



Convolutional Neural Network to Detect Deep Low-Frequency Tremors from Seismic Waveform Images

Ryosuke Kaneko^{1,2}, Hiromichi Nagao^{1,2}(✉), Shin-ichi Ito^{1,2}, Kazushige Obara², and Hiroshi Tsuruoka²

¹ Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan

`ryosuke-kaneko@g.ecc.u-tokyo.ac.jp`, `nagaoh@eri.u-tokyo.ac.jp`

² Earthquake Research Institute, The University of Tokyo, Tokyo, Japan

Abstract. The installation of dense seismometer arrays in Japan approximately 20 years ago has led to the discovery of deep low-frequency tremors, which are oscillations clearly different from ordinary earthquakes. As such tremors may be related to large earthquakes, it is an important issue in seismology to investigate tremors that occurred before establishing dense seismometer arrays. We use deep learning aiming to detect evidence of tremors from past seismic data of more than 50 years ago, when seismic waveforms were printed on paper. First, we construct a convolutional neural network (CNN) based on the ResNet architecture to extract tremors from seismic waveform images. Experiments applying the CNN to synthetic images generated according to seismograph paper records show that the trained model can correctly determine the presence of tremors in the seismic waveforms. In addition, the gradient-weighted class activation mapping clearly indicates the tremor location on each image. Thus, the proposed CNN has a strong potential for detecting tremors on numerous paper records, which can enable to deepen the understanding of the relations between tremors and earthquakes.

Keywords: Deep low-frequency tremor · Convolutional neural network · ResNet · Gradient-weighted class activation mapping

1 Introduction

The surface of the Earth comprises 15 tectonic plates, with each plate individually moving several centimeters per year. Seismic phenomena often occur at the boundary where a plate subducts beneath another one in the subduction zone by releasing strains accumulated over a long time. Figure 1 shows the schematic of a subduction zone, in which an oceanic plate is subducting beneath a continental plate. For example, the Philippine Sea Plate is subducting beneath the Eurasian Plate around southwest Japan, forming a subduction zone called the

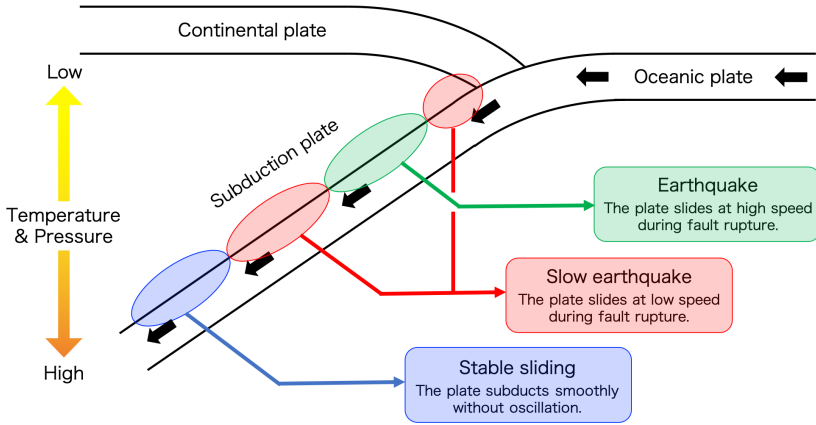


Fig. 1. Schematic of subduction zone. The ellipses represent the locations of seismic phenomena occurrence.

Nankai Trough. In a subduction zone, various seismic phenomena occur depending on the magnitude of plate friction: ordinary earthquakes (high friction), slow earthquakes (low friction), and stable sliding (no friction). The friction generally increases as the temperature decreases and the pressure increases. As both temperature and pressure increase with depth, the friction according to depth presents a complicated relation. This relation might lead to different locations of seismic phenomena occurrence, as shown in Fig. 1. Ordinary earthquakes usually occur when the strains accumulated over a long time are released instantaneously at a shallow depth, where the friction is very high. At the beginning of the 21st century, new phenomena called slow earthquakes originated from subduction zones were discovered [8]. Owing to the lower friction, slow earthquakes repeat with much shorter intervals and have smaller magnitudes than ordinary earthquakes.

