


Detecting and denoising gravitational wave signals from binary black holes using deep learning

Chinthak Murali¹ and David Lumley^{1,2}

¹*Department of Physics, The University of Texas at Dallas, Richardson, Texas 75080, USA*

²*Department of Geosciences, The University of Texas at Dallas, Richardson, Texas 75080, USA*

 (Received 5 October 2022; revised 9 June 2023; accepted 7 August 2023; published 23 August 2023)

We present a convolutional neural network, designed in the autoencoder configuration that can detect and denoise astrophysical gravitational waves from merging black hole binaries, orders of magnitude faster than the conventional matched-filtering based detection that is currently employed at advanced LIGO (aLIGO). The neural-net architecture is such that it learns from the sparse representation of data in the time-frequency domain and constructs a nonlinear mapping function that maps this representation into two separate masks for signal and noise, facilitating the separation of the two, from raw data. This approach is the first of its kind to apply machine learning based detection/denoising of gravitational waves from binary black holes in the 2D representation of gravitational wave data. We applied our formalism to the first gravitational wave event detected, GW150914, successfully recovering the signal at all three phases of coalescence at both detectors. This method is further tested on the gravitational wave data from the second observing run (*O2*) of aLIGO, reproducing all binary black hole mergers detected in *O2* at both the aLIGO detectors. This method can also interpolate and extrapolate between modeled templates and explore gravitational waves that are not present in the template bank of signals used in the matched-filtering detection pipelines. Faster and efficient detection schemes, such as this method, will be instrumental as ground based detectors reach their design sensitivity, likely to result in several hundreds of potential detections in a few months of observing runs.

DOI: [10.1103/PhysRevD.108.043024](https://doi.org/10.1103/PhysRevD.108.043024)

I. INTRODUCTION

The first direct detection of a gravitational wave (GW), GW150914 in 2015 has opened a new window into observing the universe [1,2]. In the first three observing runs (*O1*, *O2* and *O3*) of aLIGO [3] and the European advanced VIRGO (aVIRGO) detector [4] over the span of 27 months, a total of 90 CBC (Compact Binary Coalescence) events with an astrophysical origin probability $p_{\text{astro}} > 0.5$ have been detected, as reported in the gravitational wave transient catalogs, GWTC-1 [5], GWTC-2 [6] and GWTC-3 [7]. This includes 85 binary black hole (BBH) mergers, four neutron star black hole (NSBH) mergers and one binary neutron star (BNS) merger [8]. Hundreds of more detections are expected in the scheduled *O4* run of aLIGO with an enhanced sensitivity [9] and being joined by the Japanese underground detector KAGRA [10,11]. Although extremely reliable and robust, the demand to go beyond the current detection scheme and adopt faster and efficient data processing techniques are only getting higher with a number of ground based GW detectors under construction [12,13].

aLIGO uses matched-filtering based search to identify the presence of a GW in the data stream [14,15]. In this process, an enormous template bank of signals (depending on how

broad the search is) are cross-correlated with several months of detector data output, creating an alert when the matched-filter signal to noise ratio (SNR) exceeds the detection threshold of the detection pipeline. Once the presence of a GW is identified in the time segment, traditional signal processing techniques are applied to enhance the signal and suppress the noise, to retrieve the signal from the data stream. The template with the highest SNR is then used to characterize the merger event. A merging BBH system is characterized by eight intrinsic parameters, the two masses and the two spins (magnitudes and directions) and the extrinsic parameters: the luminosity distance, right ascension, declination, polarization, inclination, coalescence time, coalescence phase and two parameters of eccentricity. Sampling a fifteen dimensional parameter space to characterize each individual GW waveform and performing matched-filter analysis with each data segment is not a trivial computational task [16].

The recent advances in machine learning (ML), namely deep learning (DL), can help us navigate this intense computational problem. DL is an extremely powerful machine learning technique that can learn very complex features and functions through neural networks [17,18]. Convolutional neural network (CNN) is a special class of

neural network that performs convolution operations by means of kernel filters to the input data. During the training process, the CNN assigns weights to the filters that optimally extracts various features from the input data. As opposed to matched-filtering based search, in a DL based search, all the intensive computations are performed only during the training stage of the network, which is a one time process [19]. This procedure can also interpolate and in some cases extrapolate between waveform templates, making it more robust in detecting GW signals without necessarily training the network on a 15D parameter space. DL based search also opens up the possibility of detecting GW signals that are outside the realm of theoretical template banks currently modeled using numerical relativity (NR). Because of the ability of the neural networks to process data and detect the GWs orders of magnitude faster, it can be effectively combined with the matched-filtering based detections to enhance the confidence of a given detection and also to efficiently target data segments with positive detections for a subsequent matched-filter based analysis.

Several ML based analysis of GW data have appeared in the literature over the past few years, beginning with glitch classification and subtraction [20–25]. DL methods are first applied to direct detection of GWs by George and Huerta using simulated aLIGO noise [19] and further extended to real aLIGO noise, demonstrating the ability of DL in detecting GW events in real detector data [26]. This work has inspired several attempts to identify and locate GWs in real aLIGO data [27–33] and many other ML based studies focused on parameter estimation of real GWs [34–37] followed. Denoising GWs using DL was applied in [38], and the proposed denoising scheme was able to extract four GW events (three from Hanford, one from Livingston) with high signal overlaps. In [39,40] the authors proposed a denoising autoencoder, based on recurrent neural networks (RNN) to denoise GWs. Most recently, [41] used 1D CNN in an autoencoder configuration to denoise many of the GWs (three events from $O1$ and three events from $O2$), using whitened data from both the detectors.

Although George and Huerta [19] in the foundational article suggested that using 2D data in the context of GW detection is sub-optimal because of the extremely weak signal strength characteristic of GWs, we are using 2D representation of the data in order to separate signal from the noise. In our denoising scheme, we use 2D CNN architecture on raw detector data, with unwhitened signal templates and noise. This approach, to the best of our knowledge is the first attempt to detect and denoise GWs from binary black holes, in raw detector data in 2D representation.

In this paper we present a DL framework to detect and denoise GWs from raw strain data from the aLIGO detectors. The detector noise at both aLIGO detectors are highly nonstationary and non-Gaussian, with very high noise dominating both ends of the frequency spectrum [42]. While ambient seismic noise (from ocean, traffic,

earthquakes, etc.) dominate the low frequency end of the noise [43], photon shot noise dominate the high frequency regime. The real GW signals are extremely weak and are deeply buried inside the detector noises and occupy the same frequency band as the detector noise, making it impossible to separate from the noise using traditional filtering methods.

In this method, we implement a time-frequency denoising technique used in seismic denoising, which successfully separated earthquake signals from ambient noise in 2019 [44] using DL. The noisy strain data is first transformed into time-frequency domain using short time Fourier transform (STFT), rendering the one dimensional time series data into two dimensions, represented by Fourier coefficients. In the Fourier domain, the coefficients associated with noise are attenuated to enhance the signal coefficients. These modified Fourier coefficients are then transformed back into the time domain using inverse Fourier transform and thereby reconstructing the GW chirp signal in the time domain. The underlying idea is to promote a sparse representation of the signal in the time-frequency domain. Now, the signal can be represented by a sparse set of features which makes the separation of the signal from the noise easier in the Fourier domain. This method is especially useful when the GW signal and the detector noise occupy the same frequency band, making the filtering process using traditional signal processing techniques virtually impractical.

II. METHOD

The key here is to find a mapping function that can appropriately find a threshold to suppress the coefficients corresponding to the noise in the time-frequency domain, hence enhance signal separation. These functions are highly nonlinear and are hence difficult to construct mathematically for the GW problem. That is where DL techniques can be incredibly effective, which learns to build a high dimensional, nonlinear mapping function from the data alone during the training process. This neural-net learns the sparse representation of the data in the time-frequency domain, and builds a high-dimensional, nonlinear mapping function which maps these representations into two masks, one for the GW signal and another one for all the noises.

The raw GW data from the detector $d(t)$ undergoes STFT and is represented in the time-frequency domain as $D(t, f)$, which is a combination of the GW signal $S(t, f)$ and all the noises $N(t, f)$,

$$D(t, f) = S(t, f) + N(t, f). \quad (1)$$

The idea is to construct the mapping functions that can successfully map the detector data into a representation of the signal and a representation of the noise separately. This mapping can be accomplished through a soft thresholding in the sparse representation where the threshold is estimated by assuming a Gaussian distribution of noise [45].

From here we construct two individual masks $M_S(t, f)$ and $M_N(t, f)$ which act as the mapping functions for the signal and noise respectively, and are given by,

$$M_S(t, f) = \frac{1}{1 + \frac{|N(t, f)|}{|S(t, f)|}}, \quad (2)$$

$$M_N(t, f) = \frac{\frac{|N(t, f)|}{|S(t, f)|}}{1 + \frac{|N(t, f)|}{|S(t, f)|}}. \quad (3)$$

These masks are the targets for the supervised learning problem, and they are constructed during the training process from the training data. The neural-net is trained to construct these masks at the output layer, for every single data segment that is fed into the network as part of the training data. Once these masks are constructed, they can be multiplied with the 2D detector data $D(t, f)$ to produce the real and imaginary parts of the signal and noise separately, and combining them at the output to reconstruct the signal and noise as follows,

$$M_S(t, f) \times D(t, f) \sim S(t, f), \quad (4)$$

$$M_N(t, f) \times D(t, f) \sim N(t, f). \quad (5)$$

Since we apply the same masks to both the real and imaginary parts of the input noisy signal, the neural-net does not introduce a phase shift during denoising. The reconstructed signal and noise are further inverse transformed back into the time domain to reproduce the one dimensional time series of signal and noise,

$$STFT^{-1}(S(t, f)) = S(t), \quad (6)$$

$$STFT^{-1}(N(t, f)) = N(t). \quad (7)$$

The denoising process aims to find the true GW signal \hat{S} from the data by minimizing the mean squared error between the true signal and the estimated signal, calculated as,

$$\text{Error} = \|\hat{S} - S\|^2. \quad (8)$$

The masks $M_S(t, f)$ and $M_N(t, f)$ have the same dimension as the input data $D(t, f)$ and take values between 0 and 1 and are mutually exclusive ($M_N = 1 - M_S$).

