Geophysical Research Letters

RESEARCH LETTER
10.1029/2020GL089394

Key Points:

- A fully convolutional network is designed for real-time earthquake detection, location, and magnitude estimation
- Earthquake locations and magnitudes can be determined as early as a few seconds of earthquake signals received at very few stations
- The system evolutionarily improves and updates earthquake source parameters by receiving continuous data

Supporting Information:

- Supporting Information S1
- Movie S1
- Movie S2

Correspondence to:
M. Zhang,
Miao.Zhang@dal.ca

Citation:
Zhang, X., Zhang, M., & Tian, X. (2021). Real-time earthquake early warning with deep learning: Application to the 2016 M 6.0 Central Apennines, Italy Earthquake. *Geophysical Research Letters*, *48*, e2020GL089394. https://doi.org/10.1029/2020GL089394

Received 17 JUN 2020
Accepted 21 JAN 2021

Real-Time Earthquake Early Warning With Deep Learning: Application to the 2016 M 6.0 Central Apennines, Italy Earthquake

Xiong Zhang[1], Miao Zhang[2,3], and Xiao Tian[1]

[1]Engineering Research Center for Seismic Disaster Prevention and Engineering Geological Disaster Detection of Jiangxi Province, East China University of Technology, Nanchang, Jiangxi, China, [2]Department of Earth and Environmental Sciences, Dalhousie University, Halifax, Nova Scotia, Canada, [3]Department of Geophysics, Stanford University, Stanford, CA, USA

Abstract Earthquake early warning (EEW) systems are required to report earthquake locations and magnitudes as quickly as possible before the damaging S wave arrival to mitigate seismic hazards. Deep learning techniques provide potential for extracting earthquake source information from full seismic waveforms instead of seismic phase picks. We developed a novel deep learning EEW system that utilizes fully convolutional networks to simultaneously detect earthquakes and estimate their source parameters from continuous seismic waveform streams. The system determines earthquake location and magnitude as soon as very few stations receive earthquake signals and evolutionarily improves the solutions by receiving continuous data. We apply the system to the 2016 M 6.0 Central Apennines, Italy Earthquake and its first-week aftershocks. Earthquake locations and magnitudes can be reliably determined as early as 4 s after the earliest P phase, with mean error ranges of 8.5–4.7 km and 0.33–0.27, respectively.

Plain Language Summary Earthquake early warning (EEW) systems detect hazardous earthquakes, estimate their source parameters, and transmit warnings to the public. Conventional EEW algorithms depend on picking and analyzing the first seismic compressional wave (i.e., P wave). Seismic waveforms contain more information and can potentially be used to estimate earthquake source parameters with the fewest possible number of stations and to promptly transmit warning information. Deep learning techniques provide opportunities for extracting and exploiting the features behind seismic waveforms. In this study, we develop a fully automatic real-time EEW system by directly mapping seismic waveform data to earthquake source parameters using deep learning techniques. We apply this system to the 2016 M 6.0 Central Apennines, Italy Earthquake and its first-week aftershocks. Our results show EEW can be reliably issued as early as 4 s after the earliest P arrival.

1. Introduction

Earthquake early warning (EEW) is a practical way to mitigate seismic hazards by providing source parameters prior to significant ground shaking (Kanamori, 2005). The warning time depends on the distance between the user and earthquake epicenter, the EEW algorithm, and the station distribution. Epicentral regions suffer from the largest damage but have the least time and constraints to early warning alarms. Deep learning techniques provide potential for extracting earthquake source information from full waveforms, potentially making it possible to utilize fewer earthquake signals to reduce the early warning response time. We developed a real-time EEW system by directly mapping full waveforms to source parameters using a fully convolutional network (FCN) that was originally developed for image segmentation in computer vision.

EEW systems aim to rapidly and reliably provide earthquake source information before the damaging S wave arrival (Allen et al., 2009; Allen & Kanamori, 2003; Heaton, 1985). They detect hazardous earthquakes, estimate their source parameters, and transmit warnings to the public. Conventional EEW algorithms depend on picking and analyzing P phases and are generally divided into two categories: onsite warning and regional warning (Allen et al., 2009; Kanamori, 2005). For onsite warning, predominant periods and/or amplitudes of the first few seconds of P waves at single stations are utilized to evaluate source magnitude or ground shaking (Böse et al., 2009; Brown et al., 2009; Kuyuk & Allen, 2013; Olivieri et al., 2008; Shieh et al., 2008; Wu et al., 2007). For regional warning, arrival times and amplitudes of P waves at multiple

stations are used to estimate earthquake locations and magnitudes, providing more accurate source parameter solutions and more reliable and complete warning information (Allen et al., 2009; Cua et al., 2009; Kanamori, 2005; Picozzi et al., 2015). However, as pick-based location methods, regional warning algorithms are composed of many steps: phase detection, phase picking, phase association, earthquake location, and magnitude estimation (Grigoli et al., 2018; Satriano et al., 2011; M. Zhang et al., 2019). On the one hand, either of the above steps may potentially affect source parameter solutions. On the other hand, they require more time to receive and analyze earthquake signals at multiple stations. It turns out that early warning cannot be issued until a sufficient number of close stations are triggered, creating a large "blind zone" (Allen et al., 2009). In contrast, earthquakes usually cause the largest damage in epicentral regions. To eliminate such issues, methods based on the concept of "the triggered and not-yet-triggered stations" have been proposed and applied to constrain earthquake locations because earthquakes most likely occur near triggered stations and relatively far from not-yet-triggered stations (Horiuchi et al., 2005; Satriano et al., 2008). However, such methods only constrain earthquakes in specific regions, instead of providing accurate earthquake locations, and suffer from false alarms, especially if the number of triggered stations is small or some stations are malfunctioning. Thus, the development of a fully automatic real-time EEW system by directly mapping seismic waveform data to earthquake source parameters is critical.

