Gravitational wave signal recognition of O1 data by deep learning

He Wang,[1] Zhoujian Cao*,[2, †] Xiaolin Liu,[2] Shichao Wu,[2] and Jian-Yang Zhu[1]

[1]*Department of Physics, Beijing Normal University, Beijing 100875, China*
[2]*Department of Astronomy, Beijing Normal University, Beijing 100875, China*

Deep learning method develops very fast as a tool for data analysis these years. Such a technique is quite promising to treat gravitational wave detection data. There are many works already in the literature which used deep learning technique to process simulated gravitational wave data. In this paper we apply deep learning to LIGO O1 data. In order to improve the weak signal recognition we adjust the convolutional neural network (CNN) a little bit. Our adjusted convolutional neural network admits comparable accuracy and efficiency of signal recognition as other deep learning works published in the literature. Based on our adjusted CNN, we can clearly recognize the eleven confirmed gravitational wave events included in O1 and O2. And more we find about 2000 gravitational wave triggers in O1 data.

I. INTRODUCTION

Gravitational waves (GW) are an important prediction of Einstein's general theory of relativity, published a century ago. Gravitational wave observations of coalescing compact binaries are unique and unprecedented probes of strong-field and dynamical aspects of general relativity. More importantly, it ushers the brand new cosmic window to our universe—gravitational wave astronomy [1–8].

The capability of searching for GW signals relies on both the sensitivity of GW detectors and the theoretical waveform templates modeled for gravitational wave sources which will be used by matched-filtering data analysis technique. Currently the matched-filtering method is the standard and optimal signal processing techniques used by the gravitational wave community to find GW signals from compact BBH mergers in noisy detector data. Although the weak signal extraction and source information inversion of GW based on matched-filtering techniques are very successful, it has a great weakness as well as a potential hazard: data analysis through matched-filtering is a process of both a huge computation cost and slow computational speed. This is the major motivation for many authors to propose deep learning technique for GW data analysis [9–12]. Another possible problem with standard matched-filtering techniques is that the completeness and accuracy of GW waveform template is the prerequisite to guarantee its work. This implies a risk that we may lose the GW signals beyond the theoretical expectation. But the GW signal beyond the expectation of the theory will greatly facilitate the development of astronomy, and also provide important insight to the problems of fundamental physics such as quantum gravity and physics under extreme conditions.

The matched-filtering provides an optimal solution for identification of gravitational waves under Gaussian noise, but in practice, the data from GW detectors contains many non-Gaussian noise transients, also known as 'glitches'. A comprehensive classification and characterization of these noise features may provide valuable clues for identifying the source of noise transients, and possibly lead to their elimination. The machine learning method is becoming more and more important in various disciplines such as particle experiment [13, 14], gamma ray detection [15, 16], supernovae classification [17, 18], weak lensing data analysis [19–22], source modeling [23–26] and others. There have also been many attempts to use machine learning algorithms in gravitational wave data analysis to show promise for noise classification, categorization [27–35], and cancellation [36, 37]. And recently, an innovative project called "Gravity Spy"[1] [38, 39] combines the power of machine learning with the help of volunteers to label datasets of glitches and create a superior classifier of glitches in LIGO data. Machine learning techniques have been widely used in GW data processing, especially in the identification of signals and the classification of noises.

In recent years, deep learning, a new area of machine learning research, has been in the spotlight [40]. In the past few years, some active researchers have demonstrated empirical successes of deep neural networks in the applications to data analysis of gravitational waves [41–45]. These active researchers include George and his coworkers [9, 10], Gabbard and his coworkers [46], and others [11, 12, 47]. These published works used the convolutional neural network (CNN) from different perspectives to identify GW signals with low signal-to-noise ratio (SNR). Their works tell us that CNN architecture plays an important role for CBC GWs (also continuous GWs [48, 49] and CCSNe GWs [50]) recognition in the simulated or real noisy background from LIGO. Whereas, Gebhard et al. [51] have given an enlightening discussion on the general limitations of CNNs and proposed an alternative CNN-based architecture with proper per-

---

[1] www.gravityspy.org

formance metrics. They also claim that their trained network can cover all the GW events in both the first and second observation run (O1/O2) of LIGO, only except for GW170817, the first observation of GW from a binary neutron star inspiral.

