Machine learning experiments on velocity extraction from migration images

Zhan Niu* and Daniel O. Trad

niuz@ucalgary.ca

Abstract

In full waveform inversion (FWI), the update of velocity is obtained by calculating the gradient of the misfit between recorded and predicted data, which is defined by the cross-correlation of the reverse time of receiver wavefield and source wavefield. Benefits can be achieved by solving a direct non-linear mapping between the correlation and model update. In this report, we train a fully connected neural network with residual blocks which allows migrated images to be directly mapped into velocity models. The input images and the true velocity model comes from reverse time migration results on randomly generated 4-layer models. The training is performed with ADAM optimizer combined with L1/L2 norms as the loss function. Performance and convergence of the neural network with different hyper-parameters are also investigated systematically. We have tested the trained model with different synthetic inputs. Results show the that trained network is relatively model-dependent which performs well on the validation set but does a poor job on datasets that come from different distributions.

Theory

We use L_1 and L_2 square as the loss function, which are defined as

$$L_1(\boldsymbol{y}, \boldsymbol{y}_{pred}) = \frac{1}{n}\sum_{i}^{n}\left|y^{(i)} - y_{pred}^{(i)}\right|$$

$$L_2^2(\boldsymbol{y}, \boldsymbol{y}_{pred}) = \frac{1}{n}\sum_{i}^{n}\left(y^{(i)} - y_{pred}^{(i)}\right)^2$$

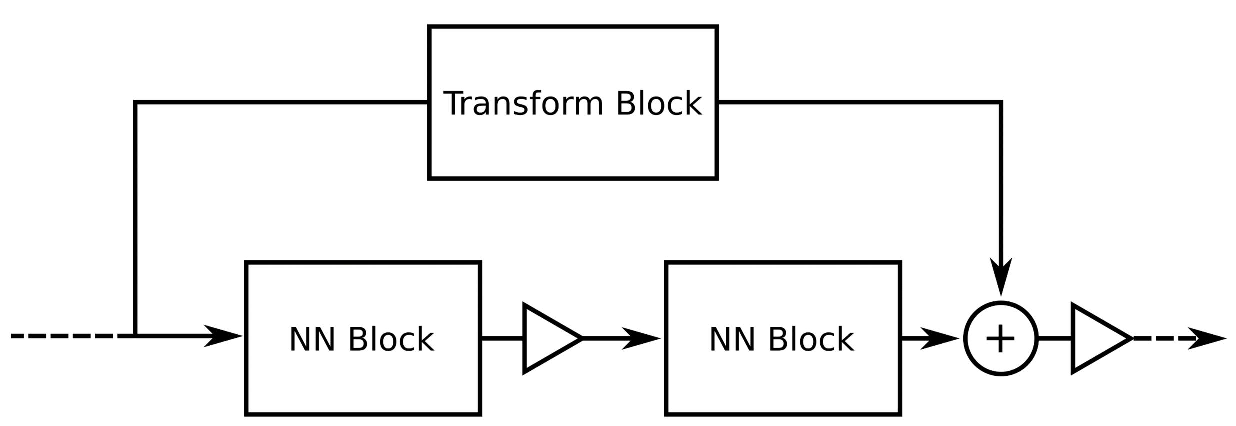

FIG. 1. A ResNet building block modified from He et al. (2016). Regular triangles refer to activation functions. Dashed arrows are connected to other blocks.

We use ResNet to help improve the accuracy of predictions. ResNet introduces shortcuts that enable the neural network to skip unnecessary steps. Since there are fewer terms when applying the chain rule. This process prevents vanishing gradient to some degree. Furthermore, the shortcut can be understood in another way. Suppose that the transformation block is identity.

Theory

Instead of fitting a function that maps from the block input x to the block output y , the ResNet block is trying to fit a function that maps from x to $(y - x)$. In other words, the ResNet is forced to focus on learning features that are non-linear to x.

