

# Array-based convolutional neural networks for automatic earthquake detection and 4D localization

Heather Shen<sup>1</sup> and Yang Shen<sup>1</sup>

<sup>1</sup>University of Rhode Island

November 22, 2022

## Abstract

The growing amounts 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 - a convolutional neural network that uses array data to seamlessly detect and localize events. 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 6 times as many earthquakes as the published catalog. Due to the enhanced detection sensitivity, localization granularity, and minimal computation costs, our solution is valuable, particularly for real-time earthquake monitoring.

1                   **Array-based convolutional neural networks for**  
2                   **automatic earthquake detection and 4D localization**

3                   **Heather Shen<sup>1\*</sup> and Yang Shen<sup>1</sup>**

4                   <sup>1</sup>Graduate School of Oceanography, University of Rhode Island, Narragansett, RI, United States

5                   **Key Points:**

- 6                   • Convolutional neural network models based on seismic arrays automatically and  
7                   accurately detect and localize earthquakes
- 8                   • Application to continuous seismic data in Hawai'i detects and localizes 6 times as  
9                   many earthquakes as the published catalog.
- 10                  • This is the first deep learning model that can automatically provide an earthquake  
11                  catalog from the continuous data of a seismic network.

---

\*As a volunteer research assistant

Corresponding author: Yang Shen, [yshen@uri.edu](mailto:yshen@uri.edu)

**Abstract**

The growing amounts 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 – a convolutional neural network that uses array data to seamlessly detect and localize events. 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 6 times as many earthquakes as the published catalog. Due to the enhanced detection sensitivity, localization granularity, and minimal computation costs, our solution is valuable, particularly for real-time earthquake monitoring.

**Plain Language Summary**

Earthquake catalogs provide baseline information about the movement of the crust and related geological hazards. Yet, catalogs are usually incomplete and fail to log smaller earthquakes undetected by seismic networks. Here, we present a new deep learning model that is computationally efficient and can seamlessly detect and locate earthquakes from continuous seismic records. When the new model is applied to Hawai‘i, it yields 6 times as many earthquakes as the published catalog, promising a more complete catalog that will help improve understanding of seismic and volcanic processes.

**1 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. Caffagni et al., 2016; Gibbons & Ringdal, 2006). Such efforts have led to a great increase in the detection of small earth-

45 quakes, yielding rich details that enable the next generation of analyses of earthquakes  
46 and faults (e.g. Ross et al., 2019). These methods are, however, computationally expen-  
47 sive and limited; detection only works for earthquakes that share similar waveforms and  
48 thus likely have the same source regions and mechanisms of the template events.

49 In the past few years, convolutional neural networks (CNNs) have been adapted  
50 for earthquake detection and location. One common feature shared by CNN approaches  
51 is that once the model is trained, it is far more computationally efficient than the waveform-  
52 similarity-based approach (Gibbons & Ringdal, 2006; Yoon et al., 2015) when it is ap-  
53 plied to new data, an advantage important for seismic network operations, particularly  
54 during periods of intense seismic activities. Perol et al. (2018) introduced a CNN model  
55 for earthquake detection and localization based on waveforms at individual stations. The  
56 localization was limited to a few subregions. Lomax et al. (2019) and Mousavi and Beroza  
57 (2020) developed CNN models for rapid earthquake characterization using single-station  
58 waveforms. Dokht et al. (2019) extended the CNN earthquake detection in the time-frequency  
59 domain. Other studies (e.g. Ross et al., 2018; W. Zhu & Beroza, 2018; L. Zhu et al., 2019;  
60 J. Wang et al., 2019) focused on seismic phase detection and picking of arrival times, which  
61 were then used in traditional travel time-based localization. Kriegerowski et al. (2019)  
62 and Zhang et al. (2020) showed it is possible to use CNNs to locate earthquakes with-  
63 out the intermediate step of phase picking; however, the former depended on manually  
64 chosen arrival times at a reference station and the latter assumed that seismic events had  
65 already been detected. Taking a different approach, Van den Ende and Ampuero (2020)  
66 used Graph Neural Networks with multi-station waveforms to locate earthquakes and  
67 estimate magnitude. Though they too applied it only to existing catalog events. Here  
68 we present a framework based on recent advances in deep learning for seamless, auto-  
69 matic detection and 4D localization of earthquakes without the intermediate steps of phase  
70 detection and picking, phase association, travel time calculation, and inversion. Our ap-  
71 proach builds upon previous work by using a network of seismic stations to first iden-  
72 tify if an earthquake has occurred, and if so, estimate the latitude, longitude, depth, and  
73 origin time of the event.

74 Specifically, we propose a two-stage seismic-array-based, convolutional neural net-  
75 work (ArrayConvNet) model where 1) earthquake detection becomes a supervised clas-  
76 sification problem and 2) earthquake localization becomes a supervised regression prob-  
77 lem. We train and test on data from 55 seismic stations on the Island of Hawai‘i – our  
78 solution not only detects earthquakes in the United States Geological Survey (USGS)  
79 catalog, but also uncovers 6 times more earthquakes missing from the catalog. Once an  
80 earthquake is detected, our model can locate an earthquake’s hypocenter to within 3-

81 4 km of the catalog. To the best of our knowledge, this is the first deep learning model  
82 that can automatically provide an earthquake catalog from the continuous data of a seis-  
83 mic network. Finally, we outline several steps that can be taken to greatly reduce the  
84 model localization errors, making it a viable solution to improve the efficiency and ac-  
85 curacy of seismic monitoring at much lower computational and human costs.

## 86 **2 Data**

87 The Hawai'i Island, USA is one of the most seismically and volcanically active re-  
88 gions in the world, a fact that was heightened by the 2018 eruption of Kilauea Volcano  
89 (Neal et al., 2019). The Hawaiian Volcano Observatory (HVO), USGS operates a per-  
90 manent seismic network (HV) on the island, providing the earthquake information and  
91 waveform data needed for this study. We use 55 seismic stations on Hawai'i Island (Fig-  
92 ure S1). Among them, 33 have three-component (north, east, and vertical) seismome-  
93 ters while the rest have single, vertical-component seismometers.

94 Both the earthquake waveforms and noise segments are downloaded from the In-  
95 corporated Research Institutions for Seismology (IRIS) Data Management Center (DMC).  
96 For each earthquake, we select a 50-second window so that the event origin time is ran-  
97 domly between 1 - 10 seconds from the trace start time. The time difference between the  
98 trace start time and the event origin time, along with the catalog hypocentral location  
99 (latitude, longitude and depth), is used to train the localization part of the model (see  
100 Method). The noise segments are chosen between the USGS reported earthquakes and  
101 are 10 to 50 seconds before the origin time of an earthquake. We visually inspect the noise  
102 windows to minimize the possible presence of unreported earthquakes in the noise seg-  
103 ments. Because the Hawaiian seismic networks have a variety of sensors, we remove in-  
104 strument response from the traces and transfer them to velocity seismograms to min-  
105 imize the effects of different instrument sensitivities to ground motion. The earthquake  
106 waveforms and noise are filtered between 3 to 20 Hz and downsampled to an uniform sam-  
107 pling rate of 50 samples per second on all channels. The frequency range is chosen for  
108 optimum earthquake signal-to-noise ratios based on visual inspection of earthquake wave-  
109 forms over a wide range of frequencies as well as previous studies of the Hawaiian earth-  
110 quake characteristics (e.g. Matoza et al., 2014). All traces are normalized individually  
111 before they are used as the inputs for the CNN model. For stations with missing records  
112 or that do not have three channels, we zero-fill the missing channels.

113 Our ArrayConvNet model has two stages: one for event detection and one for event  
114 localization (see Method). Each stage is trained on distinct training and test data sets.

## 115 2.1 Detection

116 We use 1843 analyst-reviewed earthquakes with magnitude (ml or md) ranging from  
117 0.1 to 5.28 in the 2017 USGS catalog and 1905 noise segments. The number of earth-  
118 quakes is comparable to that in Perol et al. (2018).

