

Real-time determination of earthquake focal mechanism via deep learning

Wenhuan Kuang (✉ wenhuan@stanford.edu)

Stanford University <https://orcid.org/0000-0002-4765-0823>

Congcong Yuan

Harvard University

Jie Zhang

University of Science and Technology of China <https://orcid.org/0000-0002-0530-3954>

Article

Keywords: Faulting Geometry, Stress Perturbation, Aftershock Patterns, Focal Mechanism Network, Global Waveform Characteristics

Posted Date: September 25th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-63884/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Communications on March 4th, 2021. See the published version at <https://doi.org/10.1038/s41467-021-21670-x>.

Abstract

An immediate report of the source focal mechanism with full automation after a destructive earthquake is crucial for timely characterizing the faulting geometry, evaluating the stress perturbation, and assessing the aftershock patterns. Advanced technologies such as Artificial Intelligence (AI) has been introduced to solve various problems in real-time seismology, but the real-time source focal mechanism is still a challenge. Here we propose a novel deep learning method namely Focal Mechanism Network (FMNet) to address this problem. The FMNet trained with 787,320 synthetic samples successfully estimates the focal mechanisms of four 2019 Ridgecrest earthquakes with magnitude larger than Mw 5.4. The network learns the global waveform characteristics from theoretical data, thereby allowing the extensive applications of the proposed method to regions of potential seismic hazards with or without historical earthquake data. After receiving data, the network takes less than two hundred milliseconds for predicting the source focal mechanism reliably on a single CPU.

Introduction

Mitigating the damaging level of earthquake hazards has been a long endeavor in seismology¹⁻³. When a destructive earthquake occurs, real-time reporting of the earthquake parameters is of crucial importance for immediate destruction assessment and emergency evacuations. Recent efforts have been refined towards applying AI technologies to estimate the source parameters because of its full automation, high efficiency, and human-like capability⁴⁻⁶, which has been remarkably demonstrated in numerous seismic processing tasks such as earthquake detection^{7,8}, seismic phase picking⁹⁻¹¹, magnitude estimation¹², and other tasks¹³⁻¹⁷. Besides reporting the three basic parameters of an earthquake (i.e., origin time, location, and magnitude), it is also exceedingly important to derive the source focal mechanism in time to better understand various aspects of the earthquake. For example, we can use source focal mechanisms to characterize faulting geometry and faulting mechanism^{19,20}. A group of focal mechanisms can be used to invert the spatial stress field distribution²¹⁻²³. We can also use the focal mechanism of the mainshock to calculate the static Coulomb stress changes²⁴⁻²⁶ for examining the earthquake triggering theory of the aftershocks²⁷⁻²⁹. Furthermore, the timely derived source focal mechanism can provide significant additions such as fault orientation and slipping mode to the point-source ground motion prediction model that is currently in practice³⁰⁻³² and thus has the potential to help improve the predicted ground shakings for early warning purpose. The immediate determination of the source focal mechanism is therefore of great importance to monitor and assess seismic hazards.

Compared to determining other source parameters of an earthquake (i.e., origin time, location, and magnitude), the estimation of the source focal mechanism usually requires much more human interactions and it is lack of full automation and efficiency. Approaches for conventionally resolving the focal mechanism mainly have three categories based upon the waveform information used, such as first

motions of P-waves^{33,34}, amplitudes of P- and/or S-waves^{35,36}, and full waveforms^{37,38}. After receiving the seismic data, these conventional methods usually take minutes to tens of minutes for retrieving the focal mechanism solution and hence incapable of realizing the real-time reporting. Recent progress that took advantage of an advanced search engine was performed³⁹ to estimate earthquake source parameters in less than one second. Although this approach reduces the time cost significantly, its implementation may be infeasible and impractical since it requires a tremendous search database (~hundreds of Gigabits) for each upcoming search. Besides, the search engine approach needs to reorganize the recorded waveforms as one-dimensional (1D) super trace and it is infeasible in implementation. Hence, the challenge remains in the full automation and practical implementation for real-time determination of the source focal mechanism.

In this study, we leverage the powerful advances in deep learning and propose a novel deep convolutional neural network (FMNet) for estimating the source focal mechanism rapidly using full waveforms. Unlike common applications, in which the training of supervised neural network models demand voluminous real data, the proposed FMNet here can be trained with synthetic data at first and then applied to real data directly. FMNet learns the universal characteristics of waveforms concerning the source focal mechanisms from the synthetic training data. This considers the scenarios without enough historical source focal mechanisms for training the neural network model, especially for those regions with limited seismicity but having the potential seismic hazards. For generating the large training dataset, we discretize the three-dimensional (3D) grid space of the study area of interest. We simulate theoretical waveforms with a variety of focal mechanisms at each spatial grid point. We train the FMNet model with the synthetic dataset and then apply it to predict the focal mechanisms of four real earthquakes with magnitudes larger than 5.4 of the Ridgecrest sequence which occurred in July 2019 in southern California. Additionally, we produce a by-product of the encoder, which is a sparse representation of the input waveforms, to analyze the working mechanism and robustness of the FMNet.

Results

Study area and data preparation. The study area is located in the region of Ridgecrest in southern California (Fig. 1), where a damaging earthquake sequence proceeded by an Mw 6.4 foreshock and followed by an Mw 7.1 mainshock in July 2019. Four moderate-to-large earthquakes (Mw>5.4) in the sequence are selected for this study. We collect the three-component (3-C) seismograms from 16 seismometers that are deployed by Southern Californian Seismic Network (SCSN) around the Ridgecrest area. They are utilized as the testing data for examining the validity of the proposed FMNet. Before the applications of the FMNet model, sufficient training data are vital for assuring a well-trained neural network. Here, instead of using the historic data, we simulate hundreds of thousands of synthetic data as training data since there are very limited source focal mechanisms of historical earthquakes available in this area.

As shown in Fig. 1, the study area is discretized from 35.4° to 36.2° in latitude direction, from -118.0° to -117.2° in longitude direction, and from 2 km to 20 km in depth. The intervals are 0.1°, 0.1°, and 2 km for latitude, longitude, and depth, respectively. We have $9 \times 9 \times 10 = 810$ virtual grid locations in 3D space. Assuming a double-couple source model⁴⁰ and a 1D velocity model of southern California⁴¹, we simulate the 3-C waveforms at 16 seismic stations by adopting the Thompson-Haskell propagator matrix⁴². For each virtual 3D grid, we simulate synthetic waveforms for all combinations of the strike, dip, and rake angles in the ranges of 0° to 360°, 0° to 90°, and -90° to 90°³⁷, respectively. The used intervals of the strike, dip, and rake angles are 30°, 10°, and 20°, respectively. Hence, we have $12 \times 9 \times 9 = 972$ focal mechanisms for each virtual grid and overall $810 \times 972 = 787,320$ synthetics as training samples, of which each sample contains the 3-C waveforms of 16 seismic stations with the time length of 128 seconds. We use 1 second as the sampling rate in all the simulations. Therefore each training sample has the size of 48 (three components by 16 stations) \times 128 (data length). Additionally, we have prepared another 1,000 synthetic samples as validation dataset. The validation dataset serves as unseen data to evaluate the training performance.

