This Earthquake Doesn't Exist

Anonymous Author(s)
Affiliation
Address
email

Abstract

This study applies Conditional Generative Adversarial Networks (cGAN) to the field of seismology. With GAN, realistic seismic waveforms can be created for various applications, such as augmenting limited seismic data or modeling, or generating realistic noise. A potential and alarming application of GAN is to generate realistic seismic signals that can cause disturbances to the international treaty banning nuclear explosions (CTBT). Results show that the generated seismic waves are nearly indistinguishable from real ones.

1 Introduction

Seismic waves are created when an earthquake or explosion causes the ground to shake, and they can be recorded with seismic sensors. The science that studies seismic waves is called seismology. This field has helped explore the structure of the Earth, to understand earthquakes, identify seismic hazards for various locations, and to understand and predict natural disasters. A seismic instrument can either measure the vertical component of the ground motion or both the horizontal (typically N-S and W-E motion) and vertical components. Ground motion is recorded as a waveform, where negative values mean negative motion and vice versa.

An important step in seismological data analysis is the calculation of synthetic *seismograms* (records produced by seismic sensors): one computes a forward model of the event and fits the observed waveform to the model. Such models are often slow and computationally expensive to render. It may thus be of interest to use generative adversarial networks (GAN) - a technique from the machine learning domain that can generate realistic synthetic data by training two competing neural networks against each other (one network generates synthetic data and the other judges whether it is real or fake) [1]. We apply GAN in the field of seismology and show that our approach can generate realistic seismic waveforms. While GANs are extensively used for generating images, less research has been done on other data types. We demonstrate how GAN can be trained directly on raw waveforms (1D time-series) as opposed to training on time-frequency representation (2D images).

Generating any seismic waveform is, however, of little use. To make effective use of such generated data, certain parameters must be considered (e.g. time of signal arrival to the station, azimuth of propagation, focal mechanism of the earthquake, etc.). That can be done by using Conditional GAN (cGAN) [2]. In the context of seismology, this idea has been probed by [5] in connection to detecting earthquakes in Oklahoma. They used a cGAN to generate realistic synthetic three-component seismic waveforms for two classes: arrivals from earthquakes and background noise. They found that dataset augmentation with GAN-generated seismic data could improve detection algorithms in instances when only limited amounts of labeled data are available. Despite working independently from their group, our work extends the idea of conditional generation to a number of classes for variables such as distance from source and source magnitude, which provide additional advantages for enriching datasets with relevant synthetic information and expand the range of applicability of GAN in seismology.

Submitted to NeurIPS 2021 AI for Science Workshop.

There are more potential applications for generated seismic signals: (1) GAN could potentially be used to create an earthquake response to nuclear explosions, which might be used as proof that an explosion took place when there wasn't one. The international community should therefore be aware of such possibilities, rule out hostile uses of GANs, and take preventative measures to recognize this threat. Otherwise GAN-generated seismic signals might potentially cause problems for the future Comprehensive Nuclear-Test-Ban Treaty, which will ban all nuclear explosions. (2) Despite the abundance of seismic data (tens of thousands of sensors are recording seismic activity globally) only the fraction is labeled. GAN-generated signals can augment limited seismic data, for instance for seismological tasks requiring machine learning. (3) Forward modeling of a seismic signal is computationally expensive and slow. Machine Learning inference is fast, and forward modeling can benefit from this speedup. (4) Many methods of seismology are tested on synthetic seismograms, with added synthetic noise (which is typically just normally distributed). We could replace that noise with realistic GAN-generated noise, and thus improve the capability of those methods to deal with real (recorded) seismic data. (5) The seismic hazard in a region is determined by the frequency and magnitude of past events. Geological events such as earthquakes and landslides provide a wealth of information about the region and can be used to better assess the seismic hazard in the area. Perhaps one would be able to *fine-tune* a pre-trained GAN (to take weights of a trained neural network and use it as initialization for a new model being trained on data from the same domain) to locally recorded seismic events. If that works, one would be able to generate events that have all the properties of events that already happened in the area but didn't actually happen (e.g. in some regions perhaps the only events of magnitude 4 and lower were registered. Using cGAN, one would be able to look at events of higher magnitudes as if they happened in the area). (6) Since there are so many seismic stations globally, one can perhaps select a particular earthquake event, fine-tune GAN on all available records of this event and produce "virtual receivers", displaying how this particular earthquake would have looked like if recorded from some other location.