Hi-net [9, 12], the seismometer network installed by NIED and operating since 1996, led to the discovery of deep low-frequency tremors, which are categorized as slow earthquakes. The Hi-net seismometers are more sensitive and densely located than conventional seismometer networks. Thus, they allow to observe weak oscillations that were previously unobservable and correlatively analyze records between neighboring seismometers. The envelope correlation method [8] successfully extracted evidence of tremors from Hi-net data, which has been recognized as the first discovery of tremors in the world. Such tremors are weak oscillations that occur in a deeper area than ordinary earthquakes. These tremors have an approximate dominant frequency of 2–8 Hz, last from several hours to several days, and have magnitudes below 1.3 according to the tremor catalog published by NIED [4, 11]. Thus, tremors cannot be perceived on the ground. According to the NIED catalog, more than 30,000 tremors have been observed in southwest Japan from January 2001 to April 2019, indicating their high frequency. Tremors have been detected in various subduction zones worldwide after

the first discovery (*e.g.*, [5, 15]), and they represent a research hotspot in seismology, as many studies have been indicated the relations between tremors and large earthquakes (*e.g.*, [10]). In fact, seismologists expect tremors to provide clues to predict large earthquakes and understand their mechanisms. Currently, only digital data from the last 20 years are available for studying tremors. Considering that megathrust earthquakes have periodically occurred in the Nankai Trough over intervals of 100–200 years, it is important to analyze tremor occurrences in southwest Japan over a longer period. Before the seismic records were available in digital format, seismometers continuously recorded waveforms with a pen on drum-rolled papers.

In this study, we aimed to detect tremors from seismograph paper records by using a convolutional neural network (CNN), a deep-learning method that has shown high performance for image recognition. A CNN can automatically tune its internal parameters by learning the characteristics of tremors from input images without requiring prior knowledge of tremors or manually adjusting the parameters. Training a CNN from scratch with real data polluted by a variety of noises may hinder the model construction and hyperparameter tuning. Thus, we conducted numerical experiments to construct a CNN and train it with synthetic images similar to the seismograph paper records. These experiments aim to obtain clues to improve the model and provide pretrained models for subsequent fine-tuning, which is a popular learning method to improve the model performance and learning efficiency.

2 Seismographs

Modern seismic research based on observational data strongly relies on digital records. In Hi-net, more than 1000 seismometers installed in Japan Islands continuously observe the velocity of the ground at a sampling rate 100 Hz, and NIED collects and publishes the corresponding digital data in real time. Multivariate time-series analyses on digital seismic data considering spatial correlations enable to increase the signal-to-noise ratio of detected phenomena or eventually uncover unknown phenomena such as slow earthquakes and tremors.

Past seismometers used over 50 years ago drew waveforms directly on paper. Considering the time interval between megathrust earthquakes in the subduction zones, paper records are a valuable source for research on slow earthquakes (*e.g.*, [2]). Figure 2 shows an example of a seismograph paper record. The daily records are drawn on a single piece of paper, in which each time series is drawn horizontally and contains approximately 2.5 min of data. Besides seismic waveforms, the records contain pulses that indicate a time stamp every second. The average image size of a paper record excluding its margins is approximately 7000×7000 pixels.

The digitization of paper records by tracing the waveforms is effective for investigating large earthquakes because such waveforms are extractable even from overlapping time series given the low frequencies and large amplitudes of earthquakes (*e.g.*, [6]). In contrast, tremors generally have smaller amplitudes

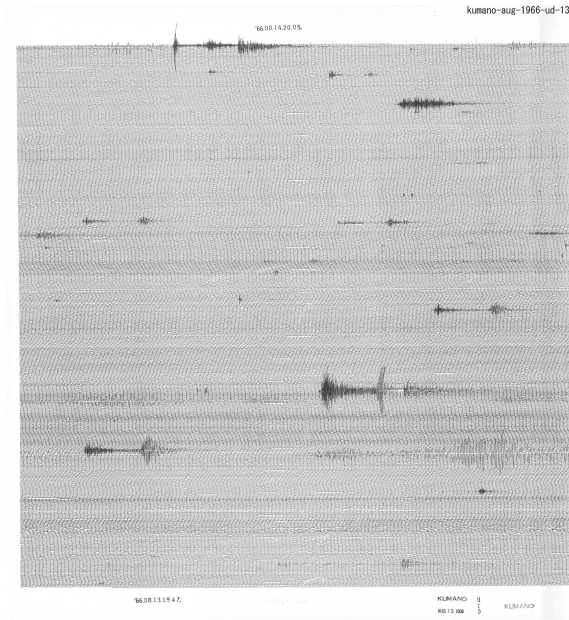


Fig. 2. Seismograph paper record from August 13, 1966, 19:47 to August 14, 1966, 20:05 in Kumano, Japan [14].

and higher frequencies than large earthquakes. Consequently, their digitization is difficult, especially for overlapping waveforms. Therefore, CNN-based image recognition is a promising alternative for analyzing tremors compared to individual waveform extraction through digitization.

