epochs with a mini-batch size of 32 samples. We set the learning rate to 5×10^{-4} and used a weight decay of 10^{-5}. These parameters are obtained using grid search. The training process is performed with the Adam optimizer [23], using the sum of mean squared errors on distance, depth, and magnitude as loss function.

B. Experimental Results

We compared our complex CNN-based approach (called complex CNN) with the baseline approach that was recently proposed in [2]. Moreover, we also compared the performance achieved by our approach with another architecture (called real CNN) that follows the same backbone as the one proposed in Fig. 2, but the complex-based layers are replaced by classical layers designed for real-valued numbers. In the latter case, the input data have six input channels, i.e., the real and imaginary parts of the STFT for each of the N, E, and Z directions. In order to assess the effectiveness of the

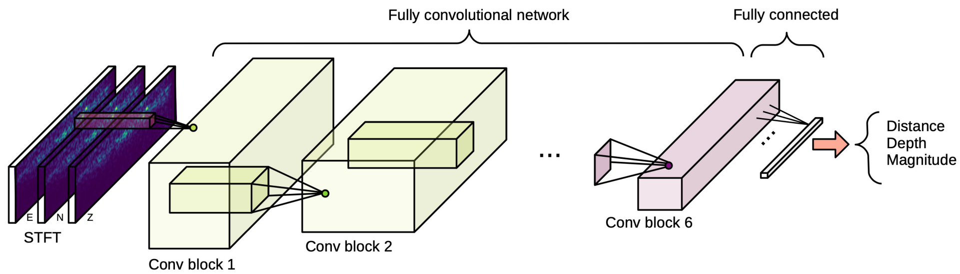

Fig. 2. General architecture of our CNN model. The input data are a tensor composed of three concatenated STFTs (one transform on each direction E, N, and Z). The input is processed through a series of convolutional blocks, both classical and residual complex convolutional blocks (as described in Fig. 1), and a fully connected layer. The output is a vector that characterizes the earthquake in terms of distance, depth, and magnitude.

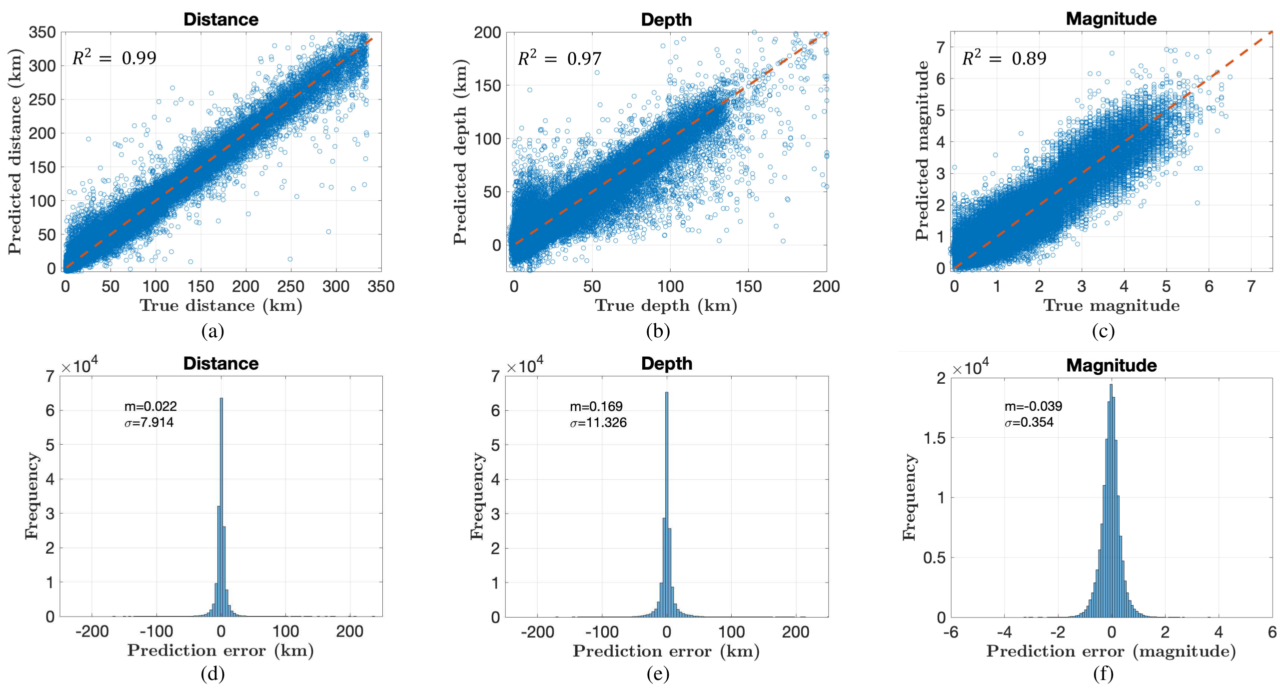

Fig. 3. Prediction results achieved by complex CNN-based architecture on the test subset of STEAD data set. The first row represents the true values versus the predicted values for (a) epicentral distance, (b) depth, and (c) magnitude. The dashed line stands for the ideal results (i.e., the error is equal to 0), whereas the coefficient of determination R^2 is provided on the plots. The second row represents the histograms depicting the distributions of errors for (d) epicentral distance, (e) depth, and (f) magnitude. The error distributions are modeled as Gaussian distributions, with the mean (m) and standard deviation (σ).

