Graph-Partitioning Based Convolutional Neural Network for Earthquake Detection Using a Seismic Array

Keisuke Yano1, Takahiro Shiina2, Sumito Kurata3, Aitaro Kato4, Fumiyasu Komaki1,5, Shin’ichi Sakai1,6, and Naoshi Hirata1,2

1The Institute of Statistical Mathematics, Tokyo, Japan, 2Geological Survey of Japan, National Institute of Advanced Industrial Science and Technology, Ibaraki, Japan, 3Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan, 4Earthquake Research Institute, The University of Tokyo, Tokyo, Japan, 5RIKEN Center for Brain Science, Saitama, Japan, 6Interfaculty Initiative in Information Studies, The University of Tokyo, Tokyo, Japan, 7National Research Institute for Earth Science and Disaster Resilience, Ibaraki, Japan

Abstract We present a deep-learning approach for earthquake detection using waveforms from a seismic array consisting of multiple seismographs. Although automated, deep-learning earthquake detection techniques have recently been developed at the single-station level, they have potential difficulty in reducing false detections owing to the presence of local noise inherent to each station. Here, we propose a deep-learning-based approach to efficiently analyze the waveforms observed by a seismic array, whereby we employ convolutional neural networks in conjunction with graph partitioning to group the waveforms from seismic stations within the array. We then apply the proposed method to waveform data recorded by a dense, local seismic array in the regional seismograph network around the Tokyo metropolitan area, Japan. Our method detects more than 97% of the local seismicity catalog, with less than 4% false positive rate, based on an optimal threshold value of the output earthquake probability of 0.61. A comparison with conventional deep-learning-based detectors demonstrates that our method yields fewer false detections for a given true earthquake detection rate. Furthermore, the current method exhibits the robustness to poor-quality data and/or data that are missing at several stations within the array. Numerical experiments using subsampled data demonstrate that the present method has the potential to detect earthquakes even when half of the normally available seismic data are missing. We apply the proposed method to analyze 1-h-long continuous waveforms and identify new seismic events with extremely low signal-to-noise ratios that are not listed in existing catalogs. We also show the potential portability of the proposed method by applying it to seismic array data not used for the training.

1. Introduction

Enormous volumes of continuous seismic data have been acquired from seismograph networks over the past decade, with these data sets consisting of observations from multiple seismic stations. Dense seismograph networks, such as the Japanese Metropolitan Seismic Observation network (MeSO-net) and the Southern California Seismic Network, monitor real-time seismicity and provide continuous waveforms from their respective network stations. Efficient and thorough analyses of these data sets should be of great benefit to seismology. The main objective of the present work, which represents a novel approach to and advance in seismic data analysis, is the development of an improved earthquake detection technique for these massive seismic network data sets.

In recent years, deep neural networks have been attracting increasing interest as tools for analyzing such complex big data in many applied fields such as image processing (Krizhevsky et al., 2012; Szegedy et al., 2015) and natural language processing (Sutskever et al., 2014). Convolutional neural networks (CNNs; Fukushima, 1980; LeCun et al., 1989) are a deep neural network architecture that differs from classically fully connected neural networks, and have recently exhibited remarkable performance in many fields. CNNs have been also used in seismic analyses. CNNs have been utilized for picking P-wave arrival times and discriminating first-motion polarities (Hara et al., 2019; Ross et al., 2018; Wang et al., 2019), detecting earthquakes and their localization (Perol et al., 2018), picking both compressional- and shear-wave (P- and S-wave) arrival times (Zhu & Beroza, 2018), detecting earthquakes and picking arrival times (Mousavi et al., 2020), discriminating between usual earthquakes and tectonic tremors (Nakano et al., 2019), locating...
hypocenters (Kriegerowski et al., 2019; Zhang et al., 2020), and determining the association of seismic phases across an seismic array (McBrearty et al., 2019). These studies have shed light on the effectiveness of employing CNNs to analyze large seismic data sets. Many other deep-learning-based studies have also been developed for various seismic analyses (Kong et al., 2019).

Here, we present a new CNN-based method for earthquake detection using waveform data recorded in a seismic array. Earthquake detection at the single-station level has a potential inability to prevent false detection, because each station records its own local noise that originates from the near-surface environment. The use of multiple seismic stations can help reduce false detection and identify very weak seismic signals (Bergen & Beroza, 2018; Li et al., 2018; Peng & Zhao, 2009). Our goal here is to develop CNNs for earthquake detection using seismographs recorded by a local seismic array.

Besides deep learning methods, there exist various types of earthquake detection methods. Energy-based methods such as the short-term average/long-term average (STA/LTA) have been implemented in the current routine (Allen, 1978; Withers et al., 1998), but have the limitation to detecting low signal-to-noise earthquake signals and have resulted in numerous false positives during an intensive seismic activity. Waveform similarity search using waveform templates, known as template matching, have demonstrated success in finding low signal-to-noise earthquake signals (Gibbons & Ringdal, 2006; Kato et al., 2016; Peng & Zhao, 2009; Ross et al., 2019; Shelly et al., 2007; Skoumal et al., 2014). Subspace detectors (Barrett & Beroza, 2014; Benz et al., 2015; Harris, 2006) generalize template matching by principal component analysis. Yet, these template-based methods have the potential limitation to identifying earthquakes with hypocenters and source mechanisms that are not similar to those of templates. Auto-correlation approach (Brown et al., 2008) searches similar waveforms in continuous records without using templates, but its scaling limit with respect to the record length has been reported (Aguir et al., 2017). Recently, Fingerprint and Similarity Thresholding (FAST) has been developed as a blind, waveform-similarity based detector that does not require templates and has a computationally efficient search algorithm (Bergen & Beroza 2018; Bergen et al., 2016; Yoon et al., 2015). Characteristically, both FAST and deep learning algorithms extract major features from seismic waveform. Compared to FAST, deep learning extracts features in a training-data set-depending way and compresses them to fixed-size parameters, implying a constant memory usage. Then it can help reduce computational runtime and memory usage (Perol et al., 2018).

