Deep Learning for Seismic Data Classification in Full-Waveform Inversion

Congyue Cui[1], Chao Song[1]

1. Department of Geosciences, Princeton University, Princeton, NJ, 08540, USA
E-mail: ccui@princeton.edu chaosong@princeton.edu

COS424 2019 Final Project

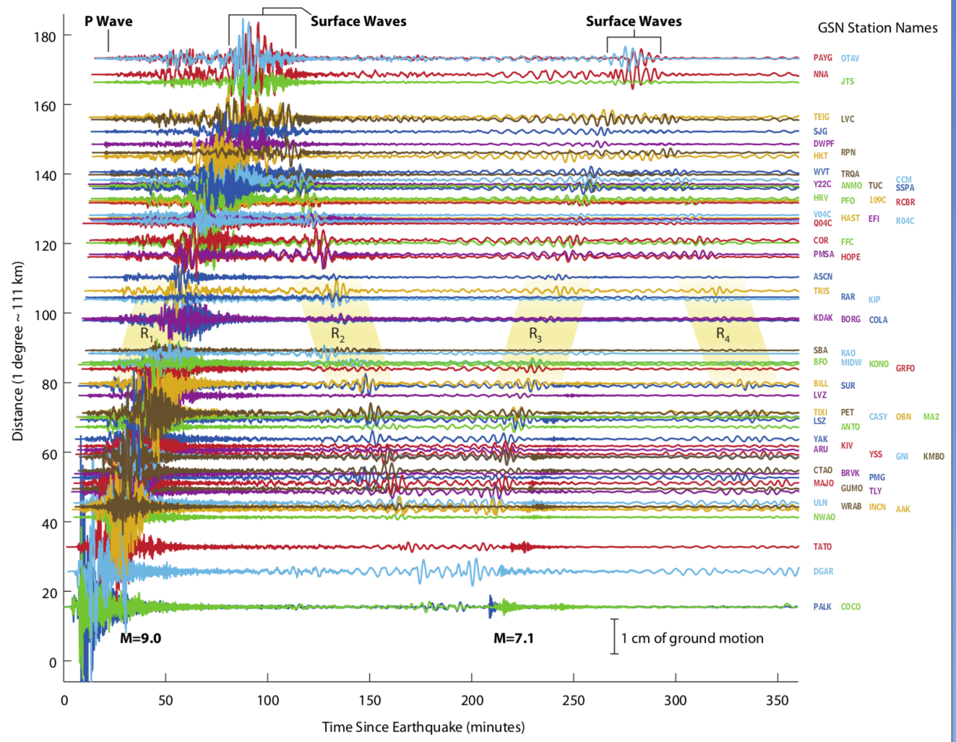
Motivation

What is our goal?

- We are using **FWI** with *seismograms* to image the *internal structure of the Earth*.
- FWI implementation needs a lot of, but only *good* data.
- High-frequency *surface wave* is messing up our data selection algorithm, so current algorithm has to discard some data that might still be useful.
- We hope *deep neural networks* classification could detect the similarity in data and synthetics unseen by humans to obtain more data.

What is a *seismogram*?

- An *earthquake* represents the shaking of the surface of the Earth, resulting from the sudden energy release of fault slip underground the Earth that creates *seismic waves*.
- A *seismogram* is the time variant wiggles recorded at the seismic measuring station, a proxy of seismic waves.



What is FWI?

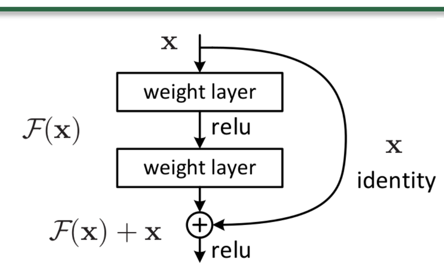
- Process from seismograms (*data*) to unknown structures (*models*) is called *inversion*.
- To fully exploit the information in data, Full-Waveform Inversion (FWI) was proposed.
- It tries to iteratively minimize the difference between the observed and synthetic seismograms.
- FWI is getting popular only recently due to the development of high-performance computing to deal with the massive computations.

