Frontiers of Physics
https://doi.org/10.1007/s11467-020-0966-4
Front. Phys.
15(5), 54501 (2020)

RESEARCH ARTICLE

Some optimizations on detecting gravitational wave using convolutional neural network

Xiang-Ru Li[1,†], Wo-Liang Yu[2], Xi-Long Fan[3,‡], G. Jogesh Babu[4]

[1] *School of Computer Science, South China Normal University, Guangzhou 510631, China*
[2] *School of Mathematical Sciences, South China Normal University, Guangzhou 510631, China*
[3] *School of Physics and Technology, Wuhan University, Wuhan 430072, China*
[4] *Pennsylvania State University, University Park, PA 16802, USA*
Corresponding authors. E-mail: †lixiangru@scnu.edu.cn, ‡xilong.fan@whu.edu.cn
Received March 25, 2020; accepted May 6, 2020

This work investigates the problem of detecting gravitational wave (GW) events based on simulated damped sinusoid signals contaminated with white Gaussian noise. It is treated as a classification problem with one class for the interesting events. The proposed scheme consists of the following two successive steps: decomposing the data using a wavelet packet, representing the GW signal and noise using the derived decomposition coefficients; and determining the existence of any GW event using a convolutional neural network (CNN) with a logistic regression output layer. The characteristic of this work is its comprehensive investigations on CNN structure, detection window width, data resolution, wavelet packet decomposition and detection window overlap scheme. Extensive simulation experiments show excellent performances for reliable detection of signals with a range of GW model parameters and signal-to-noise ratios. While we use a simple waveform model in this study, we expect the method to be particularly valuable when the potential GW shapes are too complex to be characterized with a template bank.

Keywords gravitational waves, algorithms, astrostatistics techniques

1 Introduction

Gravitational waves (GWs) potentially give a remarkable opportunity to penetrate unprecedented regions of the universe and compact astronomical objects. Direct detections of GW events by the advanced LIGO and Virgo detectors [1–5] enhance our understanding of gravitational theory and bring gravitational-wave astronomy into an observational science [6–9].

GW burst software pipelines, used by the LVC (LIGO-Virgo Collaboration) search for generic gravitational-wave transients, are implemented with minimal assumptions about the signal waveform, polarization, source direction, and occurrence time based on their time-frequency morphology [10]. Although the main targets of the burst pipelines are unmodeled transient GW events, CBC (compact binary coalescence) signals are well-discovered by the event detection procedures [11]. The parameters of some characteristics of these unmodeled GW signals, such as the central frequency, duration and bandwidth, can be estimated from the time-frequency transform of the observed time series and then converted to physical parameters by some comparison-based schemes with sophisticated waveforms computed from astrophysical models [11].

This work focuses on a procedure to detect the existence of any simulated damped sinusoid GW signals from time series with white Gaussian noises. This type waveform could be a burst of GWs from the ringdown of a perturbed black hole [12], from the post-merger of a binary neutron star (BNS) system, or from excitation of fundamental modes in neutron stars [13]. However, real detector data contains non-Gaussian noise artifacts. Therefore, this work is a preliminary investigation which assumes that such non-Gaussian disturbances have been removed from the data or from the detector [14–16]. The performance representativeness of this work on real detector data needs further studies in future.

In GW detection, a benchmark method is the matched filtering [11, 17–20]. Its typical limitation is the expensive computational requirements. Therefore, some investigations are made based on a machine learning scheme, convolutional neural networks (CNNs), for GW detection, and obtained some exciting results comparable to the matched filtering [21–24].

Based on the knowledges of machine learning and signal processing, the practical performance of a detection system depends on the configurations of such factors as a

*arXiv: 1712.00356.

detection window (duration of a time series under being processed), a sampling rate/data resolution. In related literatures on GW detection, however, the configurations of these factors were not investigated correspondingly. For example, the width of a detection window is 1 second in [21, 22, 25]; two sampling rates are 8192 HZ [21, 22, 25] and 2048HZ [26, 27]. Therefore, this work studies the influences from these factors on GW detection and their optimization, and it is shown that the performance of the CNN-based scheme can be improved evidently by optimizing these factors.

Another interesting issue in this problem is the necessity of time-frequency analysis and selection of frequency analysis methods in CNN-based GW detection scheme. Because the gravitational wave bursts are short transients of gravitational radiation and their time-frequency information is a typical characteristic for GW candidates, time-frequency analysis is a proposed procedure in some GW detection studies based on non-machine learning schemes [10, 28, 29] and many studies for glitch classification [15, 30–33], for example, wavelet decomposition and Q-transform [28, 34], coherent Waveburst (cWB) [10, 35, 36] and omicron-LIB (oLIB) [10, 37], BayesWave (BW) [10, 38, 39], and X-Pipeline [29, 40]. However, the existing investigations on CNN-based GW detections are conducted by inputting a time series into a CNN network without investigations on the effects from this kind frequency procedures. Therefore, the necessity of time-frequency analysis and the affects from time-frequency analysis methods are still two open problems to be studied in CNN-based GW detection scheme.

Based on the above two considerations, this work proposed a GW detection scheme based on CNN and a time-frequency analysis method wavelet packet decomposition, and gave an optimized configuration of a detection window (width of a detection window, overlap length between two successive detection windows) and data resolution. It is shown that the proposed optimization of this work improved the GW detection sensitivity evidently, and the accumulation effects from such factors could not be ignored in GW detection.

2 Data

Let $\boldsymbol{s} \in \boldsymbol{R}^D$ represent an observed data/sample, where $\boldsymbol{s} = \boldsymbol{n}$ or $\boldsymbol{s} = \boldsymbol{n} + \boldsymbol{h}$, $\boldsymbol{R}^D$ is a D-dimensional Euclidean space, D is a positive integer denoting the dimension of a sample, $\boldsymbol{n}$ are some noises and $\boldsymbol{h}$ is a GW signal. Suppose $y \in \{1, 0\}$ is a label of $\boldsymbol{s}$: $y = 1$ indicates that there exists a GW signal in $\boldsymbol{s}$, and $y = 0$ indicates that there is not any GW signal in $\boldsymbol{s}$. For convenience, the cases $\boldsymbol{s} = \boldsymbol{n}$ and $\boldsymbol{s} = \boldsymbol{n} + \boldsymbol{h}$ are referred to as pure noise and signal with additive noise respectively. This work is to establish a model

$$y = f(\boldsymbol{s}) \tag{1}$$

using statistical theories to detect the existence of any GW in our observed data. This is referred to as a classification problem in machine learning community.

In this kind of statistical classification scheme, the knowledge of GW detection is expressed using a set, $S_{tr} = \{(\boldsymbol{s}_i, y_i), i = 1, \cdots, N\}$, of observed/theoretical data and their labels, where $\boldsymbol{s}_i \in \boldsymbol{R}^D$ is a pure noise sample or a signal with additive noise, $y_i \in \{1, 0\}$ is its label indicating whether there exists GW in $\boldsymbol{s}_i \in \boldsymbol{R}^D$. That is to say, the empirical data set S_{tr} is a knowledge carrier. The model $y = f(\boldsymbol{s})$ is estimated from this empirical data set using some computational schemes. For convenience, this estimation is denoted by $\hat{f}$. The empirical set S_{tr} is referred to as a training set. The empirical data $(\boldsymbol{s}, y)$ is called a sample.

