CS230

Final Report for CS230 2022 Unsupervised Volcanic Seismic Events Classification with Deep Clustering

Xing Tan *
Stanford University
tanx@stanford.edu

1 Introduction

The ever-increasingly dense seismometer array and advanced seismic event detection algorithms have allowed us to detect abundant imperceptible micro-seismic events before, especially in the scenario of volcanic seismic events monitoring. However, the unprecedented amount of volcanic seismic events detected heavily increased the burden of geologists and seismologists to classify those seismic events into specific categories. Therefore, we want to propose a semi-supervised or unsupervised deep learning algorithm to classify tens of thousands of volcanic seismic events detected in Volcanic regions. We especially want to apply our methods in Mt. Etna, Italy, from December 29, 2020 - June 30, 2021, when the volcano erupted. With hopefully accurate classification results, we will be able to map the time-spatial evolution of each volcanic seismic event. Thus better understanding the dynamic processes of Mt. Etna.

Criteria for volcano signals classification is still in controversy due to the imperfect knowledge of the geophysical mechanism of volcano activities. According to the mainstream classification criteria (1), volcanic seismic events generally could be divided into transient volcano-seismic signals and continuous signals. Based on the waveform appearance and signal frequency characteristic, transient signals could further divide into Volcanic-Tectonic earthquakes (VTE), Long-Period events (LP), Hybrid events (HB), Multi-Phase events (MP), Explosion Quakes (EQ), Collapse (COL), Regional Earthquake (REG), Icequake (ICE). Continuous signals could be divided into Volcanic Tremor (VT), Surface Process (SP), and Ambient Noise (AN).

2 Previous work

Clustering, partitioning data into groups of similar objects without relying on labels, is usually referred to as a subcategory of unsupervised learning methods. Methods including K-means are common practices of clustering based on distance metrics. However, it turns out to be less efficient to perform these methods directly as the dimensions of data begin to increase. Although principle component analysis can help with dimension reduction, PCA, as a linear algorithm, works poorly in extracting highly non-linear features. Recently, clustering methods based on deep learning gradually started to emerge, which leverage neural networks' capability of non-linear mapping to extract clustering-friendly latent features which largely facilitate further clustering.

For volcanic seismic events classification, several supervised learning or unsupervised methods have already been proposed to address the auto-classification of volcanic seismic events. Successfully supervised learning approaches used to tackle this problem include neural network (2) (3), convolution

*Department of Geophysics

deep neural network(4), and random forest (5). Also, the study shows that Transfer Learning can accelerate the convergence of the supervised model(6). The biggest problem for supervised methods is their reliance on a well-labeled dataset. However, most datasets are constructed based on recordings collected from one volcano while both temporal and spectral characteristics of different volcanoes might vary significantly. So models are hard to generalize in other datasets. Also, datasets suffer from severe class imbalance as the number of recordings of events, including HB, VT, can be one-tenth of LP and VTE.

And there are some successful unsupervised schemes, firstly extracting various physics features of the signal and then implementing different clustering-based classifiers (7) (8). However, there is still a vacancy for end-to-end unsupervised deep learning methods in this scenario. And the previous methods perform poorly in classifying events from rare events like ICE etc. So our approach mainly focuses on how to implement unsupervised or semi-supervised methods, and address the problem of class imbalance.

Our contributions are:

1. We proposed a hybrid network structure and training strategy of VAE and AE
2. We achieved the SOTA performance on dataset MicSigV1 with 100 $\%$ classification accuracy with a deep clustering method.
3. We are the first to perform successful 3-class classification on dataet MicSigV1

3 Dataset

MicSigV1(11) is one public volcanic seismic dataset from the ESeismic repository. This database was collected at the Cotopaxi volcano by the Geophysical Institute at Escuela Politécnica Nacional between January and June of 2010. It has a total of 1187 seismic waveform records from two different seismic stations (VC1 and BREF) with 5 different events recorded, including LP (1044 samples), VT (101 samples), HB (8 samples), RE (27 samples), and ICE (7 samples). The sampling frequency of signals recorded at station VC1 is 50 Hz, while 100 Hz at station BREF. The length of waveforms of each event varies from 20 seconds to 60 seconds, with most events under 40 seconds. The analysis was focused on three major types of seismic events: LP, VT, and REG events.