A. Neural-net architecture

Autoencoders are a class of artificial neural networks that can learn to code patterns from unlabeled data (unsupervised learning), typically for dimensionality reduction [46]. The encoder part of the autoencoder maps the representation to a code and the decoder part reconstructs the coded representation, such that it provides outputs with the same shape as the inputs. Because of the ability of autoencoders

to learn a sparse representation of the data, our neural-net is designed as a series of 2 dimensional convolutional layers with descending and then ascending sizes, like an autoencoder [47,48], as shown in Fig. 1. This particular architecture is inspired from what is called a U-Net, where the pooling operations between convolutional layers are replaced by upsampling operators and thereby increasing the resolution of the output. The network uses the output of each convolution operation, without using any fully connected layers [49]. The presence of large number of feature channels in the upsampling part enables the network to propagate context information to high resolution layers. Skip connections are used to improve convergence during the training process [50], which are represented as overhead arrows connecting the encoder and decoder layers of same dimensions. In addition to improving the convergence, skip connections helped to minimize the signal leakage into the noise. Skip connection and Residual Neural Network (ResNet) models are detailed in the original work, [51].

The entire process of signal and noise separation is presented as a sequence in Fig. 2. The input to the neural-net is through two channels: one takes the real part of the Fourier coefficients and the other takes the imaginary part of the Fourier coefficients. This enables the neural-net to learn from both the amplitude and phase information of the data. These inputs go through a series of 2D convolutional layers of constantly decreasing dimensions, where each layer uses a Rectified Linear Unit (ReLU) activation function and are subsequently Batch Normalized. The dimensions of the convolutional layers are reduced using a stride of 2×2 and the kernel size of each layer remains at 3×3 , following [44]. Each of the convolutional layers extracts features from the data and learns to represent the data more and more sparsely as they go along the network layers, because of the autoencoder architecture. At the bottleneck layer, we have the smallest and sparsest representation of the data possible and then it goes through the deconvolutional layers which generate a high-dimensional, nonlinear mapping of this sparse representation into output masks. The output labels are the masks created for the signal (M_S) and for the noise (M_N). During the training process, the network learns to generate the masks that optimally separates signal from noise by minimizing a loss function, which is a binary cross-entropy loss function. A softmax normalized exponential function is used at the final layer to produce the output masks. Figure 1 shows the neural-net architecture which takes the inputs through two channels and generates two outputs, one of which is the familiar GW chirp signal in the time-frequency domain.

B. Data acquisition and preprocessing

The data used for training is the publicly available GW strain data from the Gravitational Wave Open Science Center (GWOSC) [52]. These are continuous recordings at both the aLIGO detectors over the course of the first three

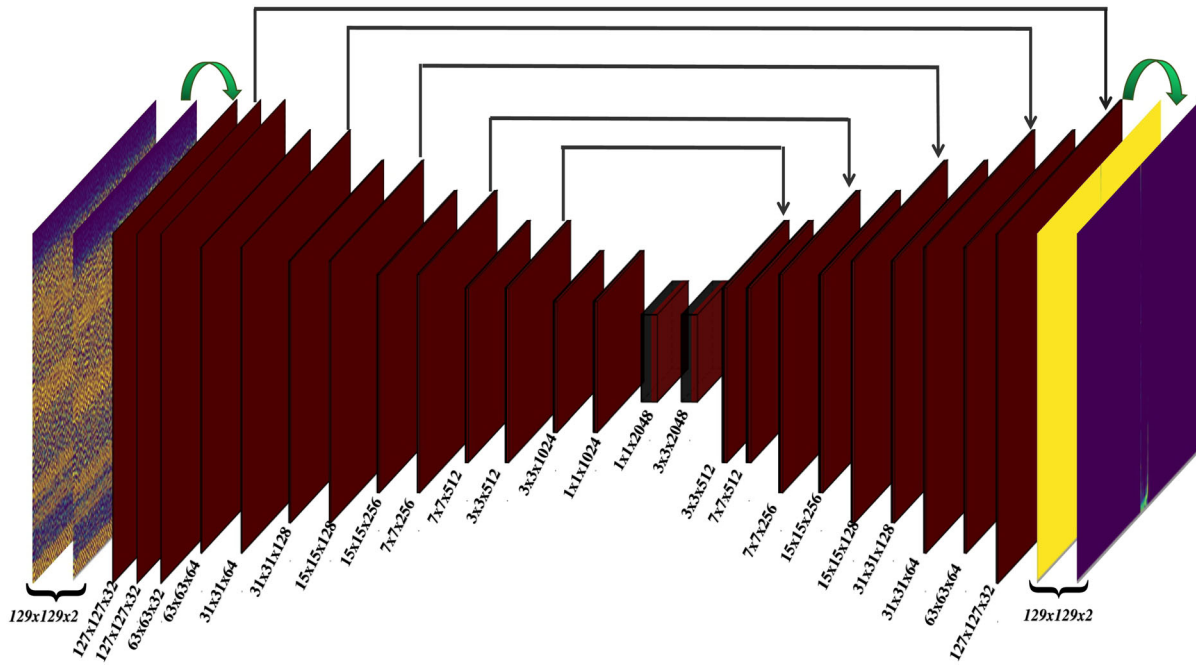


FIG. 1. Neural-net architecture. Inputs are the real and imaginary components of the Fourier coefficients, and outputs are the masks M_S and M_N . Each red rectangle is a 2D convolutional layer, with dimensions labeled at the bottom as “frequency bins \times time points \times channels”. Overhead arrows are skip connections between the encoder and the decoder parts of the network.

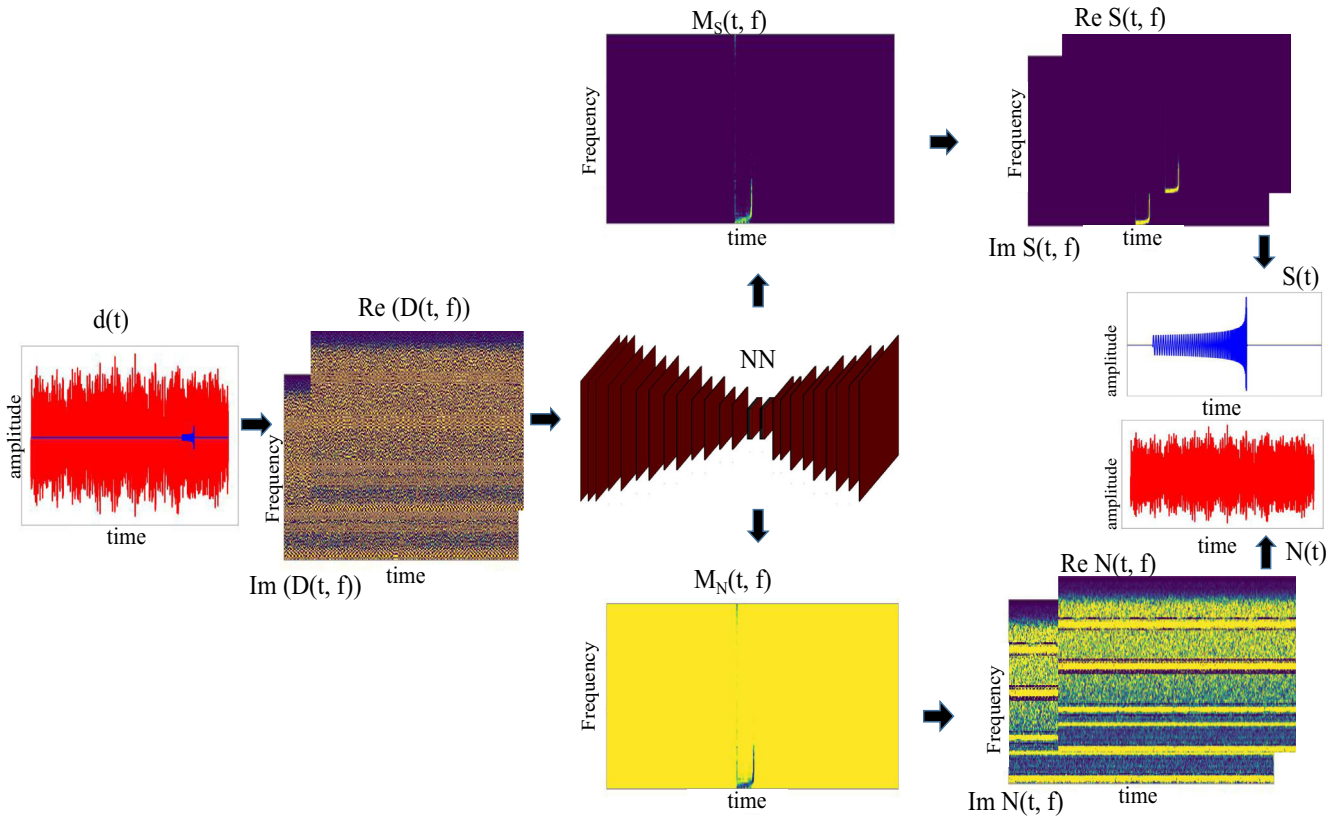


FIG. 2. Denoising flow diagram. Input is the data in the time domain, which undergoes STFT to produce the real and imaginary parts of the data, which are inputs to the neural-net. The neural-net generates masks for the signal and the noise, which are multiplied with the data itself to create the 2D representation of signal and noise separately. An inverse STFT reconstructs the GW signal and noise in the time domain. We use normalized, unwhitened data for this whole procedure.

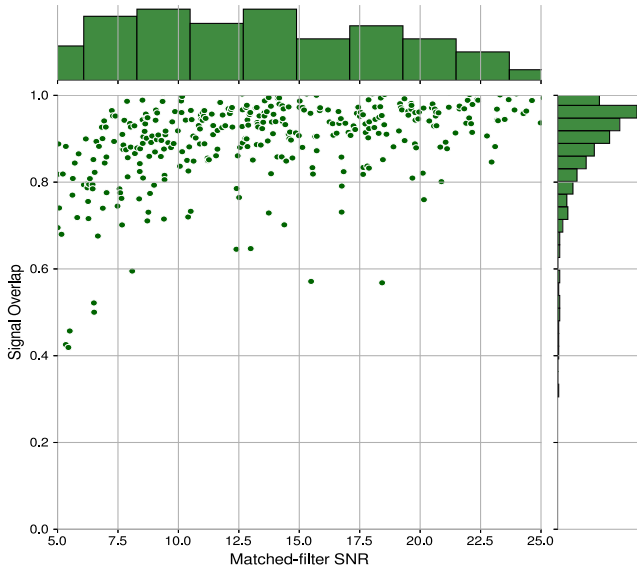


FIG. 3. Signal overlaps after denoising from the simulated noise alone, as function of the matched-filter SNR (before denoising). Top histogram shows the population sample of SNRs used for the analysis, side histogram shows the statistics of signal overlaps.

observing runs. We used the data from *O1* to train the network to detect the event GW150914 and data from *O2* to train the network for the rest of the events. From a few hours of continuous noise recordings, we generated different realizations of the noise to create a larger noise set, to make the neural-net more adaptive to variations in noise. Data from GWOSC is sampled at 4096 Hz and are 4096 s long data segments. We down-sampled the data into 2048 Hz and

divided them into eight second long data segments in the training phase. These eight second long data segments are pure noise where there is no known GW event reported so far, or any hardware injection, often used for calibration purposes. We applied a lower frequency cutoff of 30 Hz for this analysis as all the events detected so far are above this frequency. This noise is then combined with simulated GW signals, modeled using the optimized effective one body numerical relativity waveform SEOBNRv4_opt [53] (an optimized version of SEOBNRv4 [54]) sampled at 2048 Hz with masses ranging from $5M_{\odot}$ to $80M_{\odot}$ and with SNRs ranging from 5 to 25, distributed as shown in Figs. 3 and 4. For the analysis we assumed optimal orientation of both detectors with the event, also spins and eccentricities are assumed to be zero. This makes the analysis essentially two dimensional, where the signals are characterized by individual masses of the binary, which to the first order, captures the essence of the GW signal.

