This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Attention-Based Convolutional Neural Network for Earthquake Event Classification

Bonhwa Ku, Gwantae Kim, *Graduate Student Member, IEEE*, Jae-Kwang Ahn, Jimin Lee, and Hanseok Ko, *Senior Member, IEEE*

***Abstract*— This letter presents a deep convolutional neural network (CNN) with attention module that improves the performance of the classification of various earthquake events. Addressing all possible earthquake events, including not only microearthquakes and artificial-earthquakes but also large-earthquakes, requires both suitable feature expression and a classifier that can effectively discriminate seismic waveforms under adverse conditions. To robustly classify earthquake events, a deep CNN with an attention module was proposed in raw seismic waveforms. Representative experimental results show that the proposed method provides an effective structure for earthquake events classification and, with the Korean peninsula earthquake database from 2016 to 2018, outperforms previous state-of-the-art methods.**

***Index Terms*— Attention module, convolutional neural network (CNN), deep learning, earthquake classification, raw seismic waveform.**

I. Introduction

TIME series analysis is used in various fields such as speech recognition, financial analysis, seismic analysis, and bio-signal analysis. Typical approaches to time series analysis consist of feature extraction and machine learning. In traditional feature-based time series analysis, feature extraction and parameters are set according to expert empirical and experimental factors. Recently, approach to time series analysis has been shifting from traditional methods to efficient feature extraction and machine learning using deep learning. In [1], a multiscale convolutional neural network (CNN) was proposed to solve time series classification problems at varied scales. Multiscale CNN has a structure that integrates feature extraction and classification within a single framework and automatically extracts various scale and frequency features. In [2], a CNN model with a small convolution filter based on raw waveform sample data has been proposed for the music genre classification problem to overcome the problems in the frame level spectrogram model. In [3], a wavelet domain multiresolution CNN model based on ECG signals was proposed for biometric authentication in the medical engineering field. Increasingly, for diversity of data representation, randomly selected samples are transformed into wavelet space and represented in multiresolution time-frequency space before then applying multiresolution 1-D CNN.

Short time average over long time average (STA/LTA) was a typical method of detecting earthquakes through the ratio between the STA and LTA in earthquake classification or detection. STA/LTA is effective for large earthquake detection but unsuitable in low signal to noise ratio environments [4]. The autocorrelation-based waveform similarity method detected earthquakes that originate from a single region [5]. Although this autocorrelation-based approach is the most effective method, it is computationally expensive and impractical for long time series data. The template matching method, which was proposed to improve the computational overload problem of the autocorrelation-based method, detects earthquakes through correlation analysis between a template set and the input earthquake waveform [6]. For this method, detection performance is predominantly affected by the number of templates, and approaches to reducing the number of templates through principal component analysis have been proposed. Recently, the fingerprint and similarity thresholding (FAST) method, which uses the binary fingerprint feature extraction and a similarity search through locality-sensitive hashing, was proposed to reduce the computation intensity of template matching [7]. Also, various methods of classifying earthquake events using deep learning have been researched [8]–[11]. In particular, Perol proposed a CNN model using three-channel raw seismic waveform data similar to the deep learning model for the time series classification problem [8]. Its architecture consists of eight convolutional layers with convolution filter and nonlinear transform rectified linear unit (ReLU) and a fully connected (FC) layer. Mousavi's model where 30-s raw seismic waveforms were transformed into the spectrograms formulated an earthquake detection problem as sequence-to-sequence learning in spectrograms of three-channel seismograms [9]. Zhu and Beroza [10] proposed a modified U-net with an encoder–decoder structure in order to pick arrival-time of P-wave and S-wave from the input of 30-s seismic wave.