Synthetic Data Examples

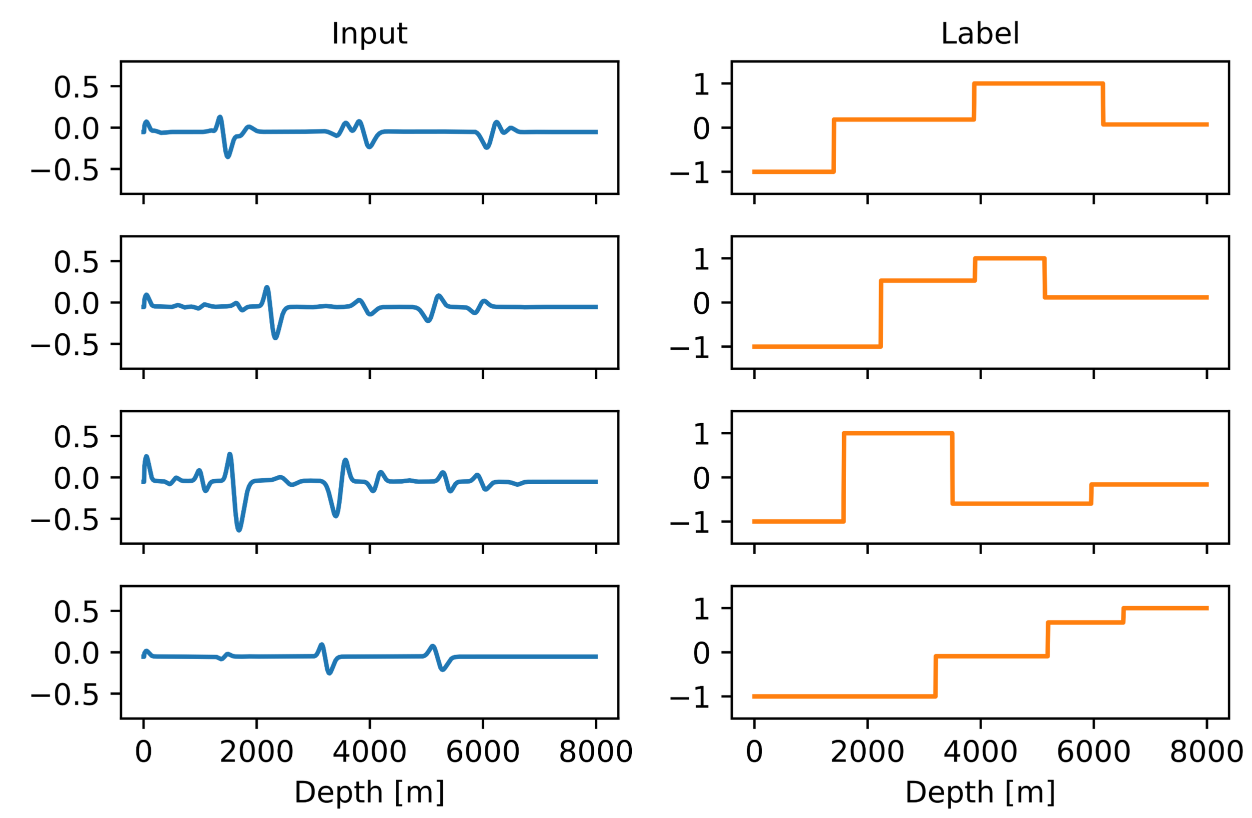

FIG. 2. Four random examples of input and label pairs.