Compared to seismic picks, seismic waveforms contain more information and can potentially be used to estimate earthquake source parameters with the fewest possible number of stations and to promptly transmit warning information. Deep learning techniques provide opportunities for extracting and exploiting the features behind seismic waveforms (Bergen et al., 2019; Kong, Trugman, et al., 2018b). Recent applications for seismological studies include phase picking (Zhu & Beroza, 2018), phase association (Ross et al., 2019), earthquake location (Perol et al., 2018), seismic discrimination (Li et al., 2018), waveform denoising (Zhu et al., 2019), and magnitude estimation (Mousavi & Beroza, 2020). One remarkable application is to detect and locate earthquakes from continuous waveforms through earthquake classification using the convolutional neural network, resulting in an earthquake detection rate 17 times higher than that of traditional methods (Perol et al., 2018). However, in the study by Perol et al. (2018), the exact earthquake location was not determined because the neural network is only able to classify earthquakes into groups. Additionally, an enormous number of samples would be required to train the classification network to obtain the location with precision comparable to that of traditional location methods (Szegedy et al., 2015; X. Zhang et al., 2020).

Precision issues can be addressed by analogizing the earthquake location to the image segmentation problem using a FCN developed for image processing (Chen et al., 2018; Long et al., 2015; Ronneberger et al., 2015). X. Zhang et al. (2020) demonstrated that despite the small number of available training samples, the FCN method shows promise for the determination of accurate earthquake location. However, the neural network was designed for earthquake location alone and could not handle real-time monitoring and magnitude estimation, making it unsuitable for the practical application of EEW. In this study, we designed a multibranch FCN for real-time EEW. One branch network is designed for earthquake location and the other for magnitude estimation. Earthquake source parameters are solved starting from the earliest stations receiving effective earthquake signals. The solutions are then improved by receiving more data in an evolutionary way. As an application, we apply this EEW system to the 2016 M 6.0 Central Apennines, Italy mainshock and its first-week aftershocks.

2. Data and Method

A series of moderate-to-large earthquakes struck Central Apennines, Italy, from August 2016 to early 2017. The 2016–2017 earthquake sequence resulted in 299 casualties and more than 20,000 homeless (Chiaraluce et al., 2017). Most damages come from the August 24, 2016 M 6.0 mainshock, which lacked distinct foreshocks and EEW. The study region is selected based on the distribution of mapped faults and background seismicity, focusing on a 48 × 96 km rectangular region with a depth range of 0–32 km (Figure 1). To maintain consistency with station recordings of historical and future earthquakes, we adopted the 12 nearest permanent broadband seismic stations with a station interval of 16–32 km. We collected 773 cataloged earthquakes with magnitude ranging from 2.0 to 3.7 as base samples to train the neural network, which occurred in the region from January 15, 2010 to August 8, 2016 (Figure S1). The M 6.0 mainshock and its

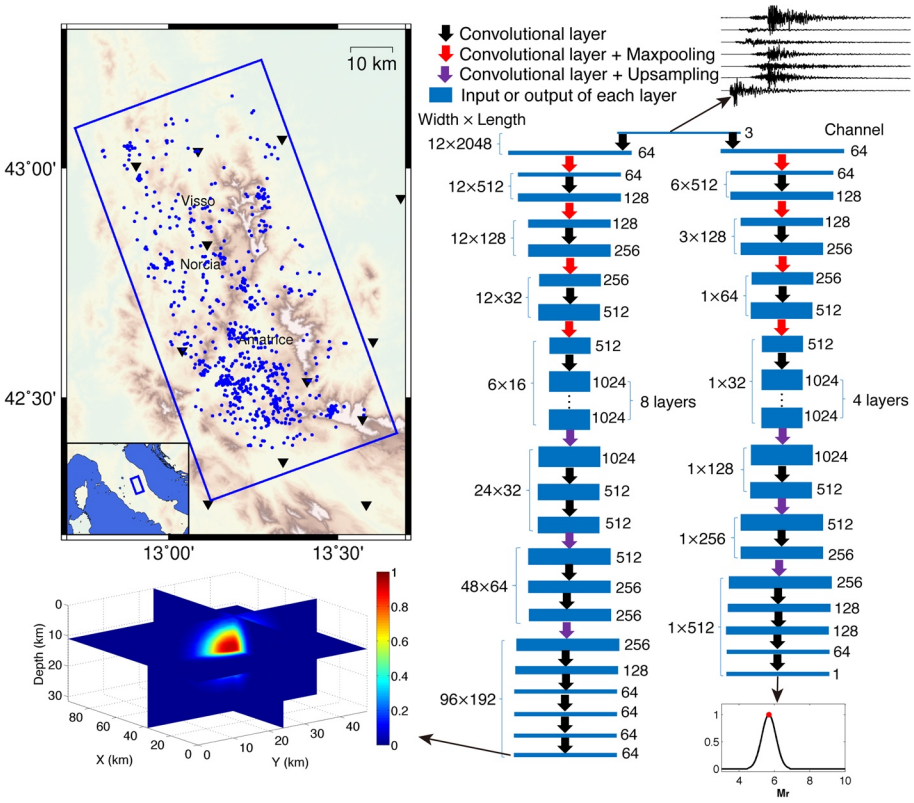

Figure 1. Map showing the study region, training data set, and neural network architecture. Detailed neural network settings are listed in Table S1. Earthquake location and magnitude are projected to 3D and 1D Gaussian distribution probabilities, respectively. Blue dots represent earthquakes used for neural network training; black triangles denote stations used in this study, and the rectangular box marks the study region. 1D, one-dimension; 3D, three-dimension.