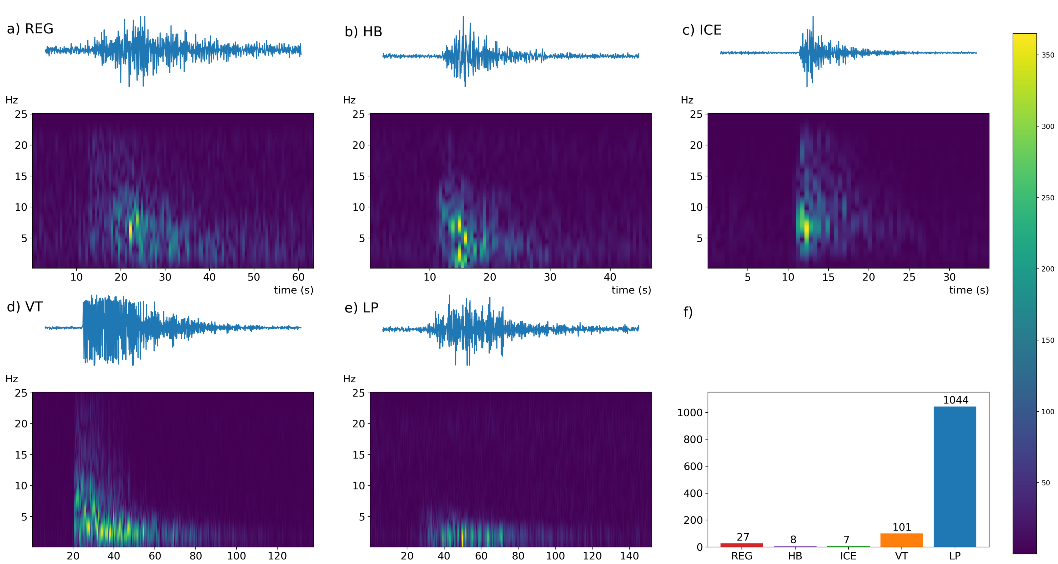

Figure 1: Sub-figures from a to e divided represent different volcanic seismic events in dataset MigSigV1. f shows the event number of 5 classes.

4 Methodology

4.1 Autoencoder

The autoencoder architecture is illustrated in figure 2. It consists of a contracting encoder (left side), a bottleneck block and an expansive decoder (right side). The contracting encoder follows a typical architecture of a convolutional network. It consists of the repeated blocks of 1 1x1 convolution that doubles the number of feature channels (except the first block increases the number 4 times), 2 3x3 convolutions (padded, stride 1) with residual skip connections, and max pooling operations with stride 2 for downsampling. A bottleneck block is followed by the encoder. The convolution hidden layers are flattened and embedded into two 20-dimensional vectors μ and σ through two fully connected layers. Assuming μ and σ represent 20-dimensional normal distribution characterized by $N(\mu, \sigma)$, we reparameterize the distribution as:

$$z = \mu + \zeta\sigma, \quad \zeta \sim N(0, 1) \tag{1}$$

In the expansion decoder, the 20-dimensional parameterization vector z will be further expanded into a 1024-dimensional vector and fed into the first block of the decoder. Each block of the decoder except the final block consists of one 1x1 convolution that halves the number of feature channels, two 3x3 convolutions (padded, stride 1) with residual skip connections, and one 2x2 transpose convolution. 1x1 convolution is used at the final layer to output the reconstructed spectrogram X' given the input spectrogram X.