In this paper we aim to use deep learning technique to find out all of the known GW events in the O1 and O2 data. In addition we would like to mark out other possible GW events candidates. We have tried CNN like the ones used in [9–12]. Although these neural networks can find GW150914, none of them can find out other events in O1. So we adjust the usual CNN a little bit. Based on our adjusted CNN networks, we can find out all 11 GW events reported by LSC [8]. Besides these confirmed events, we have also found out 2242 triggers in O1.

The plan of this paper is as follows: In Sec. II we describe the adjustment of the CNN neural network. Based on our adjusted CNN architecture, training data samples, test data samples, training strategy and search methodology on the real LIGO recording are described in Sec. III. After that we apply our trained network to the O1 data in Sec. IV. At last the Sec. V is devoted to a summary.

II. ADJUSTED CONVOLUTIONAL NEURAL NETWORKS

The inner product operation in frequency domain

$$\langle d|h \rangle = 4 \int_0^\infty \frac{\tilde{d}(f)\tilde{h}^*(f)}{S_n(f)} e^{2\pi i f t_c + \phi_c} df \qquad (1)$$

can also be expressed as a convolution in time domain. We divide the first layer of our adjusted CNN network into N_t groups. The coefficients $h_{ti}, i = 1, ..., N_t$ of each group are fixed and correspond to a whitened theoretical waveform. After the convolution operation between the input data and each group neural (we leave the detail operation to the App. A), we output the maximal value of each convolution. Then we collect these maximal values as the output of the first layer. The rest layers of our adjusted network is the same as usual CNN network [9, 10].

The coefficients h_{ti} are analogous to the template waveforms in matched filtering data analysis (Here we use index t to remind template). Intuitively yes, we even chose some template waveforms used by LSC to set these h_{ti}. Our basic idea is using these specific template waveforms to sense the GW signal deeply buried in the noise. But different to usual matched filtering technique, we need only tens of templates here instead of millions of templates in usual matched filtering technique. In the current work, we use $N_t = 35$ which works quite good to find signal in the O1 data.

Logically our adjusted CNN works in the following way. The coefficients h_{ti} we chosen span a subspace of the function space of the GW signal. After the first layer,
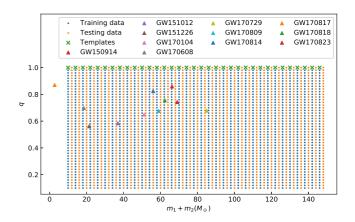

FIG. 1: The total mass and mass ratio of training/test data and templates. The eleven GW events for both O1 and O2 are also shown.

the essential matched filtering operation dig out the signal buried in the noise and project it into the subspace. Definitely such projection will admit some feature structure which may not be recognized by human but can be recognized by the following CNN layers. Such feature can be used to distinguish GW signal from pure noise.

Regarding to the 35 templates used in the first layer of our adjusted CNN network, we chose spinless identical binary black holes with total masses $M = 5 + 2i\mathrm{M}_\odot, i = 0, ..., 34$.

Before we send the data into our CNN, a tukey window[2] with $\alpha = 1/8$ is applied to remove edge effects at the beginning and end of the data stretch.

III. TRAINING AND TEST OF THE NEURAL NETWORK

A. Training data set and test data sets

Each data sample include noise part and possibly signal part. We obtain the O1 data from the Gravitational Wave Open Science Center (GWOSC) [52]. The background noises for training/testing are sampled from O1 data without GW150914, GW151012 and GW151226 events. Together with simulated GW signal we construct 3220 samples for training data and test data respectively. Among these 3220 samples, half of them are pure noise and the rest half include signals.

