

Array-Based Convolutional Neural Networks for Automatic Detection and 4D Localization of Earthquakes in Hawai'i

Heather Shen¹ and Yang Shen^{*2}

Abstract

The growing amount of seismic data necessitates efficient and effective methods to monitor earthquakes. Current methods are computationally expensive, ineffective under noisy environments, or labor intensive. We leverage advances in machine learning to propose an improved solution, ArrayConvNet—a convolutional neural network that uses continuous array data from a seismic network to seamlessly detect and localize events, without the intermediate steps of phase detection, association, travel-time calculation, and inversion. When testing this methodology with events at Hawai'i, we achieve 99.4% accuracy and predict hypocenter locations within a few kilometers of the U.S. Geological Survey catalog. We demonstrate that training with relocated earthquakes reduces localization errors significantly. We outline several ways to improve the model, including enhanced data augmentation and use of relocated offshore earthquakes recorded by ocean-bottom seismometers. Application to continuous records shows that our algorithm detects 690% as many earthquakes as the published catalog, and 125% as many events than the Hawaiian Volcano Observatory internal catalog. Because of the enhanced detection sensitivity, localization granularity, and minimal computation costs, our solution is valuable, particularly for real-time earthquake monitoring.

Cite this article as Shen, H., and Y. Shen (2021). Array-Based Convolutional Neural Networks for Automatic Detection and 4D Localization of Earthquakes in Hawai'i, *Seismol. Res. Lett.* **XX**, 1–11, doi: [10.1785/0220200419](https://doi.org/10.1785/0220200419).

Supplemental Material

Introduction

Recent advances in instrumentation have provided an exponential increase in seismic data. Yet, detecting and localizing earthquakes at scale remains expensive and inefficient. Traditional earthquake detection methods used by many seismic network operators (e.g., Allen, 1982; Withers *et al.*, 1998) do not perform well for small earthquakes in noisy environments. In addition, network operations often involve human review of earthquake arrivals and time picks as well as iterative tuning of hypocenter estimates. To improve detection, methods based on waveform similarity (matched filter or template matching) have been developed and widely applied (e.g., Gibbons and Ringdal, 2006; Caffagni *et al.*, 2016). Such efforts have led to a great increase in the detection of small earthquakes, yielding rich details that enable the next generation of analyses of earthquakes and faults (e.g., Ross *et al.*, 2019). These methods are, however, computationally expensive and limited; detection only works for earthquakes that share similar waveforms and thus likely have the same source regions and mechanisms of the template events.

In the past few years, convolutional neural networks (CNNs) have been adapted for earthquake detection and location. One common feature shared by CNN approaches is that

once the model is trained, it is far more computationally efficient than the waveform-similarity-based approach (Gibbons and Ringdal, 2006; Yoon *et al.*, 2015) when it is applied to new data, an advantage important for seismic network operations, particularly during periods of intense seismic activities. Perol *et al.* (2018) introduced a CNN model for earthquake detection and localization based on waveforms at individual stations. The localization was limited to a few subregions. Lomax *et al.* (2019) and Mousavi and Beroza (2020) developed CNN models for rapid earthquake characterization using single-station waveforms. Dokht *et al.* (2019) extended the CNN earthquake detection in the time-frequency domain. Other studies focused on seismic event or phase detection and picking of arrival times, which were then used in traditional travel-time-based localization (e.g., Kong *et al.*, 2016; Ross *et al.*, 2018; Zhu and Beroza, 2018; Wang *et al.*, 2019; Zhou *et al.*, 2019; Zhu *et al.*, 2019; Johnson *et al.*, 2020; Walter *et al.*, 2020).

1. East Greenwich, Rhode Island, U.S.A., <https://orcid.org/0000-0002-3101-5753> (HS); 2. Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island, U.S.A., <https://orcid.org/0000-0002-4185-6089> (YS)

*Corresponding author: yshen@uri.edu

© Seismological Society of America

Kriegerowski *et al.* (2019) and Zhang, Zhang, Yuan, *et al.* (2020) showed it is possible to use CNNs to locate earthquakes without the intermediate step of phase picking and association; however, the former depended on manually chosen arrival times at a reference station, and the latter assumed that seismic events had already been detected. Zhang, Zhang and Tian (2020) extended the full convolutional network model of Zhang, Zhang, Yuan, *et al.* (2020) to continuous data for earthquake early warning, though the event origin time was not a direct output of the deep learning model. It was determined instead by picking the first *P* phase and calculating the travel time from the source location to the station location. Taking a different approach, van den Ende and Ampuero (2020) used graph neural networks with multistation waveforms to locate earthquakes and estimate magnitude, though they too applied it only to existing catalog events. Here, we present a framework based on recent advances in deep learning for seamless, automatic detection and 4D localization of earthquakes for continuous data from a seismic network without the intermediate steps of phase detection and picking, phase association, travel-time calculation, and inversion. Our approach builds upon previous work by using a network of seismic stations to first identify if an earthquake has occurred, and if so, estimate the latitude, longitude, depth, and origin time of the event.

Specifically, we propose a two-stage seismic-array-based, CNN (ArrayConvNet) model where: (1) earthquake detection becomes a supervised classification problem, and (2) earthquake localization becomes a supervised regression problem. We train and test on data from 55 seismic stations on the Island of Hawai'i—our solution not only detects earthquakes in the U.S. Geological Survey (USGS) catalog, but also uncovers six times as many earthquakes as the published catalog and 25%–61% more events than in the Hawaiian Volcano Observatory (HVO) internal catalog. Once an earthquake is detected, our model can locate an earthquake's hypocenter to within 3–4 km of the catalog. To the best of our knowledge, this is the first deep learning model that can automatically provide an earthquake catalog from the continuous data of a seismic network. The deep learning model is highly computationally efficient, without phase picking, association, travel-time calculation, and inversion. Finally, we outline several steps that can be taken to greatly reduce the model localization errors, making it a viable solution to improve the efficiency and accuracy of seismic monitoring at much lower computational and human costs.

Data

The Island of Hawai'i, U.S.A. is one of the most seismically and volcanically active regions in the world, a fact that was heightened by the 2018 eruption of Kilauea Volcano (Neal *et al.*, 2019). HVO operates a permanent seismic network (the network code HV) on the island, providing the earthquake information and waveform data needed for this study. We

use 55 seismic stations on the Island of Hawai'i (Fig. 1), including stations in the HV and PT networks. Among them, 33 have three-component (north, east, and vertical) seismometers, while the rest have single, vertical-component seismometers.

Both the earthquake waveforms and noise segments are downloaded from the Incorporated Research Institutions for Seismology Data Management Center. For each earthquake, a 50 s window is selected, and the trace start time is set randomly between 1–10 s before the event origin time. The time difference between the trace start time and the event origin time, along with the catalog hypocentral location (latitude, longitude, and depth) are used to train the localization part of the model (see the [Method](#) section). The noise segments are chosen between the USGS reported earthquakes and are 10–60 s before the origin time of an earthquake. We visually inspect all noise windows and discard those with the possible presence of unreported earthquakes in the segments. Because the Hawaiian seismic networks have a variety of short-period and broadband sensors, we remove instrument response from the traces and transfer them to velocity seismograms to minimize the effects of different instrument sensitivities to ground motion. The earthquake waveforms and noise are filtered between 3 and 20 Hz and downsampled to a uniform sampling rate of 50 samples per second on all channels. The frequency range is chosen for optimum earthquake signal-to-noise ratios (SNRs) based on visual inspection of earthquake waveforms over a wide range of frequencies as well as previous studies of the Hawaiian earthquake characteristics (e.g., Matoza *et al.*, 2014). All traces are normalized individually before they are used as the inputs for the CNN model. For stations with missing records or that do not have three channels, we zero-fill the missing channels.

