Supporting Information for

Using a Deep Neural Network and Transfer Learning to Bridge Scales for Seismic Phase Picking

Chengping Chai[1], Monica Maceira[1,2], Hector J. Santos-Villalobos[1], Singanallur V. Venkatakrishnan[1], Martin Schoenball[3], Weiqiang Zhu[4], Gregory C. Beroza[4], Clifford Thurber[5], and EGS Collab Team

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA.

[2]Department of Physics and Astronomy, University of Tennessee, Knoxville, Tennessee, USA.

[3]Lawrence Berkeley National Laboratory, Berkeley, California, USA.

[4]Department of Geophysics, Stanford University, Stanford, California, USA.

[5]Department of Geology and Geophysics, University of Wisconsin-Madison, Madison, Wisconsin, 53706, USA

Contents of this file

Introduction

The supporting information includes a paragraph (Text S1) that explains precision, recall, and F1 score, a figure (Figure S1) of sensor and borehole locations, a figure (Figure S2) with example seismograms, a figure (Figure S3) showing number of data used for transfer learning (TL), histograms (Figure S4-S5) that compare data fit for both P and S waves and for manual picks and TL-derived picks, example manual phase picks and TL-derived picks with corresponding seismograms (Figure S6), checkerboard test results for P and S waves (Figure S7-S8), a figure (Figure S9) showing the neural network architecture, two figures (Figure S10-S11) comparing weights of two convolution layers, a figure (Figure S12) showing hidden convolutional features

using the PhaseNet model (W. Zhu & Beroza, 2018), and a figure (Figure S13) showing hidden features using the TL model.

Text S1.

For classification tasks, a prediction is defined as true positive when both the predicted label and ground truth are positive; as true negative when both the predicted label and ground truth are negative; as false positive when the predicted label is positive but the ground truth is negative; as false negative when the predicted label is negative but the ground truth is positive. For our case, true positive means the difference between a deep-learning-derived (or TL-derived) pick (P-wave or S-wave) and the corresponding manual pick is less than the assumed measurement error. True negative means both the deep learning model (or TL model) and the human analyst were not able to find a pick from a seismic waveform. False negative means that the human analyst found a pick from a seismic waveform but the deep learning model (or TL model) was not able to find a pick. False positive means the deep learning model (or TL model) found a pick from a seismic waveform but the human analyst did not, or the time difference is larger than the assumed error. Precision, recall, and F1 score were computed using the following

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

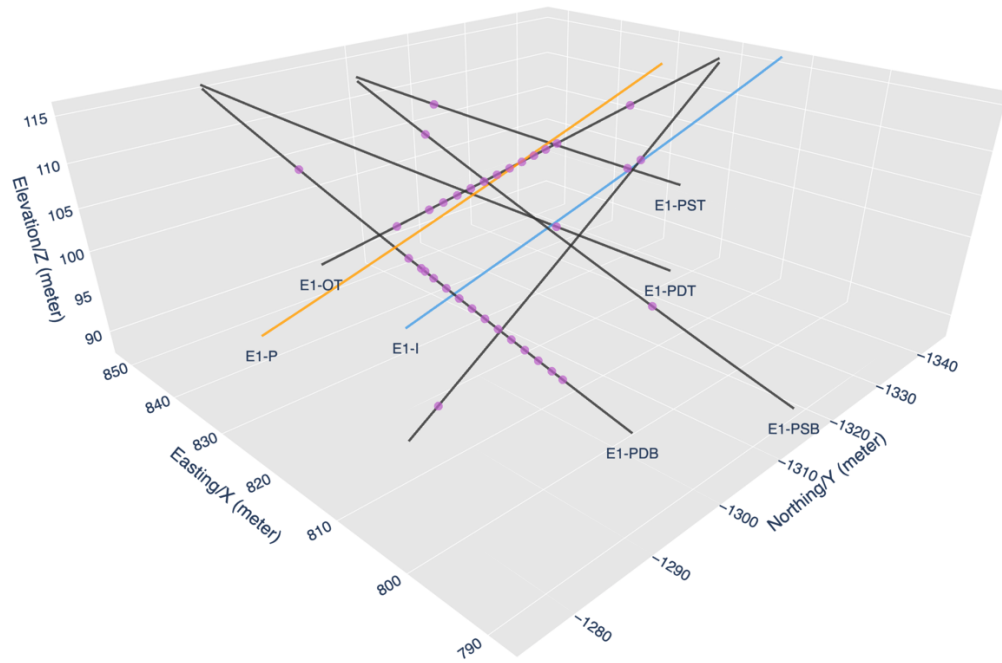$$F1 = \frac{2\ \times\ Precision\ \times\ Recall}{Precision + Recall}$$

