TraML: separation of seismically-induced ground-motion signals with autoencoder architecture

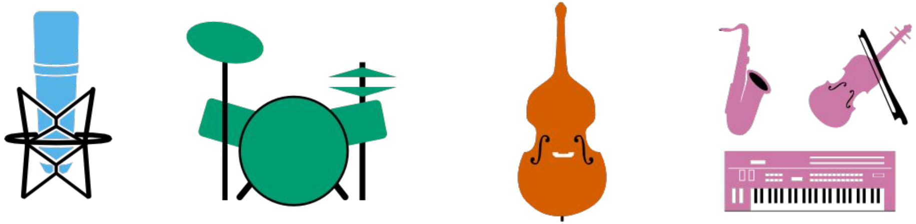
Novoselov A., Hein G., Fuchs F., Bokelmann G.

imgw
Institut für Meteorologie und Geophysik

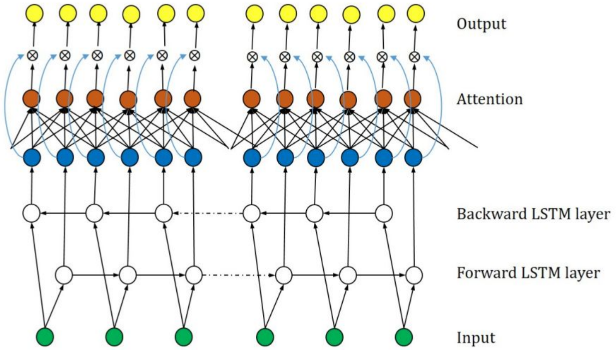
universität wien

Intro

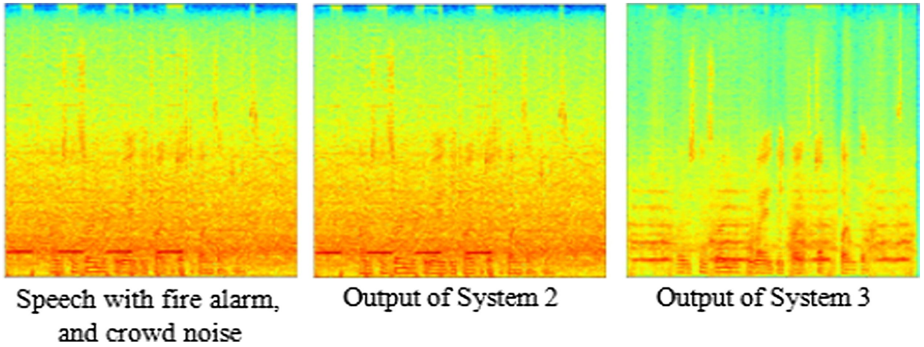
Source separation is a known problem in the Machine Learning domain; there are successful applications to music, hearing aids, and speech enhancements.



[Deezer](#) - open-source ML for instrument separation

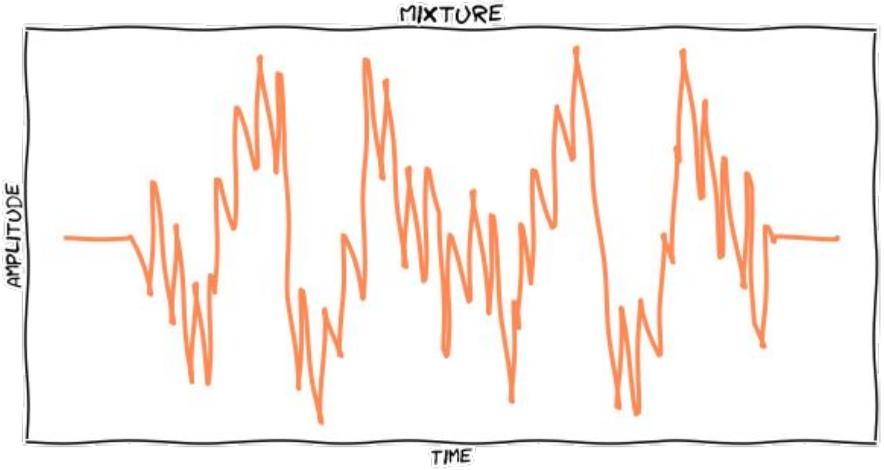


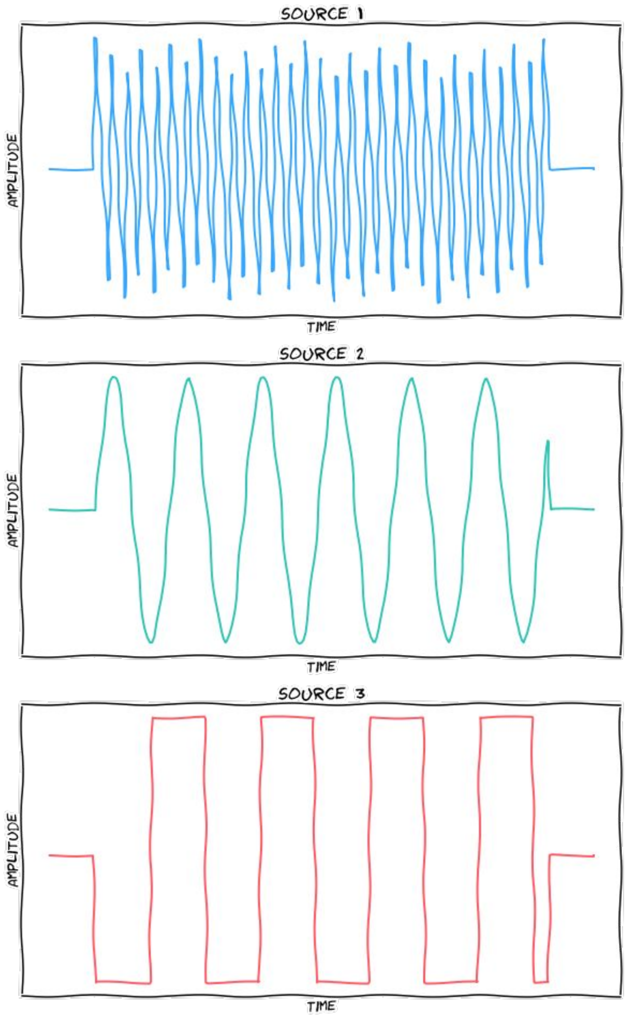
La Furca End-to-End Monaural Speech Separation ([Shi et al., 2020](#))