Both digital and analog seismic data include a wide variety of phenomena such as earthquakes, tremors, pulsations excited by oceanic waves, teleseisms (distant earthquakes), oscillations due to meteorological events, and noise. Thus, identifying tremors from seismic data becomes difficult when large earthquakes or other signals in similar spectra pollute the measurements.

3 Methods

To detect tremors by applying CNNs to real data, we conducted numerical experiments based on synthetic images to obtain a trained model that determines correctly the presence of tremors in an input image. In this section, we detail these experiments, including image synthetization and CNN construction.

3.1 Generation of Synthetic Images

We generate synthetic images based on seismograph paper records from Kumano, Japan (Fig. 2). Figure 3 shows examples of synthetic images. Each image is

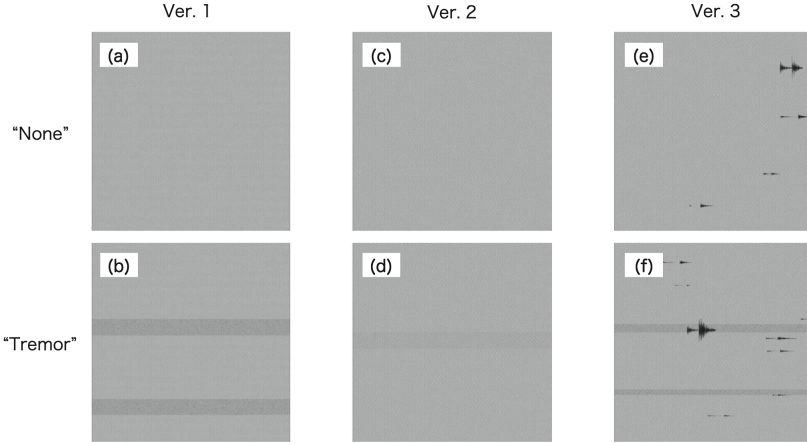


Fig. 3. Representative examples of synthetic images. Details of each image are available on the text and Table 1.

Table 1. Versions of synthetic images. The checkmarks represent the waveforms contained in each version. The values next to the checkmarks indicate the ratio of the waveform amplitude with respect to Ver. 1.

Waveforms	Ver. 1	Ver. 2	Ver. 3
Observation noise (Gaussian noise)	✓	✓	✓
Time stamps (pulses every second)	✓	✓	✓
Pulsations (Gaussian noise + bandpass filter)	✓ (1.0)	✓ (2.0)	✓ (2.0)
Tremors (Gaussian noise + bandpass filter)	✓ (1.0)	✓ (0.5)	✓ (1.0)
Earthquakes (P waves & S waves)			✓

7000×7000 pixels and corresponds to a daily record that contains 576 time series of 2.5 min vertically stacked. The synthetic images correspond to one of three versions, Ver. 1, Ver. 2, or Ver. 3, according to the included types of signals and noise listed in Table 1. For each version, 100 images were generated without tremors (labeled “none”), and 100 images were generated with tremors (labeled “tremor”). Ver. 1 images (Figs. 3(a) and 3(b)) contain observation noise, time stamps, and pulsations, as well as tremors only for “tremor” images. Ver. 1 represents a simple case that allows straightforward tremor detection. In Ver. 2 images (Figs. 3(c) and 3(d)), the tremors are smaller and the pulsations are larger than those in Ver. 1 images. In Ver. 3 images (Figs. 3(e) and 3(f)), we change the amplitudes of tremors and pulsations and insert earthquakes. Figure 4 confirms that Ver. 3 images suitably resemble the paper records regarding both their overview and details.