The final one second of each modeled waveform is used as the GW signal that the neural-net is trained to identify and extract from the detector noise. We inject this one second long signal in every eight seconds of the noise to realize the data used in the analysis. For the entire subset of the signals used for the training purpose, there is enough signal in the inspiral, merger and ringdown phases, for effective pattern recognition by the neural-net. The one second long signal is added to the eight second long noise by randomly time translating the signal within the noise to make the neural-net more robust to look for signals at different parts of the data, rather than being constrained at a single position where it is most likely to spot the signal. Another equivalent set is created with only noise and no

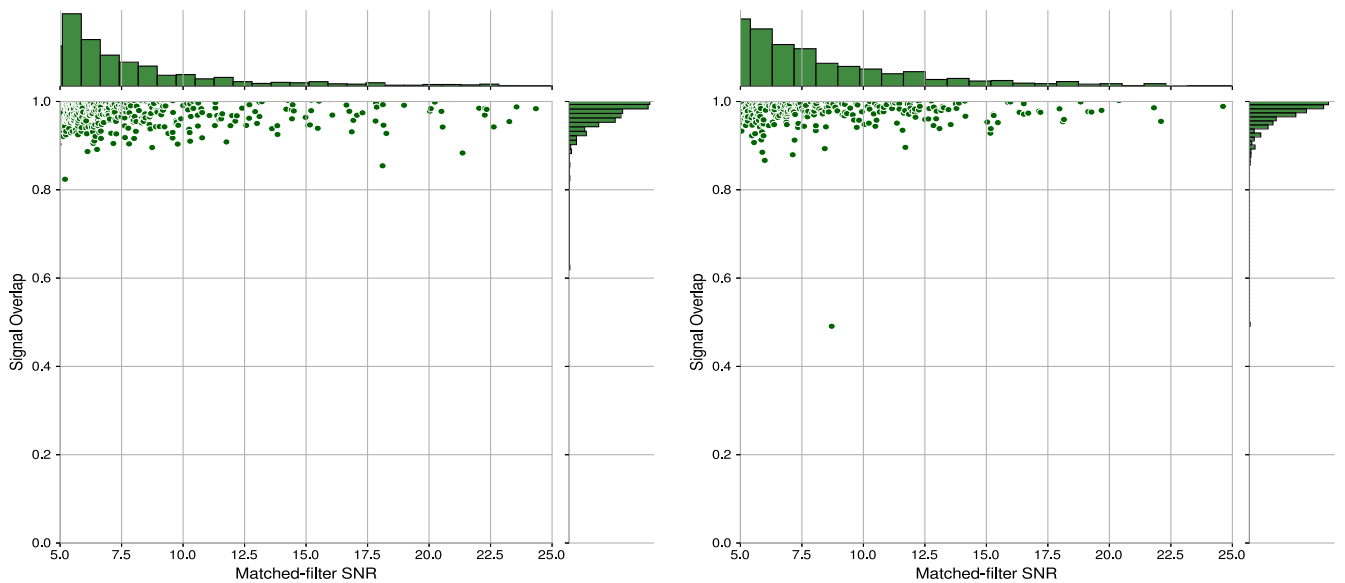


FIG. 4. Signal overlaps after denoising, at Hanford (left) and Livingston (right) as function of the matched-filter SNR (before denoising). Top histograms show the population sample of SNRs used for the analysis, side histogram shows the statistics of signal overlaps.

GW signal, whose signal output is null and the noise output is the entire data. This helped reduce the false positives considerably of the test set. A total of 22 500 waveforms are generated within the mass limits and combined with noise to make the first 22 500 samples in the training data. Another 22 500 samples are pure noise from the detectors with no signal injection. These 45 000 samples are normalized, shuffled and are ready to be fed into the network for training and validation. The spectrogram samples are normalized by subtracting the mean and dividing by the standard deviation. The network denoised spectrograms can be optionally reverted back to original energy before comparing with the template waveform.

C. Training strategy

Our training of the neural-net is a two step process. In step one, the simulated GW signals are combined with simulated noise of aLIGO detectors. We use the “zerodetunedaLIGO” [55] model within the PyCBC python package of aLIGO to simulate the noise. The neural-net is trained exclusively on modeled signal combined with modeled noise, and the parameters of the network are optimized solely based on this. The architecture and hyperparameters that resulted in the maximum convergence and signal recovery is adopted as the optimal configuration of the neural-net. The signal injection process in simulated noise is exactly the same as in the real detector noise described above. We use 45 000 samples for the simulated-noise part of the analysis as well. In step two, the weights of the neural-net from step one are transferred to the new neural-net whose inputs are created by combining simulated GW signals with real detector noise from both aLIGO detectors separately. This technique, called *transfer learning* has been shown to be very robust in enhancing convergence and improving the accuracy of neural networks in many areas, including GW detection [26]. Transfer learning has improved the performance of the neural-net by both reducing the false positives and by improving signal extraction with higher overlap with the template of a given GW. Training the network with both simulated noise and real noise together did not improve the denoising capability as transfer learning, even though in both cases the neural-net is essentially presented with the same training dataset. As noted in [56], networks trained on lower SNR signals generally outperformed the networks with higher SNRs. This has been tested with a few different train/test samples before adopting the SNR population used in the final analysis. A subset of SNRs used in training is presented in Fig. 3 (step one with simulated noise) and Fig. 4 (step two, with real noise). The optimization is performed over 10 epochs using the adam optimizer and a small subset of the training set is used for testing. Mean squared error (MSE) metric is used to monitor the performance of the denoising network. We used Google’s deep learning library, tensorflow [57] for the entire analysis.

In our analysis, we define peak amplitude (PA) of the denoised signal as the detection threshold. Our detection strategy of an unknown signal is as follows. The data segments of 4096s duration are fed into the network as eight second long data segments, and the denoising is performed on all the data segments. The segment with the presence of a real GW will ideally have the highest PA, while the PA of segments with no real GW event will be orders of magnitude lower than the PA of real GW. This is expected because the training of the neural-net was carried out in a way to output zero if no GW was injected into the data stream. In each one hour long data, we decide the cutoff PA as the one with utmost one order of magnitude lower than the strongest signal (with the highest PA), and all other data segments with lower PAs are categorized as noise. In essence, the detection and denoising are performed in a single step, which is the separation of signal from the noise. In effect, the PA of the denoised signal will act as the SNR detection threshold of a matched-filter based detection.

Matched-filter based analysis defines the false alarm rate (FAR) of each detection as the number of false positive detections with an equal or higher ranking statistic, where ranking is assigned to each positive trigger that passes the SNR threshold of the detection pipeline and inter-site travel time requirement [5]. While the FAR threshold used in $O3$ is $< 2/\text{day}$ [7], most detections have an FAR that is of the order of years. This is orders of magnitude smaller than the FARs that are typical to neural networks. In [58], it is noted that the FARs of matched-filter pipelines do not directly translate into the false alarm probability (FAP) characteristic of ML based GW search. In our analysis, we define the FAP based on the peak amplitude (PA) of a denoised GW signal. We define the FAP of a given detection as the ratio of number of denoised signals with a PA greater than or equal to the PA of a real GW signal, to the total number of denoised signals. For instance, if the PA of a real GW signal is 10 (as strain amplitude) and there are four denoised signals that have a PA greater than 10 in a total of 10,000 denoised data segments, the FAP of this detection would be 0.1%.

III. RESULTS

Table I summarizes the FAPs and overlaps of all the detected events from $O2$. In the case of real events reported in this paper, FAP of an event is calculated using several hours of data before and after the event. This is calculated separately at both detectors, and are presented as “single detector FAP” in Table I. The single detector FAPs can be further improved by analyzing more data around the event and even beyond, hence it is not necessarily a fixed number, nor can it be elevated as a detection standard. Two-detector FAPs are defined as the false alarms from the same data segment at the same time at both the detectors. The two detector FAPs are found to be zero for all real GW events detected in our analysis. This means, no false alarms are

TABLE I. Summary of the real GW events analyzed from both aLIGO detectors. Table shows the optimal matched-filter SNRs, single detector FAPs and signal overlaps.

Event	GPS	Matched-filter SNR		Single detector FAP (%)		Signal overlaps (%)	
		Hanford	Livingston	Hanford	Livingston	Hanford	Livingston
GW150914	1126259462.4	20.6	14.2	0.01	0.2	86.1	80
GW170814	1186741861.5	9.3	14.3	0.1	0.25	90	41
GW170104	1167559936.6	9.5	9.9	0.4	0.15	77	85
GW170823	1187529256.5	6.8	9.2	2	1	60	30.5
GW170809	1186302519.8	5.9	10.7	0.05	0.05	68.8	85.6
GW170608	1180922494.5	12.1	9.2	4	0.1	31	48
GW170729	1185389807.3	5.9	8.3	0.6	4	60	67.2
GW170818	1187058327.1	4.6	9.7	...	5	...	70

produced under the condition that the detection needs to be registered at both the detectors at the same time. We used 10,000 data segments of eight second duration to calculate the FAPs presented in Table I. The PA of the real GW event is typically the highest among the analyzed data, and occasional false positives are counted toward the FAP of the event.

The overlap between the template signal h and the denoised signal h^d , with N data points is calculated using the equation [41]:

$$O = \sqrt{\frac{\sum_{i=0}^N h_i h_i^d}{\sum_{i=0}^N h_i h_i}} \quad (9)$$

Figure 3 shows the overlaps of signals after denoising, as a function of the matched-filter SNRs before denoising, with simulated noise. Figure 4 shows the same with real aLIGO noise from both detectors. The increase in signal overlaps between simulated noise and real noise can be seen, which is a result of transfer learning from the simulated noise analysis. For all the real events described in the following part, the comparison and the calculation of overlaps between templates and denoised signals are performed after both being whitened using the same whitening metrics, hence both the template and the denoised signal have similar PSD.

