Seismic Event Classification Using Convolutional Neural Networks on Multichannel Waveform Data

Nicholas Laureano

University of Oregon

laureano@uoregon.edu

Abstract:
Earthquake early warning systems (EEWS) rely on rapid and accurate classification of seismic waveforms to provide timely alerts. Such systems can be improved with the use of advanced machine learning techniques. This study investigates the use of convolutional neural networks (CNNs) for classifying seismographic data into four categories: background **Noise** or earthquakes classified by magnitude (**Low**, **Medium**, or **High**). Using a curated subset of the STanford EArthquake Dataset (STEAD), I trained and evaluated several CNN models, optimizing key hyperparameters such as learning rate and weight decay. The final model achieved an **overall accuracy** of **74.76%**, with strong performance in distinguishing medium and high-magnitude events, and a high **binary accuracy** (Earthquake vs. Noise) of up to **84.53%**. However, classification of low-magnitude events proved challenging due to waveform similarity with background noise. These results suggest that while CNNs offer robust performance for high-certainty seismic events, additional refinement—potentially through hybrid models or expanded datasets—is necessary to improve performance on ambiguous signals. This research highlights the promise of deep learning in enhancing real-time earthquake detection systems.

I. Introduction

Earthquakes remain among the deadliest and most unpredictable natural disasters in modern society. The increasing concentration of populations in urban centers—often with aging infrastructure not designed to withstand strong seismic activity—amplifies the potential for catastrophic loss. Historical events such as the 1976 Tangshan earthquake in China, which caused an estimated 242,000 to 655,000 deaths [6], and the 2010 Haiti earthquake, with death toll estimates ranging from 85,000 to 316,000 [5], underscore the urgent need for accurate earthquake forecasting and early warning systems.

Despite this demand, the effectiveness of earthquake "prediction" (See Note 1) has seen only marginal improvements in recent decades. Earthquake forecasting—typically long-term and statistical in nature—relies heavily on historical seismicity and recurrence intervals of past events. While this approach can identify regions of elevated risk over years or decades, it offers limited utility for practical, life-saving interventions.

In contrast, short-term earthquake prediction aims to provide actionable information such as the likely time, location, and magnitude of an impending event. However, these predictions often rely on elusive or poorly understood precursors, anecdotal correlations, or data that is difficult to verify and reproduce [3]. This has led to widespread skepticism regarding the feasibility and reliability of short-term earthquake prediction.

Research in this domain generally falls into two categories: the study of **precursors**—potential warning signs such as radon emissions, electromagnetic anomalies, or foreshocks—and the analysis of **statistical trends**. More recently, machine learning (ML), particularly deep learning, has emerged as a promising tool for detecting subtle patterns in seismic data that may precede earthquakes. Still, the field faces significant challenges, including high false alarm rates, limited generalizability, and the ethical burden of communicating predictions with uncertain outcomes.

Given these challenges, the focus has increasingly shifted toward **improving earthquake early warning systems** (EEWS). Unlike prediction, EEWS do not attempt to forecast earthquakes in advance. Instead, they aim to detect the initial signs of an earthquake in real time and provide nearby population centers with several seconds of lead time before the most damaging waves arrive. These systems rely on rapid identification of seismic signals amid constant background noise, a task that grows increasingly complex

Note 1. The term earthquake prediction in this paper is used here to refer to forecasting and probabilistic predictions with scientific bases. Not to be confused with largely unfounded pseudoscientific prophecies.

as urban environments become noisier and sensor networks more sensitive [8].

In this context, **improving the speed** and **reliability** of earthquake detection—particularly distinguishing true seismic events from background noise—has become a critical research goal. Enhancing these systems can directly translate into saved lives and reduced infrastructure damage, making them a vital area of investigation in seismological research.

II. Background

The proposed exercise in seismic waveform classification aided by deep learning techniques utilizes the following technologies and concepts:

A. Deep Learning & Convolutional Neural Networks:

Deep learning is a subfield of machine learning that employs artificial neural networks composed of multiple layers. These networks, inspired by the structure of biological neurons, consist of interconnected nodes (or "neurons") that process and transmit information. As these layers are stacked, the network gains the capacity to model increasingly abstract and complex patterns from raw data [2].