We use our SEOBNRE model [53] to generate simulated gravitational waveforms. In the current work we only consider circular, spinless binary black holes. Corresponding to LIGO detectors we adopt the configuration from LALSimulation[3] that set all the binary sources at

---

[2] en.wikipedia.org/wiki/Window_function#Tukey_window
[3] lscsoft.docs.ligo.org/lalsuite/lalsimulation/group_

right-ascension 1h 23m 45s, declination 45 degrees, and polarization angle 30 degrees and consider the total mass of the two black holes fall in the range $10 \sim 150 M_{\odot}$ in steps of 2 and the mass ratio $q = m_1/m_2$ from 1 to 10 in steps of 0.1. Regarding the orbital plane direction we set $\iota = 0$. For the source luminosity distance D, we determine it through signal-to-noise ratio assumption. Thus, we create a training and testing data set both with 1610 waveforms. Each data sample consists two time-series of 5 seconds long through a sampling rate of 4096 Hz. The two time-series corresponds to detector H1 and L1 respectively. For the samples with GW signal we set the peak location of the signal at the center of the time-series. The mass distribution of templates, training/test waveforms and also the eleven GW events in O1/O2 are plotted in Fig. 1.

B. Training strategy

The coefficients of our adjusted CNN network except the first layer will be determined through training process. Firstly, we use the "Xavier" initialization [54] to assign initial random values to these to be determined CNN parameters. This initializer is designed to keep the scale of gradients roughly the same in all layers. Then we use the binary output scores s from our network to calculate the confidence for a GW signal by sigmoid function:

$$p = \frac{1}{1 + e^{-s}}. \tag{2}$$

Then we use binary cross-entropy loss function to evaluate deviation between the predicted values and the actual values in the training data. Based on this estimation mini-batch Adam algorithms [55] is applied to optimize the kernel entries in CNN. Here we caution that this confidence value can not be interpreted as the statistical significance of a detection [51].

Within every training epoch (i.e., a full pass over all training data) not only the entire training/test data set is randomly shuffled, but also the background noises are newly sampled in random manner from O1 data except the three GW events. At the end of every epoch, the performance of the network during the training is evaluated by average accuracy for the networks on each mini-batch. We set the learning rate 0.003 and batch size 16. During the curriculum learning, we gradually decrease the signal SNR of both training data and test data from 1 to 0.02. The training process is done within 6 hours on 4 NVIDIA GeForce GTX 1080Ti GPUs, each with 11GB of memory. All the implementations of current work were coded with Python based on the MXNet framework [56].
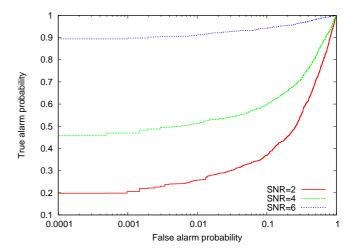

_lalsim__detector__strain.html

FIG. 2: The ROC curves for test data sets containing signals with matched filtering SNR 2, 4 and 6 respectively. We plot the true alarm probability versus the false alarm probability estimated by our adjusted CNN network.
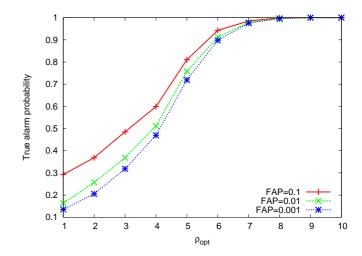

FIG. 3: Efficiency curves for false alarm probabilities 0.1, 0.01 and 0.001 respectively. The true alarm probability is plotted as a function of the optimal SNR.

C. Accuracy and efficiency test of the neural network

The authors in [46] have compared with the GW signal recognition accuracy and efficiency between CNN network and usual matched filtering technique. Based on the test data set described in the above subsection, we can calculate corresponding true alarm probability and false alarm probability. Then we can construct the receiver operator characteristic (ROC) curves in the Fig. 2.

Stronger signal is easier to be recognized. For a given SNR, the true alarm probability can be used to describe the recognition efficiency. Respect to representative false alarm probability 0.1, 0.01 and 0.001, we plot out the efficiency curves in the Fig. 3. Both the ROC curves and

the efficiency curves are comparable to the results got in [46] which means our adjusted neural network is powerful to recognize the GW signal buried in the noise.

IV. SEARCH RESULTS OF THE REAL LIGO DATA

After being trained, our adjusted CNN is ready to trigger GW signal for O1 data. Each time we take a 5-seconds long data segment from O1 data. After one segment, we move forward one second to get another 5-seconds long data segment. So each one second long data will appear in 5 segments and be processed 5 times by our network. If we assume our network can recognize GW signal within one second time duration, our network will alert continuously 5 times for a true signal.

If the output confidence value is bigger than a given threshold value $p > p_c$, our network gives an alert. If more than 5 continuous alerts happen, a trigger for GW signal will be given.

