Deep Embedded Clustering on Seismic Data

Group 23
Heather Huntley, John O’boyle, Cameron Lewis

Background

- There are massive amounts of unlabeled seismic data that are not being used, and that is accumulating over time.
- Using unsupervised machine learning methods and dimensionality reduction to compare the performance of different clustering methods.

Preferable ML Approach

- Can be done manually
- ML can analyse small sections in higher detail via NN models.
- We want to explore different edits to the existing DEC model
  - Type of Clustering
  - Encoder architecture

Literature Survey

manuscript submitted to *JGR: Solid Earth*

Unsupervised Deep Clustering of Seismic Data: Monitoring the Ross Ice Shelf, Antarctica

William F. Jenkins II[1], Peter Gerstoft[1], Michael J. Bianco[1], Peter D. Bromirski[1]

[1]Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA

[1]

- DEC in seismic data
  - Paper that was foundation for our experiment
  - Original paper found dimensionality reduction produced better clustering
- UNET with skip
  - More accurate reconstruction

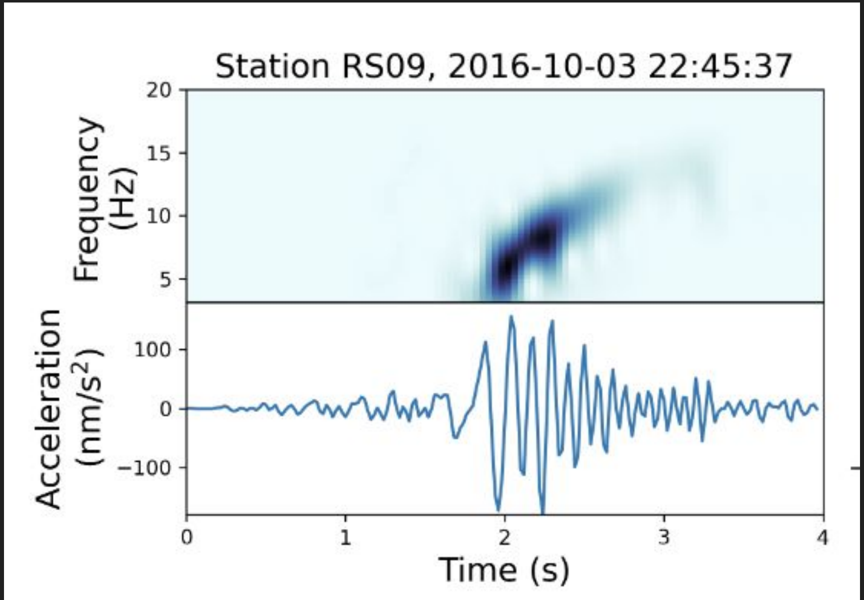
Literature Survey

- PCA in Chile Earthquake Prediction
  - Combined with other classification networks (artificial neural networks, classification trees, and random forest)
- Hierarchical clustering using in Earthquake magnitude prediction
  - Unlike k-means it is not sensitive to initial seeding or outliers
  - Comes at the cost of increased computational cost

Data

Seismic Data Recorded on the Ross Ice Shelf from 2014-2017

- Seismic data gathered via seismology auto-detection algorithms

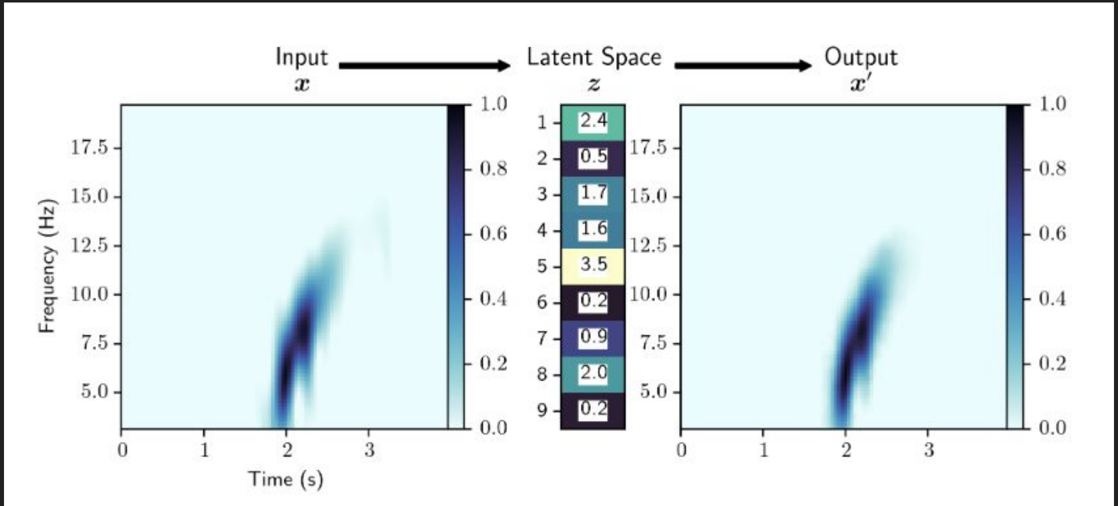


[1]