A particularly effective architecture for pattern recognition in spatial or temporal data is the **convolutional neural network** (CNN). CNNs apply convolutional filters across input data to extract local features while preserving spatial (or temporal) relationships. This approach has seen success in domains such as image classification and speech recognition, where local patterns and context are essential for accurate interpretation [2].

In the context of seismic waveform analysis, CNNs are employed to **distinguish earthquakes from background noise by identifying characteristic waveform features** indicative of seismic activity. By learning these features directly from waveform data, CNN-based classifiers can outperform traditional threshold-based or hand-engineered feature approaches in both speed and accuracy [7].

B. Seismographic Waveforms

Seismographic waveforms are continuous time-series recordings of ground motion, typically captured by sensitive instruments known as seismometers. These waveforms reflect a combination of natural and anthropogenic (human-generated) vibrations, including earthquakes, volcanic activity, mining operations, traffic, and other sources of noise [8].

A seismic event produces a distinct pattern of waves—commonly categorized as P-waves (primary), S-waves (secondary), and surface waves—which propagate through the Earth at different speeds and amplitudes. The identification and classification of these waveforms form the foundation of earthquake detection [1].

In machine learning contexts, these waveforms are typically preprocessed (e.g., normalized, windowed, and labeled) and then fed into models such as CNNs to train classifiers capable of detecting the presence of seismic events. The key challenge lies in differentiating earthquake signals from the wide variety of non-earthquake background noise present in raw seismographic data [7].

C. STEAD (STanford EArthquake Dataset)

STEAD serves as the foundational dataset for training and evaluating these deep learning models [4]. Developed to address the challenges of inconsistent labeling and limited availability of high-quality benchmark datasets in seismology, STEAD provides a large-scale, well-curated collection of seismic waveform data. The dataset includes approximately **1.2 million labeled time series** divided into two categories: local **earthquake waveforms** and seismic **background noise** without earthquake activity.

These labels offer a reliable ground truth essential for supervised learning models like convolutional neural networks. By training on this diverse and accurately labeled dataset, each model is intended to learn to distinguish between earthquake events and background noise.

III. Methods

Due to the immense size of STEAD, practical training on the entire dataset is computationally intensive and generally restricted to high-performance computing environments. To work within available resources while maintaining representativeness, I implemented a carefully constructed subsampling strategy.

STEAD contains **1,265,657** labeled seismic waveform examples, with about **81.4%** labeled as **earthquakes** (1,030,231 examples) and **18.6%** labeled as **noise** (235,426 examples). While this imbalance mirrors real-world distributions, for training a robust classifier, it was determined that increasing the proportion of **noise** in the sampled dataset would encourage better model performance when detecting **low**-signal or **non-seismic** events. Therefore, a **40-60 noise-to-earthquake** ratio was selected as a practical compromise between dataset fidelity and computational feasibility.

To further stratify the earthquake data, events were divided into three magnitude-based groups:

- **Low** ($x \leq 2.0$): 773,278 examples
- **Medium** ($2.0 < x \leq 4.0$): 237,384 examples
- **High** ($x > 4.0$): 19,569 examples

To keep the total sample size manageable, **50,000 examples** were selected, ensuring an even spread of earthquake magnitudes and a substantial proportion (**40%**) of **noise**.

	Totals	Training	Valid.	Testing	% of Pop.
Low	10,000	7,000	1,000	2,000	~1.29% of Low
Med	10,000	7,000	1,000	2,000	~4.21% of Med
High	10,000	7,000	1,000	2,000	~51.1% of High
Noise	20,000	14,000	2,000	4,000	~8.49% of Noise
Total	50,000	35,000	5,000	10,000	~0.039% of All
	100%	70%	10%	20%	

Figure 1. Testing, Validation, & Testing Sample Breakdown

These 4 categories were filled with the first matching occurrences in order within the STEAD dataset. From there, the samples were further divided into their respective **training**, **validation**, and **testing** sets randomly.

The structure of the CNN is backed by the underlying idea that we can extract features from the data as layers convolute downward, with each layer extracting more complex features from higher simpler features [2]. To compromise between processing time and feature extraction, the following design was proposed. (See Figure 2)

Training, built using the Pytorch library for machine learning in Python, consists of 10 epochs. Each epoch finished with a round of validation testing to ensure that the model being produced has similar performance on untrained data across all epochs. The **learning rate** and **weight decay** values, both **0.0001**, were decided due to the analysis in the following section. The final model produced was then evaluated against the testing set designated earlier.