To evaluate the performance of $\hat{f}$, some samples independent of S_{tr} are needed. These evaluation samples constitute a set S_{te}, which is referred to as a test set. The independence here is to ensure the objectivity of the model evaluation.

One typical characteristic of the problem is that the GW is usually very weak and the observed signal is contaminated with various disturbances. All of the existing disturbances are collectively referred to as noises. Actually, Gabbard *et al.* [21], George *et al.*[25], and this work are of investigating the feasibility of detecting GW using a machine learning scheme and its potential. In these works, a fundamental assumption is that glitches have been removed from the data or from the detector [14–16, 30, 33, 41]. Therefore, this work modeled the noises approximately using a series of independent Gaussian random variables with zero mean, without any considerations of influences from glitches, other transient sources of detector noise. That is to say, the assumed noise model is a white Gaussian noise. In real applications, however, the intensity of noise is non-stationary and the signal-to-noise ratios (SNRs) on different segments of observed data are inconstant. To simulate these cases, this work computed Gaussian noises with their corresponding variance depending on randomly chosen SNRs (Table 1). Therefore, the noises are non-stationary in this work. These are of earlier stage investigations, and we will study the performance of this kind of scheme based on real detector noise in future works.

This work uses three data quality measures: the optimal SNR ρ_{opt} [21], the ratio between the maximum amplitude of the waveform and the white noise standard deviation $A^{S/N}$, and the matched filtering SNR $\mathtt{SNR}_{mf}$ [42]

$$\rho_{opt} = 2\sqrt{\int_{f_{min}}^{\infty} \frac{|\tilde{h}(f)|^2}{\hat{S}_s(f)} \mathrm{d}f}, \tag{2}$$

$$A^{S/N} = \max(\boldsymbol{h})/\sigma(\boldsymbol{n}), \tag{3}$$

$$\mathtt{SNR}_{mf} = \frac{\langle s, h\rangle}{\sqrt{\langle h, h\rangle}}, \tag{4}$$

where

$$\langle s,h\rangle = 4\ Re\int_0^\infty \frac{\tilde{s}^*(f)\tilde{h}(f)}{\hat{S}_s(f)}\mathrm{d}f, \tag{5}$$

$\boldsymbol{h}(t)$ is a signal/template, $\tilde{\boldsymbol{h}}(f), \tilde{a}(f)$ and $\tilde{b}(f)$ are the Fourier transform of $\boldsymbol{h}(t)$, $a(t)$ and $b(t)$ respectively, $\tilde{a}^*(f)$ is the complex conjugate of $\tilde{a}(f)$, $\hat{S}_s(f)$ is the power spectral density (PSD) estimated using the Welch's method [43–45], f_{min} is the frequency at which we start to accumulate SNR, and $\sigma(\boldsymbol{n})$ is the standard deviation of noise $\boldsymbol{n}$. $\mathtt{SNR}_{mf}$ and ρ_{opt} are commonly used measures in gravitational wave community and their usages are helpful in comparing experimental results with related literatures. $A^{S/N}$ also makes the data simulation easy.

In this work, the GW time series is computed using a damped sinusoid waveform [46]:

$$\boldsymbol{h}(t|t_0) = \begin{cases} 0, & t < t_0 \\ \mathrm{e}^{-\frac{t-t_0}{\tau}}\sin(2\pi f_0(t-t_0)), & t \geq t_0, \end{cases} \tag{6}$$

where t_0 is the occurrence time of a GW, f_0 a central frequency and $\tau = Q/(\sqrt{2}\pi f_0)$ a decay parameter. The frequency f_0 covers a range [40, 200] Hz uniformly, and quality factors Q randomly take three values 3, 9 and 100 [46] as the *fiducial* models in simulated data (this work also tested the detection sensitivity on more Q values).

In a supervised learning scheme, a training set and a test set are two requisites. The training set is a knowledge carrier in this kind GW detection scheme and a proposed scheme should be established by learning from them. The test data set acts as a referee to be used in objectively evaluating the performance of the established model. In application, however, it is difficult to make sure that training data and test data accurately share the same noise contamination, occurrence time t_0, f_0 or SNR. Therefore, each training data and test data are computed based on randomly and independently generated noise, t_0, f_0 and SNR in this work.

To evaluate the performance comprehensively and illustrate the methods optimizing the proposed scheme, this work constructs a series of training sets and test sets with various parameter configurations. More about the purposes can be found in Section 4. To be clear, configurations of these data sets are presented in Table 1.

In each practical case, a GW detection scheme only can learn from a limited number of training data/templates, whose parameters cover a limited number of values. For example, in this work, the training set^1 covers 25 000 cases of t_0 and SNR. It is impossible to enumerate all possible configurations in a training set. In application, however, there are infinite number of potential configurations of parameters, e.g., occurrence time t_0, f_0, $A^{S/N}$, even in a restricted range. The potential cases to be dealt with should have more diversities than that of training data in both theory and application. Therefore, some data with different parameter configurations from that of training data should be processed satisfactorily. To evaluate this performance as good as possible, this work used much more test data than training data, and the parameter ranges of the test data are larger than that of training data.

Table 1 Configurations of data sets used in this work. Occurrence time t_0 of a GW signal is unknown beforehand, therefore t_0 is randomly generated from a given time window and is different from sample to sample. The data sets are computed from the model (6). s: Second, NS: Number of samples, $NS\{m,n\}$: m samples consisting of Gaussian noise and n samples consisting of GWs injected in Gaussian noise. In this table, a range $[a, b]$ of t_0 refers to t_0 being a random number greater than or equal to a seconds and less than or equal to b seconds for each signal in corresponding data set; the ranges of $A^{S/N}$ is defined similarly.

Data sets	$A^{S/N}$	t_0 (s)	NS	$NS\{m,n\}$
Training set^1	[0.2, 0.8]	[0, 0.03]	50000	{25000, 25000}
Test set$^1_1$	[0.15, 1.05]	[0, 0.026]	190000	{95000, 95000}
Training set^2	[0.2, 0.8]	[0, 1]	50000	{25000, 25000}
Test set$^2_1$	[0.15, 1.05]	[0, 1]	190000	{95000, 95000}

In applications, data are obtained continuously in a form of streams. A GW detection software pipeline needs to move a time window on the data stream, extract a data segment from every window, and determine whether there exists any GW (Fig. 1) in this data segment. This work proposed a way to move the detection window on the data stream with some overlapping to increase detection performance. A segment of the data from every window is an above-mentioned sample $\boldsymbol{s}$ and each sample should be input into the proposed scheme to determine the existence of any GW signal.