A. GW150914

In both detectors, GW150914 was identified as the signal with the highest PA at the exact time as the real detection. This merger of black holes of masses $35.6M_\odot$ and $30.6M_\odot$ was detected by all three search pipelines operational during O1 (PyCBC, GSTLaL and cWB) with a network SNR of 26. We analyzed data, several hours before and after the merger event to calculate the FAP of the detection, the FAP at Hanford is estimated to be 0.01% and the FAP at Livingston is estimated to be 0.2%. The denoised signals from both detectors are shown in Fig. 5. The overlap between the template of GW150914 which is simulated using the

parameters estimated by Bayesian analysis by aLIGO [2] and the denoised signal is calculated using the signal overlap formula. Denoised GW150914 at Hanford matches very closely with the model template with 86% overlap. At Livingston, the model template matches 80% with the denoised signal. The better denoising performance at Hanford is most likely the result of the higher SNR of the detection at Hanford for this event, although this trend of higher signal overlap on account of higher SNR is not observed for all the events, hence cannot be generalized. For this event, we recovered a few cycles of inspiral phase followed by the merger and ringdown at both detectors. At Livingston, the denoised event is not exactly in phase with the template at early part of the inspiral, but later on it comes to be in phase. At both detectors, the ringdown phase appears to be longer and also appears to be deviating from the signal template at the final stage of ringdown. This trend of “ringing” after ringdown is observed for many of the recovered signals, although this requires further analysis to infer if it is indeed a real physical phenomenon or an artifact. This feature was not observed in either the simulated noise tests or the real noise tests performed for demonstrating the signal overlaps (Figs. 3 and 4). The residuals are shown in separate panels for both detectors, which arise primarily due to the amplitude mismatch and phase differences between the template and the denoised signal, as opposed to the presence of unsubtracted noise. The frequency content of the residual is similar to the frequency content of the event itself, at all three phases event, which also is a consequence of the direct signal extraction, characteristic of the denoising scheme. While in a matched-filtering based detection, any unsubtracted noise can be attributed to realistic noise source, and the residual therefore is a real noise, in the denoising scheme described in this paper, the residual is not necessarily a realistic noise. The residual here is simply the difference between the template waveform and the denoised waveform that the network deems to be signal. Hence it can contain noise and/or real signal that is not represented by the template waveform. The frequency content of the template and the denoised signal is almost the same in all three

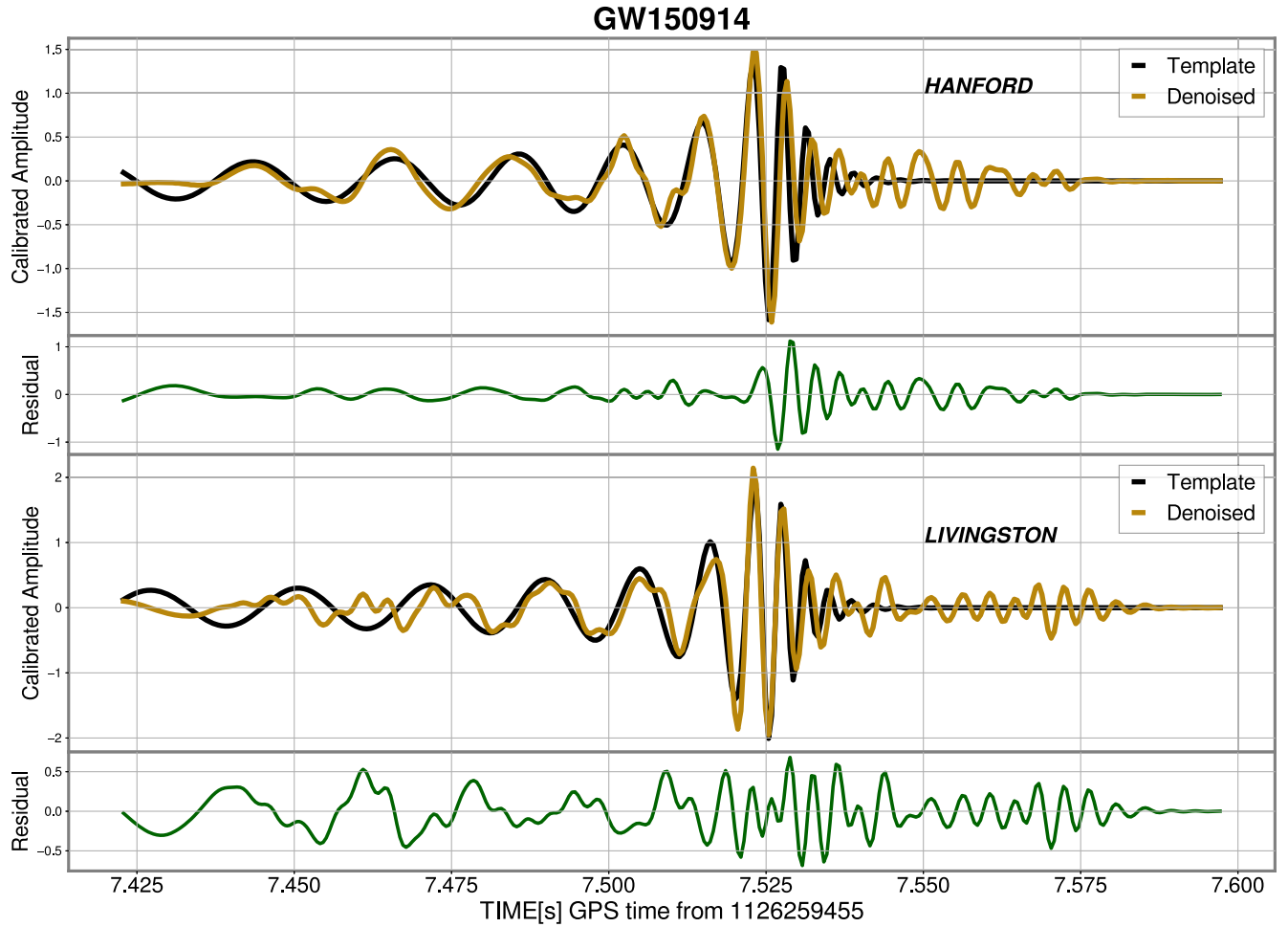


FIG. 5. GW150914 template and denoised waveform at Hanford (top panel) and Livingston (bottom panel). Amplitudes are scaled by a constant scaling factor, to match with the signal template.

phases of the merger. Hence the difference between the two (residual) is also expected to have similar frequency content, as opposed to something drastically different from the template and the denoised waveform. This, again is different from matched filtering based signal extraction, where the frequency content of the residual can be any frequency that belongs to an unsubtracted noise.

B. GW170104

GW170104 is a BBH merger of masses $30.8M_{\odot}$ and $20M_{\odot}$, detected by all three search pipelines with a network SNR of 13.8. At Hanford, the denoised signal is out of phase with the template at the early phase of inspiral and the ringdown phase deviates from the template and is longer, with an overall signal overlap of 77%. At Livingston, the denoised signal matches perfectly with the template at all three phases of the merger, with an overlap of 85%. The single detector FAPs are estimated to be 0.4% and 0.15% respectively and the two detector FAP is zero. See Fig. 6.

C. GW170809

GW170809 is a BBH merger of masses $35M_{\odot}$ and $23.8M_{\odot}$, detected by PyCBC and GSTLal pipelines with a network SNR of 12.8. At Hanford, the signal overlaps fairly well with the template at late stages of inspiral and merger, with an overall overlap of 68%. The ringdown phase deviates from the template toward the end and displays the ringing after ringdown, which sustains for a few milliseconds. At Livingston, the signal overlaps very well with the template at all three phases of the merger, with an overlap of 85%. Single detector FAP at both detectors are estimated to be about 0.05% and two detector FAP is zero. See Fig. 7.

D. GW170814

GW170814 is a BBH merger of masses $30.6M_{\odot}$ and $25.2M_{\odot}$, detected by all three search pipelines with a network SNR of 17.7. At Hanford, the denoised signal overlaps well with the template at all three phases with an overall signal overlap of 90%. There appears to be a

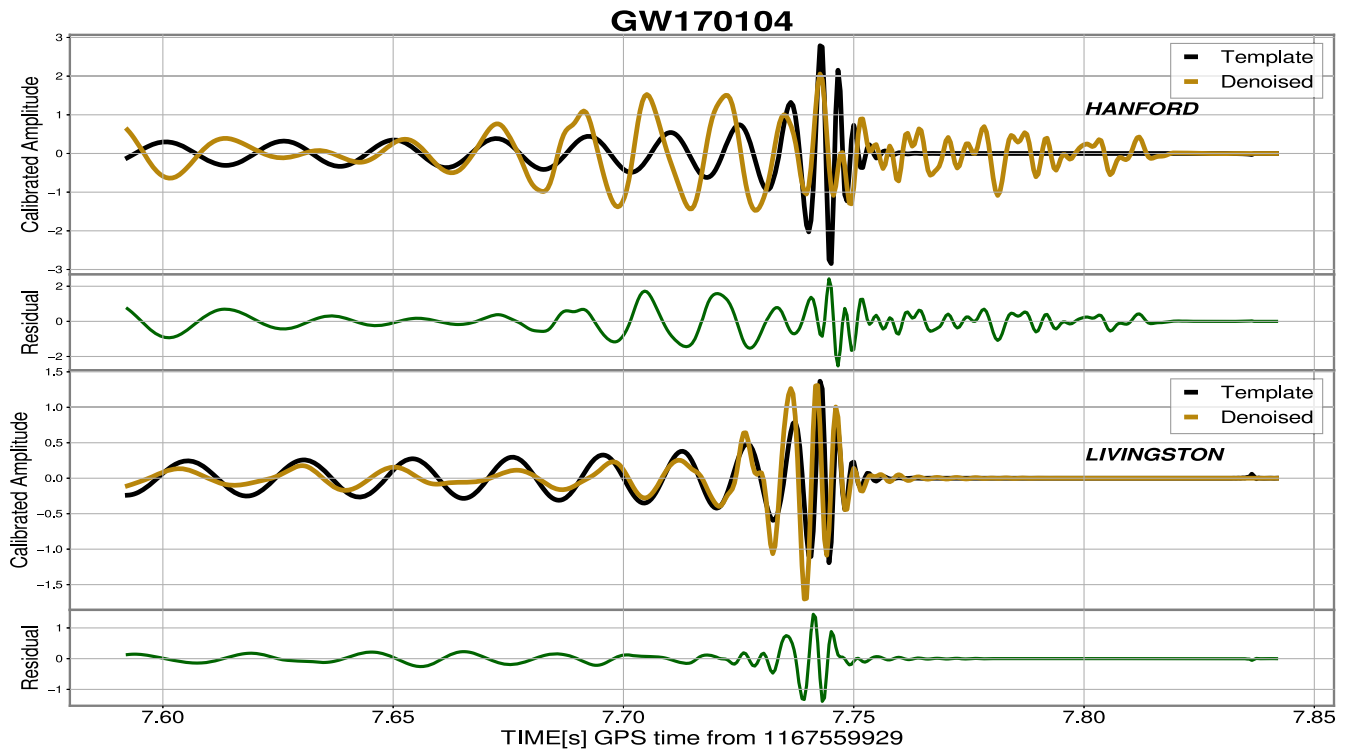


FIG. 6. GW170104 template and denoised waveform at Hanford (top panel) and Livingston (bottom panel). Amplitudes are scaled by a constant scaling factor, to match with the signal template.