Enhanced smart hearing aid using deep neural networks ([Nossier et al. 2019](#))

Source separation

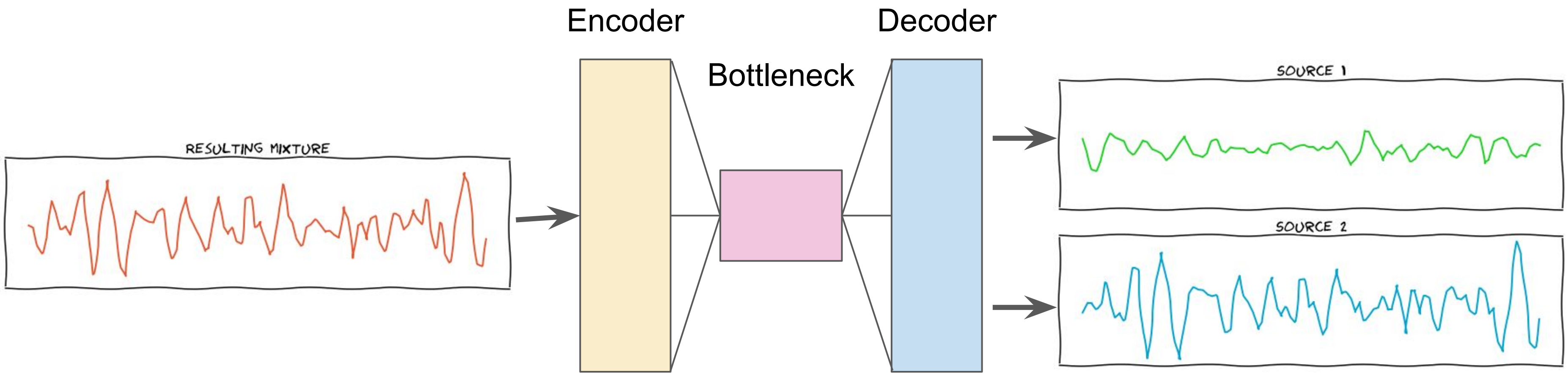




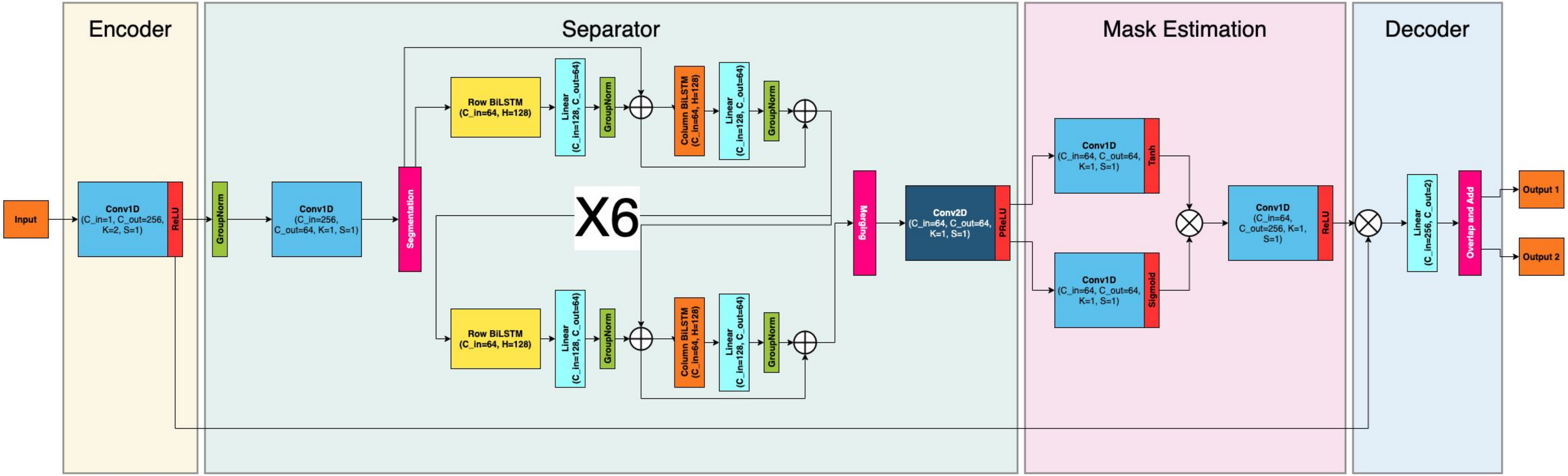
Several signals can be added together to obtain a mixture, but a mixture can also be separated back into individual signals

Autoencoders

Autoencoders are a type of machine learning neural network consisting of 3 major components: an **encoder**, which ensures efficient compression of the data, a **bottleneck** that performs re-representation of the compressed data, and a **decoder** - that decompress the data into the desired output.