formulas.

Figure S1. Seismic sensors (purple dots) and boreholes (lines). The blue line is the injection well (E1-I), and the yellow line is the production well (E1-P). Black lines represent monitoring wells.
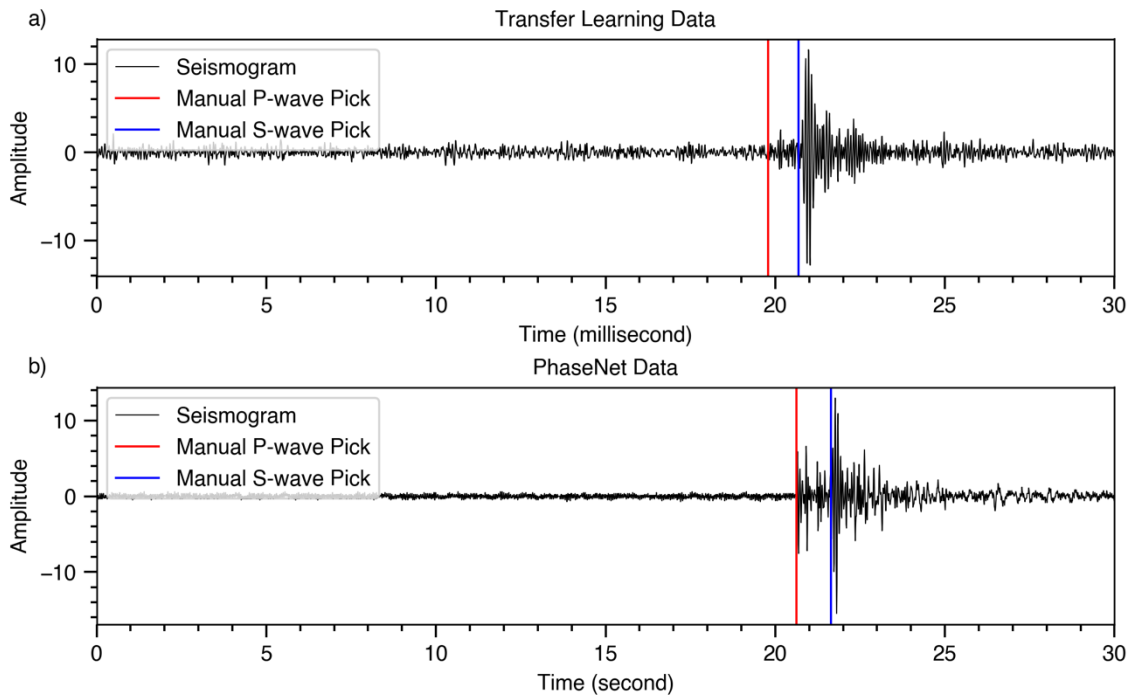