119 We explored several ways of arranging the input trace data for the detection model  
120 and chose the following approach based on the robustness of the results when the model  
121 is applied on unseen, continuous data: for both earthquakes and noise segments, we sort  
122 the 55 station traces in order of the time of the vertical component’s largest amplitude  
123 and take the absolute of the traces so all values are between 0 and 1. Therefore, for an  
124 earthquake, we see a clear propagation of earthquake arrivals through the seismic net-  
125 work in a visually easily recognizable pattern (Figure S2). The general pattern is con-  
126 sistent from earthquake to earthquake, regardless of the earthquake location and mag-  
127 nitude (and thus signal-to-noise ratio), as the wave always propagates from the lower left  
128 to the upper right in the maximum-amplitude-sorted waveform images (Figure S2). For  
129 each sorted station, the cross-station features – the information learned by the convo-  
130 lutional kernel, which often has a small size – are local in the time-and-trace-number space  
131 in a well-defined trend. In contrast, the unsorted waveforms, arranged alphabetically by  
132 station names, do not have an easy-to-follow pattern from event to event. Depending on  
133 the source-receiver geometry, a station that has an early earthquake arrival for one event  
134 may have a late arrival for the next event. The cross-station features are highly variable  
135 and may span the entire time-and-trace-number space, thus requiring a deep and large  
136 network to capture. While the sorted and unsorted waveforms do not show substantial  
137 differences in terms of model precision, recall, and the receiver operating characteristic  
138 (ROC) curve (Figure S3), they yield significantly different numbers of detections when  
139 the resulting model is applied to continuous data, indicating differences in the robust-  
140 ness of detection of small earthquakes in noisy data; the unsorted waveforms result in  
141 higher number of false detection from visual inspection of the corresponding seismic traces.  
142 Intuitively, the sorted waveforms have simpler, lower-order features, which require less  
143 complicated neural networks and thus less training data to achieve robust models.

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

147

## 2.2 Localization

148

149

150

151

152

153

154

155

156

157

For the localization part of the model, we use the earthquake locations and origin times from the same 1843 earthquakes in the 2017 USGS catalog. We augment these original waveforms by performing 7 cuts of 50-s-long waveforms for the same earthquake, with each cut starting at a random time between 1-10 s before the event origin time. The seven cuts of each earthquake have the same hypocenter (latitude, longitude, 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. Such data augmentation is commonly used in deep learning (J. Wang & Perez, 2017) and, in our case, helps to train the model to better localize the event origin time (see more in Discussion), which is crucial when the model is applied to continuous data.

158

159

160

161

162

163

164

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  $19.5^\circ$ , longitude  $-155.5^\circ$ , 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.

165

166

167

168

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.

169

## 3 Method

170

### 3.1 Network Architecture

171

172

173

174

175

176

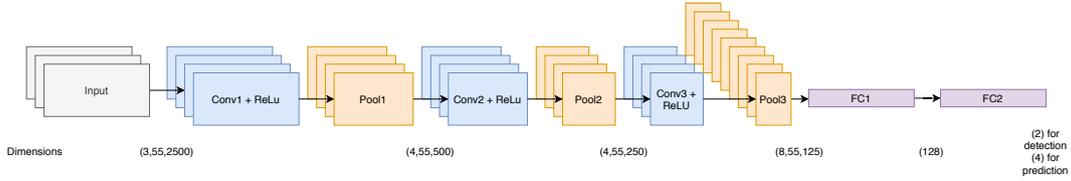
177

Traces for each event are arranged as a three-dimensional (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 (Figure 1).

178

179

After each convolutional layer, we use a rectified linear unit (ReLU) layer to apply an element-wise activation function and then a max pooling layer to perform a down-



**Figure 1.** Architecture of array-based CNN model. Conv and ReLu stand for the convolutional layer and rectified linear unit layer, while 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.

180 sampling operation and decrease the number of parameters. Convolutions are also zero-  
 181 padded to maintain input shape.

182 The kernel of the first convolutional layer has a dimension of width of 9 and height  
 183 of 1. The kernels in the second and third layers have the same dimension of width of 3  
 184 and height of 5. The motivation behind the 1D filter in the first layer is to isolate learn-  
 185 ing of temporal features among the three input channels of each station, as in Kriegerowski  
 186 et al. (2019), while the 2D filters in later layers are designed to extract cross-station in-  
 187 formation. Pooling after the first convolutional layer has a size of (1,5) with a stride of  
 188 (1,5), while pooling after the second and third convolutional layers has a size of (1,2) with  
 189 a stride of (1,2). Thus pooling in our model is designed primarily to downsample in the  
 190 time dimension.

191 We note that our number of convolutional layers (3) and the number of channels  
 192 in each layer (4, 4, and 8) are substantially smaller than in previous studies (e.g., 8 con-  
 193 volutional layers with 32 channels each in Perol et al., 2018). To determine the optimal  
 194 network architecture, we explored a range of the number of convolutional layers (2-5),  
 195 number of channels (2-32), and number of features/neurons of the first fully connected  
 196 layer (64-1024). Our guiding principle in selecting the optimum models is to find the small-  
 197 est network that yields better or comparable results in detection precision. Fewer than  
 198 3 convolutional layers and smaller than 4 channels per convolutional layer yield lower  
 199 precision, as the model may be too simple to capture the full complexity of the data. Greater  
 200 than or equal to 4 convolutional layers, larger than 8 channels, and larger than 128 neu-  
 201 rons in the first fully connected layer yield detection precision comparable to that of our  
 202 preferred network, with the training loss far below (in most cases more than an order  
 203 of magnitude smaller than) the test loss, which suggests overfitting.

### 204 **3.2 Training the Network**

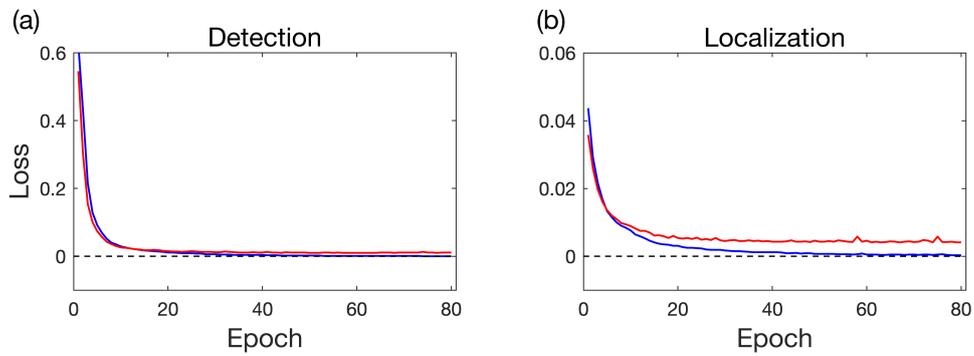
205 The two parts of the CNN model can be trained separately and then connected for  
206 examining continuous data. For detection, we optimize the network parameters by min-  
207 imizing a cross-entropy loss function. This measures the average discrepancy between  
208 our predicted distribution and the true class probability distribution in the training set  
209 and is proven for standard classification problems (e.g. Perol et al., 2018). For localiza-  
210 tion, we optimize the parameters by minimizing a mean-squared error loss between our  
211 predicted and given location and time values.

212 Given our training data set, we are able to minimize our loss functions using a batch  
213 approach. We use a typical 75-25% split for the training and test data sets, respectively.  
214 At each training step, we feed a batch of 32 inputs to the network, evaluate the expected  
215 loss on the batch, and update the network parameters accordingly using backpropaga-  
216 tion. We cycle through all training data in batches as an epoch, and after each epoch,  
217 we calculate the loss for both the training and test data sets. This is repeated until the  
218 loss stops decreasing significantly (80 epochs for both detection and localization, Fig-  
219 ure 2).

220 For optimization, we used the AdamW algorithm (Loshchilov & Hutter, 2017), which  
221 builds on the well-known Adam algorithm (Kingma & Ba, 2014) but separates the weight  
222 decay from the learning rate. The result of this distinction is that the weight decay and  
223 learning rate can be optimized separately, and has been proven to substantially improve  
224 generalization performance. For detection, we use the default learning rate,  $2e-5$ ; for  
225 localization, given the increase in training data due to augmentation, we use a larger learn-  
226 ing rate of  $5e-5$ .

### 227 **3.3 Computational Implementation**

228 We implemented our ArrayConvNet model in Pytorch (Paszke et al., 2019) and per-  
229 formed all model training, testing, and application to continuous data on an iMac with  
230 a 3.8 GHz 8-core Intel Core i7 CPU and 128 GB memory. Model training and testing  
231 in 80 epochs took about 1.3 and 5.7 hours for the detection and localization parts of the  
232 model, respectively. Application of the model to 31-day continuous seismic data took about  
233 5.5 hours.



**Figure 2.** Training (blue line) and test (red line) losses as the function of epochs for the detection and localization parts of the model. It takes less 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 the zero loss.

## 234 4 Results

### 235 4.1 Detection

236 Within 20 epochs, both the training and test losses decrease rapidly and the test  
 237 loss remains small and relatively stable as the number of epochs increases (Figure 2). For  
 238 comparison, Perol et al. (2018) used 32,000 epochs to train their model. In Dokht et al.  
 239 (2019), it took over 10,000 epochs for the earthquake detection learning to approach an  
 240 asymptotic and stable flat line. We attribute the rapid learning of our model to, at least  
 241 partially, the relative simplicity of the network architecture.