500 aftershocks with M > 2.5 that occurred in following first week (i.e., from August 24 to August 31, 2016) were selected as the testing set.

We designed a multibranch neural network for earthquake location and magnitude estimation (detailed network settings can be found in Text S1 and Table S1). Similar to X. Zhang et al. (2020), we adapt the earthquake locations and magnitudes to the application of the FCN with three-dimension (3D) and one-dimension (1D) Gaussian distribution probability labels, respectively, instead of assigning exact values (Figure 1). The labels are centered at the cataloged earthquake locations and normalized magnitudes (see later definition). For earthquake location, inputs of our neural network are three-component velocity waveforms from the 12 permanent broadband seismic stations, with a size of 2,048 (time samples) × 12 (number of stations) × 3 (number of components). We discretize the study region and set the grid interval as 0.5 km in the horizontal and vertical directions; thus, the output for the location branch of our neural network is a 3D matrix (192 × 96 × 64) representing the monitoring region (i.e., length: 96 km, width: 48 km, depth: 32 km). For magnitude estimation, we normalize input waveforms by the maximum amplitude A_{max} among the 12 stations, and define a normalized magnitude $M_r = M - \log\left(A_{max}\right)$ by deducing the waveform amplitude (A_{max}) contribution from the cataloged magnitude M. Therefore, the neural network is able to solve for magnitude with normalized waveforms. The normalized magnitude M_r is independent of magnitude and represents the constant correction term and the falloff curve with the distance (Richter, 1935). In addition to three-component waveforms, we need to keep the maximum amplitude A_{max} among stations for both the training and testing samples to convert the normalized magnitude M_r to the true magnitude M. We label normalized magnitudes M_r as 1D Gaussian distribution probabilities centered at the normalized magnitudes of training earthquakes. In the monitoring stage, the neural network outputs a 3D probability distribution for the location of an earthquake and a 1D probability distribution for its normalized magnitude M_r. If the maximum location probability is larger than our preset empirical threshold (0.9), an earthquake

is detected, and the optimal location and normalized magnitude M_r are determined according to the positions of their maximum probabilities. By memorizing the maximum amplitude of the input waveforms, we convert the normalized magnitude M_r to the true magnitude M. We apply a deep learning phase picker, PhaseNet, to pick the earliest P phase within the current time window (Zhu & Beroza, 2018). The origin time is estimated by calculating the travel time difference between the station and the earthquake location. Therefore, the system simultaneously determines the location and the magnitude, as well as the origin time.

We trained our neural network on Tensorflow (Abadi et al., 2016) before applying it to continuous waveform streams. For the earthquake location, waveforms were filtered from 2 to 8 Hz. For the magnitude estimation, we filter waveforms in the range of 0.5–9 Hz and keep the maximum amplitude A_{max} among stations. To generalize our neural network, we apply two different data augmentation strategies to simulate (1) real-time earthquake monitoring by truncating the waveforms of training samples in different time windows and (2) station malfunction during the earthquake by randomly select one station and replacing its waveforms by zeros. To fully utilize those well-recorded earthquakes, we select 553 events with at least 10 operational stations including the nearest one. We randomly cut waveforms of the 553 training samples five times from 0.5 to 27 s relative to their first theoretical P phase with a total length of 30 s to cover various cases with few stations containing earthquake signals in truncated time windows. Theoretical travel times are calculated based on a local velocity model (Chiaraluce et al., 2017). To simulate station malfunction, for each of the 773 training samples we randomly set one station as a malfunctional station, and cut other waveforms two times randomly starting from 8 to 27 s relative to their first theoretical P phases. Therefore, the total number of training samples after augmentation is 4,311 (i.e., $553 \times 5 + 773 \times 2$). Additionally, to distinguish seismic noises from effective seismic signals, we empirically add 20 noise samples to the training set and label their location probabilities as zeroes (see discussion on noise sample selection in Text S2). We utilized 200 epochs with a batch size of four to train the network and select the model with the smallest validation loss as our final neural network model. We selected 15% of the training set as validation samples and adopted the model with the lowest validation loss.

3. Results

3.1. Pseudo Real-Time Monitoring of the M 6.0 Mainshock

The August 24, 2016 M 6.0 mainshock was responsible for most of the damages during the 2016–2017 earthquake sequence. We first focus on the M 6.0 mainshock to test our system. To simulate real-time earthquake monitoring, continuous waveforms at 12 stations are continuously input into the well-trained neural network with a 30 s time window and 0.5 s interval. When the predicted maximum location probability within the truncated time window is larger than our preset threshold, our network detects an earthquake and outputs the corresponding location and magnitude. The system immediately responded to the M 6.0 mainshock as soon as 2.0 s after the earliest theoretical P phase arrival (hereinafter called “first P phase”; Figure S2). The location and magnitude were continually updated and stabilized as more data was received. We show three snapshots at 4, 9, and 15 s after the first theoretical P arrival (Figure 2). The epicentral location of the M 6.0 mainshock is preliminarily determined at 4 s with a 4.6 km error, perturbing around the target location over time. Depth changes little with time but is systemically 1.5–2 km deeper than the target depth. Magnitude is underestimated as M 4.7 at 4 s and increases to M 5.2 and M 5.7 at 9 and 15 s, respectively, approaching the expected magnitude M 6.0. Overall, both the location and magnitude of the M 6.0 mainshock become acceptable 4 s after the first P phase.