Furthermore, the experiments of Fig. 7 used the matched filtering SNR [Eq. (4)] which is computed based on a template bank. In this work, this template bank consists of the 25 000 GW signals in Training set^1 (Table 1).

3 Principles of the detection scheme

This section introduces the fundamental principles of the proposed scheme, Section 4.1 and Table 2 present its configurations.

3.1 A framework of the proposed scheme

Theoretically, we could directly use the model in Eq. (1) to determine whether there exists a GW signal in $\boldsymbol{s}$. However, it is shown that there are some evident differences between a GW signal and noises (Fig. 1) in frequency space (Fig. 2). Therefore, we can project the observed data $\boldsymbol{s}$ to be analyzed into a frequency space,

$$\boldsymbol{z} = g(\boldsymbol{s}), \tag{7}$$

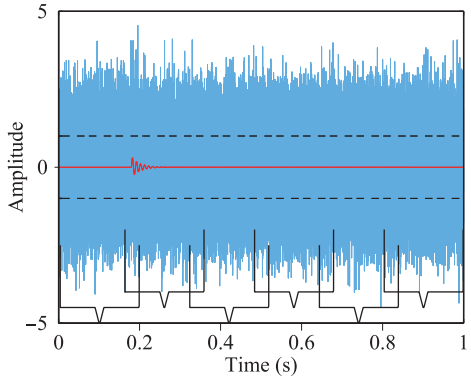

Fig. 1 A sample of simulated data. In this figure, the blue/outer curve is a sample to be analyzed, which consists of GW signal and noises. The red/inner curve is the theoretical GW signal. $A^{S/N}$ of this work covers a range [0.15,1.05]. For visibility, this figure shows a sample with $A^{S/N}$= 0.3. In application, data is obtained in stream. The GW detection system needs to move the detection window on the stream to take the data segment from it to analyze. This diagram shows our proposed scheme for moving detection window (indicated with brackets). Two other parameters of this signal are f_0=100 Hz, Q=9. Two horizonal and dashed lines indicate the standard deviation of noise.

before determining whether there exist any GW components in $\boldsymbol{s}$ as follows:

$$y = h(\boldsymbol{z}) = h(g(\boldsymbol{s})). \tag{8}$$

$\boldsymbol{z}$ is a representation of $\boldsymbol{s}$ in a frequency space. In this work, the function g is a WP decomposition, and the function h is estimated using a CNN network from a training set.

Actually, the scheme based on Eqs. (7) and (8) is a special implementation of model (1). This scheme splits model (1) into two sequential procedures: preprocess the observed data $\boldsymbol{s}$ by $\boldsymbol{z} = g(\boldsymbol{s})$, and determine the existence of any GW signal in $\boldsymbol{s}$ by $y = h(\boldsymbol{z})$. This splitting scheme can simplify a complex work to some degree. In machine learning and data mining communities, the procedure $\boldsymbol{z} = g(\boldsymbol{s})$ is called feature extraction. The feature extraction is to simplify the establishing of model (8) by finding a more approximate representation of the data to be analyzed, $\boldsymbol{s}$ here, and removing irrelevant or weakly-related data components. A flow chart of the proposed scheme is presented in Fig. 3. We will elaborate it in the following parts of this section.

In this proposed scheme (Fig. 3), the WP decomposition is a time-frequency analysis method. Actually, there are a series of investigations on the applications of time-frequency for GW data analysis. For example, Chatterji *et al.*[28] studied the methods for detecting gravitational-wave bursts based on wavelet transform and Q-transform; Klimenko *et al.* [35, 36] and Abbott *et al.* [10] investigated a coherent Waveburst (cWB) algorithm for detection and reconstruction of gravitational wave bursts; Abbott *et al.* [10] and Lynch *et al.* [37] researched a framework, omicron-LIB (oLIB), for detecting short-duration gravitational-wave bursts; Cornish and Littenberg [38], Littenberg and Cornish [39] and Abbott *et al.* [10] studied a BayesWave (BW) approach for searching "un-modeled" transient signals; Chatterji *et al.* [40] and Sutton *et al.* [29] presented a fully automated search software package, X-Pipeline, for detecting bursts associated with gamma-ray bursts (GRBs) and other astrophysical triggers by coherently analyzing the data from interferometer networks. However, the existing CNN-based investigations for GW signal detection are conducted by inputting a time series into a CNN network without studies on the effects from this kind frequency procedures. Therefore, the necessity of time-frequency analysis and the affects from time-frequency analysis methods are still two open problems to be studied in CNN-based GW detection scheme, and this work studied the WP decomposition for the CNN-based GW signal detection scheme.

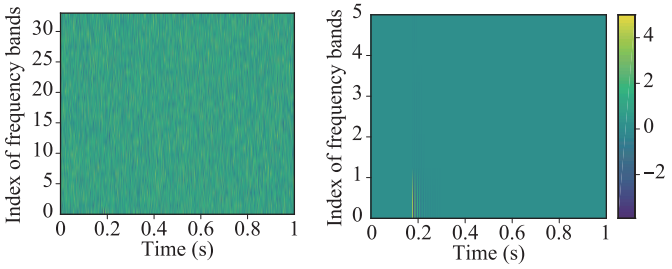

Fig. 2 Differences of the noise and a GW signal in a time-frequency space. The two subfigures are the time-frequency analysis of the noise and GW in Fig. 1, respectively. In this figure, a color indicates the intensity of the corresponding data component. The horizontal axis is time, and the vertical axis is the index of frequency bands in its WP decomposition using Daubechies basis [47]. The frequency in this figure and that of the GW signal are different although there are some positive correlation between them. The GW signal is too weak to be observed in its original time-frequency space. To increase the visibility of the GW signal, therefore, we reduce the range of vertical axis from [0, 32] to [0, 5] in the second subfigure.

Apart from the WP transform, Fourier transform [48], wavelet transform [47, 49, 50], and principal component analysis (PCA) are three typical frequency analysis methods. For example, Rampone *et al.* [51] investigated the glitch-burst discrimination and glitch classification problems based on PCA and forward neural network (FNN), and Vinciguerra *et al.* [52] studied a scheme to classify waves emitted in CBC using PCA and an average of FNNs with three hidden layers. As an earlier stage investigation, however, this work focuses on the applica-

Fig. 3 A flowchart of the proposed scheme. In this flowchart, the "Data" are the data segments in Fig. 1. WP: Wavelet packet; CNN: Convolutional neural network.

tion of WP decomposition in the CNN-based GW signal detection, and postpones the systematic studies on effects of various frequency analysis methods to future works. Our experiments show that WP is applicable in the GW detection problem and there is not significant difference on GW detection between typical wavelet bases: Biorthogonal basis (bior), Coiflets (coif), Daubechies basis (db), Haar (haar), ReverseBior (rbio), and Symlets (sym) [47, 49, 50].

Therefore, we use a WP transform with a db1 base in Eq.(7). There are multiple variants for basis function db in the implementation of Matlab wavelet toolbox [53]. The number behind db is the index of a specific variant. A brief introduction to the principle and implementation of WP transform can be found in Ref. [54] That introduction needs very little mathematical knowledge to read. For WP decomposition has both time analysis and frequency analysis capabilities, its decomposition result of an observed data (Fig. 1) is an image in a two-dimensional space (Fig. 2).