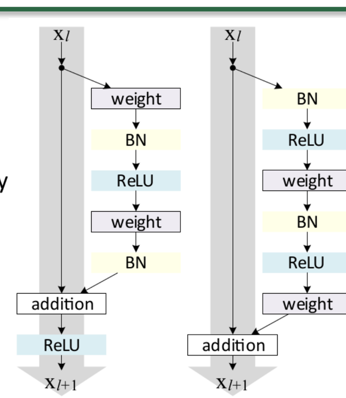
Method

What is ResNet and Why?

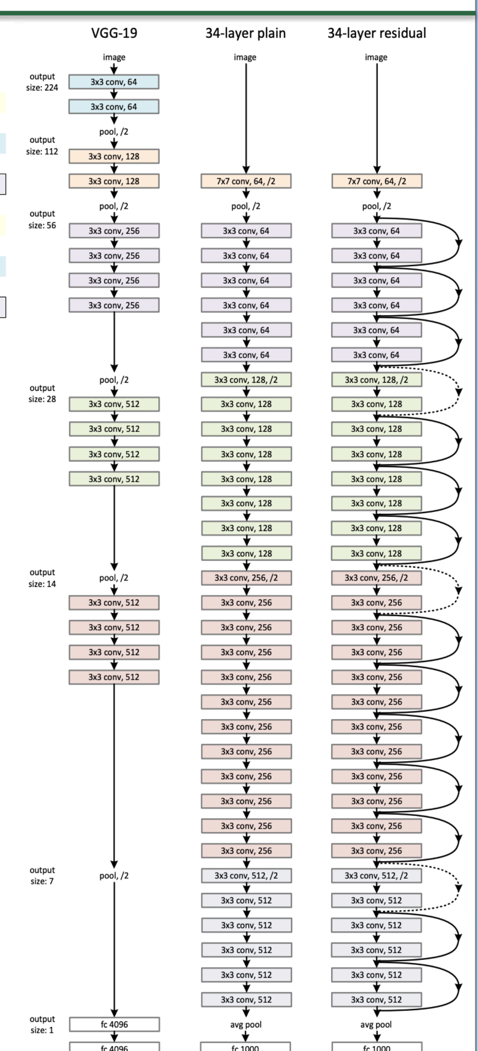
- *deep residual network* (ResNet) is one kind of *convolutional neural network* (CNN).
- CNN is a deep neural network involving convolution, inspired by the biological connectivity between animal neuron resembling.
- CNNs have already many successful applications in *Seismology,* one kind of time series analysis.

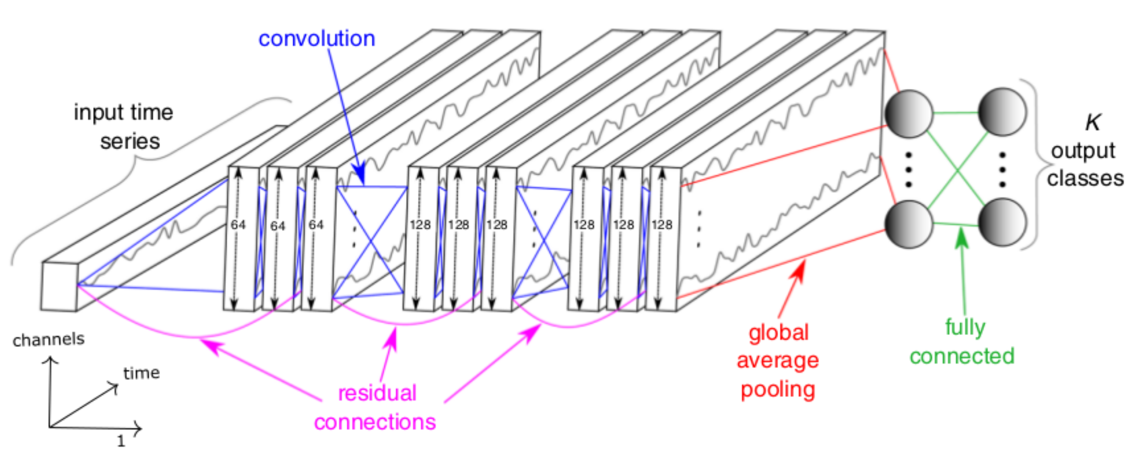
- ResNet is an ultra-deep CNN (hundreds of layers) proposed by Kaiming He et al. (2016)[1,2] in 2015.
- It solved the degradation problem by two core ideas: **residual learning** and **identity mapping**.
- This state-of-art has surpassed its ancestors in both image and time series analysis.





