ADAPTIVE FILTERING FOR EVENT RECOGNITION FROM NOISY SIGNAL: AN APPLICATION TO EARTHQUAKE DETECTION

Zhongping Zhang, Youzuo Lin, Zheng Zhou

Los Alamos National Laboratory
Los Alamos, NM, 87544

Tianlang Chen

University of Rochester
Rochester, NY, 14627

ABSTRACT

Seismic event classification and detection have been important research topics because of their significance and wide applications on hazard assessment and global security. In the real world, seismic data acquisition are always impacted by unavoidable nature factors, which will introduce low-frequency noise to the seismic events of interests. Preprocessing of seismic signal using denoising techniques can be critical to the detection of the seismic events. In our work, we develop an end-to-end framework which can automatically learn the hyper-parameter in the denoising algorithm so that we do not need to manually set the hyper-parameter. Specifically, our network structure consists of two modules, an adaptive filtering module for signal denoising, and a classification module for signal classification. We further develop two mechanisms of the adaptive filtering module, namely, sample-specific mechanism and dataset-specific mechanism. We validate the performance of our detection method using a series of field seismic datasets. The classification results show that our framework can not only remove signal noise effectively but also improve the classification accuracy.

Index Terms— adaptive filtering, neural network, signal denoising, seismic event recognition

1. INTRODUCTION

Monitoring and detection of seismic event play an important role to prevent nature hazards and detection of potential threats. Methods based on abrupt change detection [1, 2] are firstly proposed to detect seismic events of interests. These approaches produce low detection accuracy when the seismic events are buried in modest noise levels. To alleviate this issue, template matching and its modified techniques [3, 4, 5, 6, 7] are developed to provide a more robust and reliable detection. Although the robustness is improved, these methods are computationally expensive, and thus not practical for detecting events in long time-series signal. Recently, detection methods based on convolutional neural network (CNN) have been developed and achieved promising results [8, 9, 10]. Since CNN is sensitive to input data, researchers usually apply a denoising algorithm to reduce the influence of noise.

There exists various denoising algorithms including preprocessing methods [11, 12], robust feature algorithms [13], back-end approaches [14, 15], and so forth. More recently, several works [16, 17, 18] have shown that the deep neural network (DNN) can be a promising tool to denoise time-series signals. These works re-formulate signal denoising as a regression problem and their DNN frameworks are applied to recover clean signals from noisy data. The DNN-based methods often have a significant advantage in terms of time efficiency and regression accuracy. However, their major drawback is the heavy dependence on clean-noisy data pairs. When it comes to seismic signal, it is unrealistic and expensive to obtain large-scale clean-noisy signal pairs. Considering the components of seismic noise are mostly low-frequency nature noise [19], researchers in geophysics community usually apply a Fourier frequency high-pass filter to remove noise from seismic signal. The cutoff threshold of frequency high-pass filter is always based on previous experience and people have to try several times to get an appropriate value. To solve this problem, we develop an end-to-end framework which can automatically generate cutoff threshold and accomplish seismic classification based only on noisy data and classification labels.

In particular, our method consists of two blocks: an adaptive filtering block and a classification block as shown in Fig. 1. The parameters of adaptive filtering block and classification block are all trained by the classification loss. We design a novel structure which incorporates the frequency filters with learnable cutoff thresholds. Comparing with those existing DNN-based signal detection approaches, there are three major differences of our method. Firstly, our detection method incorporates the adaptive filtering block together with the classification block to improve the overall detection accuracy. Since the nature noise in the seismic dataset mostly resides in the low-frequency bands [19], we build a smooth function to approximate the high-pass frequency filter. Secondly, instead of using clean-noisy data pairs to train DNN algorithms, our denoising target only relies on noisy data and classification labels. In other words, the adaptive filtering block can be seen as a weakly supervised mechanism and the parameters are trained based on classification loss. Last but not least, unlike conventional CNN framework whose parameters are static, the cutoff thresholds of our high-pass frequency filter can vary with the input signal (sample-specific). Therefore, the sample-specific adaptive filtering block is a dynamic neural network block.

Our proposed approach is evaluated on two field seismic datasets collected at Oklahoma, USA and Decatur, USA, respectively. A series of experiments are performed to evaluate the effectiveness of our model on denoising and classification. In summary, the main contributions of our work are:

- We design a novel signal denoising structure named adaptive filtering. The adaptive filtering block can automatically generate cutoff thresholds for frequency filters according to the input samples.