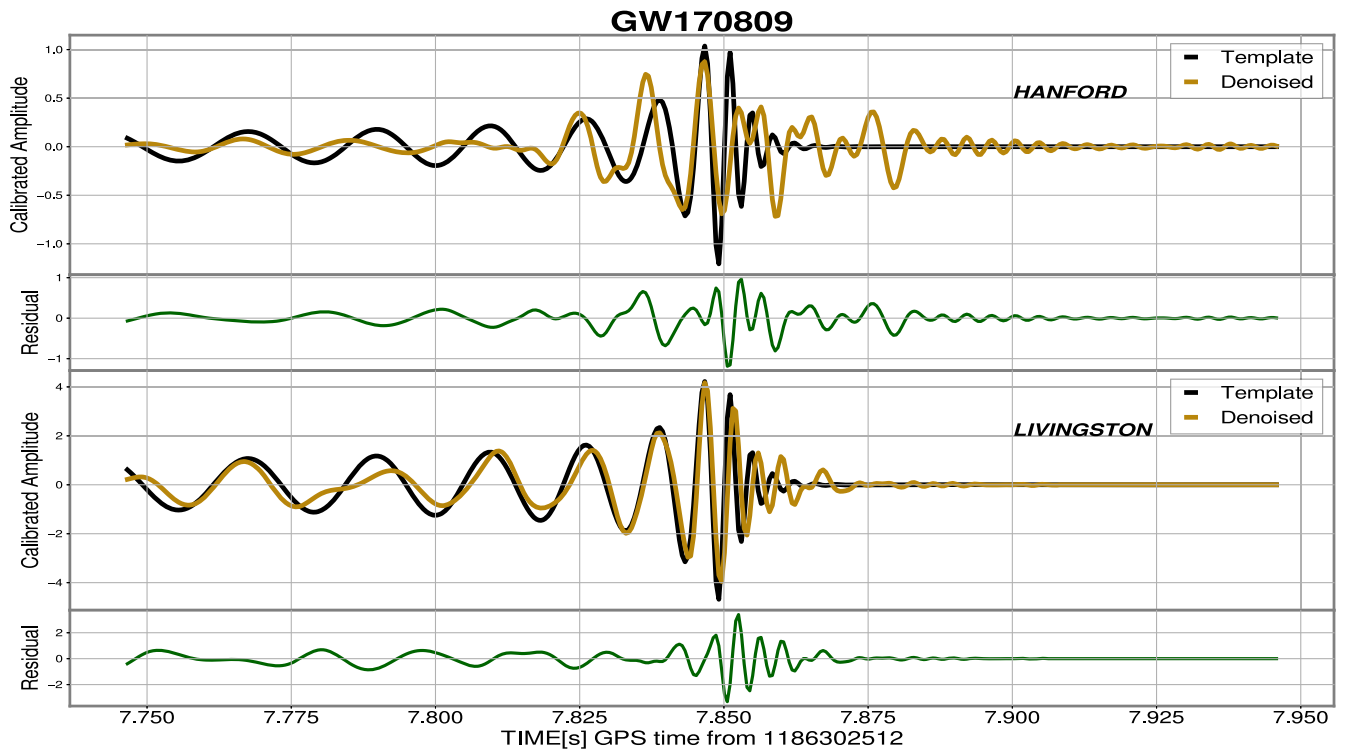


FIG. 7. GW170809 template and denoised waveform at Hanford (top panel) and Livingston (bottom panel). Amplitudes are scaled by a constant scaling factor, to match with the signal template.

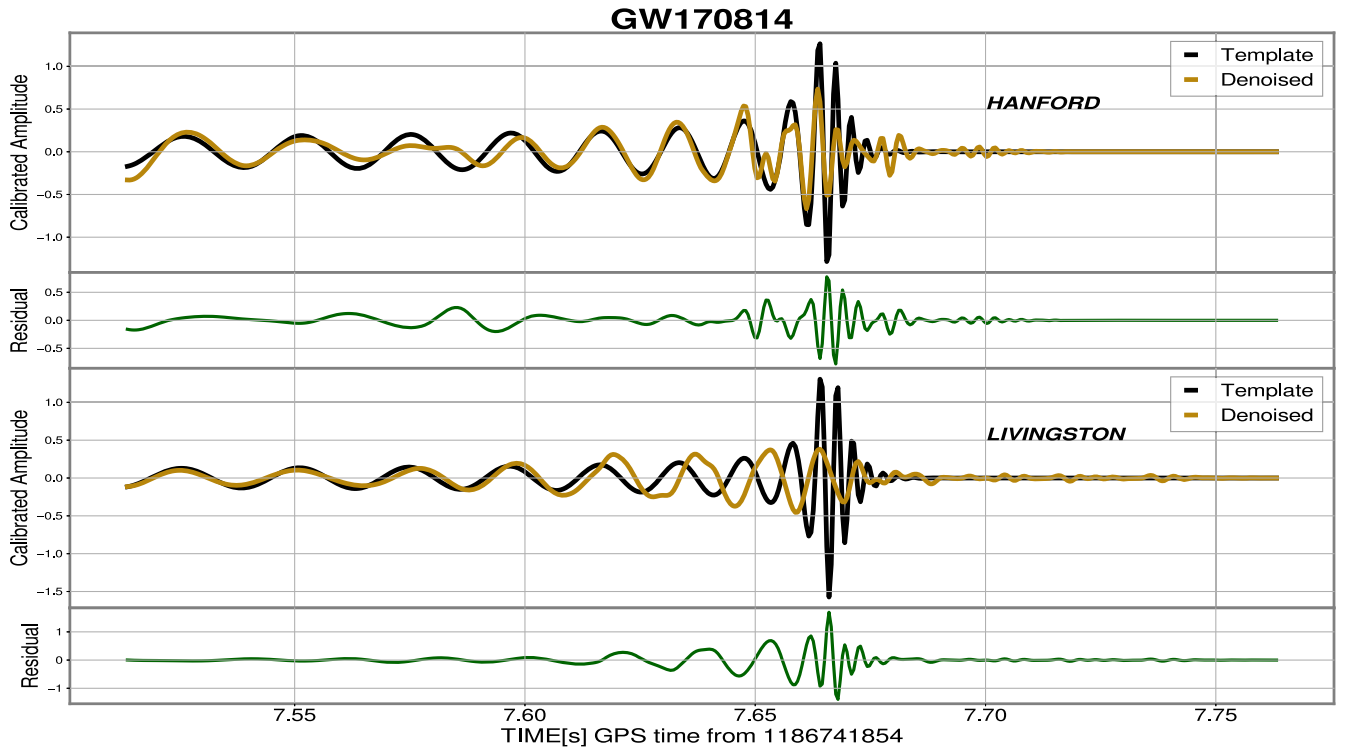


FIG. 8. GW170814 template and denoised waveform at Hanford (top panel) and Livingston (bottom panel). Amplitudes are scaled by a constant scaling factor, to match with the signal template.

“ringing” immediately after the ringdown phase. At Livingston, the signal overlaps well only at early inspiral phase, and the overall signal overlap is only 41%. The denoised merger does not recover the highest frequencies at the merging phase, but the frequency is close at the ringdown phase, which appears to persist longer than the template ringdown. The single detector FAPs are estimated to be 0.1% and 0.25% respectively and the two detector FAP of the detection is zero. See Fig. 8.

E. GW170729

GW170729 is BBH merger of masses $50.2M_{\odot}$ and $34M_{\odot}$, detected by all three search pipelines with a network SNR of 10.8. This is the heaviest system detected in $O2$. The denoised signal has the same frequency at all three phases at the Hanford detector, although out of phase at some portion of the inspiral and the overall signal overlap is calculated as 60%. The extra ringing after the ringdown is clearly visible. At Livingston, the inspiral part largely has the same frequency as the template, although there appears to be higher frequencies riding on top of the lower frequencies, which is observed only for this event. The waveform is in phase with the template at late inspiral and merger. The overall signal overlap is 67%. The ringing after ringdown is visible, with a higher amplitude than the Hanford detector. Single detector FAPs are 0.6% and 4% respectively and the two detector FAP is zero. See Fig. 9.

F. GW170608

GW170608 is a BBH merger event of masses $11M_{\odot}$ and $7.6M_{\odot}$, detected by all three search pipelines with a network SNR of 15.4. This is the lightest event of $O2$ with the highest frequencies at all three phases. At both detectors, the denoised signals deviate from the template as the neural-net fails to recover the highest frequencies at the merger and ringdown, characteristic of the event. At both detectors, the merger appears to be longer than what the template suggests and the ringdowns arrive late, and persists longer than the template. The signal overlaps are only 31% and 48% at Hanford and Livingston, and the single detector FAPs of 4% and 0.1% respectively. The two detector FAP of the detection is zero. See Fig. 10.