3.2. Pseudo Real-Time Monitoring Earthquakes on August 24, 2016

We systematically investigated the performance of real-time earthquake monitoring focusing on the day of the M 6.0 mainshock. To remove common events appearing in nearby truncated time windows, we only keep the event within 30 s after a triggered detection, which most reliably possesses the maximum location probability. We further select earthquakes by only keeping events with a maximum location probability larger than 0.93. Based on the above criteria, we detect 1,671 earthquakes in one day and they show a consistent distribution pattern with the 1,475 earthquakes in the catalog (Figure 3). Among them, there are 1,658 and 1,470 events of $M \geq 1.0$ in our results and the catalog, respectively. We focus on 34 $M \geq 3.5$

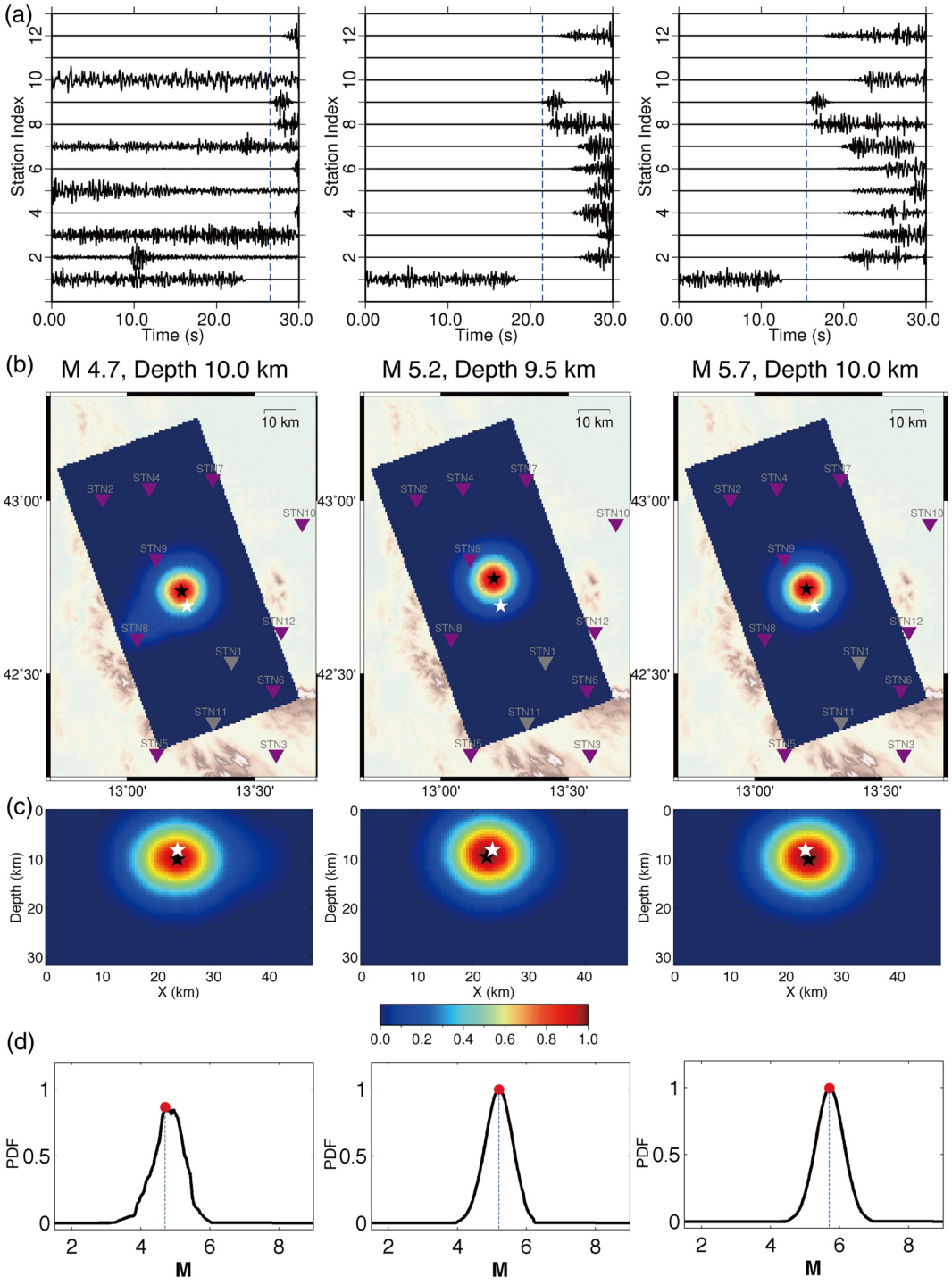

Figure 2. Snapshots of real-time monitoring of the M 6.0 mainshock. (a) Self-normalized vertical waveforms, (b) map view, (c) depth view of location probabilities, and (d) probability distribution of estimated magnitudes at 4, 9, and 15 s after the first P arrival (from left to right). The dashed lines denote the first P arrival. The station number in (b) corresponds to the station index in (a). Triangles denote seismic stations, and two malfunctioning stations are marked in gray. Black stars represent the optimal locations determined by our system, and white stars mark the cataloged locations.