The network's loss function is only composed of the reconstruction loss of the auto-encoder, which measures the mean square root of the deviation of the input and output:

$$L = MSE(X, X') \tag{2}$$

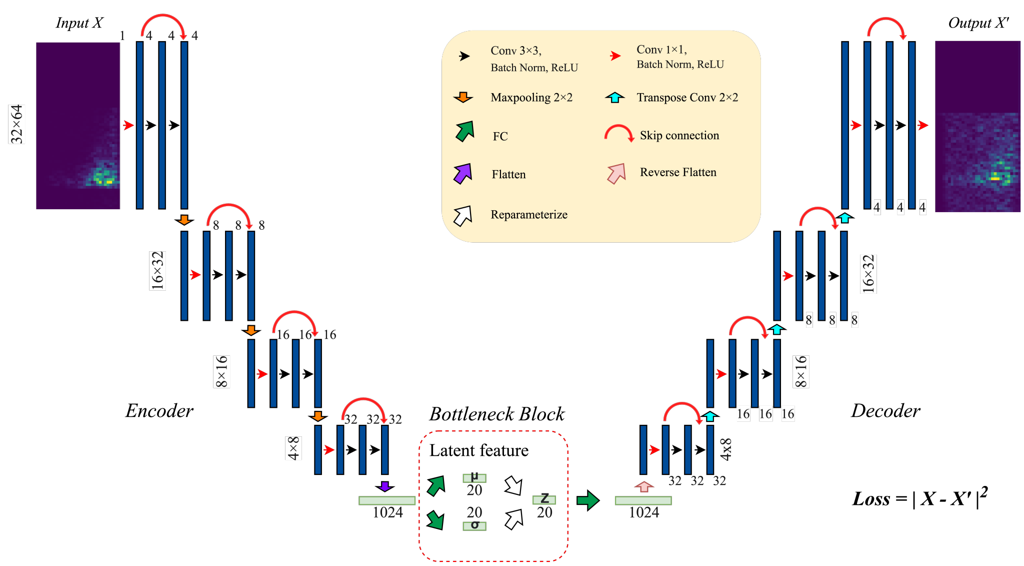

Figure 2: Autoencoder architecture. Each encoder block consists of 1x1, 3x3 convolution, max-pooling layer, and residual skip connection. Each decoder block consists of 1x1, 3x3 convolution, transpose convolution, and residual skip connection. Reconstruction loss is given in the figure.

4.2 Training Workflow

Training workflow is illustrated in the subfigure a) of figure 3. We first apply the random crop to the signal waveforms at training time to solve the class imbalance problem. REG events will be randomly cropped 50 times, VT events 20 times, and LP events three times to ensure reaching a relatively balanced dataset. All waveforms are further clipped or zero-padded into 42 seconds windows. Then all waveforms will be resampled to 50Hz, so each waveform consists of 2100 data points. Then we calculate the spectrogram of all waveforms and resize each waveform into a matrix with 32×64 size.

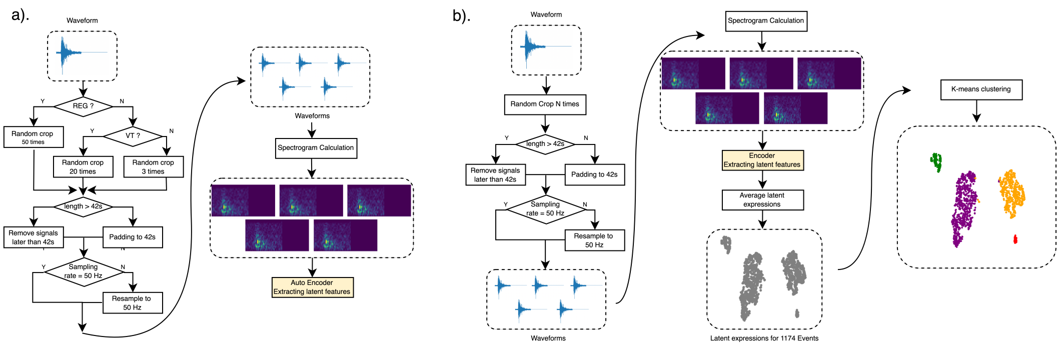

Figure 3: a). the training workflow. b). the predicting workflow

So far, we can establish the training dataset with 3978 spectrograms in total. The training dataset will be fed into the auto-encoder.