However, we need to carefully assess CNN approaches to waveforms retrieved by a local seismic array, as CNN inputs are designed to be equispaced (see Chapter 9 in Goodfellow et al., 2016), whereas the actual station distribution in an array is usually nonequispaced. CNN is a neural network tailored for processing objects with a known grid-like topology such as 2D images, and the operations in CNN such as convolution and pooling (described in Section 2) assume the inputs to have a regular grid topology. It is known that the violation of this assumption leads to degrading the performance and there is a room for the improvement (Bruna et al., 2014; Duvenaud et al., 2015; Niepert et al., 2016). We employ a graph-partitioning method, modularity maximization, to accommodate such a nonequispaced distribution by treating the seismograph network as a graph, with each node representing a seismic station in the array. Modularity maximization is a node-clustering algorithm that automatically produces sub-graphs from a given graph. We first group the seismic stations into clusters via modularity maximization and then stack the waveforms, which are convolved with learned filters in CNNs, from each partition. This step is regarded as average pooling with respect to the seismic stations. In conjunction with the max pooling with respect to time, this step reduces effective parameters in CNNs and leads to the improvement of the performance.

We apply our method to waveform data recorded by MeSO-net (Hirata et al., 2009; Kano et al., 2017; Kasahara et al., 2009; Sakai & Hirata, 2009), with a focus on the earthquakes that occurred in the Kanto district, Japan, from September 4–16, 2011. MeSO-net is an online high-density seismic network that consists of approximately 300 accelerometers installed in and around the metropolitan area in the Kanto district (Figure 1). We adopt the 13 seismic stations that are located approximately in the center of the seismic activity around the Kanto district as the seismic array for our analysis. We compare our method with CNNs for a single station and a seismic array consisting of multiple stations without any pooling with respect to the stations. This comparison illustrates that our method markedly improves the detection accuracy of CNNs for analyzing seismic signals, particularly those with a low signal-to-noise ratio (SNR). We also demonstrate...
the robustness of our method for analyzing incomplete seismic records (e.g., due to sensor problems) and temporary increases in noise levels at several stations.

2. Methods

An outline of our method is presented in Figure 2. The input consists of two-dimensional (2D) arrays of the three-channel waveforms (Up–Down, North–South, and East–West) recorded at the seismic stations within our local seismic array, where the horizontal axis of the 2D arrays is the elapsed time from a reference time before the $P$-wave onset, and the vertical axis corresponds to the station alignment. The output is the probability of the input data containing an earthquake record. Our method uses an undirected weighted graph that represents the geometry of the seismic stations within the array. The nodes of the graph represent the stations within the seismic array, and the weights of the edges represent the closeness of the stations.
which are discussed below. We first introduce the two building blocks (CNN and graph partitioning) of our method, following which we present the resultant architecture that is a combination of these two blocks.

2.1. Convolutional Neural Networks

CNNs are one of the most powerful nonlinear feature extractors and can learn a large number of digital filters in combination with fully connected neural networks. CNNs use these learned filters to also provide a function that maps an input (e.g., three-component 2D arrays that represent the observed waveforms from a seismic array) into an output value (e.g., the probability that the input data contains an earthquake).

CNNs build upon a series of layers that sequentially process the input data as follows: (a) convolve (or, more precisely, cross correlate) the input with a set of learnable filters; (b) downsample the convolution output; and (c) apply a nonlinear transformation referred to as the “activation function” to the downsampled output. The first step (a) is termed the “convolution layer”: Let \( \{ h^{(j)}_s : s = 1, \ldots, F \} \) be a total number \( N_F \) of learnable filters. An input waveform \( \{ w_t : t = 1, \ldots, T \} \) of length \( T \) and the \( j \)-th learnable filter \( h^{(j)}_s : s = 1, \ldots, F \) of length \( F \) return a convolution layer that consists of the waveform convolved with the filter, \( \sum_{t=1}^{T} h^{(j)}_s w_{t+s} : t = 1, \ldots, T \). A set of filters \( \{ h^{(j)}_s : s = 1, \ldots, F \} \) is learned during the training step (i.e., a learnable filter), and detects the presence of specific features in the original input. The second step (b) is called the pooling layer. This step drastically reduces the dimensionality, which is the number of parameters and computations in the network, to avoid overfitting caused by excess parameters. This step also introduces a degree of local translation invariance and robustness to local perturbations (see Section 9.3 in Goodfellow et al., 2016 and references therein). There are two common pooling methods: max pooling and average pooling. Max pooling is a pooling operation that calculates the maximum value for each point of each waveform convolved with a learnable filter, whereas average pooling is a pooling operation that calculates the local average for each point of each waveform convolved with a learnable filter. These operations can be written in the following mathematical forms: For the waveform \( \{ \hat{w}_t : t = 1, \ldots, T \} \) convolved with a filter, max pooling with size \( P \) returns \( \max \{ \hat{w}_{t+m} : m = 1, \ldots, P \} \): \( t = 1, \ldots, T \), and average pooling with size \( P \) returns \( \frac{1}{P} \sum_{m=1}^{P} \hat{w}_{t+m} : P = 1, \ldots, T \). This study uses zero padding, where \( \hat{w}_{t+i} = 0 \) for \( i = 1, \ldots, P \) to maintain the same output dimension of the pooling operation as its input dimension. These pooling operations have different roles: Max pooling highlights the sharp changes in the convolved waveform, whereas average pooling smooths these changes. The third step (c) is the activation layer, which makes the output nonlinear and therefore enables the network to express complex features.

However, the challenge of accommodating a CNN to the spatial distribution (geometry) of the stations to boost the power of the CNN remains when applying the CNN to seismograms from a seismic array. Here, we take advantage of graph partitioning to address this challenge, as discussed below.