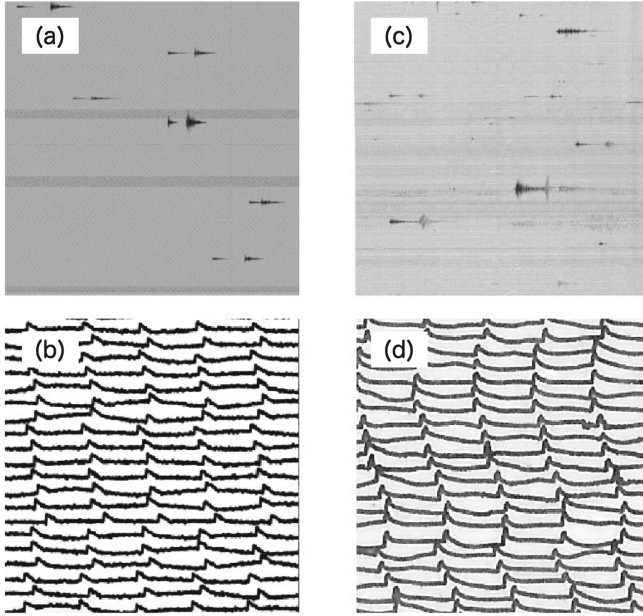


Fig. 4. (a) Ver. 3 “tremor” image. (b) Magnified view of (a). (c) Image of paper record shown in Fig. 2. (d) Magnified view of (c).

3.2 Preprocessing of Synthetic Images

For preprocessing, we divide each image vertically into five strips of 7000×1400 pixels and then resize each strip to 2000×400 pixels. The vertical division allows to easily distinguish between tremors and noise. As a tremor usually lasts several hours, the five strips must include its evidence, unlike temporary noise. This feature is useful to decide the presence of tremors on the images. For example, consider five strips extracted from an image with unknown label, either “none” or “tremor.” If a CNN assigns “none” to four of the five strips and “tremor” to the remaining strip, we can assume that the image corresponds to “none”, and the misjudgment for the last strip is due to noise. The reduced resolution after resizing aims to reduce the number of model parameters and consequently the computational cost. After preprocessing, each version comprises 500 “none” images and 500 “tremor” images. For each version, we use 800 images for training and the remaining 200 for validation, as detailed in the next subsection.

3.3 CNN for Tremor Detection

The CNN is a representative deep learning method that has exhibited high performance in tasks such as image recognition and handwriting recognition. A CNN has two distinctive layers, the convolutional layer and the pooling layer, which mathematically describe the function of human visual cells. These layers

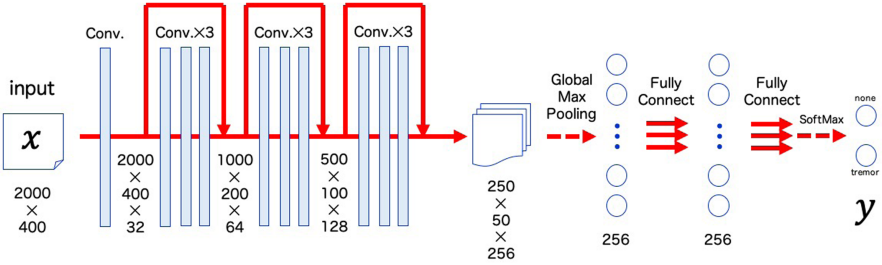


Fig. 5. Architecture of proposed CNN.

allow the CNN to extract features from input images. In recent years, CNNs have begun to be used in seismology for tasks such as detection or discrimination of seismic phenomena (e.g., [7, 13]).

Below, we define the key terms and formulas of CNNs. Let \mathbf{x} be an input image for a CNN. Image \mathbf{x} has a true label ℓ in set S of labels. Let \mathcal{M} be a CNN with internal parameters θ . For an input \mathbf{x} , CNN $\mathcal{M} = \mathcal{M}(\theta)$ provides predictions $\mathbf{y} = (p_\ell)_{\ell \in S}$, where p_ℓ is the predicted probability that the true label of \mathbf{x} is ℓ . In a CNN, the loss is a function that expresses its performance. The loss returns a non-negative real number for a pair (ℓ, \mathbf{y}) of the same \mathbf{x} . If the loss value for \mathbf{x} is close to 0, the prediction performance for \mathbf{x} is favorable. On the other hand, the accuracy is the agreement rate between true label ℓ and predicted label $\hat{\ell} = \arg \max_{\ell \in S} p_\ell$ for the inputs. Training is the process to optimize parameters θ by minimizing the loss value for inputs with known labels. Then, validation calculates the loss and accuracy at fixed parameters θ for inputs with known labels but not used for training. Thus, validation evaluates the predictive performance of the CNN for previously unseen inputs. Finally, test performs prediction on inputs with unknown labels.

