# **DAS Anomaly Detection**

***Problem Statement***

(DAS is Distributed Acoustic Sensing data)

Detect and localize anomalies in spatial and temporal domain using the gather plots.

We have around 15k images of the gather plots that are generated from different segy files and for different channel ranges. The images are all unlabeled, I just use the “Obspy” python library to read the segy files and specify the sampling rate and channel range to generate the following images.

*The flow for anomaly detection is as follows:*

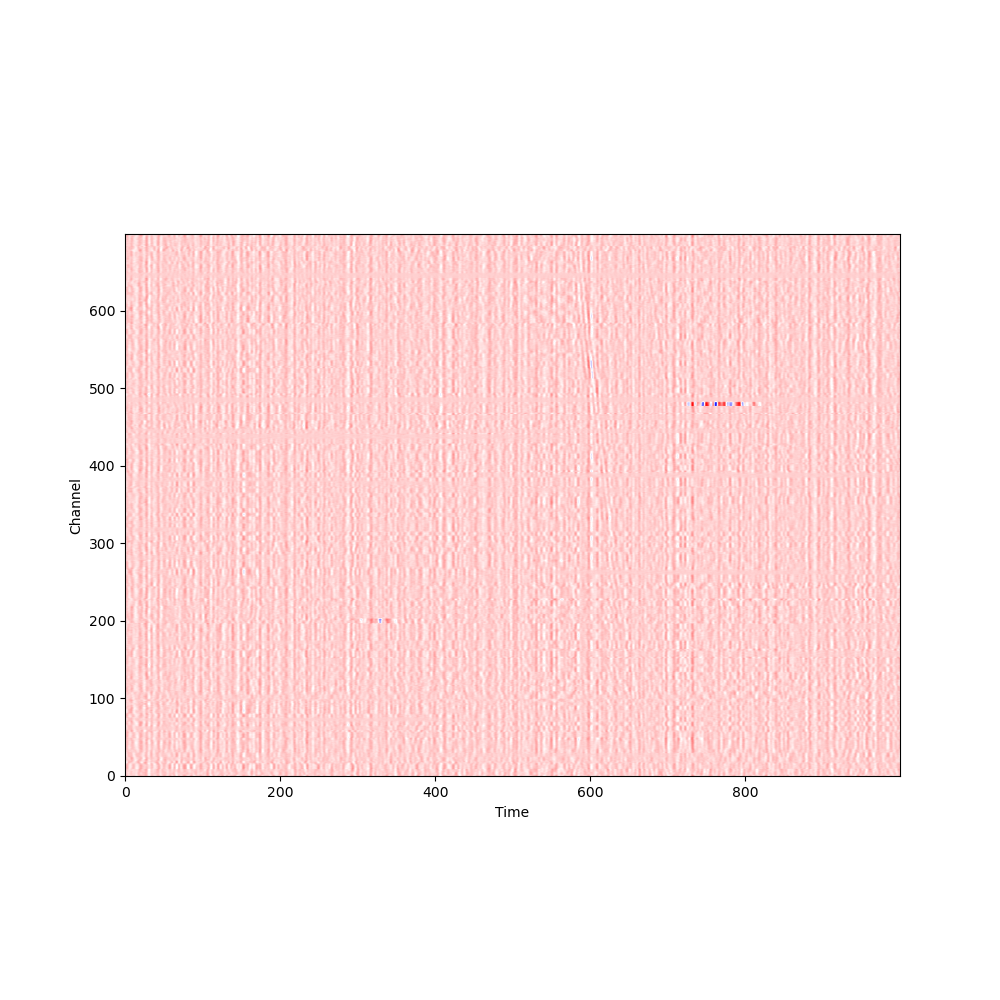
**Step 1.** Generate the training images using the Obspy Module

**Step 2.** Use Image DC Clustering Algorithm to get different clusters of the plots and segregate the normal and abnormal images.