Our ArrayConvNet model has two stages: (1) event detection and (2) event localization (see the [Method](#) section). Each stage is trained on distinct training and test data sets.

Detection training and test data

We use 1843 analyst-reviewed earthquakes (Fig. 1) in the 2017 USGS Combined Catalog (ComCat, hereinafter the USGS published catalog or published catalog) and 1905 noise segments. The number of earthquakes is comparable to that in Perol *et al.* (2018). The magnitude of the selected events ranges from 0.1 to 5.28 (ml or md). We note that HVO maintains an internal earthquake catalog (hereinafter the HVO internal catalog or internal catalog). The unpublished events are generally smaller and recorded by fewer stations (see more discussion in the [Application to Continuous Hawaiian Seismic Data](#) section). For network training and testing, we avoid automatically determined earthquakes and use only the high-quality, analyst-reviewed events in the published catalog.

We explored several ways of arranging the input trace data for the detection model and chose the following approach based on the robustness of the results when the model is

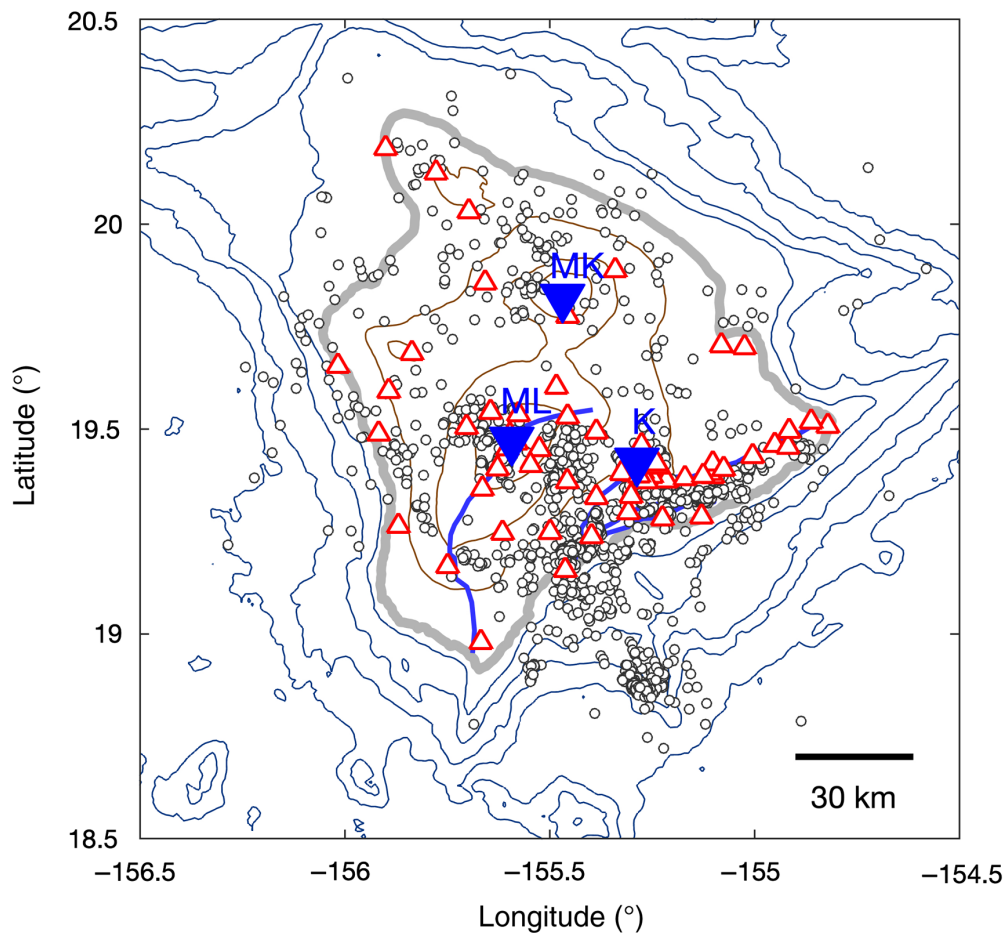


Figure 1. Map of the seismic stations (red triangles) and earthquakes (small circles) used in the study. The coast of Island of Hawai'i is outlined by the gray contour, while topography and bathymetry are contoured in 1000 m intervals. Blue lines represent the major Quaternary faults and fault systems (Cannon *et al.*, 2007). The inverted triangles and ML, K and MK stand for Mauna Loa, Kilauea, and Mauna Kea Volcanoes, respectively. The color version of this figure is available only in the electronic edition.

applied on unseen, continuous data. For both earthquakes and noise segments, we sort the 55 station traces in order of the time of the vertical component's largest amplitude and take the absolute of the traces so all values are between 0 and 1. Therefore, for an earthquake, we see a clear propagation of earthquake arrivals through the seismic network in an easily recognizable visual pattern (Fig. S1, available in the supplemental material to this article). The general pattern is consistent from earthquake to earthquake, regardless of location and magnitude (and thus SNR), as the wave always propagates from the lower left to the upper right in the maximum-amplitude-sorted waveform images (Fig. S1). For each sorted station, the cross-station features (e.g., a strong *P* or *S* arrival) are adjacent or local in the time-and-trace-number space in a well-defined trend. This is easier to learn for the convolutional kernels, which often have a small size for computational efficiency. In contrast, the unsorted waveforms arranged alphabetically by

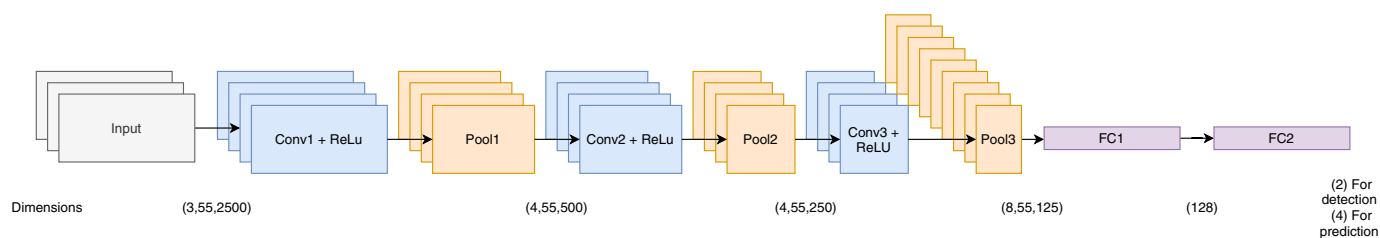
station names do not have an easy-to-follow pattern from event to event. Depending on the source–receiver geometry, a station that has an early earthquake arrival for one event may have a late arrival for the next event. The cross-station features are highly variable and may span the entire time-and-trace-number space, thus requiring a deep and large network to capture. Following the general practice in machine learning, we define precision = true positive/(true positive + false positive) and recall = true positive/(true positive + false negative). Although the sorted and unsorted waveforms do not show substantial differences in terms of model precision, recall, and the receiver operating characteristic curve (Fig. S2), they yield significantly different numbers of detections when the resulting model is applied to continuous data, indicating differences in the robustness of detection of small earthquakes in noisy data; the unsorted waveforms result in higher number of false detection from visual inspection of the corresponding seismic traces.