2.2. Graph Partitioning

Graph partitioning is a way to reduce a graph to smaller sub-graphs by partitioning a set of nodes into mutually exclusive groups and an option of clustering methods. Several methods for graph partitioning exist (Bichot & Siarry, 2013; Fortunato & Castellano, 2012). Here, we employ modularity maximization (Newman, 2006; Newman & Girvan, 2004), which finds a graph partition that maximizes the modularity score. The modularity score for a given graph partition is defined by:

\[
Q = \frac{1}{2m} \sum_{i,j} \left( A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j),
\]

where \( A = (A_{ij}) \) is the adjacency matrix of the graph, \( m \) is the number of edges of the graph, and \( k_i \) is the degree of node \( i \) (i.e., the number of connections that node \( i \) has to other nodes), \( c \) is the partition to which
node \( i \) belongs, and \( \delta(a,b) = 1 \) if \( a = b \) and 0 otherwise. It is shown that \( kk/2m \) is the expected number of edges between nodes \( i \) and \( j \) under the assumption that a graph is randomly generated according to the configuration model. Therefore, the modularity score represents the number of edges within given partitions minus the expected number of randomly generated edges in an equivalent graph. Modularity maximization for graph partitioning stems from the idea that each partition within the “best” graph partition has edges that do not appear in the configuration model. This approach has been applied to a variety of research fields, including bioinformatics, brain science, and social science (Mill et al., 2008; Xia et al., 2013).

We use modularity maximization in the CNN pooling layer via the following process: (i) We first create a graph that represents the seismic array by defining the weighted undirected graph with nodes that correspond to the stations and an adjacency matrix:

\[
A_{i,j} := \exp\left(-\alpha d(i,j)\right) \text{ if } d(i,j) \leq L \text{ and } A_{i,j} := 0 \text{ otherwise, (2)}
\]

where \( d(i,j) \) is the distance between stations \( i \) and \( j \), \( \alpha \) is a sensitivity parameter, and \( L \) is a cut-off distance parameter. We note that the determination of \( d(i,j) \) is important. In this study, we use the usual Euclidean distance between stations because this distance is generally robust for earthquake patterns that have various epicenters and ray-paths. The role of the sensitivity parameter \( \alpha \) is to control weights between stations that produce an adjacency matrix. That of the cut-off distance \( L \) is to sparsify an adjacency matrix. (ii) We then apply modularity maximization to the weighted undirected graph to obtain the graph partition. Different values of hyperparameters \((L,\alpha)\) yield different partitions. Applying modularity maximization to a station graph with larger value of \( L \) and smaller value \( \alpha \) yields a station partitioning that has more stations within the same partition. (iii) We finally convolve the waveforms with learnable filters in the CNN at each station and stack the waveforms at the stations belonging to the same partition. We essentially conduct average pooling with respect to the station axis by using the partitions.

The core idea of this operation is that the waveforms convolved with a learned filter at the closer stations become similar to each other. Therefore, stacking (average pooling) along with graph partitioning can further improve the quality of the learned features. The choices of \( d \), \( \alpha \), and \( L \) may be important, but these can be chosen in a data-driven way, as discussed in the following subsection.

### 2.3. Resultant Architecture

We propose the following architecture, which is a combination of the CNN and graph partitioning. We first convolve the input waveform at each station with filters, followed by waveform processing at each station via max pooling with respect to the time axis. We then conduct average pooling, along with graph partitioning (as described in the previous subsection) to the waveforms. We finally apply the fully connected neural network to the stacked waveforms and obtain the output probability.

In this study, we specified the hyperparameters in the CNN as follows. The number of learnable filters in the convolution layer was 30, with each being 1 s in length. The activation function was a rectified linear unit (Fukushima, 1980), \( x \mapsto \max\{x,0\} \). The length of max pooling was 0.25 s. The number of units in the fully connected layer was 40. The batch size during the training step was 20. The filters in the CNN were optimized during the training step that was conducted by minimizing the cross-entropy loss using stochastic gradient descent optimization (Robbins & Monro, 1951) with momentum (Qian, 1999), where the learning rate was 0.005 and the momentum was 0.9. The cross-entropy loss is a function that maps a pair of true label \( y \) (\( y = 1 \) for earthquake and \( y = 0 \) for noise) and output probability \( p \) for an earthquake into \(-y \log p - (1-y) \log(1-p)\). Stochastic gradient descent optimization with momentum uses a linear combination of the gradient multiplied by the learning rate and the previous update multiplied by the momentum as the next update. The \( l_2 \) penalty with the regularization parameter 1.0 (Ng, 2004) was added to the loss, which reduces the potential overfitting.

The hyperparameters \( \alpha \) and \( L \) in Equation 2 were determined via hold-out validation. We set the candidates for \( \alpha \) and \( L \) to \{0.1,0.5,1.0,5.0\} and \{3,4,5,6,7,8,9,10,11,12\}, respectively, with \( \alpha = 5.0 \) and \( L = 8 \) determined. This set, in conjunction with the modularity maximization, generates sufficient varieties of station
partitions (Figure S1), so we used it. We randomly split the training data set into the training part (80% of the data set) and the test part (the remaining 20%). The training part was used for the training step, and $\alpha$ and $L$ were chosen to minimize the cross-entropy loss of the test part.

3. Data

We applied the proposed method to seismic records from a dense regional seismograph network, MeSO-net, deployed in the Tokyo metropolitan area, Japan, to demonstrate its performance. MeSO-net consists of 296 accelerometers at an average station interval of approximately 5 km. The station intervals in the central part of the station network are in the 2–3 km range. The observed waveforms contain various types of noise signals even though the accelerometers were installed approximately 20 m below the surface. The main source of noise signals in MeSO-net is human activities. Human activities generate a lot of environmental noises such as trains, automobiles, aircrafts, and factories. The exhaustive study on noises observed at MeSO-net stations (Kawakita & Sakai, 2009) reports representative noise waveforms with their spectrograms. These noise signals appear only randomly in time at individual stations, whereas the seismic signals do not. Therefore, analysis of the information commonly recorded in the dense station network potentially improves the capability of earthquake detection.