3.2 Convolutional neural networks

One typical characteristic of the GW detection problem is that the occurrence time of GW signal is uncertain in the detection window (Fig. 1). After a WP decomposition, a sample to be analyzed is represented using a two-dimensional image with some translation uncertainties (Fig. 2). For this kind of problem, a typical scheme is the CNN [55, 56]. The CNN has been widely investigated in image understanding and computer vision [57].

The structure of a CNN used in this wok is presented in Fig. 4. This CNN network consists of an input layer, several composite computing units (CCUs), a fully connected layer and a logistic regression layer [Fig. 4(a)]. The input layer receives the WP decomposition (Fig. 2) of a data segment to be processed (Fig. 1). A CCU consists of a convolutional layer, a pooling layer, and an activation layer [Fig. 4(b)]. This work used a CNN with two CCUs.

There are two key concepts, convolution kernel (CK) and pooling for CNN, which represent the computations in the convolutional layer and the pooling layer respectively. A CK characterizes a discriminant pattern in a problem to be investigated (Fig. 5). In case of the data inputted to a convolutional layer being a matrix or a tensor, the CK is defined using a matrix and a tensor respectively, where a tensor is a stack of some matrices. In a convolutional layer, we move a CK on the data to be analyzed, and do convolution computation between the CK and the data segment under the CK. By the convolution operation, the presence of some discriminant characteristics of GW signals can be evaluated. The results of a convolutional operation are referred to as convolution responses. The input into a convolutional layer is the output from its previous layer of the neural network. The previous layer can be the input layer of the CNN or an activation layer (Fig. 4).

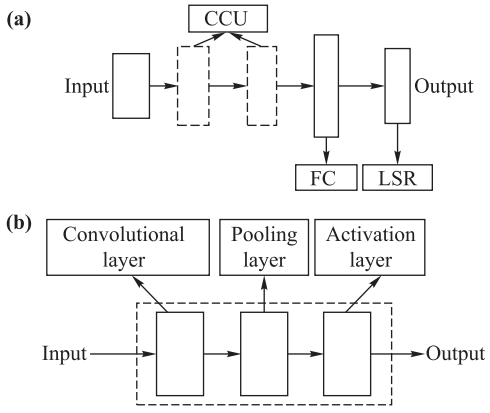

Fig. 4 The structure of a convolutional neural network (CNN). **(a)** The overall structure of a CNN; **(b)** The structure of a composite computing unit (CCU). In this work, we used two CCUs, one fully connected layer (FC) and one logistic regression layer (LSR). Therefore, this work uses two convolutional layers (b). Every convolutional layer consists of a series of convolutional kernels [Fig. 5(a)]. The configurations of this work's CNN are described in Table 2.

In a CNN, multiple CKs can be used for each convolutional layer, the size and number of the CKs on different layer are independent from each other in case of more than one convolutional layer existing. This work uses two convolutional layers, and there are 15 CKs and 20 CKs on these two layers respectively. Furthermore, selections of the CK(s) have some fundamental influences on detection performance, and should be made based on the characteristics of the problem to be investigated. For the knowledge of GW is embodied in training data in this kind of statistical schemes, configurations of the CK are learned from training data. The learning method is a back propagation algorithm [58].

In the outputs of a convolutional layer, there is usually much redundancy. The existence of redundancy can result in evident degradation of detection performance and computational efficiency on test data. To overcome this kind of problem, an operation, pooling, is adopted. The pooling reduces the redundancies from data by merging the convolution responses in every window [Fig. 5(b)], which is referred to as a pooling window. The merging is implemented by computing the maximum of the convolution responses in this pooling window in this work. This pooling window is a 4×4 matrix in this work. More about pooling can be found in Ref. [59].

4 Experimental evaluations

4.1 Learning of the proposed scheme and its specific configuration

The structure of the GW detection system is introduced in Fig. 3 and Fig. 4. In this system, the following parameters need to be determined: number of CCUs, number of

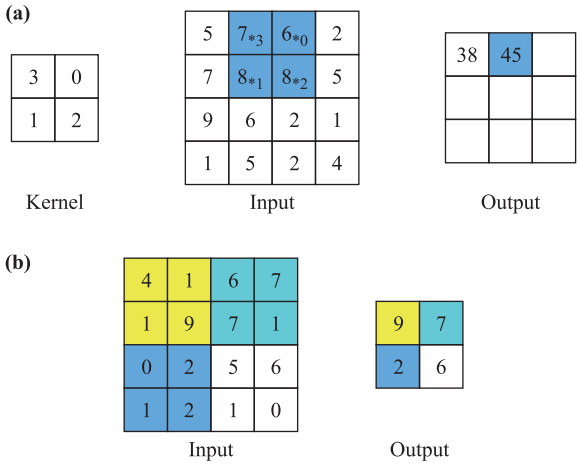

Fig. 5 Two sketch maps of a convolutional operation and a max pooling. **(a)** The output 38 is computed as $3 \times 5 + 0 \times 7 + 1 \times 7 + 2 \times 8 = 38$ by aligning the kernel with input at 5, where 3, 0, 1 and 2 come from the kerenl, and 5, 7, 7 and 8 come from the input; the output 45 is computed as $3 \times 7 + 0 \times 6 + 1 \times 8 + 2 \times 8 = 45$ by aligning the kernel with input at 7. **(b)** Max pooling is to obtain an output by applying a max filter to some subregions of the initial representation: output 9 is computed by maximizing 4, 1, 1, 9 (yellow part), 7 computed by maximizing 6, 7, 7 and 1, and so on. The input in (a) is a time-frequency image (Fig. 2). This work used two convolutional layers [Fig. 4(b)] with 15 convolution kernels and 20 convolution kernels respectively.

CKs in every convolutional layer, size and configurations of every CK, size of pooling window and the parameters in the logistic regression layer.

Theoretically, the above-mentioned parameters can be selected based on some optimization theories using training data. However, it is a hybrid, complex optimization problem consisting of both discrete parameters and continuous parameters. For example, the number of CCUs, the number of CKs in every convolutional layer, the size of every CK and the size of a pooling window are discrete parameters, should be positive integers. Configurations of every CK and the parameter of logistic regression are from real space and continuous parameters. Therefore, it is a complex and hybrid optimization problem to learn the proposed scheme.

To make its computational complexity be acceptable, the discrete parameters are chosen based on experiences. We used two CCUs, 15 and 20 convolution kernels respectively for the two convolutional layers from the input end to the output end of the CNN [Fig. 4(a)]. The continuous parameters are estimated using the back propagation algorithm [58] from training data (Section 2).

Some detailed configurations of the proposed scheme are presented in Table 2. The used training sets are described in Table 1. Figure 6 shows the architecture of a proposed implementation for the scheme in this work.