Following the workflow above, input spectrograms are 32x64 matrix. Next, we select the Adam optimizer to optimize the network and parameters, except the learning rate is set as default. The initial learning rate is 0.002 with the cosine annealing scheduler. The batches size is 32, with 100 epochs running. Experiments are conducted on Apple MacbookPro M1-Max using PyTorch deep learning framework.

4.3 Predicting Workflow

Predicting workflow is quite different from the training workflow and is illustrated in the subfigure b) of figure 3. At predicting time, the dataset for performing the final unsupervised classification differs from the training dataset. All 1172 events are equally random-cropped N times and then following the resizing, resampling, and spectrogram calculating processes. N turns out to be a crucial hyper-parameter that determines the performance, which is determined through experiments. So predicting dataset consists of $1172 \times N$ spectrograms and is only fed into the well-trained encoder to extract latent features. As each event is augmented N times, we assume that the latent features of one particular event should represent the same event no matter how augmented. So, for each event, we average the ten latent features of 10 augmentations and use its result, a 20-dimensional vector, as the final latent representation of the event. Then, we apply K-means clustering algorithms to all 1172 averaged latent features. The partition result given by the clustering algorithm can be mapped into the final classification.

5 Analysis

For clustering, we need to find the best match between the class labels and the clustering label assignment. We use clustering accuracy to evaluate the performance of the model is evaluated, which is defined as:

$$ACC = \max_{m} \frac{\sum_{i=1}^{n} 1\{y_i = m(c_i)\}}{n} \quad (3)$$

where c_i is the cluster assignment, y_i is the ground-truth label, and m is a mapping function that ranges over all possible one-to-one mappings between assignments and labels(9).

5.1 Hyper-parameter search: N

Introduced in 4.3, predicting workflow, all 1172 events are equally random-cropped N times, and we found N significantly influences the convergence of the model and the accuracy. The deep clustering prediction of different N is given in figure 4. When $N < 10$, the model struggles to distinguish the minority classes REG and VT. When N grows bigger, the boundaries of several clusters get much clear, and when $N = 10$, our model is able to cluster all 3 classes without fault. It is worth noting that LP events are partitioned into 2 sub-clusters, and we find this partition is very likely the result of the recordings coming from different stations. All signals are recorded by 2 seismic stations. And the

