ANC Project Report 1 - Linus

Motivation

To find clusters using dimension reduction techniques given the data collected and provided.

Data downloaded and overview

The downloaded files from One-Drive that was provided

was stored in the following path:

```
smb://10.97.31.159/SNTL_2/Linus/ANC-Project/ANC-Data/
```

Data preprocessing

The preprocessing of the provided data can be broken into 5 steps:

1. Reading
2. Converting
3. Standardizing
4. Labelling
5. Exporting

Reading .asc files and converting to numpy arrays (.npy)

```
# convert asc files to npy files
path = '../ANC-Data/Spectrogram-A/Spectrogram (A)/'
npy_dest_path = '../ANC-Data-NPY/Spectrogram-A/'
files = glob.glob(path + '*')

for i in files:
    with codecs.open(i, encoding='utf-8-sig') as f:
        X = np.loadtxt(f, usecols=(0,1))
        np.save(npy_dest_path + i.split('/')[-1][:-3] + 'npy', X)
```

The converted and exported npy files are stored in the following path:

```
smb://10.97.31.159/SNTL_2/Linus/ANC-Project/ANC-Data-NPY/
```

Standardization

Labelling

Exporting

The exported pickle and csv files are stored in the following path:

```
smb://10.97.31.159/SNTL_2/Linus/ANC_Project/Scripts/
```

Overview of dataframe

Dataframe:

- 918 entries (rows)
  - 700 Planes
  - Planes (Raining)
  - Planes (High altitude)
  - Helicopters
  - Rumble
- 9 features (columns)
  - Class
    - Planes
    - Planes_Raining
    - Planes_HighAlt
    - Helicopter
    - Rumble
  - FFT-A
  - FFT-Z
  - SPL-A
  - SPL-Z
  - 3rd-Octave-A
  - 3rd-Octave-Z
  - Spectrogram-A
  - Spectrogram-Z

Figure below shows a snippet of the exported dataframe.

Clustering

Several clustering techniques have been experimented. The techniques include dimension reduction techniques like PCA and TSNE, and others, like using variational autoencoder (VAE) for the deep model to learn and cluster with the latent space.

PCA

TSNE

PCA with TSNE

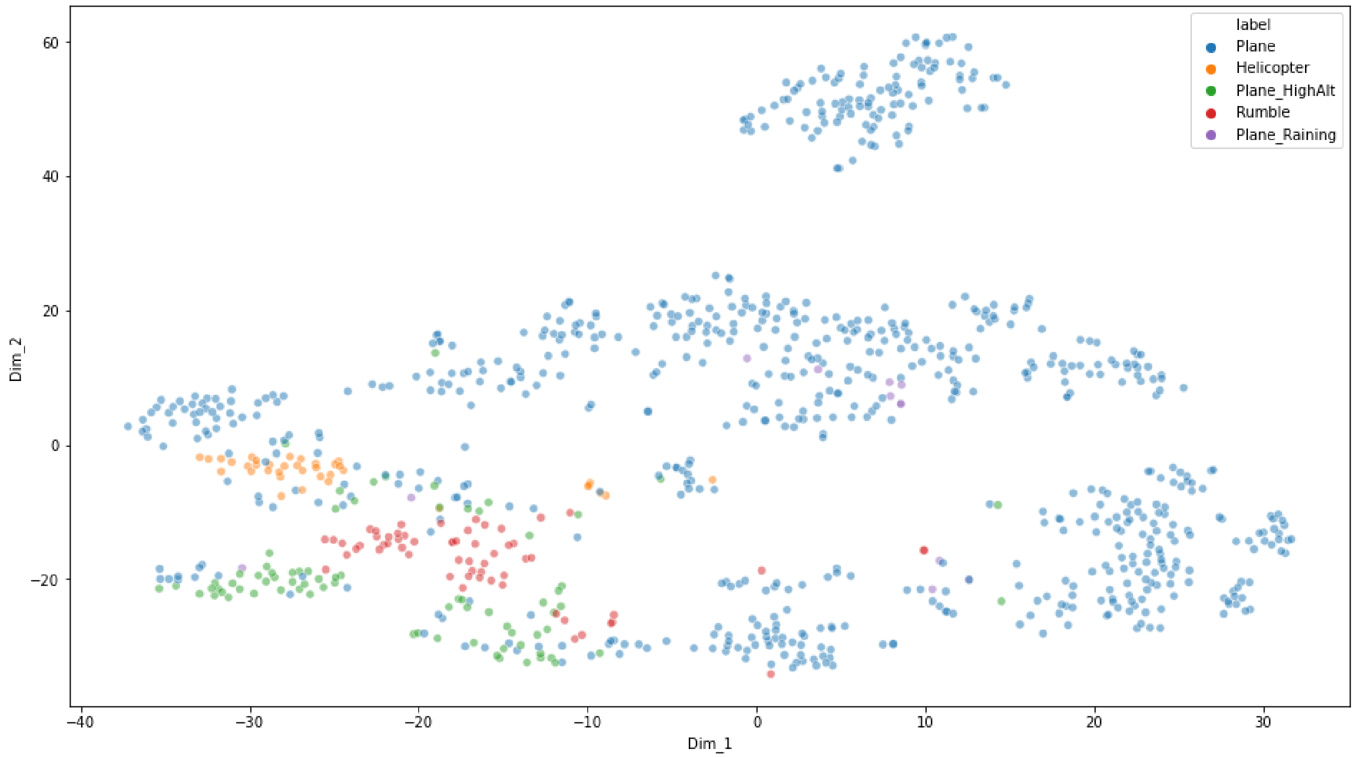
The idea between using both PCA and TSNE to PCA here takes in the high dimensional features of say the FFT-A with 918 entries of shape 2000 bins values and calculates the primary components