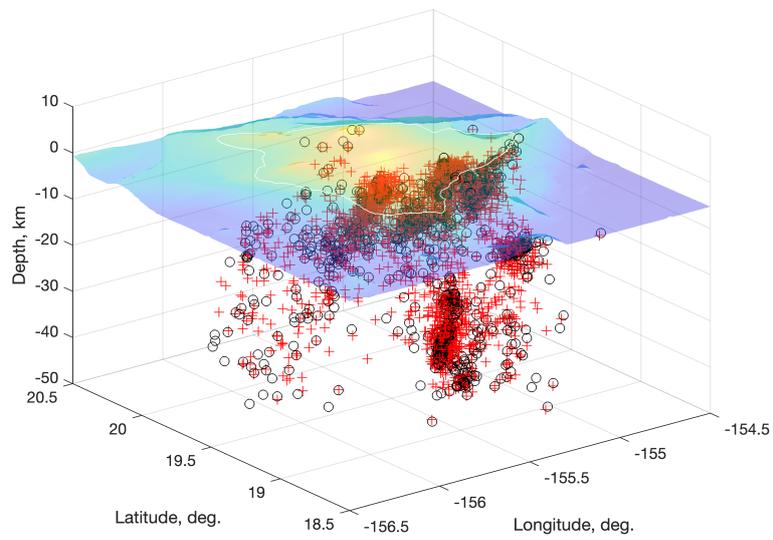
242 Our detection accuracy on the test data, defined as the percentage of events that  
 243 are correctly classified as an earthquake or noise, is 99.4% at 0.5 classification (proba-  
 244 bility) threshold. Between 0.5 and 0.7 classification threshold for earthquakes, the pre-  
 245 cision is 99.6% while recall is 99.2-99.0% (Figure S3). Above 0.7 classification thresh-  
 246 old, the precision is 100% while recall is 99-98%, suggesting that above this detection  
 247 threshold ArrayConvNet does not label any noise as earthquakes, at least in the test data,  
 248 and rarely mis-classifies earthquakes as noise. For comparison, the precision and recall  
 249 reported by Perol et al. (2018) are 94.8% and 100%, respectively, and those by Dokht  
 250 et al. (2019) are 99.6% and 99.9%, respectively.

### 251 4.2 Localization

252 Similar to the detection part of the model, the training and test losses of the lo-  
 253 calization part of the model decrease rapidly within 20 epochs. While the training loss  
 254 continues to decrease towards zero with increasing epochs, the test loss remains flatlined  
 255 (Figure 2), suggesting that the network has enough neurons or complexity to fit the train-  
 256 ing data nearly completely, but uncertainty or random noise in the data keeps the test  
 257 loss at a certain level; more epochs or a larger network likely would not improve the fit  
 258 of the test data.

259 Overall, our model is able to predict the location of an earthquake in the test data  
 260 within  $-0.08 \pm 4.5$  km in the north-south direction,  $0.07 \pm 4.1$  km in the east-west di-  
 261 rection and  $-0.02 \pm 3.5$  km in depth (Figure 3 and Figure S4). The values following the  
 262  $\pm$  sign (and hereinafter) represent one standard deviation. The difference between the  
 263 predicted and catalog origin times is  $-0.06 \pm 0.81$  s.

264 Some of the location and origin time differences may be attributed to errors in the  
 265 USGS catalog. Synthetic tests by Zhang et al. (2020) show that adding a location er-  
 266 ror to the catalog location results in their CNN model prediction error of a similar size.  
 267 Lin et al. (2014) relocated earthquakes with magnitude greater than 1.0 between 1992



**Figure 3.** Hypocenter locations of earthquakes from the USGS catalog (circles) and model predictions (red crosses) in a three-dimensional 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.

268 and 2009, using their 3D velocity model and source-specific station term corrections. Their  
269 resulting catalog thus represents a subset of the HVO events with the best location qual-  
270 ity (Lin et al., 2014). Comparing the earthquake locations in Lin et al. (2014) with the  
271 USGS catalog locations, we find a lateral location offset of  $1.1 \pm 1.8$  km and a depth off-  
272 set of  $1.0 \pm 2.1$  km. So a significant portion of the hypocenter location differences be-  
273 tween our model predictions and the USGS catalog may stem from errors in the data  
274 used to train the model (see more in Discussion).

## 275 **5 Application to Continuous Hawaiian Seismic Data**

276 Earthquake catalogs usually represent a subset of earthquakes that occurred, with  
277 detection and localization limited by signal-to-noise ratios in seismic records, number of  
278 detected stations, and other factors. The USGS catalog for Hawai'i is no exception. So  
279 while our ArrayConvNet performs well for the test data set (Figure S3), further tests on  
280 continuous data, combined with expert reviews of the results, are required to evaluate  
281 its true efficacy.

282 For seismic network operators generating earthquake catalogs, one may wish to min-  
283 imize false detection by using a higher confidence threshold (Ross et al., 2018). Here we  
284 follow this approach, using a probability threshold of 0.95 (95% confidence) in the fol-  
285 lowing discussion unless otherwise stated. Based on the precision and recall character-  
286 istics (Figure S3), the model should rarely mis-classify earthquakes as noise, and almost  
287 never identify noise as an earthquake at this confidence level.

288 We input continuous seismic data from the same 55 seismic stations in Hawai'i, which  
289 are unseen in the development of our CNN model and preprocessed in the same way as  
290 the data used to train the models. The model runs through the data in 50-s-long mov-  
291 ing windows at 3-s increments. When the detection stage of the model finds that the prob-  
292 ability of an earthquake is above a specified confidence threshold, we determine the ex-  
293 act 50-s window by choosing the one that has the highest detection probability. We then  
294 feed the window to the localization stage and calculate the event location. To be con-  
295 sistent with the localization training data, where traces start 1-10 s before the origin time,  
296 the declared event must also have a predicted origin time within 1-10 s after the start  
297 of the traces (Figure S5).

298 Using a continuous data stream from January 2018, our model detects and locates  
299 1603 earthquakes, which is approximately 6.1 times the number reported in the USGS  
300 catalog. Figure S6 presents the number of earthquakes reported by USGS and our model  
301 detection, showing a weak correlation between the daily event numbers. Comparing the

302 USGS catalog and our model outputs on selected days, we find that most but not all of  
303 the catalog events are recovered by our model, consistent with the precision and recall  
304 characteristics at the 0.95 threshold (Figure S3). The missing ones are low magnitudes  
305 ( $<0.7$ ) and have low numbers of reporting stations ( $<10$ ) in the catalog. The events de-  
306 tected and localized by our model have a similar epicentral distribution as those of the  
307 USGS catalog events for January 2018 (Figure S7). As a measure of the sensitivity to  
308 the detection probability threshold, the model detects and locates 1915 and 1542 earth-  
309 quakes with the probability thresholds of 0.68 and 0.997, respectively, or 7.3 and 5.9  
310 times the number in the published USGS catalog.

## 311 6 Discussions

312 As with any supervised machine learning, the more accurate and greater the train-  
313 ing data, the better the resulting model. In our case, the training data can be improved  
314 in several ways:

315 The first is to include the USGS catalog earthquakes from the many years of mon-  
316 itoring by HVO. A greater number of earthquakes plus a correspondingly large number  
317 of noise (visually inspected or automatically screened to minimize the presence of earth-  
318 quakes in the noise segments) should further improve the accuracy and robustness of the  
319 model.

320 The second is to use relocated earthquakes with more accurate locations (e.g. Got  
321 & Okubo, 2003; Wolfe et al., 2004; Matoza et al., 2013; Lin et al., 2014; Shelly & The-  
322 len, 2019). Lin and Okubo (2020) relocated over 48,000 earthquakes between July 2015  
323 and August 2018. With the caveat that all the relocated events in Lin and Okubo (2020)  
324 are onshore, we found that using 1806 earthquakes in the year 2017 relocated by Lin and  
325 Okubo (2020) to train localization in the same way as we discussed above for the USGS  
326 catalog reduces the location difference between the model prediction and the catalog by  
327 25-45% (from  $\pm 4.5$  to  $\pm 2.4$  km in the north-south direction, from  $\pm 4.1$  to  $\pm 2.5$  km  
328 in the east-west direction, and from  $\pm 3.5$  to  $\pm 2.6$  km in depth) and the origin time dif-  
329 ference by 13% (from  $\pm 0.8$  to  $\pm 0.7$  s), demonstrating the effects of relocated catalogs  
330 with lower location errors. For offshore earthquakes, those located with additional ocean-  
331 bottom seismometer records (Anchieta et al., 2011; Merz et al., 2019) may see large im-  
332 provements as the catalog based on the onshore HVO networks may contain higher er-  
333 rors. Relocation of earthquakes recorded by the ocean bottom seismic array deployed  
334 shortly after the 2018 Kilauea eruption is currently underway (Wei et al., 2019). The re-  
335 sults, together with relocated earthquakes onshore (Shelly & Thelen, 2019; Lin & Okubo,  
336 2020), will be used to update our ArrayConvNet model.