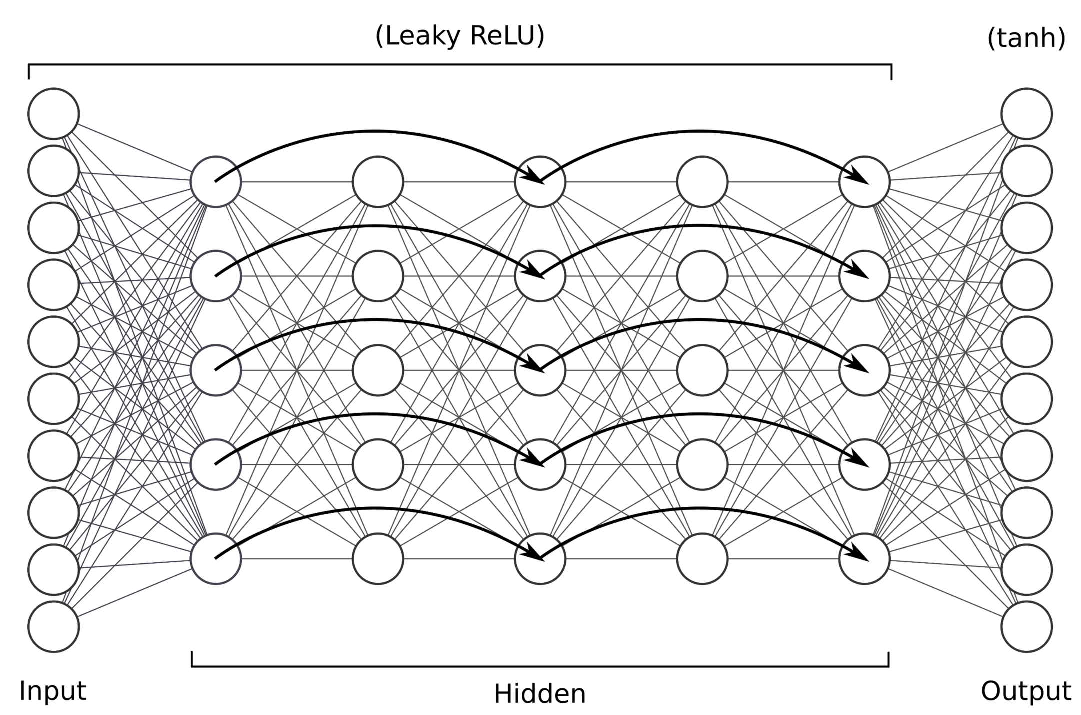

FIG. 3. A ResNet based on a fully connected network

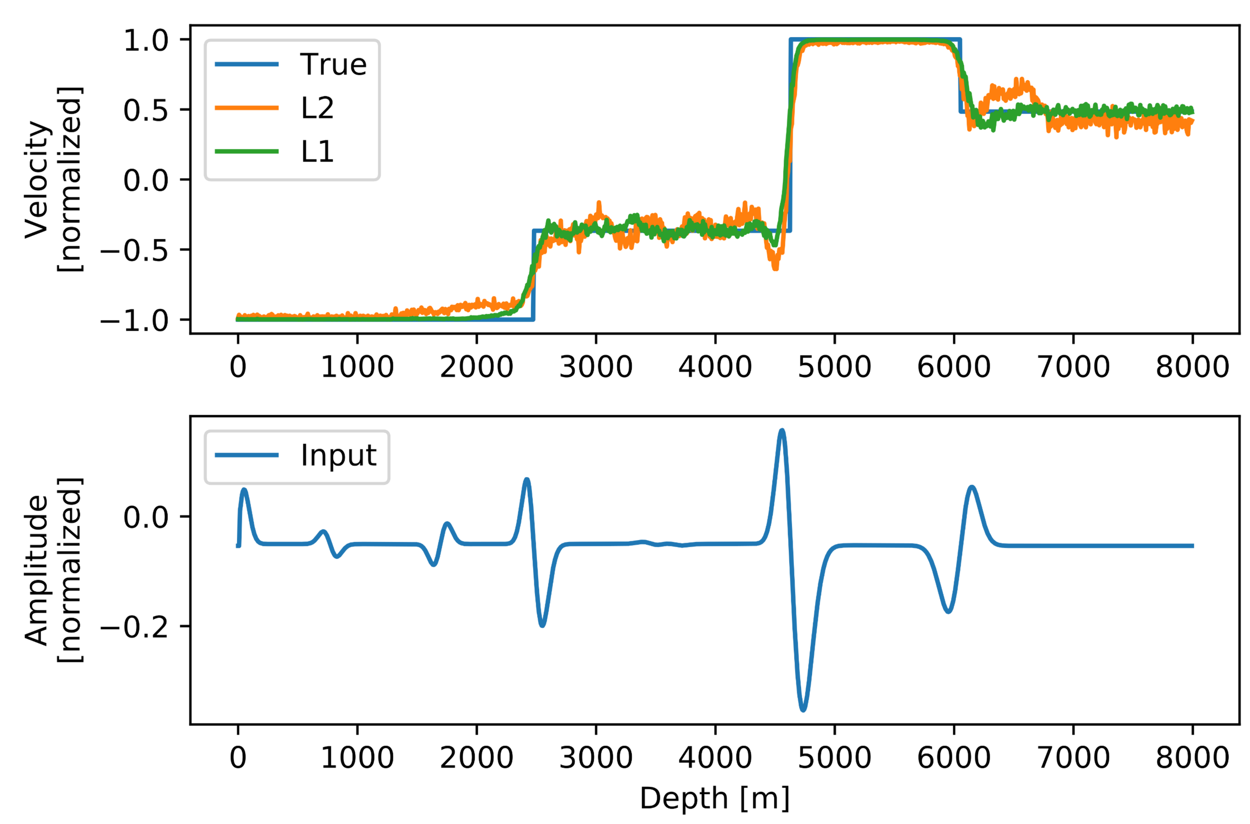

FIG. 4. Predictions made by models with L1 and L2 loss function, respectively.