In this letter, we propose an attention-based CNN structure that enables the classification of varied earthquake events. First, we propose an attention-based CNN structure to improve performance when classifying varied earthquake events. Attention mechanisms can make feature map more discriminative so that the model can focus more on the important features. In this letter, we proposed an improved attention module based on modeling interdependencies between the channels of extracted feature map. Second, we propose a CNN structure in which batch normalization and dropout techniques are suitably arranged for earthquake event classification. Batch normalization prevents the input distribution from differing between

Manuscript received October 4, 2019; revised March 29, 2020 and May 12, 2020; accepted July 31, 2020. This work was supported by the Meteorological/Earthquake See-At Technology Development Research under Grant KMI2018-09610. *(Corresponding author: Hanseok Ko.)*

Bonhwa Ku, Gwantae Kim, and Hanseok Ko are with the School of Electrical Engineering, Korea University, Seoul 02841, South Korea (e-mail: bhku@ispl.korea.ac.kr; gtkim@ispl.korea.ac.kr; hsko@korea.ac.kr).

Jae-Kwang Ahn and Jimin Lee are with Korea Meteorological Administration, Seoul 07062, South Korea (e-mail: propjk@korea.kr; jiminlee@korea.kr).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2020.3014418

1545-598X © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: BEIJING UNIVERSITY OF TECHNOLOGY. Downloaded on August 21,2020 at 01:32:37 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

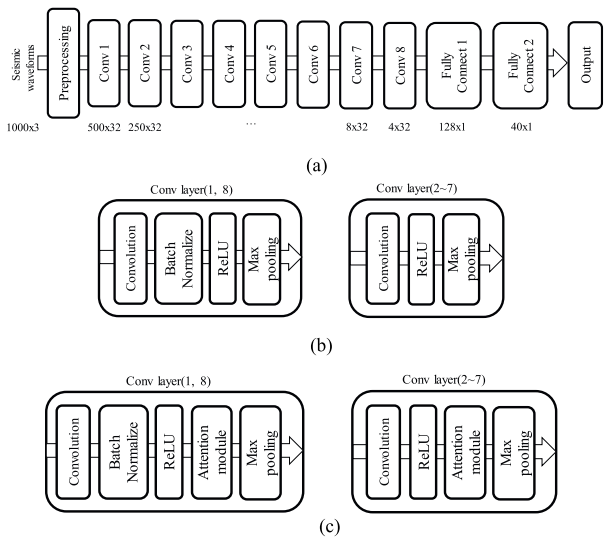

Fig. 1. Proposed CNN structure for earthquake classification. (a) Proposed CNN structure. (b) Convolution layer without attention module. (c) Convolution layer with attention module.

the layers and dropout regularization suppresses overfitting, potentially enhancing the stability and performance of the proposed model. We evaluated its performance for the Korean peninsula earthquakes from 2016 to 2018 and the experimental results indicated that compared to conventional algorithms the proposed algorithm achieves superior performance.

II. Attention-Based Convolutional Neural Net

The proposed CNN structure for earthquake classification consists of three stages: the preprocessing, feature extraction, and classification stages. As shown in Fig. 1, the feature extraction stage consists of eight convolution layers with and without the attention module, and the classification stage consists of two FC layers and a softmax function. The proposed method is trained on a data set of labeled three-channel raw seismic waveforms consisting of 1000 samples with a 10-s window recorded at 100 Hz.

A. Preprocessing

Since seismic waveforms from different stations have different biases, we perform data preprocessing according to (1). The preprocessing has the effect of geometrically moving the center of data to the origin

$$y_m = x_m - \frac{1}{M}\sum_{i=1}^{M} x_i, \quad m = 1, \ldots, M \tag{1}$$

where x_m is the seismic waveform data, y_m is transformed seismic waveform data, and M is the total number of samples in the occurrence event.