- We develop two adaptive filtering mechanism: dataset-specific and sample-specific. Dataset-specific mechanism means that the neural network can generate cutoff thresholds suitable for the whole dataset. Sample-specific mechanism means that the neural network is able to output various thresholds for every particular sample.

- We develop an end-to-end framework which incorporates both signal denoising and seismic event classification. The experiment results show that our framework can not only achieve satisfying denoising results but also boost the classification accuracy.

2. METHOD

2.1. Basic Scheme of Adaptive Filtering

Nature seismic noise generated by local-scale meteorological conditions (wind, rain), large-scale meteorological perturbations (monsoon), volcanic tremor, exists in the low-frequency bands of seismic datasets [19]. Therefore, we mainly discuss our adaptive filtering as a high-pass frequency filter.

Conventionally, seismologists usually employ a pre-processing step to denoise the seismic signal by applying a high-pass frequency filter. The selection of the cutoff threshold for high-pass filter can be subjective. In addition, it can be challenging to find the best threshold to denoise signal and maintain as much useful information as possible at the meanwhile. We, on the other hand, design the adaptive filtering module to guide the neural network to automatically select the best cutoff threshold value. We denote the input signal as $\{x_n\} := x_0, x_1, ..., x_{N-1}$ and compute the frequency representation $\{X_k\} := X_0, X_1, ..., X_{N-1}$ by

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-\frac{2\pi i}{N} kn}. \tag{1}$$

Denoting the discrete Fourier transform (DFT) as $\mathbb{F}$, Eq. (1) can be written as

$$X = \mathbb{F}(x). \tag{2}$$

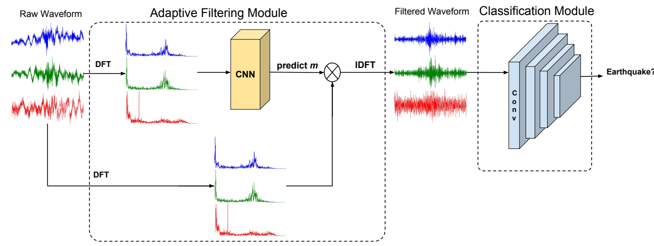

Fig. 1. The overall pipeline of our denoising and classification framework. Adaptive filtering block is shown on the dotted box to the left, and the classification network is shown on the dotted box to the right.

Suppose the cutoff threshold of an ideal high-pass filter is m, the new representation of X is

$$\hat{X}_k = \begin{cases} 0, & k \leq m, \\ X_k, & k > m. \end{cases} \tag{3}$$

Given $\hat{X}$, the filtered signal is calculated by

$$\hat{x} = \mathbb{F}^{-1}(\hat{X}), \tag{4}$$

where $\mathbb{F}^{-1}$ represents the operator of inverse discrete Fourier transform (IDFT) and $\hat{x}$ means the filtered signal in time domain.

If we can directly pass m as a parameter in the neural network, the high-pass filter should strictly filter out the low-frequency noise. However, Eq. (3) is not a differentiable function. That means the gradient cannot be back propagated through Eq. (3) and thus the parameter m is not trainable under this circumstance. To address this problem, we construct a smooth function to approximate Eq. (3)

$$\hat{X}_k = (1 + e^{-(\lambda k - m)})^{-1} X_k, \tag{5}$$

where λ is a hyper-parameter to control the slope of this function. Under this circumstance, m becomes a trainable parameter and the neural network can automatically find out the most suitable m for high-pass frequency filter.

2.2. Dataset-specific versus Sample-specific Mechanism

If we directly apply the basic scheme of adaptive filtering in the training process, threshold m is considered as a special parameter in our framework and therefore becomes uniform for the whole dataset. We name this training strategy as "dataset-specific mechanism". Dataset-specific mechanism works well when the noise distributes in a solid frequency range. However, when the signal noise distributes in various frequency ranges, dataset-specific mechanism may not be appropriate. To handle this problem, we develop a "sample-specific mechanism".

Sample-specific mechanism means our framework is able to generate a particular cutoff threshold for each input sample. In this scenario, we consider parameter m as an output of the input signal to realize the sample-specific mechanism. Since the cutoff threshold m is influenced by the noise component in the frequency domain, we take the frequency representation $\{X_k\}$ as input for m generation. $\{X_k\}$ is firstly fed into a convolutional block consists of a convolutional layer, a BatchNormalization layer, and a ReLU layer. The global max pooling layer and fully connected layer are then applied to generate the expected cutoff threshold m. The operation can be expressed in the following steps:

$$\begin{aligned} h &= \{X_k\} \circledast W_c, \\ m &= \sigma(W_f \cdot h + b_f), \end{aligned} \tag{6}$$

where W_c, W_f, and b_f are the weight parameters. The operator of $\circledast$ is the convolution operation. We omit the BatchNormalization layer and ReLU layer to avoid redundancy.

