Extending the reach of gravitational-wave detectors with machine learning

Nonlinear noise regression

Tri Nguyen [1] Michael Coughlin [2] Rich Ormiston [3] Rana Adhikari [2]

September 16, 2018

[1] University of Rochester

[2] California Institute of Technology

[3] University of Minnesota – Twin Cities

Table of contents

Background and Motivation

What is LIGO?

- **LIGO** is the world's largest gravitational-wave detector.
- LIGO detects gravitational waves by measuring the **differential arm length** (DARM) between the its perpendicular arms.

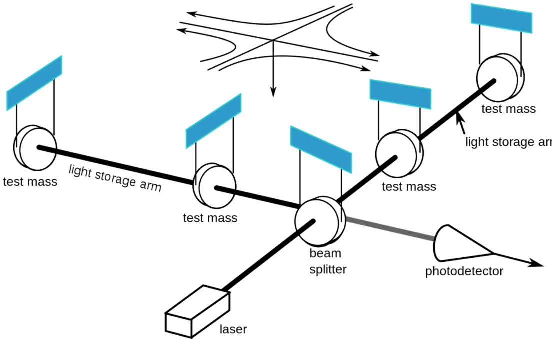

Figure 1: Schematic of the basic LIGO interferometer design.

Where is LIGO?

Figure 2: The twin LIGO detectors at Livingston, Louisiana and Hanford, Washington are separated by a distance of 3002 km.

- **Gravitational waves** (GWs) are ripples in spacetime created by any asymmetrical acceleration.
- Strong GWs are created by the most violent events in the universe (e.g. the Big Bang, black-hole mergers, etc.).
- GWs carry information (e.g. mass, spin, etc.) about their progenitors.

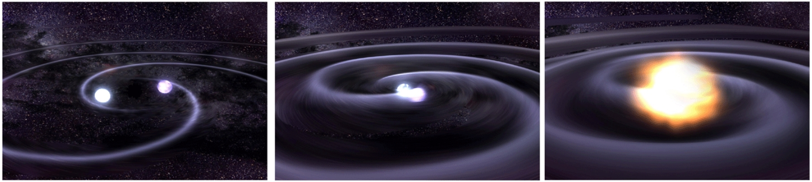

Figure 3: Gravitational waves created by an inspiral compact binary black hole system.

Detecting gravitational waves

- Strong GWs produce a displacement of 10^{-18} m, about 1000 times smaller than the diameter of a proton.
- Although LIGO has observed GWs from black-hole and neutron-star mergers, many more still lie below the sensitivity limit.

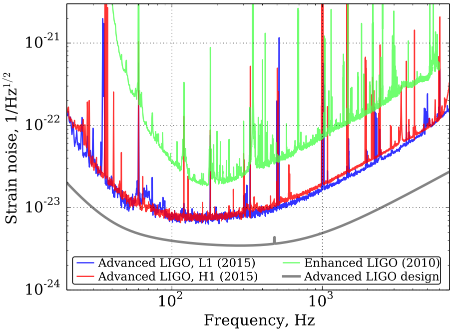

Figure 4: The sensitivity of LIGO Livingston (L1) and LIGO Hanford (H1) during the first observation run O1

Objectives

Nonlinear noise regression

- Noise sources are recorded as time series data by physical environmental monitor channels, or **witness channels**.
- Based on the detector's response history, noise regression predicts future response and subtracts the sources.
- Noise sources couple into the output through some physical processes described by a **transfer function**.
- If the transfer function is nonlinear, traditional noise regression can be challenging because the coupling mechanism is often sophisticated.

Goal: Train neural networks to perform linear and nonlinear subtraction on time series data.

Removable v. Non-removable noises

- **Removable noises** define the baseline sensitivity limit. They can only be reduced by improving the detector design.
- Examples: quantum noise, thermal noise, etc.
- **Non-removable noises** are instrumental and environmental effects. They can be subtracted given there are witness channels monitoring them.
- Examples: seismic noise, magnetic noise, etc.

Goal: Subtract the non-removable noises while keeping the signals and the removable noise intact.