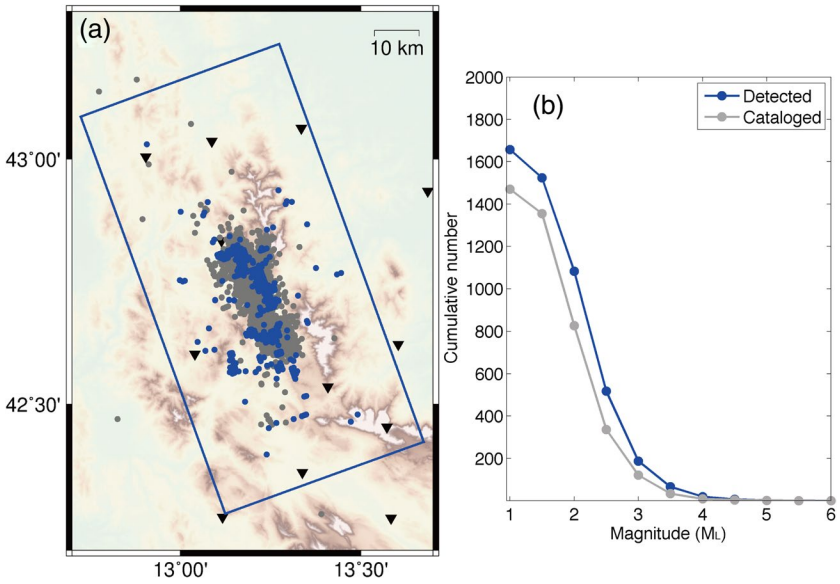

Figure 3. Source parameter comparison for earthquakes on August 24, 2016 between our detection and the catalog. (a) Epicentral location and (b) magnitude. Blue and gray dots represent earthquakes detected by the system and cataloged events, respectively. Only M ≥ 1.0 events were used for magnitude comparison.

cataloged earthquakes and compare our results with theirs for reliability and accuracy analysis. We recovered 32 of the 34 cataloged events with mean errors of epicentral location and magnitude of 4.4 km and 0.24, respectively (Figure S3). The two missed earthquakes are overlaying with other events or interfered by coda waves of large events (Figure S4). Among the 32 common events, there is only one event with origin time error larger than 2 s due to waveform overlapping with another event (Figures S5–S6). The magnitudes of the M 6.0 mainshock and the largest M 5.4 aftershock are underestimated because of waveform clipping (Figure S7). As a demonstration, we show real-time earthquake monitoring for one-hour starting from the M 6.0 mainshock (i.e., seismic activation period; Movie S1) and first hour of the day (i.e., seismic quiescence period; Movie S2), respectively.

3.3. Source-Parameter Robustness Analysis

Large earthquakes potentially cause more damage than small earthquakes. We focus on relatively large earthquakes to analyze their location and magnitude results. We adopted 42 M ≥ 3.5 earthquakes from the 500 testing samples. The mean epicentral location error of the 42 large earthquakes is 6.3 km at 4 s after the first P phase, reduces to 5.0 km at 9 s, and stabilizes at 4.0 km at 15 s (Figure S8). Compared with the cataloged magnitude, our magnitude is systemically underestimated at 4 s, because of the very limited signals received at the beginning, and approaches the cataloged magnitude as more data is received. Although there are no M > 3.7 earthquakes in the training set and the number of large earthquakes is much less than that of small ones (Figure S1), the FCN model is still applicable for predicting the locations and magnitudes of large earthquakes.

To comprehensively test the performance and robustness of our system, we symmetrically analyze 500 testing samples. To simulate real-time earthquake monitoring, we randomly cut waveforms of the 500 testing samples to form various time windows with different waveform lengths at stations. The results show that early warning could be activated as early as 3–4 s after the first P phase if only very few stations (<3) receive earthquake signals. As more data is received, the mean errors of epicentral location, depth, and magnitude decrease from 10.1, 1.7, 0.33 km to 4.7, 1.4, and 0.28 km, respectively (Figure 4). All source parameters are significantly improved during the first 5 s. Earthquake epicenters, depths and magnitudes start perturbing after 12.5, 5 and 5 s, respectively (Figure 4a, 4c, and 4e), which may be because the features of later complex scattered waves were not learned very well by the FCN. The mean magnitude error slightly increases between 5 and 10 s is likely because waveforms are in the transition period from initial S wave to full S wave

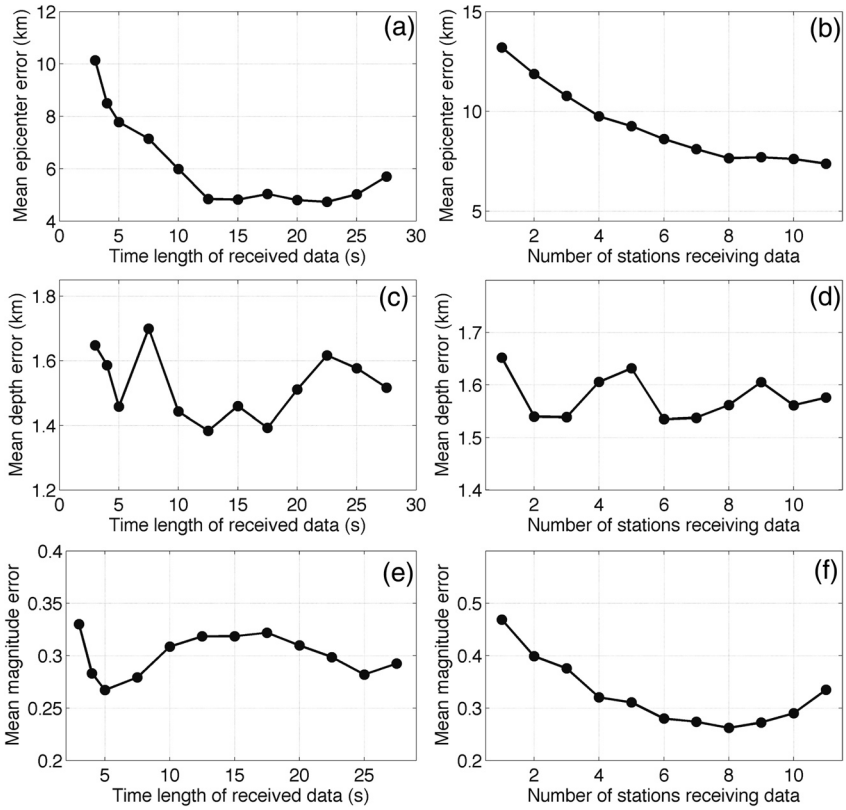

Figure 4. Error analysis of source parameters for 500 testing earthquakes. Mean errors of (a), (b) epicentral location, (c), (d) depth, and (e), (f) magnitude change with the time length of the received data and the number of stations receiving data.