DPRNN TasNet - Dual Path Recurrent Neural Network



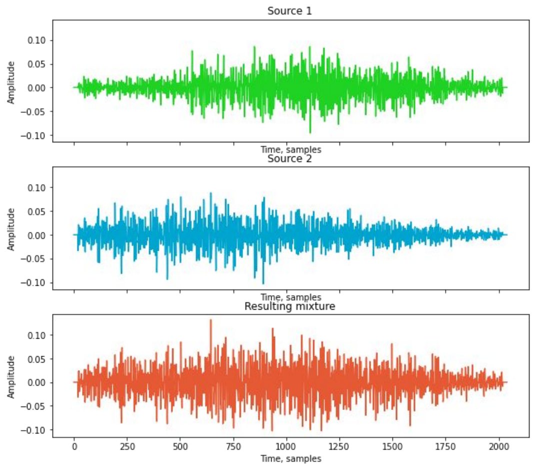
*Architecture of Dual Path Recurrent Neural Network TasNet (Luo et al. 2020). **Conv1D** and **Conv2D** - 1D and 2D convolution operations, correspondingly; **PReLU**, **ReLU**, **Tanh** and **Sigmoid** - activation functions; **Linear** - Fully-Connected layer; **GroupNorm** - Group Normalization, Row and Column **BiLSTM** - row-wise and column-wise bidirectional Long-Short-Term-Memory Cells; **Separation**, **Merging**, **Overlay and Add** - array manipulations. Arrows indicate an order of operations applied to the input. **+** is element-wise summation; **x** is element-wise multiplication.*

Data

Data recorded with the Raspberry Shake seismic sensor at the University of Vienna, located ~20 meters from railway tracks (S40, U4, Spittelau station), at sampling frequency 100 Hz.

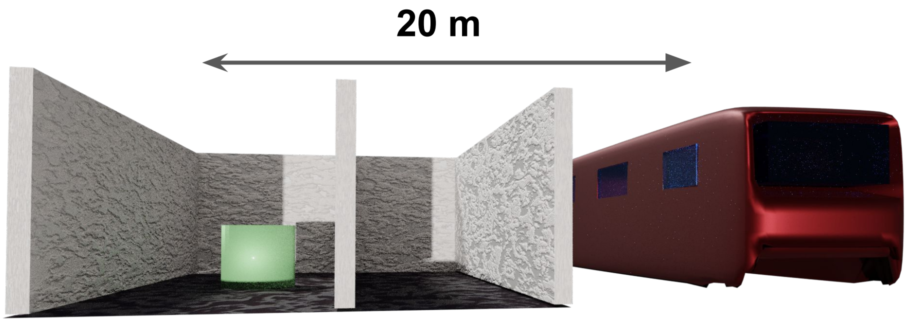
10x24 hour-long waveforms were recorded. These waveforms (which essentially are the records of passing trains) were split into consecutive windows of 2000 samples each with a 50% overlap. No bandpass filter was applied. Samples containing just the noise were rejected

Windows were randomly shuffled and combined into pairs. Separate channels in each pair were summed to obtain a waveform mixture.



Data example: Source 1, Source 2, Mixture

[Raspberry Shake sensor]()



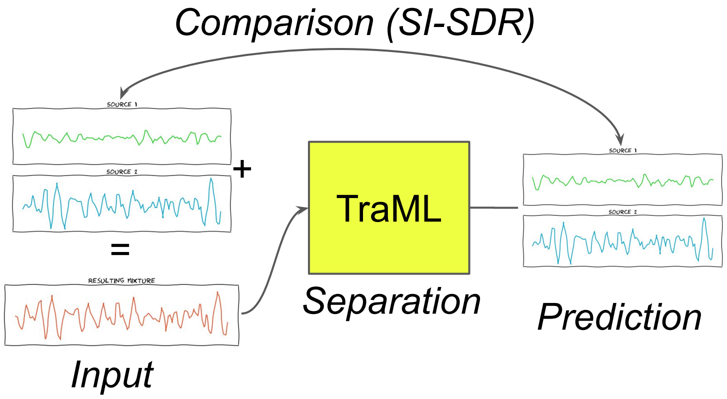
Room with RS sensor   *Corridor*   *Railway tracks*

Training

PyTorch was used as a framework for model building (Shi et al. 2020) and training. The model was trained on GPU provided by Google Colaboratory.

The training objective was to minimize Scale-Invariant Source to Distortion Ratio (SI-SDR) between original individual sources and predicted by the model waveforms since this metric is widely used as a source separation performance indicator. Waveform mixtures were L2-normalized before summation, for each epoch, 10000 sample combinations were drawn uniformly from the database consisting of 5644 unique data windows.

The initial learning rate of 1e-3 was decaying by a factor of 0.98 every 2 epochs. After 100 epochs average SI-SDR of ~ **-4.5 dB** was achieved

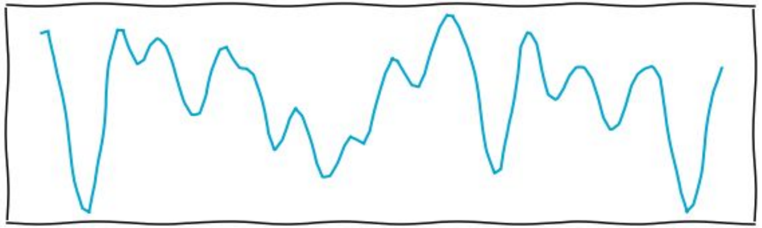


Training pipeline.