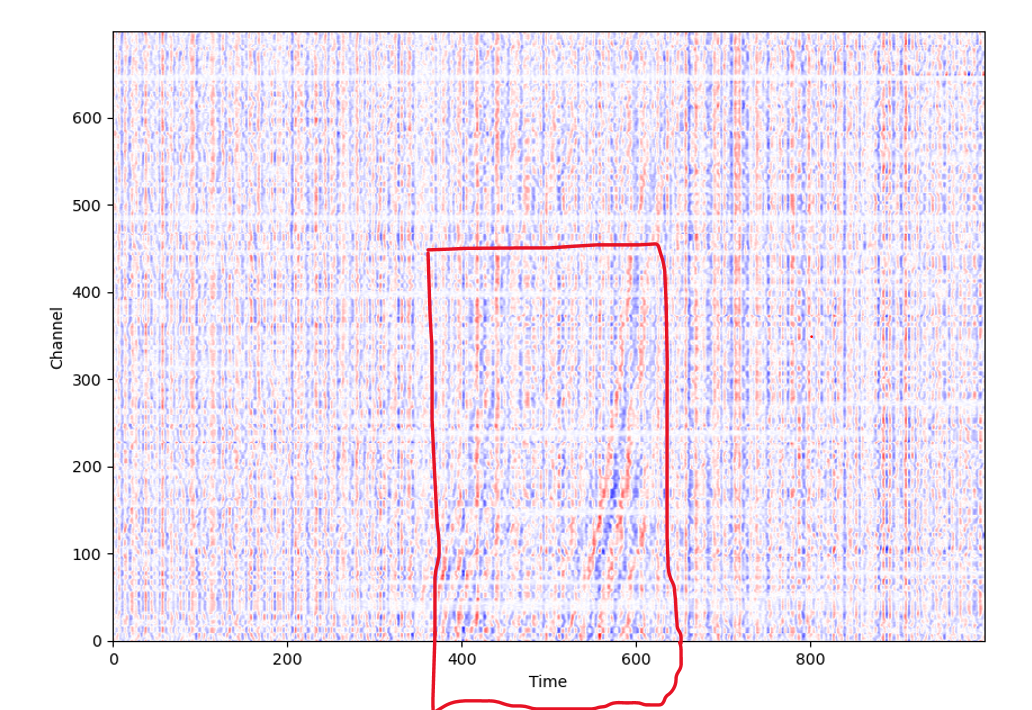
**Step 3.** Train a CNN based Auto Encoder Model to detect anomalies in images using reconstruction error and kernel density estimation threshold.

**Step 4.** Localize the anomalies using image segmentation/deep learning model.

1. The image below is normal image without any anomaly.



1. The image below is considered as an anomaly and the highlighted region is the one with anomalous behaviour.

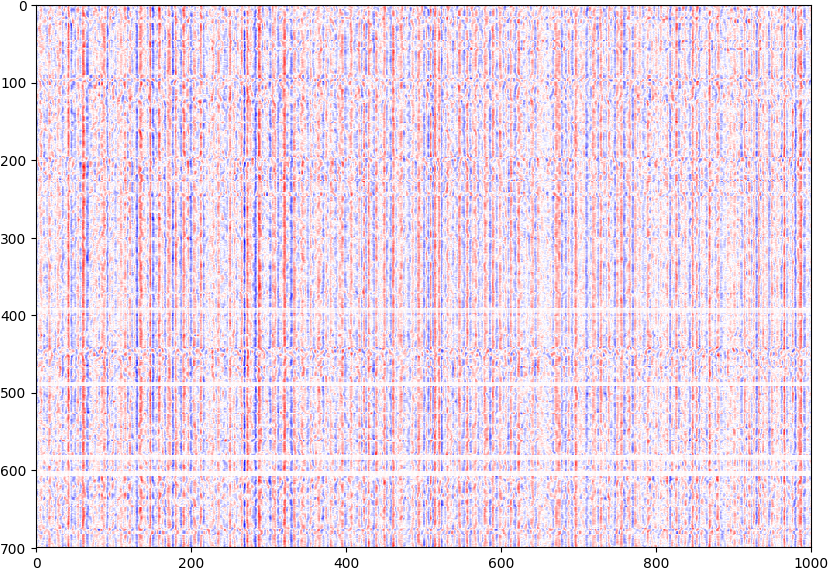


localizing the anomalies!!