Sorting by the maximum amplitude is an imperfect proxy of the source–station distance as variations in focal mechanisms and local scattering, which are often unknown or poorly constrained for small earthquakes and in a monitoring setting, may affect the amplitude. However, the effect of this imperfect sorting on event detection is likely minor as sorting only serves the purpose of organizing the traces in an orderly fashion. Intuitively, the sorted waveforms have simpler, lower-order features, which require less-complicated neural networks and thus less training data to achieve robust models.

Each input event is labeled with a “0” or a “1” to indicate whether it is a noise or an earthquake event, respectively. This now transforms detection into a well-understood classification problem.

Localization training and test data

For the localization part of the model, we use the earthquake locations and origin times from the same 1843 earthquakes in



the 2017 published USGS catalog. We augment these original waveforms by performing seven cuts of 50 s long waveforms for the same earthquake, with each cut starting at a random time between 1 and 10 s before the event origin time. The seven cuts of each earthquake have the same hypocenter (latitude, longitude, and depth) but different offsets between the trace start time and the origin time. The total number of earthquakes used to train and test the localization part is thus 12,901 (7×1843). Such data augmentation is commonly used in deep learning (Wang and Perez, 2017) and, in our case, helps to train the model to better localize the event origin time (see more in the Discussion section), which is crucial when the model is applied to continuous data.

Given the different units and scales for the hypocenter and origin time, we normalize the latitude, longitude, depth, and time values so that they are all comparable in magnitude (within -1 to 1). For the hypocenter, we subtract a reference location (latitude 19.5° , longitude -155.5° , and depth 0 km) from the catalog location and then divide the depth by 50; for the time difference between the trace start time and the origin time, we divide by 10. Therefore, we avoid the situation where one variable (e.g., depth) dominates the loss function.

Unlike the inputs for the detection part of the model, the input traces for localization are arranged alphabetically by station names. This is necessary as localization requires that the station geometry remains a constant. We now may treat localization as a supervised regression problem.

Method

Network architecture

Traces for each event are arranged as a 3D tensor $Z(c, s, t)$. The depths of Z for $c \in \{1, 2, 3\}$ correspond to three channels of seismic records, the rows for $s \in \{1, \dots, 55\}$ represent various stations, and $t \in \{1, \dots, 2500\}$ represents the time index of trace values. Inputs are then processed in a feed-forward stack of three convolutional layers, followed by two fully connected layers that in the detection model, output class scores and in the localization model, output latitude, longitude, depth, and time offset between the trace start time and earthquake origin time (Fig. 2).

After each convolutional layer, we use a rectified linear unit layer to apply an element-wise activation function and then a max pooling layer to perform a downsampling operation and decrease the number of parameters. Convolutions are also zero-padded to maintain input shape.

Figure 2. Architecture of array-based convolutional neural network (CNN) model. Conv and ReLu stand for the convolutional layer and rectified linear unit layer, whereas Pool represents max pooling. FC1 and FC2 are the two fully connected layers. Numbers within the parenthesis represent the dimensions of the input or output data at the various stages. The color version of this figure is available only in the electronic edition.

The kernel of the first convolutional layer has a dimension of width of 9 and height of 1. The kernels in the second and third layers have the same dimension of width of 3 and height of 5. The motivation behind the 1D filter in the first layer is to isolate learning of temporal features among the three input channels of each station, as in Kriegerowski *et al.* (2019), whereas the 2D filters in later layers are designed to extract cross-station information. Pooling after the first convolutional layer has a size of (1,5) with a stride of (1,5), while pooling after the second and third convolutional layers has a size of (1,2) with a stride of (1,2). Thus, pooling in our model is designed primarily to downsample in the time dimension.

We note that our number of convolutional layers (3) and the number of channels in each layer (4, 4, and 8) are substantially smaller than in previous studies (e.g., eight convolutional layers with 32 channels each in Perol *et al.*, 2018). To determine the optimal network architecture, we explored a range of the number of convolutional layers (2–5), number of channels (2–32), and number of features or neurons of the first fully connected layer (64–1024). Our guiding principle in selecting the optimum models is to find the smallest network that yields better or comparable results in detection precision. Fewer than three convolutional layers and smaller than four channels per convolutional layer yield lower precision, as the model may be too simple to capture the full complexity of the data. Greater than or equal to four convolutional layers, larger than eight channels, and larger than 128 neurons in the first fully connected layer yield detection precision comparable to that of our preferred network, with the training loss far below (in most cases more than an order of magnitude smaller than) the test loss, which suggests overfitting.

Training the network

The two parts of the CNN model can be trained separately and then connected for examining continuous data. For detection,

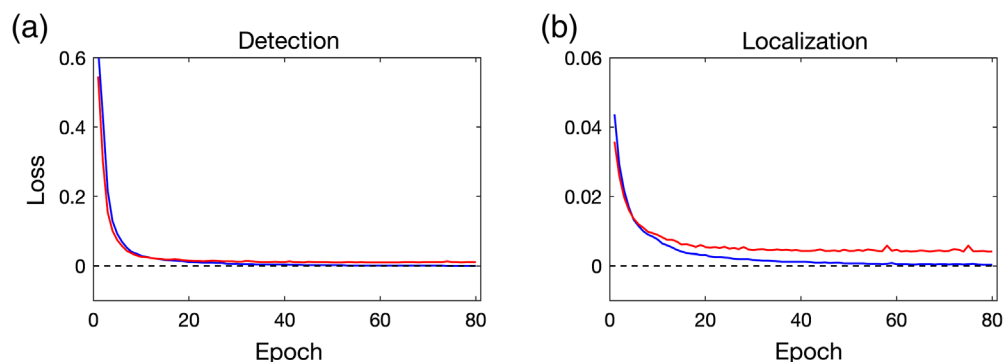


Figure 3. Training (blue line) and test (red line) losses as the function of epochs for the (a) detection and (b) localization parts of the model. It takes fewer than 20 epochs for both detection and localization for the loss to decrease rapidly and for the test loss to approach a small and relatively stable value. The dashed line marks zero loss. The color version of this figure is available only in the electronic edition.

we optimize the network parameters by minimizing a cross-entropy loss function. This measures the average discrepancy between our predicted distribution and the true class probability distribution in the training set and is proven for standard classification problems (e.g., [Perol et al., 2018](#)). For localization, we optimize the parameters by minimizing a mean-squared error loss between our predicted and given location and time values.

Given our training data set, we are able to minimize our loss functions using a batch approach. We use a typical 75%–25% split for the training and test data sets, respectively. At each training step, we feed a batch of 32 inputs to the network, evaluate the expected loss on the batch, and update the network parameters accordingly using backpropagation. We cycle through all training data in batches as an epoch, and after each epoch, we calculate the loss for both the training and test data sets. This is repeated until the loss stops decreasing significantly (80 epochs for both detection and localization, Fig. 3).

For optimization, we used the AdamW algorithm ([Loshchilov and Hutter, 2017](#)), which builds on the well-known Adam algorithm ([Kingma and Ba, 2014](#)) but separates the weight decay from the learning rate. The result of this distinction is that the weight decay and learning rate can be optimized separately and has been proven to substantially improve generalization performance. For detection, we use the default learning rate, 2×10^{-5} ; for localization, given the increase in training data due to augmentation, we use a larger learning rate of 5×10^{-5} .

Computational implementation

We implemented our ArrayConvNet model in Pytorch ([Paszke et al., 2019](#)) and performed all model training, testing, and application to continuous data on an iMac with a 3.8 GHz 8-core Intel Core i7 central processing unit and 128 GB memory. Model training and testing in 80 epochs took about

1.3 and 5.7 hr for the detection and localization parts of the model, respectively. Application of the model to 90-day continuous seismic data took about 16.5 hr (see the [Application to Continuous Hawaiian Seismic Data](#) section).