B. CNN Without Attention Module

In all convolution layers, convolution, ReLU, and max-pooling were performed, and feature maps were produced as the output. In the convolution, we convolve the feature map of previous layer with filter weights as shown in (2), and then perform the nonlinear transform ReLU as shown in (3) and the max-pooling to reduce the size of the feature map

$$\boldsymbol{F}^{l}_{c,t} = R\left(\sum_{c'}^{C}\sum_{t'}^{3} \boldsymbol{F}^{l-1}_{c',t+t'-1} \cdot \boldsymbol{W}^{l}_{c,c',t'}\right), \quad l = 1, \ldots, 8 \tag{2}$$

$$R(\boldsymbol{F}) = \max(0, \boldsymbol{F}) \tag{3}$$

where $\boldsymbol{F^l}$ represents the feature map of layer l, $\boldsymbol{F^0}$ represents the input waveform, $\boldsymbol{W^l}$ is the filter weight for layer l with dimensions $C_l \times C_{l-1} \times 3$, and C_l is the number of channels of layer l. We used 32 channels for $C_l, l = 1, \ldots, 8$, whereas the channel of the input waveform, C_0, had three channels. Max-pooling was down-sampled by the max operation on a mask with the feature map split into two strides.

While convolution layers 1 and 8 performed the same process as the convolution layers described above, they also achieve batch normalization between Convolution and ReLU as shown in Fig. 1. The batch normalization [12] was devised to prevent internal covariance shifts that the input distribution and the output distribution are different in each layer. It has been found that batch normalization is effective for gradient explosion and gradient vanishing problems in the deep learning. Batch normalization initially standardized the feature maps of each layer using mean and standard deviation during training on mini-batch units, and then transformed the feature maps by adding a scale factor and a shift factor to the standardized values. Since the same earthquake event may be recorded differently at each station, applying a batch normalization can improve classification performance due to its normalization effect on the raw seismic data input to the network. While batch normalization is frequently used to convolution layers for spectrogram input [9], it is not a common practice for earthquake classification task when using raw seismic waveforms [8], [10]. To address this, we present a CNN structure for earthquake classification which applies batch normalization to the first and last layers. After passing through the last convolution layer, the feature map created was transformed into a feature vector by the flattening process. The feature vector was input to two FC layers and then earthquake classification was performed using a softmax function. To prevent overfitting, a dropout regularization [13] that activates each neuron randomly depending on the probability value was applied to the first FC layer.

C. CNN Structure With Attention Module

The CNN structure with the attention module is similar to that described above, but with an attention module between ReLU and max pooling as shown in Fig. 1. Attention can be interpreted as a means to enhance features by extracting the most informative components of a signal. After calculating the attention value based on importance or saliency in the feature map extracted from the convolution layer, we can be obtain refined feature map weighting by the attention value. The attention mechanism can be largely divided into what to focus and where to focus. In this letter, we proposed an improved attention module based on squeeze and excitation networks (SENET) [14] that effectively represents what should be focused on. SENET improves the representation capability of the network by modeling interdependencies between the channels of the feature map, as shown in Fig. 2. SENET consisted of two operations: a squeeze operation and an excitation operation. The former compressed the feature map while preserving their global information and the latter adaptively recalibrated them according to the importance of each channel they contain. The squeeze operation used global average pooling (GAP) taking an average per channel of the feature map and the excitation operation employed two FC layers and a simple gating mechanism with a sigmoid activation. The first FC was a dimensionality-reduction layer with reduction ratio r, and the second FC was a dimensionality-increasing

Authorized licensed use limited to: BEIJING UNIVERSITY OF TECHNOLOGY. Downloaded on August 21,2020 at 01:32:37 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

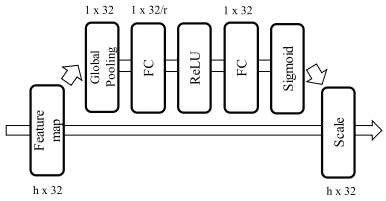

Fig. 2. Structure of SENET.

layer restoring the channel dimension of the feature map. The gating mechanism with a sigmoid activation regulated the relative importance of the channels. Function "scale" referred to channel-wise multiplication between the gating value and the feature map.

