Seismic Event and Phase Detection Using Time–Frequency Representation and Convolutional Neural Networks

4 authors:

Ramin Dokht
Natural Resources Canada
15 PUBLICATIONS   **60** CITATIONS

SEE PROFILE

Honn Kao
Natural Resources Canada
190 PUBLICATIONS   **4,323** CITATIONS

SEE PROFILE

Ryan Visser
Natural Resources Canada
20 PUBLICATIONS   **53** CITATIONS

SEE PROFILE

Brindley Smith
McGill University
14 PUBLICATIONS   **50** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Seismic structures of the Western Canada Sedimentary Basin View project

Migration Imaging of the Java Subduction Zone View project

Ⓔ

Seismic Event and Phase Detection Using Time–Frequency Representation and Convolutional Neural Networks

by Ramin M. H. Dokht, Honn Kao, Ryan Visser, and Brindley Smith

ABSTRACT

The availability of abundant digital seismic records and successful application of deep learning in pattern recognition and classification problems enable us to achieve a reliable earthquake detection framework. To overcome the limitations and challenges of conventional methods, which are mainly due to an incomplete set of template waveforms and low signal-to-noise ratio, we design a generalized model to improve discrimination between earthquake and noise recordings using a deep convolutional network (ConvNet). Exclusively based on a dataset of over 4900 earthquakes recorded over a period of 3 yrs in western Canada, a multilayer ConvNet is trained to learn general characteristics of background noise and earthquake signals in the time–frequency domain. In the next step, we train a secondary network using the wavelet transform of the major seismic arrivals to separate P from S waves and estimate their approximate arrival times. The results of validation experiments demonstrate promising performance and achieve an average accuracy of nearly 99% for both networks. To investigate the applicability of our algorithm, we apply the trained model on an independent dataset recently recorded in northeastern British Columbia (NE BC). It is found that deep-learning-based methods are superior to traditional techniques in detecting a higher number of seismic events at significantly less computational cost.

Electronic Supplement: Tables reporting the performance of convolutional networks trained directly in the time domain, and figures showing the accuracy of the validation set and P- and S-wave error measurements.

INTRODUCTION

The recent increase in the rate of earthquake occurrence in western Canada, a region with a historically low level of background seismicity, has been largely attributed to the development of unconventional hydrocarbon resources (Horner *et al.,* 1994; Schultz *et al.,* 2014; Farahbod *et al.,* 2015; Rubinstein and Mahani, 2015; Atkinson *et al.,* 2016). Reliable ground-motion analyses and seismic hazard assessments require a complete earthquake catalog containing both natural and induced earthquakes. However, most of the conventional automated techniques fail to identify induced events of low magnitudes, which makes characterizing and locating these events a challenging task. In comparison, manual picking of seismic events has a higher detection rate, but the process is extremely laborious and remains subjective to the analyst's experience.

Among the various detection algorithms, the short-term average/long-term average (STA/LTA) technique, which measures the signal-to-noise ratio (SNR) function, has been widely used for detecting moderate-to-large earthquakes if a certain triggering threshold is exceeded (Allen, 1978; Withers *et al.,* 1998). On the other hand, cross-correlation-based techniques (also called template matching) have been extensively used to identify repeating earthquakes of lower magnitudes based on similarity measurements of the entire waveforms (Gibbons and Ringdal, 2006; Skoumal *et al.,* 2015; Caffagni *et al.,* 2016). Although less sensitive to high noise levels, template matching is computationally intensive and its application is limited to identifying earthquakes sharing the same source region and mechanism (Eisner *et al.,* 2006). This can be partially remedied by generalizing template matching using clustering of similar waveform fingerprints through a set of hash functions (Yoon *et al.,* 2015; Bergen *et al.,* 2016), and subspace analysis of a set of representative waveforms (Barrett and Beroza, 2014). Recent studies take advantage of spectral information of seismic records to improve the accuracy associated with detection of weak microseismic events (Galiana-Merino *et al.,* 2008; Vaezi and Van der Baan, 2015; Mousavi *et al.,* 2016), though *a priori* knowledge of noise characteristics is required.

Deep learning is a set of representation-learning algorithms with multiple layers of nonlinear transformations (Bengio *et al.,* 2013; LeCun *et al.,* 2015; Najafabadi *et al.,* 2015). Unlike conventional machine-learning techniques, which rely on carefully hand-engineered features (Wang and Teng, 1995; Gentili and Michelini, 2006; Maity *et al.,* 2014), deep learning allows a model to learn a general-purpose representation of raw data

using a training set (Dahl *et al.*, 2013; Wang and Yeung, 2013). Deep representation learning has been widely applied to several research areas such as natural language understanding (Collobert *et al.*, 2011; Mikolov *et al.*, 2013), image classification (Krizhevsky *et al.*, 2012), and speech recognition (Dahl *et al.*, 2012; Hinton *et al.*, 2012). Recent studies have investigated the application of deep-learning techniques for earthquake detection (e.g., Perol *et al.*, 2018; Ross *et al.*, 2018; Zhu and Beroza, 2018) and seismic imaging (Araya-Polo *et al.*, 2018; Moseley *et al.*, 2018).

Our research is motivated by the successful application of deep convolutional networks (ConvNets) to overcome the limitations of traditional techniques in studying induced seismicity in Oklahoma, United States (Perol *et al.*, 2018). The basic idea is to train a ConvNet model on a large dataset of previously recorded earthquakes, so that the classifier can be generalized to identify seismic events different from those used in training. In comparison with the model proposed by Perol *et al.* (2018), the current study presents additional improvements by including both temporal and spectral information of three-component seismograms to enhance the detection accuracy. We divide the automatic earthquake detection process into two steps. A pretrained model is first adopted to separate earthquakes from nonearthquake signals in the time–frequency domain. Then, we build up a secondary supervised classification system using higher-resolution spectral images of earthquake records to discriminate between P and S waves.