The training samples are processed by filtering between 0.05 Hz to 0.1 Hz, aligning with the theoretical P-wave first arrivals, and normalizing to the maximum amplitude. These preprocessing procedures are important because they help us get rid of the effect from other source parameters such as location and magnitude and mitigate the dependence on the heterogeneity of velocity medium. Considering the real data may present noise and picking errors, we resemble realistic scenarios by adding realistic noise and a random time shift (<10 s) to the synthetics (Supplementary Fig. S1). We process all the synthetic data in the same way and use them to train the network. After the FMNet is well trained, in case that one real earthquake is identified with the existing algorithms of automatic detection and phase picking⁷⁻¹¹, we first remove their instrument responses and then perform the bandpass filtering, arrival-time alignments, and amplitude normalizations on the data prior to feeding them to the FMNet.

FMNet training and prediction. The framework of the real-time determination of the source focal mechanism is presented in Fig. 2. It consists of two parts: FMNet training and prediction. For the training part, we train the FMNet with the synthetic data prepared previously along with the corresponding training labels. We describe the architecture of the FMNet, training labeling, and the associated training parameters in the method section. In the training process, both the training and validation losses, and the goodness of fitting between true and predicted labels of validation data are viewed as metrics to evaluate the performance of the training process (Supplementary Fig. S2 and Fig. S3). The stabilized training and validation loss curves after 50 iterations with sufficiently low resultant values and the high fitting level of between true and predicted labels both indicate that the FMNet has been stably trained. When it comes to

the prediction part, we can directly feed the processed recordings of a real earthquake into the trained FMNet to predict the source focal mechanism. The time cost of the training process may take hours to days, which depends on the size of the training samples, the complexity of the neural network, and computer capability. However, once well-trained, the designed FMNet can output a focal mechanism solution in only 196 milliseconds on a single CPU. Moreover, the trained network model can be deployed to estimate the source focal mechanisms in areas of interest permanently.

FMNet prediction results. The source focal mechanisms of four large earthquakes ($M_w > 5.4$) in the Ridgecrest sequence are estimated with the trained FMNet. We show these results as red beach balls in Fig. 3. The predicted focal mechanisms generally reveal the strike-slip faulting with very steeply dipping fault planes. Among them, the three focal mechanisms in the southern region, including the Mw 6.4 foreshock and Mw 7.1 mainshock, demonstrate pressure axes in the north-south direction and tension axes in the east-west direction. The other one in the northernmost region shows a slight rotation in the fault plane azimuth. For comparison, we also plot the focal mechanism results from the SCSN moment tensor catalog as reference solutions (in black) in Fig. 3. We can see that the predicted focal mechanisms by the FMNet and the reference focal mechanisms from the SCSN catalog are essentially consistent for the three earthquakes in the southern region, considering the differences in methods, parameterization, velocity model, and the amount of recording stations used. The northernmost event is not included in the SCSN catalog for comparison. For this event, we conduct the widely used generalized Cut-and-Paste method (gCAP)³⁸ to invert its focal mechanism as shown in grey. We observe that the inverted focal mechanism and the predicted focal mechanism match well for this event. Moreover, the slight rotation of fault azimuth is consistent with the distribution pattern of the aftershock event locations (grey dots). Comparing to other studies regarding this earthquake sequence^{20,23,43}, the predicted focal mechanisms by our FMNet are essentially consistent with previous results. All these results demonstrate that the proposed FMNet enables us to determine the source focal mechanisms effectively. Additionally, the trained FMNet only takes 196 milliseconds with a minimum requirement of computing resources and memory storage, which outperforms both the conventional methods and the fast search method.

The comparison of waveforms is the most straightforward way to evaluate the predicted results. For this purpose, we simulate the synthetic waveforms using the predicted source focal mechanisms by our FMNet and analyze the similarity between real waveforms and synthetic waveforms (Supplementary Fig. S4). After comparison, we find that both the amplitude and phase information of waveforms across different seismic stations are overlapped well and the computed cross-correlation coefficients reach 0.86, which indicates that the FMNet has learned the ability to recognize the waveforms and mapping them to the corresponding source focal mechanism solution reliably.

Interpreting the FMNet using the encoder. To further investigate the working mechanism of our FMNet, by adopting a similar idea in face recognition of which the network learns a mapping from face images to a compact Euclidean space where distances directly correspond to a measure of face similarity⁴⁴⁻⁴⁵, we output the extracted features to analyze the reliability and robustness. The last layer of the compression part of the FMNet is exported as a by-product of the encoder (see Fig. 6 and Method section for details). After training, this encoder can take any training input with the size of 1 48 128 and output the extracted feature with the size of 128 1 1. With the encoder, we verify the hypothesis that a measure of feature similarity in feature domain is equivalent to a measure of waveform similarity in data domain through adopting the following steps: first, we calculate the extracted features using the encoder for the whole training dataset to build an encoded database in feature domain. Then, we calculate the extracted features of the data that records a real earthquake. Finally, we measure the L2-norm misfits between the encoded database of training data and the encoded features of the real data in feature domain. For comparison, we also calculate the L2-norm misfits in data domain measuring the waveform differences between real data and training database. By finding the smallest L2-norm misfit, if the retrieved best solution in feature domain corresponds to the best solution retrieved in data domain, we can therefore validate the above hypothesis.

We take the Mw 6.4 foreshock as an example. With the steps illustrated above, we display the comparison of L2-norm misfit distributions that are calculated in data domain (in red) and in feature domain (in black) in Fig. 4a, after ranking in ascending order. Since the whole training dataset is too large, we plot only the first 5,000 smallest misfits for clarification. We can see that the L2-norm misfit distributions calculated in data domain and feature domain present a similar shape. Meanwhile, Fig. 4b, 4c, and 4d show the corresponding training labels of the strike, dip, and rake angles for the L2-norm misfits in feature domain (the black curve in Fig. 4a). By finding the smallest L2-norm misfit, the retrieved best solution of the strike, dip, and rake angles in feature domain are highlighted as magenta circles. Then we compare the best solutions retrieved in data domain and feature domain as shown in Fig. 5. We can observe that the best solution retrieved in feature domain (in magenta) matches well with the best solution that is retrieved in data domain (in red). These analyses and comparison results validate our hypothesis that the extracted features in feature domain maintain the essential information of the original waveforms in data domain under the least-square sense, thus the extracted features are sufficient to identify its corresponding source focal mechanism. Moreover, the 10 best solutions (in magenta and black) retrieved in feature domain are generally consistent with minor variations, which illustrates the stability of the trained network.

From the above analysis, the compression part of our FMNet (i.e., the encoder) can be interpreted as a sparse transformation of the input waveforms, where the input data have been compressed from 1 48 128 to 128 1 1 in size by a decreasing factor of 48 times while keeping the key information in the data.