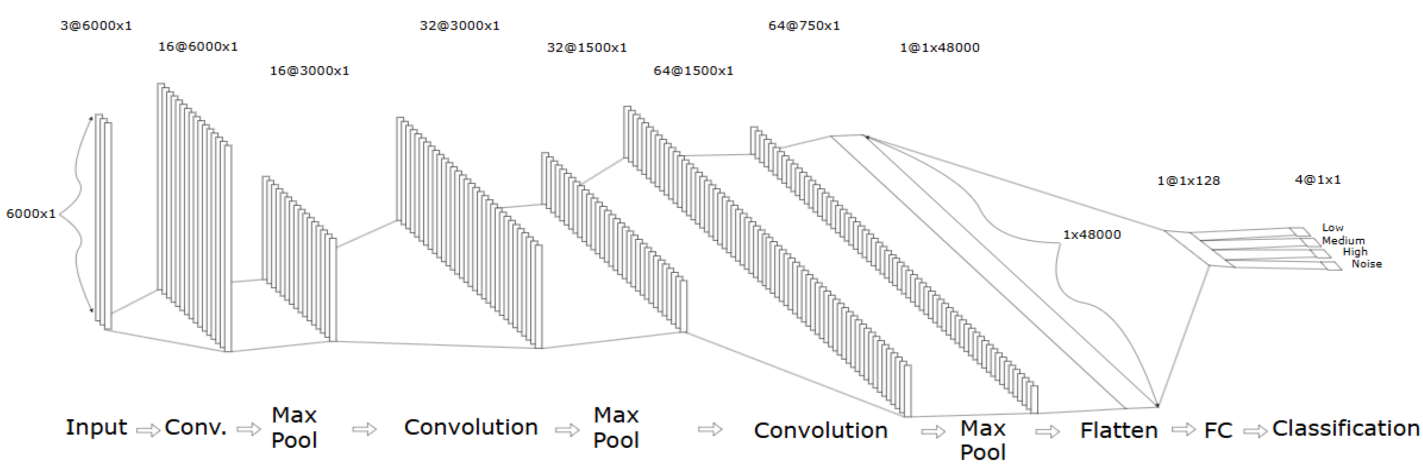

Figure 2. Seismic CNN Model Architecture

IV. Experimentation & Analysis

All training and evaluation processes were conducted with **GPU acceleration** using an **NVIDIA GeForce RTX 3070 Ti.** Over the course of experimentation, eight models (including one baseline model used in both benchmark comparisons) were trained and evaluated. Total training time amounted to approximately **14.2 hours**, averaging 1.7 hours per model.

The primary focus of this phase was to identify the optimal hyperparameters to maximize performance on the training and validation datasets. Among all candidates, **learning rate** and **weight decay** were determined to have the most significant impact on model generalization and accuracy.

Learning Rate:

A range of learning rates was evaluated:

- 1e−3 (0.001) [Default]
- 1e−4 (0.0001)
- 5e−4 (0.0005)
- 3e−3 (0.003)

Performance metrics, including accuracy and loss curves for both training and validation data, were recorded across 10 epochs. The results showed that (See Figure 3.4):

- 1e-4 consistently outperformed other rates in terms of validation accuracy and stability.
- Higher learning rates, particularly 3e-3, led to erratic performance and convergence issues.
- Even though 1e-3 (the default) showed reasonable learning behavior, it underperformed on validation metrics compared to 1e-4.

Conclusion: A learning rate of **1e-4** (0.0001) is optimal.

Weight Decay Rate:

The evaluated weight decay values included:

- 1e-5 (0.00001) [Default]
- 0
- 1e-6 (0.000001)
- 1e-4 (0.0001)
- 1e-3 (0.001)

Key observations over the 10-epoch runs (See Figure 4.4):

- Both 1e-4 and 1e-6 achieved comparable validation accuracy after 10 epochs.
- However, 1e-4 produced steadier gains and less variability in validation accuracy.
- 1e-6 exhibited inconsistent improvements, with accuracy ranging between **68%** and **82%**, suggesting instability despite reaching the same validation accuracy after the tenth epoch.

Conclusion: Given its stability and consistent trend, a weight decay of **1e-4** (0.0001) is optimal.