METHOD AND DATASET

Convolutional Neural Networks

ConvNets are feed-forward, multilayer neural networks, which were introduced to process multidimensional arrays (LeCun *et al.*, 1998). A ConvNet transforms an input volume of data features to output class probabilities through a sequence of several hidden units consisting of convolutional, pooling, and fully connected layers. Each convolutional layer contains a bank of linear filters to extract local features at all parts on the previous layer (Cireşan *et al.*, 2012) and passes the resulting convolutional responses through a nonlinear activation unit. ConvNets have been found to train several times faster by mapping all negative responses to zero using a rectified linear unit (Nair and Hinton, 2010; Glorot *et al.*, 2011):

$$x_j = \max(0, b_j + \sum_i w_{ij} x_i), \qquad (1)$$

in which b_j and w_{ij} are the bias and weights of the jth neuron passing the input from the previous layer x_i to the output feature map x_j, respectively. It is common to spatially downsample the resulting activations by merging similar local features into one (LeCun *et al.*, 2015). The subsampling, also known as pooling, operation can significantly reduce the number of free parameters, thereby improving the performance of the network and avoiding overfitting (Krizhevsky *et al.*, 2012). Max pooling is the most widely used subsampling operator; this calculates the

maximum activation over nonoverlapping local neighborhoods (Serre *et al.*, 2005; Cireşan *et al.*, 2012), and thus reduces the variability to small temporal/spatial transformations (LeCun *et al.*, 1990; Yang *et al.*, 2010; Farabet *et al.*, 2013). The output feature maps, resulted from a sequence of convolutional, nonlinear activation, and pooling layers, are concatenated and passed to a fully connected layer in which every neuron is linearly connected to all activations in the previous layer. The output of the last fully connected layer is fed to a normalized exponential function (softmax classifier), which calculates a probability distribution over C different possible classes (Peterson and Söderberg, 1989):

$$p_j = \frac{\exp(x_j)}{\sum_{k=1}^{C} \exp(x_k)} \qquad k = 1, ..., C. \qquad (2)$$

The objective of the current study is to find a set of learnable free parameters, which minimizes the misfit between the predicted p and ground-truth scores g of N instances using a L2-regularized multinomial logistic loss function:

$$J = \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{C} -g_k^{(n)} \log(p_k^{(n)}) + \lambda \sum_i \|\mathbf{W}_i\|_2^2. \qquad (3)$$

The regularization parameter λ controls the trade-off between the data misfit and model constraints, and $\mathbf{W}_i$ are the model parameters (bias and weights) of the ith layer. The solution to the objective function is found using the gradient descent technique, which calculates the updates on network parameters at each iteration t by a linear combination of the negative gradient of the loss function and the model update from the previous iteration $t - 1$:

$$\Delta \mathbf{W}_t = \mu \Delta \mathbf{W}_{t-1} - \alpha \nabla J(\mathbf{W}_t), \qquad (4)$$

in which the learning rate α is the weight of the negative gradient, and the momentum μ controls the network parameter update at each iteration. For a large training set, it would be more efficient to estimate the stochastic approximation of the cost function by drawing a small random selection (minibatch) of the training set at each iteration.

Earthquake Detection

In the first step of our earthquake detection framework, we design a ConvNet to scan continuous seismic records and separate between noise and earthquake signals. Exclusively based on a comprehensive earthquake catalog from western Canada, a classifier is trained to identify coherent high-power earthquake signals between three-component seismograms. In this region, the Geological Survey of Canada has reported 4914 earthquakes with local magnitudes M_L ranging from 0.1 to 4.9 between January 2014 and December 2016 (Fig. 1; Visser *et al.*, 2017). After deconvolving the instrument response from the ground velocity records, we apply a 2 Hz high-pass filter and resample the data to 20 Hz. The resulting waveforms are visually
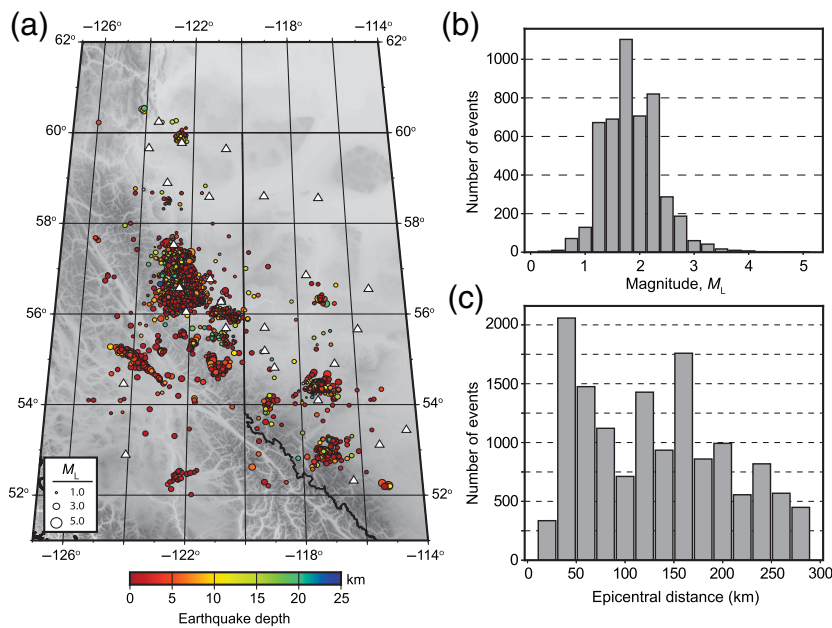
▲ **Figure 1.** (a) Distributions of earthquakes (circles) and seismic stations (white triangles) used in this study. The warm and cold colors correspond to shallow and deep earthquakes, respectively. The circle sizes represent the earthquake magnitudes. (b) A histogram of event local magnitude distribution with mean and standard deviation values of 1.8 and 0.55, respectively. (c) Distribution of source–receiver distances using a 20 km bin size.

inspected and low-quality data are discarded. The final dataset contains 13,949 earthquake–receiver pairs with epicentral distances up to 275 km (Fig. 1c) and a total of 148,000 noise records selected from time segments free of earthquake arrivals. To avoid overfitting due to the uneven number of observations per category, the earthquake signals are randomly repeated, time-shifted, and contaminated with Gaussian noise.

Unlike the previous studies (Perol *et al.*, 2018; Ross *et al.*, 2018), which perform the classification task in the time domain, the current model attempts to learn the general characteristics associated with the earthquake and nonearthquake signals in the time–frequency domain. Because seismic event and noise recordings have intrinsically different spectral contents, the performance of the network can be further improved by learning from the spectro-temporal representations of waveform data (Wang and Teng, 1995; Vaezi and Van der Baan, 2015). The first layer of the ConvNet is represented by spectrograms of 100-s-long data segments calculated using the short-time Fourier transform (STFT; Portnoff, 1980). The spectrograms are computed using a moving Hanning window of 40 samples with 50% overlap between successive windows (Fig. 2). The resulting spectrograms are then normalized by their maximum spectral values and therefore become independent of earthquake magnitude.