In summary, the input of the proposed scheme (Fig. 6) is a segment of a time series (Fig. 1), and the output indicates whether there exists any GW signal in a detection window where the segment is extracted from. Therefore, if an output implies an occurrence of a GW signal, the time interval of the data segment window is an estimation of the GW occurrence time, and it is not necessary to combine the CNN results from multiple segments together. Otherwise, if we combine the outputs computed from several segments from different windows by maximizing them, then we can only know whether there exist some GW signals in these windows, but lose the number of times GW event occuring and the occurrence time estimation(s).

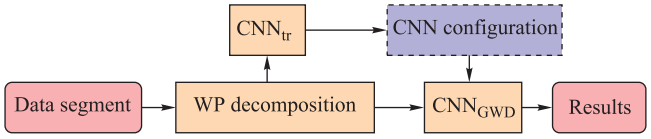

Fig. 6 A diagram to show the modules for a software pipeline implementing the proposed scheme and data flow in it. The $\mathrm{CNN}_{\mathrm{tr}}$ is to learn the configurations of CKs, the connections in step 10 of Table 2, and the parameters in logistic regression. $\mathrm{CNN}_{\mathrm{GWD}}$ detects the existence of any GW in a data segment.

4.2 Evaluation measures

This work uses sensitivity and a receiver operating characteristic (ROC) curve to evaluate the proposed scheme. The ROC curve is defined by sensitivity and False Positive Rate (FPR). In some literatures, the sensitivity is also referred to as recall, true positive rate (TPR) or true-alarm probability. To give their definitions, we introduce several notations.

Suppose S is a set of observed data,

$$S_1 = \{(s,y) : (s,y) \in S \text{ and there exists GW in } s\},$$
$$S_2 = \{(s,y) : (s,y) \in S \ \texttt{and} \ (s,y) \notin S_1\},$$

and $\hat{S}_1$ and $\hat{S}_2$ are respectively the estimations of S_1 and S_2 based on the proposed GW detection system. That is to say, S_2 consists of only noise cases, and each sample in S_1 is superimposed by noise and GW. Therefore, the GW detection problem can be dealt with by determining whether an observed data segment belongs to S_1 or S_2 [Eq. (1)].

Using the above-mentioned notations, the sensitivity and FPR on S are defined as follows:

$$\texttt{Sensitivity}(S) = |S_1 \cap \hat{S}_1|/|S_1|, \tag{9}$$

$$\texttt{FPR}(S) = \frac{|\hat{S}_1 \cap S_2|}{|S_2|}, \tag{10}$$

where $|\cdot|$ is to count the number of samples in a data set.

Table 2 The architecture of the proposed scheme. For the experiments in Section 4.2, the inputs come from the training set$_1^1$ in Table 1 in learning stage, and test set$_1^1$ (Table 1) in the performance evaluation stage. WP: Wavelet packet, CK: Convolutional kernel, ReLU: Rectified Linear Unit, DSt: Data structure. DCI: In the layer with index 3 (Layer 3), the result is a 32×126 matrix from computations based on each CK, the output of Layer 3 is a tensor by stacking the results of 15 CKs; in the output of this layer, there is a third dimension, induced by the index of CKs, which is referred to DCI (Dimension of CK index). The size of the input is 4096.

Index	Procedures	DSt (data size)	Description
1	Input	Vector (4096)	Extract a segment as a sample from a data stream
2	WP decomposition	Matrix (32×128)	Conduct a 5-level WP decomposition using a Daubechies basis
3	Convolutional layer	Tensor (32×126×15)	15 CKs, size of every CK is 1×3 (frequency axis×time axis)
4	Pooling layer	Tensor (8×32×15)	Size of the pooling window is 4×4
5	Activation layer	Tensor (8×32×15)	Using the ReLU [59] as an activation function
6	Convolutional layer	Tensor (8×31×20)	20 CKs, size of every CK is $1 \times 2 \times 15$, (frequency axis×time axis × DCI)
7	Pooling layer	Tensor (2×8×20)	Size of the pooling window is 4×4
8	Activation layer	Tensor (2×8×20)	Using the ReLU as an activation function
9	Flatten layer	Vector (320)	Reshape data from a tensor into a vector by stacking its elements
10	Full connected layer	Vector (10)	Consists of 10 hidden units with the ReLU as activation function
11	Activation layer	Vector (10)	Using the ReLU as an activation function
12	Logistic regression layer	Vector (2)	Estimate the probability of a sample belonging to each class

Therefore, FPR is an indicator of false alert (false-alarm probability in GW community language).

4.3 Eperimental evaluations

In literatures, two typical methods for GW detection are matched filtering [11, 17–20] and CNN [15, 21, 22, 30]. The matched filtering does GW detection by comparing an observed data with each template in a bank. In comparing evaluation, the matched filtering uses each data element both of noise and interesting GW signal. On the other hand, the CNN is a machine learning method which can establish a mapping from some experiences (training data). This mapping of the CNN with a logistic regression output layer can give a result indicating whether there exists any GW signal in the input data. And the CNN-based scheme in literatures shows comparable detection performance with the benchmark scheme matched filting [15, 21, 22, 30].

In theory, however, the performance of a GW detection system depends on a series of factors, for example, the configuration of detection window, data resolution, time-frequency analysis, and configurations of the CNN. Therefore, this work did some improvements upon these factors based on the pioneering works [22] and [21]. This paper proposed a detection window with a duration 0.03 seconds and an overlap 0.004 seconds between two successive detection windows. A sample is represented using a series of amplitude values sampled evenly on a time axis in this work. Data resolution (DR) is measured using Hertz (Hz), which refers to the number of pixels/amplitude values per second (pixels are evenly sampled on time axis in this work). This work uses the DR $= 1.37 \times 10^5$Hz and wavelet decomposition for time-frequency analysis. The experimental results in Fig. 7 show the effectiveness of these improvements.

To observe the effectiveness of each improvement individually, this work conducted a comprehensive evaluation by gradually changing the detection scheme from that of Gabbard *et al.* [21] to this work's framework (Table 2 and Section 4.1) (Thank H. Gabbard, M. Williams, F. Hayes, and C. Messenger very much for their enthusiastic discussions and sharing of their experimental code). The configuration and results of this evaluation experiment are presented in Table 4, Fig. 8 and Table 3 respectively.

In this evaluation, experiment SE1 utilize the configurations of Gabbard *et al.* [21], for example, CNN structure, data resolution, detection window. The experiments SE2, SE3, SE4, SE5 test the performance of CNN structure and detection window width optimization, data resolution, wavelet packet decomposition and detection window overlap scheme (Table 4). The results in Table 3 and Fig. 8 show that the proposed optimization scheme have evident improvement on detection performance. This work and Gabbard *et al.* [21] used different theoretical model for simulating gravitation wave signal. Therefore, there exist some differences between the detection results of SE 1 and Gabbard *et al.* [21]. After these optimizing strategies on time-frequency analysis, detection window, data resolution and pooling scheme, the CNN-based method shows some evident potentials to detect the GW much more sensitively than the matched filtering (Fig. 7).