2.3. Training Strategy

Following the adaptive filtering block, we deploy a CNN to predict the final classification results. The parameters of both adaptive filtering block and CNN module are trained based on cross entropy of classification results. The complete framework with the proposed adaptive filtering is shown in Fig. 1. We train the adaptive filtering and classification module together but with different learning rates.

3. EVALUATION

We evaluate the effectiveness of our new event detection method based on two seismic datasets, namely, "Oklahoma dataset" and "Decatur dataset". In this section, datasets and experiment settings are first described. We then perform a series of experiments and provide detection results and explanations accordingly.

3.1. Datasets & Experiment Settings

Our first test dataset, Oklahoma dataset, is collected by United States Geological Survey (USGS). The dataset recorded seismic events from 2010-03-13 18 p.m. to 2012-02-20 0 a.m. Since the labels only involve major seismic events and we would like to recognize both major and small earthquakes, we manually labeled 2,654 seismic events started from 2010-03-13 6 p.m. to 2011-11-23 3 p.m. to construct our classification dataset. We take the 2,654 seismic events as positive samples and randomly picked 7,962 negative samples from non-event signals. The ratio between positive and negative samples is 1:3. We randomly select 70% of the samples as the training set and 20% of the samples as the testing set. The other samples are set as our validation set to adjust hyper-parameters. Our second test dataset, Decatur dataset, is acquired during the period of time from 2013-11-15 2 a.m. to 2014-09-16 3 a.m. The dataset is also provided by USGS. In this dataset, we mainly focus on major seismic events. Therefore, we use the default 88 positive samples. Correspondingly, we picked 264 negative samples from non-event signals. Since the positive samples are too small and can easily cause overfitting for a deep learning framework, data augmentation is used here to expand the positive samples to 264. Similarly, we use 70% of data for training, 10% for validating, and 20% for testing.

Table 1. Comparison of precision, recall, F-score, and accuracy on Oklahoma dataset

	Precision	Recall	F-score	Accuracy
Single Channel (BHE)				
CNN	73.77	65.16	69.20	87.79
CNN+Wavelet	60.94	71.37	65.74	84.34
CNN+Manually	71.72	77.40	74.45	88.82
CNN+Adap	72.61	81.36	76.73	89.62
CNN+Adap+SP	**74.45**	**82.86**	**78.43**	**90.41**
Multi-channel (BHE/BHN/BHZ)				
CNN	84.62	72.50	78.09	91.44
CNN+Wavelet	91.84	82.67	87.02	94.81
CNN+Manually	**94.88**	87.38	90.98	96.35
CNN+Adap	94.67	87.01	90.68	96.23
CNN+Adap+SP	94.32	**90.77**	**92.51**	**96.91**

Table 2. Comparison of precision, recall, F-score, and accuracy on Decatur dataset

	Precision	Recall	F-score	Accuracy
CNN	76.78	79.63	78.18	77.14
CNN+Manually	94.11	88.89	91.43	91.43
CNN+Adap	94.23	90.74	92.45	92.38
CNN+Adap+SP	**94.34**	**92.59**	**93.46**	**93.33**

The magnitude of seismic signal distributes in a large range. For instance, the magnitude of a major earthquake may span from -20,000 to 20,000 while the magnitude of a small earthquake only span from -2,000 to 2,000. The large distribution makes the neural network hard to converge. As a result, we normalize every single sample to range (-1 $\sim$ 1) to address the problem. The batch size is set to 256. We set the learning rate of Adam optimizer to 10^{-3} and 10^{-4} respectively for adaptive filtering and classification network.