G. GW170823

GW170823 is a BBH merger event of masses $35.8M_{\odot}$ and $29M_{\odot}$, detected by all three search pipelines with a network SNR of 12.2. At Hanford, the denoised signal overlaps well with the template at the inspiral phase but does not recover the highest frequencies at the merger. The ringdown deviates from the template and lasts longer and the overall signal overlap is about 60%. At Livingston, the signal overlaps with template well at inspiral phase and reasonably well at merger and ringdowns. The overall signal overlap is calculated to be of only 30%. The single

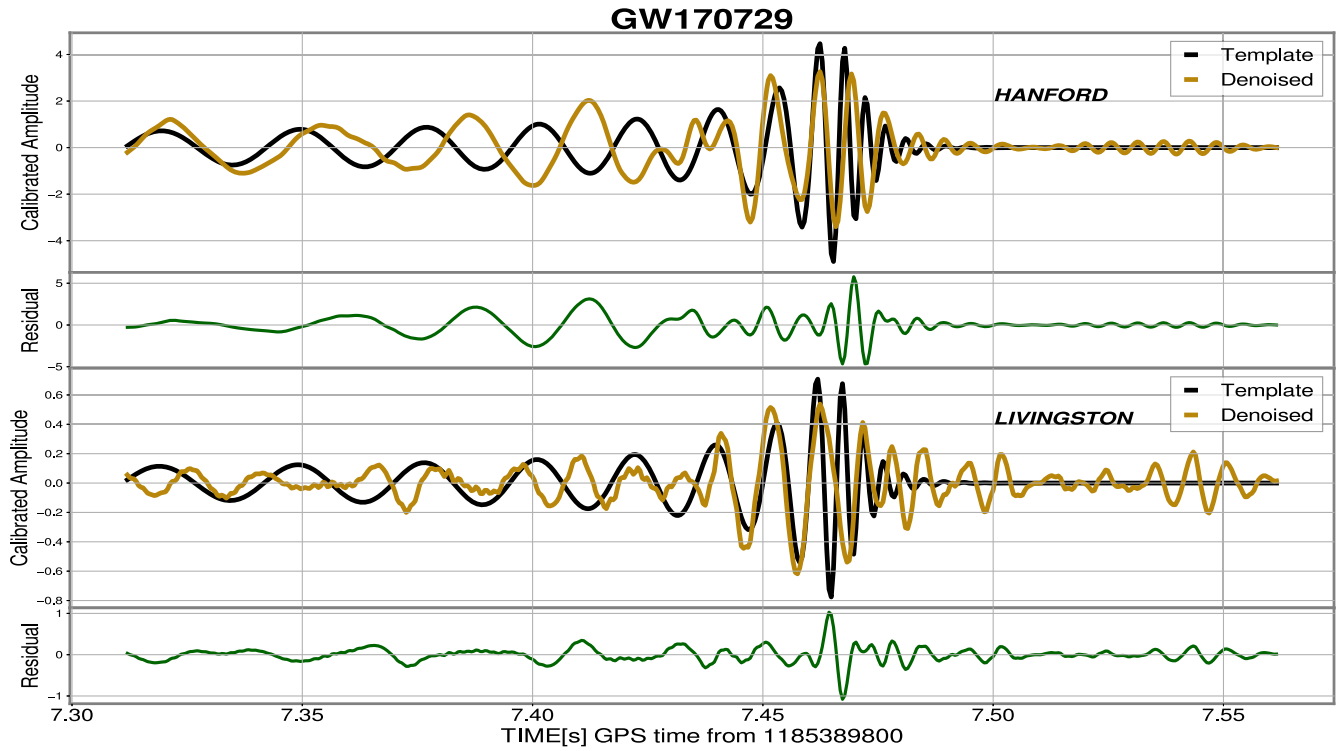


FIG. 9. GW1700729 template and denoised waveform at Hanford (top panel) and Livingston (bottom panel). Amplitudes are scaled by a constant scaling factor, to match with the signal template.

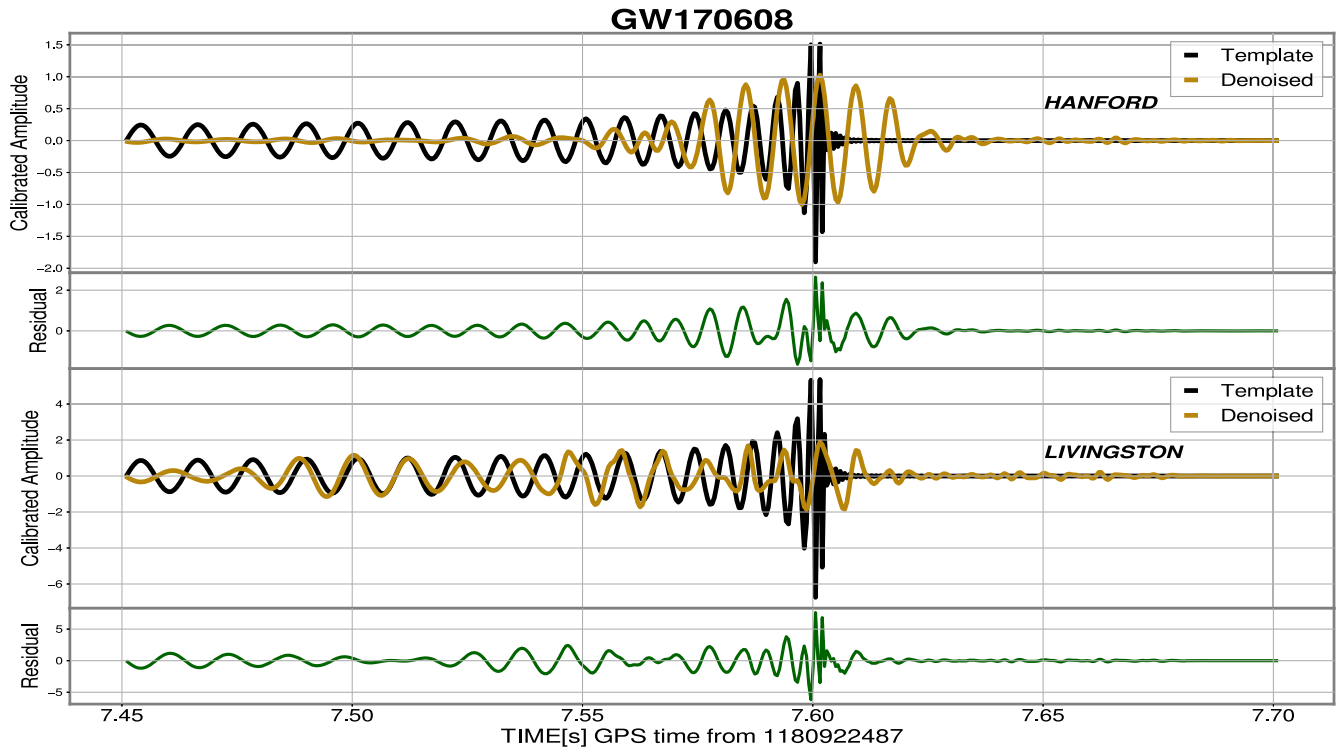


FIG. 10. GW170608 template and denoised waveform at Hanford (top panel) and Livingston (bottom panel). Amplitudes are scaled by a constant scaling factor, to match with the signal template.

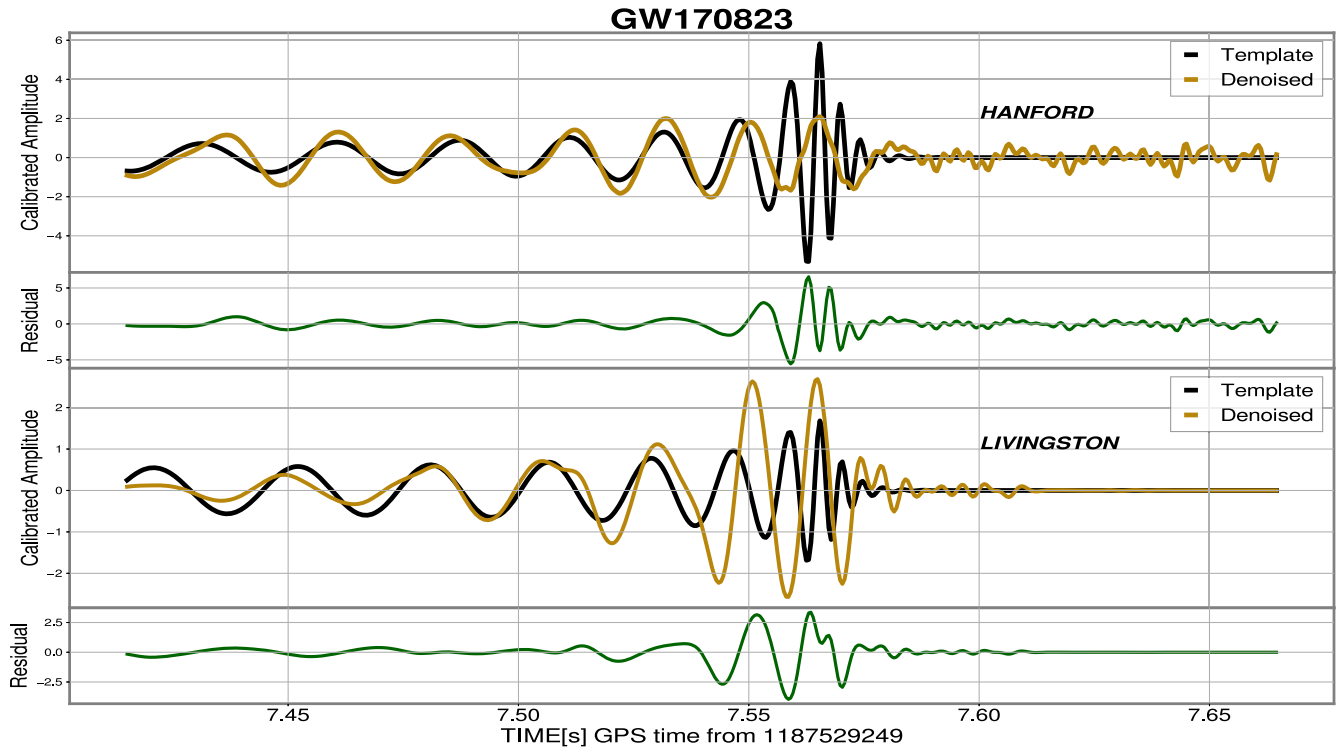


FIG. 11. GW170823 template and denoised waveform at Hanford (top panel) and Livingston (bottom panel). Amplitudes are scaled by a constant scaling factor, to match with the signal template.

detector FAPs are 2% and 1% respectively and the two detector FAP of the detection is zero. See Fig. 11.