337 The third is to use enhanced data augmentation. Due to limited computing resources,  
338 we have not explored the asymptotic limit of the number of cuts per earthquake in im-  
339 proving localization. Our tests show that using 7 cuts of the same earthquake with ran-  
340 dom offsets between the trace start time and event origin time improves the hypocen-  
341 ter depth from the case with no data augmentation by more than a factor of 2 (from  
342  $\pm 7.8$  km to  $\pm 3.5$  km), and from the case with 3 cuts per earthquake ( $\pm 4.7$  km) by 26%.  
343 This form of data augmentation is clearly effective in improving localization of (origin)  
344 time and reducing its tradeoff with the location and the event depth in particular. An-  
345 other computationally more expensive form of data augmentation is to generate realis-  
346 tic synthetic earthquake waveforms that may account for topography, 3D velocity het-  
347 erogeneities, and attenuation (e.g. N. Wang et al., 2018). Such synthetic waveforms are  
348 Earth-model dependent, but have the advantage that the sources can be placed anywhere,  
349 filling the gaps of the catalog earthquake distribution.

350 Our model focuses on typical catalog earthquakes with short-period and high-frequency  
351 energy. However, there are volcanic and magmatic activities that generate long-period  
352 (LP) and very-long-period (VLP) seismicity with frequencies below the frequencies used  
353 in this study (e.g. Battaglia et al., 2003; Dawson & Chouet, 2014; Matoza et al., 2014;  
354 Wech et al., 2020). Because the frequencies of LP and VLP events overlap with micro-  
355 seism, broadening the frequency range to the LP and VLP frequencies may cause an over-  
356 all decrease of trace signal-to-noise ratios. We suggest that LP and VLP events should  
357 be processed differently and modeled separately from the typical catalog earthquakes.

358 Beyond these improvements to the model, we suggest that this approach can be  
359 generalized for other areas. Although the limitation of this methodology is the size of  
360 the training set and number of stations, transfer learning may be applicable in this con-  
361 text. Starting from an existing, well-performing model, it is common to only retrain the  
362 last layers of the model and apply it to a different application. Utilizing transfer learn-  
363 ing decreases the requirement of having thousands to millions of labeled earthquake events  
364 to orders of magnitude less, making our suggested methodology much more accessible.  
365 On the other hand, as we demonstrated with the Hawaiian data, training of ArrayConv-  
366 vNet with more or less stations and events requires only moderate computational resources  
367 that are accessible to nearly everyone. Thus, ArrayConvNet may be useful in other seis-  
368 mically active locations, where earthquake catalogs already exist. The unique potential  
369 values of ArrayConvNet are 1) its computational efficiency, which facilitates real-time  
370 seismic monitoring; 2) its sensitivity and robustness in detecting and localizing small earth-  
371 quakes under noisy conditions, which may enable next generation of analyses of earth-  
372 quakes and faults (Ross et al., 2019); and 3) its independence from template earthquakes

373 (as opposed to waveform similarity based methods), which allows it to uncover events  
 374 with source locations and mechanisms that have not been cataloged before.

## 375 Acknowledgments

376 We thank Quqing Lin for providing the relocated earthquakes used to assess the  
 377 effects of catalog location errors. Special thanks to the staff at HVO for collecting the  
 378 seismic data and providing the earthquake catalog used in this study. Y. Shen is sup-  
 379 ported by the U.S. NSF Grant No. 1949620. There is no financial conflicts of interests  
 380 of any kind with respect to the results of this paper for any author.

381 The USGS earthquake catalog is obtained from ([https://earthquake.usgs.gov/  
 382 earthquakes/search/](https://earthquake.usgs.gov/earthquakes/search/), last accessed March 23, 2020). The waveform data is available  
 383 from the IRIS DMC (<https://ds.iris.edu/ds/nodes/dmc/>, last accessed March 2020).

384 Python codes and the trained models are available at [https://github.com/seismolab/  
 385 ArrayConvNet](https://github.com/seismolab/ArrayConvNet)

## 386 References

- 387 Allen, R. (1982). Automatic phase pickers: Their present use and future prospects.  
 388 *Bull. Seismol. Soc. Am*, *72*(6B), 225–242.
- 389 Anchieta, M., Wolfe, C., Pavlis, G., Vernon, F., Eakins, J., Solomon, S., ... Collins,  
 390 J. (2011). Seismicity around the hawaiian islands recorded by the plume seis-  
 391 mometer networks: Insight into faulting near Maui, Molokai, and Oahu. *Bull.*  
 392 *Seismol. Soc. Am*, *101*(4), 1742–1758. doi: 10.1785/0120100271.
- 393 Battaglia, J., Got, J.-L., & Okubo, P. (2003). Location of long-period events below  
 394 Kilauea Volcano using seismic amplitudes and accurate relative relocation. *J.*  
 395 *Geophys. Res*, *108*(B12), 2553. doi: 10.1029/2003JB002517.
- 396 Caffagni, E., Eaton, D., Jones, J., & Baan, M. (2016). Detection and analysis of  
 397 microseismic events using a matched filtering algorithm (mfa). *Geophys. J. Int.*,  
 398 *206*(1), 644–658. doi: 10.1093/gji/ggw168
- 399 Dawson, P., & Chouet, B. (2014). Characterization of very-long-period seismicity ac-  
 400 companying summit activity at Kilauea Volcano. *J. Volcanology and Geother-*  
 401 *mal Research*, *278-279*, 59–85. Retrieved from [http://dx.doi.org/10.1016/  
 402 j.jvolgeores.2014.04.010](http://dx.doi.org/10.1016/j.jvolgeores.2014.04.010). doi: 10.1016/j.jvolgeores.2014.04.010.
- 403 Dokht, R., Kao, H., Visser, R., & Smith, B. (2019). Seismic event and phase detec-  
 404 tion using time-frequency representation and convolutional neural networks.  
 405 *Seismol. Res. Lett.*, *90*(2A), 481–490. doi: 10.1785/0220180308

- 406 Gibbons, S., & Ringdal, F. (2006). The detection of low magnitude seismic events  
 407 using array-based waveform correlation. *Geophys. J. Int.*, *165*(1), 149–166. doi:  
 408 10.1111/j.1365-246X.2006.02865.x
- 409 Got, J.-L., & Okubo, P. (2003). New insight into Kīlauea’s volcano dynamics  
 410 brought by large-scale relative relocation of microearthquakes. *J. Geophys.*  
 411 *Res.*, *108*(B7), 2337. doi: 10.1029/2002JB002060.
- 412 Kingma, D., & Ba, J. (2014). *Adam: A method for stochastic optimization*. Re-  
 413 trieved from <https://arxiv.org/abs/1412.6980>.
- 414 Kriegerowski, M., Petersen, G., Vasyura-Bathke, H., & Ohrnberger, M. (2019).  
 415 A deep convolutional neural network for localization of clustered earth-  
 416 quake based on multistation full waveforms. *Seismol. Res. Lett.* doi:  
 417 10.1785/0220180320
- 418 Lin, G., & Okubo, P. (2020). Seismic evidence for a shallow detachment beneath  
 419 Kīlauea’s south flank during the 2018 activity. *Geophys. Res. Lett.* doi: 10  
 420 .1029/2020GL088003
- 421 Lin, G., Shearer, P., Matoza, R., Okubo, P., & Amelung, F. (2014). Three-  
 422 dimensional seismic velocity structure of Mauna Loa and Kīlauea volcanoes  
 423 in Hawai‘i from local seismic tomography, tomography. *J. Geophys. Res. Solid*  
 424 *Earth*, *119*, 4377–4392. doi: 10.1002/2013JB010820.
- 425 Lomax, A., Michelini, A., & Jozinović, D. (2019). An investigation of rapid earth-  
 426 quake characterization using single-station waveforms and a convolutional neu-  
 427 ral network. *Seismol. Res. Lett.*, *90*(2A), 517–529. doi: 10.1785/0220180311
- 428 Loshchilov, I., & Hutter, F. (2017). *Decoupled weight decay regularization*.
- 429 Matoza, R., Shearer, P., Lin, G., Wolfe, C., & Okubo, P. (2013). Systematic relo-  
 430 cation of seismicity on Hawai‘i Island from 1992 to 2009 using waveform cross  
 431 correlation and cluster analysis. *J. Geophys. Res. Solid Earth*, *118*, 2275–2288.  
 432 doi: 10.1002/jgrb.50189.
- 433 Matoza, R., Shearer, P., & Okubo, P. (2014). High-precision relocation of  
 434 long-period events beneath the summit region of Kīlauea Volcano, Hawai‘i,  
 435 from 1986 to 2009. *Geophys. Res. Lett.*, *41*, 3413–3421. doi: 10.1002/  
 436 2014GL059819.
- 437 Merz, D., Caplan-Auerbach, J., & Thurber, C. (2019). Seismicity and velocity struc-  
 438 ture of Loihi submarine volcano and southeastern Hawai‘i. *J. Geophys. Res.*  
 439 *Solid Earth*, *124*(2). doi: 10.1029/2019JB018168.
- 440 Mousavi, S., & Beroza, G. (2020). Bayesian-deep-learning estimation of earthquake  
 441 location from single-station observations. *IEEE Transactions on Geoscience*  
 442 *and Remote Sensing*, 1–14. doi: 10.1109/TGRS.2020.2988770.