Synthetic Data Examples

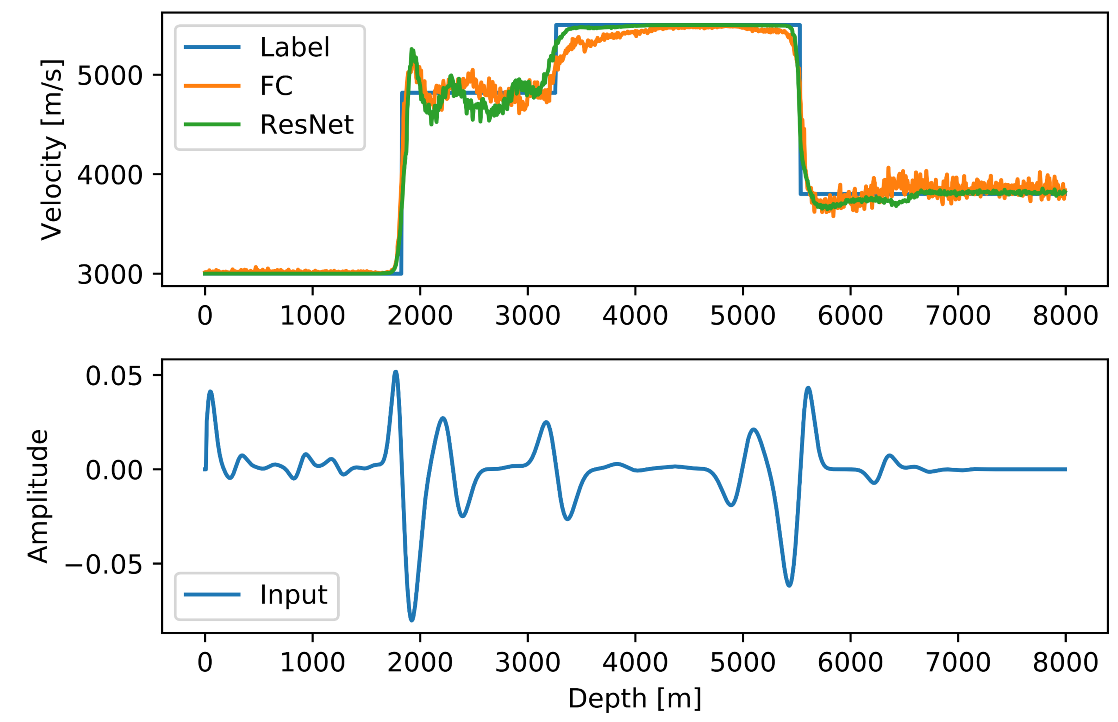

FIG. 5. A comparison between predictions from the fully connected (FC) model and the ResNet.