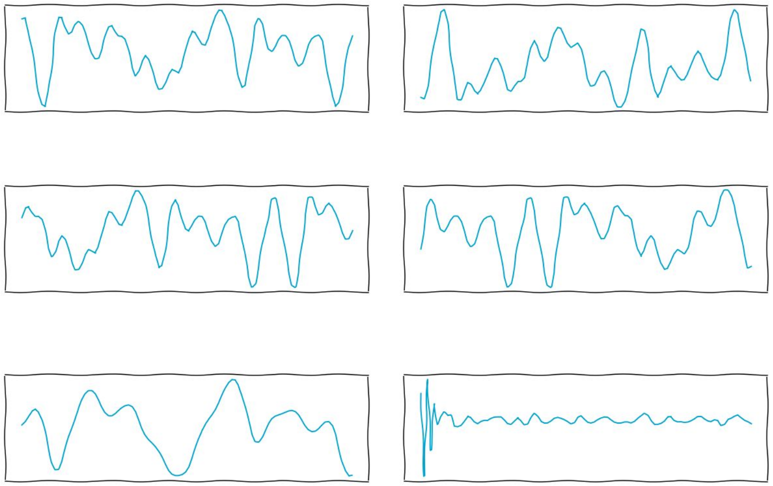
Data augmentation

To achieve better generalization, the following augmentations were applied to each signal composing the mixture:

- Random polarity change
- Apply random roll (shift in samples)
- Apply either low-pass or high-pass with random frequency below Nyquist frequency, or keep signal intact



