EQConvMixer: A Deep Learning Approach for Earthquake Location From Single-Station Waveforms

Article *in* IEEE Geoscience and Remote Sensing Letters · September 2023

DOI: 10.1109/LGRS.2023.3312324

EQConvMixer: A Deep Learning Approach for Earthquake Location from Single-station Waveforms

Hagar S. Elsayed, Omar M. Saad, M. Sami Soliman, Yangkang Chen, and Hassan A. Youness

***Abstract*—We present a novel deep-learning method using the ConvMixer network for automatic earthquake location. The proposed ConvMixer network utilizes three-component waveform recordings of single stations for estimating the hypocenter location. The ConvMixer network is a patch-based architecture that combines depthwise and pointwise convolutions to extract the global and local information of the earthquake waveforms. We train and test the proposed method using the Italian seismic dataset (INSTANCE). The ConvMixer network estimates the earthquake hypocenter locations with high accuracy, reaching a mean absolute error (MAE) of 2.71 km for the epicenter distance, and 1.15 km for the depth. In addition, we use the global STanford EArthquake Dataset (STEAD) to further evaluate the performance of the ConvMixer. As a result, the ConvMixer network achieves MAEs of 2.27 km and 1.19 km for the distance and the depth, respectively. The proposed ConvMixer network is compared to the benchmark methods, i.e., ResNet, AlexNet, MobileNet, and Xception, and outperforms all of them.**

***Index Terms*—Earthquake location, deep learning, ConvMixer.**

I. Introduction

EARTHQUAKE location is a crucial task in observational seismology and plays a fundamental role in numerous seismological studies, such as fault zone characterization, seismic hazard assessment, seismic tomography, and earth deformation. The location of earthquake sources can be expressed as a typical inverse problem, which aims to estimate the location coordinates (longitude, latitude, and depth) of the earthquakes from seismic recordings. Due to the rapidly growing quantity of seismic data worldwide, the development of automated earthquake location approaches is strongly demanded. Generally, there are two ways for earthquake location, i.e., the single-station approach and the network approach. The single-station approach needs only a single waveform for locating earthquakes, while the network approach requires at least three seismic stations to locate the earthquakes and hence could become completely useless in situations when only a single seismic station is available. The single-station approach is mandatory in the seismic network that is characterized by a small number of seismic stations, while the network approach is valid for a denser seismic network containing a large number of seismic stations.

Recently, deep learning techniques have been widely applied in various seismological problems, e.g., earthquake detection [1], seismic data denoising [2], seismic phase picking [3], magnitude estimation [4], earthquake location [5, 6], seismic data compression [7], and seismic source discrimination [8]. Several deep learning approaches have been proposed for estimating the earthquake location directly from three-component seismograms recorded by a single station [9–13]. Perol et al. [9] proposed a ConvNetQuake, which is based on a convolutional neural network (CNN), for detecting earthquakes and classifying the epicenter distance into six geographic regions based on single station waveforms. Lomax et al. [10] expanded the ConvNetQuake model to estimate earthquake parameters such as epicenter distance, azimuth, depth, and magnitude. Mousavi and Beroza [11] used two distinct Bayesian neural networks to estimate the epicenter distance, P-wave travel time, and azimuth from single-station observations. Ristea and Radoi [12] proposed a complex CNN model to estimate the epicenter distance, depth, and magnitude using time-frequency representations of single station waveforms. Yoon et al. [13] utilized a multi-tasking learning (MLT) architecture for estimating the magnitude and epicenter distance using earthquake waveforms from a single station.

Most of the state-of-the-art earthquake location algorithms are based on CNN architecture. CNN is a common intrinsic deep-learning architecture that extracts meaningful representative features of the input data using filters. While traditional CNN utilizes raw input data for learning, recent CNN approach such as ConvMixer has demonstrated the effectiveness of employing patch-based representation for learning from visual data [14]. ConvMixer is a simplistic form of CNN architecture, which divides the input into a group of non-overlapping patches [14]. ConvMixer preserves the same size and resolution representation across all layers in the network. Additionally, ConvMixer dissociates spatial mixing from channel mixing of information for exploring global and local features using only regular convolutions. Despite its simplicity, ConvMixer has the capability of learning effectively from small training datasets [14]. Using ConvMixer can enhance earthquake location accuracy due to its ability to extract the spatial relation between the different seismic phases from the single-station seismic records.