(a) original (b) proposed





- An similar architecture of ResNet in *Seismology* could be like above, but has not be tried yet in previous researches.

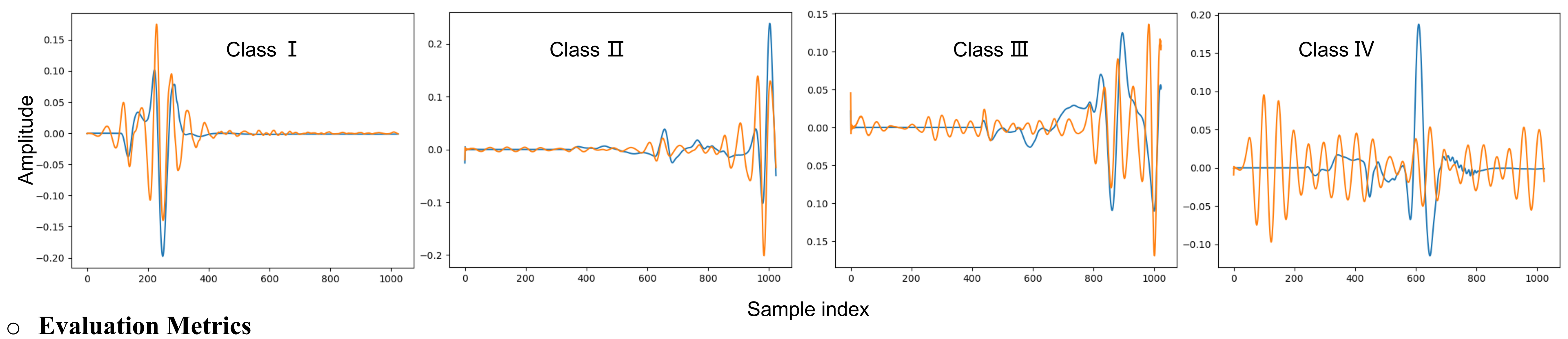
Data and Implementation

- **Data Description**
  - Use ObspyDMT to retrieve, process and manage seismic data[3]
  - All data used comes from IRIS (Incorporated Research Institutions for Seismology) Data Services[4]
  - Z component of 20,000 seismograms in total; each 30-min long; 50-s low-pass filtered; down-sampled from 18,000 to 1024.
  - Use the time-frequency misfit between the observed and synthetic seismogram[5] as the proxy to label our data in the train and test set

$$\chi_e^2(u_i^0, u_i) := \int_{\mathbb{R}^2} W_e^2(t,\omega)[|\,\tilde{u}_i(t,\omega)\,| - |\,\tilde{u}_i^0(t,\omega)\,|]^2 dt d\omega \qquad \chi_p^2(u_i^0, u_i) := \int_{\mathbb{R}^2} W_p^2(t,\omega)[\phi_i(t,\omega) - \phi_i^0(t,\omega)]^2 dt d\omega$$