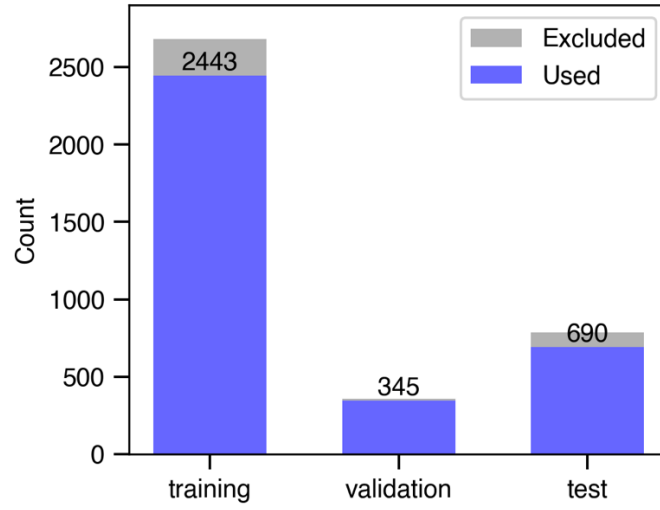**Figure S2.** A comparison of seismograms used for (a) TL and (b) the original PhaseNet.

Figure S3. Number of waveform samples of training, validation and test set used for TL.
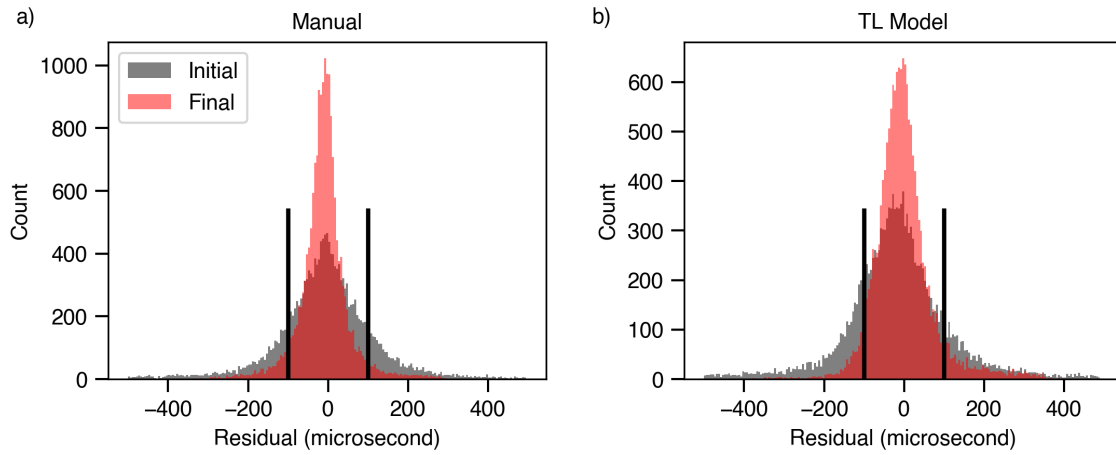