The encoder also provides an alternative way to rapidly retrieve the best-matched source focal mechanism by searching in the dataset with encoded features that are prepared with the training data in advance. The expansion part of the FMNet mainly takes these extracted features to reconstruct a mapping function that yields the Gaussian distributions to represent the three angles of a focal mechanism. We address all these analyses presented in this section are for understanding the working mechanism of the FMNet and also for robustness analysis. When the proposed deep learning methodology is applied in a real case, we can directly feed the real data into the well-trained FMNet and output the focal mechanism rapidly. The intermediate output of the extracted feature maps can be used to further evaluate the reliability of the solution.

Discussions

The proposed FMNet is a novel deep-learning-based intelligent algorithm that allows us to estimate the source focal mechanism in less than 200 milliseconds after receiving the data with full automation. Compared to currently available algorithms, the proposed FMNet can not only determine the focal mechanism much faster but also independent of the heavy database for searching. The FMNet extracts and learns the essential features of the waveforms from the training dataset, thereby memorizing all the information into the neural network, and hence re-visiting the database is unnecessary. The proposed FMNet only stores the neural weights of a few Megabytes for the memory usage, which makes the FMNet much more feasible and applicable for automatic real-time applications. Our study also reveals that we can train the FMNet on the synthetic dataset and it is independent of historical earthquake data, which shall allow the extensive applications of the proposed method to areas with low seismicity but high risks of potential earthquake hazards. Besides, the successful application of AI in determining source focal mechanism shows that advanced AI technologies can handle much more complex data patterns other than those in detection, picking, location, etc.⁷⁻¹¹, and hence it may significantly provoke further studies in resolving the complexity of earthquake process.

Although our case study has used the long-period waveforms, the proposed deep learning method can be theoretically extended to early P-waves with a finer 3D earth structure and an accurate modeling tool for preparing the training data of early P-waves, which shall make this method more practical in real-time monitoring or early warning applications. Increasing the size of the monitoring area or increasing the number of recording stations will certainly increase the training time, but it will not increase the computation time significantly in prediction since the FMNet only requires extremely few computing resources for implementing the prediction. Combining with current efficient algorithms^{7-11; 15-16} for other source parameters (i.e. detection, location, and magnitude), we are able to estimate all the source parameters at once for real-time seismic hazard assessment.

Beyond the point-source assumption used, adopting the advanced modeling techniques⁴⁶⁻⁴⁸, we may also use the finite-fault rupture model to simulate the theoretical waveforms of larger earthquakes ($M > 6.5$) to train the network. Similarly, the trained network could have the potential to rapidly recognize the finite-fault mechanism and retrieve valuable information of the rupturing process such as the slip velocity, fault dimension, and rupture directivity⁴⁹⁻⁵¹ in real-time. The proposed methodology may provide an alternative way to largely accelerate the timeliness of solving these parameters and hence help predict the expected ground shakings more accurately for issuing an early warning.

Rather than other deep learning applications⁷⁻¹⁶ where numerous earthquakes are tested, the current FMNet is only evaluated on four earthquakes with magnitudes larger than 5.4. This results from the limitation of historical moderate-to-large earthquakes that occurred in the study area. Hundreds of earthquakes below M_w 5.4 could not be used for testing because of their low-quality signal in the used frequency band. These smaller earthquakes generally contain higher frequencies and it is challenging to model the high-frequency theoretical waveforms with a simple 1D velocity model. After testing, we find the conventional methods (such as gCAP) also fail to invert the source focal mechanisms with a 1D velocity model for these smaller earthquakes. With a finer 3D velocity structure and an efficient waveform modeling tool²³, our method shall be extended to smaller earthquakes as long as the synthetic training data can be well modeled to match the real data. Nonetheless, moderate-to-large earthquakes ($M_w > 5$) especially those that might cause destructive damages are our main target of this study. Besides, the reliability toward evaluating the FMNet output is very important for reporting source focal mechanisms. We have used the by-product of the encoder to interpret the working mechanism of the network and we find the network is explainable. We can employ the encoder analysis as introduced previously to monitor the effectiveness and robustness of the FMNet. We are further able to assess the predicted solutions of FMNet through the waveform comparison between the real and synthetic data in practice.

In the proposed method, the FMNet does not require the pre-knowledge of the location or depth of a real earthquake as long as it is within the monitoring area, since we have included all 3D grid locations in the training data. After testing, we find that it is currently challenging for the proposed FMNet to recognize the focal depth. This is mainly because the waveform differences concerning different focal depths are much smaller than concerning different source focal mechanisms. Variations of earthquake depth cause very minor changes in waveforms (Supplementary Fig. S5). The designed FMNet learns the general energy pattern in the waveforms, but it is currently not able to recognize the minor waveform differences to distinguish different focal depths. Determining the best focal depth via deep learning would require an independent effort and it is out of the scope in this study.

Method

FMNet architecture. The neural network we design is in the category of Fully Convolutional Network (FCN). FCN is a supervised deep learning network mainly based on convolutional layers but without fully connected layers and it has the merit of fewer model parameters and high computing efficiency⁵². Fig. 6 shows the architecture of the FMNet containing a bunch of Convolutional layers, MaxPooling layers, UpSampling layers, and also with some necessary operations such as LeakyRelu and BatchNormalization. There are 16 trainable layers. The input to the FMNet is a two-dimensional (2D) array representing the 3-C waveforms from 16 seismic stations. The output of the FMNet is three Gaussian probability distributions of the three angles of the focal mechanism. FMNet consists of two parts: a compression part that extracts the features of the input waveforms and an expansion part that transforms the extracted features to yield the output label of the focal mechanism.

In the compression part, the FMNet gradually compresses the input data from the size of 48 (three components by 16 stations) 128 (data length) to 1 1, by downsampling the input data layer by layer. At the same time, the number of filter channels gradually increases from 1 (channel of input) to 128. The data size has been changed from 1 (channel of input) 48 (three components by 16 stations) 128 (data length) to 128 (channels after compression) 1 1. The compression part of the FMNet can be regarded as an encoder process that compresses the data size by a factor of 48 times. The encoder is also exported as a by-product in this study for interpreting the FMNet. The encoder can take any waveform as input and output the extracted data features with the size of 128 (Supplementary Fig. S6). On the contrary, in the extension part, the FMNet gradually expands the extracted features by upsampling the features layer by layer. The upsampling operation is only carried out in the first dimension of the data, thus the size of the data in each channel is expanded from 1 1 to be 128 1. Meanwhile, the number of filter channels gradually decreases from 128 to 3. For the expansion part, the data size has been altered from 128 (channels after compression) 1 1 to 3 (channels) 128 (output length) 1. Each output channel has a size of 128 1, representing a 1D Gaussian distribution. All layers use the same configuration when employing convolutional and pooling operations. Filter sizes are 3 3 for the compression part and 3 1 for the expansion part since we only expand along the first dimension of data.