Comparing with the matched filtering method, one remarkable characteristic of the proposed scheme is its efficiency [21, 22]. Although the learning of the configuration for a CNN is the most time-consuming stage in the proposed scheme, it can be done beforehand. In test stage, we only need to compute the output of a CNN using the learned parameters. However, matched filtering method will evaluate the similarity of every sample to be checked with each template in a bank. For convenience, this characteristic of matched filtering is referred to as every-

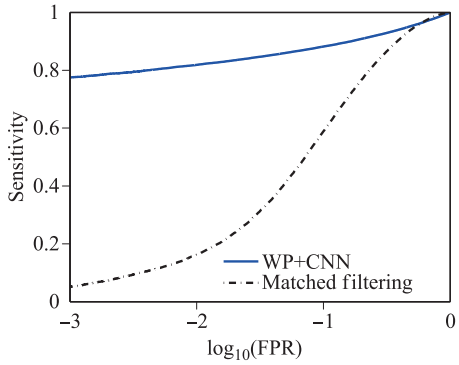

Fig. 7 Performance comparison between the proposed scheme and the matched filtering. The evaluation results in this figure is computed on Test set$_1^1$ (Table 1).

template-evaluation-for-each-test-sample (ETEETS). To have an accurate detection performance, the bank should span a large astrophysical parameter space and consist of many samples/templates. The huge volume of the bank and the ETEETS characteristic result that the matched filtering method is computationally inefficient. The experiments of this work show that the proposed CNN-based scheme takes 18.765 milliseconds to process a sample on a computer using twelve Inter(R) Xeon(R) CPUs (E7-4830) with frequency 2.31GHz, while the matched filtering 1899.350 milliseconds to process a sample. One additional note, in the experiment SE 5 (Fig. 8), the training time of the proposed CNN model is approximately 430 minutes, and the Wavelet Packet decomposition time is approximately 25 minutes for the training set[1].

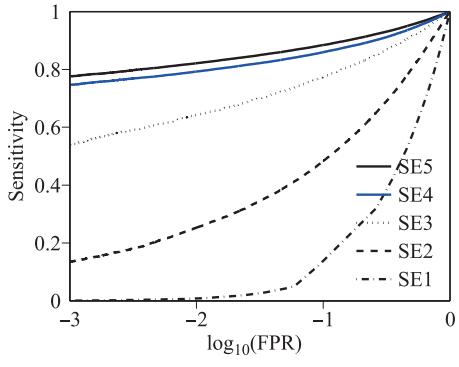

Fig. 8 ROC curve: A comprehensive evaluation on the performance of the proposed improvement schemes. This evaluation was conducted by gradually changing the detection scheme configuration from that of Gabbard *et al.* [21] to this work's proposal (Table 2 and Section 4.1), the detailed configurations of this experiment are presented in Table 4. In this experiment, the SE1 was learned from Training set[2] (Table 1), evaluated on Test set$_1^2$ (Table 1), and SEi ($i = 2, 3, 4, 5$) were learned and tested on Training set[1] and Test set$_1^1$ respectively (Table 1).

Table 3 The sensitivities in case of FPR= 0.001, 0.01, 0.1 and $\rho_{opt} = 2$. Considering the lowest SNR ρ_{opt} of the test data in two closely related works [22, 25] is 2, Table 3 presents the evaluation results on the data with $\rho_{opt} = 2$ from Test set$_1^2$ and Test set$_1^1$. That is to say, the test data in experiments of this table is a subset of those of Fig. 8. Exp. No.: Experimental identification. Some detailed configurations of this experiments are presented in Table 4.

Exp. No.	FPR=0.001	FPR=0.01	FPR=0.1
SE1	0.001	0.005	0.300
SE2	0.067	0.365	0.555
SE3	0.660	0.748	0.835
SE4	0.853	0.883	0.918
SE5	0.947	0.967	0.984

5 Conclusions and discussion

This work investigated the optimization techniques for the CNN-based GW detection scheme and proposed a GW detection scheme based on CNN and wavelet packet decomposition method. The motivation is to fill in the lack of the research on the dependencies of CNN-based GW detection scheme on vital factors such as frequency analysis, configurations of detection window, and data resolution [21, 22, 25–27]. Experimental results show that the proposed optimization of this work improved the GW detection sensitivity evidently. Therefore, the accumulation effects from such factors couldnot be ignored in GW detection.

This proposed scheme consists of three essential procedures: first, extract a segment from a data stream; second, decompose this segment using WP; third, detect the existence of any GW signal using a CNN. It is shown that the WP spotlights the characteristics of GW in a time-frequency space, and can improve the detection performance. In reality, the observed data comes as a data stream. Therefore, this work proposes an overlapping window scheme, by which the GW detection problem is reformulated as a classification paradigm in machine learning.

Actually, this is a very flexible scheme as a reference for related scientists in this field. For example, if preparing some training samples for one or more classes of glitches and replacing the logistic regression with a softmax regression, the proposed scheme can be explored in glitch detection; if replacing the logistic regression with a linear regression, we can study the application of the proposed scheme in GW parameter estimation in theory, for example, parameters f_0 and τ in Eq. (6).

However, the principle of the CNN indicates that there exists more error in estimating parameter t_0 in GW model of Eq. (6), the occurrence time of a GW in a data steam, than any other parameter. This error comes from the time translation invariance effect of the pooling operation. The time translation invariance effect means that

Table 4 Configurations of the comprehensive evaluation experiments on the proposed optimization schemes. WPD: wavelet packet decomposition, DS: data structure, CK: convolutional kernel, ReLU: Rectified Linear Unit [59], ELU: exponential linear unit, configuration of a convolutional layer $(i_1, i_2, \cdots, i_k)$: i_1 is convolutional kernels (CKs) and $i_2 \times \cdots \times i_k$ ($k = 3$ or 4) is size of every CK, configuration of a pooling layer (j_1, j_2): size of the pooling window is $j_1 \times j_2$, configuration of a fully connected layer i: i is the number of hidden units with the ReLU as activation function. In experiment SE1, there isn't any overlap between two successive pooling window as proposed in Gabbard *et al* [21], but there exist some overlaps in experiments SEi ($i \geq 2$).