Results

Detection

Within 20 epochs, both the training and test losses decrease rapidly and the test loss remains small and relatively stable as the number of epochs increases (Fig. 3). For comparison, [Perol et al. \(2018\)](#) used 32,000 epochs to train their model. In [Dokht](#)

[et al. \(2019\)](#), it took over 10,000 epochs for the earthquake detection learning to approach an asymptotic and stable flatline. We attribute the rapid learning of our model to, at least partially, the relative simplicity of the network architecture.

Our detection accuracy on the test data, defined as the percentage of events that are correctly classified as an earthquake or noise, is 99.4% at 0.5 classification (probability) threshold for earthquakes. Between 0.5 and 0.7 classification threshold, the precision is 99.6% while recall is 99.2%–99.0% (Fig. 4 and Fig. S2). Above 0.7 classification threshold, the precision is 100% while recall is 99%–98%, suggesting that above this detection threshold ArrayConvNet does not label any noise as earthquakes, at least in the test data, and rarely misclassifies earthquakes as noise. For comparison, the precision and recall reported by [Perol et al. \(2018\)](#) are 94.8% and 100%, respectively, and those by [Dokht et al. \(2019\)](#) are 99.6% and 99.9%, respectively.

Localization

Similar to the detection part of the model, the training and test losses of the localization part of the model decrease rapidly within 20 epochs. While the training loss continues to decrease toward zero with increasing epochs, the test loss remains flat-lined (Fig. 3), suggesting that the network has enough neurons or complexity to fit the training data nearly completely, but uncertainty or random noise in the data keeps the test loss at a certain level; more epochs or a larger network likely would not improve the fit of the test data.

Overall, our model is able to predict the location of an earthquake in the test data within 0.08 ± 4.5 km in the north–south direction, 0.07 ± 4.1 km in the east–west direction, and -0.02 ± 3.5 km in depth (Fig. 5 and Fig. S3). The values following the $\pm$ sign (and hereinafter) represent one standard deviation. The difference between the predicted

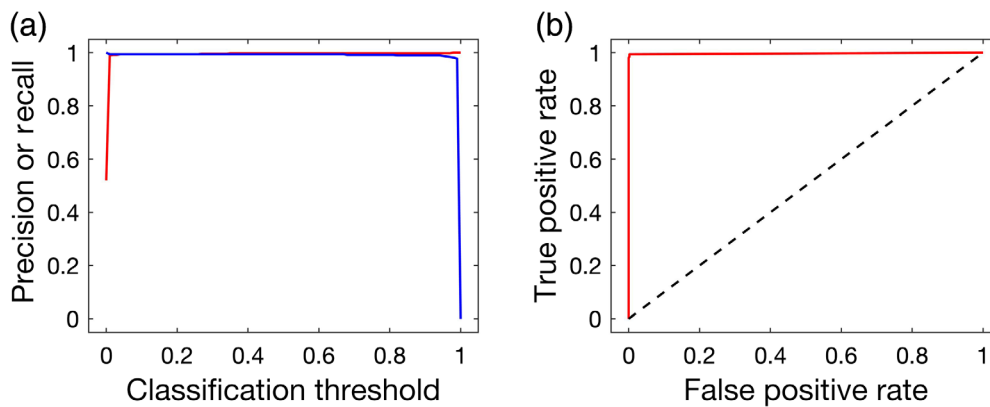


Figure 4. (a) Precision (red) and recall (blue) as a function of classification threshold for the CNN model using the maximum-amplitude-sorted waveforms. (b) Receiver operating characteristics curve (red line) for the model. The dashed line is for a model with no predicting skill. The color version of this figure is available only in the electronic edition.

and catalog origin times is -0.06 ± 0.81 s. Some of the location and origin-time differences may be attributed to errors in the USGS catalog. Synthetic tests by Zhang, Zhang, Yuan, *et al.* (2020) show that adding a location error to the catalog location results in their CNN model prediction error of a similar size. Lin *et al.* (2014) relocated earthquakes with magnitude > 1.0 between 1992 and 2009, using a 3D velocity model and source-specific station term corrections. Their resulting catalog thus represents a subset of the HVO events with the best location quality (Lin *et al.*, 2014). Comparing the earthquake locations in Lin *et al.* (2014) with the USGS catalog locations, we find a lateral location offset of 1.1 ± 1.8 km and a depth offset of 1.0 ± 2.1 km. Hence, a significant portion of the hypocenter location differences between our model predictions and the USGS catalog may stem from errors in the training data (see more in the Discussion section).

Application to Continuous Hawaiian Seismic Data

Earthquake catalogs usually represent a subset of earthquakes that occurred, with detection and localization limited by SNRs in seismic records, number of detected stations, and other factors. The USGS catalog for Hawai'i is no exception. Although our ArrayConvNet performs well for the test data set (Fig. 4), further tests on continuous data, combined with expert reviews of the results, are required to evaluate its true efficacy.

For seismic network operators generating earthquake catalogs, one may wish to minimize false detection by using a higher confidence threshold (Ross *et al.*, 2018). Here, we follow this approach using a probability threshold of 0.95 (95% confidence) in the following discussion unless otherwise stated. Based on the precision and recall characteristics (Fig. 4), the model should rarely misclassify earthquakes as noise, and almost never identify noise as an earthquake at this confidence level.