Feature extraction

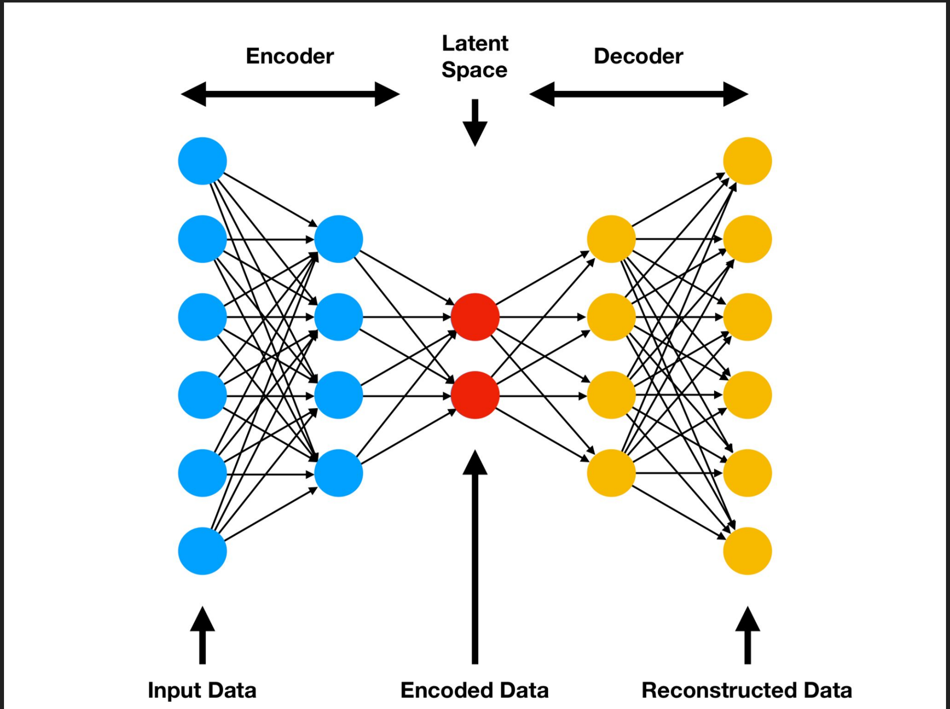
- Time-series data → Fourier Transform → Spectrogram
- CNN autoencoders are known to run well on images
- Dimensions of each spectrogram:
  - 1x87x100

[1]