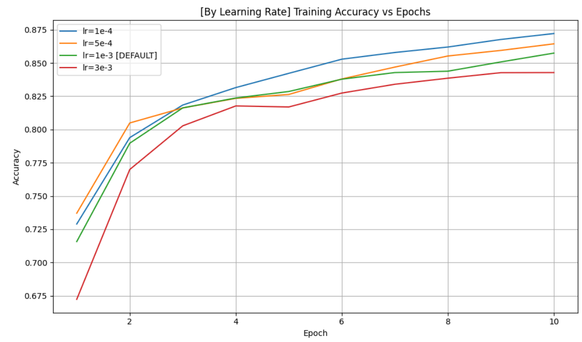

Figure 3.1 Learning Rates on Training Acc. over Epochs

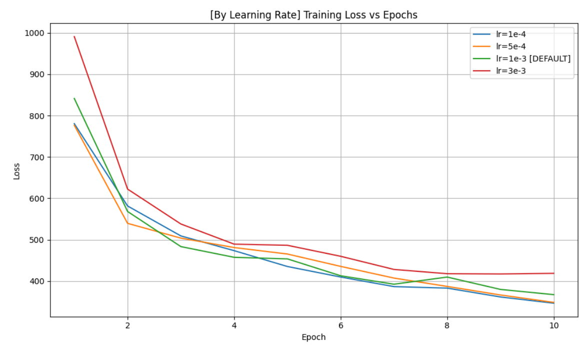

Figure 3.2 Learning Rates on Training Loss over Epochs

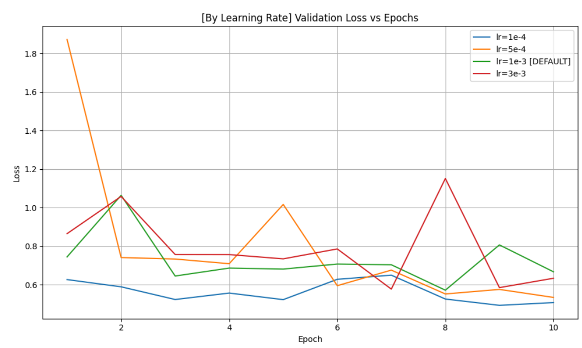

Figure 3.3 Learning Rates on Validation Loss over Epochs

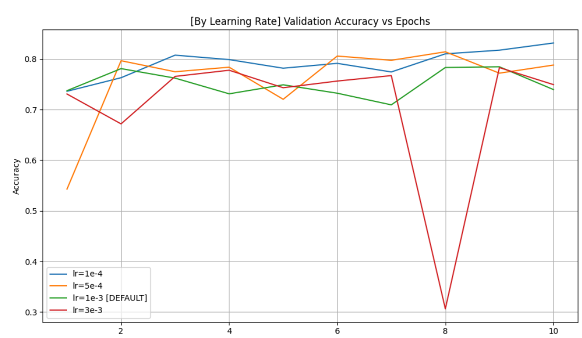

Figure 3.4 Learning Rates on Validation Acc. over Epochs

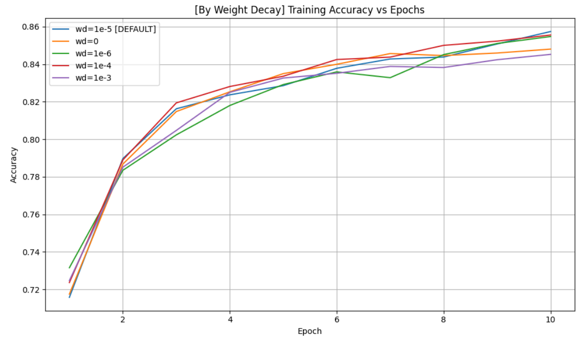

Figure 4.1 Weight Decay Rates on Tra. Acc. over Epochs

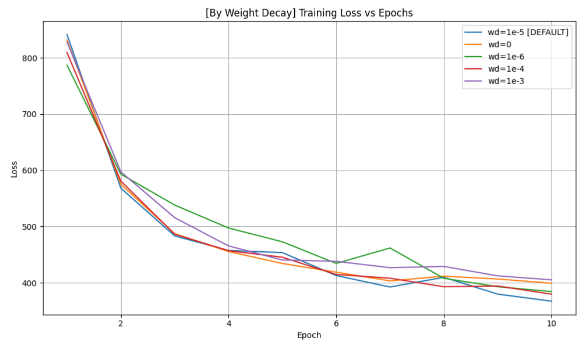

Figure 4.2 Weight Decay Rates on Tra. Loss over Epochs

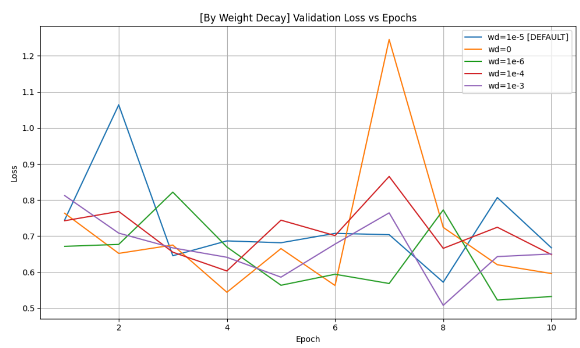

Figure 4.3 Weight Decay Rates on Val. Loss over Epochs

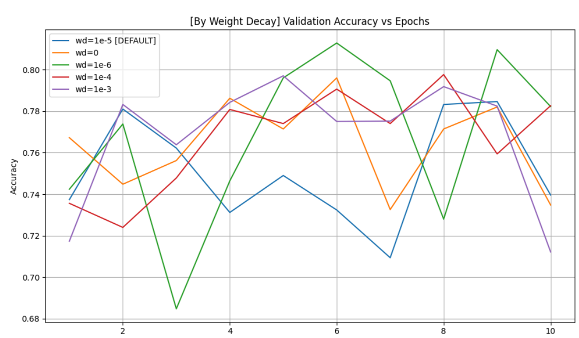

Figure 4.4 Weight Decay Rates on Val. Acc. over Epochs

Final Model Evaluation:

With the model architecture and hyperparameters tuned as previously discussed, I conducted a final evaluation using the held-out test set of **10,000 labeled examples**. The primary hypothesis guiding this evaluation states that the model would likely show a high degree of misclassification between the **Noise** and **Low** earthquake classes. This stems from the inherent similarity in waveform features between background seismic **Noise** and **Low**-energy seismic events, which can exhibit weak or indistinct **P** and **S wave** arrivals. In contrast, **Medium** and **High**-magnitude earthquakes typically display clearer, more intense wave patterns, making them easier to distinguish.