TABLE II

MAEs (for the Test Set) Achieved by Our Complex CNN Architecture and Comparisons With Other Methods. The Symbol ↓ Means That Lower Values Are Better

Method	Distance [km]↓	Depth [km]↓	Magnitude↓
Mousavi et al. [2]	7.3	6.7	-
Real CNN (ours)	5.53	6.67	0.33
Complex CNN (ours)	**4.51**	**6.15**	**0.26**

proposed approach, we use the mean absolute error (MAE) between the label epicentral distance, magnitude, and depth and the corresponding predicted values. The performance results are shown in Table II.

As already mentioned, the approach proposed in [2] does not provide magnitude estimation and uses additional information regarding the arrival time of P and S waves in order to direct the attention of the network toward the most informative part of the data. The experiments performed on the same data set show that the real CNN model outperforms the solution for both distance and depth estimation. Our best model, complex CNN, attains the best results in terms of MAE by a considerable margin for all estimated variables. The MAE on distance for the complex CNN is with 2.79 km smaller compared with [2] and with 1.02 km smaller than the same architecture, but for real-valued numbers. The complex CNN obtained the smallest MAE for depth, 6.15 km, with 0.55 km better than [2]. In addition to estimating the distance and depth of the seismic event, our model predicts also the magnitude of the earthquake. This estimation may be helpful for a preliminary classification and an early damage assessment since our complex model can predict the earthquake magnitude with an MAE of 0.26.

In order to have a more comprehensive understanding of the results achieved by the complex model, we show the distribution error on the test subset of STEAD data set for all estimations (distance, depth, and magnitude) in Fig. 3. As it can be depicted from Fig. 3(a), in terms of distance

TABLE III

INFERENCE PROCESSING TIME, EXPRESSED IN MILLISECONDS (ms), OF OUR PROPOSED MODELS ON CPU (i.e., INTEL I7) AND GPU DEVICES (i.e., GPU 1 IS QUADRO M4000 AND GPU 2 IS RTX2080TI). THE RESULTS ALSO INCLUDE THE STFT COMPUTING TIME

Architecture	CPU [ms]	GPU 1 [ms]	GPU 2 [ms]
Real CNN	123.5	10.1	9.8
Complex CNN	861.3	13.1	11.3

estimation, all errors are distributed closely along the dotted line, which refers to an ideal case of errors equal to 0. This fact is highlighted in Fig. 3(d), where we observe that the error follows a Gaussian distribution, with a mean of 0.022 and a standard deviation of 7.914. The depth estimation error is similar, with the mention that the errors are wider spread in comparison with the distance errors, i.e., the standard deviation of the depth distribution errors is 11.326.

A key factor in the seismic domain is the processing time. In this regard, we analyzed the inference time for the real CNN and the complex CNN on three different processing units, Intel i7 CPU, Nvidia Quadro M4000 GPU, and Nvidia RTX2080Ti GPU. The corresponding results are listed in Table III. As expected, the complex CNN is slower in comparison with the real CNN approach, but the difference is almost negligible on the GPUs. We observe that our best model, complex CNN, has a response time of 13.1 ms on GPU 2 and 11.3 ms on GPU 2, including the STFT computing time, which is approximately 5.5 ms. The processing time is shorter than in the case of using other time–frequency representations (i.e., 8.5 ms for the S-transform and 234.3 ms for the continuous wavelet transform with the Morlet base function). These results show that our solution is suitable for applications requiring timely information for disaster management and risk mitigation.

Overall, the proposed complex CNN-based model achieves an estimation accuracy that outperforms current approaches for this data set. In this sense, the proposed approach can be considered as a candidate for accurate and rapid seismic source characterization by providing close estimates regarding the parameters of a seismic event based only on single-station raw observation data.

IV. CONCLUSION

In this letter, we propose a new complex convolutional model that can provide a rapid characterization of the seismic activity in terms of epicentral distance, depth, and magnitude, based on information acquired from a single station. In line with the recent findings in the literature, our experiments show that a single seismic station is enough to accurately characterize the seismic activity for a certain area. Furthermore, our architecture can provide close estimates from raw seismograms, without relying on information regarding the arrival of the P and S waves. We compared our model in a comprehensive experiment and obtained state-of-the-art results on the STEAD data set. One interesting direction that could be further investigated is the integration of other time–frequency representations (e.g., S-transform or generalized S-transform) in the estimation procedure. However, this could lead to a longer processing time and a slower response compared to the proposed method.