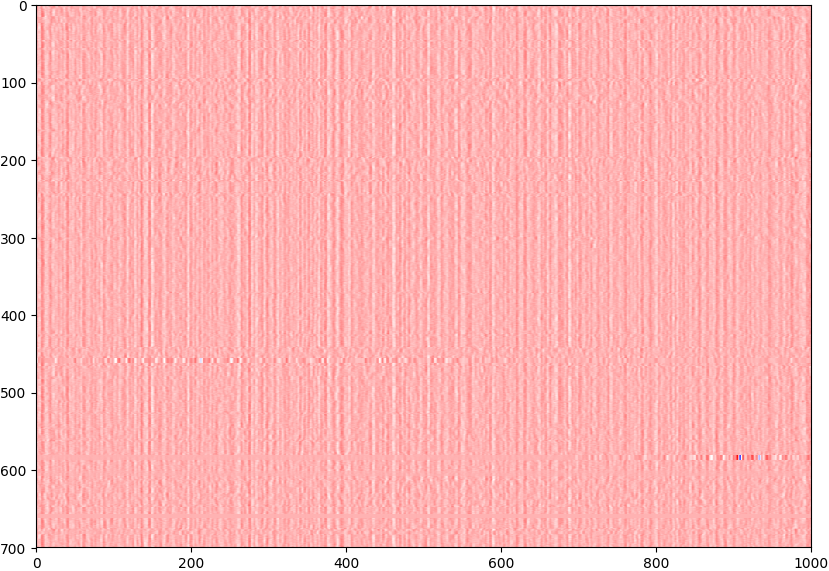
Currently I have used CNN based Auto Encoder for anomaly detection. I have trained the model on normal images and when we have any image that has any kind of outlier, the model predicts it as an anomaly based on the reconstruction error. But once we classify the image as an anomaly, I need to localize/highlight/segment the part that has anomaly and need to get the channel no. and time duration in which that anomaly occurred. I am not sure what kind of algorithm or deep learning model should be used for this task.

**Images from different clusters**

**Cluster 0**



**Cluster 1**



## Data Description

DAS : Distributed Acoustic Sensing data can be acquired from University of Wisconsin Porotomo Ftp server ftp://roftp.ssec.wisc.edu/porotomo/PoroTomo2/DATA/DASH/20160321/

Specifically, we are interested in detecting the earthquake event located in this dataset

<ftp://roftp.ssec.wisc.edu/porotomo/PoroTomo2/DATA/DASH/20160321/PoroTomo_iDAS16043_160321000721.sgy>