including reflections and scattered phases, leading to the maximum amplitude cannot be estimated stably. With more stations receiving effective signals, the mean errors of epicentral location, depth, and magnitude decrease from 13.2, 1.6, 0.47 km to 7.3, 1.53, 0.26 km, respectively (Figure 4b, 4d, and 4f). The mean errors of epicentral location and magnitude are 11.9 km and 0.4 when only two stations record effective signals with all other noises and 8.5 km and 0.28 at 4 s after the first P arrival, potentially meeting EEW requirements. We also notice that the mean depth error is less sensitive to the data length and the number of stations compared with the mean epicenter error, which may be caused by larger errors of depths in the training data set.

4. Discussions

4.1. Unbalanced Magnitude Distribution in the Training Set

The percentage of small earthquakes is much greater than that of large earthquakes in the training set because specific regions have many more small events (Figure S1). This problem potentially affects both location and magnitude estimations in deep learning-based methods. Waveform features in large and small earthquakes may be different because of their distinct rupture scales (i.e., point source vs. finite fault). For location, our neural network is able to handle large events with an acceptable error range. Other studies point out more serious issues with the magnitude estimation (Mousavi & Beroza, 2020; Ochoa et al., 2018). It can be mitigated by increasing the percentage of large events in the training set by, for instance, generating synthetic waveforms of large events. However, this is challenged by accurate 3D velocity models and high-frequency waveform simulation. To eliminate this problem, we introduce a normalized magnitude M_r to label earthquake samples by excluding the contribution of absolute amplitudes and equalizing the labeled magnitude range between large and small earthquakes. With the exception of the M 6.0 mainshock and M 5.4 aftershock that were affected by waveform clipping, our predicted magnitudes for large events are as good as those for small events; waveform clipping can be addressed by utilizing strong motion recordings if

available. Whereas we notice that our magnitudes are usually underestimated during the first seconds due to very limit information input for the network, which can be offset by combinedly utilizing a conventional EEW algorithm (e.g., Wu and Zhao, 2006) although it may be not stable as well when only one station is available. Additionally, our result may potentially suffer from magnitude saturation for large earthquakes (e.g., M > 6), just like all other amplitude-based magnitude estimation (e.g., local magnitude) and EEW algorithms (e.g., the Pd method; Brown et al., 2009; Yamada et al., 2009).

4.2. Potential Improvements

We demonstrated source parameters can be automatically and evolutionarily solved in real-time from continuous seismic streams using the FCN method and cataloged earthquakes (see discussion on the FCN method in Text S3). However, the cataloged earthquakes for training and testing are not ground truth and inevitably contain uncertainties regarding the location and magnitude, which may originate from phase picking and velocity models and from the algorithms themselves. Thus, the network model may be further improved by adopting high-precision earthquake relocations for network training (Liu et al., 2020; Waldhauser & Ellsworth, 2000). Additionally, the monitoring ability and accuracy are expected to be improved by upgrading seismic station coverage.

In our system, the earthquake origin time is obtained based on the travel time difference between the earliest available P phase arrival and the source location. Our neural network automatically outputs the earthquake location when an earthquake is detected; then, the first P phase is picked by a deep learning picker, PhaseNet, within the current time window (Zhu & Beroza, 2018). A more straightforward way is to design another branch of the neural network to directly output the origin time, which will be tested in our future work.

Results show that our neural network generally works well for real-time EEW. Whereas one potential drawback is that the system may not distinguish multiple events when they occur closely one after another within the truncating time window (e.g., 30 s), potentially leading to event missing or source parameter mixing up between the events. This is challenging in most available EEW algorithms as well. The reliability and accuracy of early warning could be further improved by jointly working with other state-of-the-art EEW algorithms. For instance, deep learning-based single station location methods have shown potential for EEW in station sparse regions (Lomax et al., 2019; Mousavi & Beroza, 2019, 2020) and smartphones have been approved to significantly improve seismic data coverage for EEW in high-population regions (Kong, Inbal, et al., 2018a; Minson et al., 2015).

We use historical earthquakes to train the neural network in the monitoring zone and predict the location and magnitude for future earthquakes. The potential assumption is that a certain number of historical earthquakes are available in the target region. Thus, it is anticipated to work well for earthquake-prone areas (e.g., California and Japan) but not for regions with few background seismicity. Such problem may be solved if we train the neural network using world-wide seismic data and replace absolute station geometry information by utilizing a graph neural network (e.g., van den Ende et al., 2020).

5. Conclusions

We propose a deep-learning EEW system for providing earthquake source parameters in real time. Without phase picking, our deep learning neural network directly extracts earthquake location and magnitude information from seismic waveform streams and evolutionarily updates the solutions. Our system could provide reliable earthquake source parameters within a few seconds of effective signals received at very few stations (e.g., 2). We successfully applied the system to the 2016 Central Apennines, Italy mainshock and its subsequent aftershocks. Our results demonstrate the feasibility of using deep learning techniques for real-time EEW.

Data Availability Statement

The figures in our paper were made using Generic Mapping Tools (Wessel et al., 2013) and MATLAB (https://www.mathworks.com/products/matlab.html). Seismic data were downloaded from Italy's National Institute of Geophysics and Volcanology (INGV) through the International Federation of Digital Seismograph Networks (FDSN) web services. The earthquake catalog used in this study can be downloaded from http://cnt.rm.ingv.it/ (last accessed May 2020).