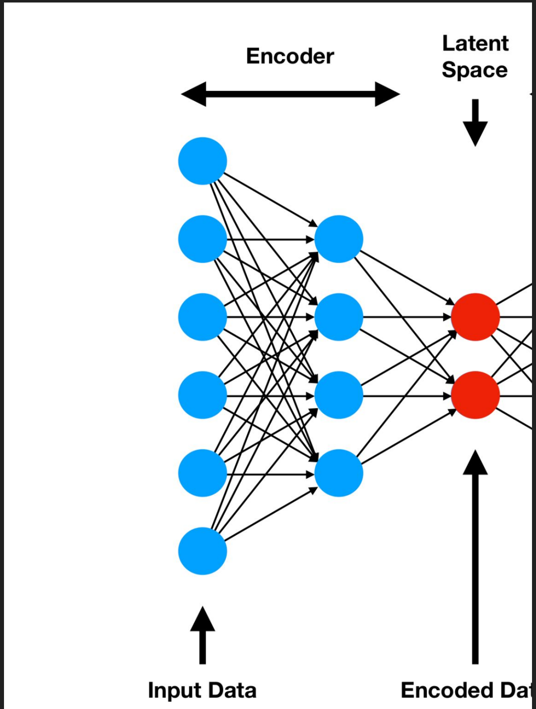


Pretraining the model - Autoencoder

- Autoencoder & Clustering component. DEC → GMM

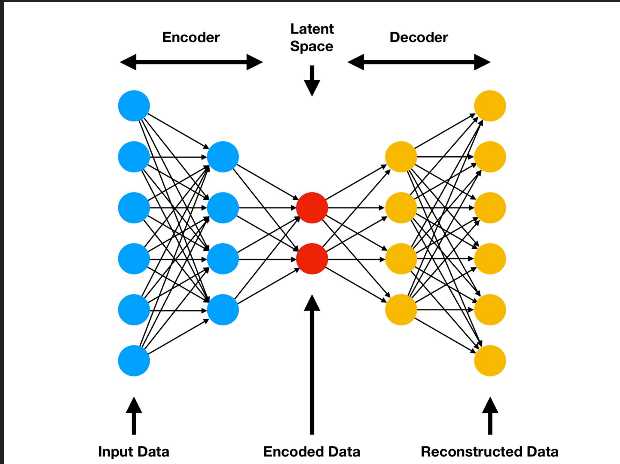


[4]

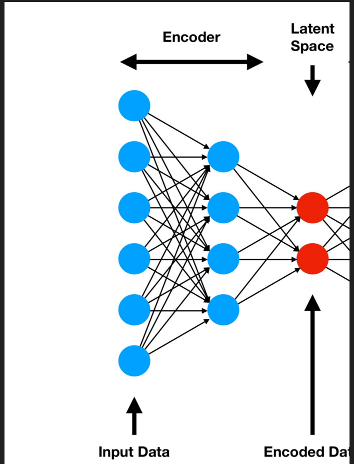


Pretraining the model - Autoencoder

- 5x conv-relu - ENCODER
- Flatten-linear-relu - LATENT SPACE
- 5x convTranspose-relu - DECODER
- 10 epochs, lr =0.001
- MSE Loss

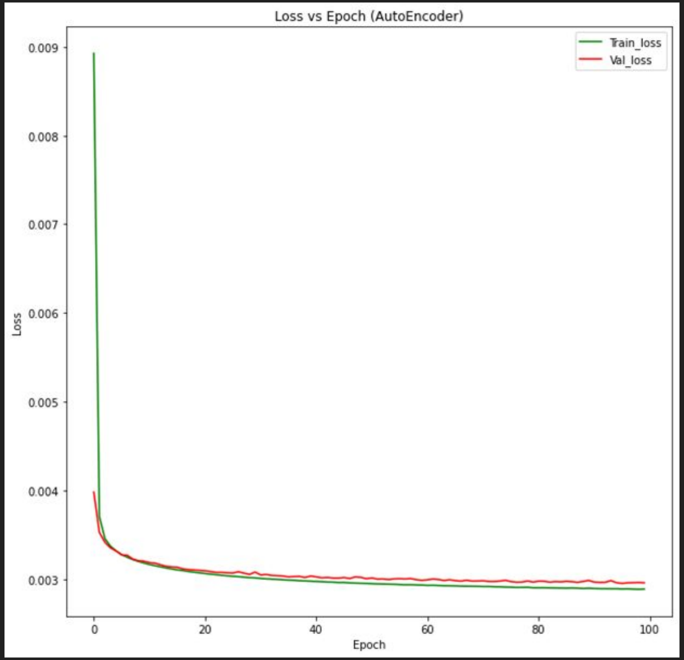


- Sklearn GMM is run on the latent space features
- Dataset → latent space dataset
- Selection of 8 clusters