In this study, we propose a ConvMixer network architec-

Hagar S. Elsayed is with the ENSN Laboratory, Seismology Department, National Research Institute of Astronomy and Geophysics (NRIAG), Helwan 11421, Egypt, and also with the Department of Computers and Systems Engineering, Faculty of Engineering, Minia University, Minia 61519, Egypt (e-mail: hagar_saad@nriag.sci.eg). (Corresponding author: Hagar S. Elsayed)

Omar M. Saad is with the ENSN Laboratory, Seismology Department, National Research Institute of Astronomy and Geophysics (NRIAG), Helwan 11421, Egypt (e-mail: omar.saad@nriag.sci.eg).

M. Sami Soliman is with the ENSN Laboratory, Seismology Department, National Research Institute of Astronomy and Geophysics (NRIAG), Helwan 11421, Egypt, and also with the Department of Communication and Computer Engineering, Faculty of Engineering, Nahda University, Beni Suef 65211, Egypt (e-mail: mahmoud.salam@nriag.sci.eg).

Yangkang Chen is with the Bureau of Economic Geology, Jackson School of Geosciences, The University of Texas at Austin, Austin, TX 78713 USA.

Hassan A. Youness is with the Department of Computers and Systems Engineering, Faculty of Engineering, Minia University, Minia 61519, Egypt (e-mail: hassan_youness@mu.edu.eg).

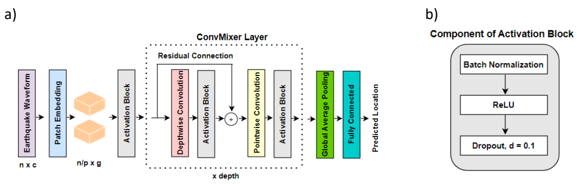

Fig. 1: (a) Framework of the proposed ConvMixer architecture. (b) Component of activation block.

ture for estimating the earthquake hypocenter (i.e., longitude, latitude, and depth) using single-station records without providing information about the S-phase arrival time or velocity model to the ConvMixer. We use the Italian seismic dataset (INSTANCE) [15] for training and testing the model. The test results indicate that the ConvMixer model is capable of estimating the hypocenter locations accurately using only a single-station waveform, which is extremely beneficial for seismic networks with insufficient stations.

II. Methodology

A. Dataset

We use a subset of the INSTANCE [15] for training and testing the proposed ConvMixer model. The INSTANCE dataset contains labeled three-component seismograms of earthquake and noise waveforms recorded in the Italian peninsula. Each waveform has a length of 120 s with a 100 Hz sampling rate. First, we select earthquake waveforms recorded by stations that each have more than 1,000 earthquake events at epicenter distances within one degree (≈112 km) with a signal-to-noise ratio (SNR) greater than 15 dB. The SNR is determined as follows [15]:

$$SNR = 20\log_{10}\frac{|E|}{|Q|}, \tag{1}$$

where $|E|$ and $|Q|$ are the 95th percentiles of amplitudes in a 5 s window after the S-phase arrival time and before the P-phase arrival time, respectively. Then, we extract only earthquakes whose waveforms are of length equal to 60 seconds, i.e., 3 s before and 57 s after the P-phase arrival time. Accordingly, 104,575 earthquake waveforms are extracted with a local magnitude range from 0 to 6.5 M_L, and a depth range between 0.1 to 71.7 km. The longitude varies between $11.39°$ and $15.10°$ and the latitude between $41.18°$ and $44.42°$.

B. The Proposed Method

We develop a deep learning approach to estimate the hypocenter location of earthquakes using a single station. The network architecture of the proposed ConvMixer is shown in Fig. 1a. The ConvMixer network consists of several layers, i.e., the input layer, patch embedding layer, activation block, ConvMixer layer, global average pooling layer, and fully connected layer. The three-component earthquake waveforms of one minute (6,000 samples) length are used as input for the ConvMixer network.