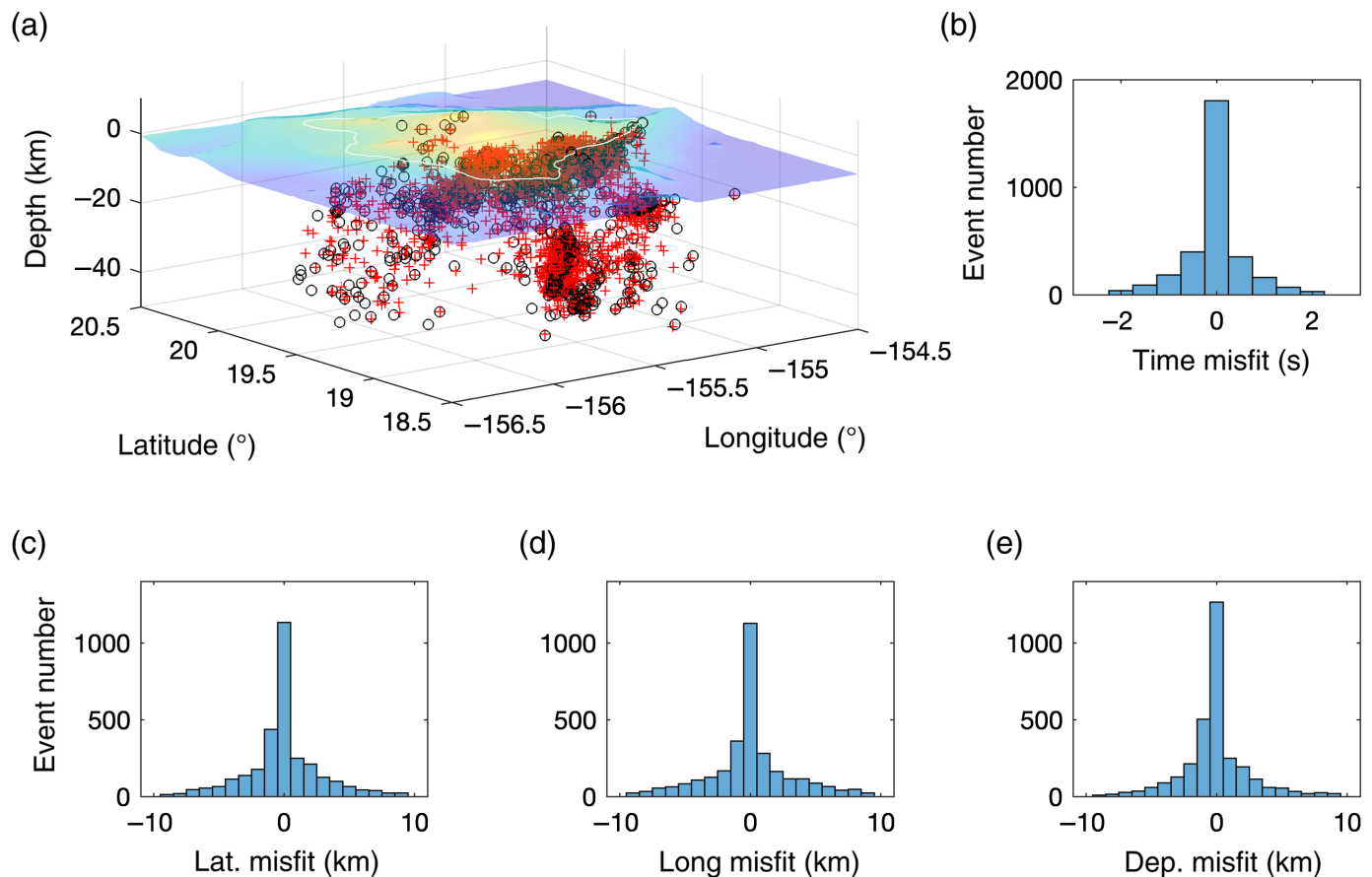
We input continuous seismic data from the same 55 seismic stations in Hawai'i, which are unseen in the development of our CNN model and preprocessed in the same way as the data used to train the models. The model runs through the data in 50 s long moving windows at 3 s increments. When the detection stage of the model finds that the probability of an earthquake is above a specified confidence threshold, we determine the exact 50 s window for localization by choosing the one that has the highest detection probability from the contiguous windows

above the threshold. We then feed the chosen window to the localization stage and calculate the event location. To be consistent with the localization training data, where traces start 1–10 s before the origin time, the declared event must also have a predicted origin time within 1–10 s after the start of the traces (Fig. 6).

Using a continuous data stream from January to March 2018, our model detects and locates 5547 earthquakes, which is approximately 6.9 times the number reported in the published USGS catalog (807). The HVO internal catalog lists 3440 events with magnitude > 0.1 and 4434 events for all located events for the same period (B. Shiro, personal comm., 2020). As a measure of the sensitivity to the detection probability threshold, the model detects and locates 5943 and 5042 earthquakes with the probability thresholds of 0.68 and 0.997, respectively.

There is a weak correlation between the daily event numbers in the published USGS catalog and our model detection (Fig. 7). The correlation is substantially higher (correlation coefficient 0.68) between our model result and the HVO internal (all-event) catalog. On relatively quiet days when the daily event counts are 20–30, the number of earthquakes in our model is similar to that in the HVO internal catalog. On days with more intense activities, our model detects and locates substantially more events than in both the published and internal catalogs. For example, on 25 January 2018, our model yields 194 events versus 84 events in the internal catalog and 21 events in the published catalog. The events detected and localized by our model have a similar epicentral distribution as those of the HVO internal catalog events from January to March 2018 (Fig. S4). The positive correlation in the daily event numbers and similarity in the event distributions validate the overall effectiveness of our model.

Visual inspection of the events in our model in the first five days of 2018 verified that all 254 events are earthquakes.



In addition, we found that most but not all of the catalog events are recovered by our model, consistent with the precision and recall characteristics at the 0.95 threshold (Fig. 4), where the model may rarely misclassify earthquakes as noise, but almost never identify noise as an earthquake. On 10 out of 90 days, our model contains fewer events than the HVO internal (all-event) catalog. The missing ones are mostly low magnitudes (<0.7) and have low numbers of reporting stations (<10) in the catalog. For example, on 22 March 2018 the majority of the events listed in the HVO internal catalog (78 out of 83) have magnitudes lower than 1.7, the nominal magnitude cutoff for events included in the published USGS catalog (B. Shiro, personal comm., 2020). Statistically, earthquakes with magnitude greater than or equal to 2.8 in the published and internal catalogs are 100% recovered by our ArrayConvNet model (Fig. S6). At the minimum magnitude of 1.7, the event recovery rate is 90% and 88% for the published and internal catalogs, respectively. At the minimum magnitude of 0.7, the event recovery rate is 88% for the published catalog and 75% for the internal catalog.

Discussions

As with any supervised machine learning, the more accurate and greater the training data, the better the resulting model. In our case, the training data can be improved in several ways.

Figure 5. (a) Hypocenter locations of earthquakes from the U.S. Geological Survey (USGS) catalog (black circles) and ArrayConvNet model predictions (red crosses) in a 3D view looking from the southwest direction. Clusters of earthquakes in the catalog and model predictions are clearly visible. The topography and bathymetry of the island are shown as a semi-transparent surface. The coastline is marked by the white line. (b–e) The histograms of the origin time, latitude, longitude, and depth differences between the USGS catalog and our model predictions, respectively. The color version of this figure is available only in the electronic edition.

The first way is to include the catalog earthquakes from the many years of monitoring by HVO. We trained our model using only the earthquakes from the published USGS catalog in 2017. The HVO internal catalog contains several times more earthquakes, most of them are smaller earthquakes with magnitudes lower than 1.7. A greater number and a wider magnitude range of earthquakes plus a correspondingly large number of noise (visually inspected or automatically screened to minimize the presence of earthquakes in the noise segments and ensure the quality of earthquakes in the catalog) should further improve the accuracy and robustness of the model.

The second way is to use relocated earthquakes with more accurate locations (e.g., [Got and Okubo, 2003](#); [Wolfe et al.,](#)