The architecture of earthquake detection ConvNet consists of a sequence of four convolutional layers, each followed by a pooling layer (Table 1). The supervised classifier is trained on a random subset consisting of 75% of spectrograms (training set) and its performance is evaluated in terms of detection

accuracy of the remaining 25% observations (validation set). The training is run for 50,000 iterations with a trade-off parameter $\lambda = 10^{-4}$, an initial learning rate of 0.01 decaying to 0.005, and a fixed moment rate of 0.9. The training process is performed using a mini-batch of 200 samples per iteration and takes approximately 4 hrs on an Intel Core i7 CPU (4 GHz).

Phase Identification

The goal of the next experiment is to design a secondary classifier using the existing earthquake catalog for separating between different body waves. Robust measurements of phase arrivals are achieved by analyzing the nonstationary, multi-component seismic signals using the wavelet transform (WT) (Zhang *et al.*, 2003; Ahmed *et al.*, 2007; Galiana-Merino *et al.*, 2008). Compared with the STFT spectrogram, the WT has higher temporal resolution, but it suffers from the energy spread out along the frequency axis. The synchrosqueezing wavelet transform (SWT) is a new time–frequency analysis technique that combines the conventional WT with a frequency reassignment method to enhance the time–frequency localization (Daubechies and Maes, 1996; Daubechies *et al.*, 2011). Assuming that the energy of a seismic event is concentrated in a few high-amplitude wavelet coefficients, Mousavi *et al.* (2016) introduced an SWT-based technique to simultaneously suppress nonstationary random noise and detect onset times of weak microseismic events using a characteristic function of the thresholded wavelet coefficients.

In this step, we train a separate network using a dataset of high-resolution wavelet power spectra of major seismic arrivals picked by expert analysts. The new dataset consists of 11,500 *P*-wave, 11,500 *S*-wave, and 45,000 noise windows of 5 s long each. The windows containing both *P* and *S* waves are repeated and centered on the arrival time of each phase separately. We applied the same data augmentation strategy described in the Earthquake Detection section to ensure that there is an equal number of observations in each class. Then, we employed a Morlet wavelet, product of a complex exponential and a Gaussian envelope, as the basis function to calculate the SWT of windowed data (Fig. 3). The final dataset is eventually divided into the training and validation sets including time–frequency representations of waveform data normalized by the maximum value of wavelet coefficients of all three components. The architecture of phase identification network and its model parameters are similar to those of earthquake detection. However, the input layer consists of three maps of 80×101 neurons, and all convolutional layers have a fixed filter size of 3×5. The pooling layers have local receptive fields of size 2×2 with a constant stride of 2 in both dimensions, the first fully connected layer has 192 neurons and the output layer is a vector of 3 class labels (Table 2).
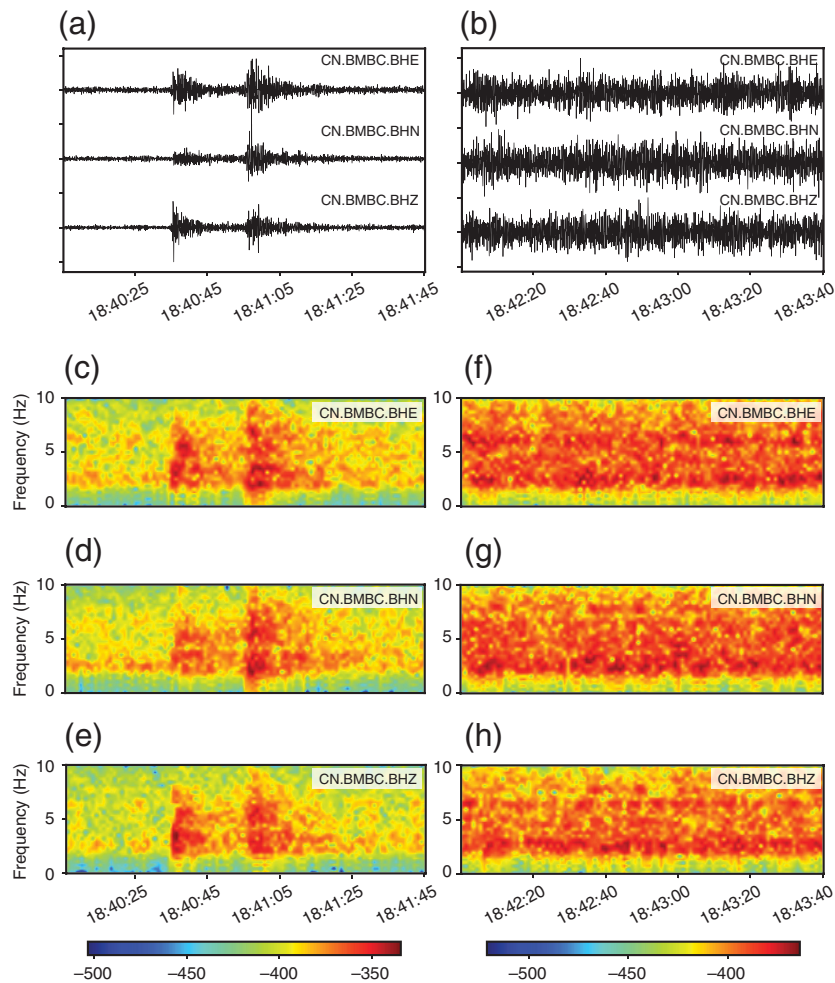
▲ **Figure 2.** Three-component waveforms of (a) a 2.4 magnitude earthquake recorded by CN.BMBC at 164 km distance on 25 February 2014, and (b) a postevent noise window recorded at the same station. (c–e) The short-time Fourier transform of earthquake records in (a). (f–h) Same as (c–e) but calculated for the noise window in (b).

RESULT

We evaluate the performance of the proposed framework on validation datasets selected independent of training sets. While the solver tries to minimize the objective function using the training samples (learning curves in Fig. 4a,b), validation data are employed to measure the generalization capability of the trained network for correct classification of new samples that have never been considered during the training phase (Table 3). The learning curves show that the earthquake detection algorithm rapidly converges to the optimal solution after 20,000 iterations, but it takes ∼40,000 iterations until the objective function becomes flat for the seismic phase identification model. The neural network produces a vector of confidence scores for all class labels, indicating the probability of a particular input belonging to a given category. The accuracy of correct classification is calculated every 5000 iterations and the model with the highest accuracy on the validation set is selected as the final classifier (Ⓔ Fig. S1, available in the electronic supplement to this article).