In order to improve SENET, not only excitation operation, but also squeeze operation is important. In the original SENET, the GAP in which all feature values in the channels are assigned an identical weight is used for the squeeze operation, as shown in (4). However, a global weighted average pooling (GWAP) that assigns weights according to the relative importance of each feature value may preserve the global information of the feature map better than the GAP. In order to calculate the attention value effectively, we proposed two squeeze operations using the GWAP as shown in the following equations:

$$s_{\mathbf{GAP}}(\boldsymbol{c}) = \frac{\mathbf{1}}{\boldsymbol{h}} \sum_{i=1}^{h} \boldsymbol{F_c}(i) \tag{4}$$

$$s_{\mathbf{GWAP_1}}(\boldsymbol{c}) = \sum_{i=1}^{h} \alpha_{\boldsymbol{c}}(i) \boldsymbol{F_c}(i) \tag{5}$$

$$s_{\mathbf{GWAP_2}}(\boldsymbol{c}) = \mathbf{1}/\boldsymbol{h} \sum_{i=1}^{h} \beta_{\boldsymbol{c}}(i) \boldsymbol{G}_{\mathbf{max}_c} \tag{6}$$

where $\alpha_{\boldsymbol{c}}(i) = (\boldsymbol{F_c}(i)/\sum_{i=1}^{h} \boldsymbol{F_c}(i))$, $\beta_{\boldsymbol{c}}(i) = (\boldsymbol{F_c}(i) - \boldsymbol{G}_{\mathbf{min}_c}/\boldsymbol{G}_{\mathbf{max}_c} - \boldsymbol{G}_{\mathbf{min}_c})$, $\boldsymbol{F_c}$ is the feature in the channel c, h is feature length in the channel c, $\boldsymbol{G}_{\mathbf{min}_c}$ and $\boldsymbol{G}_{\mathbf{max}_c}$ are min value and max value in the channel c of the feature map, respectively. In GWAP_1 case, we assigned weights that represent relative importance as normalized feature value in the channels. The GWAP_1 has a value between GAP and global maximum pooling (GMP). The GWAP_2 calculated weights based on a contrast stretching method [15] that was employed in image enhancement in order to increase sharpness of the feature map. Contrast stretching is a method of improving the visibility of an image by spreading the histogram of a low contrast image distribution to achieve a wider range of contrast values. Applying this concept to the feature map, we can obtain the squeeze operation that embed more distinguishable channel-wise statistics.

III. Experiments

In this section, we performed experiments to investigate the effectiveness of proposed CNN architectures and attention module. The simulation was performed on an NVIDIA Geforce GTX 1080Ti GPU and Tensorflow, and the network training was performed over 300 epochs. For model training, the adaptive moment estimation (ADAM) optimization was used.

A. Data Set

Raw seismic data were collected from January 2016 to July 2018 at about 200 stations in South Korea. Raw seismic data from each station are recorded at 100 Hz on three channels: Z oriented vertically, N oriented north–south, and E oriented west–east. As shown in Fig. 3, the database sets are constructed based on 24-h continuous raw seismic data and the earthquake event catalog file provided by National Earthquake Comprehensive Information System (NECIS) [16]. The top of Fig. 3 shows a 24-h continuous data waveform from a station, and the bottom an example of an earthquake event cropped to a 13-s window including a 3 s margin before P-wave arrival. Using only velocity-meter sensors of HH and EL types, we extract 13-s windows showing three event types: large earthquakes (equal to or more than magnitude 2.0), microearthquakes (less than magnitude 2.0), and artificial earthquakes caused by underground nuclear tests or blasting sites. In the case of noise event, we obtained the data set containing only seismic noise by two approaches: extracting random 10-s windows of seismic noise from dates not including the cataloged events or extracting random 10-s windows of seismic noise except for the cataloged events.

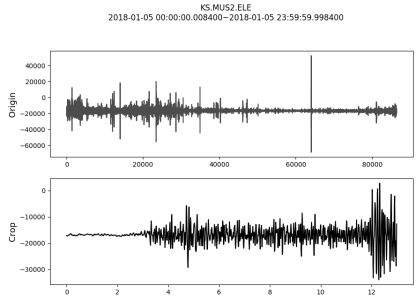

Fig. 3. Continuous waveform and extracted event waveform.