2 Data and Methods

We propose a model based on Auxiliary Classifier GAN [4] (see Fig. 1) to generate seismic signals that do not exist. We called this network - TEDE GAN (This Earthquake Doesn't Exist GAN). We train the model on STanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI [3] - earthquake waveform dataset consisting of nearly 1.5 mln 3-channel samples (one vertical displacement and two horizontal displacements) as well as $\sim$0.5 mln samples of seismic noise (no earthquake is recorded). This data were preprocessed as follows: we removed the instrument response, converted the data to displacement, and normalized the channels on the global maximum (for each sample). Since this dataset is labeled (Receiver Type; Polarization; P- arrival time; S- arrival time; Source Depth, km; Source Magnitude; Source Mechanism (Strike Dip Rake); Source Distance, km; Back Azimuth, Deg; Signal-to-Noise ratio), we can exploit labels for a conditional generation. We achieve this by converting continuous labels into discrete bins of size N (number of classes we want to train the model on). Each bin is constructed in such a way as to contain an equal number of samples to avoid the label imbalance problem. We condition our model on the aforementioned properties and demonstrate that we can generate realistic seismic waveforms with any conditional properties.

3 Results

First, we generate seismic noise with no conditions (see Fig. 2A). This noise has all the properties of what would be recorded on a seismometer in the absence of an earthquake. We then generate 3-component earthquake records with a conditional label and demonstrate how TEDE GAN can be conditioned on the source distance, i.e. how far away, from the receiver, the earthquake happened (see Fig. 2C). The farther the receiver is from the source, the longer the signal should travel. We can see how with increasing distance in the figure, the first arrivals (large amplitudes) are shifting towards the right of the window. The TEDE GAN can also be conditioned on the magnitude of the source, i.e. how strong the earthquake is (see Fig. 2B). One way to tell if the earthquake was strong is to look at the amplitudes of the seismic signal, however, since our TEDE GAN generates signals with normalized amplitudes (in the range [-1,+1]), we cannot use this as a visual cue about the quality of the generated conditioned signal. Nevertheless, the higher the magnitude of the earthquake - the more energy is released over time, and this is something that we indeed can observe on the generated