Here, we selected 13 stations located at the eastern part of MeSO-net as the target seismic array (Figure 1) because they are distributed approximately in the center of the seismically active area in and around the Kanto district (Figure 3). Therefore, this seismic array provided the most earthquakes determined by visual picking that are available for supervised learning by the CNNs. Three-component continuous data were recorded at a 200 Hz sampling rate at each seismic station in the array. We downsampled the records to 25 Hz and then applied a 2–8 Hz bandpass filter to remove background noises. Background noise levels at this frequency range (2–8 Hz) are relatively small (Nishida et al., 2008) and signals of earthquakes are clearly identified in stations of MeSO-net. We detrended an input waveform and normalized it using the median of the maximums of each waveform.

The Earthquake Research Institute, The University of Tokyo, identified 599 earthquakes from MeSO-net data that spanned the September 4–16, 2011 time period (Figure 3a). A total of 6,241 noise signals were also obtained as a by-product of applying STA-LTA method to the continuous waveform data. We manually eliminated all of the signals with very small SNRs at the target array after checking all of the potential earthquakes recorded by the array. This elimination is necessary to mitigate potential deterioration of the performance of
of the onset of each noise in Equation 1, but there is no significant difference. The true positives are events that have been classified as earthquakes from true earthquakes, and false positives are events that have been classified as earthquakes from true noise. We used 318 true earthquakes and randomly selected 500 true noise signals. Abbreviation: MDP, minimum detection probability.

| MDP  | True positives | False positives |
|------|----------------|-----------------|
| 0.5  | 312 (98.1%)    | 38 (7.6%)       |
| 0.61 | 310 (97.5%)    | 17 (3.4%)       |
| 0.9  | 289 (90.9%)    | 2 (0.4%)        |

Notes. The true positives are events that have been classified as earthquakes from true earthquakes, and false positives are events that have been classified as earthquakes from true noise. We used 318 true earthquakes and randomly selected 500 true noise signals. Abbreviation: MDP, minimum detection probability.

We briefly mention the results from applying modularity maximization to the target seismic array when L and α in Equation 1 are fixed to values that are specified via hold-out validation. The resultant partition is illustrated in Figure 2. Modularity maximization using the Euclidean distance achieves a reasonably equispaced station distribution within the same partition. Furthermore, histograms of the S-P arrival-time differences are quite similar to each other within the same partition, as shown in Figure S2, indicating that graph partitioning with the optimized hyperparameters works well.

### 4. Results

#### 4.1. Performance of Our Method

We begin with reporting the performance of our method by varying minimum threshold values for the output earthquake probability. Hereafter, we refer to this threshold as the “minimum detection probability (MDP).” Both true positives (classifying a true earthquake as an earthquake) and false positives (classifying a true noise as an earthquake) become smaller as MDP increases.

As typical cases we present results calculated for three MDPs (0.5, 0.61, 0.9) in Table 1. Our method detects almost all of the earthquakes in the case where MDP is set to 0.5, but it has a relatively large number of false positives. The number of false positives is significantly reduced when MDP is set to 0.9, but fewer earthquakes are detected in the process. MDP of 0.61 results in ≥97% true positives among the true earthquakes and <4% false positives among the true noise signals. We chose 0.61 of MDP as an optimal MDP for our method due to its high performance in the receiver operating characteristics (ROC) curve in Figure 4. The ROC curve measures the classifier performance as the tradeoff between the true positive and false positive along the range of MDPs. The choice of 0.61 is the closest to the upper-left corner, which indicates that the proposed method yields nearly the ideal/best performance at this MDP. The ROC curve of our method also shows that the false positives are <1% when MDP is large (>0.8), <10% when MDP is >0.5, and increases markedly when MDP is <0.35.
We also compared the ROC curve of our method with those of two different CNNs, namely, a CNN for a single station (CNN@E.IIDM) and the conventional CNN for multiple seismic stations (conventional CNN), as shown in Figure 4. The conventional CNN means a CNN without any pooling operation along the station axis, which is commonly employed in multi-station set-ups. We note that our method covers the conventional CNN as a specific case when $L$ is less than the minimum of the distances between stations, and also the CNN for a single station when applied to a single station. We adopt E.IIDM as the single station used in a CNN for a single station because E.IIDM has the highest SNR among the stations in the seismic array (Figure S3).

The ROC curves of the three methods highlight that our method outperforms the other methods (Figure 4). Our method has the lowest false positive rate among the three methods when the same true positive rate is given. Furthermore, a comparison of the three methods with the optimal values of MDPs (the closest points to the upper-left corner in the ROC curves) indicates that our method attains both the most true positives and fewest false positives. Figure S4 entails more detailed comparison with different CNNs.

4.2. Temporal Variation in Output Probability When Detecting an Event

We checked time variation of the performance of our method. We calculated the output probability by applying our method to 2–8 Hz-bandpass-filtered waveform records. We obtained a time series of output probabilities by shifting the 20-s-long window in 0.02 s increments. Figure 5 shows a time variation of the output probability, which shows that the output of our method exhibits reasonable behavior sensitive enough to detect seismic signals: The output probability suddenly increases and exceeds 0.61 when $P$-waves have arrived in the time window at several stations (1), 0.95 when $P$-waves have arrivals in the time window at almost all of the stations (2), and 0.98 when both $P$- and $S$-waves appear in the window at almost all of the stations (3), and gradually reduces to <0.61 when the $S$-wave arrivals are outside of the time window at several stations (4).

4.3. Effects of the Signal-to-Noise Ratio on Event Detection

We introduced an SNR-based metric, the median of the ratios of the in-event median absolute deviances (MADs) to the pre-event MADs, to quantify the impact of the SNR on the performance of the method. Specifically, we defined the median MAD ratio (MADR) as:
where $T$ is the time length, $N$ is the number of stations, $\{a_{\text{pre},i}(t); t = 1, \ldots, T\}$ is the pre-event recorded amplitude at station $i$, $\{a_{\text{in},i}(t); t = 1, \ldots, T\}$ is the in-event recorded amplitude at station $i$, $\bar{a}_{\text{pre},i}$ is the median of $a_{\text{pre},i}(t)$ with respect to time $t$, and $\bar{a}_{\text{in},i}$ is the median of $a_{\text{in},i}(t)$ with respect to time $t$. We used the 20-s-long waveforms that started 30 s prior to $P$-wave arrivals as the pre-event waveforms. We defined the three-component
median MADR as the median of three median MADRs that were calculated from the Up–Down, North–South, and East–West components. The median MADR is a median version of the root-mean-square (RMS) deviation ratio and is more robust than the RMS deviation ratio when the waveforms are contaminated (see Chapter 5 of Huber & Ronchetti, 2009).