- We perform our analysis on mock and real data.
- Real data are LIGO Hanford (LHO) data during the first observation run on August 14, 2017.
- Mock data are generated by coupling white noises $w(t)$ into the DARM $h(t)$ via the **resonance function**:

$$h(t) = \mathcal{F}^{-1}\left[\frac{\mathcal{F}[w(t)]}{\omega_0^2 - \omega^2 + i\frac{\omega_0\omega}{Q}}\right]$$

where ω_0 and Q are the angular resonant frequency and the quality factor. $\mathcal{F}$ denotes the Fourier transform.

Visualizing the data

- The input consists of time series from multiple channels.
- Each channel has a duration of 2048 seconds and a sample rate of 512 Hz (total of 1,048,576 samples).

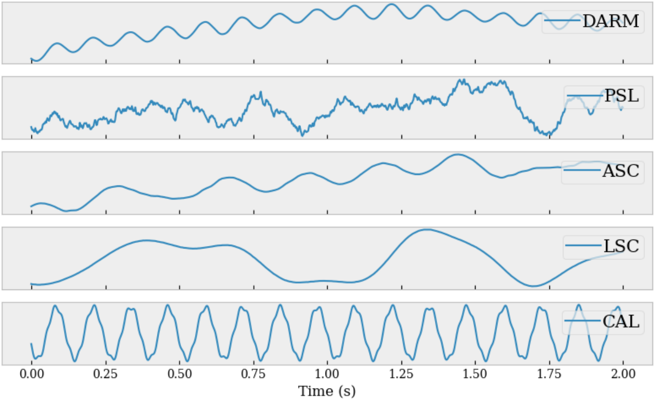

Figure 5: Sample witness channels by subsystems from LHO data on August 14, 2017.

Visualizing the data

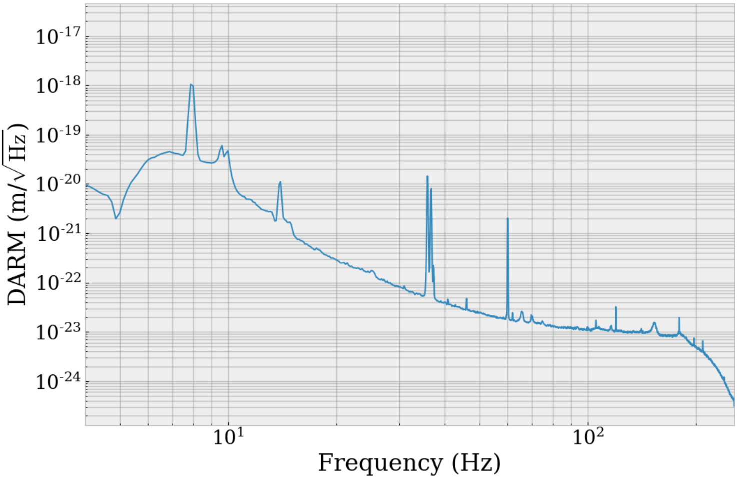

Figure 6: The power spectrum density of LHO data on August 14, 2017.

Methods

Data preprocessing

- Sample-rate conversion: Up-sample or down-sample so all channels have the same sample rate of 512 Hz.
- Bandpass filter: Attenuate frequencies outside a certain range.
- Standard scaling: Normalize the mean and standard deviation of each channel to 0 and 1.
- Lookback window: Divide each channel into short and overlapping series. There are 1,048,546 mini series, each with 16 data points (or 0.03125 seconds).

Neural networks

- **Neural networks** are a set of non-parametric, supervised machine learning algorithms.
- Neural networks can learn and perform tasks such as data classification and regression.
- A neural network may consist of thousands of interconnected nodes. Each node has parameters to characterize the data.

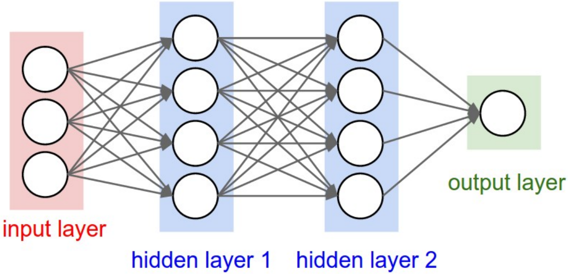

Figure 7: Neural networks block diagram.

Long Short-Term Memory networks

- **Long Short-Term Memory** (LSTM) networks are special neural networks to process sequential inputs, like time series data.
- A major strength of LSTMs is the ability to store and use information from past inputs.