FMNet labeling and training parameters. We design the network as a regression problem. The training label is three Gaussian probability distributions, in which the maximum probability of each distribution corresponds to one component of the source focal mechanism (i.e. strike, dip, and rake). Thus the output training label has a size of 3 128 1 (Supplementary Fig. S7). The predicted focal mechanism of the real data can be retrieved by finding the peak values of the three output Gaussian probability distributions. This formalization of training labels greatly helps the convergence when training the network and the standard deviation of the Gaussian probability distribution affects the training convergence¹⁵. After testing, we find the standard deviation of 10 achieves a stable training convergence for our neural network and thus is used in this study.

Since we have designed the network as a regression problem, the Mean Square Error (MSE) option is chosen as the training loss function⁵³. For the training process, the Adam method⁵⁴ is tested to be effective in our FMNet and thus it is chosen as the optimizer, though other optimizers may also work decently. Totally 50 iterations with the batch size of 16 are implemented during the training. Besides, the learning rate is another crucial parameter that affects the level of final convergence⁵⁵. We test different learning rates and keeping other factors the same (Supplementary Fig. S8). The performance of each learning rate is evaluated by the final training loss after the same training iterations. After testing, we find the learning rate of 0.001 achieves the smallest convergence and is therefore used for the training in this study.

Declarations

Acknowledgments

Wenhuan Kuang thanks the helpful discussions from Prof. Mark Zoback at Stanford. The three-component (3-C) seismograms from 16 seismometers can be downloaded from the Southern Californian Seismic Network (SCSN) website. Open source platform of Keras is used to design the deep learning neural network. GMT and Matlab softwares are used to generate the figures.

Author contributions

Wenhuan Kuang designed the project, processed/analyzed the data, and wrote the manuscript. Congcong Yuan helped analyze the results and revise the manuscript. Jie Zhang helped revise the manuscript.

Competing financial interests

The authors declare no competing financial interests.

References

1. Kanamori, H., Hauksson, E. & Heaton, T. Real-time seismology and earthquake hazard mitigation. *Nature* **390**(6659), 461-464 (1997).
2. Allen, R. M. & Kanamori, H. The potential for earthquake early warning in southern California. *Science* **300**(5620), 786-789 (2003).
3. Kanamori, H. Real-time seismology and earthquake damage mitigation. *Annual Review of Earth and Planetary Sciences* **33**, 195-214 (2005).

4. Hinton, G. E., Osindero, S. & Teh, Y. W. A fast learning algorithm for deep belief nets. *Neural computation* **18**(7), 1527-1554 (2006).
5. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**(7553), 436 (2015).
6. Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M. & Dieleman, S. Mastering the game of Go with deep neural networks and tree search. *Nature* **529**(7587), 484 (2016).
7. Perol, T., Gharbi, M. & Denolle, M. Convolutional neural network for earthquake detection and location. *Science Advances* **4**(2), e1700578 (2018).
8. Meier, M. A., Ross, Z. E., Ramachandran, A., Balakrishna, A., Nair, S., Kundzicz, P., Li, Z., Andrews, J., Hauksson, E. & Yue, Y. Reliable Real-time Seismic Signal/Noise Discrimination with Machine Learning. *Journal of Geophysical Research: Solid Earth* **124**(1), 788-800 (2019).
9. Ross, Z. E., Meier, M. A. & Hauksson, E. P wave arrival picking and first-motion polarity determination with deep learning. *Journal of Geophysical Research: Solid Earth* **123**(6), 5120-5129 (2018).
10. Zhu, W. & Beroza, G. C. PhaseNet: a deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International* **216**(1), 261-273 (2018).
11. Mousavi, S.M., Ellsworth, W.L., Zhu, W., Chuang, L.Y. & Beroza, G.C. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications* **11**(1), pp.1-12 (2020).
12. Mousavi, S. M. & Beroza, G. C. A machine-learning approach for earthquake magnitude estimation. *Geophysical Research Letters* **47**(1), p.e2019GL085976 (2020).
13. Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J. & Johnson, P. A. Machine learning predicts laboratory earthquakes. *Geophysical Research Letters* **44**(18), 9276-9282 (2017).
14. DeVries, P. M., Viégas, F., Wattenberg, M. & Meade, B. J. Deep learning of aftershock patterns following large earthquakes. *Nature* **560**(7720), 632-634 (2018).
15. Holtzman, B. K., Paté, A., Paisley, J., Waldhauser, F. & Repetto, D. Machine learning reveals cyclic changes in seismic source spectra in Geysers geothermal field. *Science advances* **4**(5), eaao2929 (2018)
16. Zhang, X., Zhang, J., Yuan, C., Liu, S., Chen, Z., & Li, W. Locating induced earthquakes with a network of seismic stations in Oklahoma via a deep learning method. *Scientific reports* **10**(1), 1-12 (2020).
17. Kong, Q., Trugman, D. T., Ross, Z. E., Bianco, M. J., Meade, B. J., & Gerstoft, P. Machine learning in seismology: Turning data into insights. *Seismological Research Letters* **90**(1), 3-14 (2019).
18. Bergen, K. J., Johnson, P. A., Maarten, V., & Beroza, G. C. Machine learning for data-driven discovery in solid Earth geoscience. *Science* **363**(6433), eaau0323 (2019).
19. Moore, G. F., Bangs, N. L., Taira, A., Kuramoto, S., Pangborn, E. & Tobin, H. J. Three-dimensional splay fault geometry and implications for tsunami generation. *Science* **318**(5853), 1128-1131 (2007).
20. Ross, Z. E., Idini, B., Jia, Z., Stephenson, O. L., Zhong, M., Wang, X., Zhan, Z., Simons, M., Fielding, E. J., Yun, S. H. & Hauksson, E. Hierarchical interlocked orthogonal faulting in the 2019 Ridgecrest