Figure 6a shows that the lower bound of the output probabilities of events with the same median MADR increases with increasing median MADR. Our method generally misses earthquakes with a small median MADR (i.e., a small SNR). Similar behaviors are observed in the conventional CNN (Figure 6b). Yet, we find that our method generally yields higher output probabilities compared with the conventional CNN results. In addition, the performance deterioration of our method is slower than that of the conventional CNN, which is confirmed by comparing the largest values of the median MADRs that have the output probabilities below the optimal MDPs. The combination of a higher output probability (Figure 6) and lower false detection rate (Figure 4) indicates that the appropriate clustering of stations in a seismic array greatly improves CNN-based earthquake detection.

It is worth noting that events with median MADR values of <1.0 can in some cases be detected with relatively high output probabilities. This means that changes in the pre- and in-event amplitudes are not significant for such events. Further analysis of the waveform and/or amplitude features of these events may improve the detection accuracy of earthquakes with small SNRs.

5. Discussion

5.1. Robustness With Respect to Missing Data

Continuous seismic records may be incomplete owing to temporary power outages, as well as failures in seismic sensors and/or data loggers. Therefore, we examined the robustness of our proposed method with respect to such incomplete data records to investigate its performance and stability. We created subsampled

![Figure 6. Output probabilities and three-component median MADRs for (a) our method and (b) the conventional CNN. The dotted horizontal lines denote the optimal MDPs of 0.61 and 0.82 in our method and the conventional CNN, respectively. MADR, median absolute deviance ratio; MDP, minimum detection probability.](image)
data that consisted of waveforms with missing data using the validation data set (318 earthquakes). We also randomly removed seismic stations and created 10 subsampled data sets with different numbers of missing stations for the earthquakes in the validation data set.

Figure 7a illustrates how the number of missing stations changes output probability for 3,180 trials, with a drop in performance as the number of missing stations increases. Increasing the number of missing stations from one to six results in 83%, 80%, 78%, 74%, 68%, and 58% of the trials possessing output probabilities of >0.61, respectively. It also shows that the median output probability remains above 80%, regardless of the number of missing stations, which implies that performance deterioration is relatively low for half of the trials. These two observations suggest that data quality, in the form of missing data, impacts on performance deterioration. We selected three events with different median MADRs to illustrate this point (Figure 7b). The event with the highest median MADR (=95.30) possessed an output probability of approximately 1.0, even when the data loss arising from missing stations increased. The event with the median MADR nearly 1.0 possessed an output probability that decayed as a function of data loss and remained high (>0.9) until the number of missing stations increased to six. The event with the low median MADR (=0.77) possessed an output probability that decayed substantially, falling below 0.61 in cases where the number of missing stations increased to five. Given the high quality of the data, our CNN-based method can reliably evaluate earthquakes, even if up to half of the seismic data are missing.

5.2. Continuous Application to the September 9, 2011 Data

We applied our method to 1-h-long continuous waveform data, starting at 0:00 JST on September 9, 2011, to explore the feasibility of the proposed method in detecting earthquakes from continuous streaming data. We used the data because a large number of earthquakes are excited by the 2011-March-11 Tohoku-oki earthquake (M9) even in the Kanto district (e.g., Ogata et al., 2019). We calculated the temporal change in the output probabilities by shifting the 20-s-long window in 0.02 s increments over the entire hour.
picked event candidates that possessed output probabilities of >0.61 for more than 1 s (yellow stars and orange stars in Figure 8b).

A comparison of the time series of the median MAD at the stations within the array and the output probabilities from our method (Figure 8) reveals several sharp peaks in the output probability sequence that are synchronous with local peaks in the median MAD sequence, with four event candidates being detected with a high degree of confidence. Two of these event candidates correspond to earthquakes that had been detected by JMA and MeSO-net (orange stars in Figure 8), thereby implying the validity of our method in detecting earthquakes using the seismic array. Although the other two candidates (yellow stars in Figure 8b) were not detected in the routine MeSO-net detection procedure, we consider these to be actual earthquakes, as seismic signals can be identified in the nationwide seismic network in the Kanto district (e.g., high-sensitivity seismograph network in Japan (Hi-net); Obara et al., 2005; Okada et al., 2004) at the time of these two events. The relocated hypocenters of these events, which were determined from arrival times picked visually, fall within the lateral range of the training earthquakes (Figure 8e). The event at approximately 00:50:20 JST has a very high (>0.99) output probability (Figure 8d). This high output probability and the proximity of the event to a cluster of training earthquakes imply that events occurring near training earthquakes can be detected with a high output probability. MeSO-net consists of a large, dense distribution of seismic stations around the Tokyo metropolitan area compared with the nationwide seismic network. Therefore, the level of performance obtained in this study suggests that implementation of a suitable method can improve the ability to detect and monitor small earthquakes that occur in and around metropolitan areas.

5.3. Application of Our Method to Seismic Array Data Not Used for the Training

In order to demonstrate the performance in applying our method to untrained seismic arrays, we checked the concordance of the output probabilities with the earthquake signal arrivals in several seismic arrays not used for the training.

We used four untrained arrays, each of which consists of 13 seismic stations, as shown in Figure 9a. These four untrained arrays have different distances from the trained array, so we chose them for the following evaluations. We note that our detection method can be applied to an untrained array when the following two conditions hold. The first condition is that a seismic array contains at most 13 stations. The second condition is that an array has to be divided into 3 partitions. In this experiment, we selected the hyperparameters \( L, \alpha \) in Equation 1 for the untrained arrays as in Figure S5. We used a waveform of 100-s long window starting from 23:29:56 09/09/2011 (JST) as depicted in Figure 9b. Within this time window, two successive earthquakes are contained. These events with magnitudes \( M_{3.9} \) and \( M_{2.9} \) closely occurred with an origin time difference of \( \sim 30 \) s; their origin times are 23:29:52 (JST) and 23:30:20 (JST), respectively. The source depths are 4.4 km (±0.9) and 5.2 km (±0.5), respectively. We calculated the temporal change in the output probabilities by shifting the 20-s-long window in 0.02 s increments.