Each dataset includes data recorded in 30 sec

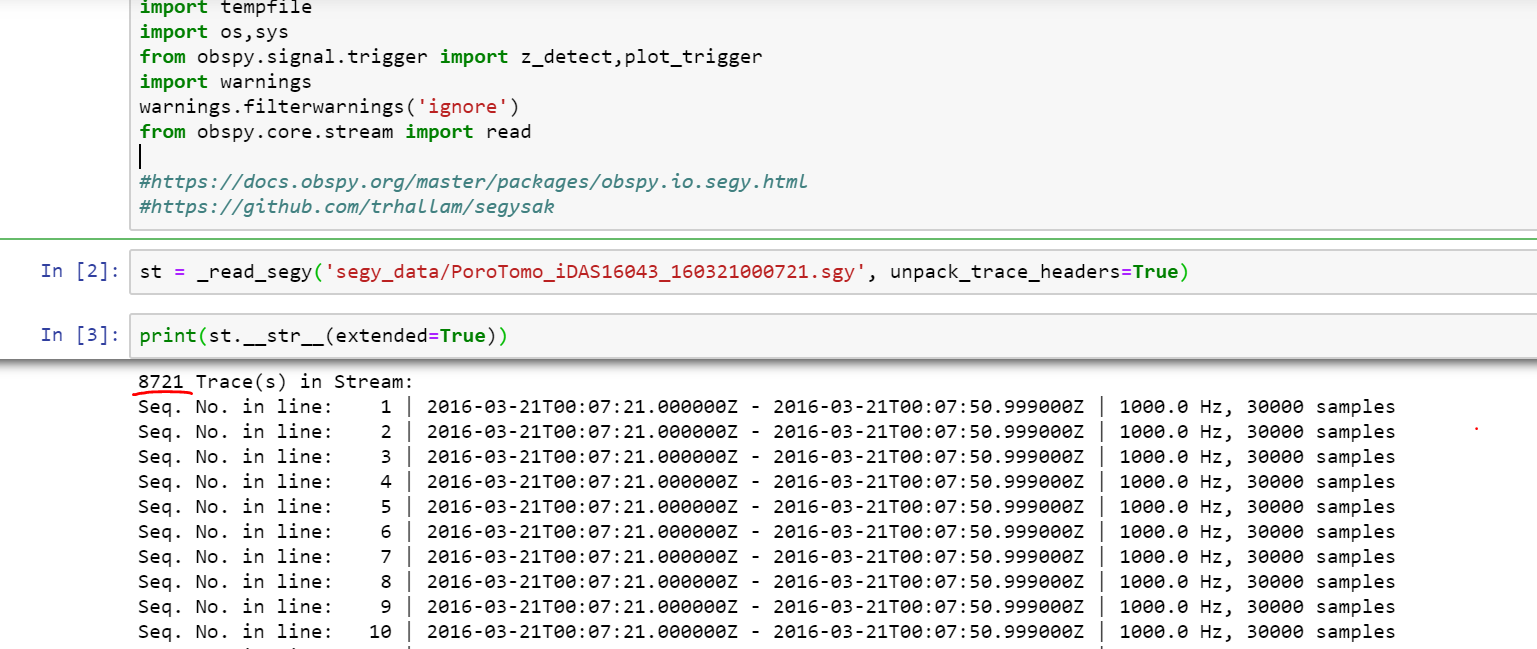
## Data Format and Data Processing

DAS data is usually given in SEGY format.

Python “Obspy” library is used to read the “segy” files and generate different kind of plots such as Spectrogram, CWT (Continuous Wavelet Transform) and gather plots.

The following screenshot shows how we read the “segy” files using Obspy module in python. DAS data is mostly seismic data that is recorded using sensors and is used for earthquake onset detection.

We can observe that there are “8721 Traces” in the obspy stream object. It means that there are “8721” different channels (sensors).

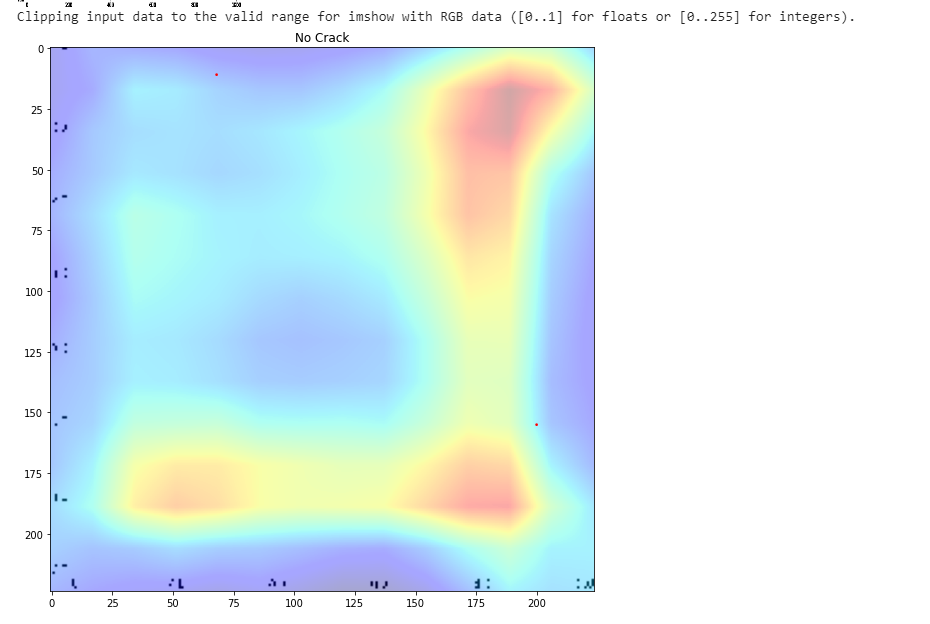


## ML Task

The idea is to cut the DAS stream into small image frames, train a deep learner (CNN or ConvLSTM) on base line data so that it can be used to identify anomalies. For example, this can be formulated as supervised classification problem.