REFERENCES

[1] T. Perol, M. Gharbi, and M. Denolle, "Convolutional neural network for earthquake detection and location," *Sci. Adv.*, vol. 4, no. 2, Feb. 2018, Art. no. e1700578.

[2] S. M. Mousavi and G. C. Beroza, "Bayesian-deep-learning estimation of earthquake location from single-station observations," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 8211–8224, Nov. 2020.

[3] O. M. Saad, A. G. Hafez, and M. S. Soliman, "Deep learning approach for earthquake parameters classification in earthquake early warning system," *IEEE Geosci. Remote Sens. Lett.*, early access, Jun. 9, 2020, doi: 10.1109/LGRS.2020.2998580.

[4] S. M. Mousavi and G. C. Beroza, "A machine-learning approach for earthquake magnitude estimation," *Geophys. Res. Lett.*, vol. 47, no. 1, Jan. 2020.

[5] M. P. A. Ende and J. Ampuero, "Automated seismic source characterization using deep graph neural networks," *Geophys. Res. Lett.*, vol. 47, no. 17, Sep. 2020, Art. no. e2020GL088690.

[6] M. Malfante, M. Dalla Mura, J.-P. Metaxian, J. I. Mars, O. Macedo, and A. Inza, "Machine learning for volcano-seismic signals: Challenges and perspectives," *IEEE Signal Process. Mag.*, vol. 35, no. 2, pp. 20–30, Mar. 2018.

[7] A. Payani, A. Abdi, X. Tian, F. Fekri, and M. Mohandes, "Advances in seismic data compression via learning from data: Compression for seismic data acquisition," *IEEE Signal Process. Mag.*, vol. 35, no. 2, pp. 51–61, Mar. 2018.

[8] W. Zhu, S. M. Mousavi, and G. C. Beroza, "Seismic signal denoising and decomposition using deep neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9476–9488, Nov. 2019.

[9] Z. E. Ross, M. Meier, E. Hauksson, and T. H. Heaton, "Generalized seismic phase detection with deep learning," *Bull. Seismol. Soc. Amer.*, vol. 108, no. 5, pp. 2894–2901, Aug. 2018.

[10] N. Liu, T. He, Y. Tian, B. Wu, J. Gao, and Z. Xu, "Common-azimuth seismic data fault analysis using residual UNet," *Interpretation*, vol. 8, no. 3, pp. SM25–SM37, Aug. 2020.

[11] A. Lomax, A. Michelini, and D. Jozinovi, "An investigation of rapid earthquake characterization using single-station waveforms and a convolutional neural network," *Seismol. Res. Lett.*, vol. 90, no. 2, pp. 517–529, Mar. 2019.

[12] S. M. Mousavi, Y. Sheng, W. Zhu, and G. C. Beroza, "STanford EArthquake dataset (STEAD): A global data set of seismic signals for AI," *IEEE Access*, vol. 7, pp. 179464–179476, 2019.

[13] J. B. Allen and L. R. Rabiner, "A unified approach to short-time Fourier analysis and synthesis," *Proc. IEEE*, vol. 65, no. 11, pp. 1558–1564, Nov. 1977.

[14] M.-F. Guo, X.-D. Zeng, D.-Y. Chen, and N.-C. Yang, "Deep-learning-based Earth fault detection using continuous wavelet transform and convolutional neural network in resonant grounding distribution systems," *IEEE Sensors J.*, vol. 18, no. 3, pp. 1291–1300, Feb. 2018.

[15] N. Liu, J. Gao, B. Zhang, F. Li, and Q. Wang, "Time–frequency analysis of seismic data using a three parameters S transform," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 1, pp. 142–146, Dec. 2018.

[16] N. Liu, J. Gao, B. Zhang, Q. Wang, and X. Jiang, "Self-adaptive generalized S-transform and its application in seismic time–frequency analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7849–7859, Oct. 2019.

[17] N.-C. Ristea, A. Anghel, and R. T. Ionescu, "Fully convolutional neural networks for automotive radar interference mitigation," in *Proc. VTC-Fall*, 2020, pp. 1–5.

[18] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. ICML*, 2010, pp. 807–814.

[19] M.-I. Georgescu, R. Tudor Ionescu, N.-C. Ristea, and N. Sebe, "Nonlinear neurons with human-like apical dendrite activations," 2020, *arXiv:2003.03229*. [Online]. Available: http://arxiv.org/abs/2003.03229

[20] C. Trabelsi *et al.*, "Deep complex networks," in *Proc. ICLR*, 2018, pp. 1–19.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[22] H. Li, Z. Xu, G. Taylor, C. Studer, and T. Goldstein, "Visualizing the loss landscape of neural nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 6389–6399.

[23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, 2015, pp. 1–15.