Input signal

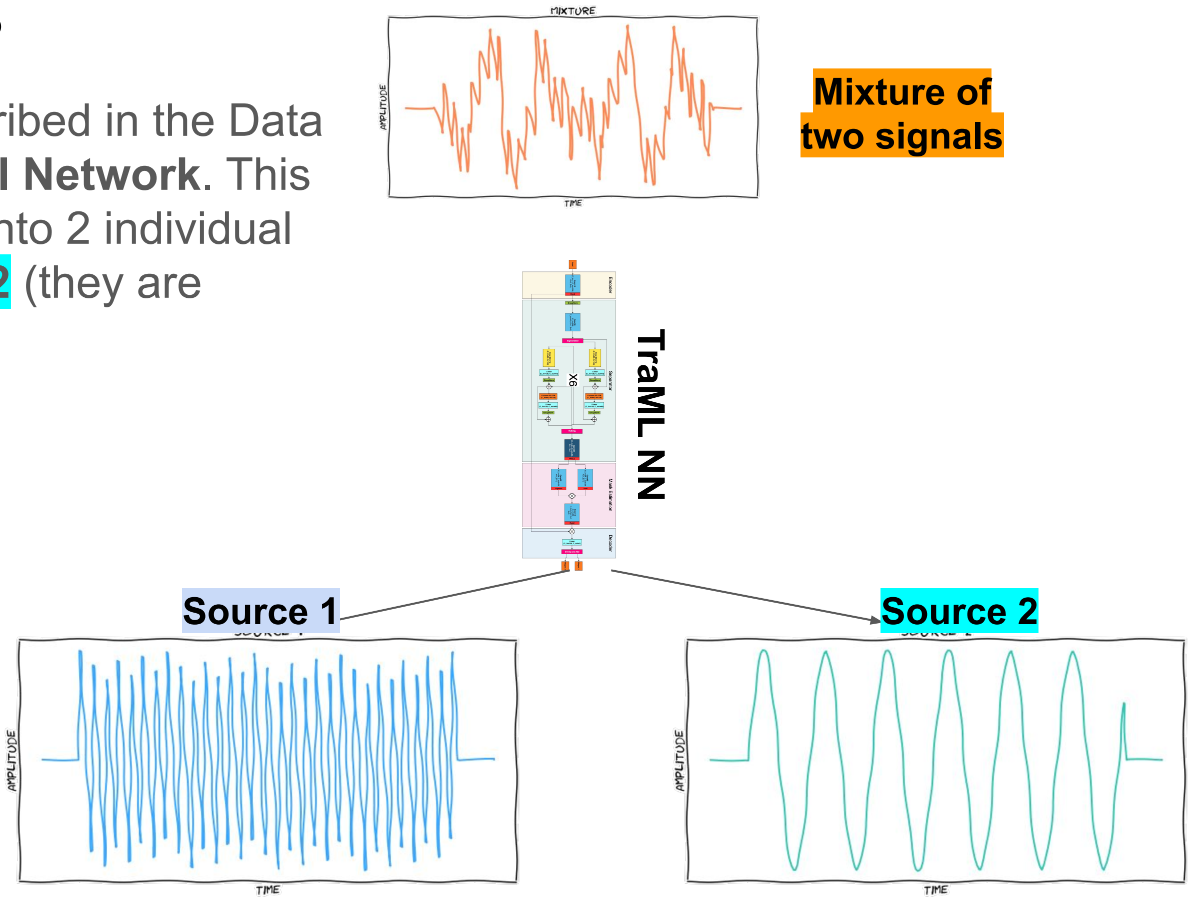


Augmented signal

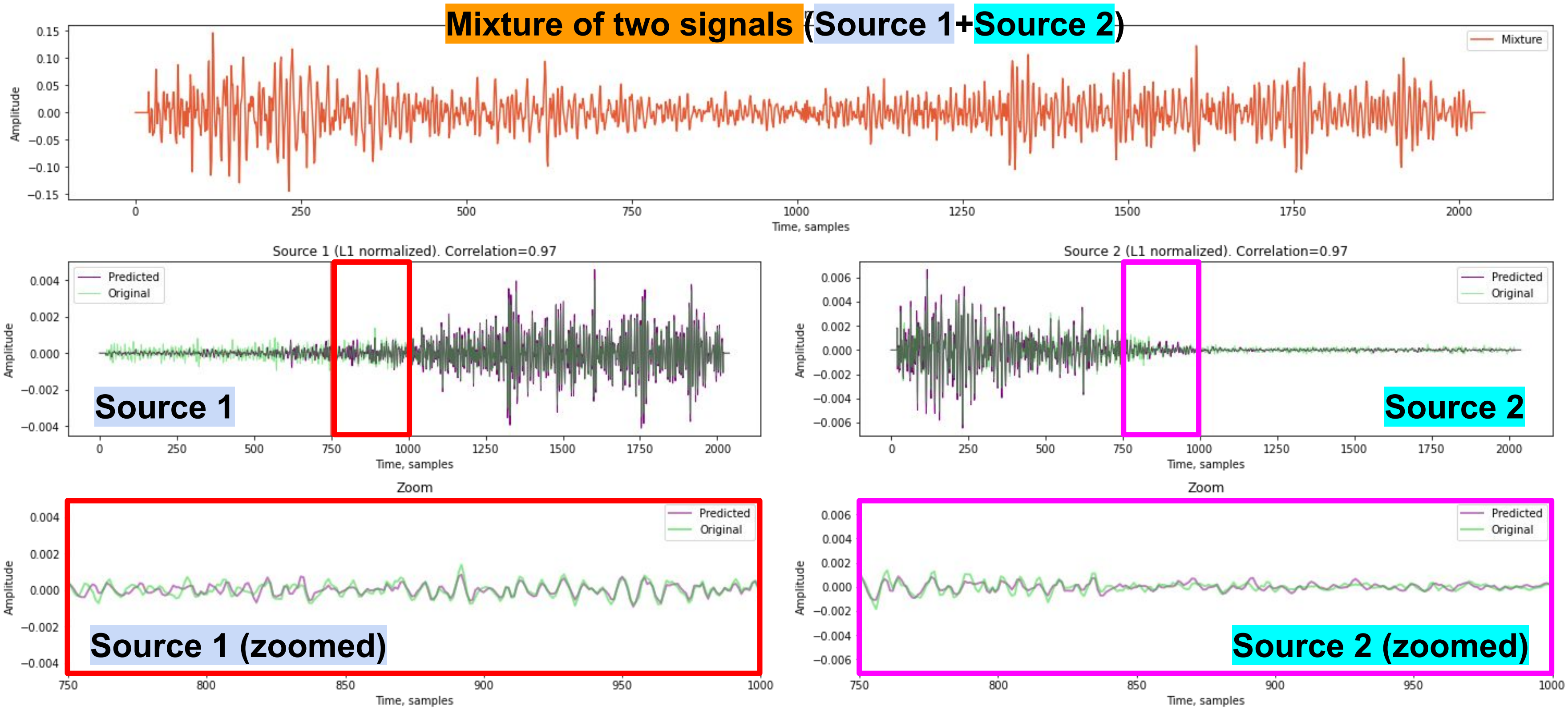
Separation process

A mixture of two signals described in the Data section is fed into **TraML Neural Network**. This network separates the mixture into 2 individual signals: **Source 1** and **Source 2** (they are generated by the network).

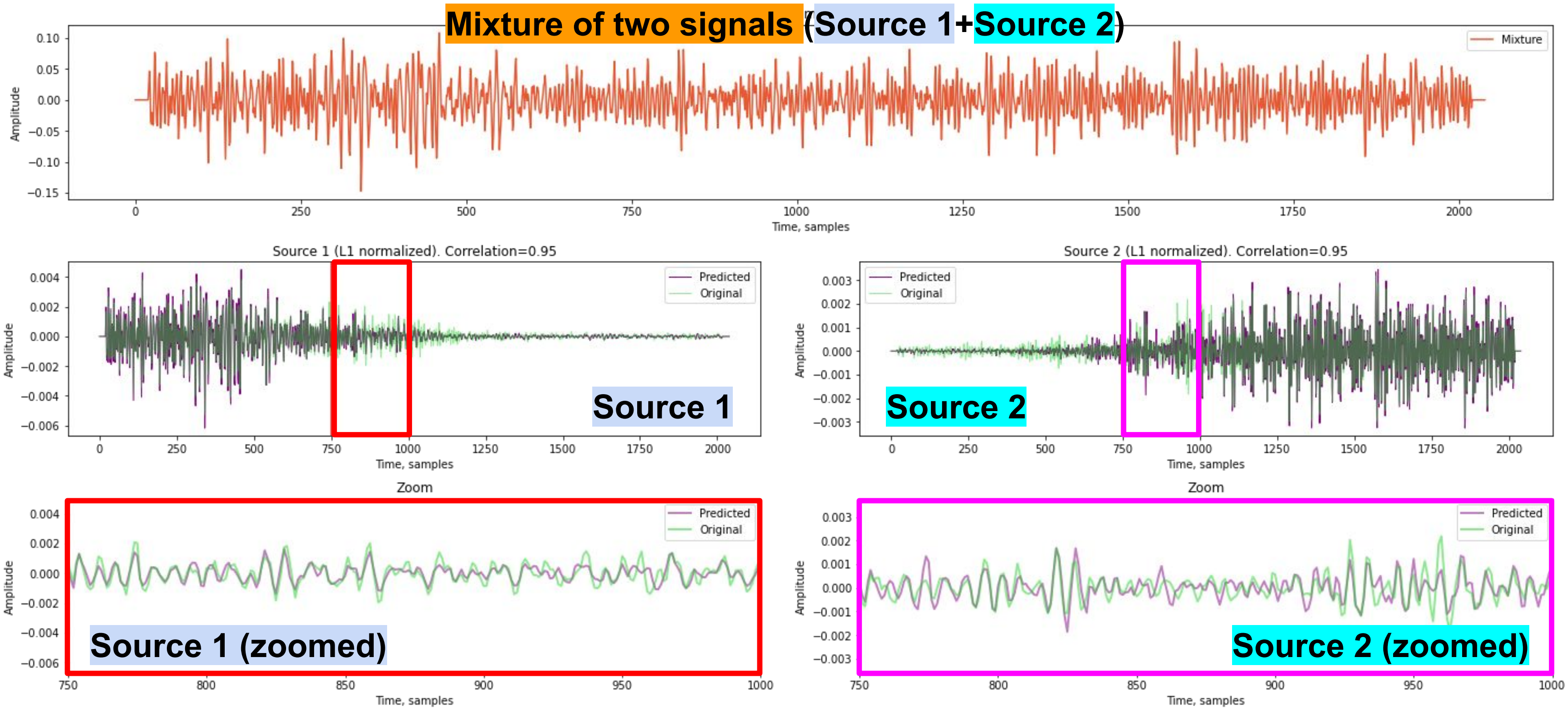
Those separated signals (predicted by the Neural Network) are then compared against their original counterparts. A correlation coefficient is demonstrated in the figure caption. Please note, that since TraML operates on Scale Invariant SDR, the amplitude output of the network is not to scale, and therefore has to be L2-normalized for a fair comparison.