**Sample Image**

(gather plot having samples from channels starting from 100-300)

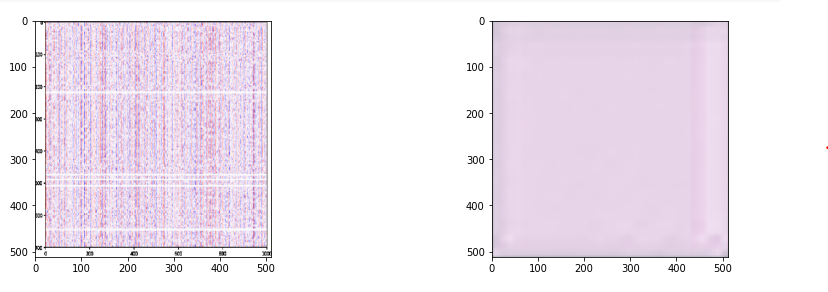


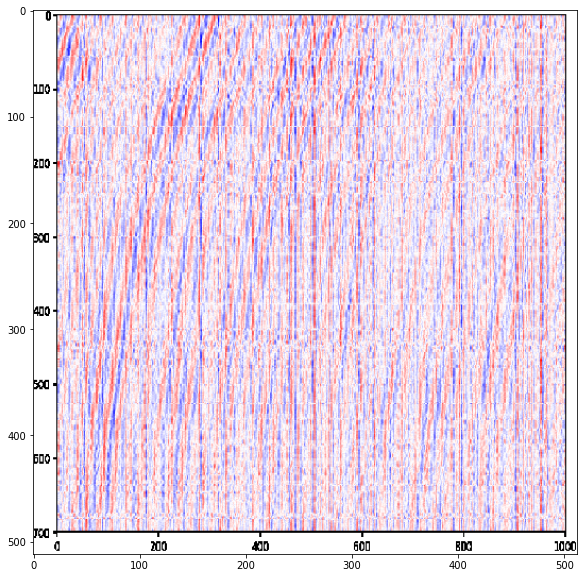
Model Training for Unsupervised Anomaly Detection and Localization

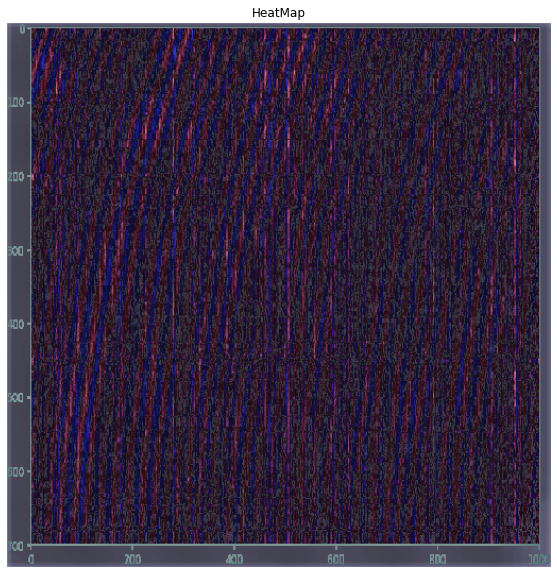
1. **Image Clustering**

Perform image vectorization to extract features using pre-trained model in Keras (Functional API) ie VGG16 and RESNET50 and then pass these vectors to K-Means for Clustering to generate 4 diff clusters.

1. Anomaly Detection and Localization using CNNAE







1. Assigning labels to images to segregate the images into anomaly and normal data. Build a machine learning to detect the abnormal image and localize the anomaly into it. Use VGG19/16 model to extract the image features and train the last few layers of the model architecture to specialize in the given classification task. *block5\_conv3*