- earthquake sequence. *Science* **366**(6463), 346-351 (2019).
21. Michael, A. J. Use of focal mechanisms to determine stress: a control study. *Journal of Geophysical Research: Solid Earth* **92**(B1), 357-368 (1987).
 22. Nanjo, K. Z. Were changes in stress state responsible for the 2019 Ridgecrest, California, earthquakes?. *Nature communications* **11**(1), e3082-e3082 (2020).
 23. Wang, X. & Zhan, Z. Seismotectonics and Fault Geometries of the 2019 Ridgecrest Sequence: Insight From Aftershock Moment Tensor Catalog Using 3-D Green's Functions. *Journal of Geophysical Research: Solid Earth* **125**(5), e2020JB019577 (2020).
 24. Kilb, D., Gomberg, J. & Bodin, P. Aftershock triggering by complete Coulomb stress changes. *Journal of Geophysical Research: Solid Earth* **107**(B4), ESE-2 (2002).
 25. Toda, S., Stein, R.S., Beroza, G.C. & Marsan, D. Aftershocks halted by static stress shadows. *Nature Geoscience* **5**(6), 410-413 (2012).
 26. Sumy, D.F., Cochran, E.S., Keranen, K.M., Wei, M. & Abers, G.A. Observations of static Coulomb stress triggering of the November 2011 M5. 7 Oklahoma earthquake sequence. *Journal of Geophysical Research: Solid Earth* **119**(3), 1904-1923 (2014).
 27. King, G.C., Stein, R.S. & Lin, J. Static stress changes and the triggering of earthquakes. *Bulletin of the Seismological Society of America* **84**(3), 935-953 (1994).
 28. Kilb, D., Gomberg, J. & Bodin, P. Triggering of earthquake aftershocks by dynamic stresses. *Nature*, **408**(6812), 570-574 (2000).
 29. Gomberg, J., Reasenber, P.A., Bodin, P.L. & Harris, R.A. Earthquake triggering by seismic waves following the Landers and Hector Mine earthquakes. *Nature*, **411**(6836), 462-466 (2001).
 30. Allen, R. M. & Melgar, D. Earthquake early warning: Advances, scientific challenges, and societal needs. *Annual Review of Earth and Planetary Sciences* **47**, 361-388 (2019).
 31. Minson, S. E., Meier, M. A., Baltay, A. S., Hanks, T. C. & Cochran, E. S. The limits of earthquake early warning: Timeliness of ground motion estimates. *Science advances* **4**(3), eaaq0504 (2018).
 32. Thakoor, K., Andrews, J., Hauksson, E. & Heaton, T. From Earthquake Source Parameters to Ground-Motion Warnings near You: The ShakeAlert Earthquake Information to Ground-Motion (eqInfo2GM) Method. *Seismological Research Letters* **90**(3), 1243-1257 (2019).
 33. Snoke, J. A. FOCMEC: Focal mechanism determinations. *International Handbook of Earthquake and Engineering Seismology* **85**, 1629-1630 (2003).
 34. Kilb, D. & Hardebeck, J. L. Fault parameter constraints using relocated earthquakes: A validation of first-motion focal-mechanism data. *Bulletin of the Seismological Society of America* **96**(3), 1140-1158 (2006).
 35. Hardebeck, J. L. & Shearer, P. M. Using S/P amplitude ratios to constrain the focal mechanisms of small earthquakes. *Bulletin of the Seismological Society of America* **93**(6), 2434-2444 (2003).
 36. Tan, Y. & Helmberger, D. A new method for determining small earthquake source parameters using short-period P waves. *Bulletin of the Seismological Society of America* **97**(4), 1176-1195(2007).

37. Zhao, L. S. & Helmberger, D. V. Source estimation from broadband regional seismograms. *Bulletin of the Seismological Society of America* **84**, 91–104 (1994).
38. Zhu, L. & Helmberger, D. V. Advancement in source estimation techniques using broadband regional seismograms. *Bulletin of the Seismological Society of America* **86**, 1634–1641 (1996).
39. Zhang, J., Zhang, H., Cheng, E., Zheng, Y., Kuang, W. & Zhang, X. Real-time earthquake monitoring using a search engine method. *Nature Communications* **5**, 5664 (2014).
40. Aki, K. & Richards, P. G. *Quantitative seismology* (2002).
41. Kamer, Y., Kissling, E., Ouillon, G. & Sornette, D. KaKiOS-16: A Probabilistic, Nonlinear, Absolute Location Catalog of the 1981–2011 Southern California Seismicity. *Bulletin of the Seismological Society of America* **107**(5), 1994-2007 (2017).
42. Zhu, L. & Rivera, L. A. A note on the dynamic and static displacements from a point source in multilayered media. *Geophysical Journal International* **148**(3), 619-627 (2002).
43. Chen, K., Avouac, J. P., Aati, S., Milliner, C., Zheng, F. & Shi, C. Cascading and pulse-like ruptures during the 2019 Ridgecrest earthquakes in the Eastern California Shear Zone. *Nature communications* **11**(1), 1-8 (2020).
44. Schroff, F., Kalenichenko, D. & Philbin, J. Facenet: A unified embedding for face recognition and clustering. *In Proceedings of the IEEE conference on computer vision and pattern recognition* 815-823 (2015).
45. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the inception architecture for computer vision. *In Proceedings of the IEEE conference on computer vision and pattern recognition* 2818-2826 (2016).
46. Motazedian, D. & Atkinson, G. M. Stochastic finite-fault modeling based on a dynamic corner frequency. *Bulletin of the Seismological Society of America* **95**(3), 995-1010 (2005).
47. Pacor, F., Cultrera, G., Mendez, A. & Cocco, M. Finite fault modeling of strong ground motions using a hybrid deterministic–stochastic approach. *Bulletin of the Seismological Society of America* **95**(1), 225-240 (2005).
48. Mai, P. M. & Thingbaijam, K. K. S. SRCMOD: An online database of finite-fault rupture models. *Seismological Research Letters* **85**(6), 1348-1357 (2014).
49. Bouchon, M. & Vallée, M. Observation of long supershear rupture during the magnitude 8.1 Kunlunshan earthquake. *Science* **301**(5634), 824-826 (2003).
50. Ammon, C. J., Ji, C., Thio, H. K., Robinson, D., Ni, S., Hjorleifsdottir, V., Kanamori, H., Lay, T., Das, S., Helmberger, D. & Ichinose, G. Rupture process of the 2004 Sumatra-Andaman earthquake. *Science* **308**(5725), 1133-1139 (2005).
51. Corradini, M., McBrearty, I., Rouet-Leduc, B. P., Hulbert, C., Satriano, C., Johnson, P. A. & Bernard, P. High-Frequency Radiation and Earthquake Rupture Complexities: From Back Projection to a Machine Learning Approach. *Geophysical Research Abstracts* **21** (2019).

52. Long, J., Shelhamer, E. & Darrell, T. Fully convolutional networks for semantic segmentation. *In Proceedings of the IEEE conference on computer vision and pattern recognition* 3431-3440 (2015).
53. Wang, Z. & Bovik, A.C. Mean squared error: Love it or leave it? A new look at signal fidelity measures. *IEEE signal processing magazine* 26(1), 98-117 (2009).
54. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint* arXiv:1412.6980 (2014).
55. Zeiler, M. D. ADADELTA: an adaptive learning rate method. *arXiv preprint* arXiv:1212.5701 (2012).

