G2Net

Gravitational Wave Detection

Find gravitational wave signals

from binary black hole collisions

Kaggle competition

Table of Content

[Introduction 3](#_Toc76826386)

[Story 3](#_Toc76826387)

[Objective 3](#_Toc76826388)

[Inventory 4](#_Toc76826389)

[Kaggle 4](#_Toc76826390)

[Local 4](#_Toc76826391)

[Framework 4](#_Toc76826392)

[Data 4](#_Toc76826393)

[Workplan 4](#_Toc76826394)

[Data ingestion 4](#_Toc76826395)

[Analysis 4](#_Toc76826396)

[notebooks 4](#_Toc76826397)

[Modeling 4](#_Toc76826398)

[Submission 4](#_Toc76826399)

[Conclusion 4](#_Toc76826400)

[Data ingestion 5](#_Toc76826401)

[Analysis 5](#_Toc76826402)

[EDA 5](#_Toc76826403)

[Spectrogram 5](#_Toc76826404)

[Notebooks 5](#_Toc76826405)

[Starter 5](#_Toc76826406)

[Inference 5](#_Toc76826407)

[Modeling 5](#_Toc76826408)

[Submission 5](#_Toc76826409)

[Conclusions 5](#_Toc76826410)

[Conclusions 6](#_Toc76826411)

[Annexes 6](#_Toc76826412)

# Introduction

## Story

Inspired from:

<https://en.wikipedia.org/wiki/First_observation_of_gravitational_waves>

In 2015 September the 14th, the first direct observation of gravitational waves happened and was made official the 11th of February 2016.

This event had been recorded by the two new Laser Interferometer Gravitational-Wave Observatory (LIGO) located in the United States of America (Hanford and Livingston).

A third Large Interferometer named Virgo located in Europe was upgraded shortly after improving the Gravitational Waves potentialities.

Since then, these 3 Larges Interferometers are actively searching for detectable Gravitational Waves (GW).

In short, the interferometers are built around an L structure where laser emits photon along the few kilometer’s arms, these photons are reflected at the extremities, then is measured their time of arrival.

When a strong enough event happens the GW impacts photons travel time (rather curve the space hence modify the travel length, hence modify the time of arrival).

By strong event here we are talking of black hole collision or at least neural star collision. Yes, this universe is kinda scary…

## Objective

Long story short, the objective is predicting whether 3 channels recorded events match gravitational wave. A simple yes or no binary classification problem.

Quoting Kaggle:

<https://www.kaggle.com/c/g2net-gravitational-wave-detection/data>

In this competition you are provided with a training set of time series data containing simulated gravitational wave measurements from a network of 3 gravitational wave interferometers (LIGO Hanford, LIGO Livingston, and Virgo). Each time series contains either detector noise or detector noise plus a simulated gravitational wave signal. The task is to identify when a signal is present in the data (target=1).

The parameters that determine the exact form of a binary black hole waveform are the masses, sky location, distance, black hole spins, binary orientation angle, gravitational wave polarization, time of arrival, and phase at coalescence (merger). These parameters (15 in total) have been randomized according to astrophysically motivated prior distributions and used to generate the simulated signals present in the data, but are not provided as part of the competition data.

## Inventory

### Kaggle

Kaggle provide with cloud (remote) Jupyter notebooks, 16GB RAM, multi-processor, 20GB local space.

Kaggle notebooks can be GPU enhanced but limited to 35 hours per week, counter is running for the notebooks not the GPU usage time; therefore, to be activated cautiously, that’s why the below local environment will be leveraged for most of the trainings.

Also note when running in GPU mode, less CPU are available so care with heavy CPU preprocessing, in fact we’ll rely on intermediate preprocessed data.

### Local

A local computer based on CPU Intel 4790K, 16GB RAM, GPU RTX2060, and SSD.

### Framework

Jupyter python, with TensorFlow Framework chosen as per my competencies.