H. GW170818

GW170818 is a BBH merger of masses $35.4M_{\odot}$ and $26.7M_{\odot}$, detected by only one of the search pipeline (GstLAL) with a network SNR of 12. The detection of

this event was extremely weak at the Hanford detector with an SNR of only 4.6, which is only slightly above the detection threshold of both GstLAL and PyCBC search pipelines. Our neural-net detected this event only at the Livingston detector, and could not reliably provide a peak amplitude at the Hanford site, hence not included in this article. At the Livingston detector, the frequency largely matched with the template at the inspiral phase, although

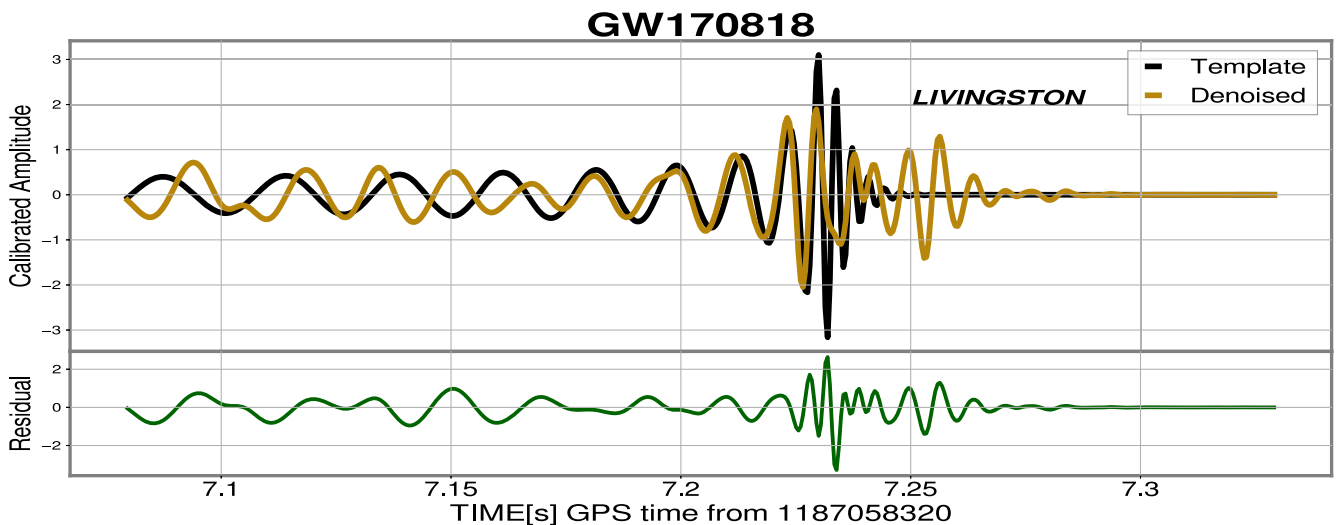


FIG. 12. GW170818 template and denoised waveform at Hanford (top panel) and Livingston (bottom panel). Amplitudes are scaled by a constant scaling factor, to match with the signal template.

not in phase with the template at early inspiral. The merger and ringdown phases are in phase with the template and there is an additional ringing immediately after the ringdown, but not as long as many of the other denoised signals. The signal overlap and FAP are 70% and 5% respectively. See Fig. 12.

IV. DISCUSSION

The results presented in this article are highly encouraging for this innovative CNN technique. This optimization scheme is demonstrated to have the ability to detect and separate GW signals from highly nonstationary and non-Gaussian noise, hence warrants further exploration.

The signal nonoverlap of some of the detected events with the model template is possibly due to the absence of spin effects that is missing in this analysis, especially for the events where the denoised signal is missing a complete cycle, in the inspiral. Spin projections along the direction of orbital angular momentum affect the inspiral rate of the binary. The spin components aligned with the orbital angular momentum increase the number of orbits while the spin components antialigned with the orbital angular momentum decrease the number of orbits from any given separation to the merger with respect to nonspinning case [59,60]. In some of the denoised waveforms, we see certain cycles partially missing, or attenuated in amplitude (at 7.56s and 7.57s of GW170823—Fig. 11) rapidly, without reconstructing the complete cycle. This is more likely an effect related to the CNN itself, as opposed to a physical effect, and can in principle be tackled by further optimizing the hyperparameters of the neural-net or by incorporating some of the latest features developed in various deep learning libraries. Using an expanded set of waveform approximants and a broader parameter space in waveform modeling can also potentially rectify such anomalies. The two-detector FAP is zero for all the detected events, which is promising from the detection and early alert point of view.

There are a multitude of ways to improve this detection strategy. Firstly, this analysis is two dimensional. The only source property that is used in characterizing the signals used for training the neural-net is the black hole masses. Although ML methods are equipped with the capacity to interpolate between waveform templates, we expect the performance of the neural-net to reliably get better by training with a few more source properties such as spins of individual black holes, inclinations and eccentricities. In addition, the training was exclusively based on one family of NR waveforms (SEOBNRv4), which can be expanded with more NR waveform families with higher modes and precession effects included, such as IMRPhenomXPHM [61] and SEOBNRv4PHM [62]. This extended analysis is currently underway and will be communicated in the future. Enhancing the parameter space with more properties of the binary does not compromise on the fast performance of this method to detect and denoise GW, since all the

intensive computations are performed during the training stage, which is a one time process. This is different from the matched-filter analysis, where the computations need to be repeated over and over again for the entire observing run. Although the neural-net is expected to have high fidelity to search for real events after being trained on noise from a previous observing run, the best results can be expected by retraining the network with the current observing run, as the sensitivity of the detector is being improved constantly, and hence the characteristics of the noise is changing over time. Another area of interest is, expanding this formalism to include BNS and NSBH mergers.

Although beyond the scope of this work, there have been discussions in the literature in the past on “echoes” that are reminiscent of the observed ringing phenomenon. These echoes arise as a result of the introduction of structures near the event horizon of black holes. These echoes are late, repeated ringings of the ringdown phase of a black hole merger as a result of waves trapped between the near-horizon structure and the angular momentum barrier, as first discussed in [63,64]. The observational evidence of echoes in a real GW event is further discussed and debated in [65–67]. It requires a careful analysis of the characteristics of these patterns and comparison to the theoretically proposed features of the echoes to confirm if these are indeed observational signatures of the echoes. In the absence of such studies, we do not make any claims in that regard in this work.

V. CONCLUSION

In this article we presented a deep learning framework and its potential to detect and denoise GW signals from black hole binaries. The network learns a sparse representation of the data and separates signal from noise by generating two adaptively thresholding masks. Effectively, we are subtracting noises from all frequency bands including where the frequency content of the noise overlaps with that of the signal. From the unprocessed raw strain data of the detector, the neural-net successfully detected all the black hole binary mergers from the second observing run of aLIGO. This ML based detection strategy is a strong candidate to be incorporated into the search, analysis and parameter estimation of merger events in the upcoming observing runs of ground based detectors. Understanding the nature of various complex noise sources and studying them statistically can be extremely challenging in the context of GW detectors, hence this task can virtually be outsourced to deep neural networks which are equipped to learn the patterns that are characteristics of the noise themselves. It is our view that a neural-net based detection and denoising will work best in conjunction with the matched-filtering method, for both detection and parameter estimation.

With the prospect of detecting hundreds of GW events in the nearby future with current and emerging ground based detectors, analyzing months of detector data output is a

very challenging task. A deep learning search pipeline can be immensely helpful in faster detections, efficient data processing and increase/reduce the confidence of a given detection. An analysis of marginal detections [68] using ML methods is also crucial, which can help confirm or reject the marginal detection, in addition to looking for potentially missed real GW signals.

The transfer learning technique that we adopted to improve the performance can be used to retrain the network with data from the most recent observing run. Retraining or transfer learning the network with new data is fairly easy given the computational resources available today that are exclusively powerful for ML based analysis. Retraining this network takes only a few minutes in a GPU based HPC system. Also, the network can always be updated with new template signals developed that are potential candidates including other astrophysical events beyond CBC.

The online, low-latency search for events which are common place during the observing runs of aLIGO can be made more efficient by incorporating a neural-net based early alert for a multimessenger counterpart. The emerging

hardware and software infrastructure of AI and GPU based parallel computing are on an accelerated trajectory today, which are very promising developments for the GW astronomy in the coming years.

ACKNOWLEDGMENTS

The authors would like to thank the computational resources granted for this research by the Texas Advanced Computing Center (TACC—project AST22014) and Optane cluster resource provided by the Seismology research group (David Lumley, Hejun Zhu) of University of Texas at Dallas. C. M. is supported through the research assistantship program by David Lumley.

Both authors contributed to the conceptualization and design of the work. C. M. performed the initial analysis, interpretation and drafted the manuscript. D. L. supervised the work and contributed to the final analysis and interpretation of the results. Both authors reviewed the results, revised and approved the final version of the manuscript.

-
- [1] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Observation of Gravitational Waves from a Binary Black Hole Merger, *Phys. Rev. Lett.* **116**, 061102 (2016).
 - [2] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Properties of the Binary Black Hole Merger GW150914, *Phys. Rev. Lett.* **116**, 241102 (2016).
 - [3] J. Aasi *et al.* (LIGO Scientific Collaboration), Advanced LIGO, *Classical Quantum Gravity* **32**, 074001 (2015).
 - [4] F. Acernese *et al.* (VIRGO Collaboration), Advanced Virgo: A second-generation interferometric gravitational wave detector, *Classical Quantum Gravity* **32**, 024001 (2015).
 - [5] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs, *Phys. Rev. X* **9**, 031040 (2019).
 - [6] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run, *Phys. Rev. X* **11**, 021053 (2021).
 - [7] R. Abbott *et al.* (LIGO Scientific, VIRGO, and KAGRA Collaborations), GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run, [arXiv:2111.03606](https://arxiv.org/abs/2111.03606).
 - [8] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, *Phys. Rev. Lett.* **119**, 161101 (2017).
 - [9] B. P. Abbott *et al.* (KAGRA, LIGO Scientific, and VIRGO Collaborations), Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA, *Living Rev. Relativity* **21**, 3 (2018).
 - [10] K. Somiya (KAGRA Collaboration), Detector configuration of KAGRA: The Japanese cryogenic gravitational-wave detector, *Classical Quantum Gravity* **29**, 124007 (2012).
 - [11] T. Akutsu *et al.* (KAGRA Collaboration), KAGRA: 2.5 Generation interferometric gravitational wave detector, *Nat. Astron.* **3**, 35 (2019).
 - [12] I. B. *et al.*, LIGO-India (2011).
 - [13] F. Amann *et al.*, Site-selection criteria for the Einstein Telescope, *Rev. Sci. Instrum.* **91**, 9 (2020).
 - [14] S. A. Usman *et al.*, The PyCBC search for gravitational waves from compact binary coalescence, *Classical Quantum Gravity* **33**, 215004 (2016).
 - [15] C. Messick *et al.*, Analysis framework for the prompt discovery of compact binary mergers in gravitational-wave data, *Phys. Rev. D* **95**, 042001 (2017).
 - [16] R. Smith, S. E. Field, K. Blackburn, C.-J. Haster, M. Pürrer, V. Raymond, and P. Schmidt, Fast and accurate inference on gravitational waves from precessing compact binaries, *Phys. Rev. D* **94**, 044031 (2016).
 - [17] J. Schmidhuber, Deep learning in neural networks: An overview, *Neural Netw.* **61**, 85 (2015).
 - [18] Y. Lecun, Y. Bengio, and G. Hinton, Deep learning, *Nature (London)* **521**, 436 (2015).
 - [19] D. George and E. A. Huerta, Deep neural networks to enable real-time multimessenger astrophysics, *Phys. Rev. D* **97**, 044039 (2018).
 - [20] M. Zevin *et al.*, Gravity Spy: Integrating Advanced LIGO detector characterization, machine learning, and citizen science, *Classical Quantum Gravity* **34**, 064003 (2017).

