SnT 2025
8 SEPTEMBER
ONLINE DAY
9 TO 12 SEPTEMBER
AT HOFBURG PALACE, VIENNA & ONLINE
CTBT: SCIENCE AND TECHNOLOGY CONFERENCE

Machine-Learning Event Classification in North-East China using augmented data

Alemayehu L Jemberie (PhD)

United States of America Air Force Technical Applications Centre (AFTAC, ajemberie@aftac.gov)

INTRODUCTION AND MAIN RESULTS

A transfer learning approach was adopted, using a VGG16 neural network model to classify earthquakes from non-earthquake events in North-East China close to the North Korean test site.

- Deeper events (depth > 5 km, assumed to be Earthquakes) are identified at 83% accuracy.
- Shallow events (depth < 5km, assumed to be explosions) are fewer than deeper events in number and have low identification rate at 62%.
- Synthetics were generated for Shallow events only. Scattering effect needs to be included in the waveform synthesis. More explosion (shallow) data is needed for better training.

Machine-Learning Event Classification in North-East China using augmented data

Alemayehu L Jemberie (PhD)

P2.1-836

Abstract

A transfer learning approach was adopted, using a VGG16 neural network model to classify earthquakes from non-earthquake events in North-East China close to the North Korean test site. Magnitudes below 4 are considered for classification of tectonic events (earthquakes) from explosions. Because of the scarcity of observed explosion data, synthetic seismograms were generated using SW4 to augment the observed data so that overfitting doesn't happen during validation. Our preliminary experiment shows promising results which may help us identify any potential future small event.

A Simple Neural Network

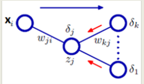
A simple Neural Network

Given: $a_j = \sum_{i=0}^{D} w_{ji}^{(1)} x_i$

$z_j = h(a_j)$

$a_k = \sum_{j=0}^{M} w_{kj}^{(2)} z_j$

$y_k = \sigma(a_k)$



The Loss/Error/Objective function:

Loss = $E = \frac{1}{2}\sum_{k=0}^{N}(y_k - t_k)^2 = \left(\frac{1}{2}\sum_{k=0}^{N}(\sigma(a_k) - t_k)^2\right)$

Objective: minimize the Loss function

t_k are known labels

$\delta_j = \delta_k w_{kj}^{(2)} h'$ and $\nabla E = \frac{\partial E}{\partial w_{ji}} = \delta_j x_i$

Hidden layer:

Weight change: $\Delta w_{ji} = -\tau_j \nabla E(w_{ji})$

But, $\nabla E(w_{ji}) = \frac{\partial E}{\partial w_{ji}} =$

$\delta_j x_i = \sum_{k=0}^{N} \delta_k w_{kj}^{(2)} h' x_i$

$\Delta w_{ji}^{(1)} = -\tau_j \delta_j x_i = -\tau_j \sum_{k=0}^{N} \delta_k w_{kj}^{(2)} h' x_i$

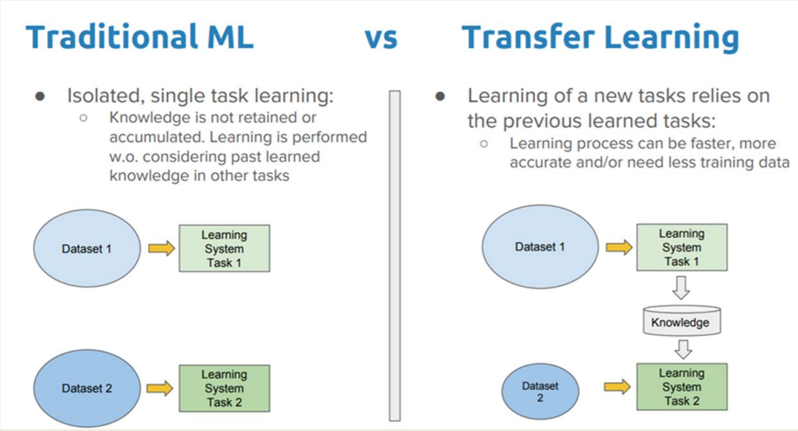
Output layer:

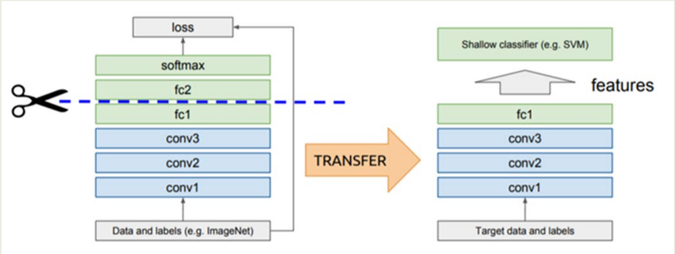
$\Delta w_{kj}^{(2)} = -\tau_k \delta_k z_k$

Deep Neural Network (L hidden layers)

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}\left[\mathbf{a}^{(L+1)}\left(\mathbf{h}^{(L)}\left(\mathbf{a}^{(L)}\left(\ldots\left(\mathbf{h}^{(2)}\left(\mathbf{a}^{(2)}\left(\mathbf{h}^{(1)}\left(\mathbf{a}^{(1)}(\mathbf{x})\right)\right)\right)\right)\right)\right)\right)\right]$$

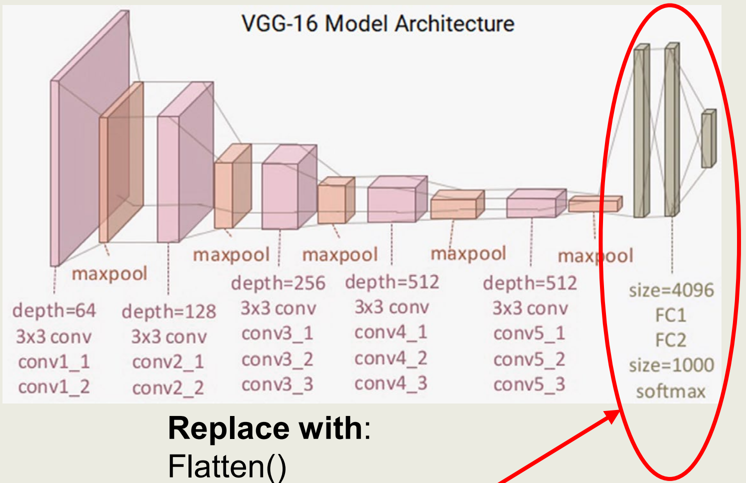
Why Transfer Learning?





"idea: use outputs of one or more layers of a network trained on a different task as generic feature detectors. Train a new shallow model on these features." Credit: Hands-on Transfer Learning with Python (2018, Dipanjan Sarkar, Raghav Bali and Tamoghna Ghosh)

Transfer Learning using VGG16 architecture



Replace with:
Flatten()
Dense(256, activation = 'relu')
Dense(1,activation = 'sigmoid')

VGG16:

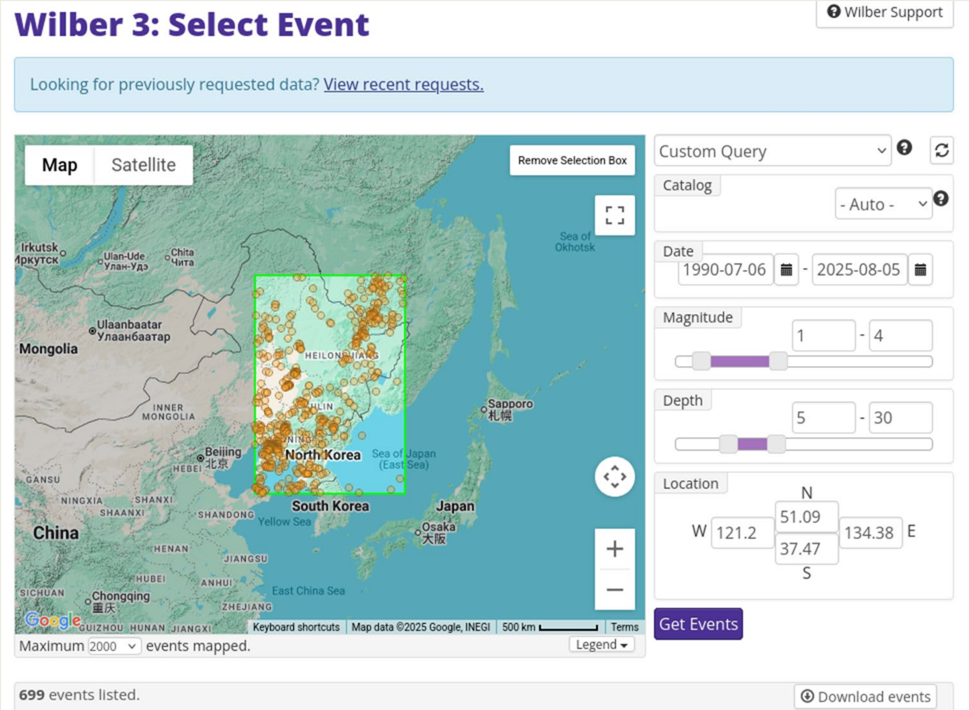
- A 16-layer deep convolution networks (ConvNets) learning architecture (Simonyan and Zisserman, 2015)
- Winner of ImageNet Challenge in 2014

Machine-Learning Event Classification in North-East China using augmented data

Alemayehu L Jemberie (PhD)

P2.1-836

Data



More than 600 events that occurred at depths > 5 km (assumed to be earthquakes) with magnitudes between 1 and 4 for the past 35 years in the study area. (Source: https://ds.iris.edu/wilber3/find_event)

Seismogram pre-processing

Pre-processing of seismograms from both earthquakes and explosions:

- Correct effect of recording instrument
- Remove mean
- Remove trend
- Band-pass seismograms between 1 and 20 Hz.
- Images were calculated for each seismogram.

Training, Validating and Testing Data

- Training data: 3800 seismograms from deeper events (assumed earthquakes, all observed), and 3800 from shallow (assumed explosions, half observed and half synthetic)
- Validation: about 1900 seismograms from each class
- Testing data: about 1200seismograms from each class

SnT 2025
8 SEPTEMBER ONLINE DAY
9 TO 12 SEPTEMBER AT HOFBURG PALACE, VIENNA & ONLINE
CTBT: SCIENCE AND TECHNOLOGY CONFERENCE

Machine-Learning Event Classification in North-East China using augmented data

Alemayehu L Jemberie (PhD)

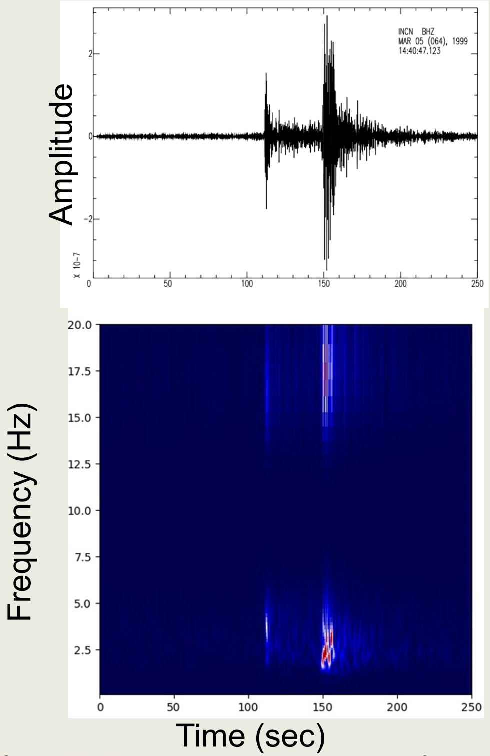
P2.1-836

Continuous Wavelet Transform (CWT)