We built the proposed CNN by incorporating the residual connections used in the ResNet [1]. Figure 5 shows the architecture of the proposed CNN, which establishes a binary classifier that determines whether the input image has label “none” or “tremor.” The CNN output has the form $\mathbf{y} = (p_{\text{“none”}}, p_{\text{“tremor”}})$ corresponding to $S = \{\text{“none”}, \text{“tremor”}\}$. We used the categorical cross-entropy as the loss function and optimized it using the Adam method [3] with a batch size of 16 in this study.

4 Results

Figure 6 shows the accuracy and loss throughout learning of Ver. 1 images. An epoch (horizontal axis) indicates learning iterations, that is, a set of training using all the training images and validation using all the validation images. Although the validation accuracy does not improve in the initial learning stage, it rapidly increases after 13 epochs to reach almost 1.0, possibly after the model parameters leave a local optimum. Figure 7 shows the gradient-weighted class

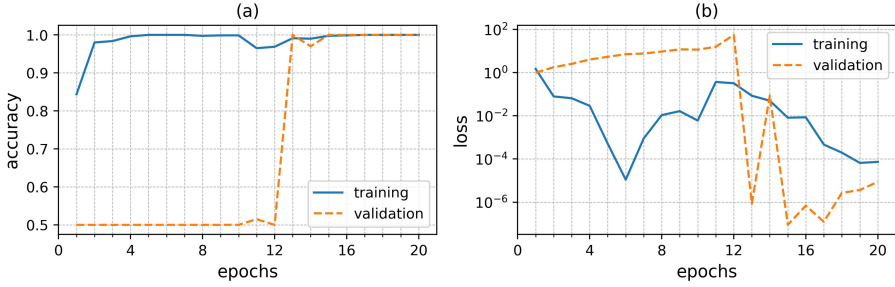


Fig. 6. (a) Accuracy and (b) loss for Ver. 1 images according to number of training epochs.

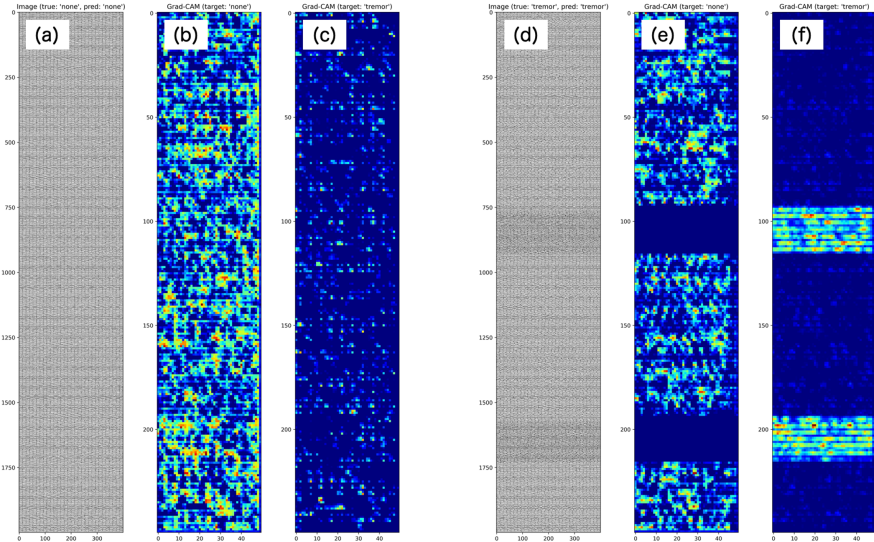


Fig. 7. (a) Ver. 1 “none” image and corresponding Grad-CAMs for prediction of labels (b) “none” and (c) “tremor.” (d) Ver. 1 “tremor” image and corresponding Grad-CAMs for prediction of labels (e) “none” and (f) “tremor.”

activation maps (Grad-CAMs) [16] of Ver. 1 images. Each Grad-CAM indicates the image regions that influence the prediction calculation, with red (light) indicating the highest influence and blue (dark) indicating no influence. For prediction of “none”, the response is uniform on the entire image, except for the areas containing tremors (Figs. 7(b) and 7(e)). In Fig. 7(f), remarkable responses clearly appear in areas containing tremors. Therefore, the model correctly detects tremors and correctly identifies the “tremor” image.

Figures 8 and 9 show the learning performance and Grad-CAMs for Ver. 2 images, respectively. Although the tremors in Ver. 2 are smaller than those in Ver. 1, model training and tremor detection are successful.