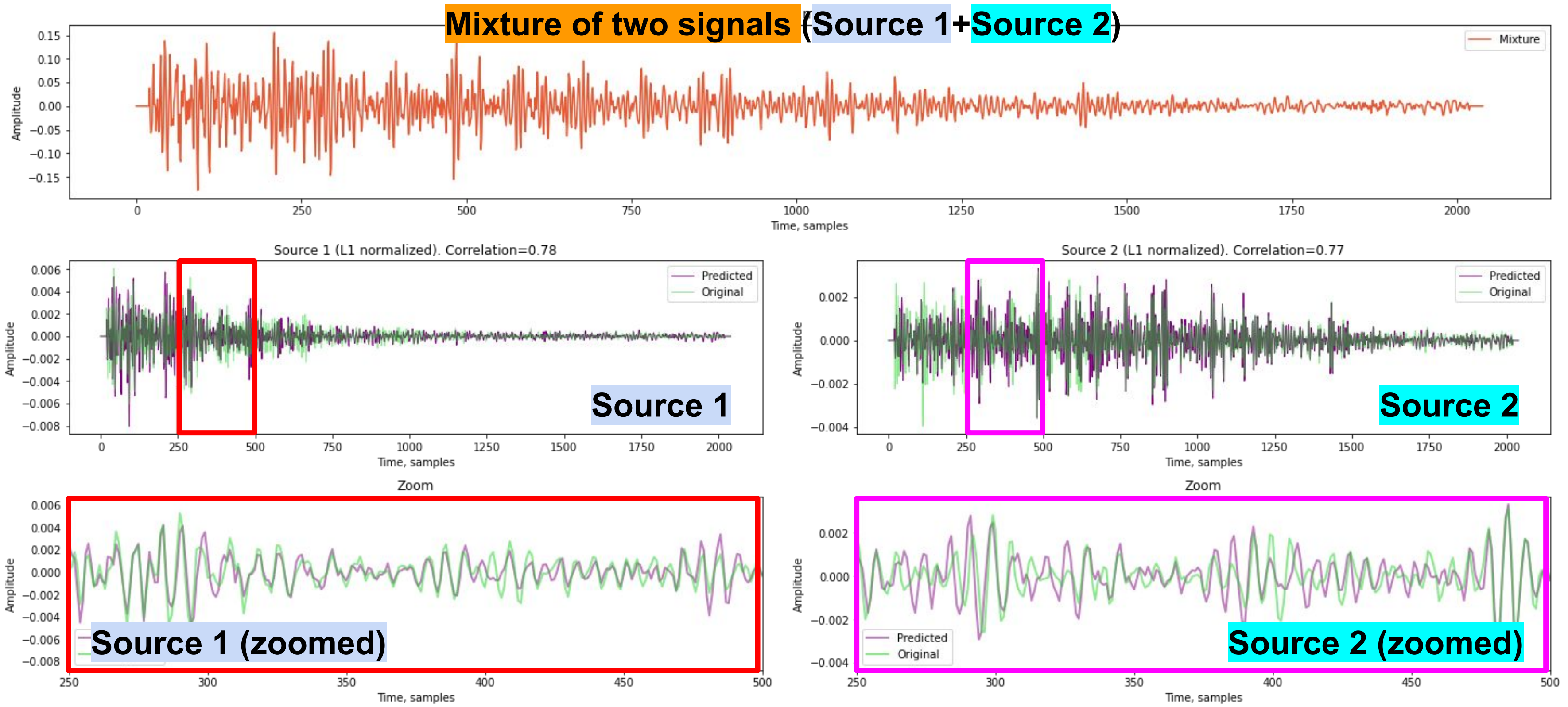
Results (Not overlapping sources)