The Continuous Wavelet Transform (CWT), $F_W(\sigma,\tau)$, is defined as the inner product of a mother wavelet $\psi_{\sigma,\tau}(t)$ with the seismogram $f(t)$:

$$F_W(\sigma,\tau) = \int_{-\infty}^{\infty} f(t)\frac{1}{\sqrt{\sigma}}\bar{\psi}\left(\frac{t-\tau}{\sigma}\right)dt$$

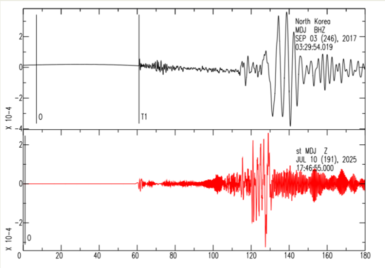
Where $\bar{\psi}$ is the complex conjugate of ψ, σ is a dilation parameter and τ is a translation parameter (Sinha et al, 2005).



CWT of a shallow (< 5km) event

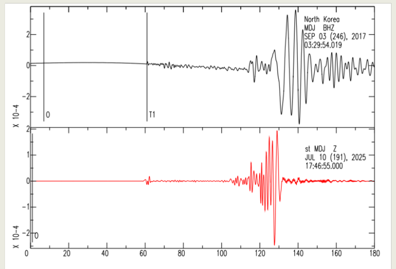
Waveform Synthesis of the September 02, 2017, M6.3 N. Korean Nuclear Explosion as recorded by station MDJ.

- The synthetic waveforms are generated using SW4 Code (3D) (Peterson et.al., 2023).
- 3D velocity Structure (Vs, Vp and density) from Julia et al. 2021 was used.
- Station MDJ is an IMS station.



Observed

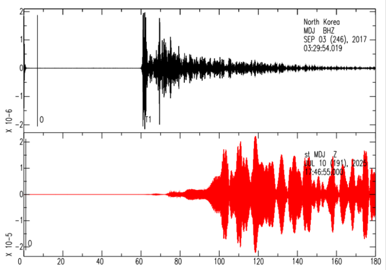
Synthetic



Observed

Good match at low frequencies

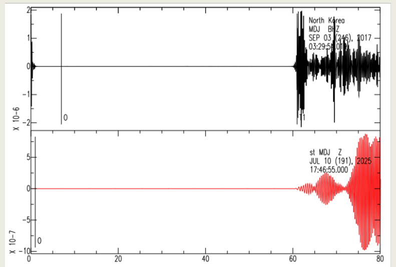
Synthetic



Observed

Difficult to synthesize the high frequency part

Synthetic (notice the aliasing effect)



Observed

Zoomed bn 0 and 80 sec

Synthetic

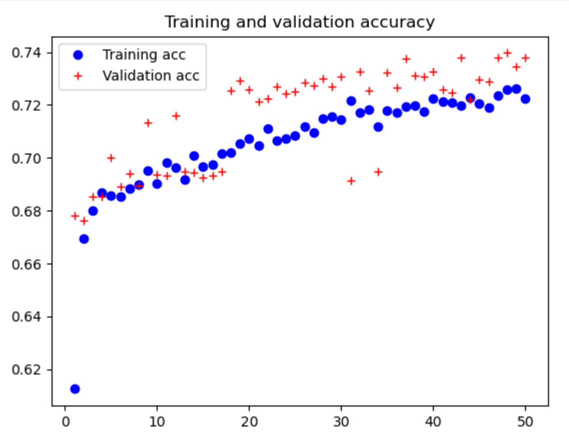
SnT 2025 CTBT: SCIENCE AND TECHNOLOGY CONFERENCE
8 SEPTEMBER ONLINE DAY
9 TO 12 SEPTEMBER AT HOFBURG PALACE, VIENNA & ONLINE