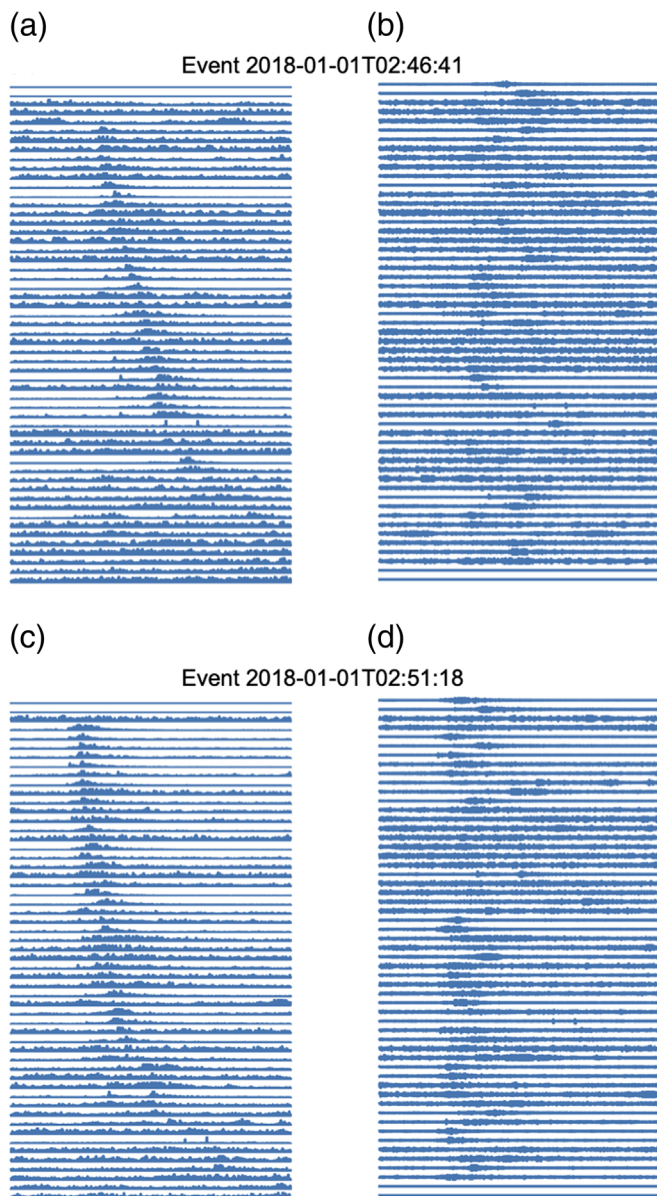


Figure 6. Two examples of earthquakes detected using the ArrayConvNet model in the first hours of 2018, which are not in the published or internal catalogs. The (a) maximum-amplitude-sorted and (b) unsorted vertical-component waveforms for an event occurred a few seconds after 2018-01-01T02:46:41 (1 January 2018) (the trace start time). The sorted traces are arranged from the top to bottom and are taken as absolute values. The total window length is 50 s. Panels (c) and (d) are the same as (a) and (b) for an event shortly after 2018-01-01T02:51:18. The color version of this figure is available only in the electronic edition.

2004; Matoza *et al.*, 2013; Lin *et al.*, 2014; Shelly and Thelen, 2019). Lin and Okubo (2020) relocated over 48,000 earthquakes between July 2015 and August 2018. With the caveat that all the relocated events in Lin and Okubo (2020) are onshore (Fig. S5), we found that using 1806 earthquakes in

the year 2017 relocated by Lin and Okubo (2020) to train localization in the same way as we discussed earlier in this article for the USGS catalog reduces the location difference between the model prediction and the Lin and Okubo catalog by 25%–45% (from ± 4.5 to ± 2.4 km in the north–south direction, from ± 4.1 to ± 2.5 km in the east–west direction, and from ± 3.5 to ± 2.6 km in depth) and the origin-time difference by 13% (from ± 0.8 to ± 0.7 s), demonstrating the effects of relocated catalogs with lower location errors (Fig. S5). For offshore earthquakes, those located with additional ocean-bottom seismometer records (Anchietta *et al.*, 2011; Merz *et al.*, 2019) may see large improvements as the USGS catalog, which is based on the onshore HVO networks, may contain higher errors. Relocation of earthquakes recorded by the ocean-bottom seismic array deployed shortly after the 2018 Kilauea eruption is currently underway (Wei *et al.*, 2020). The results, together with relocated earthquakes onshore (Shelly and Thelen, 2019; Lin and Okubo, 2020), will be used to update our ArrayConvNet model.

The third way is to use enhanced data augmentation. Because of limited computing resources, we have not explored the asymptotic limit of the number of cuts per earthquake in improving localization. Our tests show that using seven cuts of the same earthquake with random offsets between the trace start time and event origin time improves the hypocenter depth from the case with no data augmentation by more than a factor of 2 (from ± 7.8 km to ± 3.5 km), and from the case with three cuts per earthquake (± 4.7 km) by 26%. This form of data augmentation is clearly effective in improving localization of (origin) time and reducing its tradeoff with the location and the event depth in particular. Another computationally more expensive form of data augmentation is to generate realistic synthetic earthquake waveforms that may account for topography, 3D velocity heterogeneities, and attenuation (e.g., Wang *et al.*, 2018). Such synthetic waveforms are Earth-model dependent but have the advantage that the sources can be placed anywhere, filling the gaps of the catalog earthquake distribution.

Our model focuses on typical catalog earthquakes with short-period and high-frequency energy. However, there are volcanic and magmatic activities that generate long-period (LP) and very-long-period (VLP) seismicity with frequencies below the frequencies used in this study (e.g., Battaglia *et al.*, 2003; Dawson and Chouet, 2014; Matoza *et al.*, 2014; Wech *et al.*, 2020). Because the frequencies of LP and VLP events overlap with microseism, broadening the frequency range to the LP and VLP frequencies may cause an overall decrease of trace SNRs. We thus suggest that LP and VLP events should be processed differently and modeled separately from the typical catalog earthquakes.

Our model assumes detection of a single event in a selected 50 s moving window (Fig. 2). After an event is detected and located, the model moves the window forward by two intervals

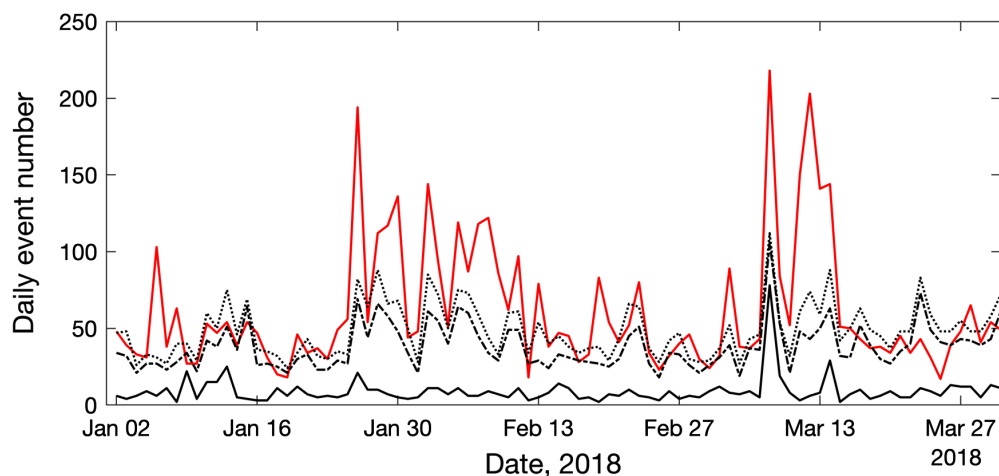


Figure 7. Comparison of daily number of earthquakes detected and located by our model (red line) to those in the published USGS catalog (black line) and in the Hawaiian Volcano Observatory internal catalog (dotted line represent all events; dashed and dotted line represent those with magnitude > 0.1). The color version of this figure is available only in the electronic edition.

(6 s) and starts detection for the next event. In real earthquake monitoring, multiple earthquakes may occur within a short time frame. Such complexity is not considered in the current model. One way to deal with overlapping earthquakes is to expand the output of the neural network (Fig. 2) to include cases with two or more events. For example, in the detection stage, the output can have three or more elements with element 0 for noise, element 1 for a single event, element 2 for double events, and so on. In the localization stage, the output can then include the source location(s) and origin time(s) for one or more events. The training data for overlapping earthquakes can be synthesized from the addition of variably weighted waveforms of catalog earthquakes.

Beyond these improvements to the model, we suggest that this approach can be generalized for other areas. Although the limitation of this methodology is the size of the training set and number of stations, transfer learning may be applicable in this context. Starting from an existing, well-performing model, it is common to only retrain the last layers of the model and apply it to a different application. Utilizing transfer learning decreases the requirement of having thousands to millions of labeled earthquake events to orders of magnitude less, making our suggested methodology much more accessible. On the other hand, as we demonstrated with the Hawaiian data, training of ArrayConvNet with different number of stations and events requires only moderate computational resources that are accessible to nearly everyone and it is straightforward to re-train the network for different regions with different station distributions. Thus, ArrayConvNet may be useful in other seismically active locations where earthquake catalogs already exist.