We first confirm that in the trained array, two events have high SNR and the output probabilities are concordant with the arrivals of these events. In Untrained Arrays 1, 2, and 3, our method correctly detects two events with the high output probabilities (>0.61). In Untrained Array 4, our method successfully detects the first event but misses the second event. The missing of the second event in Untrained Array 4 is due to a low SNR (Figure 9f). In summary, our method can be applicable to untrained networks under the above two conditions.

5.4. Response to Teleseismic Earthquakes

Here, we scrutinized the response of our method to waveforms generated from distant earthquakes with the moderate-to-large magnitude. As an example, we show the time series of the obtained output probability with the 2–8 Hz-bandpass-filtered waveform records from the 2011 \( M_w 7 \) Vanuatu earthquake (the epicentral distance is about 6,400 km) in Figure 5. The output probability fluctuates around the optimal MDP line (0.61) when the direct P-waves arrive at the array. Although our detection method reacts when the seismographs have sufficient amplitudes in 2–8 Hz, the output probability does not indicate a large value that far exceeds the optimal MDP line (0.61). We obtained the similar temporal changes in the output prob-
Figure 8. Application of our method to 1-h-long waveform data starting at 0:00 JST on September 9, 2011. (a) Time series of the median MAD for the 13-station array. (b) Time series of the output probabilities, with the detected events indicated (orange stars if the catalogs exist and yellow stars if not). (c and d) Waveforms of two newly detected events. (e) Epicenters of detected events (orange and yellow stars) and training events during the time period from September 4–11, 2011 and the 2011–2016 time period (gray dots) relative to the 13-station array (inverted red triangles). MAD, median absolute deviance.
Figure 9. Application of our method to untrained seismic arrays. (a) Locations of arrays (inverted colored triangles) and epicenters of the target earthquakes (yellow stars). (b) Time series of the output probabilities and the vertical components of seismographs observed at the trained array. (c–f) Time series of the output probabilities and the vertical components of seismographs observed at four untrained arrays (Untrained Arrays 1, 2, 3, 4).
abilities for the other distant events. Thus, our method results in a weak detection at the timing of arrival of waveform from a distant event.

To improve the performance of our method, an idea is to learn teleseismic earthquakes as one category of noise signals in the framework of CNNs. It is thus necessary to develop an appropriate classification method for noise signals before applying the supervised learning. Such idea would be implemented in combination with an unsupervised learning (Johnson et al., 2020).

6. Conclusion

We developed a CNN that incorporated graph partitioning for earthquake detection using a multi-station local seismic array. The performance of our method is validated by both its high true detection rate and low false detection rate. Our method outperforms two different conventional types of CNN-based detector. We found that the temporal variations in the output earthquake probability of our method during event detection are concordant with the P- and S-wave arrivals in the array. We used the median MADR as a measure of the SNR to confirm that the SNR has some influence on the output earthquake probability of our method; however, some low-SNR events can be detected with relatively high output probabilities. Numerical experiments using subsampled data demonstrated that our method exhibits robustness when some data are missing. We also identified new seismic events with extremely low SNRs that are not listed in existing earthquake catalogs by applying our method to 1-h-long continuous waveform data. We applied our method to untrained arrays to show the potential portability of our method.

Data Availability Statement

Seismic waveform data used in the present study are available via the website (http://www.eri.u-tokyo.ac.jp/project/iSeisBayes/MeSO-net_Narita-Array_2020).

Acknowledgments

The Hi-net data (National Research Institute for Earth Science and Disaster Resilience, 2019) were used for relocating detected events. GMT software package (Wessel & Smith, 1998) was used for creating geographical figures. TensorFlow was used for training and applying CNN. NetworkX (Hagberg et al., 2008) was used for conducting graph partitioning. The authors sincerely thank Hiroichii Nagao, Shigeki Nagawa, Hiroshi Tsuruoka, and Miwa Yoshida for their help in processing the MeSO-net data. The authors sincerely thank the AE and two anonymous referees for their valuable comments and suggestions. This work is supported by JST CREST Grant Number JPM-JCR1763, Japan.