Procedures	SE1	SE2	SE3	SE4	SE5
Optimization schemes					
Data resolution (Hz)	8192	8192	1.37×10^5	1.37×10^5	1.37×10^5
Window width	1s	0.03s	0.03s	0.03s	0.03s
WP decomposition	No	No	No	Yes	Yes
Window overlap	No	No	No	No	0.004s
Configurations of CNN (based on the order from input layer to output layer)					
Convolutional layer	(8, 1, 64)	(15, 1, 3)	(15, 1, 3)	(15, 1, 3)	(15, 1, 3)
Pooling layer	No	**(1, 16)**	**(1, 16)**	(4, 4)	(4, 4)
Activation layer	ELU	ReLU	ReLU	ReLU	ReLU
Convolutional layer	(8, 1, 32, 8)	(20, 1, 2, 15)	(20, 1, 2, 15)	(20, 1, 2, 15)	(20, 1, 2, 15)
Pooling layer	(1, 8)	**(1, 16)**	**(1, 16)**	(4, 4)	(4, 4)
Activation layer	ELU	ReLU	ReLU	ReLU	ReLU
Convolutional layer	(16, 1, 32, 8)	No	No	No	No
Activation layer	ELU	No	No	No	No
Convolutional layer	(16, 1, 16, 16)	No	No	No	No
Pooling layer	(1, 6)	No	No	No	No
Activation layer	ELU	No	No	No	No
Convolutional layer	(32, 1, 16, 16)	No	No	No	No
Activation layer	ELU	No	No	No	No
Convolutional layer	(32, 1, 16, 32)	No	No	No	No
Pooling layer	(1, 4)	No	No	No	No
Activation layer	ELU	No	No	No	No
Flatten layer	Yes	Yes	Yes	Yes	Yes
Full connected layer	64	10	10	10	10
Activation	ELU	ReLU	ReLU	ReLU	ReLU
Full connected layer	64	No	No	No	No
Activation layer	ELU	ReLU	ReLU	ReLU	ReLU
Logistic regression layer	Yes	Yes	Yes	Yes	Yes

a small change of gravitational wave occurrence time in the detection window will not affect the calculation results of CNN. Therefore, the significance of this error depends on the width of the detection window. To alleviate the problem, we can firstly estimate parameter t_0 using a CNN-based scheme with an optimized window width and obtain an initial estimation, then reestimate it using a CNN method with a shorter detection window around the initial estimation.

When this paper was under review, we became aware of the paper [60] which conducted an interesting investigation by introducing a novel observation data decomposition method based on theoretical waveforms. In principle, the theoretical waveforms in Wang *et al.* [60] play a similar role with the wavelet basis in the proposed scheme of this work. The theoretical waveforms have attractive physical background, the wavelet packet decomposition has concrete theoretical foundations and widely applications. However, it is an interesting topic to study the existence of their different characteristics on the detection completeness of GW events, false alarm rate.

Acknowledgements Authors thank H. Gabbard, M. Williams, F. Hayes, and C. Messenger very much for their enthusiastic discussions and sharing of their experimental code. We are grateful for valuable suggestions and corrections from anonymous reviewers, Eric D. Feigelson, Dr. Jin Li and B. S. Sathyaprakash. X. L. and W. Y. were supported by the National Natural Science Foundation of China (Grant Nos. 11973022 and U1811464), the Natural Science Foundation of Guangdong Province (No. 2020A1515010710), and China Scholarship Council (No. 201706755006), and the Joint Research Fund in Astronomy (No. U1531242) under cooperative agreement between the National Natural Science Foundation of China (NSFC) and Chinese Academy of Sciences (CAS). Xilong Fan was supported by the National Natural Science Foundation of China (Grant Nos. 11673008 and 11922303) and Hubei Province Natural Science Fund for the Distinguished Young Scholars.

References and notes

1. B. P. Abbott, R. Abbott, T. D. Abbott, et al., Observation of gravitational waves from a binary black hole merger, *Phys. Rev. Lett.* 116(6), 061102 (2016)
2. B. P. Abbott, R. Abbott, and T. D. Abbott, Binary black hole mergers in the first advanced LIGO observing run, *Phys. Rev. X* 6(4), 041015 (2016)
3. B. P. Abbott, R. Abbott, T. D. Abbott, et al., GW170104: Observation of a 50-solar-mass binary black hole coalescence at redshift 0.2, *Phys. Rev. Lett.* 118, 221101 (2017)
4. B. P. Abbott, R. Abbott, T. D. Abbott, et al., GW170817: Observation of gravitational waves from a binary neutron star inspiral, *Phys. Rev. Lett.* 119(16), 161101 (2017)
5. B. P. Abbott, R. Abbott, T. D. Abbott, et al., GW170814: A three-detector observation of gravitational waves from a binary black hole coalescence, *Phys. Rev. Lett.* 119(14), 141101 (2017)
6. B. P. Abbott, R. Abbott, and R. X. Adhikari, et al., Multi-messenger observations of a binary neutron star merger, *Astrophys. J. Lett.* 848(2), L12 (2017)
7. B. P. Abbott, R. Abbott, T. D. Abbott, et al., Gravitational waves and gamma-rays from a binary neutron star merger: GW170817 and GRB 170817A, *Astrophys. J. Lett.* 848(2), L13 (2017)
8. B. P. Abbott, et al., A gravitational-wave standard siren measurement of the Hubble constant, *Nature* 551(7678), 85 (2017)
9. S. Adrián-Martínez, M. G. Aartsen, B. Abbott, et al., High-energy neutrino follow-up search of gravitational wave event GW150914 with ANTARES and IceCube, *Phys. Rev. D* 93, 122010
10. B. Abbott, R. Abbott, T. D. Abbott, et al., All-sky search for short gravitational-wave bursts in the first advanced LIGO run, *Phys. Rev. D* 95, 042003 (2017)
11. B. P. Abbott, G. Cagnoli, J. Degallaix, et al., Observing gravitational-wave transient GW150914 with minimal assumptions, *Phys. Rev. D* 93, 122004 (2016)
12. C. Vishveshwara, Scattering of gravitational radiation by a Schwarzschild black-hole, *Nature* 227, 936 (1970)
13. O. Benhar, V. Ferrari, and L. Gualtieri, Gravitational wave asteroseismology revisited, *Phys. Rev. D* 70, 124015 (2004)
14. J. Powell, D. Trifirò, E. Cuoco, et al., Classification methods for noise transients in advanced gravitational-wave detectors, *Class. Quantum Grav.* 32, 215012 (2015)
15. M. Zevin, S. Coughlin, et al., Gravity spy: Integrating advanced LIGO detector characterization, machine learning, and citizen science, *Class. Quantum Grav.* 34, 064003 (2017)
16. J. Powell, A. Torres-Forné, et al., Classification methods for noise transients in advanced gravitational-wave detectors II: Performance tests on advanced LIGO data, *Class. Quantum Grav.* 34, 034002 (2017)
17. B. Allen, W. G. Anderson, P. R. Brady, D. A. Brown, and J. D. E. Creighton, FINDCHIRP: An algorithm for detection of gravitational waves from inspiraling compact binaries, *Phys. Rev. D* 85(12), 122006 (2012)
18. S. Babak, R. Biswas, et al., Searching for gravitational waves from binary coalescence, *Phys. Rev. D* 87, 024033 (2013)
19. K. Cannon, R. Cariou, A. Chapman, et al., Toward early-warning detection of gravitational waves from compact binary coalescence, *Astrophys. J.* 748(2), 136 (2012)
20. S. A. Usman, A. H. Nitz, I. W. Harry, et al., The PyCBC search for gravitational waves from compact binary coalescence, *Class. Quantum Grav.* 33(21), 215004 (2016)
21. H. Gabbard, M. Williams, F. Hayes, and C. Messenger, Matching matched filtering with deep networks for gravitational-wave astronomy, *Phys. Rev. Lett.* 120(14), 141103 (2018)
22. 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)
23. B. J. Lin, X. R. Li, and W. L. Yu, Binary neutron stars gravitational wave detection based on wavelet packet analysis and convolutional neural networks, *Front. Phys.* 15(2), 24602 (2020)
24. H. M. Luo, W. B. Lin, Z. C. Chen, and Q. G. Huang, Extraction of gravitational wave signals with optimized convolutional neural network, *Front. Phys.* 15(1), 14601 (2020)
25. D. George and E. A. Huerta, Deep neural networks to enable real-time multimessenger astrophysics, *Phys. Rev. D* 97, 044039 (2018)
26. T. D. Gebhard, N. Kilbertus, G. Parascandolo, I. Harry, and B. Schlkopf, CONVWAVE: Searching for gravitational waves with fully convolutional Neural Nets, in: Workshop on Deep Learning for Physical Sciences (DLPS) at the 31st Conference on Neural Information Processing Systems (NIPS), 2017
27. T. D. Gebhard, N. Kilbertus, I. Harry, and B. Schlkopf, Convolutional neural networks: A magic bullet for gravitational-wave detection? *Phys. Rev. D* 100(6), 063015 (2019)
28. S. Chatterji, L. Blackburn, G. Martin, and E. Katsavounidis, Multiresolution techniques for the detection of gravitational-wave bursts, *Class. Quantum Grav.* 21(20), S1809 (2004)
29. P. J. Sutton, G. Jones, S. Chatterji, et al., X-Pipeline: An analysis package for autonomous gravitational-wave burst searches, *New J. Phys.* 12(5), 053034 (2010)
30. S. Bahaadini, N. Rohani, S. Coughlin, M. Zevin, V. Kalogera, and A. K. Katsaggelos, Deep multi-view models for glitch classification, *IEEE ICASSP*, 2931–2935 (2017)
31. S. Bahaadini, V. Noroozi, N. Rohani, S. Coughlin, M. Zevein, J. R. Smith, V. Kalogera, and A. Katsaggelos, Machine learning for Gravity Spy: Glitch classification and dataset, *Information Sciences* 444, pp 172–186 (2018)