The patch embedding divides the input waveform into several patches of size p, and dimension g [14]. This is achieved by applying a convolutional layer with c input channels, g output channels, and kernel size and stride equal to the patch size p. This transforms a waveform with n samples and c channels to $\frac{n}{p}$ samples and g channels. The output of the patch embedding can be represented as follows:

$$y_0 = \text{Conv}_{c\rightarrow g}(y, \text{stride} = p, \text{kernel\_size} = p), \tag{2}$$

where Conv represents the convolution operation, and y denotes the input waveform.

Afterward, the activation block is applied, which consists of three different layers, including batch normalization (BN) [16] followed by a rectified linear units (ReLU) activation function, and a dropout layer [17] with a rate of 0.1 as depicted in Fig. 1b. The BN is applied for accelerating the training process and stabilizing the network as follows [16]:

$$BN(y_0) = \lambda\,\frac{y_0 - \mu(y_0)}{\sqrt{\sigma^2(y_0)}} + \gamma, \tag{3}$$

where μ indicates the mean, σ^2 represents the variance, λ and γ are the learnable scale and shift parameters of the BN layer, respectively, which are tuned during the training process. Afterward, the ReLU is used for achieving non-linear transformation, the output of which is determined as follows:

$$ReLU(\ BN(y_0)\ ) = max(\ 0,\ BN(y_0)\ ), \tag{4}$$

Next, the dropout layer is utilized to avoid the overfitting problem [17]. The output of the activation block can be formulated as:

$$y_1 = d(ReLU(BN\{y_0\})), \tag{5}$$

where d represents the dropout layer.

The following part of the network is the ConvMixer layer [14], which is repeated for a specified number of times. The ConvMixer layer consists of a residual block comprising a depthwise convolution, and the activation block. The input of the depthwise convolution is concatenated with the output of the activation block. Then, the concatenated output is followed by pointwise convolution and the activation block. The depthwise convolution is a grouped convolution where the number of the groups equals the number of channels g. The depthwise convolution is used to mix the spatial locations among patches. ConvMixer network utilizes large kernel sizes (e.g., 9 9) for the depthwise convolution layer to mix the distant spatial locations and capture the global information . The pointwise convolution, which is a convolution with a kernel size of 1, is utilized to mix channel locations within patches for exploring local features. Both of the activation blocks applied after the depthwise and pointwise convolutions have the same architecture as shown in Fig. 1b. The overall process of the ConvMixer layer can be expressed as follows :

$$y_l' = d(ReLU(BN\{\text{ConvDepthwise}(y_{l-1})\})) + y_{l-1}, \tag{6}$$

$$y_{l+1} = d(ReLU(BN\{\text{ConvPointhwise}(y_l')\})), \tag{7}$$

where y_l' represents the input of the ConvMixer block, and y_{l+1} represents the output feature map.

Following the ConvMixer layer, the global average pooling layer is applied to minimize the feature map size by computing the average value for every feature map to produce a single feature vector of size g, resulting in a decreased number of network parameters.

Finally, the output feature vector is fed into the fully connected layer (output layer) with a linear activation function for estimating the earthquake location. The longitude and latitude difference between the earthquake event and the seismic station, and the normalized depth value are the output of the ConvMixer network.

C. ConvMixer Network Architecture for Earthquake Location

The input of the ConvMixer network is 60-sec three-component seismograms recording at a 100 Hz sampling rate (6000×3). We use a convolution layer with 512 feature maps. The filter size and the stride are equal to the patch size of 10 samples to serve as a patch embedding layer, resulting in patches of length 600 with 512 channels (600×512). The kernel size of the depthwise convolution is 13, and the depth of the ConvMixer layer is set to 10. The output size is the same for all the ConvMixer layers (600×512) except the global average pooling layer has an output size of 600. The fully connected layer is the final output layer with 3 neurons and a linear activation function, which represents the predicted longitude, latitude, and depth corresponding to the ConvMixer network. The mean square error (MSE) between the catalog values and the predicted values of the ConvMixer network is utilized as the loss function. We use Adam optimizer [18] with a learning rate of 0.001 for tuning the ConvMixer network parameters by minimizing the loss function.