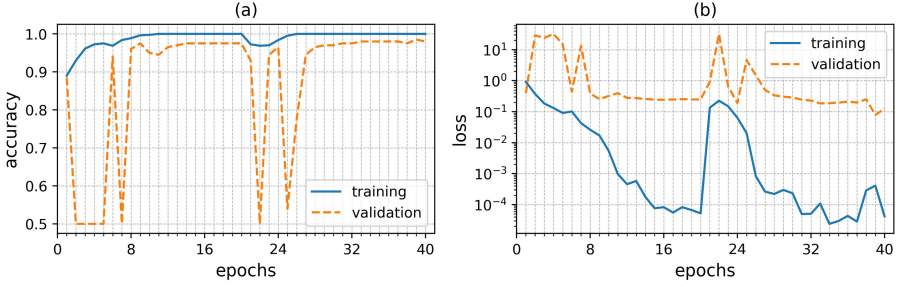


Fig. 8. (a) Accuracy and (b) loss for Ver. 2 images according to number of training epochs.

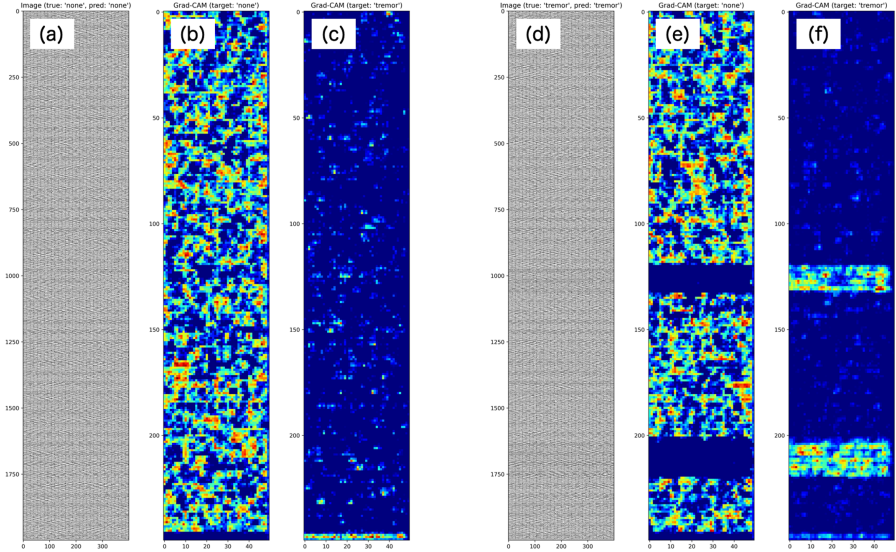


Fig. 9. (a) Ver. 2 “none” image and corresponding Grad-CAMs for prediction of labels (b) “none” and (c) “tremor.” (d) Ver. 2 “tremor” image and corresponding Grad-CAMs for prediction of labels (e) “none” and (f) “tremor.”

Figures 10 and 11 show the learning performance and Grad-CAMs for Ver. 3 images, respectively. In Ver. 3 images, the Grad-CAM responses to earthquake waveforms are notable. For small earthquake waveforms, the Grad-CAMs do not show any remarkable response. This result may be due to the input image shrinking as it passes through the convolutional layers, and small waveforms eventually disappear during shrinking. On the other hand, in Figs. 11(b) and 11(e), the large earthquake waveforms cause the most influential responses. This result suggests that the CNN can distinguish earthquake waveforms from tremors, observation noise, and pulsations. Moreover, the proposed CNN may be able to achieve multinomial classification of labels such as “none,” “tremor,” “earthquake,” and

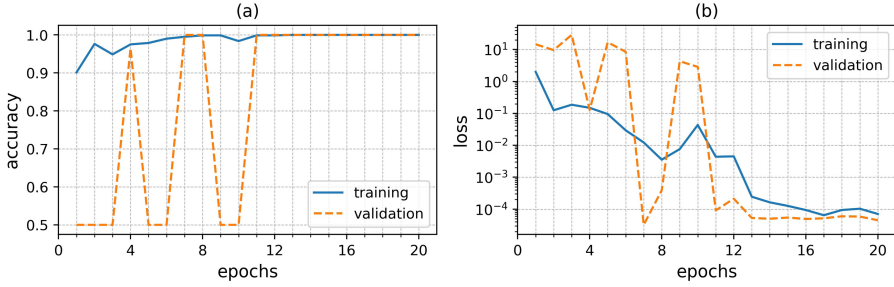


Fig. 10. (a) Accuracy and (b) loss for Ver. 3 images according to number of training epochs.