For earthquake detection, the network accuracy reaches 99.8% on the validation set and a total number of only 40 earthquake windows are mislabeled. This is comparable to that obtained from the training dataset, which yields respective accuracies of 99.7% and 99.95% for earthquake and nonearthquake categories (see Table 3). On the other hand, the phase identification network results in a slightly reduced accuracy on average. The second classifier gives an average error of 0.7% for the entire training set and predicts

Table 1
Architecture of Convolutional Network (ConvNet) Used for Seismic Event Detection

Layer	Type	Kernel Size	Stride	Output
1	Input data	—	—	3 maps of 33 × 99 neurons
2	ConvReLU	5 × 7	—	16 maps of 29 × 93 neurons
3	Maxpool	1 × 2	(1, 2)	16 maps of 29 × 47 neurons
4	ConvReLU	5 × 5	—	16 maps of 25 × 43 neurons
5	Maxpool	1 × 2	(1, 2)	16 maps of 25 × 22 neurons
6	ConvReLU	3 × 3	—	16 maps of 23 × 20 neurons
7	Maxpool	2 × 2	(2, 2)	16 maps of 12 × 10 neurons
8	ConvReLU	3 × 3	—	16 maps of 10 × 8 neurons
9	Maxpool	2 × 2	(2, 2)	16 maps of 5 × 4 neurons
10	FC	—	—	320 neurons
11	FC	—	—	2 neurons

ConvReLU, a convolutional layer followed by a rectified linear unit; Maxpool, a max pooling layer; FC, a fully connected layer.
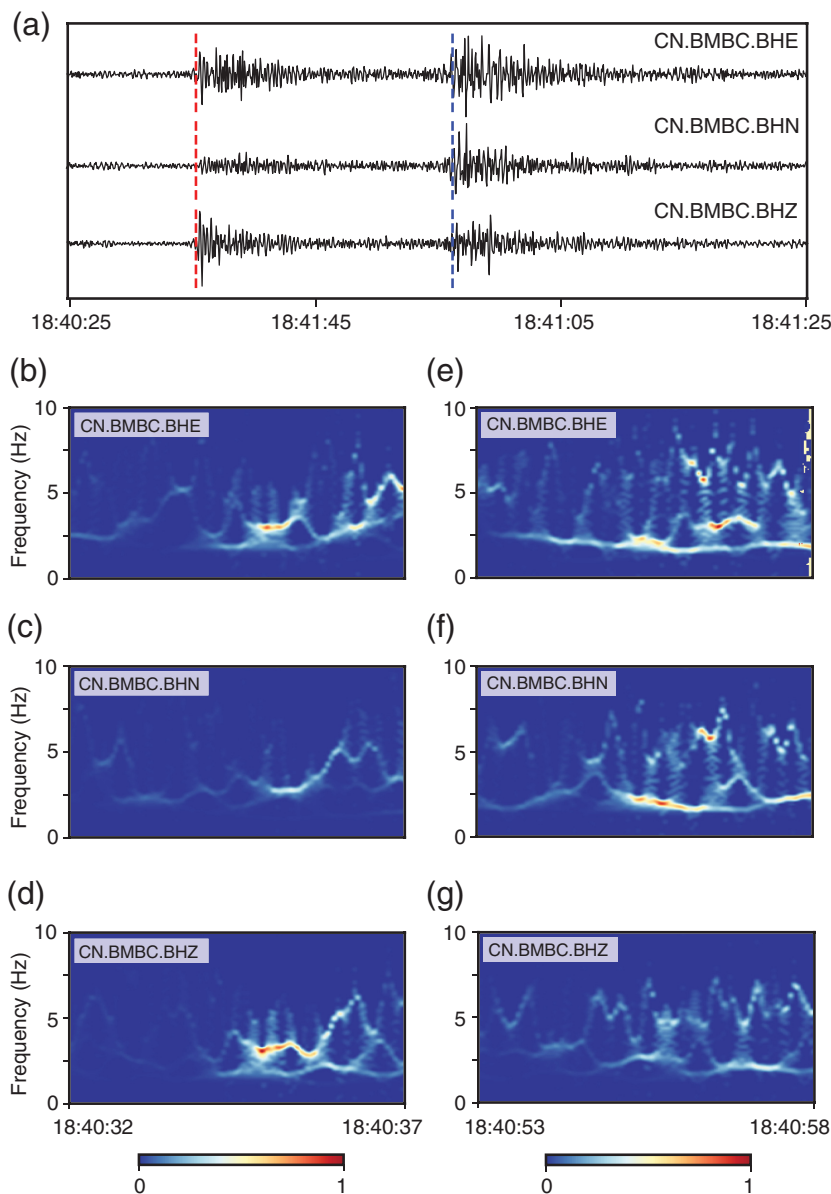
▲ **Figure 3.** (a) Three-component waveforms of the earthquake shown in Figure 2a. Red and blue dashed lines indicate the *P*- and *S*-wave arrival times, respectively. (b–d) The normalized synchrosqueezing wavelet transform (SWT) of a 5-s-long window centered on the *P*-wave arrival. (e–g) The normalized SWT of a 5-s-long window centered on the *S*-wave arrival.

correct labels for approximately 98.7% and 98.4% of *P*- and *S*-wave windows in the validation set, respectively.

The performance of the proposed model is compared with that obtained from a model that is trained directly in the time domain using the same dataset. Although the difference between the earthquake detection networks is insignificant, the total number of misclassified samples obtained from our model is nearly 40% of that of the time-domain classifier (see classification accuracies in Table 3 and Ⓔ Table S1 and confusion matrices in Table 4 and Ⓔ Table S2). However, the difference becomes more pronounced for the phase identification process. On average, the model trained using the time–frequency features

improves the accuracy by up to 3.2%. Although the time-domain model misidentifies 580 of *P* phases and 588 of *S* phases in the validation set, our model provides a higher recall rate and misses only 137 and 79 of *P* and *S* waves, respectively (see confusion matrices in Table 5 and Ⓔ Table S3).

The earthquake detection and phase identification results indicate that the classification performance can be affected by the minimum accepted confidence score (Fig. 4). We define a hard-thresholding function, which discards earthquakes whose scores are less than a given probability threshold (LeCun *et al.*, 1990). A careful selection of probability threshold is required because a low threshold value can increase the number of false detections, while higher threshold values may result in erroneous rejection of true events. The distribution of the number of earthquakes as a function of detection score shows a gradual increase for confidence scores greater than 0.6 with a sudden rise at a detection score of 0.95 (Fig. 4c). The same observations are made for the phase separation results as nearly 98.5% of correctly predicted examples fall in the last probability bin (Fig. 4d).