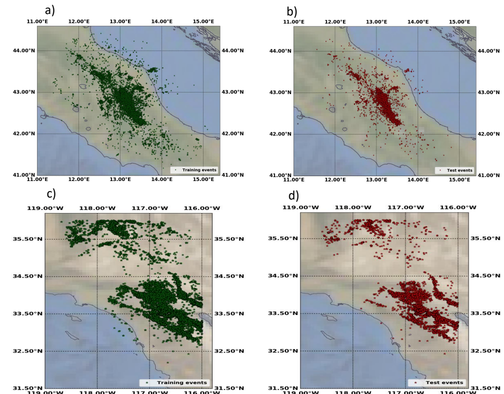

Fig. 2: Distribution of earthquakes used in this study. Training earthquakes distribution of the (a) INSTANCE dataset and (c) STEAD dataset. Testing earthquakes distribution of the (b) INSTANCE dataset and (d) STEAD dataset.

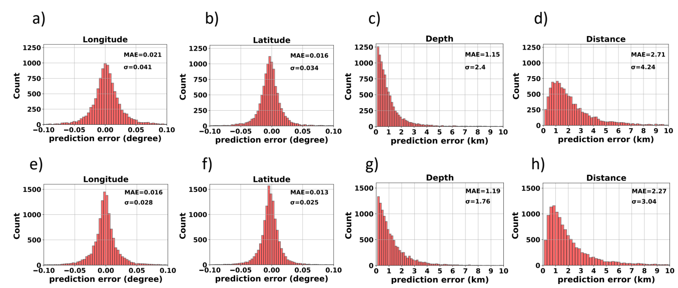

Fig. 3: Error distributions of the ConvMixer network for the (a)-(d) INSTANCE dataset and (e)-(h) STEAD dataset.

III. Results

A. The INSTANCE Dataset

Each of the three-component earthquake waveforms in the selected portion of the INSTANCE dataset is band-pass filtered between 1 and 45 Hz and normalized. We randomly divide the seismic dataset into 90% (94,117) for training and validation and 10% (10,458) for testing. The locations of the training and testing earthquakes are shown in Figs. 2a and 2b. We use 200 epochs with a batch size of 64 to train the ConvMixer model, and the training process is terminated when the validation loss does not decrease for the last 25 consecutive epochs. Afterward, we employ the pre-trained ConvMixer model to predict the earthquake hypocenter locations of the testing set. For quantifying the accuracy of the predictions, we compute the mean absolute errors (MAEs) of the longitude, latitude, and depth between the values given in the INSTANCE dataset and the values predicted by the Convmixer model. Accordingly, the MAEs are $0.021°$, $0.016°$, and 1.15 km for the estimated longitude, latitude, and depth, respectively, and the corresponding standard deviations (σ) are $0.041°$, $0.034°$, and 2.4 km. The MAE of the distance between the predicted and the ground-truth locations is 2.71 km with σ of 4.24 km. The error distributions of the predicted longitude, latitude, depth and distance are shown in Figs. 3a-d. We use The R^2 score (regression score function) to verify the effectiveness of the predicted epicenter location. Accordingly, we obtain R^2 of 0.98, which implies that the estimated and the ground-truth location are highly consistent. For evaluating the location performance of the Convmixer network, we compute the accuracy as follows:

$$Accuracy = \frac{TP}{TP + FP}, \tag{8}$$

where TP and FP represent the true positive, and false positive, respectively. Earthquakes with a distance error less than $0.1°$ ($\approx$11.2 km) are considered as TP; otherwise are considered as FP. Accordingly, the accuracy is 97.37%. To evaluate the performance of the proposed Convmixer network, we compare the results of the Convmixer network to the results of the benchmark networks such as ResNet [19], AlexNet [20], MobileNet [21], and Xception [22] as shown in Table I. To conduct a fair comparison, we utilize the same training and testing sets for all the networks. As a result, the ConvMixer network outperforms the other networks in estimating the earthquake location as it achieves the least MAE values for the depth and distance. Moreover, the ConvMixer network has the fewest parameters, i.e., 2,758,659 parameters, compared with other models. However, the ConvMixer network has the largest training time because of the large kernel size of the depthwise convolution. The training process occurred only once, while in real-time, the prediction time is the most important. The prediction time for one waveform of all the networks is milliseconds as shown in Table I, which are comparable with each other.