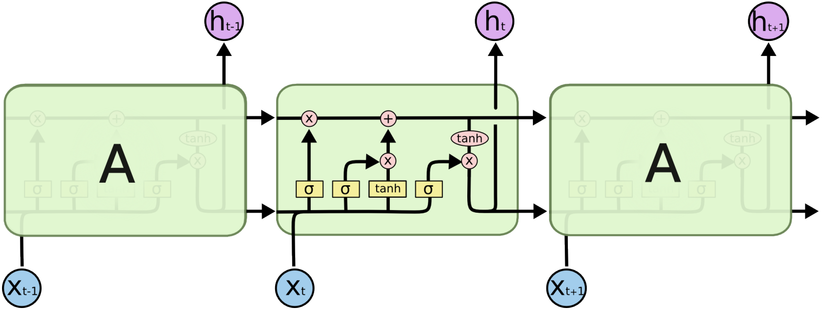

Figure 8: A typical LSTM layer contains four interacting sub-layers.

Why do we need LSTMs?

- LIGO noise budget includes slowly migrating signals.
- Example: seismic waves at test masses take several seconds to get to the witness channels.
- Unlike standard neural networks, LSTMs can capture the long-term dependencies in the data.

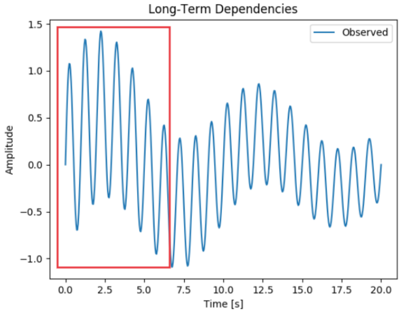

Figure 9: Seeing only the data inside the red box, an LSTM can still predict the time series.

$$\text{Observed} = e^{-0.1t}\left(\frac{3}{5}\sin(10t) + \sin(t)\right)$$

DEEP CLEAN

Basic workflow

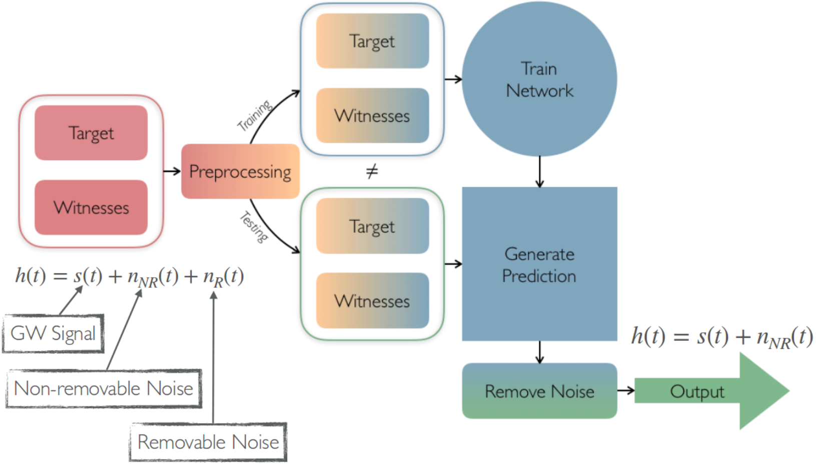

Figure 10: The noise regression pipeline of DeepClean.

Inside each network

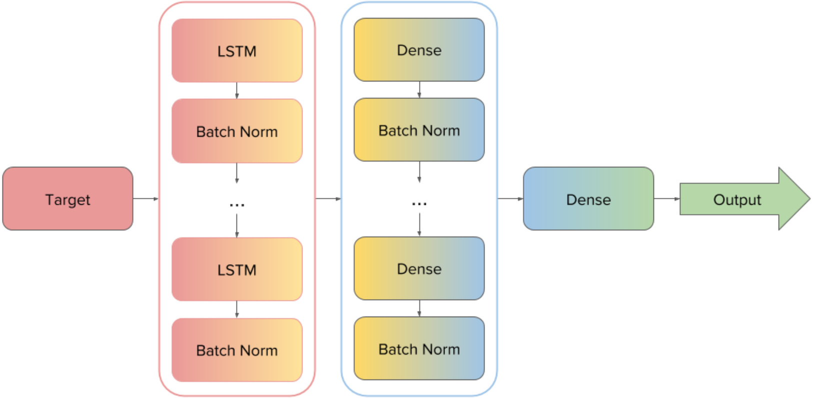

Figure 11: The architecture of each DeepClean network.