In the Fig. 4 we plot the confidence values outputted by our adjusted CNN near the GW150914, GW151012 and GW151226 events for O1. We can see our network can mark out all of these three events clearly. For GW150914 and GW151226 there are 5 continuous confidence values approaching 1 while other confidence values are much smaller. For GW151012 the confidence values marking out the signal are not as big as GW150914 and GW151226, but still bigger than 0.5 which is also clearly bigger than the nearby values.

It is interesting to check the effect of the noise used in the training data set and the real data. The authors in [51] used the network trained with O1 noise to treat the O2 data. We do the same test here. We apply the trained network described in the above section directly to O2 data. We show the confidence values for the eight GW events for O2 in the Fig. 5. All of the eight events have been clearly marked out. The authors in [51] shows that the network trained with binary black hole coalescence signal may fail to find out the binary neutron star inspiral signal. This is not true to our adjusted network. We can mark out the GW170817 event very clearly.

Besides the three confirmed GW events, we have also applied our network to all of the O1 data. In the current work we set $p_c = 0.5$. If there are more than 5 continuous $p > p_c$, we output a signal alarm at the time of these first $p > p_c$.

If leaving the data quality alone, we find 3955 triggers in O1 which include GW150914, GW151012 and GW151226. If we consider only the data which pass the CBC-CAT3 test there are 2242 triggers in O1. Averagely there are one trigger about every one and half hours. As noted by [51], deep learning can not assign a significance to each trigger. So we can not tell which triggers are more believable than others. We can only suggest these triggers are deserved to be checked in more detail with other analysis methods. We have listed these triggered data with center time and time duration on Github [4].

The authors of GWTC-1 reported 3 GW events candidates besides the three known GW events in O1 in the table II of [8]. All these 3 candidates are different to our triggers.

The authors of [57] reported the first open gravitational wave catalog (1-OGC), where 20 GW event candidates are reported including the three known GW events. All these 17 subthreshold candidates are also different to our triggers.

We have also checked the consistency between our triggers and the gamma ray burst (GRB) events listed in [58]. There are 1209 GRB events recorded during the O1 run of LIGO. But there is no consistent events found between our triggers and these GRB records.

V. SUMMARY

There are many works published about applying deep learning technique to gravitational wave data analysis in the past few years. Most of these works used simulated data. In the current paper we designed an adjusted CNN and applied it to the whole O1 data of LIGO.

Being trained with noise taken from O1 data and simulated binary black hole coalescence waveform our network can find out the three confirmed GW events clearly. In addition, we used this trained network directly to the 8 GW events found in O2 and we can also mark out all these 8 events clearly. Although this test is inconsistent due to different noise behavior between O1 and O2, our test results indicate that our network and method are robust respect to the training data set. Similar investigation has been done in [51].

Besides the three GW events included in O1 we have also found more than 2000 candidates. As noted by the authors of [51], we can not assign a significance to each candidate. So we simply call these candidates triggers. We mean these trigged data segments are deserved to be investigated in detail through other means.

We have compare our more than 2000 triggers with the subthreshold events reported in GWTC-1 [8] and in 1-OGC [57]. No consistent events are found between these subthreshold events and our triggers. And more we have also checked the reported GRB events listed in [58] during the O1 observation time of LIGO. We have not found consistent events between these GRB events and our triggers either.
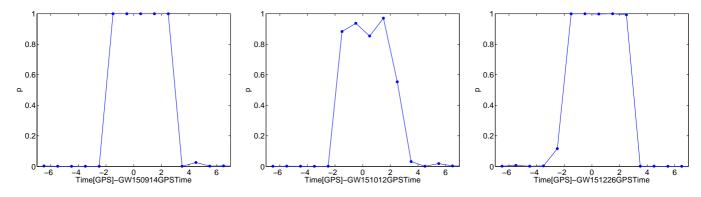
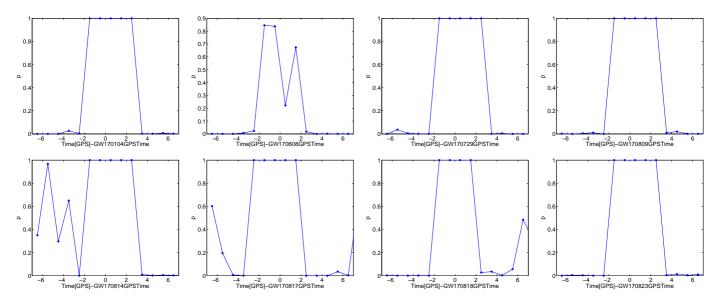FIG. 4: The output confidence values of our adjusted CNN near the three GW events of O1.