TABLE I: Results comparison of the earthquake location using different models. The reported time is using NVIDIA Quadro P4000 GPU.

Model	**Depth MAE (km)**		**Distance MAE (km)**		**Accuracy**		**Training Time Per Epoch (sec)**		**Prediction Time Per One Waveform (ms)**	**Parameters**
	Italian	California	Italian	California	Italian	California	Italian	California		
ResNet	1.55	1.61	6.42	6.32	86.89%	86.19%	1404	1640	32	10,124,803
AlexNet	1.32	1.25	4.67	3.0	92.78%	97.12%	127	137	25	51,913,295
MobileNet	1.70	1.78	9.36	7.83	74.06%	79.19%	615	741	35	7,976,899
Xception	1.46	1.44	4.55	4.12	92.85%	93.54%	960	1437	36	3,523,891
ConvMixer	**1.15**	**1.19**	**2.71**	**2.27**	**97.37%**	**98.52**	**1433**	**1672**	**29**	**2,758,659**

B. The STanford EArthquake Dataset (STEAD)

For evaluating the generalization ability of the ConvMixer, we train and test the ConvMixer using a portion of the STEAD dataset [23]. We extract the earthquake waveforms recorded in the California region, where ($-120°$E $<$ Longitude $< -116°$E and $32°$ N $<$ Latitude $< 36°$N). We only use waveforms that match the same criteria as the INSTANCE dataset, e.g., epicenter distances <= 112 km. Accordingly, we extract 119,263 waveforms with a depth range from 0.1 to 24.69 km, and a magnitude range from 0.1 to 5.5 M_L. The waveforms are bandpass filtered between 1 and 45 Hz and normalized. The seismic dataset is randomly split into training and validation (90%), and testing (10%) sets. Figs. 2c and 2d show the locations of training and testing events. The training process of the ConvMixer model is stopped when the validation loss fails to decrease for the last 25 successive epochs. We utilize a batch size of 64 and the number of epochs is set to 200. Then, we use the pre-trained ConvMixer model for predicting the earthquake hypocenter locations of the testing set. As a result, the MAEs of the longitude, latitude, and depth corresponding to the ConvMixer model are $0.016°$, $0.013°$, and 1.19 km, respectively, and the corresponding σ are $0.028°$, $0.025°$, and 1.76 km. The MAE of the distance is 2.27 km with a corresponding σ of 3.04 km. The error distributions corresponding to the ConvMixer model are shown in Figs. 3e-h. The results comparison between the Convmixer network and the benchmark networks are shown in Table I. The proposed algorithm achieves the minimum MAE values for both depth and distance while achieving the highest accuracy of 98.52%. We can notice that the proposed algorithm outperforms the deep learning benchmark methods.

IV. Discussion

A. Parameter Selection of the Proposed ConvMixer Network Architecture

The architecture of the ConvMixer network is mainly determined by four parameters [14]:

1) Width: the embedding dimension (number of filters in the convolutional layers),
2) Depth: the number of ConvMixer layers,
3) Patch size: the size of the patch, which controls the inherent resolution of the ConvMixer model,
4) Kernel size: the size of the depthwise convolution kernel.

Accordingly, to design the optimal ConvMixer network architecture, we tune the network architecture and determine the MAEs of the predicted longitude, latitude, and depth for each case using the INSTANCE dataset. First, we adjust both the width and the depth of the ConvMixer model, while fixing the patch size to 60 and the kernel size to 7. We employ different combinations, i.e., the width varies between 64, 128, 256, and 512, and the corresponding depth numbers change between 4, 6, 8, 10, and 12. As a result, we find that the optimal width and depth values are 512 and 10, respectively. After that, we tune the patch size using the optimal parameters achieved from the previous step. We utilize different patch sizes, e.g., 60, 40, 30, 20, and 10. As a result, we find that the patch size of 10 gives the minimum MAE values. Finally, using the optimal network parameters obtained by the previous steps, we adjust the kernel size of the depthwise convolution, e.g., 7, 9, 11, 13, and 15. As a result, the optimal kernel size is 13. consequently, the optimal hyperparameters of the ConvMixer model are 512, 10, 10, and 13 for the width, depth, patch size, and kernel size, respectively.