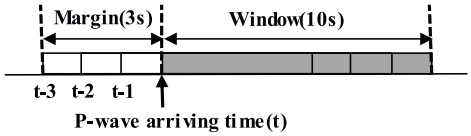

Fig. 4. Data augmentation using window slicing.

Generally, data augmentation is used to increase the volume of a data set, which prevent overfitting and improve generalization of model. Using the window slicing method [1] we increase the data set by sliding a 10-s window at 1-s interval from 0 to 3 s on the extracted data set, as shown in Fig. 4. By using the window slicing method, we increased the data sets related to earthquake events by four times, while data augmentation is not applied to the data set related to noise event. To evaluate the performance of various types of earthquake event classification, the proposed model is trained and tested using the event data sets summarized in Table II. We split the data set for two years (2016–2017) in Table I into 80% and 20% as training set and validation set, respectively and use the data set for 2018 year in Table I as test set. In Table II, the classification targets are set to "earthquake versus noise," "large earthquake versus noise," "microearthquake versus noise," "artificial earthquake versus noise," "natural earthquake versus artificial earthquake versus noise." "Earthquake" consisted of large-earthquake, microearthquake and artificial earthquake and "natural earthquake" consisted of large-earthquake and microearthquake. In Table II, the numbers in boldface indicate the number of test sets.

Authorized licensed use limited to: BEIJING UNIVERSITY OF TECHNOLOGY. Downloaded on August 21,2020 at 01:32:37 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I

DATA SETS SHOWS THE NUMBER OF EARTHQUAKE EVENTS EXTRACTED AT EACH STATION FOR THE PERIOD FROM JANUARY 2016 TO JULY 2018 AND THE NUMBERS IN PARENTHESES INDICATE THE NUMBER OF EARTHQUAKE EVENTS ACCORDING TO THE SENSORS OF HH AND EL TYPES, RESPECTIVELY. LE, ME, AE, AND N STAND FOR LARGE EARTHQUAKE, MICROEARTHQUAKE, ARTIFICIAL EARTHQUAKE, AND NOISE EVENT, RESPECTIVELY

Events/ Years	2018	2017	2016
LE	172[109, 63]	993[627,366]	1700[890,810]
ME	483[292,191]	1237[696,541]	845[492,353]
AE	157[114,43]	236[171,65]	324[242,82]
Noise	9852[7312,2540]	12523[8329,4194]	10000[5000,5000]

TABLE II

VARIOUS TYPES OF EARTHQUAKE EVENT CLASSIFICATION. E/N, LE/N, ME/N, AE/N, AND NE/AE/N STAND FOR EARTHQUAKE VERSUS NOISE, LARGE EARTHQUAKE VERSUS NOISE, MICROEARTHQUAKE VERSUS NOISE, ARTIFICIAL EARTHQUAKE VERSUS NOISE, NATURAL EARTHQUAKE VERSUS ARTIFICIAL EARTHQUAKE VERSUS NOISE, RESPECTIVELY

	(1)	(2)	(3)
E(1)/N(2)	21340/**3248**	22523/**5752**	·
LE(1)/N(2)	10772/**688**	22523/**2312**	·
ME(1)/N(2)	8328/**1932**	22523/**4068**	·
AE(1)/N(2)	2240/**628**	22523/**2372**	·
NE(1)/AE(2)/N(3)	19100/**2620**	2240/**628**	22523/**5752**

TABLE III

HYPERPARAMETERS ACCORDING TO METHODS

	Baseline method[8]	Proposed methods
Filter size/Stride/ No of channels	3/2/32	3/1/32
Pooling size/Stride	·	2/2
No of Layers	8	8
Learning rate/ L_2 Regularization parameter	$10^{-4}/10^{-3}$	$5x10^{-3}/\cdot$
Batch size/Dropout	128/·	128/0.5