FIG. 5: The output confidence values of our adjusted CNN trained with O1 noise near the eight GW events of O2.

Appendix A: Convolution operation in the first layer of the adjusted CNN network

For a template $h(t)$ and a detected strain data $d(t)$, the matched-filter SNR is defined as

$$\rho^2(t_c) \equiv \frac{|\langle d|h\rangle(t_c)|^2}{\langle h|h\rangle}, \tag{A1}$$

$$\langle d|h\rangle(t_c) = 4\int_0^\infty \frac{\tilde{d}(f)\tilde{h}^*(f)}{S_n(f)} e^{2\pi i f t_c} df, \tag{A2}$$

$$\langle h|h\rangle = 4\int_0^\infty \frac{\tilde{h}(f)\tilde{h}^*(f)}{S_n(f)} df, \tag{A3}$$

where $|\cdot|$ means taking the absolute value, $*$ means complex conjugate and $\tilde{\ }$ is the Fourier transformation like

$$\tilde{d}(f) = \int_{-\infty}^\infty d(t) e^{-2\pi i f t} dt. \tag{A4}$$

The quantity $S_n(f)$ is the one-sided average power spectral density of the detector noise. Due to the convolution theorem and the relationship between convolution and correlation, the correlation in Eq. (A2) can be rewritten as

$$\begin{aligned}
\langle d|h\rangle(t) &= 4\int_0^\infty \frac{\tilde{d}(f)\cdot\tilde{h}^*(f)}{S_n(f)} e^{2\pi i f t} df \\
&= 4\int_0^\infty [\tilde{d}(f)\bar{S}_n(f)]\cdot[\tilde{h}(f)\bar{S}_n(f)]^* e^{2\pi i f t} df \\
&= 4\int_0^\infty \tilde{\bar{d}}(f)\cdot\tilde{\bar{h}}^*(f) e^{2\pi i f t} df
\end{aligned}$$

$$= 2\,\bar{d}(t) * \bar{h}(-t)\,, \tag{A5}$$

$$\bar{d}(t) = d(t) * \bar{S}_n(t)\,, \tag{A6}$$

$$\bar{h}(t) = h(t) * \bar{S}_n(t)\,, \tag{A7}$$

$$\bar{S}_n(t) = \int_{-\infty}^{+\infty} S_n^{-1/2}(f) e^{2\pi i f t} df\,. \tag{A8}$$

where $d(t) * h(t)$ means the convolution of functions $d(t)$ and $h(t)$. Similarly, the $\langle h|h \rangle$ can also be calculated in this way as $\langle h|h \rangle = 2\,[\bar{h}(t) * \bar{h}(-t)]\big|_{t=0}$.

[1] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. **116**, 061102 (2016), URL https://link.aps.org/doi/10.1103/PhysRevLett.116.061102.

[2] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. **116**, 241103 (2016), URL https://link.aps.org/doi/10.1103/PhysRevLett.116.241103.

[3] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. X **6**, 041015 (2016), URL https://link.aps.org/doi/10.1103/PhysRevX.6.041015.

[4] B. P. Abbott et al. (LIGO Scientific and Virgo Collaboration), Phys. Rev. Lett. **118**, 221101 (2017), URL https://link.aps.org/doi/10.1103/PhysRevLett.118.221101.

[5] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. **119**, 141101 (2017), URL https://link.aps.org/doi/10.1103/PhysRevLett.119.141101.

[6] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. **119**, 161101 (2017), URL https://link.aps.org/doi/10.1103/PhysRevLett.119.161101.

[7] B. P. Abbott et al., The Astrophysical Journal Letters **851**, L35 (2017), URL http://stacks.iop.org/2041-8205/851/i=2/a=L35.

[8] B. P. Abbott et al. (LIGO Scientific, Virgo) (2018), 1811.12907.

