

Automatic Digitization of Seismograms in Smoked Paper using Convolutional Neural Network

*Mitsuko Furumura¹, Yoko Ogawa², Kazuo Sakamoto², Ritsuko S. Matsu'ura¹

1. Research Division, Earthquake Research Center, Association for the Development of Earthquake Prediction, 2. KOZO KEIKAKU ENGINEERING Inc.

1. Introduction

There are many analog seismograms of past destructive earthquakes in the world, and some researchers have already started digitization projects (e.g., Michelini et al., 2005; Paulescu et al., 2016). However, the digitization technique for unclear seismograms in smoked papers is still in progress.

Here, we tried experiments to digitize seismograms automatically through machine learning.

2. Used data and outline of the experiments

Two types of images were used in our experiments.

One type is synthesized from K-NET data (synthetic images). First, accelerograms of the earthquakes with various magnitudes, focal depth, and hypocenter distances are converted into displacement records. Then, the records were curved along the movement of the needle mechanism of the seismographs and background images were combined in order to make realistic seismograms. As the background of the synthetic images, we use scanned images described below, including characters, scratches or folds that sometimes hinder the tracing. The resulting 561 images were used as the synthetic images and the corresponding displacement records were used as supervised data.

The other type is scanned images of analog seismograms which are acquired using large flatbed color scanners with mechanical resolutions of 400 or 500 [dpi] (real images). As supervised data for these images, we used 21 open data digitized by the National Research Institute of Fire and Disaster (NRIFD) and 11 data that we digitized manually.

For training and testing our neural network, fixed height (1200 [pixel]) images containing target waveforms were clipped out from the synthetic and real images, and morphed to remove curvature due to the needle mechanism.

Our network directly outputs a Y coordinate for each X coordinate of a drawn waveform from a 2-D image input. We structure the neural network without full-connection layers for acceptance of various sizes of images and add a pooling layer at bottleneck to get 1-D outputs. The network is trained using mean squared error as the loss function.

3. Results

The convolutional neural networks are trained using three data sets; (1) 561 synthetic images only, (2) combination of 561 synthetic and 31 real images and (3) 31 real images only. We tested 22 real images of seismograms which were not used for the training and compared the output Y values with the true one generated by corresponding NRIFD data for verifying its performance. RMSEs (root mean square errors) of the three cases were 20.67, 16.41 and 15.48 [pixel], respectively, and case (3) was the best. Fig. 1 shows a result of case (3). It seems that automatic digitization of the seismograms works well in many parts, but the errors are still large at peaks of the waveforms. Fig. 2 shows comparison of the three cases. The errors of case (1) and case (2) at the peaks are less than one of case (3). However, case (1) and case (2) are susceptible to vertical edges in characters or folds, thereby the RMSEs of case (1) and (2) are larger than case (3).

4. For future developments

We almost succeeded automatic digitization of the seismograms in the smoked paper. The scanned images of seismograms are quite useful for training the network but preparing many supervised data is not easy. Therefore, training using more synthetic images with the noises mentioned above is required to improve the performance.

Acknowledgements

This project has been supported by the Headquarters for Earthquake Research Promotion (HERP) of the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan. We used K-NET data by NIED and digital strong-motion data by the National Research Institute of Fire and Disaster.

Michellini et al. (2005). Eos Trans. AGU 86, 261–266.

Paulescu et al. (2016). Acta Geophys. 64, no. 4, 963–977.

Keywords: automatic digitization, analog seismograms, convolutional neural network

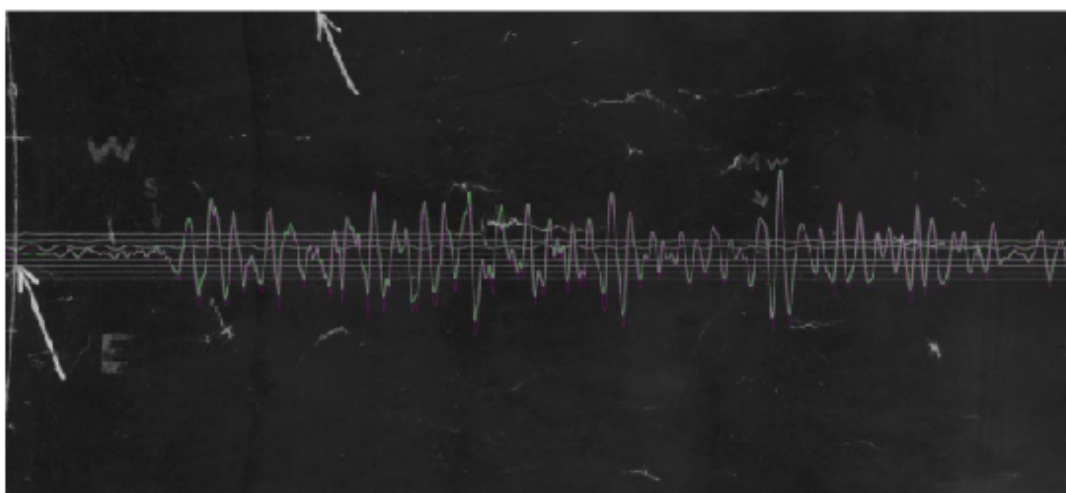


Fig. 1 Scanned image of a seismogram used for verification, manually digitized data (green) and output data of the neural network (magenta).

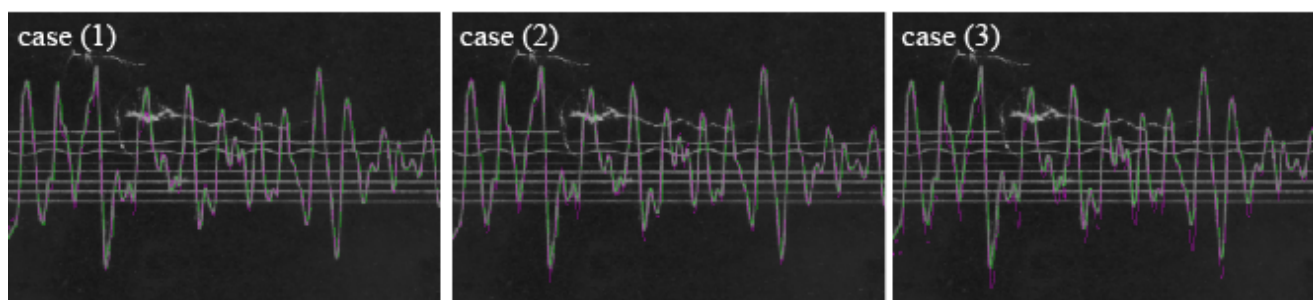


Fig. 2 Results of the three cases: manually digitized data (green) and output data of the neural network (magenta).