Induced Seismicity in Northeast British Columbia

After successful training and validation, the ConvNet models were employed for monitoring the seismic activity in Fort St. John and near Dawson Creek, an area with an increased number of potentially induced earthquakes in recent years (Horner *et al.*, 1994; British Columbia Oil and Gas Commission, 2014; Visser *et al.*, 2017). The network was implemented on continuous records from an array of nine broadband stations, deployed in August 2017, which are not included in either training or validation set. For each station, the daily records are first divided into 100-s-long windows (with 50% overlap between windows) and the STFT representations of windowed data are calculated. If the first classifier detects a seismic event at a probability threshold of 0.95, the SWT of the corresponding time window will be calculated and scanned by the second network to identify *P* and *S* waves. The earthquake detection step takes ∼5 s for a three-component daily recording, which is ∼30 times faster than the STA/LTA technique. To assess the detection performance and generalization power of the model, all detected events are visually inspected. It turns out that only 13 events are false detections (2% of the total number of detected events), 9 of which are eventually eliminated by the phase identification classifier. In comparison, the time-domain model returns ∼25% more false
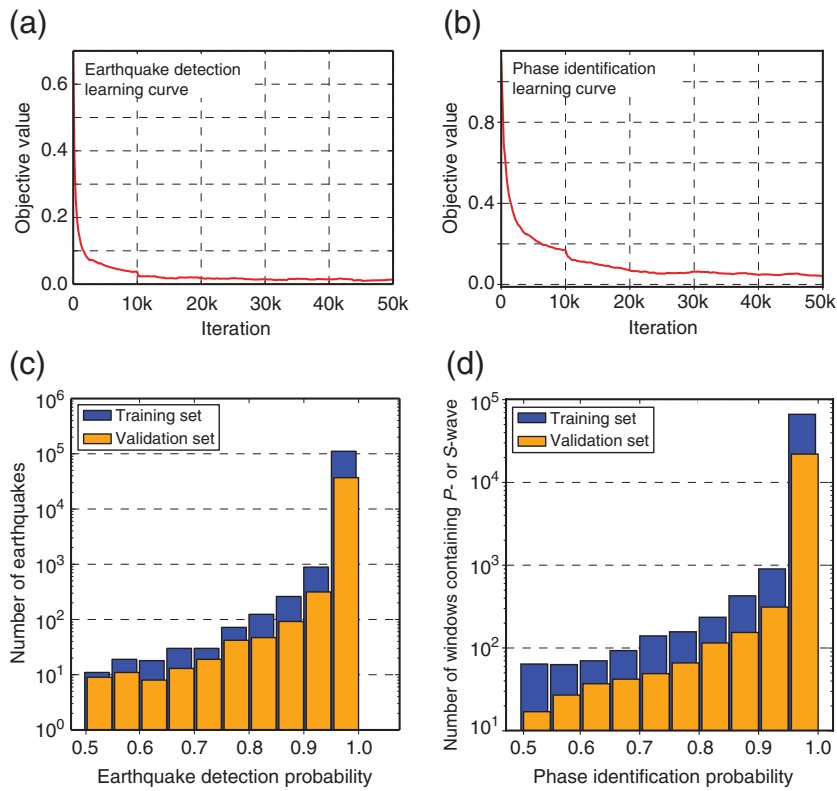
▲ **Figure 4.** Learning curves of the (a) earthquake detection, and (b) phase identification networks showing the reduction in objective function value during the training process. Distributions of correctly labeled (c) earthquakes and (d) seismic phases as a function of detection score.

positives than its time–frequency counterpart. The phase identification network is capable of detecting multiple P- and S-wave arrivals if more than one event exists within a time window (Fig. 5 and Ⓔ Fig. S2). Our technique finds 652 events with epicentral distances ranging from 1 to 28 km on 5 September 2017 (Fig. 6). The local magnitudes of detected earthquakes vary from −0.5 to 1.1, which remain below the

minimum magnitude in the training catalog for nearly 30% of events.

The automatic detection results corroborate the observations obtained from manual picking and show significant temporal variations (Fig. 6). The ConvNet model detects ∼20% more earthquakes than previously reported by an expert analyst, though P waves are identified for only 40% of detected events. A possible explanation for this observation includes the difficulties in separating P and S phases for earthquakes recorded at shorter distances. For the remaining 60%, the P-wave arrivals are included mostly in the S-wave windows because the differential times between the two phases are less than 1.5 s (nearly 80% of detected events are recorded at epicentral distances ≤ 10 km). It is worth noting that the time–frequency phase classifier identifies 52 more P-wave windows than the model trained in the time domain. The ConvNet models trained in the time–frequency and time domains miss 31 and 39 of the events identified by a human analyst, respectively. A lower probability threshold allows the network to detect smaller events, but this may result in false detections.

DISCUSSION

Most of the existing earthquake detection techniques are poorly suited for identifying low-magnitude events, and their application in areas with no record of seismicity is limited by the absence of template earthquake waveforms (Yoon *et al.*, 2015). The main goal of this research is to present a detection framework independent of earthquake magnitude, epicentral distance, and noise level. We propose a method utilizing multiresolution time–frequency analysis and advances in deep learning to achieve a robust earthquake detection. In comparison with the time-domain analysis, the

Table 2
Architecture of ConvNet Used for Seismic Phase Identification

Layer	Type	Kernel Size	Stride	Output
1	Input data	—	—	3 maps of 80 × 101 neurons
2	ConvReLU	3 × 5	—	16 maps of 78 × 97 neurons
3	Maxpool	2 × 2	(2, 2)	16 maps of 39 × 49 neurons
4	ConvReLU	3 × 5	—	16 maps of 37 × 45 neurons
5	Maxpool	2 × 2	(2, 2)	16 maps of 19 × 23 neurons
6	ConvReLU	3 × 5	—	16 maps of 17 × 19 neurons
7	Maxpool	2 × 2	(2, 2)	16 maps of 9 × 10 neurons
8	ConvReLU	3 × 5	—	16 maps of 7 × 6 neurons
9	Maxpool	2 × 2	(2, 2)	16 maps of 4 × 3 neurons
10	FC	—	—	192 neurons
11	FC	—	—	3 neurons

ConvReLU, a convolutional layer followed by a rectified linear unit; Maxpool, a max pooling layer; FC, a fully connected layer.