- 443 Neal, C. A., Brantley, S. R., Antolik, L., Babb, J. L., Burgess, M., Calles, K., ...  
 444 Damby, D. (2019). The 2018 rift eruption and summit collapse of Kilauea  
 445 Volcano. *Science*, *363*(6425), 367–374. doi: 10.1126/science.aav7046
- 446 Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., & Chin-  
 447 tala, S. (2019). Pytorch: An imperative style, high-performance deep  
 448 learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. Alch,  
 449 E. e Buc, & R. Garnett (Eds.), *Advances in neural information process-*  
 450 *ing systems 32* (p. 8024–8035). Curran Associates, Inc. Retrieved from  
 451 [http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style](http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf)  
 452 [-high-performance-deep-learning-library.pdf](http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf) (Retrieved from)
- 453 Perol, T., Gharbi, M., & Denolle, M. (2018). Convolutional neural network for  
 454 earthquake detection and location. *Science Advances*, *4*(2), 1700578. doi: 10  
 455 .1126/sciadv.1700578
- 456 Ross, Z., Meier, M.-A., Hauksson, E., & Heaton, T. (2018). Generalized seis-  
 457 mic phase detection with deep learning. *Bull. Seismol. Soc. Am*, *108*(5A),  
 458 2894–2901. doi: 10.1785/0120180080.
- 459 Ross, Z., Trugman, D., Hauksson, E., & Shearer, P. (2019). Searching for hidden  
 460 earthquakes in southern California. *Science*, *364*, 767–771. doi: 10.1126/science  
 461 .aaw6888
- 462 Shelly, D., & Thelen, W. (2019). Anatomy of a caldera collapse: Kilauea 2018  
 463 summit seismicity sequence in high resolution. *Geophys. Res. Lett*, *46*,  
 464 14395–14,403. doi: 10.1029/2019GL085636.
- 465 Van den Ende, M., & Ampuero, J.-P. (2020). Automated seismic source character-  
 466 ization using deep graphic neural networks. *Geophys. Res. Lett*, *47*(17), 2020  
 467 088690. Retrieved from <https://doi.org/10.1029/2020GL088690> doi: 10  
 468 .1029/2020GL088690
- 469 Wang, J., & Perez, L. (2017). *The effectiveness of data augmentation in image clas-*  
 470 *sification using deep learning*. (arXiv:1712.04621v1 [cs.CV] 13 Dec 2017)
- 471 Wang, J., Xiao, Z., Liu, C., Zhao, D., & Yao, Z. (2019). Deep learning for picking  
 472 seismic arrival times. *J. Geophys. Res. Solid Earth*, *124*(7), 6612–6624. doi: 10  
 473 .1029/2019JB017536.
- 474 Wang, N., Li, J., Borisov, D., Gharti, H., Shen, Y., Zhang, W., & Savage, B.  
 475 (2018). Modeling three-dimensional wave propagation in anelastic mod-  
 476 els with surface topography by the optimal strong stability preserving  
 477 Runge-Kutta method. *J. Geophys. Res. Solid Earth*, *124*. Retrieved from  
 478 <https://doi.org/10.1029/2018JB016175>. doi: 10.1029/2018JB016175.
- 479 Wech, A., Thelen, W., & Thomas, A. (2020). Deep long-period earthquakes gen-

- 480 erated by second boiling beneath Mauna Kea volcano. *Science*, *368*(6492),  
481 775–779. doi: 10.1126/science.aba4798.
- 482 Wei, X., Shen, Y., Caplan-Auerbach, J., & Morgan, J. (2019, December). Seismicity  
483 of the Kilauea Submarine South Flank Following the 2018 Eruption and Mw  
484 6.9 Earthquake. In *Agu fall meeting abstracts* (Vol. 2019, p. V41A-02). Re-  
485 trieved from <https://ui.adsabs.harvard.edu/abs/2019AGUFM.V41A..02W>
- 486 Withers, M., Aster, R., Young, C., Beiriger, J., Harris, M., & SusanMoore, J. T.  
487 (1998). A comparison of select trigger algorithms for automated global seismic  
488 phase and event detection. *Bull. Seismol. Soc. Am*, *88*, 95–106.
- 489 Wolfe, C., Okubo, P., Ekstrom, G., Nettles, M., & Shearer, P. (2004). Characteris-  
490 tics of deep ( $\geq 13$  km) hawaiian earthquakes and hawaiian earthquakes west of  
491 155.55w. *Geochem. Geophys. Geosyst*, *5*, 04006. doi: 10.1029/2003GC000618.
- 492 Yoon, C., O’Reilly, O., Bergen, K., & Beroza, G. (2015). Earthquake detection  
493 through computationally efficient similarity search. *Sci. Adv*, *1*, 1501057. doi:  
494 10.1126/sciadv.1501057
- 495 Zhang, X., Zhang, J., Yuan, C., Liu, S., Chen, Z., & Li, W. (2020). Locating in-  
496 duced earthquakes with a network of seismic stations in Oklahoma via a deep  
497 learning method. *Sci. Reports*, *10*(1941). Retrieved from [https://doi.org/](https://doi.org/10.1038/s41598-020-58908-5)  
498 [10.1038/s41598-020-58908-5](https://doi.org/10.1038/s41598-020-58908-5). doi: 10.1038/s41598-020-58908-5.
- 499 Zhu, L., Peng, Z., McClellan, J., Li, C., Yao, D., Li, Z., & Fang, L. (2019). Deep  
500 learning for seismic phase detection and picking in the aftershock zone of 2008  
501 mw7.9 Wenchuan earthquake. *Physics Earth Planet. Int*, *293*, 106261. doi:  
502 10.1016/j.pepi.2019.05.004.
- 503 Zhu, W., & Beroza, G. (2018). PhaseNet: a deep-neural network-based seismic  
504 arrival-time picking method. *Geophys. J. Int*, *216*(1), 261–273. doi: 10.1093/  
505 gji/ggy423

Figure1\_model-architecture.pdf.

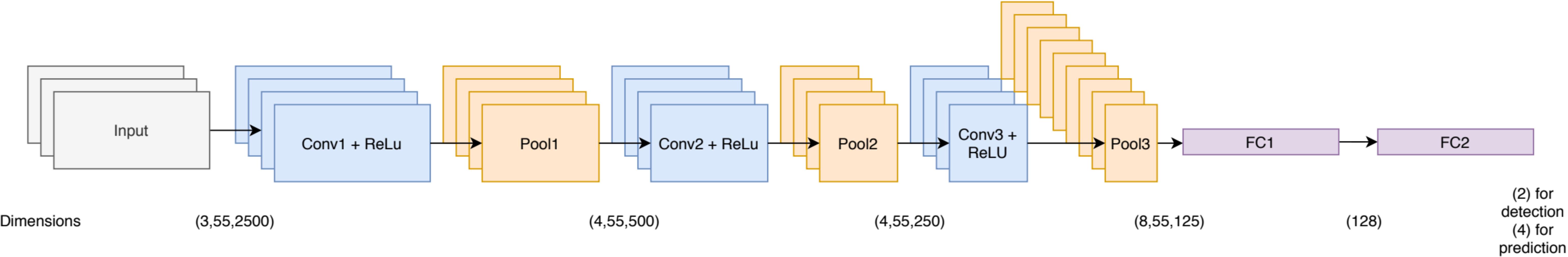


Figure2\_loss\_vs\_epoch.png.