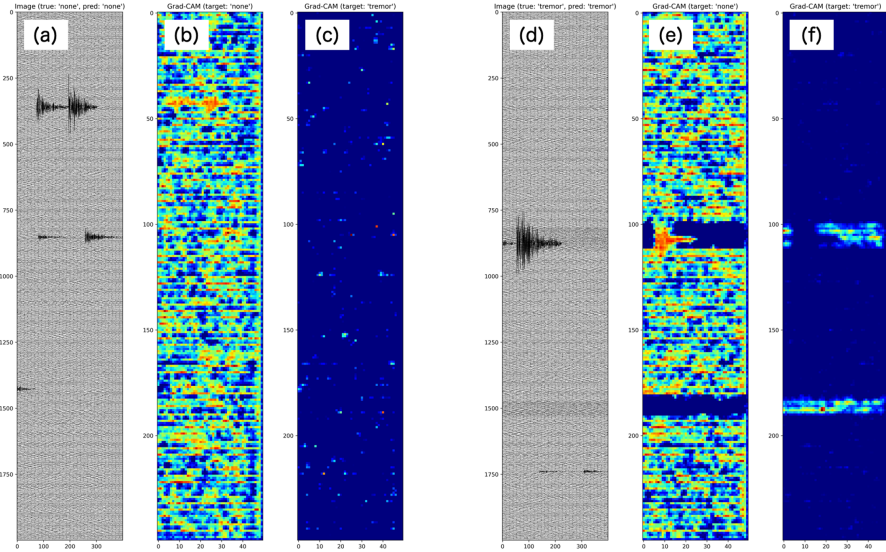


Fig. 11. (a) Ver. 3 “none” image and corresponding Grad-CAMs for prediction of labels (b) “none” and (c) “tremor.” (d) Ver. 3 “tremor” image and corresponding Grad-CAMs for prediction of labels (e) “none” and (f) “tremor.”

“both.” Figure 11(f) shows tremor responses that do not appear when tremors are masked by earthquake waveforms.

Overall, the tremor waveforms appear to show a band-like pattern in the images. This result may indicate that the CNN discriminates tremors based on a rough view of the image. To discard this possibility, we conducted an additional experiment using monochromatic and dichromatic images, as shown in Fig. 12. The set of monochromatic images consists of 256 images for grayscale values from 0 to 255. The set of dichromatic images consists of 200 images, with each image containing a monochromatic band-like pattern on a monochromatic background. By providing these images as inputs to the CNN trained on Ver. 3, all the

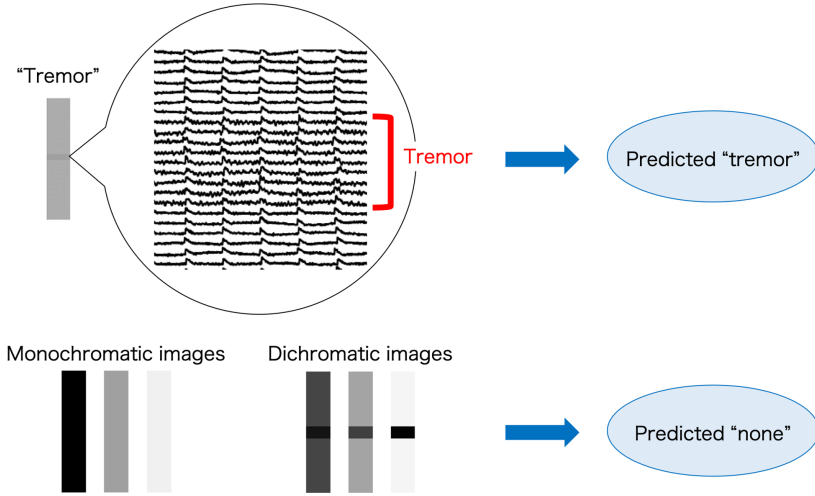


Fig. 12. Proposed CNN trained on Ver. 3 images correctly predicts label “tremor” for “tremor” images and label “none” for all monochromatic and dichromatic images.

predicted labels are “none.” Although Ver. 3 “tremor” images seem to be similar to the dichromatic images, the CNN correctly determines the presence of tremors in Ver. 3 images. This result indicates that the CNN appropriately learned tremor features.