B. Interpretation of the Earthquake Location Mechanism Inside the ConvMixer Network

The deep learning feature maps contain critical information to clarify the relevant physical characteristics for locating earthquakes. Therefore, for analyzing the performance of the proposed algorithm, we select an earthquake example from the INSTANCE dataset and examine the feature maps extracted by the ConvMixer network as shown in Fig. 4. Fig. 4a shows the input 60-sec vertical component of the selected earthquake. Fig. 4b shows the output of the patch embedding layer, which has 600 patches. Figs. 4c-l show the output feature maps at different layers of the ConvMixer. Each of the ConvMixer layers aims to extract both local and global features of the earthquake patches. We observe that the extracted features preserve the P- and S-phase arrival times, which is suitable for estimating the earthquake location fundamentally.

C. The Application in Earthquake Early Warning

The primary purpose of the Earthquake Early Warning System (EEW) is to determine the earthquake parameters and generate an alert within a few seconds immediately after the arrival time of the P-wave. Therefore, the estimation of earthquake location should be accomplished utilizing a small portion of the waveform following the P-phase arrival time. Consequently, for ensuring that the proposed system satisfies the efficiency criteria of the EEW, we use 6-sec three-component earthquake waveforms (2 s before and 4 s after the P-phase arrival time). Those 6-sec seismograms are extracted from the same training and testing sets of the INSTANCE dataset. Additionally, we use the same ConvMixer

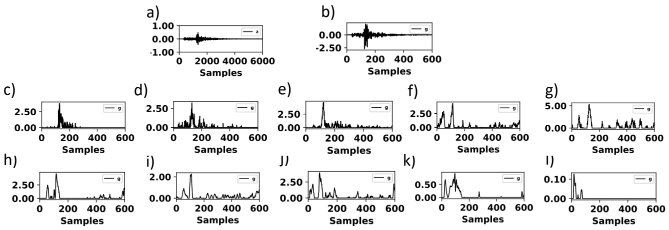

Fig. 4: Earthquake location mechanism analysis inside the ConvMixer network. (a) Example of the vertical component of an earthquake. (b) The output of the patch embedding layer. (c)-(l) The extracted feature maps from the first to the tenth ConvMixer layer, respectively.

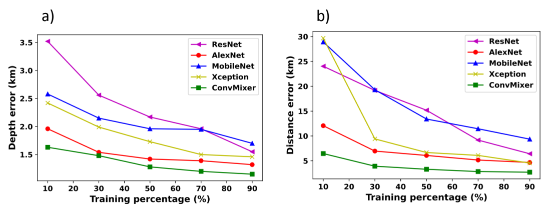

Fig. 5: Result comparison of ConvMixer and benchmark networks using different training/testing ratios of the INSTANCE dataset. (a) Depth error. (b) Distance error.

network architecture. As a result, the MAEs of the testing set are $0.03°$, $0.03°$, 1.16 km, and 4.52 km for the longitude, latitude, depth, and distance, respectively, with corresponding σ of $0.057°$, $0.056°$, 2.58 km, and 6.3 km. The ConvMixer network achieves promising location results despite using only 4 s after the P-phase arrival time recorded from single stations, which can benefit the EEW systems to rapidly estimate source location.

D. The Performance of the Proposed Algorithm Using Small Training Set

For further evaluation of the performance of the ConvMixer, we train the ConvMixer model using different training/testing ratios, e.g., 10%, 30%, 50%, 70%, and 90% of the INSTANCE dataset while using the same test set for all the cases, i.e., 10% of the INSTANCE dataset. Figs. 5a and 5b show the performance of the ConvMixer and the deep learning benchmark models using different training/testing ratios. Accordingly, the ConvMixer has the least MAE values for the depth and distance, especially when training with 10% of the dataset, i.e., the depth and distance errors are 1.63 km and 6.46 km, which indicates a good generalization ability.