Figures

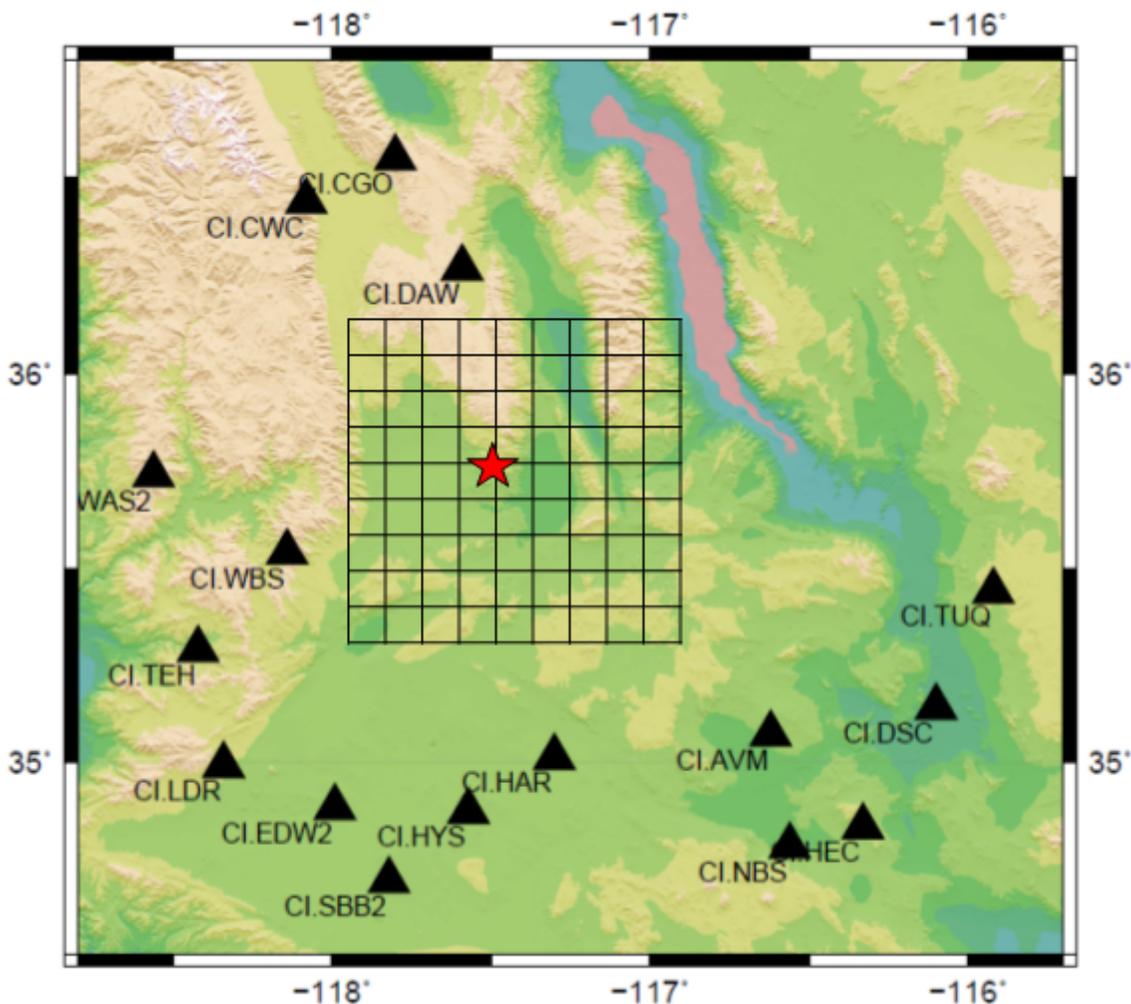


Figure 1

Grid discretization of the study area. The study area locates in the Ridgecrest region of southern California. The range of monitoring area is about 100 km by 100 km in both latitude and longitude directions. The 3D grid discretization has a depth range from 2 km to 20 km. 16 seismic stations (black triangles) within 150 km are used to model the 3-C synthetic training data. Red star denotes the mainshock in the Ridgecrest sequence.

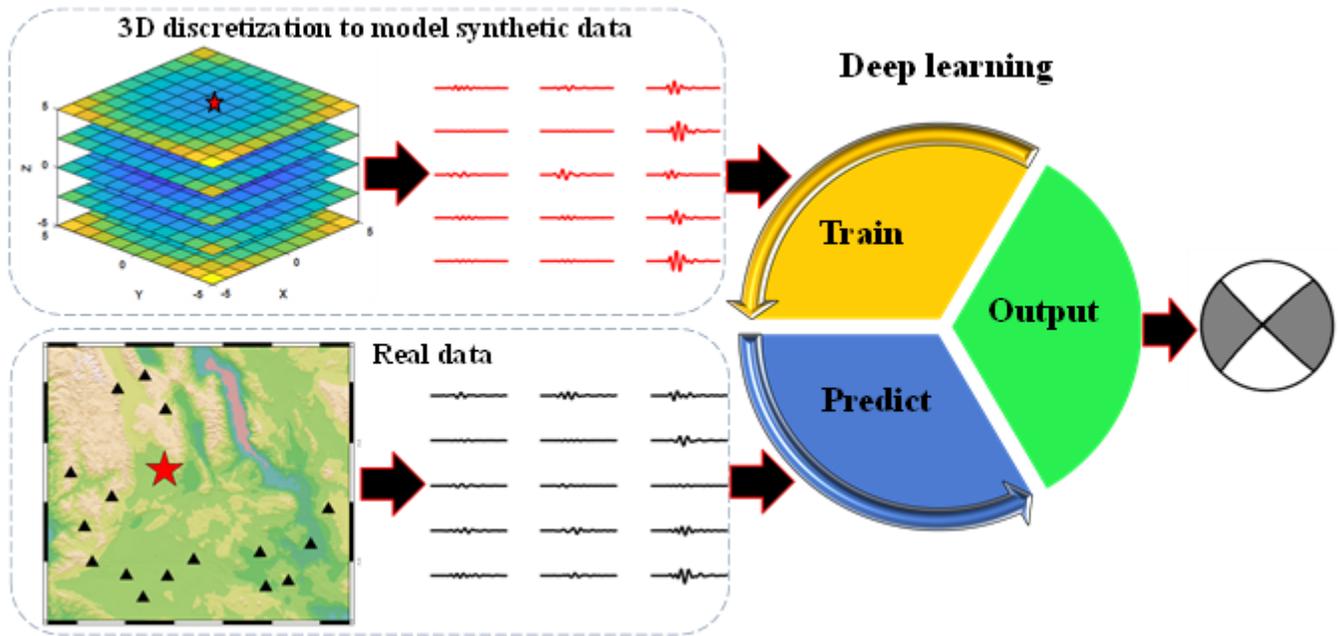


Figure 2

Schematic flowchart illustrating the framework of determining source focal mechanism via deep learning. First, we discretize the monitoring area of interest into 3D grids and simulate the theoretical waveforms as training data to train a designed FMNet. Then, when a real earthquake occurs, we directly feed the recorded waveforms into the well-trained FMNet and output the earthquake focal mechanism directly.