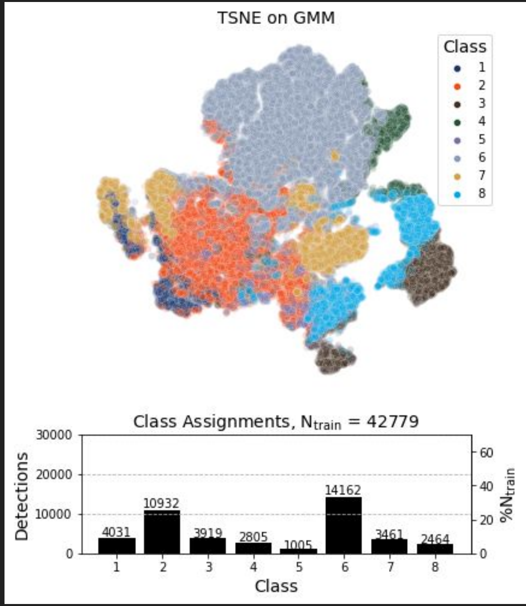
Results: AutoEncoder Training

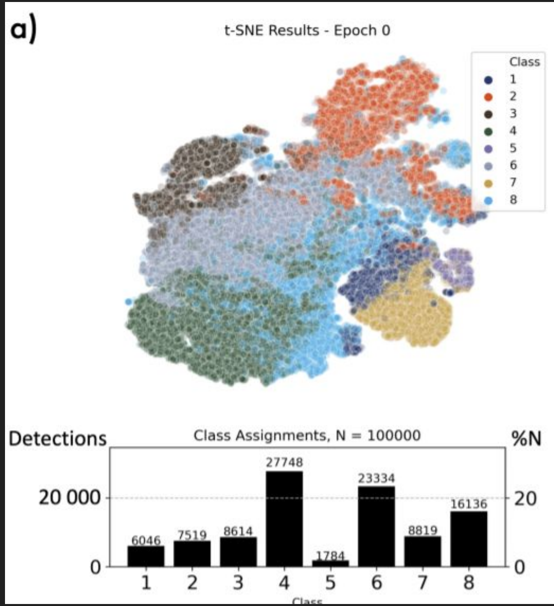
- Loss on initial training seemed initially seemed misleading, resolved with proper weight initialization
- Experimented with hyperparameters
- Found more reasonable results

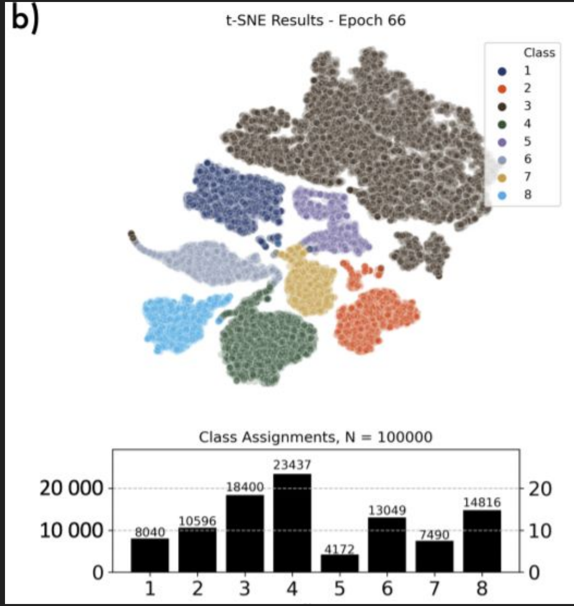


Results: Clustering

- DEC more separated clusters







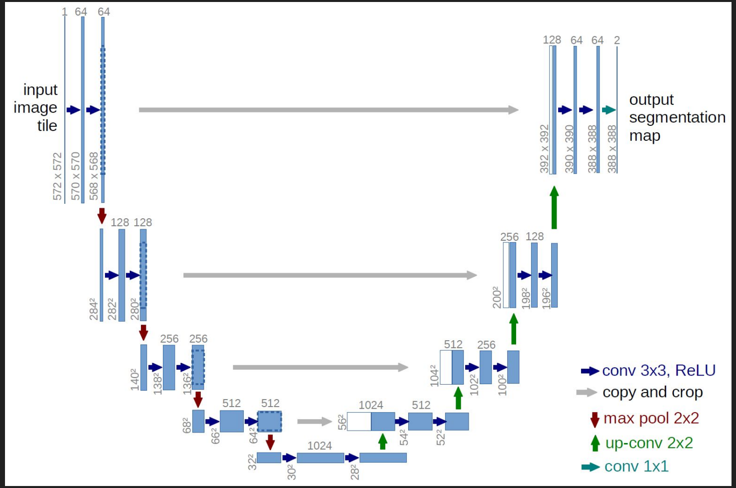
Ours

DEC GMM init [1]

DEC Trained

Further items to be completed before final report submission

- Rigorous early stop loss for autoencoder training
- Replace autoencoder network with U-net architecture
- Compare clustering results using latent space from U-Net vs latent space from AE



[5]

References

1) DEC (Jenkins et., al 2020) https://www.essoar.org/pdfjs/10.1002/essoar.10505894.1

2) DEC (Xie et., al 2016) https://arxiv.org/pdf/1511.06335.pdf

3) PCA https://www.researchgate.net/publication/274700549_Improving_Earthquake_Prediction_with_Principal_Component_Analysis_Application_to_Chile

4) GlowCrust (Trugman and Shearer 2017) https://igppweb.ucsd.edu/~shearer/mahi/PDF/2017/Trugman_growclust_2017.pdf

5) U-Net (Ronneberger et., al 2016): https://arxiv.org/pdf/1505.04597.pdf

6) AutoEncoder Images https://www.compthree.com/blog/autoencoder/