V. Conclusion

The ConvMixer network has been proposed to automatically estimate the hypocenter location of an earthquake using one-minute three-component waveforms recorded by a single station. The robust performance of the ConvMixer network results from the patch embedding mechanism and the combination of depthwise and pointwise convolutions, where the hypocenter location is estimated using both global and local features extracted from the earthquake patches. The ConvMixer network is trained and tested using the INSTANCE and STEAD datasets, which demonstrates that single-station waveforms are sufficient for the ConvMixer network to estimate the location accurately. The ConvMixer network also outperforms the benchmark methods when evaluated using the same training testing sets. Besides, the proposed ConvMixer shows promising results using only 4 s after the P-wave arrival time which can support fast and reliable decision-making in earthquake early warning systems.

References

[1] H. S. Elsayed, O. M. Saad, M. S. Soliman, Y. Chen, and H. A. Youness, “Attention-based fully convolutional densenet for earthquake detection,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–10, 2022.

[2] O. M. Saad, Y. Chen, A. Savvaidis, W. Chen, F. Zhang, and Y. Chen, “Unsupervised deep learning for single-channel earthquake data denoising and its applications in event detection and fully automatic location,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–10, 2022.

[3] O. M. Saad and Y. Chen, “Capsphase: Capsule neural network for seismic phase classification and picking,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2021.

[4] O. M. Saad, Y. Chen, A. Savvaidis, S. Fomel, and Y. Chen, “Real-time earthquake detection and magnitude estimation using vision transformer,” *Journal of Geophysical Research: Solid Earth*, vol. 127, no. 5, p. e2021JB023657, 2022.

[5] M. Kriegerowski, G. M. Petersen, H. Vasyura-Bathke, and M. Ohrnberger, “A deep convolutional neural network for localization of clustered earthquakes based on multistation full waveforms,” *Seismological Research Letters*, vol. 90, no. 2A, pp. 510–516, 2019.

[6] O. M. Saad, M. Bai, and Y. Chen, “Uncovering the microseismic signals from noisy data for high-fidelity 3d source-location imaging using deep learning,” *Geophysics*, vol. 86, no. 6, pp. KS161–KS173, 2021.

[7] E. B. Helal, O. M. Saad, A. G. Hafez, Y. Chen, and G. M. Dousoky, “Seismic data compression using deep learning,” *IEEE Access*, vol. 9, pp. 58 161–58 169, 2021.

[8] O. M. Saad, M. S. Soliman, Y. Chen, A. A. Amin, and H. Abdelhafiez, “Discriminating earthquakes from quarry blasts using capsule neural network,” *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[9] T. Perol, M. Gharbi, and M. Denolle, “Convolutional neural network for earthquake detection and location,” *Science Advances*, vol. 4, no. 2, p. e1700578, 2018.

[10] A. Lomax, A. Michelini, and D. Jozinović, “An investigation of rapid earthquake characterization using single-station waveforms and a convolutional neural network,” *Seismological Research Letters*, vol. 90, no. 2A, pp. 517–529, 2019.

[11] S. M. Mousavi and G. C. Beroza, “Bayesian-deep-learning estimation of earthquake location from single-station observations,” *arXiv preprint arXiv:1912.01144*, 2019.

[12] N.-C. Ristea and A. Radoi, “Complex neural networks for estimating epicentral distance, depth, and magnitude of seismic waves,” *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.

[13] D. Yoon, Y. Li, B. Ku, and H. Ko, “Estimation of magnitude and epicentral distance from seismic waves using deeper crnn,” *IEEE Geoscience and Remote Sensing Letters*, 2023.

[14] A. Trockman and J. Z. Kolter, “Patches are all you need?” *arXiv preprint arXiv:2201.09792*, 2022.

[15] A. Michelini, S. Cianetti, S. Gaviano, C. Giunchi, D. Jozinović, and V. Lauciani, “Instance–the italian seismic dataset for machine learning,” *Earth System Science Data*, vol. 13, no. 12, pp. 5509–5544, 2021.

[16] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *International conference on machine learning*. pmlr, 2015, pp. 448–456.

[17] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[18] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

[19] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[21] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” *arXiv preprint arXiv:1704.04861*, 2017.

[22] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.

[23] 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. 179 464–179 476, 2019.