(a)

(b)

(c)

(d)

(e)
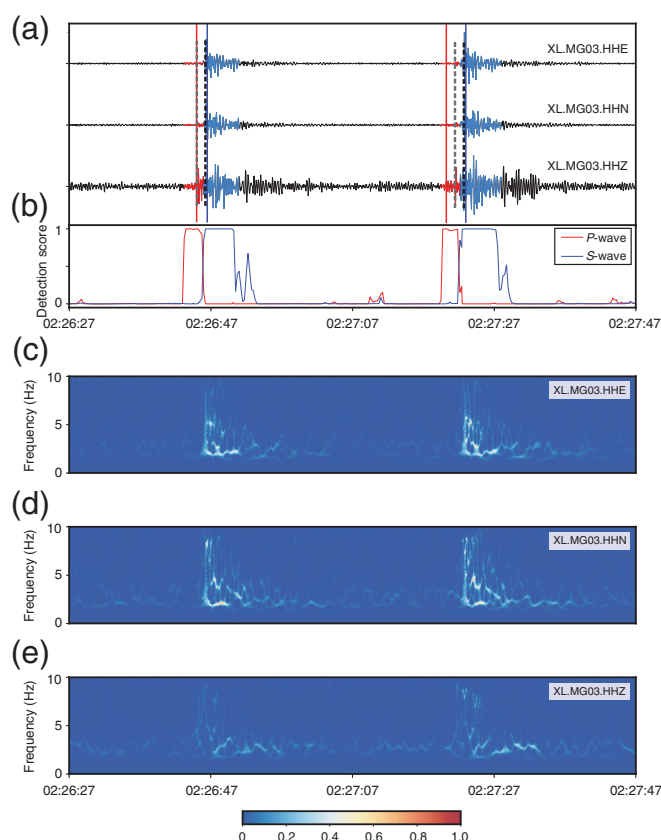
Figure 5. (a) A 100-s-long time segment labeled as seismic event by the earthquake detection network. The second network identifies two separate events within this time segment using a 5-s-long sliding window. Red and blue colors represent windows containing P and S waves, respectively. Vertical red and blue lines mark the P- and S-wave arrival times using their corresponding detection scores, respectively. Dashed gray and black lines indicate the manually picked P and S waves, respectively. (b) Continuous functions of the P- (red) and S-wave (blue) detection scores. (c–e) The SWT of waveform records presented in (a).
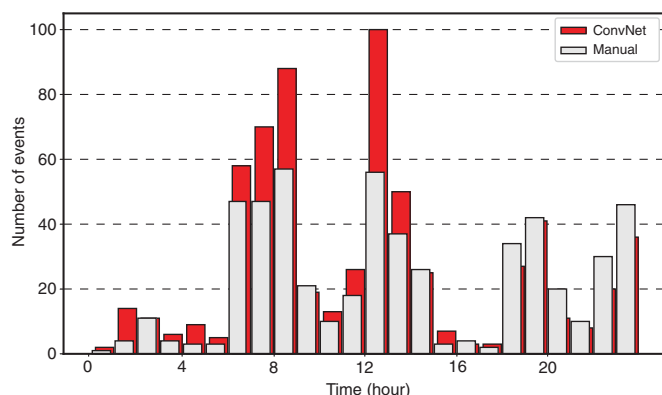
Figure 6. Distribution of detected earthquakes over the period of one day, on 5 September 2017, in the Dawson Creek area. Convolutional network (ConvNet) detects a total number of 652 events (red bars), which is ∼20% more than the manually picked earthquakes (gray bars).

Table 3
Classification Accuracies of the Earthquake Detection and Seismic Phase Identification ConvNets Trained in the Time–Frequency Domain

	Training Set (%)	Validation Set (%)
Earthquake detection accuracy		
Seismic event	99.7	99.6
Noise	99.95	99.9
Phase identification accuracy		
P-wave	99.1	98.7
S-wave	99.1	98.4
Noise	99.7	99.4

Table 4
Validation Set Confusion Matrix Calculated for the Earthquake Detection Network Trained Based on the Time–Frequency Feature Maps

		Predicted Labels	
		Seismic event	Noise
True labels	Seismic event	**37,460**	40
	Noise	142	**36,882**

Table 5
Validation Set Confusion Matrix Calculated for the Phase Identification Network Trained Based on the Time–Frequency Feature Maps

		Predicted Labels		
		P-wave	S-wave	Noise
True labels	P-wave	**11,363**	90	47
	S-wave	58	**11,421**	21
	Noise	87	91	**11,322**

application of SWT enables the network to learn the spectral structures of noise and seismic signals, which has been found to increase the accuracy of phase identification process by up to ∼3%. However, the calculation of WTs of continuous waveforms is computationally expensive and may not efficiently scale to the large-array data. To reduce the processing time, the phase identification network scans only the earthquake time windows that are detected by the preliminary ConvNet using the STFT-based spectrograms.

The learning ability of a neural network can vary with changes in its structure (Wang and Teng, 1995). The sensitivity of the proposed method with respect to the network architecture was explored by (1) increasing the number of filters in convolutional layers, and (2) replacing the pooling with strided convolutions (Perol *et al.*, 2018). We found no significant variation in model performance in both cases, though the
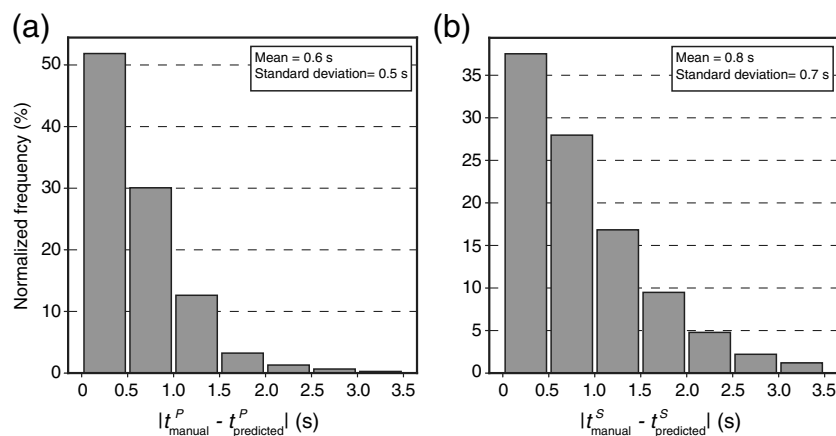
▲ **Figure 7.** Distributions of errors between manual time picks and those predicted from the detection scores for (a) *P* and (b) *S* waves. The average error in predicted *P*-wave time is less than that of the *S*-wave.

training time dramatically rises with the increasing number of free parameters (the former case). To avoid overtraining, a constant dropout rate of 10% was applied to the input and convolutional layers, which forces the network to learn more robust features by ignoring a random subset of neurons during the training process. An optimum dropout rate provides the best trade-off between the training data misfit and validation set accuracy. The near-identical performance of the proposed model on both training and validation sets reflects the generalization power of the model beyond the existing earthquake catalog.