B. Performance Evaluation

The classification performance was evaluated for accuracy, true positive rate, and false positive rate. They were defined as

$$\begin{aligned} \text{Acc} &= \frac{\boldsymbol{TP+TN}}{\boldsymbol{TP+TN+FP+FN}}, \\ \text{TPR} &= \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}}, \\ \text{FPR} &= \frac{\boldsymbol{FP}}{\boldsymbol{TN+FP}} \end{aligned} \tag{7}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

1) Result of CNN Structure Without Attention Module: We first assessed the classification performance of the proposed CNN structure without the attention module. The baseline method applies ConvNetQuake [8] which uses the same input data type as ours and the detailed hyperparameters are listed in Table III. We compared our method without the attention module (Modified CNN) with ConvNetQuake. Table IV summarizes the comparison results with our method and ConvNetQuake. From Table IV, we observe that our method provides better performance than the baseline method for all event classifications. The proposed method improves the accuracy by 3.2% on average, and the improvement in TPR yet decrease in FPR is particularly noticeable. The event classification of artificial earthquake was not as good as the other event classifications. This performance degradation can be attributed to insufficient data quantity and data length. The number of artificial earthquake events in the training data was significantly lower than other events.

TABLE IV

COMPARISON OF BASELINE AND MODIFIED CNN. ACC, TPR, AND FPR STAND FOR ACCURACY, TRUE POSITIVE RATE AND FALSE POSITIVE RATE, RESPECTIVELY

	Baseline method[8]			M-CNN		
	Acc	TPR	FPR	Acc	TPR	FPR
E vs N	93.2	89.2	4.5	**96.8**	93.1	1.1
LE vs N	95.8	89.0	2.1	**97.7**	93.0	0.9
ME vs N	93.9	84.5	1.6	**97.5**	93.1	0.4
AE vs N	93.7	74.5	1.2	**96.0**	83.0	0.5
NE vs AE vs N	89.1	84.6/ 45.2/ 95.9	6.5/ 1.2/ 14.2	**94.1**	91.2/ 62.9/ 98.8	3.0/ 1.0/ 8.0

2) Result of CNN Structures With Attention Module: To evaluate the influence of the proposed attention module, we compared the classification performance of three CNN structures featuring different integrated attention modules. Three CNN structures with the attention module all have the same structure except for squeeze operation. In the attention module, the squeeze operations used GAP, $GWAP_1$, and $GWAP_2$ and the reduction ratio r was set to 4. Table V shows the classification performance of the attention-based CNN methods according to squeeze operation. First, we compare the results in Table V with those in Table IV to see the effect of the attention module. All of the attention-based CNN methods showed better performance results than the baseline method and the M-CNN. Especially, the accuracy of the attention-based CNN using $GWAP_2$ was superior to baseline method and M-CNN by 4.2% and 0.92% on average, respectively. In addition, Table V shows that the proposed attention module CNN (AM-CNN) using $GWAP_1$ shows similar performance to AM-CNN using GAP, but the proposed AM-CNN using $GWAP_2$ shows better performance in all experiments than those of AM-CNN using GAP. AM-CNN using $GWAP_2$ showed an improvement of 0.38% and 0.42% on average when compared to AM-CNN using GAP and $GWAP_1$, respectively. That is, the most enhanced attention method we obtained used the squeeze operation incorporating contrast stretching method.

3) Result of Attention-Based CNN Structure According to Sensor Types: Finally, we compared the classification performances of the attention-based CNN model according to data sets by sensor types. In the experiments of Sections III-B1 and III-B2, we trained the attention-based CNN model using both HH type and EL type sensors. In this section, however, our model was trained and tested using the data sets from only one sensor type. In Table VI, the second column shows the results of training and testing on only HH type sensor data, the third column shows the results of training and testing on only EL type sensor data. The fourth and fifth columns show the results of training on both data types, yet testing on specific data types. From Table VI, we observe that the performance difference between training the model with individual sensor types and all sensor types was small. Since CNN can automatically extract an optimal feature set from given data sets, our method could manage data sets composed of different sensor types effectively.