Results

DeepClean successfully subtracted the resonant noise ($\omega_0 = 13.15$ rad/s, $Q = 1000$) and left out the 60-Hz AC power line and the rest of the background (bucket noise).

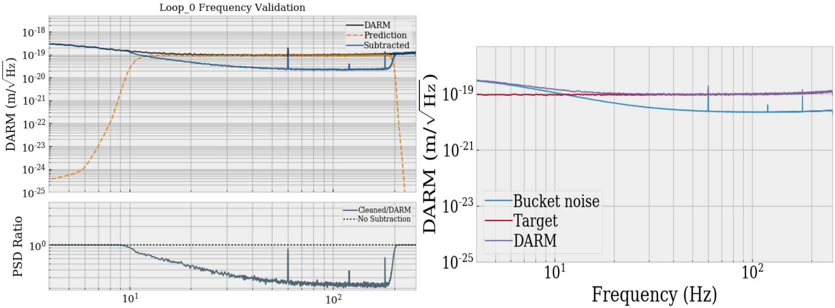

Figure 12: Left: The subtraction in the 10-250 Hz band. Right: The expected outcome.

LHO data

DeepClean successfully subtracted the 7-Hz calibration line and 60-Hz AC power line.

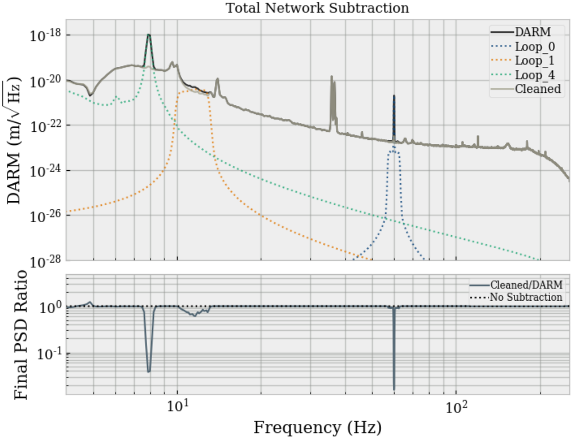

Figure 13: Total subtraction.

LHO data

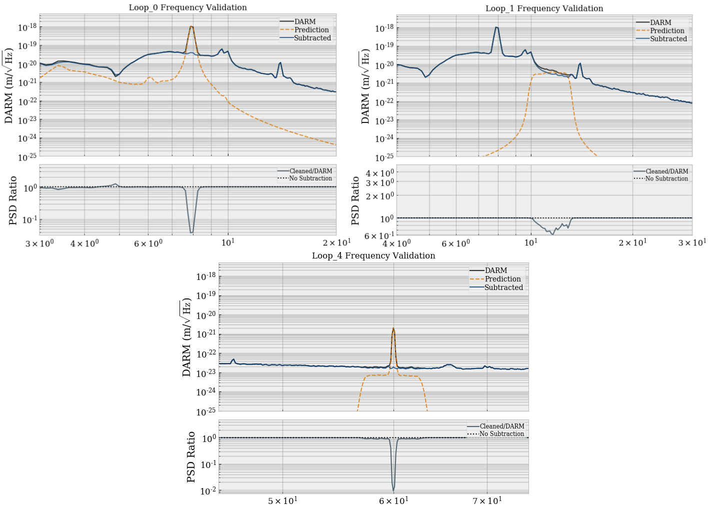

Figure 14: Subtraction on the 3-9 Hz (top left), 10-13 Hz (top right), and 57-63 Hz (bottom) bands.

Conclusion

Future work

- Obtain better subtraction in the 20-100 Hz band $\Rightarrow$ Discover more compact binaries and uncover additional information on past observations.
- Understand which witness channels are important through an ablation and/or correlation study.
- Optimize for real-time noise regression (i.e. analyze an hour of data in less than an hour).

Thank You

TO my mentors, Michael, Rich and Rana. FOR your guidance.

TO my besties, Alvin, Vinny, Sky and Disha. FOR being there.

TO my friends, the LIGO SURF group. FOR being supportive.

TO the Caltech SURF program. FOR the opportunity.

TO Los Angeles. FOR changing my life.