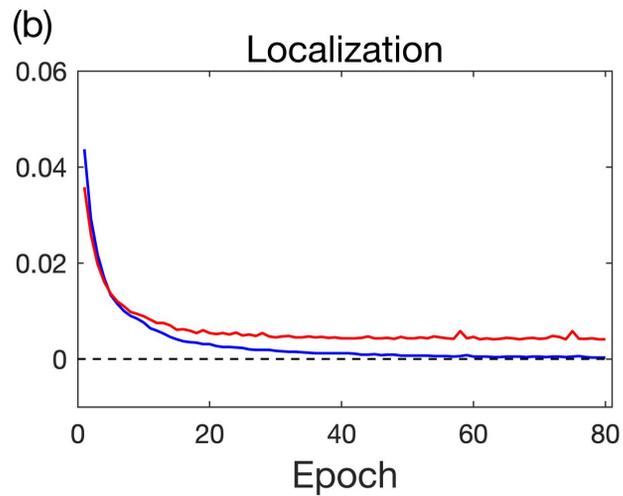
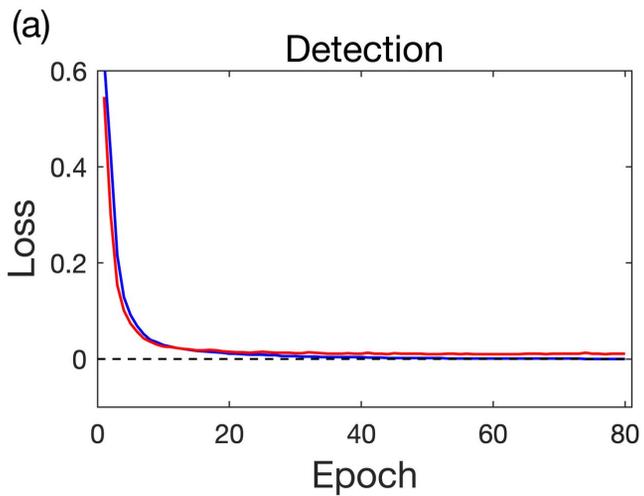
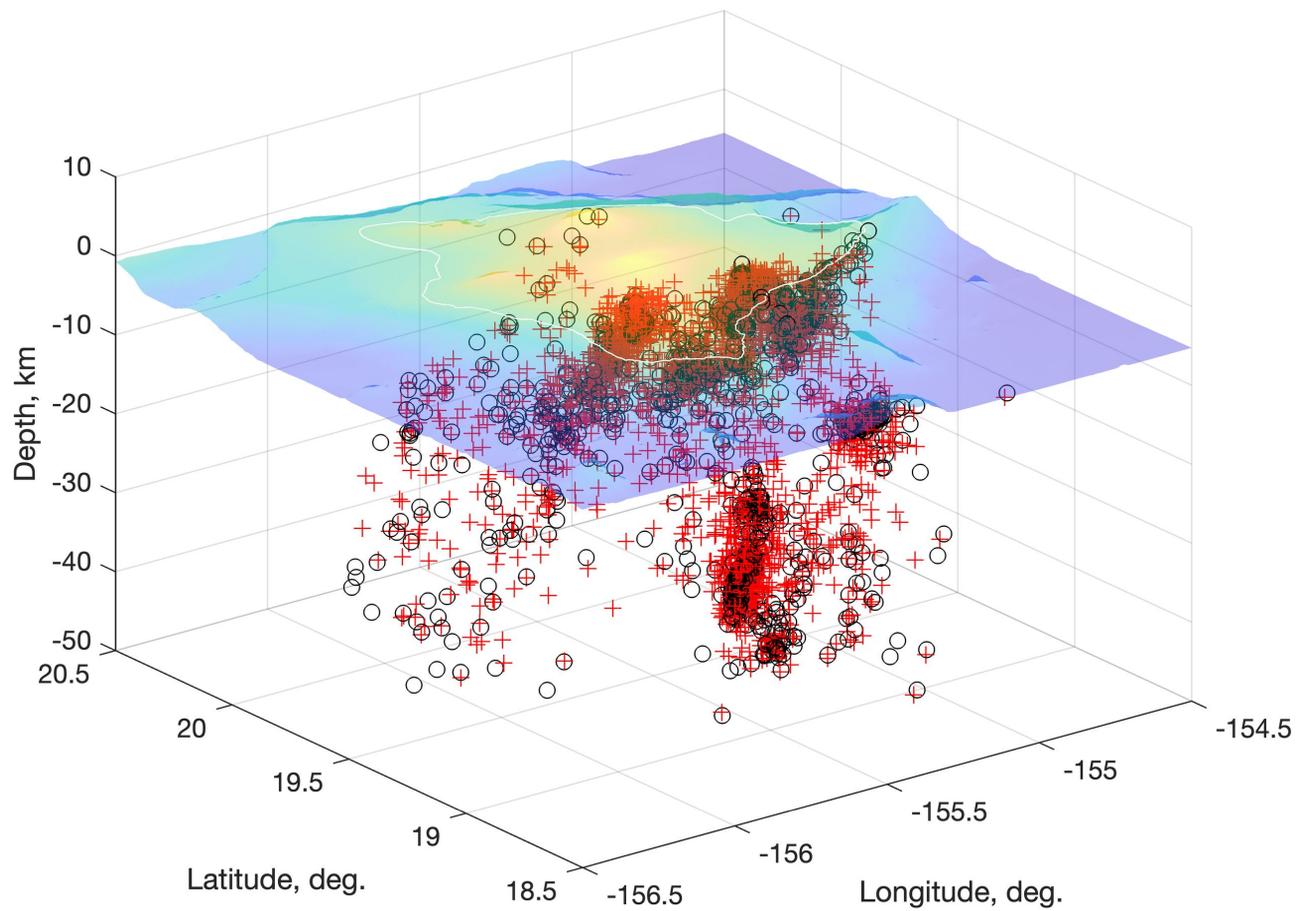


Figure3\_hvo\_vs\_pre\_3dview1SW\_r600.png.



# Supporting Information for ” Array-based convolutional neural networks for automatic earthquake detection and 4D localization”

Heather Shen<sup>1</sup> \* and Yang Shen<sup>1</sup>

<sup>1</sup>Graduate School of Oceanography, University of Rhode Island, Narragansett, RI, United States