References

Aguiar, A. C., Chao, K., & Beroza, G. C. (2017). Tectonic tremor and LFEs on a reverse fault in Taiwan. Geophysical Research Letters, 44, 6683–6691. https://doi.org/10.1002/2016gl072148
Allen, R. (1978). Automatic earthquake recognition and timing from single traces. Bulletin of the Seismological Society of America, 68(5), 1521–1532.
Barrett, S. A., & Beroza, G. C. (2014). An empirical approach to subspace detection. Seismological Research Letters, 85(3), 594–600. https://doi.org/10.1785/0220130152
Benz, H. M., McMahon, N. D., Aster, R. C., McNamara, D. E., & Harris, D. B. (2015). Hundreds of earthquakes per day: The 2014 Guthrie, Oklahoma, earthquake sequence. Seismological Research Letters, 86, 1318–1325. https://doi.org/10.1785/0220150019
Bergen, K. J., & Beroza, G. C. (2018). Detecting earthquakes over a seismic network using single-station similarity measures. Geophysical Journal International, 213(3), 1984–1998. https://doi.org/10.1093/gji/ggy3100
Bergen, K. J., Yoon, C., & Beroza, G. C. (2016). Scalable similarity search in seismology: A new approach to large-scale earthquake detection (9939, pp. 301–308). Proceedings of the 9th International Conference on Similarity Search and Applications, Lecture Notes in Computer Science, Springer. https://doi.org/10.1007/978-3-319-46759-7_23
Bichot, C., & Siarry, P. (2013). Graph partitioning. John Wiley &Sons.
Brown, J., Beroza, G., & Sherry, D. (2008). An autocorrelation method to detect low frequency earthquakes within tremor. Geophysical Research Letters, 35(16), L16305. https://doi.org/10.1029/2008gl034560
Bruna, J., Zaremba, W., Szlam, A., & LeCun, Y. (2014). Spectral networks and locally connected networks on graphs. Paper presented at the International Conference on Learning Representations (ICLR 2014), Banff, Alberta, Canada.
Duvenaud, D. K., Maclaurin, D., Iparraguirre, J., Bombarell, R., Hirzel, T., Aspuru-Guzik, A., & Adams, R. P. (2015). Convolutional networks on graphs for learning molecular fingerprints (pp. 2224–2232). Proceedings of Advances in Neural Information Processing Systems (NIPS 2015).
Fortunato, S., & Castellano, C. (2012). Community structure in graphs. In Computational complexity (pp. 490–512). Springer. https://doi.org/10.1007/978-1-4614-1809-3_3
Fréay, B., & Verleysen, M. (2014). Classification in the presence of label noise: A survey. IEEE Transactions on Neural Networks and Learning Systems, 25(5), 845–869. https://doi.org/10.1109/tnnls.2013.2292894
Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics, 36, 193–202. https://doi.org/10.1007/bf0044251
Gibbons, S. J., & Ringdal, F. (2006). The detection of low magnitude seismic events using array-based waveform correlation. Geophysical Journal International, 164(1), 149–166. https://doi.org/10.1111/j.1365-246x.2006.02865.x
Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
Hagberg, A., Swart, P., & Schult, D. (2008). Exploring network structure, dynamics, and function using NetworkX (pp. 11–15). Proceedings of the 7th Python in Science Conference (SciPy2008), Pasadena, CA.
Hara, S., Fukahata, Y., & Iio, Y. (2019). P-wave first-motion polarity determination of waveform data in western Japan using deep learning. Earth, Planets and Space, 71, 127. https://doi.org/10.1186/s40623-019-1111-x
Harris, N. (2006). Subspace detectors: Theory (Report UCRL-TR-222758). Lawrence Livermore National Laboratory, Lawrence Livermore National Laboratory.

Hirata, N., Sakai, S., Sato, H., Takeoka, K., & Koketsu, K. (2009). An outline of the Special Project for Earthquake Disaster Mitigation in the Tokyo Metropolitan Area-Subproject I: Characterization of the plate structure and source faults in and around the Tokyo Metropolitan Area. Bulletin of the Earthquake Research Institute, University of Tokyo, 84, 41–58.

Huber, F., & Rongchetti, E. (2009). Robust statistics (2nd ed.). John Wiley & Sons. https://doi.org/10.1002/9780470443469

Johnson, C., Ben-Zion, Y., Meng, H., & Vernon, F. (2020). Identifying different classes of seismic noise signals using unsupervised learning. Geophysical Research Letters, 47(15), 7647–7654. https://doi.org/10.1029/2020GL088353

Kano, M., Nagao, H., Nagata, K., Ito, S., Sakai, S., Nakagawa, S., et al. (2017). Seismic waveform imaging of long-period ground motion in the Tokyo metropolitan area. Japan. Journal of Geophysical Research: Solid Earth, 122(7), 5435–5451. https://doi.org/10.1002/2017JB014276

Kasahara, K., Sakai, S., Morita, Y., Hirata, N., Tsuruoka, H., Nakagawa, S., et al. (2009). Development of the metropolitan seismic observation network (MeSO-net) for detection of mega-thrust beneath Tokyo metropolitan area. Bulletin of the Earthquake Research Institute, University of Tokyo, 84, 71–88.

Kato, A., Fukuda, J. I., Nakagawa, S., & Obara, K. (2016). Foreshock migration preceding the 2016 Mw 7.0 Kumamoto earthquake, Japan. Geophysical Research Letters, 43(17), 8945–8953. https://doi.org/10.1002/2016gl070079

Kawakita, Y., & Sakai, S. (2009). Various types of noise in MeSO-net. Bulletin of the Earthquake Research Institute, University of Tokyo, 84, 127–139.

Kong, Q., Trugman, D. T., Ross, Z. E., Bianco, M. J., Meade, B. J., & Gerstoft, P. (2019). Machine learning in seismology: Turning data into insights. Seismological Research Letters, 90(1), 3–14. https://doi.org/10.1785/0220180025

Kriegerovský, M., Petersen, G. M., Vasyura-Batkhe, H., & Ohnberger, M. (2019). A deep convolutional neural network for localization of clustered earthquakes based on multistation full waveforms. Seismological Research Letters, 90(2A), 510–516. https://doi.org/10.1785/0220180302

Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet classification with deep convolutional neural networks (25, pp. 1097–1105). Proceedings of Neural Information Processing Systems (NIPS 2012).

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. Neural Computation, 1(4), 541–551. https://doi.org/10.1162/neco.1989.1.4.541

Li, Z., Peng, Z., Hollis, D., Zhu, L., & McClellan, J. (2018). High-resolution seismic event detection using local similarity for large-N arrays. Scientific Reports, 8(1), 1646. https://doi.org/10.1038/s41598-018-19728-w

McBreyter, I. W., Delorey, A. A., & Johnson, P. A. (2019). Pairwise association of seismic arrivals with convolutional neural networks. Seismological Research Letters, 90(6), 503–509. https://doi.org/10.1785/0220180302

Mill, J., Tang, T., Kaminsky, Z., Khare, T., Yazdanpanah, S., Bouchard, L., et al. (2008). Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. The American Journal of Human Genetics, 82(3), 696–711. https://doi.org/10.1016/j.ajhg.2008.01.008

Mousavi, S., Ellsworth, W., Zhu, W., Chung, L., & Beroza, G. (2020). Earthquake transformer—An attentive deep-learning model for simulating earthquakes. Nature Geoscience, 13(1), 1–12. https://doi.org/10.1038/s41467-020-17591-w

Nakano, M., Sugiyama, D., Hori, T., Kuwata, T., & Tsuboi, S. (2019). Discrimination of seismic signals from earthquakes and tectonic tremor by applying a convolutional neural network to running spectral images. Seismological Research Letters, 90(2A), 530–538. https://doi.org/10.1785/0220180329