The unique potential values of ArrayConvNet are: (1) its computational efficiency, which facilitates real-time seismic

monitoring and makes it computationally accessible to nearly every seismologist; (2) its sensitivity and robustness in detecting and localizing small earthquakes under noisy conditions, which may enable next generation of analyses of earthquakes and faults (Ross *et al.*, 2019); and (3) its independence from template earthquakes (as opposed to waveform-similarity-based methods), which allows it to uncover events with source locations and mechanisms that have not been cataloged before.

Data and Resources

The U.S. Geological Survey (USGS) earthquake catalog is obtained

from (<https://earthquake.usgs.gov/earthquakes/search/>, last accessed March 2020). The Hawaiian Volcano Observatory (HVO) internal catalog is provided by Brian Shiro. The waveform data are available from the Incorporated Research Institutions for Seismology Data Management Center (IRIS DMC; <https://ds.iris.edu/ds/nodes/dmc/>, last accessed May 2020). Python codes and the trained models are available at <https://github.com/seismolab/ArrayConvNet> (last accessed November 2020). The supplemental material for this article include six supporting figures mentioned in the main text.

Declaration of Competing Interests

The authors acknowledge there are no conflicts of interest recorded.

Acknowledgments

The authors thank Guoqing Lin and Brian Shiro for providing the relocated earthquakes used to assess the effects of catalog location errors and the Hawaiian Volcano Observatory (HVO) internal catalog, respectively. Special thanks to the staff at HVO for collecting the seismic data used in this study. Comments by Brian Shiro, an anonymous reviewer, and the Associated Editor greatly helped improve the manuscript. Y. Shen is supported by the U.S. National Science Foundation (NSF) Grant Number 1949620. There is no financial conflict of interests of any kind with respect to the results of this article for any author.

References

- Allen, R. (1982). Automatic phase pickers: Their present use and future prospects, *Bull. Seismol. Soc. Am.* **72**, no. 6B, 225–242.
- Anchietta, M., C. Wolfe, G. Pavlis, F. Vernon, J. Eakins, S. C. Solomon, G. Laske, and J. A. Collins (2011). Seismicity around the Hawaiian islands recorded by the plume seismometer networks: Insight into faulting near Maui, Molokai, and Oahu, *Bull. Seismol. Soc. Am.* **101**, no. 4, 1742–1758, doi: [10.1785/0120100271](https://doi.org/10.1785/0120100271).

- Battaglia, J., J.-L. Got, and P. Okubo (2003). Location of long-period events below Kilauea Volcano using seismic amplitudes and accurate relative relocation, *J. Geophys. Res.* **108**, no. B12, 2553, doi: [10.1029/2003JB002517](https://doi.org/10.1029/2003JB002517).
- Caffagni, E., D. Eaton, J. Jones, and M. Baan (2016). Detection and analysis of microseismic events using a matched filtering algorithm (MFA), *Geophys. J. Int.* **206**, no. 1, 644–658, doi: [10.1093/gji/ggw168](https://doi.org/10.1093/gji/ggw168).
- Cannon, E. C., R. Burgmann, A. J. Crone, M. N. Machette, and R. L. Dart (2007). Map and data for Quaternary faults and fault systems on the Island of Hawai'i, *U.S. Geol. Surv. Open-File Rept.* 2007-1284, 1 plate, 81 pp.
- Dawson, P., and B. Chouet (2014). Characterization of very-long-period seismicity accompanying summit activity at Kilauea Volcano, *J. Volcanol. Geoth. Res.* **278/279**, 59–85, doi: [10.1016/j.jvolgeores.2014.04.010](https://doi.org/10.1016/j.jvolgeores.2014.04.010).
- Dokht, R., H. Kao, R. Visser, and B. Smith (2019). Seismic event and phase detection using time-frequency representation and convolutional neural networks, *Seismol. Res. Lett.* **90**, no. 2A, 481–490, doi: [10.1785/0220180308](https://doi.org/10.1785/0220180308).
- Gibbons, S., and F. Ringdal (2006). The detection of low magnitude seismic events using array-based waveform correlation, *Geophys. J. Int.* **165**, no. 1, 149–166, doi: [10.1111/j.1365-246X.2006.02865.x](https://doi.org/10.1111/j.1365-246X.2006.02865.x).
- Got, J.-L., and P. Okubo (2003). New insight into Kilauea's volcano dynamics brought by large-scale relative relocation of micro-earthquakes, *J. Geophys. Res.* **108**, no. B7, 2337, doi: [10.1029/2002JB002060](https://doi.org/10.1029/2002JB002060).
- Johnson, S. W., D. J. A. Chamber, M. S. Boltz, and K. D. Koper (2020). Application of a convolutional neural network for seismic phase picking of mining-induced seismicity, *Geophys. J. Int.* **224**, no. 1, 230–240, doi: [10.1093/gji/ggaa449](https://doi.org/10.1093/gji/ggaa449).
- Kingma, D., and J. Ba (2014). Adam: A method for stochastic optimization, arXiv: 1412.6980, retrieved from <https://arxiv.org/abs/1412.6980> (last accessed June 2020).
- Kong, Q., R. M. Allen, L. Schreier, and Y.-W. Kwon (2016). MyShake: A smartphone seismic network for earthquake early warning and beyond, *Sci. Adv.* **2**, no. 2, e1501055, doi: [10.1126/sciadv.1501055](https://doi.org/10.1126/sciadv.1501055).
- Kriegerowski, M., G. Petersen, H. Vasyura-Bathke, and M. Ohrnberger (2019). A deep convolutional neural network for localization of clustered earthquake based on multistation full waveforms, *Seismol. Res. Lett.* **90**, no. 2A, 510–516, doi: [10.1785/0220180320](https://doi.org/10.1785/0220180320).
- Lin, G., and P. Okubo (2020). Seismic evidence for a shallow detachment beneath Kilauea's south flank during the 2018 activity, *Geophys. Res. Lett.* **47**, no. 15, e2020GL088003, doi: [10.1029/2020GL088003](https://doi.org/10.1029/2020GL088003).
- Lin, G., P. Shearer, R. Matoza, P. Okubo, and F. Amelung (2014). Three-dimensional seismic velocity structure of Mauna Loa and Kilauea volcanoes in Hawai'i from local seismic tomography, *J. Geophys. Res.* **119**, no. 5, 4377–4392, doi: [10.1002/2013JB010820](https://doi.org/10.1002/2013JB010820).
- Lomax, A., A. Michelini, and D. Jozinović (2019). An investigation of rapid earthquake characterization using single-station waveforms and a convolutional neural network, *Seismol. Res. Lett.* **90**, no. 2A, 517–529, doi: [10.1785/0220180311](https://doi.org/10.1785/0220180311).
- Loshchilov, I., and F. Hutter (2017). Fixing weight decay regularization in Adam, *CoRR*, *abs/1711.05101*, Retrieved from <http://arxiv.org/abs/1711.05101> (last accessed June 2020).
- Matoza, R., P. Shearer, G. Lin, C. Wolfe, and P. Okubo (2013). Systematic relocation of seismicity on Hawai'i Island from 1992 to 2009 using waveform cross correlation and cluster analysis, *J. Geophys. Res.* **118**, no. 5, 2275–2288, doi: [10.1002/jgrb.50189](https://doi.org/10.1002/jgrb.50189).
- Matoza, R., P. Shearer, and P. Okubo (2014). High-precision relocation of long-period events beneath the summit region of Kilauea Volcano, Hawai'i, from 1986 to 2009, *Geophys. Res. Lett.* **41**, no. 10, 3413–3421, doi: [10.1002/2014GL059819](https://doi.org/10.1002/2014GL059819).
- Merz, D., J. Caplan-Auerbach, and C. Thurber (2019). Seismicity and velocity structure of Lōihi submarine volcano and south-eastern Hawai'i, *J. Geophys. Res.* **124**, no. 2, doi: [10.1029/2019JB018168](https://doi.org/10.1029/2019JB018168).
- Mousavi, S., and G. Beroza (2020). Bayesian-deep-learning estimation of earthquake location from single-station observations, *IEEE Trans. Geosci. Rem. Sens.* **58**, no. 11, 8211–8224, doi: [10.1109/TGRS.2020.2988770](https://doi.org/10.1109/TGRS.2020.2988770).
- Neal, C. A., S. R. Brantley, L. Antolik, J. L. Babb, M. Burgess, K. Calles, M. Cappos, J. C. Chang, S. Conway, L. Desmither, *et al.* (2019). The 2018 rift eruption and summit collapse of Kilauea Volcano, *Science* **363**, no. 6425, 367–374, doi: [10.1126/science.aav7046](https://doi.org/10.1126/science.aav7046).
- Paszke, A., S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, and S. Chintala (2019). Pytorch: An imperative style, high-performance deep learning library, in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. Alch, E. E. Buc, and R. Garnett (Editors), Vol. 32, Curran Associates, Inc., New York, 8024–8035.
- Perol, T., M. Gharbi, and M. Denolle (2018). Convolutional neural network for earthquake detection and location, *Sci. Adv.* **4**, no. 2, 1700578, doi: [10.1126/sciadv.1700578](https://doi.org/10.1126/sciadv.1700578).
- Ross, Z., M.-A. Meier, E. Hauksson, and T. Heaton (2018). Generalized seismic phase detection with deep learning, *Bull. Seismol. Soc. Am.* **108**, no. 5A, 2894–2901, doi: [10.1785/0120180080](https://doi.org/10.1785/0120180080).
- Ross, Z., D. Trugman, E. Hauksson, and P. Shearer (2019). Searching for hidden earthquakes in southern California, *Science* **364**, no. 6442, 767–771, doi: [10.1126/science.aaw6888](https://doi.org/10.1126/science.aaw6888).
- Shelly, D., and W. Thelen (2019). Anatomy of a caldera collapse: Kilauea 2018 summit seismicity sequence in high resolution, *Geophys. Res. Lett.* **46**, 14,395–14,403, doi: [10.1029/2019GL085636](https://doi.org/10.1029/2019GL085636).
- van den Ende, M., and J.-P. Ampuero (2020). Automated seismic source characterization using deep graphic neural networks, *Geophys. Res. Lett.* **47**, no. 17, doi: [10.1029/2020GL088690](https://doi.org/10.1029/2020GL088690).
- Walter, J. I., P. Ogwari, A. Thiel, F. Ferrer, and I. Woelfel (2020). easyQuake: Putting machine learning to work for your regional seismic network or local earthquake study, *Seismol. Res. Lett.* **92**, no. 1, 555–563, doi: [10.1785/0220200226](https://doi.org/10.1785/0220200226).
- Wang, J., and L. Perez (2017). The effectiveness of data augmentation in image classification using deep learning, arXiv: 1712.04621v1 [cs.CV], retrieved from <https://arxiv.org/abs/1712.04621> (last accessed October 2020).
- Wang, J., Z. Xiao, C. Liu, D. Zhao, and Z. Yao (2019). Deep learning for picking seismic arrival times, *J. Geophys. Res.* **124**, no. 7, 6612–6624, doi: [10.1029/2019JB017536](https://doi.org/10.1029/2019JB017536).
- Wang, N., J. Li, D. Borisov, H. Gharti, Y. Shen, W. Zhang, and B. Savage (2018). Modeling three-dimensional wave propagation in anelastic models with surface topography by the optimal strong stability preserving Runge-Kutta method, *J. Geophys. Res.* **124**, doi: [10.1029/2018JB016175](https://doi.org/10.1029/2018JB016175).