## Contents of this file

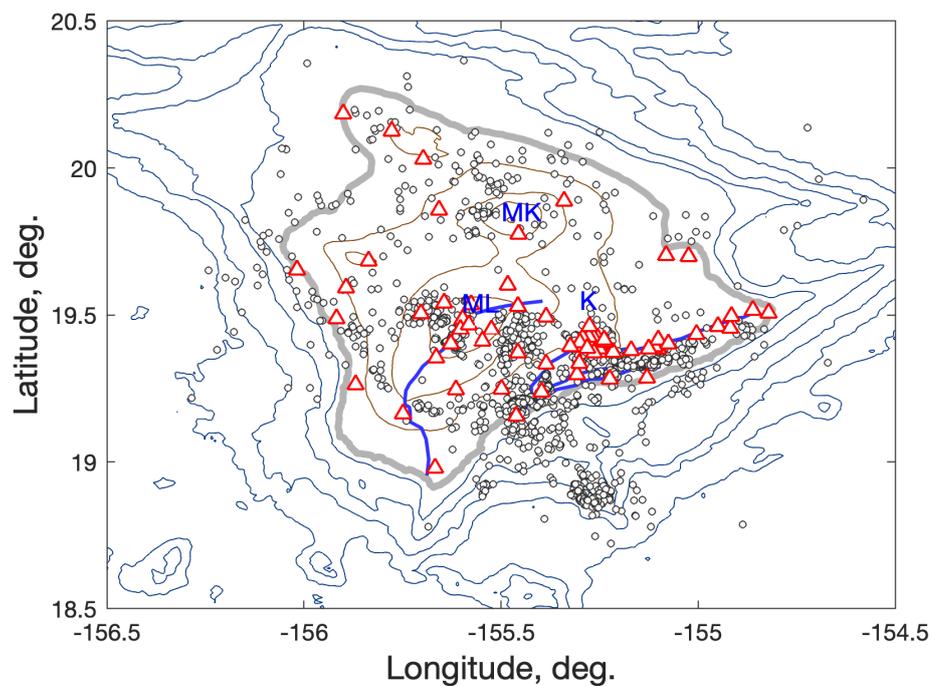
1. Figures S1 to S7

## Introduction

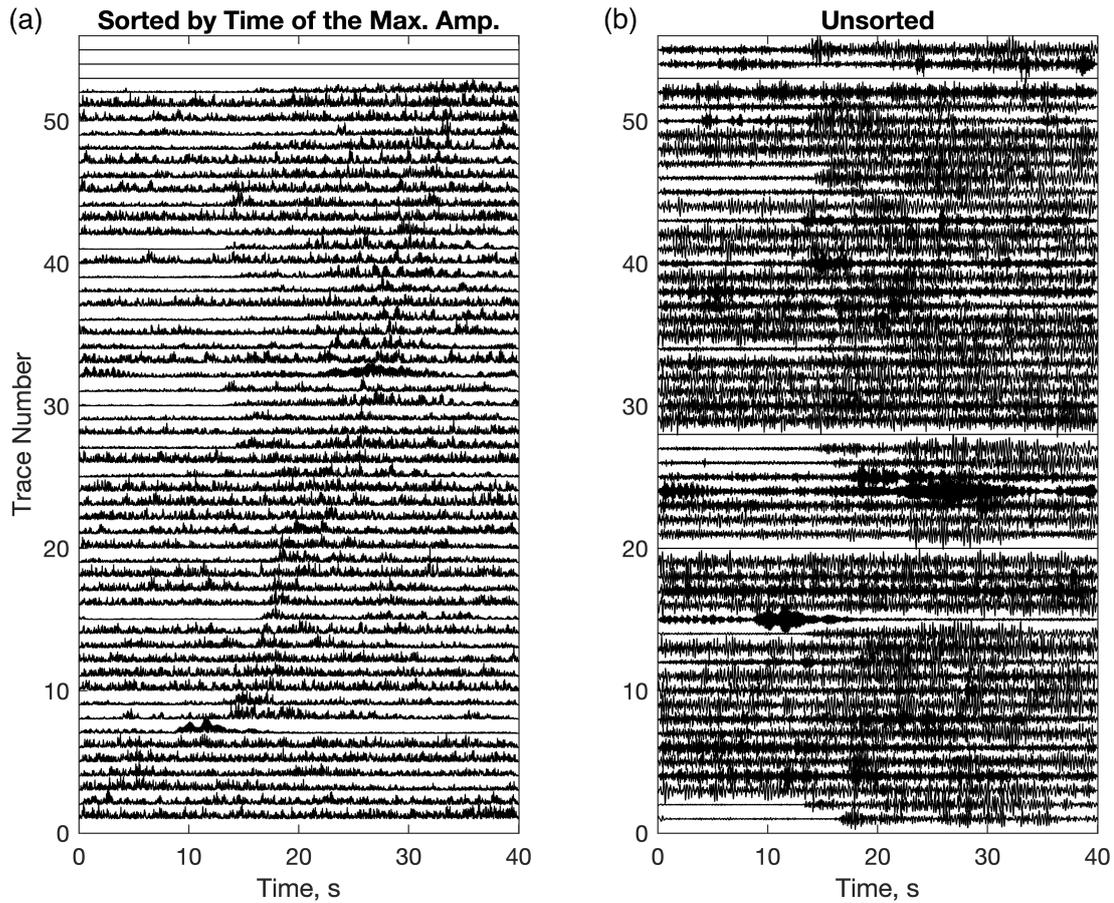
This file contains seven supplementary figures referred to in the main text.

---

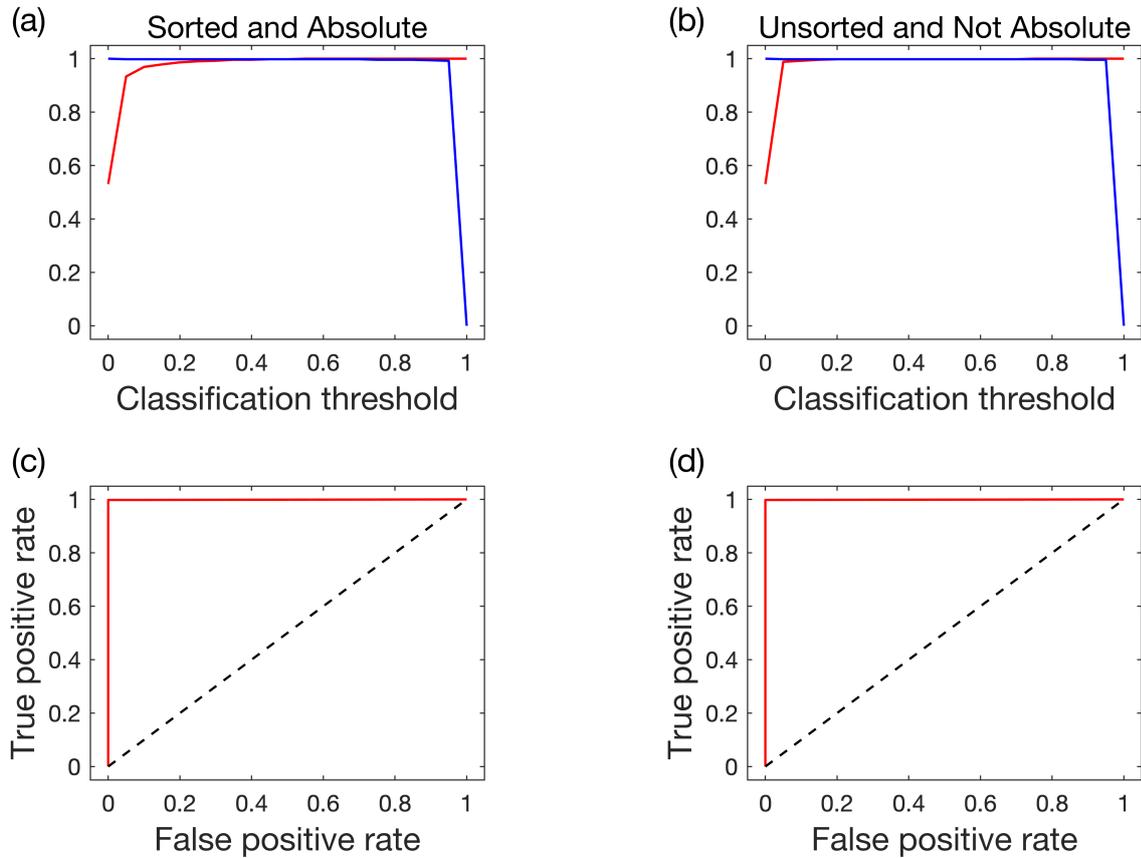
\* As a volunteer research assistant



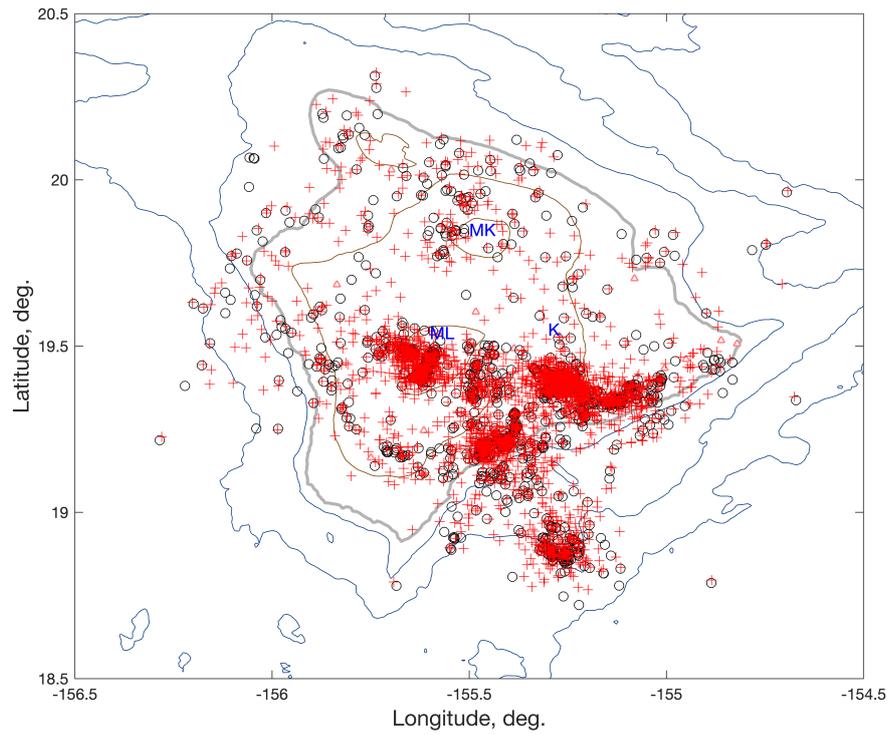
**Figure S1.** Map of the seismic stations (red triangles) and earthquakes (small circles) used in the study. The coast of Hawai'i Island is outlined by the grey contour, while topography and bathymetry are contoured in 1000-m intervals. Blue lines represent the major Quaternary faults and fault systems. ML, K and MK stand for Mauna Loa, Kilauea and Mauna Kea Volcanoes, respectively.