Leveraging at most the TensorFlow API including Datasets, addons, and integrated Keras is a personal bet I hope will be rewarding in the future, but may be tricky at some point due to the Kaggle still in 2.4.0 Tensorflow update level.

### Data

There are 560 000 numpy records available for training.

Each data sample (npy file) contains 3 time series (1 for each detector), each spanning for 2 sec, and sampled at 2,048 Hz. Something like 96KB data per sample.

That’s around 60GB to deal with when including test data.

### Competition

Root of the competition is at <https://www.kaggle.com/c/g2net-gravitational-wave-detection>

Started the 30th of June 2021

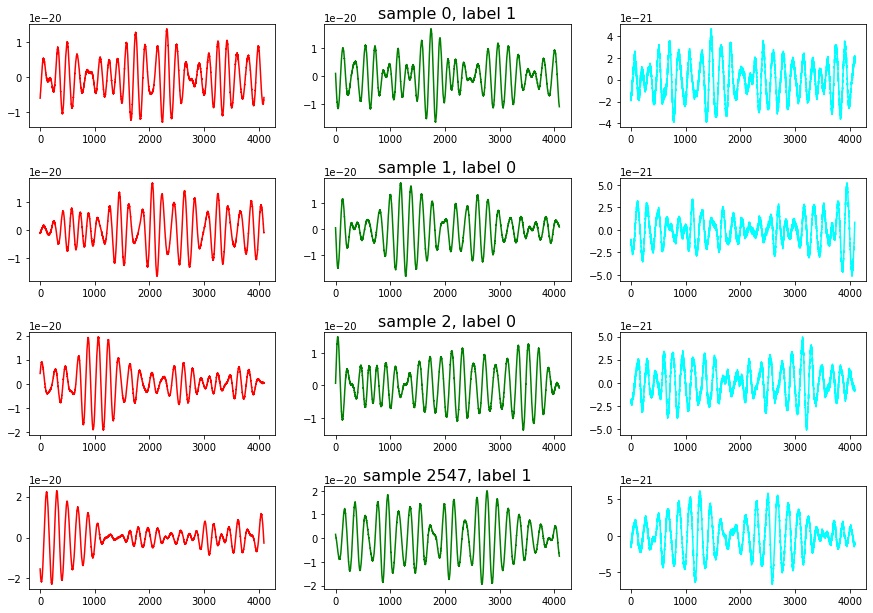
Runs for 3 months

# Data ingestion

# Analysis

## Simple EDA

Below are a few samples plotted:



## Spectrogram

Quote from Kaggle:

“The integrated signal-to noise ratio (SNR) is classically the most informative measure of how detectable a signal is and a typical level of detectability is when this integrated SNR exceeds ~8. This shouldn't confused with the instantaneous SNR - the factor by which the signal rises above the noise - and in nearly all cases the (unlike the first gravitational wave detection GW150914) these signals are not visible by eye in the time series.”

# Notebooks

## Starter

Kaggle notebook available at <https://www.kaggle.com/vincentdumetz/g2net-tf-dataset-starter>

## Inference

Kaggle notebook available at <https://www.kaggle.com/vincentdumetz/g2net-tf-dataset-inference>

# Modeling

# Submission

# Conclusions

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| function | alpha | Validation accuracy | Batch to train | Time to train | Batch time (m:ss) |
| Leaky ReLU (Baseline) | .3 | **.5608** | 32 | 2h59 | **5:59** |
| ReLU | N/A | .4271 | 25 | 2h29 | 5:96 |
| PReLU | N/A | .4192 | 27 | 3h49 | 8:48 |
| ELU | .25 | .5342 | 32 | 4h02 | 7:56 |
| GELU | N/A | .4896 | 32 | 4h24 | 8:25 |
| SELU | N/A | .4567 | 32 | 3h07 | 5:84 |
| SWISH | N/A | .4671 | 32 | 3h27 | 6:46 |