In addition to phase separation, the second classifier provides a means to estimate the phase arrival times, which are required to determine earthquake location and other source parameters. Ross *et al.* (2018) suggest that the location of maximum detection score of each phase can be used to measure its approximate travel time if multiple successive windows identify the same phase (see Fig. 5b). The maximum difference between the manual picks and those obtained from the detection scores is on the order of one half of the sliding window length (Fig. 7). The accuracy of automatic time picking is generally affected by the SNR, and weaker signals result in larger errors (Ⓔ Fig. S3). However, the implementation of the WT enables noise reduction, which can further reduce the picking error and possibly automate the earthquake location process (a topic beyond the scope of this article, but we refer readers to Mousavi *et al.*, 2016, for more detailed discussion).

CONCLUSION

Based on recent advances in deep learning, we propose a ConvNet algorithm for robust detection of seismic events to address the shortcomings of existing methods. Our technique relies on the detectability of earthquake signals in the time–frequency domain and takes advantage of the spectral characteristics of phase arrivals to separate *P* and *S* waves. In comparison with manual detection, our technique can identify ∼20% more events while significantly reducing the processing time and improving the efficiency. In addition to event detection, this approach provides initial estimates of phase onset times, which can be used to determine preliminary earthquake locations. Highly improved accuracy and reliable rejection of false detections are achieved by joint application of the event detection and phase identification ConvNets. The proposed approach can be potentially utilized to enhance real-time monitoring of both natural and induced seismicity.

DATA AND RESOURCES

The regional earthquake catalog used in this study was compiled by the Geological Survey of Canada (http://publications.gc.ca/site/eng/9.856883/publication.html, last accessed September 2018). Waveform data can be collected from the Incorporated Research Institutions for Seismology (IRIS) Data Management Center at https://ds.iris.edu/ds/nodes/dmc/ (last accessed September 2018). We used Mocha, a deep-learning framework for Julia, to train the convolutional networks (the latest version of Mocha is available at https://mochajl.readthedocs.io/en/latest/, last accessed September 2018). Some figures were generated using the Generic Mapping Tools (GMT) v.5.4.2 (www.soest.hawaii.edu/gmt, last accessed September 2018; Wessel and Smith, 1998). ⧉

ACKNOWLEDGMENTS

REFERENCES

Ahmed, A., M. Sharma, and A. Sharma (2007). Wavelet based automatic phase picking algorithm for 3-component broadband seismological data, *J. Seismol. Earthq. Eng.* **9,** nos. 1/2, 15–24.

Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces, *Bull. Seismol. Soc. Am.* **68,** no. 5, 1521–1532.

Araya-Polo, M., J. Jennings, A. Adler, and T. Dahlke (2018). Deep-learning tomography, *The Leading Edge* **37,** no. 1, 58–66.

Atkinson, G. M., D. W. Eaton, H. Ghofrani, D. Walker, B. Cheadle, R. Schultz, R. Shcherbakov, K. Tiampo, J. Gu, R. M. Harrington, *et al.* (2016). Hydraulic fracturing and seismicity in the western Canada sedimentary basin, *Seismol. Res. Lett.* **87,** no. 3, 631–647.

Barrett, S. A., and G. C. Beroza (2014). An empirical approach to subspace detection, *Seismol. Res. Lett.* **85,** no. 3, 594–600.

Bengio, Y., A. Courville, and P. Vincent (2013). Representation learning: A review and new perspectives, *IEEE Trans. Pattern Anal. Mach. Intell.* **35,** no. 8, 1798–1828.

Bergen, K., C. Yoon, and G. C. Beroza (2016). Scalable similarity search in seismology: A new approach to large-scale earthquake detection, *International Conf. on Similarity Search and Applications*, Tokyo, Japan, 24–26 October 2016, 301–308.

British Columbia Oil and Gas Commission (2014). Investigation of observed seismicity in the Montney Trend, available at http://www.bcogc.ca/investigation-observed-seismicity-montney-trend (last accessed December 2018).

Caffagni, E., D. W. Eaton, J. P. Jones, and M. van der Baan (2016). Detection and analysis of microseismic events using a Matched Filtering Algorithm (MFA), *Geophys. J. Int.* **206,** no. 1, 644–658.

Cireşan, D., U. Meier, J. Masci, and J. Schmidhuber (2012). Multi-column deep neural network for traffic sign classification, *Neural Networks* **32,** 333–338.

Collobert, R., J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa (2011). Natural language processing (almost) from scratch, *J. Mach. Learn. Res.* **12,** 2493–2537.

Dahl, G. E., T. N. Sainath, and G. E. Hinton (2013). Improving deep neural networks for LVCSR using rectified linear units and dropout, *2013 IEEE International Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, BC, Canada, 26–31 May 2013, 8609–8613.

Dahl, G. E., D. Yu, L. Deng, and A. Acero (2012). Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition, *IEEE Trans. Audio Speech Lang. Process.* **20,** no. 1, 30–42.

Daubechies, I., and S. Maes (1996). A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models, in *Wavelets in Medicine and Biology*, A. Aldroubi and M. Unser (Editors), CRC Press, Boca Raton, Florida, 527–546.

Daubechies, I., J. Lu, and H.-T. Wu (2011). Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool, *Appl. Comput. Harmon. Anal.* **30,** no. 2, 243–261.

Eisner, L., T. Fischer, and J. H. Le Calvez (2006). Detection of repeated hydraulic fracturing (out-of-zone growth) by microseismic monitoring, *The Leading Edge* **25,** no. 5, 548–554.

Farabet, C., C. Couprie, L. Najman, and Y. LeCun (2013). Learning hierarchical features for scene labeling, *IEEE Trans. Pattern Anal. Mach. Intell.* **35,** no. 8, 1915–1929.

Farahbod, A. M., H. Kao, D. M. Walker, and J. F. Cassidy (2015). Investigation of regional seismicity before and after hydraulic fracturing in the Horn River basin, northeast British Columbia, *Can. J. Earth Sci.* **52,** no. 2, 112–122.

Galiana-Merino, J. J., J. L. Rosa-Herranz, and S. Parolai (2008). Seismic P phase picking using a Kurtosis-based criterion in the stationary wavelet domain, *IEEE Trans. Geosci. Remote Sens.* **46,** no. 11, 3815–3826.

Gentili, S., and A. Michelini (2006). Automatic picking of P and S phases using a neural tree, *J. Seismol.* **10,** no. 1, 39–63.

Gibbons, S. J., and F. Ringdal (2006). The detection of low magnitude seismic events using array-based waveform correlation, *Geophys. J. Int.* **165,** no. 1, 149–166.