- [21] N. Mukund, S. Abraham, S. Kandhasamy, S. Mitra, and N. S. Philip, Transient classification in LIGO data using difference boosting neural network, *Phys. Rev. D* **95**, 104059 (2017).
- [22] J. Powell, A. Torres-Forné, R. Lynch, D. Trifirò, E. Cuoco, M. Cavaglià, I. S. Heng, and J. A. Font, Classification methods for noise transients in advanced gravitational-wave detectors II: Performance tests on Advanced LIGO data, *Classical Quantum Gravity* **34**, 034002 (2017).
- [23] J. Powell, D. Trifirò, E. Cuoco, I. S. Heng, and M. Cavaglià, Classification methods for noise transients in advanced gravitational-wave detectors, *Classical Quantum Gravity* **32**, 215012 (2015).
- [24] D. George, H. Shen, and E. A. Huerta, Classification and unsupervised clustering of LIGO data with deep transfer learning, *Phys. Rev. D* **97**, 101501 (2018).
- [25] S. Bahaadini, N. Rohani, S. Coughlin, M. Zevin, V. Kalogera, and A. K. Katsaggelos, Deep multi-view models for glitch classification, [arXiv:1705.00034](https://arxiv.org/abs/1705.00034).
- [26] D. George and E. A. Huerta, Deep learning for real-time gravitational wave detection and parameter estimation: Results with Advanced LIGO data, *Phys. Lett. B* **778**, 64 (2018).
- [27] W. Wei, A. Khan, E. A. Huerta, X. Huang, and M. Tian, Deep learning ensemble for real-time gravitational wave detection of spinning binary black hole mergers, *Phys. Lett. B* **812**, 136029 (2021).
- [28] M. B. Schäfer, F. Ohme, and A. H. Nitz, Detection of gravitational-wave signals from binary neutron star mergers using machine learning, *Phys. Rev. D* **102**, 063015 (2020).
- [29] E. A. Huerta *et al.*, Accelerated, scalable and reproducible AI-driven gravitational wave detection, *Nat. Astron.* **5**, 1062 (2021).
- [30] G. Baltus, J. Janquart, M. Lopez, A. Reza, S. Caudill, and J.-R. Cudell, Convolutional neural networks for the detection of the early inspiral of a gravitational-wave signal, *Phys. Rev. D* **103**, 102003 (2021).
- [31] G. Baltus, J. Janquart, M. Lopez, H. Narola, and J.-R. Cudell, Convolutional neural network for gravitational-wave early alert: Going down in frequency, *Phys. Rev. D* **106**, 042002 (2022).
- [32] H. Gabbard, M. Williams, F. Hayes, and C. Messenger, Matching Matched Filtering with Deep Networks for Gravitational-Wave Astronomy, *Phys. Rev. Lett.* **120**, 141103 (2018).
- [33] M. Andrews, M. Paulini, L. Sellers, A. Bobrick, G. Martire, and H. Vestal, DeepSNR: A deep learning foundation for offline gravitational wave detection, [arXiv:2207.04749](https://arxiv.org/abs/2207.04749).
- [34] S. Jadhav, N. Mukund, B. Gadre, S. Mitra, and S. Abraham, Improving significance of binary black hole mergers in Advanced LIGO data using deep learning: Confirmation of GW151216, *Phys. Rev. D* **104**, 064051 (2021).
- [35] J. D. Álvares, J. A. Font, F. F. Freitas, O. G. Freitas, A. P. Morais, S. Nunes, A. Onofre, and A. Torres-Forné, Exploring gravitational-wave detection and parameter inference using deep learning methods, *Classical Quantum Gravity* **38**, 155010 (2021).
- [36] P. G. Krastev, K. Gill, V. A. Villar, and E. Berger, Detection and parameter estimation of gravitational waves from binary neutron-star mergers in real LIGO data using deep learning, *Phys. Lett. B* **815**, 136161 (2021).
- [37] E. Cuoco *et al.*, Enhancing gravitational-wave science with machine learning, *Mach. Learn. Sci. Tech.* **2**, 011002 (2021).
- [38] W. Wei and E. A. Huerta, Gravitational wave denoising of binary black hole mergers with deep learning, *Phys. Lett. B* **800**, 135081 (2020).
- [39] H. Shen, D. George, E. A. Huerta, and Z. Zhao, Denoising gravitational waves using deep learning with recurrent denoising autoencoders, *Proceedings of the ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK* (2019), pp. 3237–3241, [10.1109/ICASSP.2019.8683061](https://doi.org/10.1109/ICASSP.2019.8683061).
- [40] H. Shen, D. George, E. A. Huerta, and Z. Zhao, Denoising gravitational waves with enhanced deep recurrent denoising auto-encoders, *Proceedings of the ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK* (2019), pp. 3237–3241, [10.1109/ICASSP.2019.8683061](https://doi.org/10.1109/ICASSP.2019.8683061).
- [41] P. Bacon, A. Trovato, and M. Bejger, Denoising gravitational-wave signals from binary black holes with dilated convolutional autoencoder, [arXiv:2205.13513](https://arxiv.org/abs/2205.13513).
- [42] A. e. a. Buikema, Sensitivity and performance of the Advanced LIGO detectors in the third observing run, *Phys. Rev. D* **102**, 062003 (2020).
- [43] S. Ballmer and V. Mandic, New technologies in gravitational-wave detection, *Annu. Rev. Nucl. Part. Sci.* **65**, 555 (2015).
- [44] W. Zhu, S. M. Mousavi, and G. C. Beroza, Seismic signal denoising and decomposition using deep neural networks, *IEEE Trans. Geosci. Remote Sens.* **57**, 9476 (2019).
- [45] D. L. Donoho and J. M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, *Biometrika* **81**, 425 (1994).
- [46] M. A. Kramer, Nonlinear principal component analysis using autoassociative neural networks, *Alcheringa* **37**, 233 (1991).
- [47] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, 2016), <http://www.deeplearningbook.org>.
- [48] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion, *J. Mach. Learn. Res.* **11**, 3371 (2010).
- [49] J. Long, E. Shelhamer, and T. Darrell, Fully convolutional networks for semantic segmentation, [arXiv:1411.4038](https://arxiv.org/abs/1411.4038).
- [50] O. Ronneberger, P. Fischer, and T. Brox, U-Net: Convolutional networks for biomedical image segmentation, [arXiv:1505.04597](https://arxiv.org/abs/1505.04597).
- [51] K. He, X. Zhang, S. Ren, and J. Sun, Deep Residual Learning for Image Recognition, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA* (2016), pp. 770–778, [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [52] R. Abbot *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Open data from the first and second

- observing runs of Advanced LIGO and Advanced Virgo, *SoftwareX* **13**, 100658 (2021).
- [53] C. Devine, Z. B. Etienne, and S. T. McWilliams, Optimizing spinning time-domain gravitational waveforms for Advanced LIGO data analysis, *Classical Quantum Gravity* **33**, 125025 (2016).
- [54] A. Bohé *et al.*, Improved effective-one-body model of spinning, nonprecessing binary black holes for the era of gravitational-wave astrophysics with advanced detectors, *Phys. Rev. D* **95**, 044028 (2017).
- [55] LIGO Scientific Collaboration, LALSuite: LIGO Scientific Collaboration Algorithm Library Suite, ascl:2012.021 (2020).
- [56] M. B. Schäfer, O. Zelenka, A. H. Nitz, F. Ohme, and B. Brügmann, Training strategies for deep learning gravitational-wave searches, *Phys. Rev. D* **105**, 043002 (2022).
- [57] M. Abadi *et al.*, TensorFlow: Large-scale machine learning on heterogeneous distributed systems, [arXiv:1603.04467](https://arxiv.org/abs/1603.04467).
- [58] T. D. Gebhard, N. Kilbertus, I. Harry, and B. Schölkopf, Convolutional neural networks: A magic bullet for gravitational-wave detection?, *Phys. Rev. D* **100**, 063015 (2019).
- [59] L. Blanchet, Gravitational radiation from post-Newtonian sources and inspiralling compact binaries, *Living Rev. Relativity* **17**, 2 (2014).
- [60] M. Campanelli, C. O. Lousto, and Y. Zlochower, Spinning-black-hole binaries: The orbital hang-up, *Phys. Rev. D* **74**, 041501 (2006).
- [61] G. Pratten *et al.*, Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes, *Phys. Rev. D* **103**, 104056 (2021).
- [62] S. Ossokine *et al.*, Multipolar effective-one-body waveforms for precessing binary black holes: Construction and validation, *Phys. Rev. D* **102**, 044055 (2020).
- [63] V. Cardoso, E. Franzin, and P. Pani, Is the Gravitational-Wave Ringdown a Probe of the Event Horizon?, *Phys. Rev. Lett.* **116**, 171101 (2016); **117**, 089902(E) (2016).
- [64] V. Cardoso, S. Hopper, C. F. B. Macedo, C. Palenzuela, and P. Pani, Gravitational-wave signatures of exotic compact objects and of quantum corrections at the horizon scale, *Phys. Rev. D* **94**, 084031 (2016).
- [65] J. Abedi, H. Dykaar, and N. Afshordi, Echoes from the Abyss: Tentative evidence for Planck-scale structure at black hole horizons, *Phys. Rev. D* **96**, 082004 (2017).
- [66] G. Ashton, O. Birnholtz, M. Cabero, C. Capano, T. Dent, B. Krishnan, G. D. Meadors, A. B. Nielsen, A. Nitz, and J. Westerweck, Comments on: Echoes from the abyss: Evidence for Planck-scale structure at black hole horizons, [arXiv:1612.05625](https://arxiv.org/abs/1612.05625).
- [67] J. Abedi, H. Dykaar, and N. Afshordi, Echoes from the Abyss: The holiday edition!, [arXiv:1701.03485](https://arxiv.org/abs/1701.03485).
- [68] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run, [arXiv:2108.01045](https://arxiv.org/abs/2108.01045).