[9] D. George and E. A. Huerta, Phys. Rev. D **97**, 044039 (2018), URL https://link.aps.org/doi/10.1103/PhysRevD.97.044039.

[10] D. George and E. Huerta, Physics Letters B **778**, 64 (2018), ISSN 0370-2693, URL http://www.sciencedirect.com/science/article/pii/S0370269317310390.

[11] X. Li, W. Yu, and X. Fan, ArXiv e-prints (2017), 1712.00356.

[12] X.-L. Fan, J. Li, X. Li, Y. Zhong, and J. Cao, SCIENCE CHINA Physics, Mechanics & Astronomy p. in press (2018).

[13] T. Cohen, M. Freytsis, and B. Ostdiek, Journal of High Energy Physics **2018**, 34 (2018), ISSN 1029-8479, URL https://doi.org/10.1007/JHEP02(2018)034.

[14] S. Chang, T. Cohen, and B. Ostdiek, Phys. Rev. D **97**, 056009 (2018), URL https://link.aps.org/doi/10.1103/PhysRevD.97.056009.

[15] S. Caron, G. A. Gomez-Vargas, L. Hendriks, and R. R. de Austri, Journal of Cosmology and Astroparticle Physics **2018**, 058 (2018), URL http://stacks.iop.org/1475-7516/2018/i=05/a=058.

[16] S. Abraham, N. Mukund, A. Vibhute, V. Sharma, S. Iyyani, D. Bhattacharya, A. R. Rao, S. Vadawale, and V. Bhalerao, *A machine learning approach for grb detection in astrosat czti data* (2019), 1906.09670.

[17] T. Charnock and A. Moss, The Astrophysical Journal Letters **837**, L28 (2017).

[18] A. Moss (2018), 1810.06441.

[19] A. Gupta, J. M. Z. Matilla, D. Hsu, and Z. Haiman, Physical Review D **97**, 103515 (2018).

[20] M. Shirasaki, N. Yoshida, and S. Ikeda, Phys. Rev. **D100**, 043527 (2019), 1812.05781.

[21] D. Ribli, B. r. Pataki, J. M. Z. Matilla, D. Hsu, Z. Haiman, and I. Csabai (2019), 1902.03663.

[22] J. Fluri, T. Kacprzak, A. Lucchi, A. Refregier, A. Amara, T. Hofmann, and A. Schneider, Phys. Rev. **D100**, 063514 (2019), 1906.03156.

[23] A. J. K. Chua, C. R. Galley, and M. Vallisneri, Phys. Rev. Lett. **122**, 211101 (2019), 1811.05491.

[24] A. Rebei, E. A. Huerta, S. Wang, S. Habib, R. Haas, D. Johnson, and D. George, Phys. Rev. **D100**, 044025 (2019), 1807.09787.

[25] A. J. K. Chua and M. Vallisneri (2019), 1909.05966.

[26] Y. Setyawati, M. Prrer, and F. Ohme (2019), 1909.10986.

[27] S. Rampone, V. Pierro, L. Troiano, and I. M. Pinto, Int. J. Mod. Phys. **C24**, 1350084 (2013), 1401.5941.

[28] S. Mukherjee, R. Obaid, and B. Matkarimov, J. Phys. Conf. Ser. **243**, 012006 (2010).

[29] J. Powell, D. Trifir, E. Cuoco, I. S. Heng, and M. Cavagli, Class. Quant. Grav. **32**, 215012 (2015), 1505.01299.

[30] J. Powell, A. Torres-Forn, R. Lynch, D. Trifir, E. Cuoco, M. Cavagli, I. S. Heng, and J. A. Font, Class. Quant. Grav. **34**, 034002 (2017), 1609.06262.

[31] D. George, H. Shen, and E. A. Huerta, Phys. Rev. D **97**, 101501 (2018), URL https://link.aps.org/doi/10.1103/PhysRevD.97.101501.

[32] N. Mukund, S. Abraham, S. Kandhasamy, S. Mitra, and N. S. Philip, Phys. Rev. D **95**, 104059 (2017), URL https://link.aps.org/doi/10.1103/PhysRevD.95.104059.

[33] M. Razzano and E. Cuoco, Classical and Quantum Gravity **35**, 095016 (2018), URL http://stacks.iop.org/0264-9381/35/i=9/a=095016.