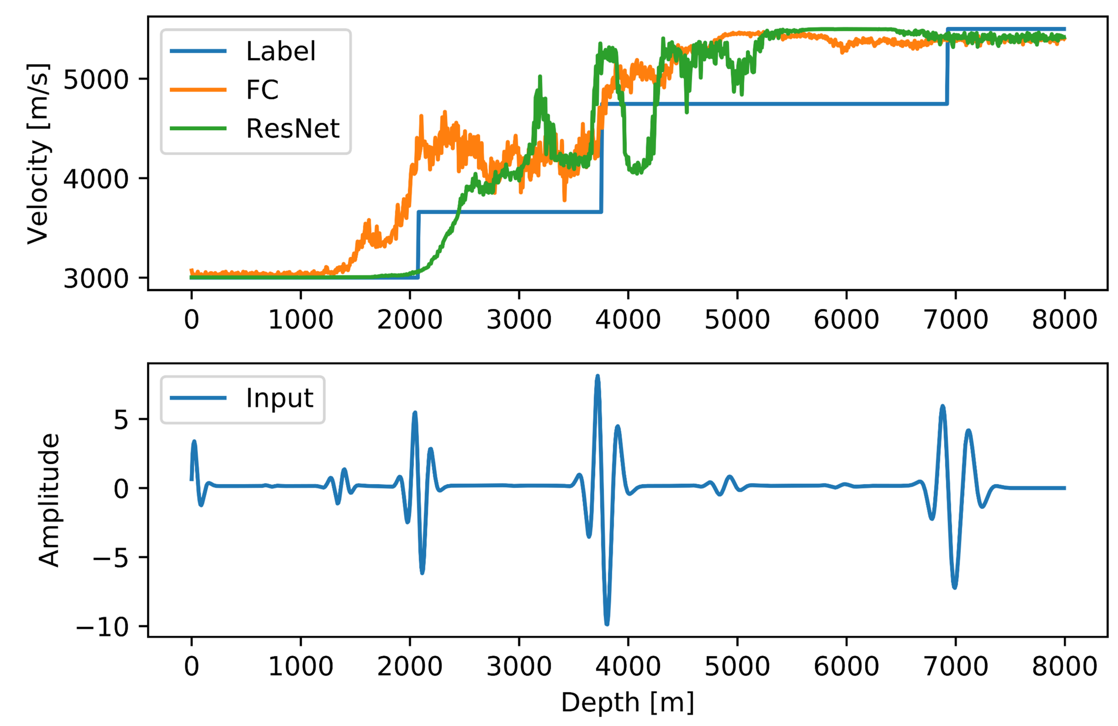

FIG. 6. A typical prediction on data with Ricker wavelet.

Conclusion

We use fully-connected networks to recover the migration images from random four 4-layer velocity models. We investigate different behaviours when using L1/L2 norms as loss function and we conclude L1 is more suitable for this type of problem. We test ResNet shortcuts to the network and they reduce fluctuations. The model performs poorly on data from different distributions of the training set. Future works may include applying more advance training techniques like gradient boosting or seeking better representations of the input and outputs. Also, we may need to try total variation instead of L1 norm.

Acknowledgements

We thank the sponsors of CREWES for continued support. This work was funded by CREWES industrial sponsors and NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 461179–13. Special thanks to Jian Sun at Penn State University and Marcelo Guarido for valuable discussions.