Authorized licensed use limited to: BEIJING UNIVERSITY OF TECHNOLOGY. Downloaded on August 21,2020 at 01:32:37 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE V

PERFORMANCE COMPARISON ACCORDING TO SQUEEZE OPERATIONS. AM-CNN REPRESENTS CNN STRUCTURE WITH ATTENTION MODULE AND EACH AM-CNN USES GAP, $GWAP_1$, AND $GWAP_2$, RESPECTIVELY, AS SQUEEZE OPERATION IN ATTENTION MODULE

	AM-CNN(GAP)			AM-CNN($GWAP_1$)			AM-CNN($GWAP_2$)		
	Acc	TPR	FPR	Acc	TPR	FPR	Acc	TPR	FPR
E vs N	97.0	95.0	1.8	**97.4**	95.2	1.4	**97.4**	95.8	1.7
LE vs N	98.2	96.4	1.3	98.4	94.5	0.4	**98.8**	97.4	0.8
ME vs N	**98.2**	95.7	0.6	97.7	93.8	0.4	**98.2**	96.3	0.9
AE vs N	96.1	83.0	0.4	96.0	81.4	0.1	**96.7**	84.9	0.2
NE vs AE vs N	95.3	93.5/ 72.9/ 98.5	2.6/ 1.1/ 5.2	95.1	92.4/ 72.1/ 98.8	2.6/ 1.1/ 5.7	**95.6**	94.6/ 74.2/ 98.5	2.7/ 1.0/ 4.2

TABLE VI

PERFORMANCE COMPARISON BY SENSOR TYPES

	HH only train/test			EL only train/test			All train/HH only test			All train/EL only test		
	Acc	TPR	FPR	Acc	TPR	FPR	Acc	TPR	FPR	Acc	TPR	FPR
E vs N	97.3	94.8	1.9	97.5	95.3	1.4	97.6	94.7	1.5	97.1	96.2	2.4

IV. CONCLUSION

In this letter, we proposed a CNN structure with an attention module for earthquake event classification. We first proposed a modified CNN structure with batch normalization and dropout to handle various earthquake events and environmental conditions and then improved the representational power of the network by an SE attention module based on novel global pooling. The experimental results indicated that the proposed method with the attention module was an effective structure for the classification of various earthquake events and that it outperformed the previous state-of-the-art method. In addition, the experimental results showed that the proposed structure could flexibly learn good feature representations for different sensor types. Finally, the proposed attention-based CNN structure may be effectively applied to systems such as seismic active monitoring systems and earthquake early warning systems.

REFERENCES

[1] Z. Cui, W. Chen, and Y. Chen, "Multi-scale convolutional neural networks for time series classification," May 2016, *arXiv:1603.06995*. [Online]. Available: http://arxiv.org/abs/1603.06995

[2] J. Lee, J. Park, K. Kim, and J. Nam, "SampleCNN: End-to-end deep convolutional neural networks using very small filters for music classification," *Appl. Sci.*, vol. 8, no. 1, p. 150, Jan. 2018.

[3] Q. Zhang, D. Zhou, and X. Zeng, "HeartID: A multiresolution convolutional neural network for ECG-based biometric human identification in smart health applications," *IEEE Access*, vol. 5, pp. 11805–11816, May 2017.

[4] M. Withers *et al.*, "A comparison of select trigger algorithms for automated global seismic phase and event detection," *Bull. Seismol. Soc. Amer.*, vol. 88, no. 1, pp. 95–106, Feb. 1998.

[5] S. J. Gibbons and F. Ringdal, "The detection of low magnitude seismic events using array-based waveform correlation," *Geophys. J. Int.*, vol. 165, no. 1, pp. 149–166, Apr. 2006.