National Research Institute for Earth Science and Disaster Resilience. (2019). NIED Hi-Net. National Research Institute for Earth Science; Disaster Resilience. https://doi.org/10.17598/NIED.0003

Newman, M. E. J. (2006). Modularity and community structure in networks. Proceedings of the National Academy of Sciences of the United States of America, 103(23), 8577–8582. https://doi.org/10.1073/pnas.0601602103

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. Physical Review E, 69(2), 026113. https://doi.org/10.1103/physreve.69.026113

Ng, A. (2004). Feature selection, L1 vs L2 regularization, and rotational invariance. Proceedings of the Twenty-First International Conference on Machine Learning (ICML’04). Banff, Alberta, Canada. https://doi.org/10.1145/1015330.1015435

Niepert, M., Ahmed, M., & Kutzkov, K. (2016). Learning convolutional neural networks for graphs. Proceedings of the 33rd International Conference on International Conference on Machine Learning (ICML 2016), Barcelona, Spain.

Nishida, N., Kawakatsu, H., & Obara, K. (2008). Three-dimensional crustal S wave velocity structure in Japan using microseismic data recorded by Hi-net tiltmeters. Journal of Geophysical Research, 113, B10302. https://doi.org/10.1029/2007JB005395

Obara, K., Kasahara, K., Hori, S., & Okada, Y. (2005). A densely distributed high-sensitivity seismograph network in Japan: Hi-net by National Research Institute for Earth Science and Disaster Prevention. Review of Scientific Instruments, 76, 021301. https://doi.org/10.1063/1.1854197

Ogata, Y., Katsura, K., Tsuruoka, H., & Hirata, N. (2019). High-resolution 3D earthquake forecasting beneath the greater Tokyo area. Earth Planets and Space, 71(123), 1–14. https://doi.org/10.1186/s40623-019-1086-7

Okada, Y., Kasahara, K., Hori, S., Obara, K., Sekiguchi, S., Fujiwara, H., & Yamamoto, A. (2004). Recent progress of seismic observation networks in Japan—Hi-net, F-net, K-NET and KIK-net. Earth Planets and Space, 56, 15–28. https://doi.org/10.1186/1880-5876-56-15

Peng, Z., & Zhao, P. (2009). Migration of early aftershocks following the 2004 Parkfield earthquake. Nature Geoscience, 2, 877–881. https://doi.org/10.1038/ngeo697

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

Qian, N. (1999). On the momentum term in gradient descent learning algorithms. Neural Networks, 12(1), 145–151. https://doi.org/10.1016/s0893-6080(98)00116-6

Robbins, H., & Monro, S. (1951). A stochastic approximation method. The Annals of Mathematical Statistics, 22(3), 400–407. https://doi.org/10.1214/aoms/1177729586

Ross, Z. E., Meier, M.-A., & Hauksson, E. (2018). PiWave arrival picking and first-motion polarity determination with deep learning. Journal of Geophysical Research: Solid Earth, 123(6), 5120–5129. https://doi.org/10.1029/2017JB015251

Ross, Z. E., Trugman, D. T., Hauksson, E., & Shearer, P. M. (2019). Searching for hidden earthquakes in Southern California. Science, 366(6442), 767–771. https://doi.org/10.1126/science.aaw6888

Sakai, S., & Hirata, N. (2009). Distribution of the Metropolitan Seismic Observation network. Bulletin of the Earthquake Research Institute, University of Tokyo, 84, 57–70.
Shelly, D. R., Beroza, G. C., & Ide, S. (2007). Non-volcanic tremor and low-frequency earthquake swarms. *Nature*, 446, 305–307. https://doi.org/10.1038/nature05666

Sietsma, J., & Dow, R. J. F. (1991). Creating artificial neural networks that generalize. *Neural Networks*, 4(1), 67–79. https://doi.org/10.1016/0893-6080(91)90033-2

Skoumal, R. J., Brudzinski, M. R., Currie, B. S., & Levy, J. (2014). Optimizing multi-station earthquake template matching through re-examination of the Youngstown, Ohio, sequence. *Earth and Planetary Science Letters*, 405, 274–280. https://doi.org/10.1016/j.epsl.2014.08.033

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks (pp. 3104–3112). Proceedings of Advances in Neural Information Processing Systems (NIPS 2014).

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). *Going deeper with convolutions* (pp. 1–9). Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR 2015).

Wang, J., Xiao, Z., Liu, C., Zhao, D., & Yao, Z. (2019). Deep learning for picking seismic arrival times. *Journal of Geophysical Research: Solid Earth*, 124(7), 6612–6624. https://doi.org/10.1029/2019jb017536

Wessel, P., & Smith, W. H. F. (1998). New, improved version of generic mapping tools released. *Eos, Transactions, American Geophysical Union*, 79(47), 579. https://doi.org/10.1029/98eo00426

Withers, M., Aster, R., Young, C., Beiriger, J., Harris, M., Moore, S., & Trujillo, J. (1998). A comparison of select trigger algorithms for automated global seismic phase and event detection. *Bulletin of the Seismological Society of America*, 88(1), 95–106.

Xia, M., Wang, J., & He, Y. (2013). BrainNet Viewer: A network visualization tool for human brain connectomics. *PLoS One*, 8(7), e68910. https://doi.org/10.1371/journal.pone.0068910

Yoon, C., O’Reilly, O., Bergen, K., & Beroza, G. (2015). Earthquake detection through computationally efficient similarity search. *Science Advances*, 1(11), e1501057. https://doi.org/10.1126/sciadv.1501057

Zhang, X., Zhang, J., Yuan, C., Liu, S., Chen, Z., & Li, W. (2020). Locating induced earthquakes with a network of seismic stations in Oklahoma via a deep learning method. *Scientific Reports*, 10(1941), 1–12. https://doi.org/10.1038/s41598-020-58908-5

Zhu, W., & Beroza, G. (2018). PhaseNet: A deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, 216(1), 261–273. https://doi.org/10.1093/gji/ggy423