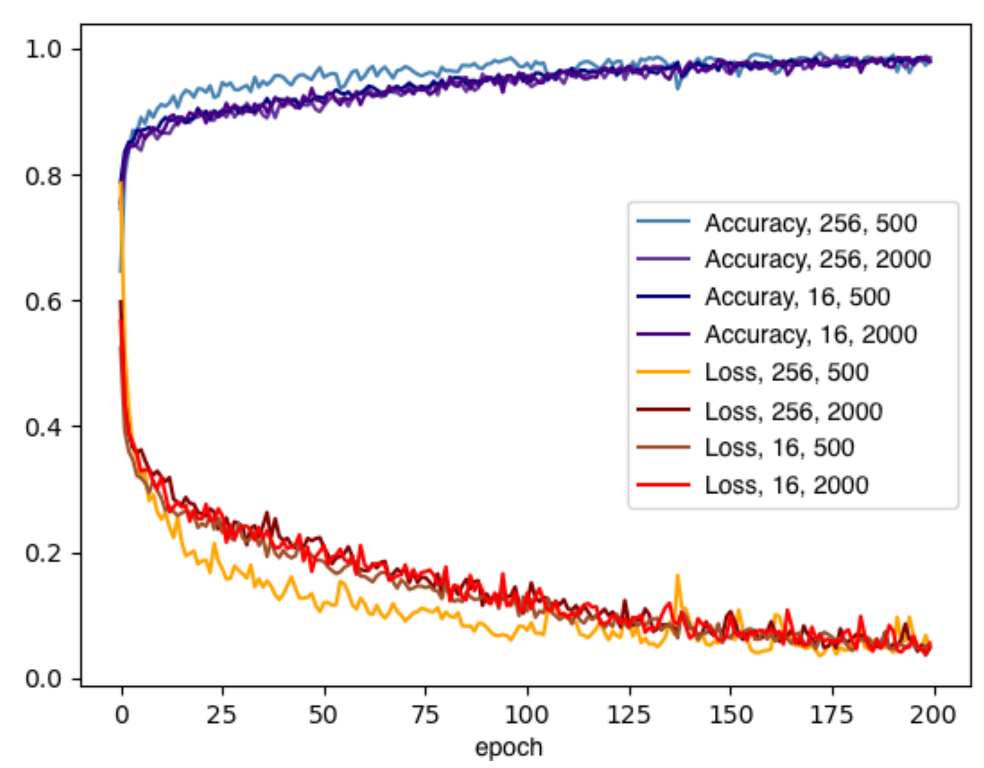
Class	I	II	III	IV
Empirical Misfit Boundary	<1.1	1.1-1.2	1.2-1.3	>1.3
Proportion	10%	15%	15%	60%
Description	good	inspection needed	gradually adding	discarded

- **Implemention**
  - Refer to the basic CNN API in Keras[6] and some time series analysis ideas in Fawaz et al. (2019)[7]
  - Grows the network by continuously adding more data and epochs to learn incrementally[8]
  - One-hot encoding; 11 layers including 9 convolutional layers in 3 blocks, 1 global average pooling layer and 1 softmax classifier layer.
  - In each block, fixed 64 (128, 128) filters in each conv layer, which equals to the number of extracted features; preceded by batch normalization and ReLU activation; mini-batch size is 16 (256)

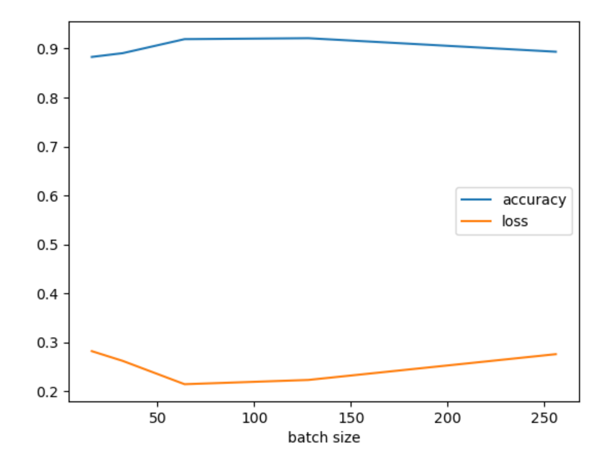


- **Evaluation Metrics**
  - Compare the results with the selection (labeling) algorithm.
  - Identify contradictory results (Fourier analysis and manual verification).
  - Add the seismogram to the inversion and check strange patterns (one of the ways we find out false positives in our current algorithm).