3.2. Oklahoma Dataset

Table 1 shows the quantitative results of our classification experiments on Oklahoma dataset. In Table 1, "Adap" is the abbreviation for adaptive filtering and "SP" is the abbreviation for sample-specific mechanism. "Manually" means we set the cutoff thresholds of Fourier transform manually. Specifically,

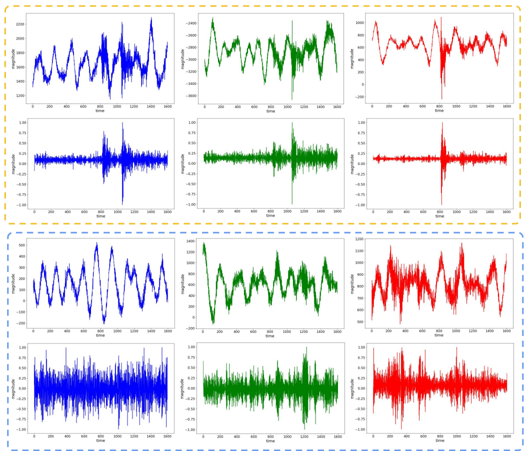

Fig. 2. Examples of our denoising network. Signals in the orange dotted box are positive samples (seismic events). Signals in the blue dotted box are negative samples. For each box, the first row illustrates the raw signals and the second row shows the processed signals generated by adaptive filtering block. From left to right: BHE, BHN, and BHZ component.

we test the performance of cutoff thresholds every 50 interval and select the threshold which can achieve the best classification results. "Wavelet" means filtering the seismic signal with Daubechies' least-asymmetric wavelet and we manually select 6 as the vanishing moment [20]. To make fair comparisons, we apply a four-layer CNN as our classification networks for all these models. We first conduct our experiments using single channel signals. In our baselines, Fourier transform outperforms Wavelet transform. The reason is that the major components of natural seismic noise exist in low-frequency band. The seismic events mostly spans in high-frequency band. Therefore, Fourier transform can generate high efficiency and satisfying results. On the other hand, because of insufficient prior knowledge, as well as more parameters needed to be manually picked, the results of Wavelet cannot outperform Fourier transform in our experiments. As shown in the table, CNN+Adap can achieve better performance than the other baselines. Based on adaptive filtering, sample-specific mechanism can further improve the classification results. Besides, we also incorporate multi-channel signals to evaluate the effectiveness of our detection method. Compared with single channel signals, the prediction results of multi-channel signals are more accurate. Our adaptive filtering block and sample-specific mechanism are still able to observably boost the classification measures.

To further substantiate the effectiveness of our model, we visualize some denoising results in Fig. 2. We select a positive sample and a negative sample respectively to demonstrate the enhancement effect. As shown in the figure, it is hard to observe dramatic differences between positive and negative

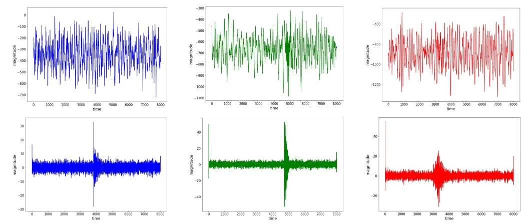

Fig. 3. Examples of denoised signals on Decatur dataset.

samples based on raw signals. However, we can observe an obvious change from the signals generated by our adaptive filtering block. Positive samples, namely the seismic events, always have a sharp pulse while negative samples are relatively plain signals. This phenomenon demonstrates explainable improvements in the classification results.

3.3. Decatur Dataset

We also evaluate our model on Decatur dataset. Since the application of data augmentation, the ratio of positive and negative sample is 1:1. Table 2 shows the performance of our models on Decatur dataset. Consistent with Oklahoma dataset results, we can observe that CNN+Adap still remarkably outperforms the CNN baselines and sample-specific mechanism further boost the classification results. Figure 3 illustrates denoised signals generated by our model. Even though the noise pattern of Decatur dataset is not exactly similar to Oklahoma dataset, adaptive filtering works well on both conditions.

4. CONCLUSION

In this paper, we present a novel seismic detection method using adaptive filtering network to incorporate both signal denoising and recognition modules. In adaptive filtering module, we further develop two mechanisms, namely, dataset-specific mechanism and sample-specific mechanism, respectively. Compared with dataset-specific mechanism, sample-specific mechanism is more flexible and can handle with noise according to particular input sample. Our proposed methods can significantly improve the classification performance without using extra ground truth information. Extensive experiments using field seismic data sets are conducted to validate the effectiveness of our approaches. It is worth to mention that we design our detection method based on high-pass filter for seismic event applications. Similar idea can be applied to low-pass or even band-pass filter for different applications.

5. ACKNOWLEDGE

We thank the support of U.S. DOE Fossil Energy office and the Center for Space and Earth Science (CSES) at Los Alamos National Laboratory.

6. REFERENCES

[1] Mitchell Withers, Richard Aster, Christopher Young, Judy Beiriger, Mark Harris, Susan Moore, and Julian Trujillo, "A comparison of select trigger algorithms for automated global seismic phase and event detection," *Bulletin of the Seismological Society of America*, vol. 88, no. 1, pp. 95–106, 1998.