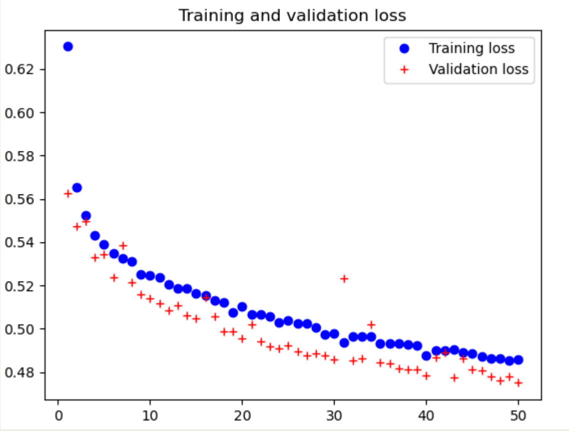
Machine-Learning Event Classification in North-East China using augmented data

Alemayehu L Jemberie (PhD)

P2.1-836

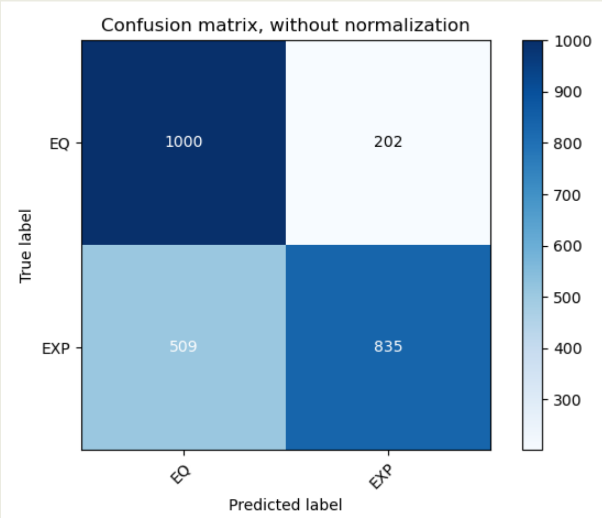
Training and Validation Accuracy, and Loss

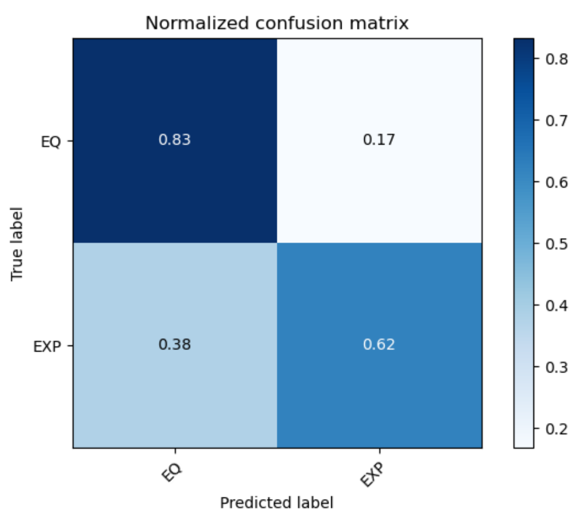




Epoch

Event Identification





Conclusions

- Deeper events (depth > 5 km, assumed to be Earthquakes) are identified at 83% accuracy.
- Shallow events (depth < 5km, assumed to be explosions) are fewer than deeper events in number and have low identification rate at 62%.
- Synthetics were generated for Shallow events only. Scattering effect needs to be included in the waveform synthesis.
- More explosion (shallow) data is needed for better training.

References

Petersson, N. Anders, Sjogreen, Bjorn, Tang, Houjun, & Pankajakshan, Ramesh. (2023, September 6). geodynamics/sw4: SW4, version 3.0. doi:10.5281/zenodo.8322590, url: https://doi.org/10.5281/zenodo.8322590

Tang, Z., Julià, J., Mai, P. M., Mooney, W. D., & Wu, Y. (2022). Shear-wave velocity structure beneath Northeast China from joint inversion of receiver functions and Rayleigh wave phase velocities: Implications for intraplate volcanism. Journal of Geophysical Research: Solid Earth, 127,.