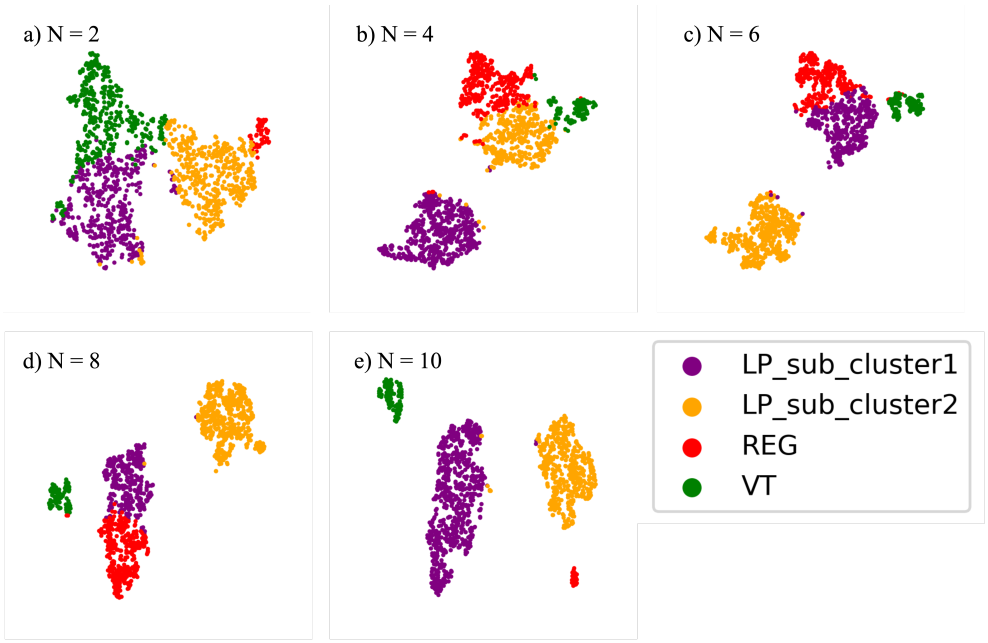

Figure 4: a) to e) Represent the t-SNE visualization of feature domain with hyperparameter N =2, 4, 6, 8, 10. 20-dimensional latent features are visualized via t-SNE into 2-dimensional vectors

Table 1: Comparison of different methods on MicSigV1 dataset. The best accuracy is highlighted in bold. The table was modified from a previous study(11)

Method	ACC
Ours	**100 (3 classes)**
Classification using DT and SVM	90 (2 classes)
VGG	94 (3 classes)
Filter-based feature selection with a Gaussian mixture classifier(12)	94.7 (2 classes)
Classification using a semi-supervised approach (10)	94.7 (2 classes)
Exhaustive comparison of 69 classifiers(13)	95 (2 classes)
Classification using an image-based descriptor(14)	96 (2 classes)
Feature selection and classification using kNN and DT(8)	99 (2 classes)

different subsurface structure and signal-to-nose ratio may result in the difference in the temporal and spectral characteristics.

5.2 Comparison with Other Methods

To our biggest surprise, our model achieved 100% accuracy of the three classes classification, which outperforms every unsupervised, semi-supervised, and supervised method tested on this dataset. Meanwhile, our model successfully performs 3 class classifications of LP, VT, and REG events. To our knowledge, all existing methods are binary classifiers focusing on LP and VT events. Also, our network is light-wise with 120k parameters compared with supervised methods.

Based on a previous study (11) of validating the performance of different methods on a MicSigV1 based on a partial dataset containing 436 seismic events, we can make the comparsion of the effectivness of deep learnong methods proposed recently based on evaluation metrics ACC , details are shown in table 1

5.3 Ablation Studies

The bottleneck layer architecture and the reparameterization strategy of our model is borrowed from the VAE. We conducted ablation studies of both the bottleneck block and the loss function to evaluate

Table 2: Ablation studies on bottleneck block and loss function

Method	ACC
Our method(bottleneck block, reconstruction loss)	**100**
Classic Auto-encoder (Fully connected layer, reconstruction loss)	97.0
Beta VAE(Bottleneck block, Construction loss + KL divergence loss)	< 95

the effectiveness of our model. And results are shown in 2. In this scenario, our architecture with bottleneck block and reparameterization strategy outperforms the classic auto-encoder model only with one fully connected layer instead. Also, we find that, in this problem, reconstruction loss exceeds the composed loss of reconstruction loss and Kullback–Leibler divergence loss (KL divergence loss), which is generally applied in VAE architecture.

The reason our proposed architecture is superior can be inferred from the following aspects. First, the bottleneck block and reparameterization can bring noise to the decoding process and force the model to learn the distribution of the dataset, which can increase the robustness of the model. For beta-VAE, KL divergence loss is necessary to ensure the distribution learned by the model characterized by μ and σ should be normal distributions with $\sigma = 1$. It successfully enables VAE to generate new samples from the dataset's existing data. However, our problem is extracting great latent features of each event, while we do not care how well our model can generate new samples. So reconstruction loss seems to be sufficient enough. Also, based on our experiment results, in the second half training process with beta-VAE, KL divergence loss will first drop and then gradually grow up. In contrast, reconstruction loss continues to decrease, which indicates KL divergence loss, while increasing the robustness, might partly scarify the model's encoder's performance.

6 Discussion and Conclusion

This project proposed a novel unsupervised method for volcanic seismic event classification and achieved state-of-the-art performance on the MicSigV1 dataset. We successfully achieve 100 $\%$ accuracy and are the first to perform successful 3-class-classification of VT, REG, and LP events on this dataset.