Glorot, X., A. Bordes, and Y. Bengio (2011). Deep sparse rectifier neural networks, *Proceedings of the Fourteenth International Conf. on Artificial Intelligence and Statistics*, Fort Lauderdale, Florida, 11–13 April 2011, 315–323.

Hinton, G., L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, *et al.* (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups, *IEEE Signal Process. Mag.* **29,** no. 6, 82–97.

Horner, R. B., J. E. Barclay, and J. M. MacRae (1994). Earthquakes and hydrocarbon production in the Fort St. John area of northeastern British Columbia, *Can. J. Explor. Geophys.* **30,** no. 1, 39–50.

Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012). Imagenet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems*, Lake Tahoe, Nevada, 3–6 December 2012, 1097–1105.

LeCun, Y., Y. Bengio, and G. Hinton (2015). Deep learning, *Nature* **521,** no. 7553, 436.

LeCun, Y., B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel (1990). Handwritten digit recognition with a back-propagation network, *Advances in Neural Information Processing Systems*, Denver, Colorado, 27–30 November 1989, 396–404.

LeCun, Y., L. Bottou, Y. Bengio, and P. Haffner (1998). Gradient-based learning applied to document recognition, *Proc. IEEE* **86,** no. 11, 2278–2324.

Maity, D., F. Aminzadeh, and M. Karrenbach (2014). Novel hybrid artificial neural network based autopicking workflow for passive seismic data, *Geophys. Prospect.* **62,** no. 4, 834–847.

Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, and J. Dean (2013). Distributed representations of words and phrases and their compositionality, *Advances in Neural Information Processing Systems*, Lake Tahoe, Nevada, 5–10 December 2013, 3111–3119.

Moseley, B., A. Markham, and T. Nissen-Meyer (2018). Fast approximate simulation of seismic waves with deep learning, available at https://arxiv.org/abs/1807.06873 (last accessed December 2018).

Mousavi, S. M., C. A. Langston, and S. P. Horton (2016). Automatic microseismic denoising and onset detection using the synchrosqueezed continuous wavelet transform, *Geophysics* **81,** no. 4, V341–V355.

Nair, V., and G. E. Hinton (2010). Rectified linear units improve restricted Boltzmann machines, *Proc. of the 27th International Conf. on Machine Learning (ICML-10)*, Haifa, Israel, 21–24 June 2010, 807–814.

Najafabadi, M. M., F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic (2015). Deep learning applications and challenges in big data analytics, *J. Big Data* **2,** no. 1, 1.

Perol, T., M. Gharbi, and M. Denolle (2018). Convolutional neural network for earthquake detection and location, *Sci. Adv.* **4,** no. 2, e1700578.

Peterson, C., and B. Söderberg (1989). A new method for mapping optimization problems onto neural networks, *Int. J. Neural Syst.* **1,** no. 1, 3–22.

Portnoff, M. (1980). Time–frequency representation of digital signals and systems based on short-time Fourier analysis, *IEEE Trans. Acoust. Speech Signal Process.* **28,** no. 1, 55–69.

Ross, Z. E., M. A. Meier, E. Hauksson, and T. H. Heaton (2018). Generalized seismic phase detection with deep learning, available at https://arxiv.org/abs/1805.01075 (last accessed December 2018).

Rubinstein, J. L., and A. B. Mahani (2015). Myths and facts on wastewater injection, hydraulic fracturing, enhanced oil recovery, and induced seismicity, *Seismol. Res. Lett.* **86,** no. 4, 1060–1067.

Schultz, R., V. Stern, and Y. J. Gu (2014). An investigation of seismicity clustered near the Cordel Field, west central Alberta, and its relation to a nearby disposal well, *J. Geophys. Res.* **119,** no. 4, 3410–3423.

Serre, T., L. Wolf, and T. Poggio (2005). Object recognition with features inspired by visual cortex, *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition 2005 (CVPR 2005)*, Vol. 2, San Diego, California, 20–25 June 2005, 994–1000.

Skoumal, R. J., M. R. Brudzinski, and B. S. Currie (2015). Earthquakes induced by hydraulic fracturing in Poland Township, Ohio, *Bull. Seismol. Soc. Am.* **105,** no. 1, 189–197.

Vaezi, Y., and M. Van der Baan (2015). Comparison of the STA/LTA and power spectral density methods for microseismic event detection, *Mon. Not. R. Astron. Soc.* **203,** no. 3, 1896–1908.

Visser, R., B. Smith, H. Kao, A. Babaie Mahani, J. Hutchinson, and J. E. McKay (2017). A comprehensive earthquake catalogue for northeastern British Columbia and western Alberta, 2014–2016, *Geol. Surv. of Canada, Open-File 8335*, doi: 10.4095/306292.

Wang, J., and T.-L. Teng (1995). Artificial neural network-based seismic detector, *Bull. Seismol. Soc. Am.* **85,** no. 1, 308–319.

Wang, N., and D.-Y. Yeung (2013). Learning a deep compact image representation for visual tracking, *Advances in Neural Information Processing Systems*, Lake Tahoe, Nevada, 5–10 December 2013, 809–817.

Wessel, P., and W. H. Smith (1998). New, improved version of Generic Mapping Tools released, *Eos Trans. AGU* **79,** no. 47, 579.

Withers, M., R. Aster, C. Young, J. Beiriger, M. Harris, S. Moore, and J. Trujillo (1998). A comparison of select trigger algorithms for automated global seismic phase and event detection, *Bull. Seismol. Soc. Am.* **88,** no. 1, 95–106.

Yang, J., K. Yu, and T. Huang (2010). Supervised translation-invariant sparse coding, *2010 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, San Francisco, California, 13–18 June 2010, 3517–3524.

Yoon, C. E., O. O'Reilly, K. J. Bergen, and G. C. Beroza (2015). Earthquake detection through computationally efficient similarity search, *Sci. Adv.* **1,** no. 11, e1501057.

Zhang, H., C. Thurber, and C. Rowe (2003). Automatic P-wave arrival detection and picking with multiscale wavelet analysis for single-component recordings, *Bull. Seismol. Soc. Am.* **93,** no. 5, 1904–1912.

Zhu, W., and G. C. Beroza (2018). PhaseNet: A deep-neural-network-based seismic arrival time picking method, available at http://arxiv.org/abs/1803.03211v1 (last accessed December 2018).

Ramin M. H. Dokht
Honn Kao
Ryan Visser
Brindley Smith
Pacific Geoscience Centre
Natural Resources Canada
Geological Survey of Canada
9860 West Saanich Road
Sidney, British Columbia
Canada V8L 4B2
ramin.mohammadhosseinidokht@canada.ca