Acknowledgments
This work was supported by the National Natural Science Foundation of China Grant (No. 41704040, 42004040), Open Fund from Engineering Research Center for Seismic Disaster Prevention and Engineering Geological Disaster Detection of Jiangxi Province (SDGD202009), National Science Foundation Grant (No. EAR-1759810), the Natural Science and Engineering Research Council of Canada Discovery Grant (No. RGPIN-2019-04297), and the Science foundation of East China University of Technology Grant (DHBK2019072). The authors thank the Editor Germán Prieto, Associate Editor, and two anonymous reviewers for insightful suggestions and comments, which greatly improved this paper.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al. (2016). Tensorflow: A system for large-scale machine learning. *Paper presented at 12th USENIX symposium on operating systems design and implementation*, Savannah, GA.

Allen, R. M., Gasparini, P., Kamigaichi, O., & Böse, M. (2009). The status of earthquake early warning around the world: An introductory overview. *Seismological Research Letters*, *80*(5), 682–693. https://doi.org/10.1785/gssrl.80.5.682

Allen, R. M., & Kanamori, H. (2003). The potential for earthquake early warning in southern California. *Science*, *300*(5620), 786–789. https://doi.org/10.1126/science.1080912

Bergen, K. J., Johnson, P. A., de Hoop, M. V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. *Science*, *363*(6433) eaau0323. https://doi.org/10.1126/science.aau0323

Böse, M., Hauksson, E., Solanki, K., Kanamori, H., Wu, Y.-M., & Heaton, T. H. (2009). A new trigger criterion for improved real-time performance of onsite earthquake early warning in southern California. *Bulletin of the Seismological Society of America*, *99*(2A), 897–905. https://doi.org/10.1785/0120080034

Brown, H. M., Allen, R. M., & Grasso, V. F. (2009). Testing elarmS in Japan. *Seismological Research Letters*, *80*(5), 727–739. https://doi.org/10.1785/gssrl.80.5.727

Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(4), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184

Chiaraluce, L., Di Stefano, R., Tinti, E., Scognamiglio, L., Michele, M., Casarotti, E., et al. (2017). The 2016 central Italy seismic sequence: A first look at the mainshocks, aftershocks, and source models. *Seismological Research Letters*, *88*(3), 757–771. https://doi.org/10.1785/0220160221

Cua, G., Fischer, M., Heaton, T., & Wiemer, S. (2009). Real-time performance of the virtual seismologist earthquake early warning algorithm in southern California. *Seismological Research Letters*, *80*(5), 740–747. https://doi.org/10.1785/gssrl.80.5.740

Grigoli, F., Scarabello, L., Böse, M., Weber, B., Wiemer, S., & Clinton, J. F. (2018). Pick- and waveform-based techniques for real-time detection of induced seismicity. *Geophysical Journal International*, *213*(2), 868–884. https://doi.org/10.1093/gji/ggy019

Heaton, T. H. (1985). A model for a seismic computerized alert network. *Science*, *228*(4702), 987–990. https://doi.org/10.1126/science.228.4702.987

Horiuchi, S., Negishi, H., Abe, K., Kamimura, A., & Fujinawa, Y. (2005). An automatic processing system for broadcasting earthquake alarms. *Bulletin of the Seismological Society of America*, *95*(2), 708–718. https://doi.org/10.1785/0120030133

Kanamori, H. (2005). Real-time seismology and earthquake damage mitigation. *Annual Review of Earth and Planetary Sciences*, *33*(1), 195–214. https://doi.org/10.1146/annurev.earth.33.092203.122626

Kong, Q., Inbal, A., Allen, R. M., Lv, Q., & Puder, A. (2018). Machine learning aspects of the myshake global smartphone seismic network. *Seismological Research Letters*, *90*(2A), 546–552. https://doi.org/10.1785/0220180309

Kong, Q., Trugman, D. T., Ross, Z. E., Bianco, M. J., Meade, B. J., & Gerstoft, P. (2018). Machine learning in seismology: Turning data into insights. *Seismological Research Letters*, *90*(1), 3–14. https://doi.org/10.1785/0220180259

Kuyuk, H. S., & Allen, R. M. (2013). A global approach to provide magnitude estimates for earthquake early warning alerts. *Geophysical Research Letters*, *40*(24), 6329–6333. https://doi.org/10.1002/2013gl058580

Li, Z., Meier, M.-A., Hauksson, E., Zhan, Z., & Andrews, J. (2018). Machine learning seismic wave discrimination: Application to earthquake early warning. *Geophysical Research Letters*, *45*(10), 4773–4779. https://doi.org/10.1029/2018gl077870

Liu, M., Zhang, M., Zhu, W., Ellsworth, W. L., & Li, H. (2020). Rapid characterization of the July 2019 Ridgecrest, California earthquake sequence from raw seismic data using machine learning phase picker. *Geophysical Research Letters*, *47*(4), e2019GL086189. https://doi.org/10.1029/2019GL086189

Lomax, A., Michelini, A., & Jozinović, D. (2019). An investigation of rapid earthquake characterization using single-station waveforms and a convolutional neural network. *Seismological Research Letters*, *90*(2A), 517–529. https://doi.org/10.1785/0220180311

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *Paper presented at 2015 IEEE conference on computer vision and pattern recognition (CVPR)*, Boston, MA.