32. 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)
33. N. Mukund, S. Abraham, S. Kandhasamy, and N. S. Philip, Transient classification in LIGO data using difference boosting neural network, *Phys. Rev. D* 95, 104059 (2017)
34. J. C. Brown, Calculation of a constant Q-spectral transform, *J. Acoust. Soc. Am.* 89(1), 425 (1991)
35. S. Klimenko, I. Yakushin, A. Mercer, and G. Mitselmakher, Coherent method for detection of gravitational wave bursts, *Class. Quantum Grav.* 25, 114029 (2008)
36. S. Klimenko, G. Vedovato, M. Drago, F. Salemi, V. Tiwari, G. A. Prodi, C. Lazzaro, S. Tiwari, F. Da Silva, and G. Mitselmakher, Method for detection and reconstruction of gravitational wave transients with networks of advanced detectors, *Phys. Rev. D* 93, 042004 (2016)
37. R. S. Lynch, S. Vitale, R. C. Essick, E. Katsavounidis, and F. Robinet, An information-theoretic approach to the gravitational-wave burst detection problem, *Phys. Rev. D* 95, 104046 (2017)
38. N. J. Cornish and T. B. Littenberg, BayesWave: Bayesian Inference for Gravitational Wave Bursts and Instrument Glitches, *Class. Quantum Grav.* 32, 135012 (2015)
39. T. B. Littenberg and N. J. Cornish, Bayesian inference for spectral estimation of gravitational wave detector noise, *Phys. Rev. D* 91, 084034 (2015)
40. S. Chatterji, A. Lazzarini, L. Stein, P. Sutton, A. Searle, and M. Tinto, Coherent network analysis technique for discriminating gravitational-wave bursts from instrumental noise, *Phys. Rev. D* 74, 082005 (2006)
41. S. Bose, S. Dhurandhar, et al., Towards mitigating the effect of sine-Gaussian noise transients on searches for gravitational waves from compact binary coalescences, *Phys. Rev. D* 94, 122004 (2016)
42. B. J. Owen and B. S. Sathyaprakash, Matched filtering of gravitational waves from inspiraling compact binaries: Computational cost and template placement, *Phys. Rev. D* 60(2), 022002 (1999)
43. pwelch: Welch's power spectral density estimate.
44. G. D. Meadors, K. Kawabe, and K. Riles, Increasing LIGO sensitivity by feedforward subtraction of auxiliary length control noise, *Class. Quantum Grav.* 31, 105014 (2014)
45. P. D. Welch, The use of Fast Fourier Transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms, *IEEE Transactions on Audio and Electroacoustics* 15(2), 70 (1967)
46. J. Abadie, B. P. Abbott, R. Abbott, et al., All-sky search for gravitational-wave bursts in the second joint LIGO-Virgo run, *Phys. Rev. D* 85, 122007 (2012)
47. S. Mallat, A Wavelet Tour of Signal Processing, Boston: Academic Press, 2009
48. K. B. Howell, Principles of Fourier analysis, CRC Press, 2016
49. I. Daubechies, Ten Lectures on Wavelets, Philadelphia: Society for Industrial and Applied Mathematics, 1992
50. S. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Trans. on Pattern Analysis and Machine Intel.* 11(7), 674 (1989)
51. S. Rampone, V. Pierro, L. Troiano, et al., Neural network aided glitch-burst discrimination and glitch classification, *Inter. J. Mod. Phys.* 24(11), 1350084 (2013)
52. S. Vinciguerra, M. Drago, G. A. Prodi, et al., Enhancing the significance of gravitational wave bursts through signal classification, *Class. Quantum Grav.* 34, 094003 (2017)
53. MATLAB and Wavelet Toolbox Release 2013b, The MathWorks, Inc., Natick, Massachusetts, United States
54. X. R. Li, Y. Lu, G. Comte, AL. Luo, Y. H. Zhao, and Y. J. Wang, Linearly Supporting feature extraction for automated estimation of stellar atmospheric parameters, *Astrophys. J. Suppl. S.* 218(1), 3(2015)
55. Y. LeCun, B. E. Boser, J. S. Denker, et al., Handwritten digit recognition with a back-propagation network, in *Advances in Neural Information Processing Systems*, 396 (1990)
56. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, *Proceedings of the IEEE* 86, pp 2278–2324 (1998)
57. Y. LeCun, Y. Bengio, and G. E. Hinton, Deep learning, *Nature* 521(7553), 436(2015)
58. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Learning representations by back-propagating errors, *Nature* 323(6088), 533(1986)
59. I. Goodfellow, Y. Bengio, and A. Courville, Deep learning, Cambridge: MIT Press, 2016
60. H. Wang, Z. J. Cao, X. L. Liu, S. C. Wu, and J. Y. Zhu, Gravitational wave signal recognition of O1 data by deep learning, arXiv: 1909.13442 (2019)