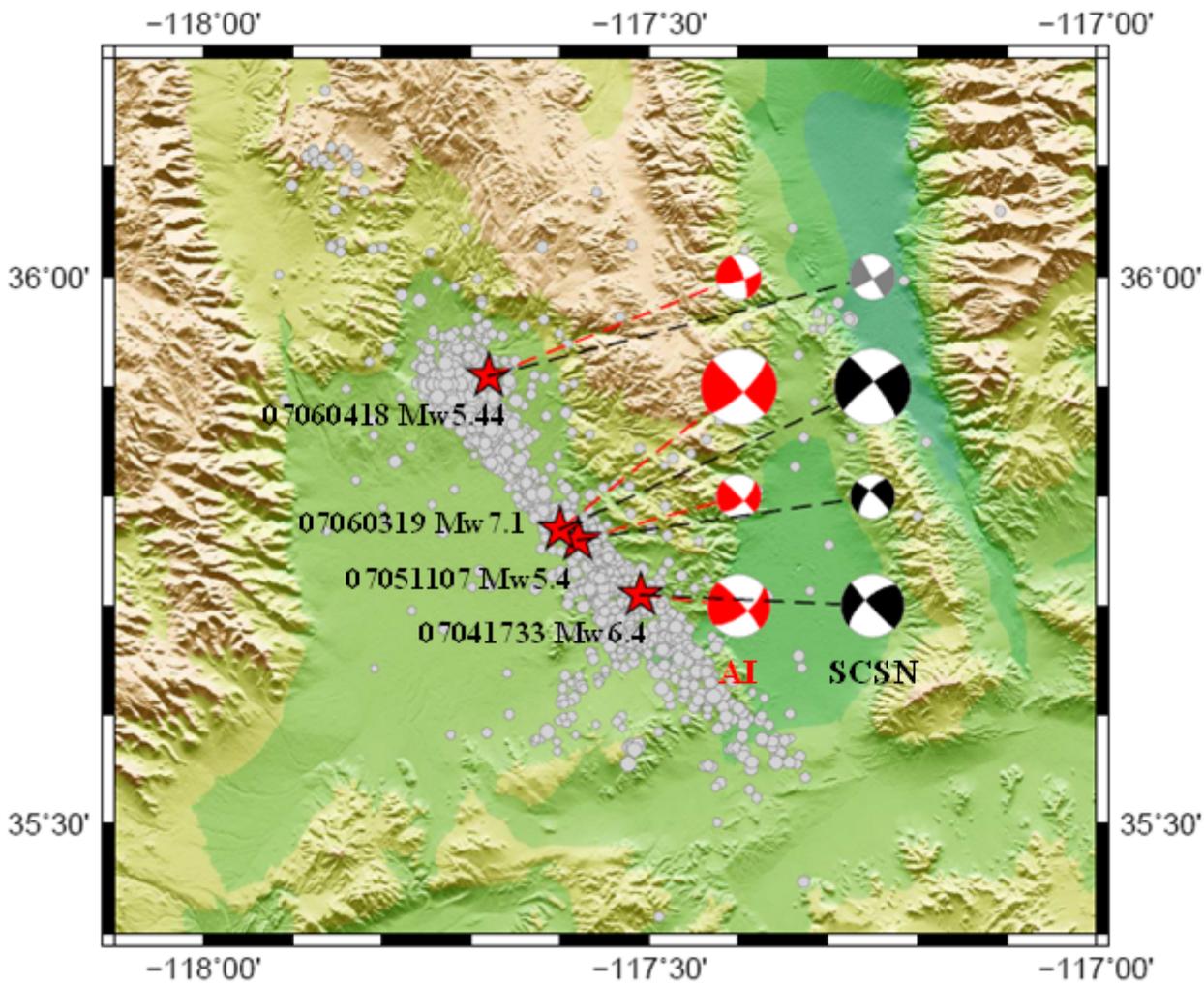


Figure 3

The application results to the July 2019 Ridgecrest sequence in southern California. Four large earthquakes with magnitudes larger than 5.4 (red stars), including the foreshock of Mw 6.4 and the mainshock of Mw 7.1, are tested. The determined focal mechanisms from FMNet are shown in red, and the reference solutions from SCSN moment tensor catalog are shown in black for comparison. The northernmost focal mechanism (grey) is inverted using gCAP method since it is not included in SCSN Moment Tensor Catalog. Grey dots show background seismicity.