In this experiment, we took 50 principal components for the features

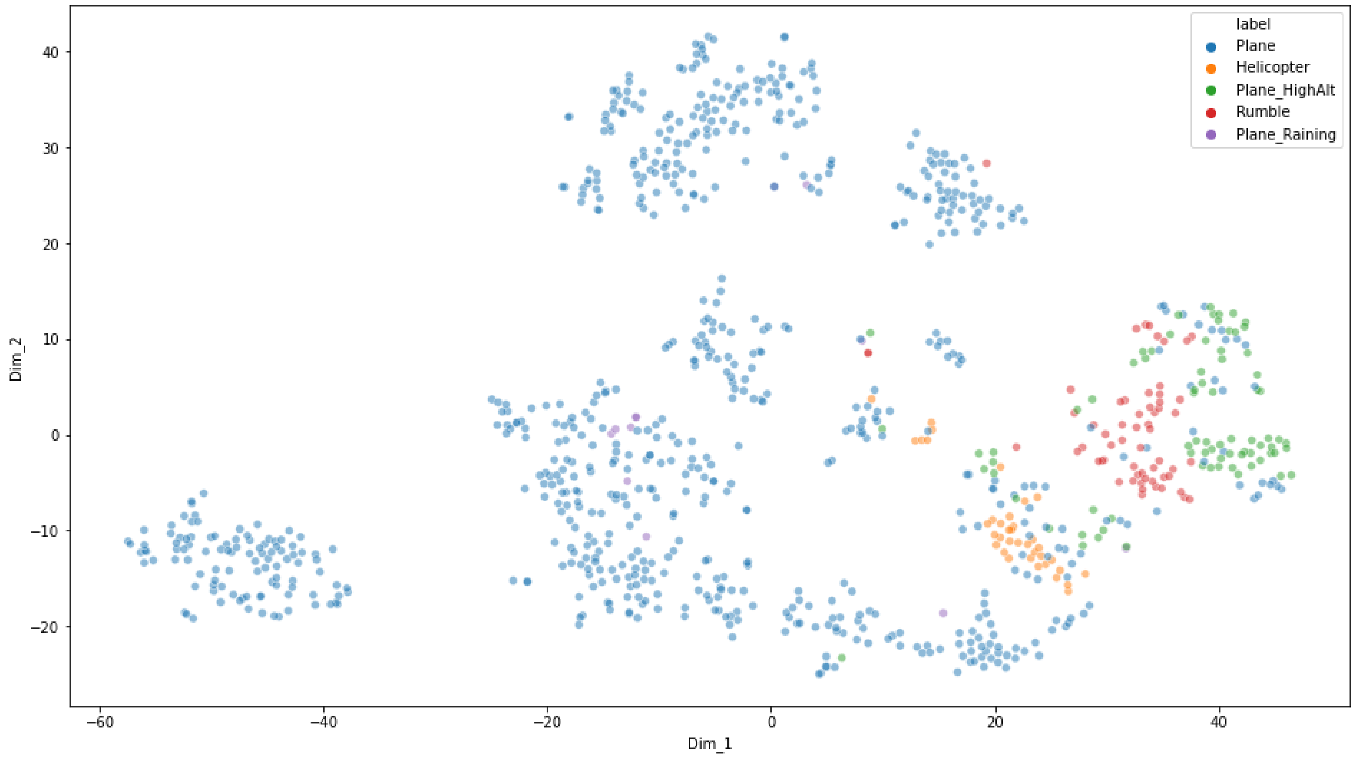
Then we further reduce the 50 principal components with TSNE into 2 dimensions for visualization purpose.

