GWSkyNet Output Guide

The LSC-Virgo-KAGRA Collaboration

Understanding the GWSkyNet output

GWSkyNet is a machine learning signal-vs-noise classifier designed to facilitate potential electromagnetic follow-up observations [1]. GWSkyNet can be used to screen for noise events or noise-influenced sky maps using a universal metric (calculated using the sky localization information and Flexible Image Transport System (FITS) file metadata) that can be interpreted in the same way for all multi-detector gravitational wave (GW) candidates across all search pipelines. GWSkyNet provides information complementary to what was available in O3 in low-latency by exploring a different area of the analysis parameter space than p_{astro} and false alarm rate from search pipelines.

GW search pipelines compute a false alarm rate and corresponding p_{astro} based on studies of how frequently the pipeline is triggered by noise[1]. While the false alarm rate from search pipelines and p_{astro} remain important metrics, GWSkyNet provides an additional metric, which is computed using the sky localization information and FITS file metadata. GWSkyNet provides a complementary perspective to help reduce false positives attributed to detector noise.

For an event candidate identified with a network signal-to-noise ratio of ≥ 7 and a signal-to-noise ratio of ≥ 4.5 in at least two detectors, GWSkyNet will provide three quantities in a `json` file: GWSkyNet score, false alarm probability (FAP) and false negative probability (FNP), explained below. The `json` file will be uploaded to and accessible on GraceDB[2].

The GWSkyNet score is a measure of the model's confidence that a candidate belongs to the noise or signal class: 0 is consistent with noise, and 1 is consistent with a signal. The score is *not* the probability that this particular candidate is astrophysical; we would need to account for the true distribution of noise events and signals (which we do not assume).

To provide context for how often noise (or a signal) would produce an equivalent GWSkyNet score, GWSkyNet also outputs the corresponding **FAP** and **FNP** for a candidate's score.

- FAP measures the fraction of noise events at or above a candidate's GWSkyNet score.
- FNP measures the fraction of astrophysical events at or below a candidate's GWSkyNet score.

This empirical measurement (see Figure 1) is derived from GWSkyNet performance on known LIGO-Virgo detector glitches and simulated gravitational wave signals (recovered by one or more searches as superevents, below search pipeline false alarm rate of 1 per hour).

[1] https://emfollow.docs.ligo.org/userguide/
[2] https://gracedb.ligo.org/

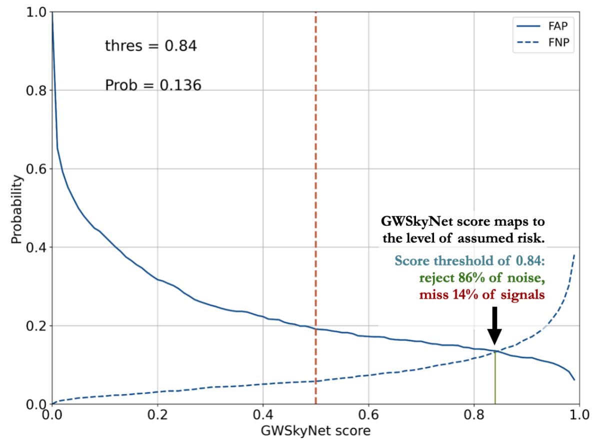

Figure 1: The mapping between GWSkyNet score, FAP (noise rejection) and FNP (signal rejection) for the current GWSkyNet model, tested on O3 data with simulated GW signals.

Interpreting GWSkyNet output - an example:

GWSkyNet score = 0.5, FAP = 20%, FNP = 6%.

This means 20% of noise events have an equal or greater score, and 6% of real events have an equal or lower score.

Examples of how to use GWSkyNet output

The following examples provide guidance for the use of GWSkyNet outputs to screen candidates for events caused or influenced by detector glitches. We assume that astronomers in these examples may be employing additional thresholds on other metrics, for example $p_{\rm astro}$ or false alarm rate from search pipelines.

- **Example 1**: Vera Rubin wants to compare Swift triggers to GW candidate alerts with false alarm rate lower than 1 per month to investigate for further follow up. She has ample telescope time available and wants to compile as large of a list of promising candidates as she can. She uses a GWSkyNet score threshold of 0.3 to screen 75% noise and catches 96% of astrophysical signals.

- **Example 2**: Nancy Grace Roman has very limited telescope time and is keen to follow up only the GW candidate alerts most likely to be astrophysical in origin. She uses a GWSkyNet score threshold of 0.75 to reject 85% noise and catches 90% astrophysical signals.

Limitations to GWSkyNet's performance

- GWSkyNet is (currently) only trained on multi-detector events, defined as event candidates identified with a network signal-to-noise ratio of ≥ 7 and a signal-to-noise ratio of ≥ 4.5 in at least two detectors. Therefore in O4, GWSkyNet will only annotate superevents that meet the same criteria.

- Cases where a glitch and signal overlap need to be interpreted carefully; GWSkyNet is trained only on simulated signals and cases where there is a real glitch in one or more detectors.

Acknowledgements

We gratefully acknowledge the valuable contributions of the following individuals to the development and success of GWSkyNet. We are grateful to Niko Lecoeuche for his insightful contributions to our understanding of GWSkyNet's performance. We also wish to thank Mario Drout, Aaron Tohuvavohu, Derek Davis, Brandon Piotrzkowski, Deep Chatterjee, Leo Singer, Beverly Berger, and Alan Knee for their valuable feedback and suggestions, which helped us refine GWSkyNet as a pipeline. We are humbled and grateful for the support and guidance provided by these individuals, without which GWSkyNet would not have been possible.

References

[1] Miriam Cabero, Ashish Mahabal, and Jess McIver. GWSkyNet: A Real-time Classifier for Public Gravitational-wave Candidates. The Astrophysical Journal Letters, 904(1):L9, November 2020. `https://iopscience.iop.org/article/10.3847/2041-8213/abc5b5`.