The model achieved an overall classification accuracy of **74.76%.** The **confusion matrix** (see Figure 5) illustrates the distribution of predictions across the four classes: **Noise**, **Low**, **Medium**, and **High**. To further contextualize the performance, I also include two accuracy metrics:

- **Exact Accuracy**: Correct prediction of the precise class label
- **Earthquake/Noise Accuracy**: Grouping all earthquake classes into a single "Earthquake" label and measuring binary classification accuracy (Earthquake vs. Noise)

Unsurprisingly, the model performed very well in identifying **Noise**, **Medium**, and **High** classes. However, as hypothesized, the **Low** class showed substantial overlap with **Noise** in predictions. Only **30.75%** of **Low** events were correctly classified, with the majority being misidentified as **Noise**. Despite this, the inverse misclassification (**Noise** being labeled as **Low**) occurred far less frequently, which suggests asymmetric confusion between these two classes.

To aid interpretation, the confusion matrix includes a color-coded severity scale:

- **Green**: Correct classifications
- **Yellow**/**Orange**/**Red**: Incorrect classifications of increasing severity

Importantly, the model showed strong resilience against severe misclassification. There were only **16** instances out of **10,000** where **High**-magnitude earthquakes were classified as **Noise**, and zero examples of the **Noise** or **Low** classes being classified as **High**. This indicates that while exact classification accuracy could be improved, the model rarely misclassifies events across opposite extremes, which is critical in a real-world EEWS.