Single feature

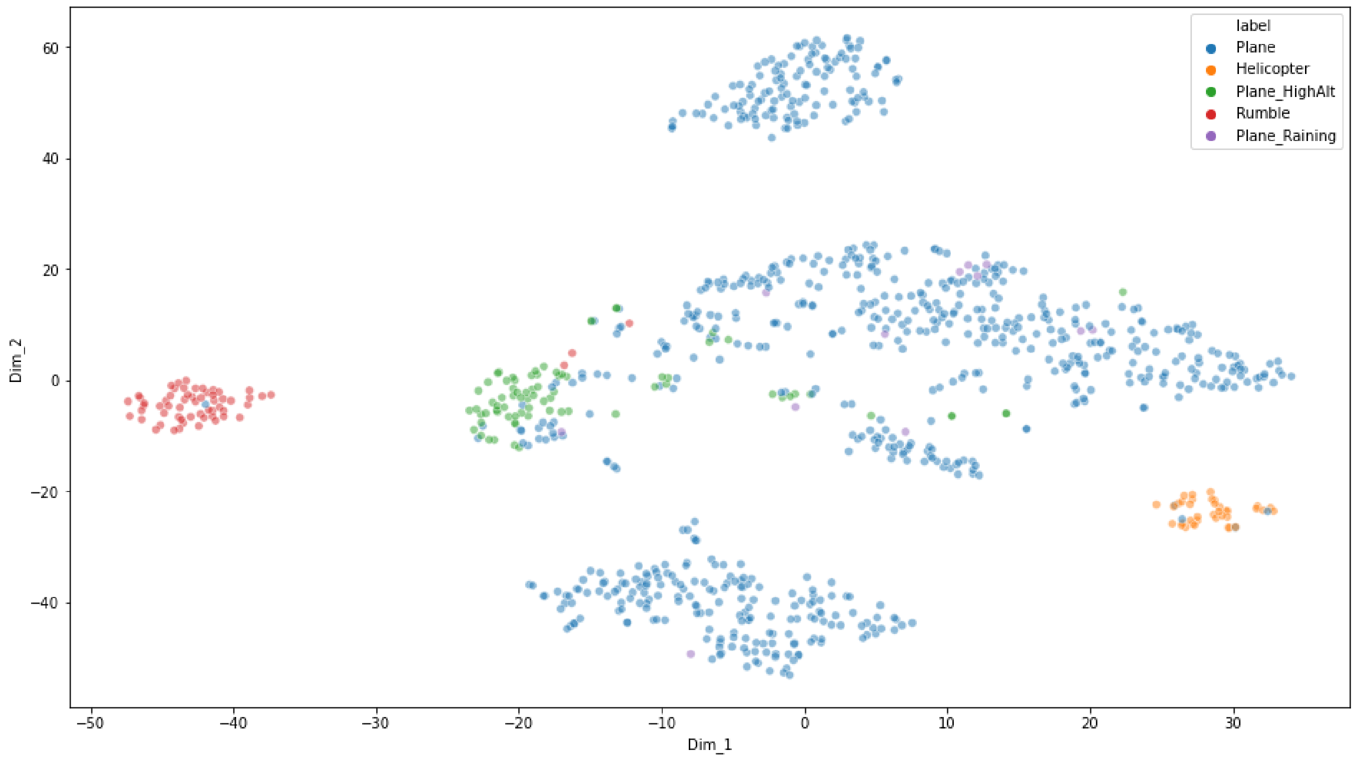
FFT-A



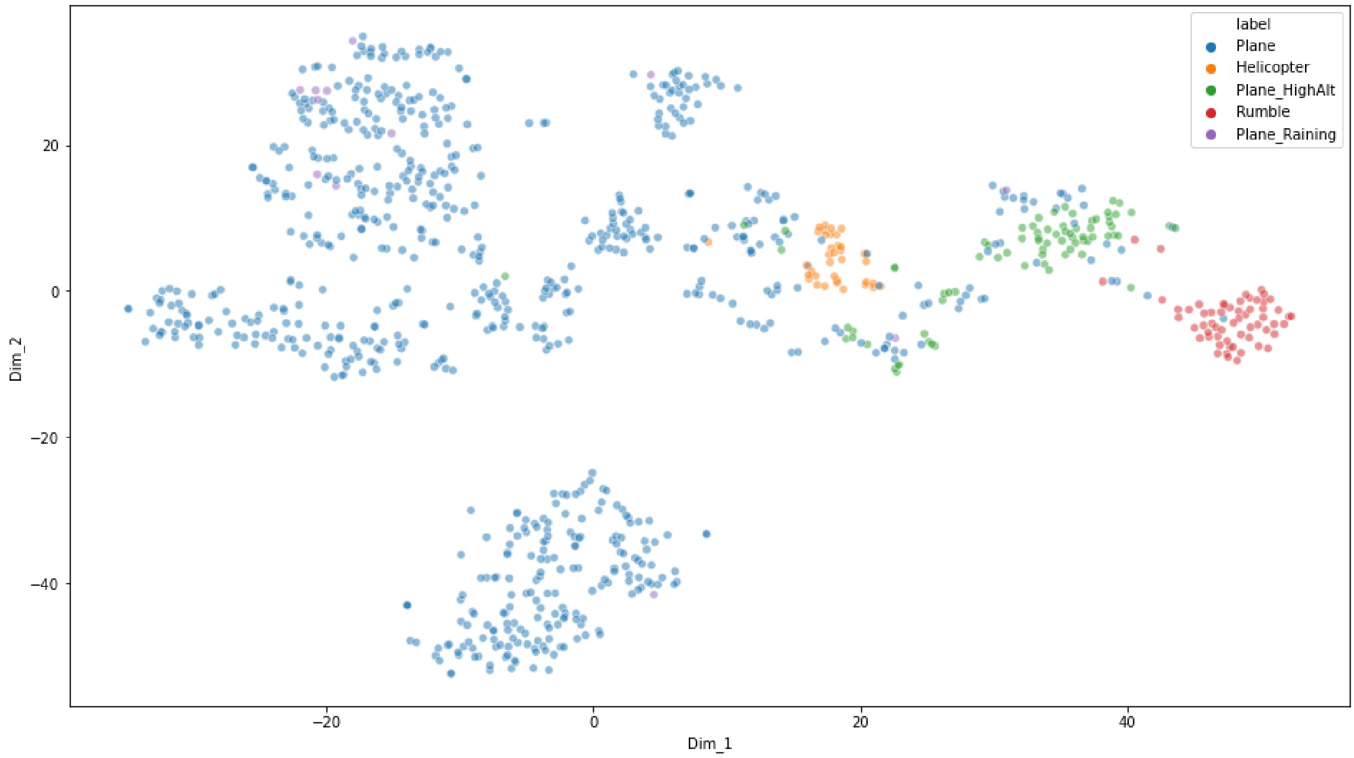
FFT-Z



3rd-Octave-A

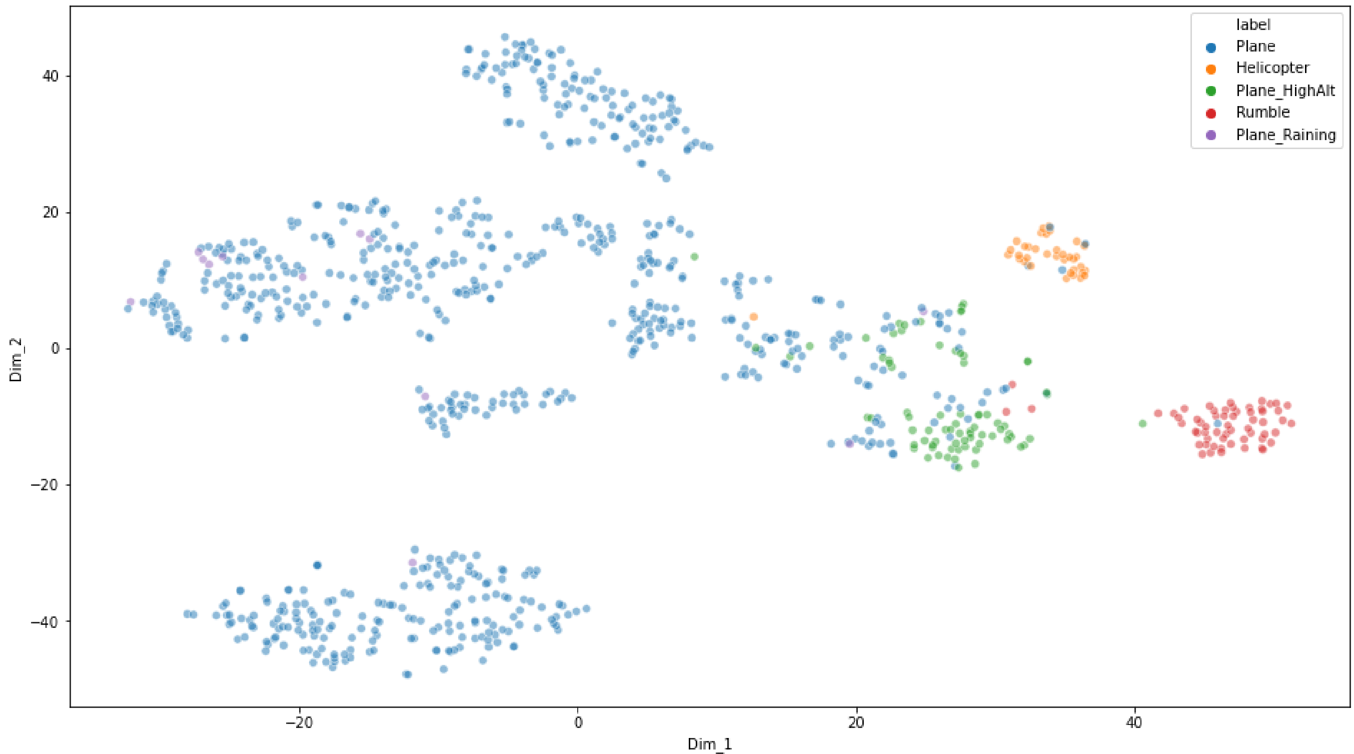


3rd-Octave-Z



Combined features

3rd-Octave(A) and 3rd-Octvate(Z)



Conclusion

For both FFT(A, Z) and 3rd-Octave(A, Z) we can observe that there are clusters being formed given the.

The 3rd-Octave(A, Z) features seems to provide the best clustering as since from the above figures.

Future work

Experiment on employing variational autoencoder (VAE) for deep model to learn and cluster with the latent space by employing reconstruction loss for optimization.