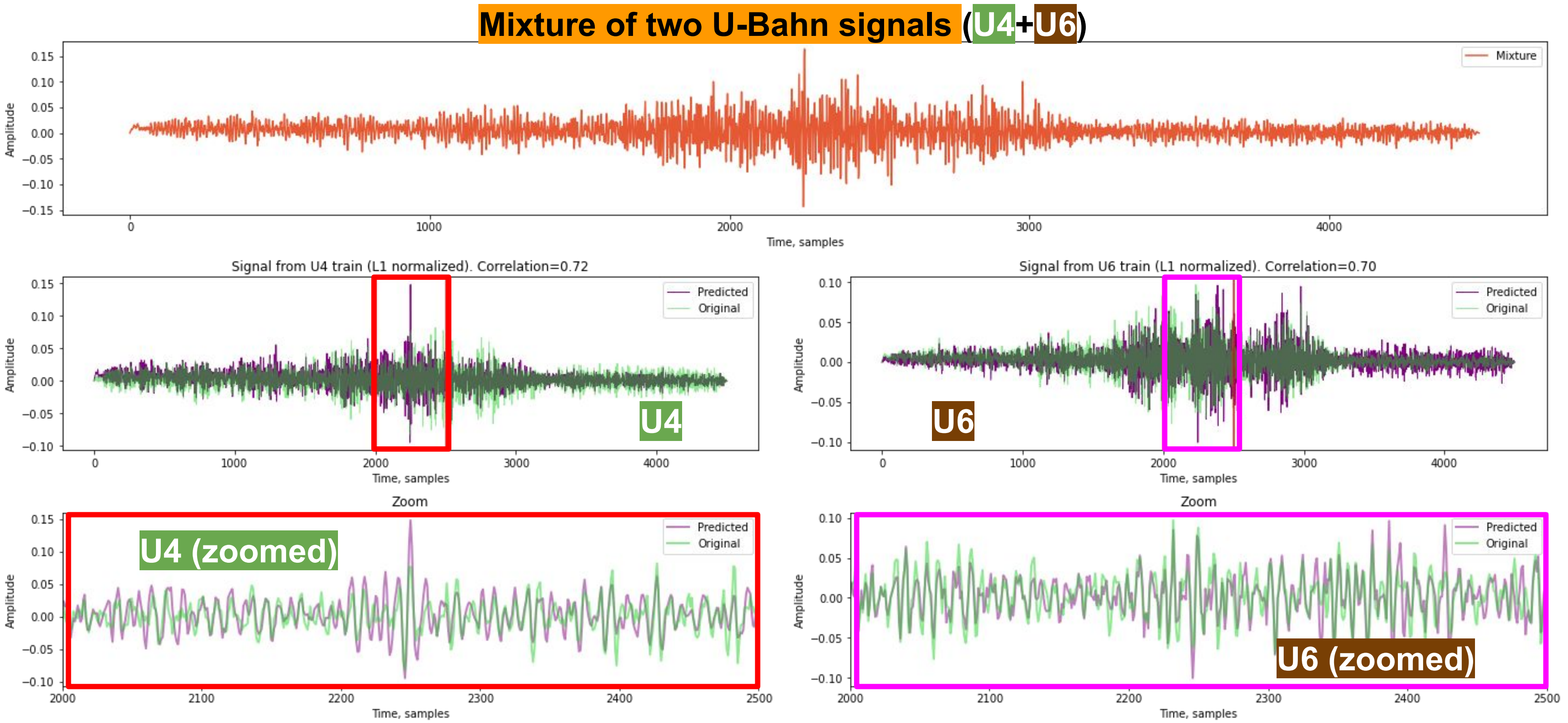
Results (Slightly overlapping sources)



Results (Strongly overlapping sources)