- Wech, A., W. Thelen, and A. Thomas (2020). Deep long-period earthquakes generated by second boiling beneath Mauna Kea volcano, *Science* **368**, no. 6492, 775–779, doi: [10.1126/science.aba4798](https://doi.org/10.1126/science.aba4798).
- Wei, X., Y. Shen, J. Caplan-Auerbach, and J. Morgan (2020). An OBS array to investigate offshore seismicity during the 2018 Kilauea eruption, *Seismol. Res. Lett.* **92**, no. 1, 603–612, doi: [10.1785/0220200206](https://doi.org/10.1785/0220200206).
- Withers, M., R. Aster, C. Young, J. Beiriger, M. Harris, S. Moore, and J. Trujillo (1998). A comparison of select trigger algorithms for automated global seismic phase and event detection, *Bull. Seismol. Soc. Am.* **88**, 95–106.
- Wolfe, C., P. Okubo, G. Ekstrom, M. Nettles, and P. Shearer (2004). Characteristics of deep (≥ 13 km) Hawaiian earthquakes and Hawaiian earthquakes west of 155.55°W , *Geochem. Geophys. Geosyst.* **5**, no. 4, Q04006, doi: [10.1029/2003GC000618](https://doi.org/10.1029/2003GC000618).
- Yoon, C., O. O'Reilly, K. Bergen, and G. Beroza (2015). Earthquake detection through computationally efficient similarity search, *Sci. Adv.* **1**, no. 11, e1501057, doi: [10.1126/sciadv.1501057](https://doi.org/10.1126/sciadv.1501057).
- Zhang, X., J. Zhang, C. Yuan, S. Liu, Z. Chen, and W. Li (2020). Locating induced earthquakes with a network of seismic stations in Oklahoma via a deep learning method, *Sci. Rep.* **10**, no. 1941, doi: [10.1038/s41598-020-58908-5](https://doi.org/10.1038/s41598-020-58908-5).
- Zhang, X., M. Zhang, and X. Tian (2020b). Real-time earthquake early warning with deep learning: Application to the 2016 Central Apennines, Italy earthquake sequence, *Geophys. Res. Lett.* **48**, e2020GL089394, doi: [10.1029/2020GL089394](https://doi.org/10.1029/2020GL089394).
- Zhou, Y., H. Yue, Q. Kong, and S. Zhou (2019). Hybrid event detection and phase-picking algorithm using convolutional and recurrent neural networks, *Seismol. Res. Lett.* **90**, no. 3, 1079–1087, doi: [10.1785/0220180319](https://doi.org/10.1785/0220180319).
- Zhu, L., Z. Peng, J. McClellan, C. Li, D. Yao, Z. Li, and L. Fang (2019). Deep learning for seismic phase detection and picking in the after-shock zone of 2008 $M_w 7.9$ Wenchuan earthquake, *Phys. Earth Planet. In.* **293**, 106261, doi: [10.1016/j.pepi.2019.05.004](https://doi.org/10.1016/j.pepi.2019.05.004).
- Zhu, W., and G. Beroza (2018). PhaseNet: A deep-neural network-based seismic arrival-time picking method, *Geophys. J. Int.* **216**, no. 1, 261–273, doi: [10.1093/gji/ggy423](https://doi.org/10.1093/gji/ggy423).

Manuscript received 8 November 2020

Published online 14 April 2021