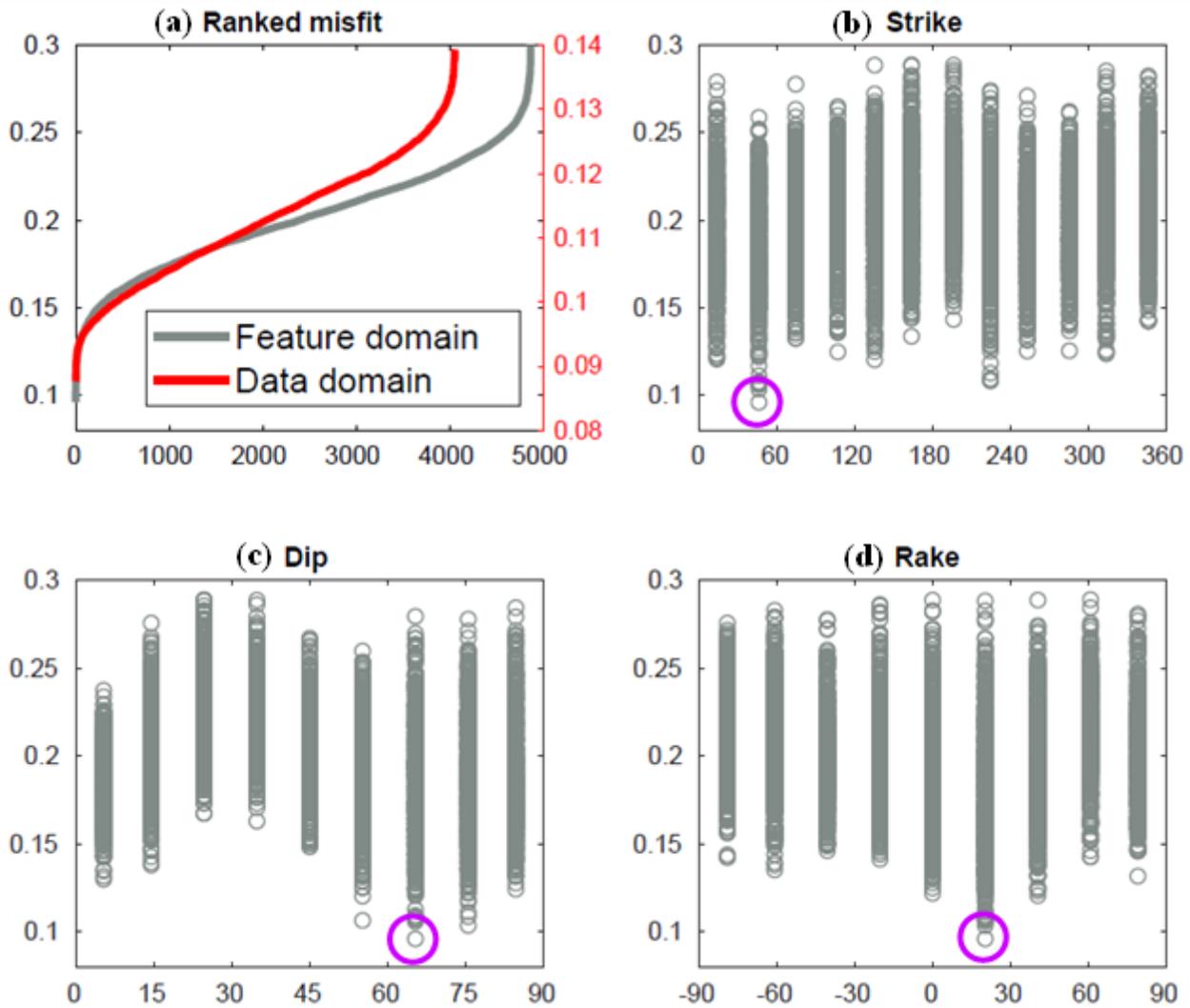


Figure 4

Interpreting the neural network using the encoder. a The comparison of the L2-norm misfit distributions that are calculated in data domain (in red) and feature domain (in black) for the Mw 6.4 foreshock, which are ranked in ascending order. For clarification, only the first 5,000 smallest misfits are plotted. b, c, and d The corresponding training labels represented in the discretized strike, dip, and rake angles, for the L2-norm misfit distribution in feature domain as shown in the black curve in a. The best solutions of the strike, dip, and rake angles retrieved by finding the smallest L2-norm misfit are highlighted in magenta circles.

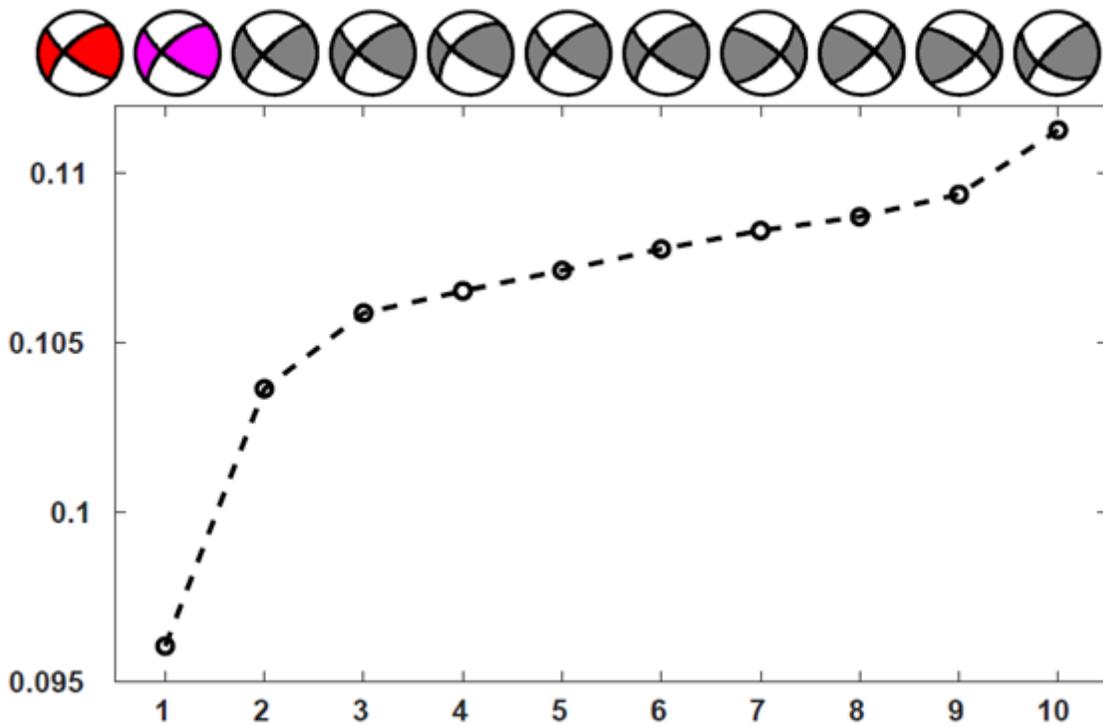


Figure 5

The comparison of best solutions retrieved in data domain (in red) and feature domain (in magenta). The 10 best solutions (in magenta and black) with the smallest L2 norm misfits retrieved in feature domain are also shown for comparison after ranked in ascending order.

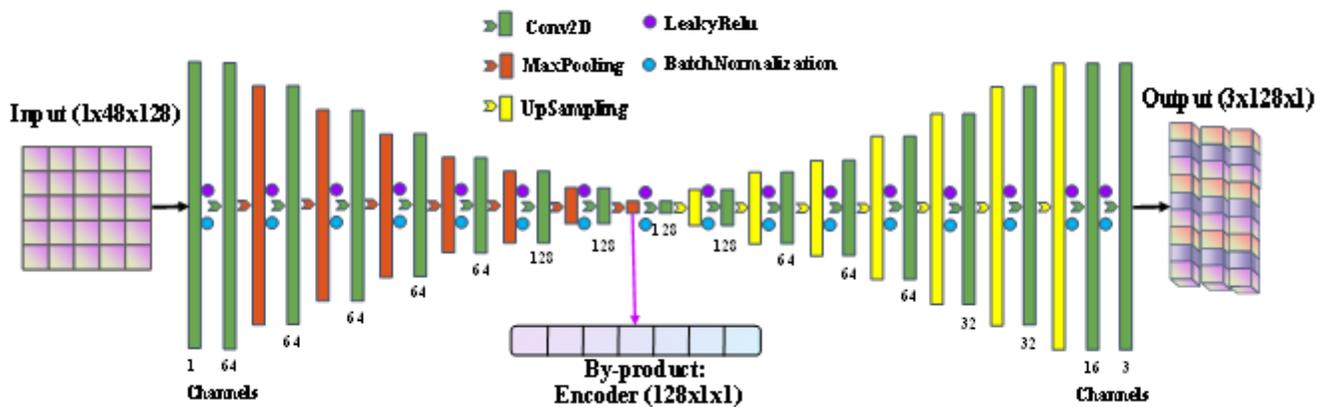


Figure 6

The FMNet architecture. The designed FMNet contains 16 trainable layers as well as MaxPooling, UpSampling, LeakyRelu, and BatchNormalization layers. The input data have the size of 1 (channel of

input) \times 48 (3-C waveforms of 16 seismic stations) \times 128 (trace length). The left part of the FMNet gradually compresses the input data from $1 \times 48 \times 128$ to $128 \times 1 \times 1$, and then the right part of the FMNet gradually expands the extracted features from $128 \times 1 \times 1$ to $3 \times 128 \times 1$ as the output of three Gaussian probability distribution representing the three angles of the focal mechanism. The intermediate layer of the encoder, with the size of $128 \times 1 \times 1$, is also exported as a by-product to interpret the neural network.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplement1.docx](#)