U4 - U6 U-Bahn separation example



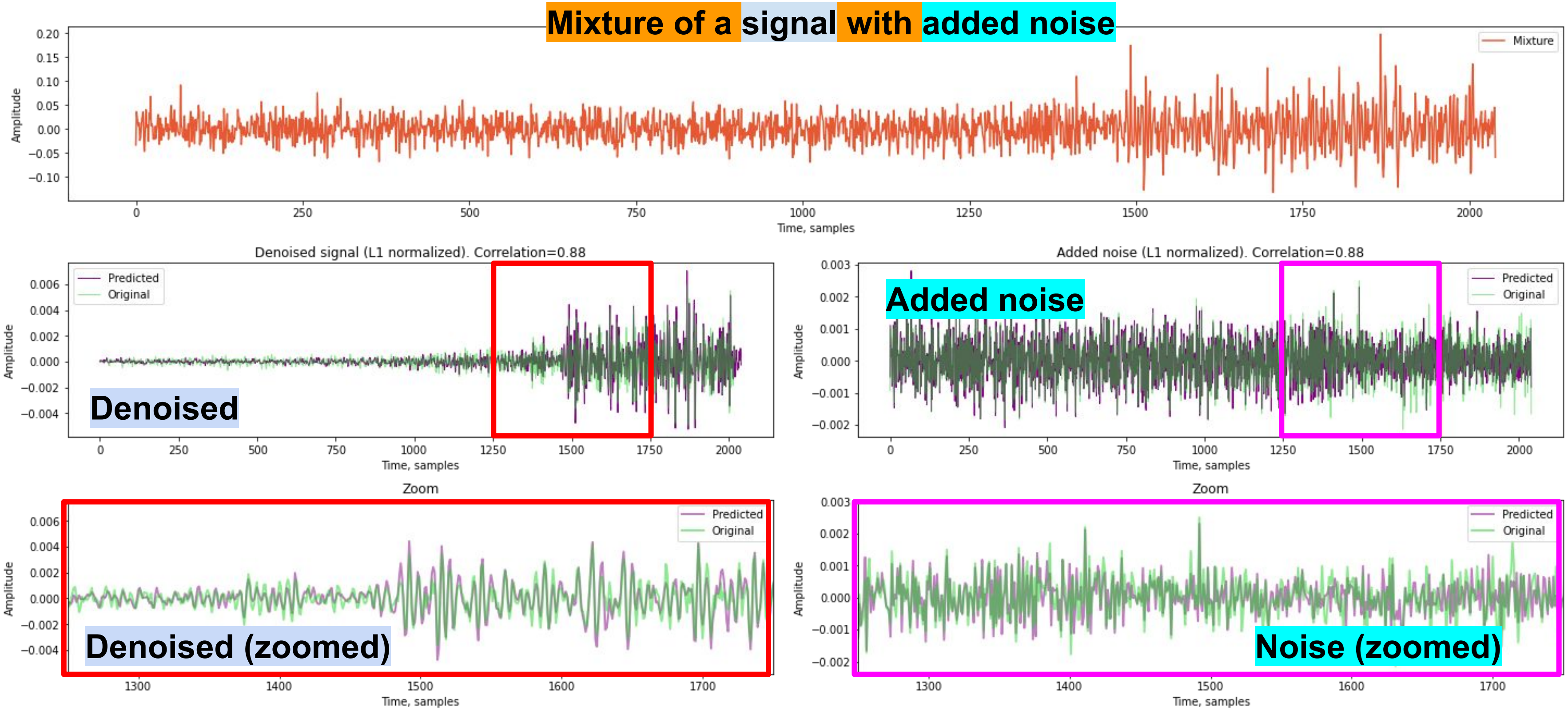
Fine-tuning to other applications

In practice, deep neural networks like the one, presented here, have a very big number of parameters (*2.633.729 in TraML*). Training Neural Networks on an insufficiently small dataset greatly affects the ability to generalize, often result in overfitting.

More often in practice, models get fine-tuned in existing pre-trained networks by continued training on the smaller dataset at hand. Provided that the nature of the dataset does not significantly differ from the original dataset, the pre-trained model will already have learned features that are relevant to the problem at hand (e.g. reading and generating waveforms).

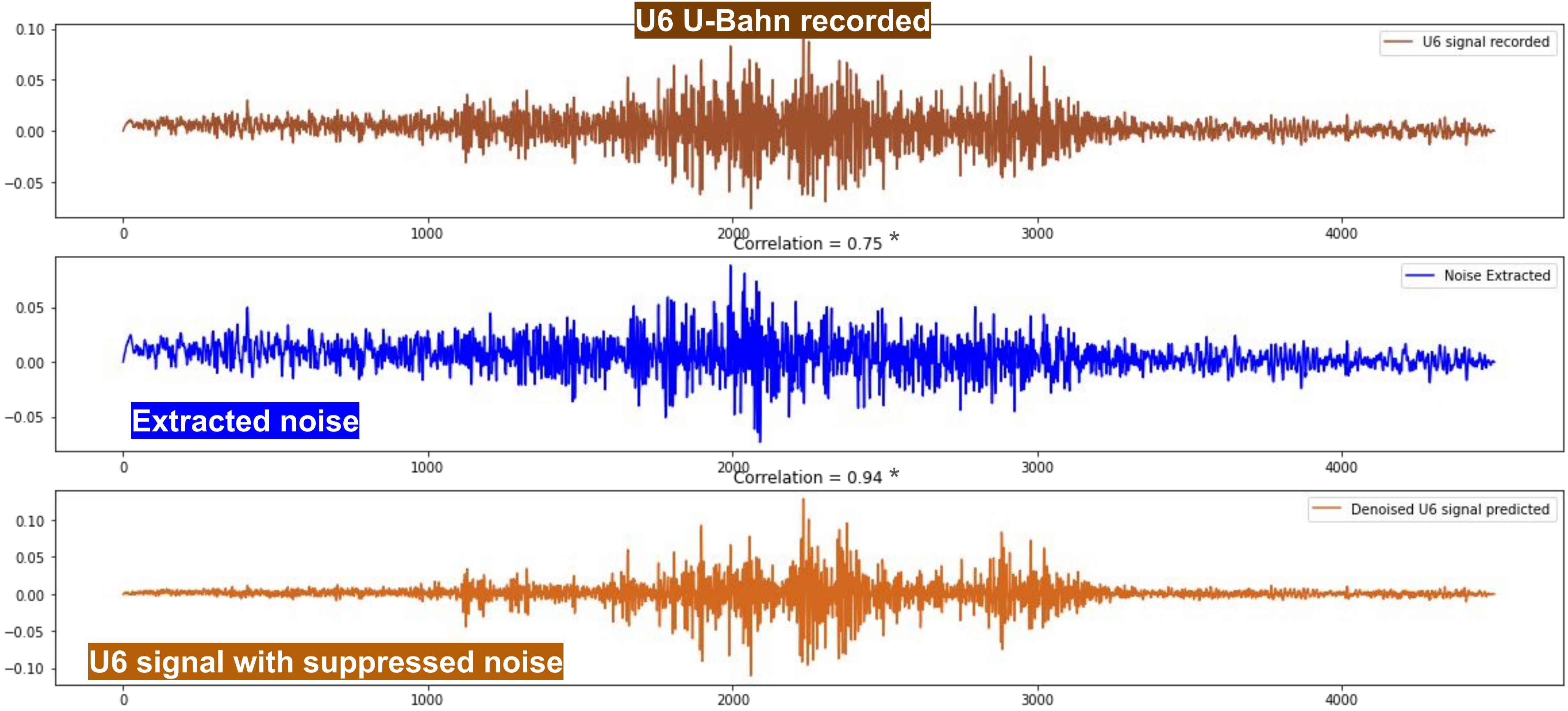
In the particular case of TraML, which was originally trained for source separation, one can fine-tune it to denoise the data.

Noise-suppression



U4 denoising example

* Correlation values are shown with the respect to the input U6 record



Conclusions

The task of source separation that was previously considered practically not solvable (conventional methods are giving very poor results), can now be solved with advanced methods such as Machine Learning. Such methods can perform not only source separation but also noise suppression.

The generative architecture of autoencoders has proven itself under various applications; it is now entering the seismology domain; it can help to better understand subsurface processes by obtaining cleaner data.

As we optimize our models with more data and compute we aim to achieve even better results in the observable future.