[6] R. J. Skoumal, M. R. Brudzinski, B. S. Currie, and J. Levy, "Optimizing multi-station earthquake template matching through re-examination of the Youngstown, Ohio, sequence," *Earth Planet. Sci. Lett.*, vol. 405, pp. 274–280, Nov. 2014.

[7] C. E. Yoon, O. O'Reilly, K. J. Bergen, and G. C. Beroza, "Earthquake detection through computationally efficient similarity search," *Sci. Adv.*, vol. 1, no. 11, Dec. 2015, Art. no. e1501057.

[8] T. Perol, M. Gharbi, and M. Denolle, "Convolutional neural network for earthquake detection and location," *Sci. Adv.*, vol. 4, no. 2, Feb. 2018, Art. no. e1700578.

[9] S. M. Mousavi, W. Zhu, Y. Sheng, and G. C. Beroza, "CRED: A deep residual network of convolutional and recurrent units for earthquake signal detection," 2018, *arXiv:1810.01965*. [Online]. Available: http://arxiv.org/abs/1810.01965

[10] W. Zhu and G. C. Beroza, "PhaseNet: A deep-neural-network-based seismic arrival time picking method," 2018, *arXiv:1803.03211*. [Online]. Available: http://arxiv.org/abs/1803.03211

[11] A. Mignan and M. Broccardo, "Neural network applications in earthquake prediction (1994–2019): Meta-analytic insight on their limitations," Oct. 2019, *arXiv:1910.01178*. [Online]. Available: http://arxiv.org/abs/1910.01178

[12] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," Mar. 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167

[13] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, Jun. 2014.

[14] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," Sep. 2017, *arXiv:1709.01507*. [Online]. Available: http://arxiv.org/abs/1709.01507

[15] R. E. Woods and R. C. Gonzalez, *Digital Image Processing*, 4th ed. London, U.K.: Pearson, 2018.

[16] (Jul. 2019). *NECIS*. [Online]. Available: http://necis.kma.go.kr/

[17] K. Xu *et al.*, "Show, attend and tell: Neural image caption generation with visual attention," 2015, *arXiv:1502.03044*. [Online]. Available: http://arxiv.org/abs/1502.03044

[18] J. Park, S. Woo, J.-Y. Lee, and I. S. Kweon, "BAM: Bottleneck attention module," 2018, *arXiv:1807.06514*. [Online]. Available: http://arxiv.org/abs/1807.06514

[19] S. Dieleman and B. Schrauwen, "End-to-end learning for music audio," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 6964–6968.

[20] D. Ardila, C. Resnick, A. Roberts, and D. Eck, "Audio deepdream: Optimizing raw audio with convolutional networks," in *Proc. Int. Soc. Music Inf. Retr. Conf.*, 2016, pp. 7–11.

[21] J. Thickstun, Z. Harchaoui, and S. Kakade, "Learning features of music from scratch," Apr. 2017, *arXiv:1611.09827*. [Online]. Available: http://arxiv.org/abs/1611.09827

[22] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[23] X. Cui, V. Goel, and B. Kingsbury, "Data augmentation for deep neural network acoustic modeling," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 9, pp. 1469–1477, Sep. 2015.

[24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Apr. 2015, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[25] A. Borovykh, S. Bohte, and C. W. Oosterlee, "Conditional time series forecasting with convolutional neural networks," Sep. 2018, *arXiv:1703.04691*. [Online]. Available: http://arxiv.org/abs/1703.04691

[26] N. Hatami, Y. Gavet, and J. Debayle, "Classification of time-series images using deep convolutional neural networks," Oct. 2017, *arXiv:1710.00886*. [Online]. Available: http://arxiv.org/abs/1710.00886

[27] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: A review," Sep. 2018, *arXiv:1809.04356*. [Online]. Available: http://arxiv.org/abs/1809.04356

Authorized licensed use limited to: BEIJING UNIVERSITY OF TECHNOLOGY. Downloaded on August 21,2020 at 01:32:37 UTC from IEEE Xplore. Restrictions apply.