[34] M. Zevin, S. Coughlin, S. Bahaadini, E. Besler, N. Rohani, S. Allen, M. Cabero, K. Crowston, A. K. Katsaggelos, S. L. Larson, et al., Classical and Quantum Gravity **34**, 064003 (2017), 1611.04596.

[35] D. George, H. Shen, and E. A. Huerta (2017), 1706.07446.

[36] W. Wei and E. A. Huerta (2019), 1901.00869.

[37] H. Shen, D. George, E. A. Huerta, and Z. Zhao (2019), 1903.03105.

[38] M. Zevin, S. Coughlin, S. Bahaadini, E. Besler, N. Rohani, S. Allen, M. Cabero, K. Crowston, A. K. Katsaggelos, S. L. Larson, et al., Classical and Quantum Gravity **34**, 064003 (2017), URL http://stacks.iop.org/0264-9381/34/i=6/a=064003.

[39] S. Bahaadini, V. Noroozi, N. Rohani, S. Coughlin, M. Zevin, J. Smith, V. Kalogera, and A. Katsaggelos, Information Sciences **444**, 172 (2018), ISSN 0020-0255, URL http://www.sciencedirect.com/science/article/pii/S0020025518301634.

[40] G. Allen et al. (2019), 1902.00522.

[41] H. Gabbard, C. Messenger, I. S. Heng, F. Tonolini, and R. Murray-Smith (2019), 1909.06296.

[42] C. Chatterjee, L. Wen, K. Vinsen, M. Kovalam, and A. Datta (2019), 1909.06367.

[43] A. L. Miller et al. (2019), 1909.02262.

[44] P. G. Krastev (2019), 1908.03151.

[45] H. Shen, E. A. Huerta, Z. Zhao, E. Jennings, and H. Sharma (2019), 1903.01998.

[46] H. Gabbard, M. Williams, F. Hayes, and C. Messenger, Phys. Rev. Lett. **120**, 141103 (2018), URL https://link.aps.org/doi/10.1103/PhysRevLett.120.141103.

[47] L.-L. Wang, J. Li, N. Yang, and X. Li, New J. Phys. **21**, 043005 (2019).

[48] C. Dreissigacker, R. Sharma, C. Messenger, R. Zhao, and R. Prix, Phys. Rev. D **100**, 044009 (2019), URL https://link.aps.org/doi/10.1103/PhysRevD.100.044009.

[49] F. Morawski, M. Bejger, and P. Ciecielag, arXiv e-prints arXiv:1907.06917 (2019), 1907.06917.

[50] P. Astone, P. Cerdá-Durán, I. Di Palma, M. Drago, F. Muciaccia, C. Palomba, and F. Ricci, Phys. Rev. D **98**, 122002 (2018), URL https://link.aps.org/doi/10.1103/PhysRevD.98.122002.

[51] T. D. Gebhard, N. Kilbertus, I. Harry, and B. Scholkopf (2019), 1904.08693.

[52] M. Vallisneri, J. Kanner, R. Williams, A. Weinstein, and B. Stephens, Journal of Physics: Conference Series **610**, 012021 (2015), URL https://doi.org/10.1088%2F1742-6596%2F610%2F1%2F012021.

[53] Z. Cao and W.-B. Han, Phys. Rev. D **96**, 044028 (2017), URL https://link.aps.org/doi/10.1103/PhysRevD.96.044028.

[54] X. Glorot and Y. Bengio, in *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (2010), pp. 249–256.

[55] D. P. Kingma and J. Ba (2014), 1412.6980.

[56] T. Chen, M. Li, Y. Li, M. Lin, N. Wang, M. Wang, T. Xiao, B. Xu, C. Zhang, and Z. Zhang, ArXiv e-prints (2015), 1512.01274.

[57] A. H. Nitz, C. Capano, A. B. Nielsen, S. Reyes, R. White, D. A. Brown, and B. Krishnan, The Astrophysical Journal **872**, 195 (2019), URL https://doi.org/10.3847%2F1538-4357%2Fab0108.

[58] The Gamma-ray Coordinates Network (GCN), TAN: Transient Astronomy Network. https://gcn.gsfc.nasa.gov.