	ACTUAL LABEL				
Predictions	**Noise**	**Low**	**Med**	**High**	**Totals**
Noise	3952	1367	116	16	5451
Low	30	615	550	2	1197
Medium	18	18	1004	77	1117
High	0	0	330	1905	2235
Totals	4000	2000	2000	2000	

Exact Acc.	98.80%	30.75%	50.20%	95.25%
EQ/Noise Acc	X	31.65%	94.20%	99.20%

Evaluation	
Exact Acc.	EQ/Noise Acc.
72.50%	X
51.37%	97.49%
89.88%	98.38%
85.23%	100.0%

Exact All	74.76%
EQ/N All	84.53%
By $(\frac{3952}{4000} \times 0.4) + (\frac{4501}{6000} \times 0.6)$	

Figure 5. Final Testing Data Evaluation Breakdown

V. Conclusions & Further Directions

In this study, I explored the application of a convolutional neural network (CNN) for classifying seismic waveform data into four distinct categories: **Noise**, or **Low-**, **Medium-**, and **High**-magnitude earthquakes, using a curated subset of the STanford EArthquake Dataset (STEAD). Our goal was to assess the feasibility of automated event classification in the context of earthquake early-warning systems.

The final model demonstrated robust performance on high-certainty classifications, particularly in distinguishing between **Noise**, **Medium**, and **High**-magnitude events. Notably, it exhibited a strong ability to avoid dangerous false negatives—with virtually no instances of **High**-magnitude events being misclassified as **Noise** or **Low**, and no cases of benign background signals being interpreted as severe events. These results are encouraging for real-time deployment scenarios, where minimizing severe misclassifications is critical for public safety and system trustworthiness. Ultimately, the model can be described under this context as having an **84.53% accuracy** when labeling earthquakes as earthquakes (regardless of intensity) and noise as noise proportional to the makeup of the testing distribution.

However, the evaluation also highlighted a significant area for improvement: the model's difficulty in reliably distinguishing low-magnitude earthquakes from noise. This challenge reflects the subtlety and similarity of waveform features in low-energy seismic events, which often lack distinct spectral markers. The model correctly classified only **30.75%** of **low-magnitude events**, suggesting the need for more refined feature extraction or alternate architectures (e.g. **LSTM** or **RNN**) to better capture these nuances.

Future directions to address this limitation may include:

- Incorporating additional neural-network architectures in predictions (e.g., **CNN** & **LSTM** for spatial and temporal feature extraction),
- **Foreshock classification** and analysis (e.g. identifying foreshocks within waveforms, in a pipeline architecture)
- **Increasing** training, validation, and testing **dataset sizes** to increase confidence in model generalizability. (Of course, this requires increased time and processing power)
- **Analyzing** final model **kernel filters** for waveform feature interpretation.

Overall, while this model represents a promising step toward more effective waveform classification, especially for early warning purposes, continued research and refinement are essential. The findings reinforce that even partial success in this domain can yield meaningful improvements in preparedness and response if integrated with broader seismic monitoring systems.

Code Availability:

The full implementation, including training scripts, model definitions, and evaluation tools, is publicly available at:

https://github.com/NLaureano/SeismicCNN

<u>**References**</u>

1. The Editors of Encyclopaedia Britannica (2025, May 3). seismic wave. Encyclopedia Britannica. https://www.britannica.com/science/seismic-wave
2. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. http://www.deeplearningbook.org
3. Yan Y. Kagan, Are earthquakes predictable?, Geophysical Journal International, Volume 131, Issue 3, December 1997, Pages 505–525, https://doi.org/10.1111/j.1365-246X.1997.tb06595.x
4. Mousavi, S.Mostafa & Sheng, Yixiao & Weiqiang, Zhu & Beroza, Gregory. (2019). STanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2947848.
5. Pallardy, R. (2025, April 21). 2010 Haiti earthquake. Encyclopedia Britannica. https://www.britannica.com/event/2010-Haiti-earthquake
6. Rafferty, J.P. (2024, August 14). Tangshan earthquake of 1976. Encyclopedia Britannica. https://www.britannica.com/event/Tangshan-earthquake-of-1976
7. Rigobert Tibi, Patrick Hammond, Ronald Brogan, Christopher J. Young, Keith Koper; Deep Learning Denoising Applied to Regional Distance Seismic Data in Utah. Bulletin of the Seismological Society of America 2021;; 111 (2): 775–790. doi: https://doi.org/10.1785/0120200292
8. Lei Yang et al. ,Toward improved urban earthquake monitoring through deep-learning-based noise suppression.Sci. Adv.8,eabl3564(2022).DOI:10.1126/sciadv.abl3564