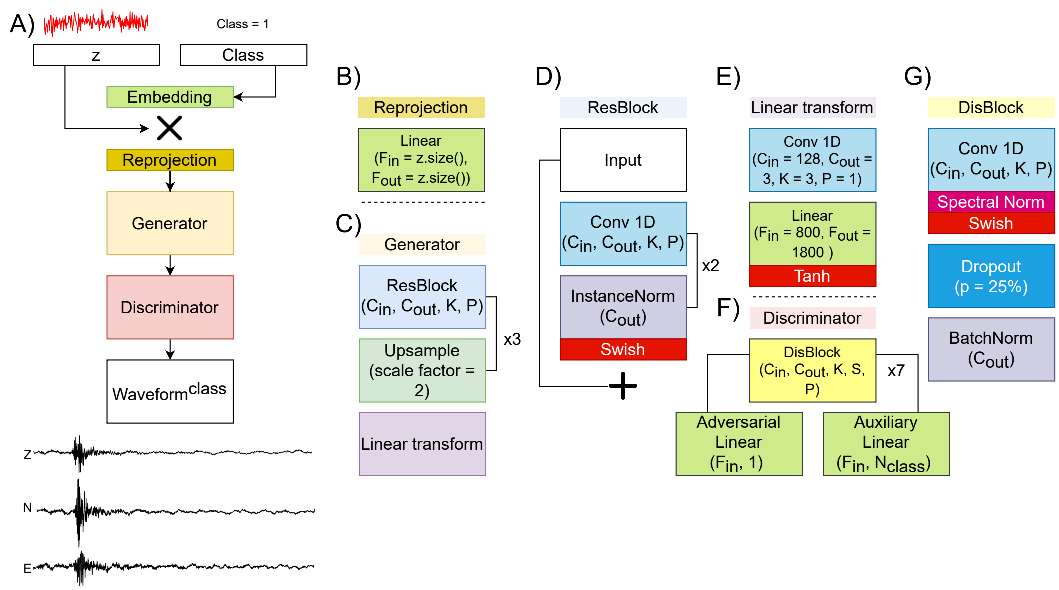

Figure 1: The proposed model uses an auxiliary classifier GAN that makes two predictions - real/fake and class. A) Architecture of the This Earthquake Doesn't Exist GAN. The network takes a latent vector z and a class label as an input. The label is then embedded with a linear embedding layer and multiplied by z. Using a linear layer, the obtained vector is re-projected to the latent dimension (see B). B) Re-projection block, used to balance the influence of z and embeddings. C) Architecture of the Generator. We adopt ResNet-like convolutional blocks (see D) to generate 3-component (Z, N, E) waveforms directly from time-domain data. D) ResBlocks for processing the latent vector. E) Schematics of the linear transform block F) Architecture of the Discriminator. Discriminator has two heads: first is predicting whether this is a real waveform or fake; second is predicting the class of this waveform. G) Schematics of the DisBlock.

signals (weaker earthquakes appear to be pulse-like, weather stronger ones are spread over the whole time window).

The approach of using conditional GAN has some limitations. In the case of images, it is rather easy to assess the quality of the image and decide whether the generation is of sufficient quality. In the case of seismic waveforms, this is much more challenging. If the generation does not obey physical laws - its application would be rather limited, and yet, this is something difficult to observe visually. There are no automatic metrics for generated seismic waveform quality as of now. One of the next steps in supporting this line of research would be to design a (physics informed) metric that measures how realistic generated seismic waveforms are.

4 Conclusions

We demonstrate that a generative model can be trained without using time-frequency representations based on raw waveforms. Using raw waveforms as input, we can generate realistic earthquake signals based on categorical labels. Based on the subjective evaluation of geophysics professionals, these generated signals are nearly identical to real ones. The future work will focus on improving the conditional capabilities of the model with a combination of multiple continuous and categorical variables, and developing a metric to quantitatively evaluate the realness of a generated seismic signal. A pre-trained model and the source code will be released as soon as possible.

Generative Adversarial Networks provide researchers with the ability to generate synthetic seismic data that can be used to model past or hypothetical events, as well as numerous other applications, mentioned in the text. Therefore, we hope that this work will provide a new approach to seismology.

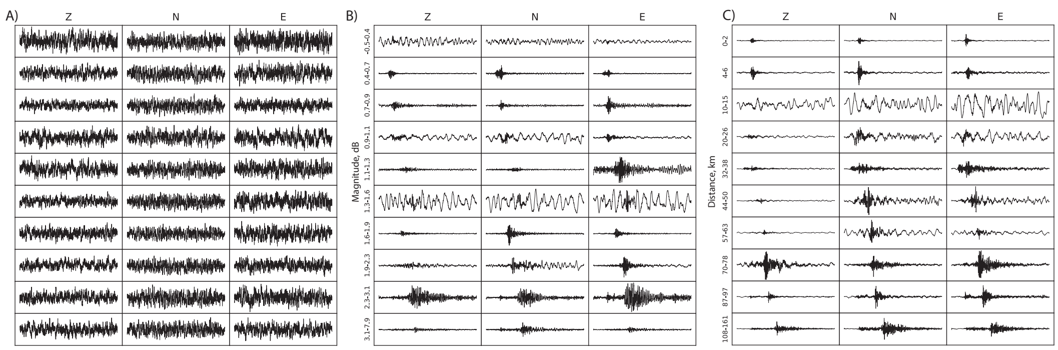

Figure 2: Samples generated with This Earthquake Doesn't Exist GAN. All 3 channels (Z-vertical and N, E - horizontal) are displayed. A) Generated seismic noise. B) Earthquake signals conditioned on the source magnitude (how strong was the earthquake). C)Earthquake signals conditioned on the source distance (how far away the earthquake is from the recorder).

References

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[2] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[3] S Mostafa Mousavi, Yixiao Sheng, Weiqiang Zhu, and Gregory C Beroza. Stanford earthquake dataset (stead): A global data set of seismic signals for ai. *IEEE Access*, 2019.

[4] Augustus Odena, Christopher Olah, and Jonathon Shlens. Conditional image synthesis with auxiliary classifier gans. In *International conference on machine learning*, pages 2642–2651. PMLR, 2017.

[5] Tiantong Wang, Daniel Trugman, and Youzuo Lin. Seismogen: Seismic waveform synthesis using gan with application to seismic data augmentation. *Journal of Geophysical Research: Solid Earth*, 126(4):e2020JB020077, 2021.