**Figure S2.** An example of earthquake waveforms sorted by (a) the time of the maximum amplitude and (b) the unsorted waveforms arranged by station names alphabetically. The earthquake occurred at time 2017-01-06T01:28:50 according to the USGS catalog. Only the vertical-component traces are shown. Each trace is normalized by its maximum amplitude and the sorted traces are taken as absolute values. The missing channels are zero-filled and placed towards the high-trace-number end in the sorted version. The traces have been bandpass filtered between 3 to 20 Hz.

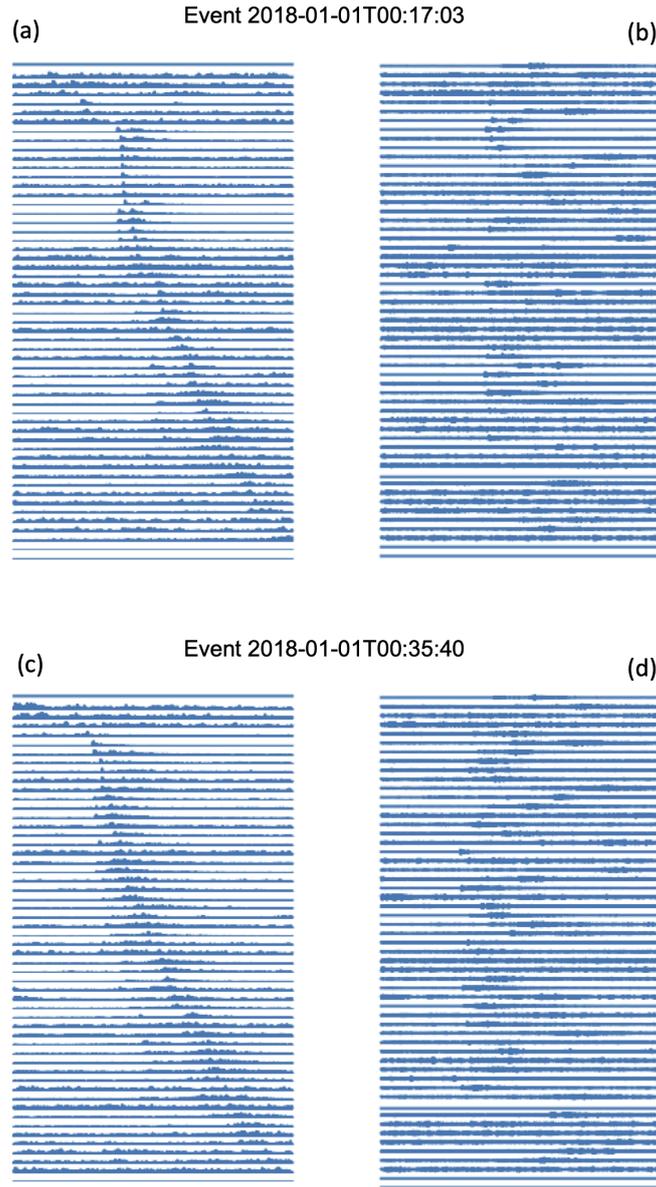


**Figure S3.** Precision (red) and recall (blue) as a function of the classification threshold for the CNN model using the maximum-amplitude-sorted waveforms (a) and that with unsorted, alphabetically arranged waveforms (b). Receiver operating characteristics (ROC) curve (red line) for the model with the sorted (c) and unsorted (d) waveforms. The dashed line is for a model with no predicting skill. The precision, recall, and ROC curves of the resulting models of the two waveform inputs are nearly identical and almost completely overlap if they are plotted together.

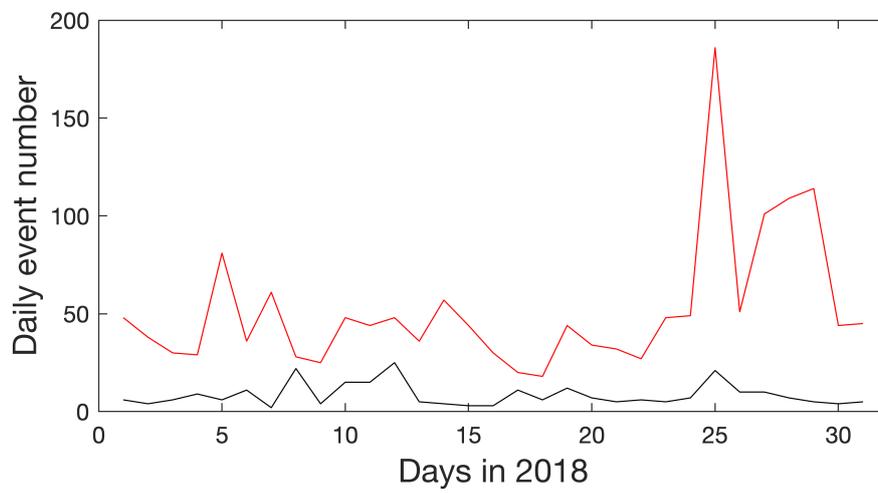


**Figure S4.** Epicenter locations of earthquakes from the USGS catalog (circles) and ArrayCon- vNet model predictions (red crosses) for the test data.

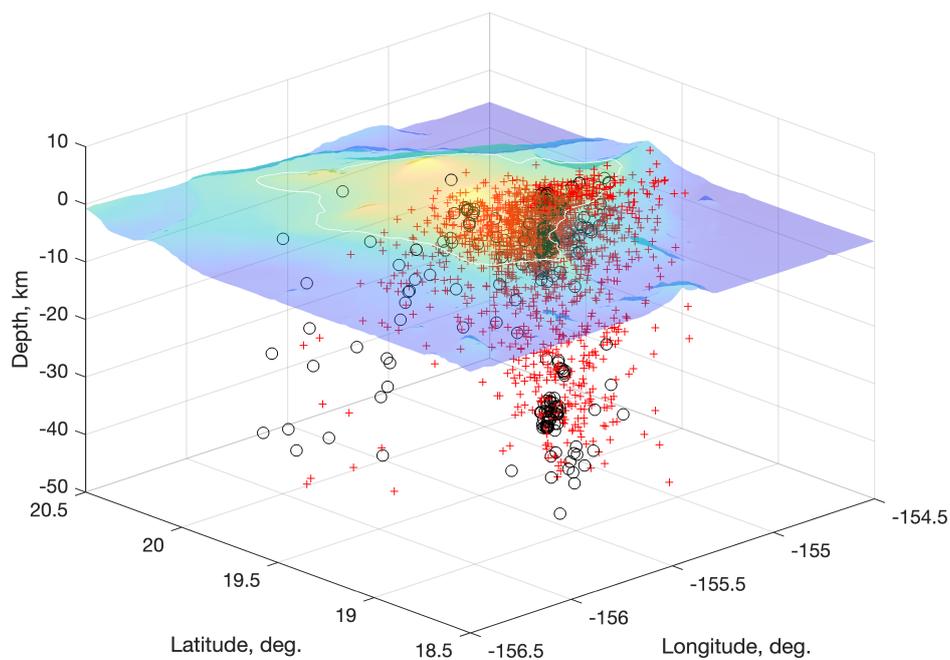
October 13, 2020, 10:19pm



**Figure S5.** Two examples of uncovered earthquakes in the first hour of 2018, which are not in the published USGS catalog. The maximum-amplitude-sorted (a) and unsorted vertical-component waveforms (b) for an event occurred a few seconds after 2018-01-01T00:17:03 (the trace start time). The sorted traces are arranged from the top to bottom and taken as absolute values. The total window length is 50 seconds. (c) and (d) are the same as (a) and (b) for an event shortly after 2018-01-01T00:35:40.



**Figure S6.** Comparison of daily number of earthquakes reported by USGS (black line) and that detected and located by our model (red line).



**Figure S7.** Hypocenter locations of 261 earthquakes from the USGS catalog during January 2018 (circles) and 1605 events detected and localized by our ArrayConvNet model (red crosses) in a three-dimensional view looking from the southwest direction. The topography and bathymetry of the Island are shown as a semi-transparent surface.