Figure S4. Histograms comparing the P-wave residual difference between the initial velocity model and the final velocity model for (a) manual picks and (b) TL model picks. The vertical lines indicate the assumed uncertainty of phase picks.
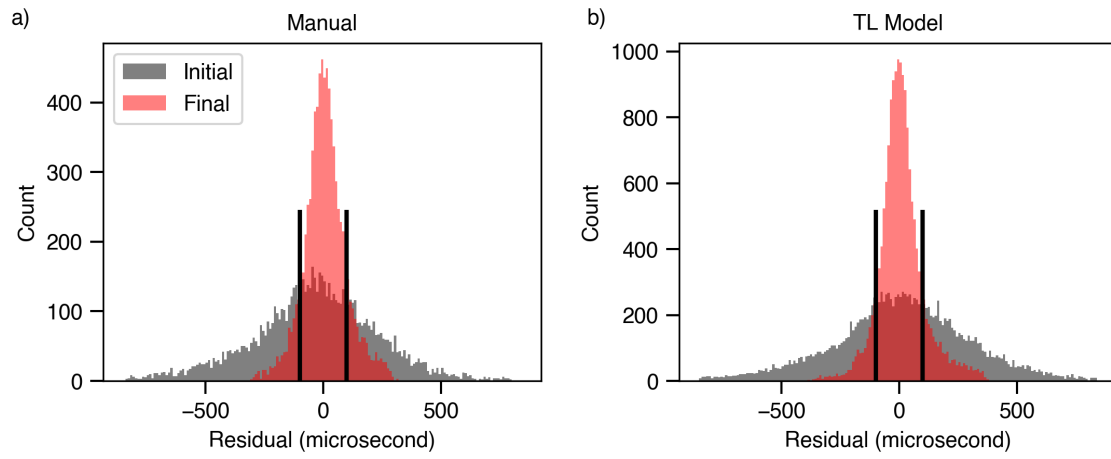
Figure S5. Histograms comparing the S-wave residual difference between the initial velocity model and the final velocity model for (a) manual picks and (b) TL model picks. The vertical lines indicate the assumed uncertainty of phase picks.
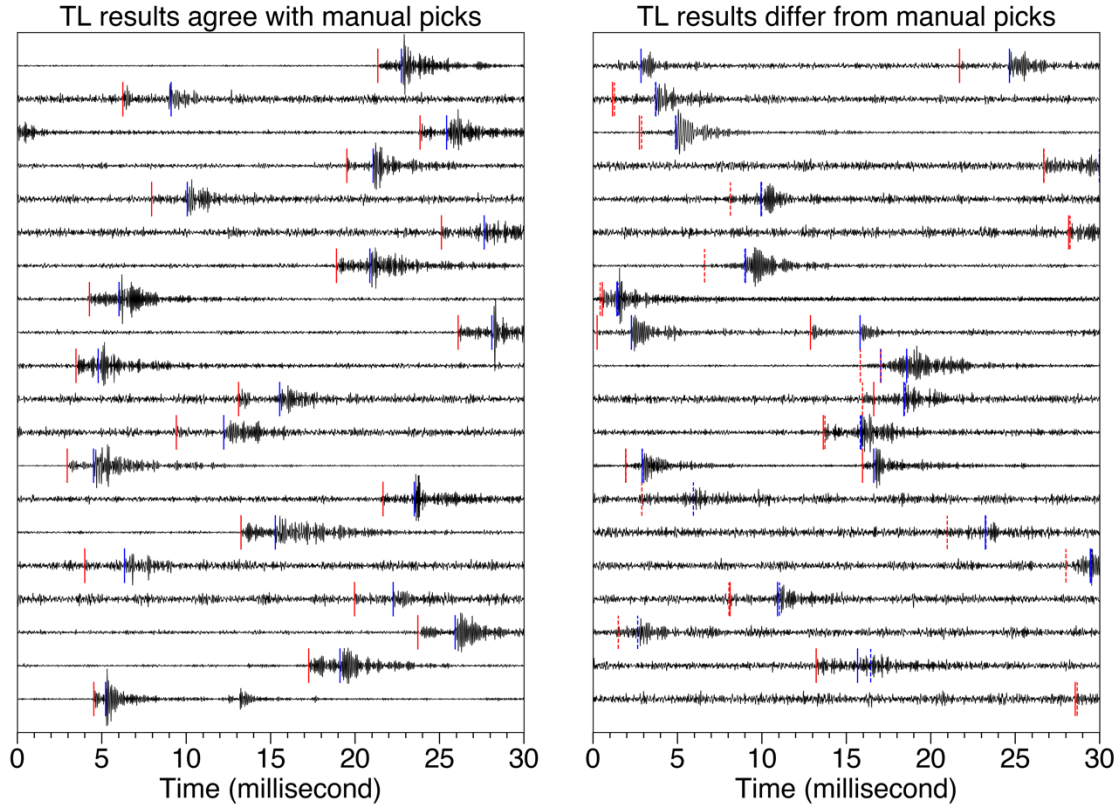
Figure S6. Randomly selected example waveforms and TL-derived phase picks that agree with (left) and differ from (right) the manual picks. Only one component seismogram is shown here, but three component seismograms were used for phase picking. Red vertical lines represent P-wave picks. Blue vertical lines are for S-waves. Solid vertical lines are TL-derived picks. Dashed lines are manual picks. When the time difference between TL results and manual picks is larger than 0.1 ms or one of the picks is missing, we consider the results to be different.