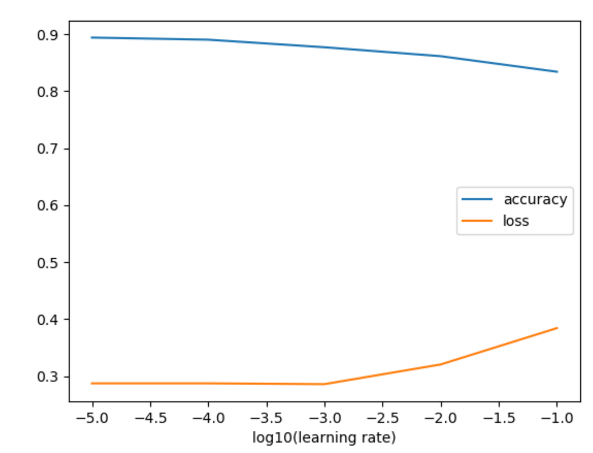
Result



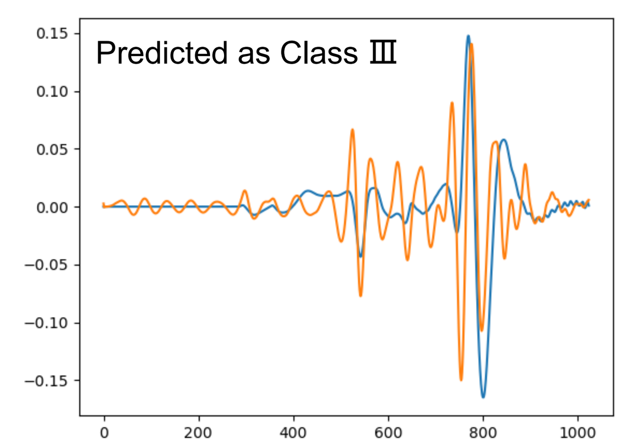
- Accuracy and Loss variation against with epoch using different batch sizes and test set size. 256 (16) means batch size. 500 (2000) means the data traces used in test set, the remaining traces are in the training set

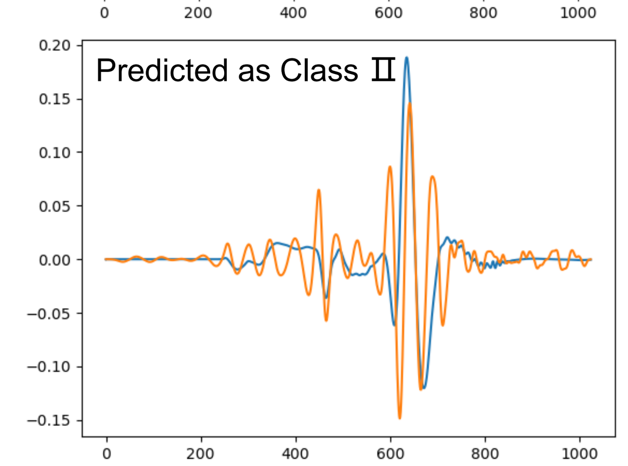


- Accuracy and Loss against with **batch size**



- Accuracy and Loss against with **learning rate**





- Suspicious seismograms seems more like what is predicted by the network? Need human intervention
  - Upper: predicted as Class III by ResNet, determined as II by T-F misfit, examined as III by human eyes
  - Lower: predicted as Class II by ResNet, determined as III by T-F misfit, examined as II by human eyes

Discussion

- **Data-preprocessing**
  - Data volume (more data in both train and test set)
  - Diversity (more types of bad observation or even synthetics)
  - Label credibility (is the current labelling algorithm reliable, could be tricky to quantify)
  - Other representations (Wavelet Transform; …)
- **Architecture**
  - Mini-batch size (size of data fed into the network every time)
  - Learning rate (cause tradeoff between accuracy and time cost)
  - Feature size (number of extracted features in each convolutional layer)
  - Depth (add more layers)
- **Other methods**
  - FCNN (fully-connected neural networks)
  - SVM (support vector machine)
  - RF (random forests and ensemble learning)

Conclusion

- Using automated Resnet classification is feasible for data selection during the data processing stage before FWI.
- Resnet has achieved an acceptable accuracy in finding qualified data for FWI, and its performance is already better than our previous algorithm.
- Several parameters in architecture design, such as mini-batch size, number of features and network depth, can be further optimized and so does the input itself, i.e. data-preprocessing.
- Although this project is rather preliminary, it might illuminate a broader application of ResNet in the future of Seismology, given the big data era is coming.

Reference

[1] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778. 2016.

[2] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Identity mappings in deep residual networks." In *European conference on computer vision*, pp. 630-645. Springer, Cham, 2016.

[3] https://github.com/kasra-hosseini/obspyDMT

[4] http://ds.iris.edu/ds/

[5] Fichtner, Andreas. *Full seismic waveform modelling and inversion*. Springer Science & Business Media, 2010.

[6] https://keras.io/examples/cifar10_resnet/

[7] https://github.com/hfawaz/dl-4-tsc

[8] Xiao, Tianjun, Jiaxing Zhang, Kuiyuan Yang, Yuxin Peng, and Zheng Zhang. "Error-driven incremental learning in deep convolutional neural network for large-scale image classification." In *Proceedings of the 22nd ACM international conference on Multimedia*, pp. 177-186. ACM, 2014.