



# G2NET KAGGLE COMPETITION

INITIAL OUTCOME

By Christopher Zerafa

# OVERVIEW

- What? Why?
- Competition
- Results
- Statistics
- Way forward

## WHAT? WHY?

- *Subsidiary of Google LLC*
- *Online community of data scientists and machine learning practitioners*
- *Allows users to find and publish data sets, explore and build models in a web-based data-science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges.*
- *In June 2017, Kaggle passed 1 million registered users, spanning 194 countries.*
- *Competitions regularly attract over a thousand teams and individuals, with cash prizes.*
- *In the past, it has led researchers to publish papers in peer-reviewed journals based on work done during these competitions.*

kaggle™

### ***Ties in with CA17137 – g2net Objectives:***



*Geographical Spread*



*Inclusiveness*



*Academic Pursuit*

# COMPETITION

Research Prediction Competition

## G2Net Gravitational Wave Detection

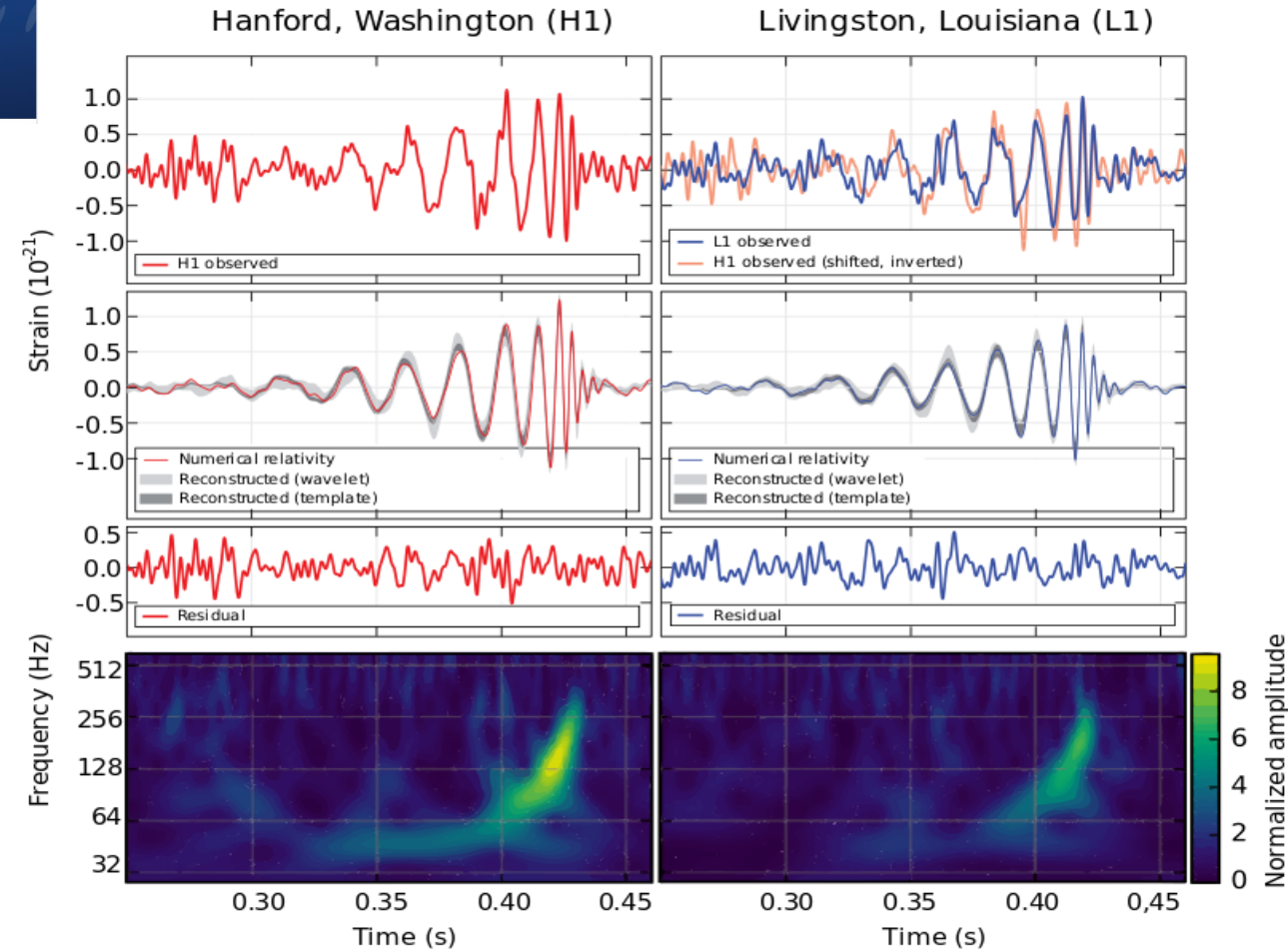
Find gravitational wave signals from binary black hole collisions



European Gravitational Observatory - EGO · 1,219 teams · 8 days ago

**\$15,000**  
Prize Money

- Held between 30 June 2021 and 30 September 2021
- Available at <https://www.kaggle.com/c/g2net-gravitational-wave-detection>
- **Task:** Build a model to analyze simulated GW time-series data from a network of Earth-based detectors
- **Evaluation:** Submissions evaluated on area under the ROC curve between the predicted probability and the observed target.



# RESULTS

0.88538, 0.88299, 0.88291 ROC result of top 3 team

## Architecture

- A lot of results used EfficientNet
- Vision Transformer (ViT)
- Simple CNN

## Approaches

- With or without whitening of data (This was not something that any of the GW specialist recommended, but the community found literature to support this)
- Data management via tfrecords
- Pre-processing using CWT, Bandpass filtering,

## Computational Power

- Some used Kaggle environments
- Some used personal 2080Tis
- Some had 10s of V100

## STATISTICS

*1,200 teams*

*1,537 individual competitors*

*29,590 submissions*

*\$15,000 prize money sponsored by Google*

77.38 GB of Data...downloaded by 1537 participants.....116TB of GW data

1305 blog post

## WAY FORWARD

- **Report Initial Outcomes**
- **Survey the participants:** A survey will be sent out to all participants to be able to gauge statistics. Topics that will be included are: Gender equality, Country of Origin, Interest in participating in Cost Action activities, Accessibility to Computational Resources Necessary for Competing
- **Evaluate the Outcome:** Following the closing of the competition, all evaluations are to be evaluated and the best performing teams will be identified. The best performing teams will provide documentation, steps to reproduce the results and necessary the code repositories.
- **Publishing of Results:** The best performing teams will then be asked for an interview to present methodology and summary of findings. We will try and do this as a Cost Action event, and have it published as part of our Outreach programme.
- **Further Commercial Applications:** The main results achieved will be investigated further and potentially expanded further with other commercial Partners. For example, NVIDIA Applied Research Accelerator Program (<https://www.nvidia.com/en-us/industries/higher-education-research/applied-research-program/>) offer a programme to accelerate academic ideas with high potential.
- **Next Competition:** Looking at the initial numbers, the outreach achieved, and the results submitted by the community, there is a likelihood that we will attempt to do another competition. The next competition will use a more challenging dataset to classify GW or for the classification of seismic noise prone to interfere with GW sensors.

# ACKNOWLEDGMENTS

- Christopher Messenger
- Michael Williams
- Maggie Demkin
- Walter Reade
- Dr Elena Cuoco





THANK YOU  
FOR YOUR  
ATTENTION

ANY QUESTIONS