[2] Michèle Basseville, Igor V Nikiforov, et al., *Detection of abrupt changes: theory and application*, vol. 104, Prentice Hall Englewood Cliffs, 1993.

[3] Steven J Gibbons and Frode Ringdal, "The detection of low magnitude seismic events using array-based waveform correlation," *Geophysical Journal International*, vol. 165, no. 1, pp. 149–166, 2006.

[4] Robert J Skoumal, Michael R Brudzinski, Brian S Currie, and Jonathan Levy, "Optimizing multi-station earthquake template matching through re-examination of the youngstown, ohio, sequence," *Earth and Planetary Science Letters*, vol. 405, pp. 274–280, 2014.

[5] David R Shelly, "A 15 year catalog of more than 1 million low-frequency earthquakes: Tracking tremor and slip along the deep san andreas fault," *Journal of Geophysical Research: Solid Earth*, vol. 122, no. 5, pp. 3739–3753, 2017.

[6] Sarah A Barrett and Gregory C Beroza, "An empirical approach to subspace detection," *Seismological Research Letters*, vol. 85, no. 3, pp. 594–600, 2014.

[7] Harley M Benz, Nicole D McMahon, Richard C Aster, Daniel E McNamara, and David B Harris, "Hundreds of earthquakes per day: The 2014 guthrie, oklahoma, earthquake sequence," *Seismological Research Letters*, vol. 86, no. 5, pp. 1318–1325, 2015.

[8] Y. Wu, Y. Lin, Z. Zhou, David Bolton, and Paul Johnson, "DeepDetect: A cascaded region-based densely connected network for seismic event detection," *IEEE Transactions on Geoscience and Remote Sensing*, 2018 (accepted).

[9] Thibaut Perol, Michaël Gharbi, and Marine Denolle, "Convolutional neural network for earthquake detection and location," *Science Advances*, vol. 4, no. 2, pp. e1700578, 2018.

[10] Dawei Mu, Pietro Cicotti, Yifeng Cui, En-Jui Lee, Florin-James Langer, Junyi Qiu, Haemin Jenny Lee, and Cody Morrin, "Deep learning for seismic template recognition," in *Proceedings of the Practice and Experience on Advanced Research Computing*. ACM, 2018, p. 58.

[11] Felix Weninger, Martin Wöllmer, Jürgen Geiger, Björn Schuller, Jort Gemmeke, Antti Hurmalainen, Tuomas Virtanen, and Gerhard Rigoll, "Non-negative matrix factorization for highly noise-robust asr: To enhance or to recognize?," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, 2012.

[12] Zbyn ek Koldovskỳ, Jirí Málek, Jan Nouza, and Miroslav Balík, "Chime data separation based on target signal cancellation and noise masking," in *Machine Listening in Multisource Environments*, 2011.

[13] Oriol Vinyals and Suman V Ravuri, "Comparing multilayer perceptron to deep belief network tandem features for robust asr," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 4596–4599.

[14] Jun Du and Qiang Huo, "A feature compensation approach using high-order vector taylor series approximation of an explicit distortion model for noisy speech recognition," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 8, pp. 2285–2293, 2011.

[15] Ramon Fernandez Astudillo, "Integration of short-time fourier domain speech enhancement and observation uncertainty techniques for robust automatic speech recognition," 2010.

[16] Qifeng Chen, Jia Xu, and Vladlen Koltun, "Fast image processing with fully-convolutional networks," in *IEEE International Conference on Computer Vision*, 2017, vol. 9, pp. 2516–2525.

[17] Tal Remez, Or Litany, Raja Giryes, and Alex M Bronstein, "Deep class-aware image denoising," in *Sampling Theory and Applications (SampTA), 2017 International Conference on*. IEEE, 2017, pp. 138–142.

[18] Xue Feng, Yaodong Zhang, and James Glass, "Speech feature denoising and dereverberation via deep autoencoders for noisy reverberant speech recognition," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 1759–1763.

[19] Guanghui Li, Yue Li, and Baojun Yang, "Seismic exploration random noise on land: Modeling and application to noise suppression," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4668–4681, 2017.

[20] Jérôme Lebrun and Martin Vetterli, "Balanced multiwavelets theory and design," *IEEE Transactions on Signal Processing*, vol. 46, no. 4, pp. 1119–1125, 1998.