However, we have to point out that our method is based on data augmentation, which uses the number of each type of event. But for unsupervised learning, the technique should get rid of this information. The reason why we apply the data augmentation strategy is that we find the dataset is highly imbalanced. The model tends to view those minority events as noise that leads to terrible performance, even though unsupervised learning should relieve considerable class imbalance. To tackle this problem, we propose dynamically updating the dataset during the training process. Assuming the minority events have a relatively larger reconstructive loss, we measure the reconstruction loss of all events and resample all events with greater weight on those with more significant loss and smaller weight on those with a minor loss. But the performance remains to be poor.

Also, we originally want to more advanced deep clustering methods with better architecture and clustering loss. But it seems that our model seems to be sufficient enough to perform well in this dataset. But it definitely worth a try on other more complex scenarios.

References

[1] Bormann, Peter. "New manual of seismological observatory practice." (2002).

[2] Lara, Fernando, et al. "A deep learning approach for automatic recognition of seismo-volcanilong-period the Cotopaxi volcano." Journal of Volcanology and Geothermal Research 409 (2021): 107142.

[3] Titos, Manuel, et al. "A deep neural networks approach to automatic recognition systems for volcano-seismic events." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11.5 (2018): 1533-1544.

[4] Salazar, Aaron, et al. "Deep-learning for volcanic seismic events classification." 2020 IEEE Colombian Conference on Applications of Computational Intelligence (IEEE ColCACI 2020). IEEE, 2020.

[5] Rodgers, Mel, et al. "Waveform classification and statistical analysis of seismic precursors to the July 2008 Vulcanian Eruption of Soufrière Hills Volcano, Montserrat." EGU General Assembly Conference Abstracts. 2016.

[6] Titos, Manuel, et al. "Classification of isolated volcano-seismic events based on inductive transfer learning." IEEE Geoscience and Remote Sensing Letters 17.5 (2019): 869-873.

[7] Duque, Adrian, et al. "Exploring the unsupervised classification of seismic events of Cotopaxi volcano." Journal of Volcanology and Geothermal Research 403 (2020): 107009.

[8] Lara-Cueva, R.A., Benítez, D.S., Carrera, E.V., Ruiz, M. and Rojo-Álvarez, J.L., 2016. Feature selection of seismic waveforms for long period event detection at Cotopaxi Volcano. Journal of Volcanology and Geothermal Research, 316, pp.34-49.

[9] S. M. Mousavi, W. Zhu, W. Ellsworth and G. Beroza, "Unsupervised Clustering of Seismic Signals Using Deep Convolutional Autoencoders," in IEEE Geoscience and Remote Sensing Letters, vol. 16, no. 11, pp. 1693-1697, Nov. 2019, doi: 10.1109/LGRS.2019.2909218.

[10] Brusil, C., Grijalva, F., Lara-Cueva, R., Ruiz, M. and Acuña, B., 2019, November. A semi-supervised approach for microseisms classification from Cotopaxi volcano. In 2019 IEEE Latin American Conference on Computational Intelligence (LA-CCI) (pp. 1-6). IEEE.

[11] Pérez, Noel, et al. "ESeismic: Towards an Ecuadorian volcano seismic repository." Journal of Volcanology and Geothermal Research 396 (2020): 106855.

[12] Venegas, P., Pérez, N., Benítez, D., Lara-Cueva, R. and Ruiz, M., 2019. I am combining filter-based feature selection methods and gaussian mixture model for the classification of seismic events from cotopaxi volcano. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(6), pp.1991-2003.

[13] Venegas, P., Pèrez, N., Benítez, D.S., Lara-Cueva, R. and Ruiz, M., 2019, November. Building machine learning models for long-period and volcano-tectonic event classification. In 2019 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON) (pp. 1-6). IEEE.

[14] Perez, N., Venegas, P., Benitez, D., Lara-Cueva, R. and Ruiz, M., 2020. A new volcanic seismic signal descriptor and its application to a data set from the cotopaxi volcano. IEEE Transactions on Geoscience and Remote Sensing, 58(9), pp.6493-6503.