5 Conclusion

The proposed CNN is expected to effectively detect tremors from seismograph paper records, as verified through numerical experiments. Based on the finding from the experiments, we will conduct CNN training with real data, which contain a wider variety of noises than synthetic images. To improve the CNN performance, we will also explore persistent parameter tuning and additional data preprocessing methods.

Acknowledgement. This work was supported by JST CREST under Grant Numbers JPMJCR1763 and JPMJCR1761. The key ideas in this study derived from the activities of JSPS KAKENHI Grant-in-Aids for Scientific Research (B) No. 17H01703, 17H01704, 18H03210, Grant-in-Aid for Scientific Research (S) No. 19H05662, Grant-in-Aid for Challenging Research (Exploratory) No. 20K21785, and Earthquake Research Institute Joint Research ERI JURP 2020-A-05.

References

1. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016). <https://doi.org/10.1109/CVPR.2016.90>

2. Kano, M., Kano, Y.: Possible slow slip event beneath the Kii Peninsula, southwest Japan, inferred from historical tilt records in 1973. *Earth Planets Space* **71**(95) (2019). <https://doi.org/10.1186/s40623-019-1076-9>
3. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization (2017). [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) [cs.LG]
4. Maeda, T., Obara, K.: Spatiotemporal distribution of seismic energy radiation from low-frequency tremor in western Shikoku, Japan. *J. Geophys. Res.: Solid Earth* **114**(B10) (2009). <https://doi.org/10.1029/2008JB006043>
5. Nadeau, R.M., Dolenc, D.: Nonvolcanic tremors deep beneath the San Andreas Fault. *Science* **307**(5708), 389 (2005). <https://doi.org/10.1126/science.1107142>
6. Nagai, R., Kikuchi, M., Yamanaka, Y.: Comparative study on the source processes of recurrent large earthquakes in Sanriku-oki region: the 1968 Tokachi-oki earthquake and the 1994 Sanriku-oki earthquake. *Zisin (J. Seismol. Soc. Jpn. 2nd ser.)* **54**(2), 267–280 (2001). <https://doi.org/10.4294/zisin1948.54.2.267>. (in Japanese)
7. Nakano, M., Sugiyama, D., Hori, T., Kuwatani, T., Tsuboi, S.: Discrimination of seismic signals from earthquakes and tectonic tremor by applying a convolutional neural network to running spectral images. *Seismol. Res. Lett.* **90**(2A), 530–538 (2019). <https://doi.org/10.1785/0220180279>
8. Obara, K.: Nonvolcanic deep tremor associated with subduction in southwest Japan. *Science* **296**(5573), 1679–1681 (2002). <https://doi.org/10.1126/science.1070378>
9. Obara, K., Kasahara, K., Hori, S., Okada, Y.: A densely distributed high-sensitivity seismograph network in Japan: Hi-net by National Research Institute for Earth Science and Disaster Prevention. *Rev. Sci. Instrum.* **76**(2) (2005). <https://doi.org/10.1063/1.1854197>
10. Obara, K., Kato, A.: Connecting slow earthquakes to huge earthquakes. *Science* **353**(6296), 253–257 (2016). <https://doi.org/10.1126/science.aaf1512>
11. Obara, K., Tanaka, S., Maeda, T., Matsuzawa, T.: Depth-dependent activity of non-volcanic tremor in southwest Japan. *Geophys. Res. Lett.* **37**(13) (2010). <https://doi.org/10.1029/2010GL043679>
12. Okada, Y., et al.: Recent progress of seismic observation networks in Japan—Hi-net, F-net, K-NET and KiK-net. *Earth Planets Space* **56**(8), xv–xxviii (2004). <https://doi.org/10.1186/BF03353076>
13. Perol, T., Gharbi, M., Denolle, M.: Convolutional neural network for earthquake detection and location. *Sci. Adv.* **4** (2017). <https://doi.org/10.1126/sciadv.1700578>
14. Satake, K., Tsuruoka, H., Murotani, S., Tsumura, K.: Analog seismogram archives at the Earthquake Research Institute, the University of Tokyo. *Seismol. Res. Lett.* **91**(3), 1384–1393 (2020). <https://doi.org/10.1785/0220190281>
15. Schwartz, S.Y., Rokosky, J.M.: Slow slip events and seismic tremor at circum-Pacific subduction zones. *Rev. Geophys.* **45**(3) (2007). <https://doi.org/10.1029/2006RG000208>
16. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: visual explanations from deep networks via gradient-based localization. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 618–626 (2017). <https://doi.org/10.1109/ICCV.2017.74>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