Minson, S. E., Brooks, B. A., Glennie, C., Murray, J. R., Langbein, J., Owen, S. E., et al. (2015). Crowdsourced earthquake early warning. *Science Advances*, *1*(3), 1–7. https://doi.org/10.1126/sciadv.1500036

Mousavi, S. M., & Beroza, G. C. (2019). *Bayesian-deep-learning estimation of earthquake location from single-station observations*. Preprint at https://arxiv.org/abs/1912.01144

Mousavi, S. M., & Beroza, G. C. (2020). A machine-learning approach for earthquake magnitude estimation. *Geophysical Research Letters*, *47*(1), e2019GL085976. https://doi.org/10.1029/2019gl085976

Ochoa, L. H., Niño, L. F., & Vargas, C. A. (2018). Fast magnitude determination using a single seismological station record implementing machine learning techniques. *Geodesy and Geodynamics*, *9*(1), 34–41. https://doi.org/10.1016/j.geog.2017.03.010

Olivieri, M., Allen, R. M., & Wurman, G. (2008). The potential for earthquake early warning in Italy using elarmS. *Bulletin of the Seismological Society of America*, *98*(1), 495–503. https://doi.org/10.1785/0120070054

Perol, T., Gharbi, M., & Denolle, M. (2018). Convolutional neural network for earthquake detection and location. *Science Advances*, *4*(2), e1700578. https://doi.org/10.1126/sciadv.1700578

Picozzi, M., Zollo, A., Brondi, P., Colombelli, S., Elia, L., & Martino, C. (2015). Exploring the feasibility of a nationwide earthquake early warning system in Italy. *Journal of Geophysical Research: Solid Earth*, *120*(4), 2446–2465. https://doi.org/10.1002/2014jb011669

Richter, C. F. (1935). An instrumental earthquake magnitude scale. *Bulletin of the Seismological Society of America*, *25*(1), 1–32.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Paper presented at International Conference on Medical image computing and computer-assisted intervention*, Munich, Germany.

Ross, Z. E., Yue, Y., Meier, M., Hauksson, E., & Heaton, T. H. (2019). PhaseLink: A deep learning approach to seismic phase association. *Journal of Geophysical Research*, *124*(1), 856–869. https://doi.org/10.1029/2018JB016674

Satriano, C., Lomax, A., & Zollo, A. (2008). Real-time evolutionary earthquake location for seismic early warning. *Bulletin of the Seismological Society of America*, *98*(3), 1482–1494. https://doi.org/10.1785/0120060159

Satriano, C., Wu, Y., Zollo, A., & Kanamori, H. (2011). Earthquake early warning: Concepts, methods and physical grounds. *Soil Dynamics and Earthquake Engineering*, *31*(2), 106–118. https://doi.org/10.1016/j.soildyn.2010.07.007

Shieh, J., Wu, Y., & Allen, R. M. (2008). A comparison of τc and τpmax for magnitude estimation in earthquake early warning. *Geophysical Research Letters*, *35*(20). https://doi.org/10.1029/2008GL035611

Szegedy, C., Wei, L., Yangqing, J., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). Going deeper with convolutions. *Paper presented at 2015 IEEE conference on computer vision and pattern recognition (CVPR)*, Boston, MA.
van den Ende, M. P. A., & Ampuero, J.-P. (2020). Automated seismic source characterization using deep graph neural networks. *Geophysical Research Letters*, *47*(17), e2020GL088690. https://doi.org/10.1029/2020gl088690
Waldhauser, F., & Ellsworth, W. L. (2000). A double-difference earthquake location algorithm: Method and application to the northern Hayward Fault, California. *Bulletin of the Seismological Society of America*, *90*(6), 1353–1368. https://doi.org/10.1785/0120000006
Wessel, P., Smith, W. H. F., Scharroo, R., Luis, J. F., & Wobbe, F. (2013). Generic mapping tools: Improved version released. *Eos, Transactions American Geophysical Union*, *94*(45), 409–410. https://doi.org/10.1002/2013EO450001
Wu, Y.-M., Kanamori, H., Allen, R. M., & Hauksson, E. (2007). Determination of earthquake early warning parameters, τc and Pd, for southern California. *Geophysical Journal International*, *170*(2), 711–717. https://doi.org/10.1111/j.1365-246X.2007.03430.x
Wu, Y. -M., & Zhao, L. (2006). Magnitude estimation using the first three seconds P-wave amplitude in earthquake early warning. *Geophysical Research Letters, 33*(16), L16312. https://doi.org/10.1029/2006GL026871
Yamada, M., Olsen, A. H., & Heaton, T. H. (2009). Statistical features of short-period and long-period near-source ground motions. *Bulletin of the Seismological Society of America*, *99*(6), 3264–3274. https://doi.org/10.1785/0120090067
Zhang, M., Ellsworth, W. L., & Beroza, G. C. (2019). Rapid earthquake association and location. *Seismological Research Letters*, *90*(6), 2276–2284. https://doi.org/10.1785/0220190052
Zhang, X., Zhang, J., Yuan, C., Liu, S., Chen, Z., & Li, W. (2020). Locating induced earthquakes with a network of seismic stations in Oklahoma via a deep learning method. *Scientific Reports*, *10*(1), 1941. https://doi.org/10.1038/s41598-020-58908-5
Zhu, W., & Beroza, G. C. (2018). PhaseNet: A deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, *216*(1), 261–273. https://doi.org/10.1093/gji/ggy423
Zhu, W., Mousavi, S. M., & Beroza, G. C. (2019). Seismic signal denoising and decomposition using deep neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(11), 9476–9488. https://doi.org/10.1109/TGRS.2019.2926772

References From the Supporting Information

Kingma, D. P., & Ba, J. (2014). *Adam: A method for stochastic optimization*. Preprint at https://arxiv.org/abs/1412.6980
Livni, R., Shalev-Shwartz, S., & Shamir, O. (2014). *On the computational efficiency of training neural networks*. arXiv preprint arXiv:1410.1141, 855–863. https://arxiv.org/abs/1410.1141
Wei, C., Lee, J. D., Liu, Q., & Ma, T. (2018). *On the margin theory of feedforward neural networks*. Preprint at https://arxiv.org/abs/1810.05369