Searching for gravitational waves from stellar-mass binary black holes early inspiral

Xue-Ting Zhang,[*] Natalia Korsakova,[†] Man Leong Chan,[‡] Chris Messenger,[§] and Yi-Ming Hu[¶]

(Dated: June 12, 2024)

The early inspiral from stellar-mass binary black holes (sBBHs) can emit milli-Hertz gravitational wave signals, making them detectable sources for space-borne gravitational wave missions like Tian-Qin. However, the traditional matched filtering technique poses a significant challenge for analyzing this kind of signals, as it requires an impractically high number of templates ranging from 10^{31} to 10^{40}. We propose a search strategy that involves two main parts: initially, we reduce the dimensionality of the simulated signals using incremental principal component analysis (IPCA). Subsequently we train the convolutional neural networks (CNNs) based on the compressed TianQin data obtained from IPCA, aiming to develop both a detection model and a point parameter estimation model. The compression efficiency for the trained IPCA model achieves a cumulative variance ratio of 95.6% when applied to 10^6 simulated signals. To evaluate the performance of CNN we generate the receiver operating characteristic curve for the detection model which is applied to the test data with varying signal-to-noise ratios. At a false alarm probability of 5% the corresponding true alarm probability for signals with a signal-to-noise ratio of 50 is 86.5%. Subsequently, we introduce the point estimation model to evaluate the value of the chirp mass of corresponding sBBH signals with an error. For signals with a signal-to-noise ratio of 50, the trained point estimation CNN model can estimate the chirp mass of most test events, with a standard deviation error of 2.48 $M_\odot$ and a relative error precision of 0.12.

I. INTRODUCTION

Stellar-mass binary black hole is a typical kind of multi-band observation sources of gravitational wave (GW), with space-borne detectors observing the early inspiral phase of the signal, and ground-based detectors observing the late inspiral, merger, and ringdown phases. Ground-based detectors have already detected these signals, opening up the era of gravitational wave astronomy [1–3]. Future space-based detectors (e.g. TianQin [4], LISA [5]) are likely to detect this type of source.

Due to the long duration ranging from months to years expected for sBBH signals observed by space-borne detectors, they offer the potential for high-precision estimation of physical parameters, including characteristics such as the spins of the binary and its localization[6, 7]. The information regarding the spins is particularly valuable for unraveling the evolutionary history of binary systems, given the notable variations in the effec-

tive spin predicted by various formation mechanisms[8–10].Additionally, the localization of these signals can contribute significantly to constraining the Hubble constant [11–15]

Indeed, the detection of GWs signal from long-lived sBBH poses a considerable challenge when employing matched filtering, especially given the need for over 10^{31} templates[16] . Additionally, Owen[17] et al. proposed an archival search algorithm that utilizes information from ground-based detectors to narrow the parameter space for the signal being searched in space-based detector data, allowing for more efficient data analysis. However, this method relies on the gravitational wave signal being observed by both ground-based and space-borne detectors [14]. Furthermore semi-coherent method has been explored in zero-noise scenarios [18] with the help of particle-swarm optimization, which has not provided sensitivity analysis yet. If a sBBH merger, occurring over approximately one year, can be detected by a space-borne detector, it holds the potential to offer a pre-merger warning for ground-based detectors. This necessitates the development of a swift and cost-effective pipeline for the search of sBBH signals in the mHz frequency band. Such a pipeline would serve as a crucial link between space-borne detectors and ground-based detectors.

In recent years, machine learning and/or deep learning algorithms have demonstrated significant potential for application in gravitational wave data analysis.[19–35]. Machine learning algorithms excel in detecting nonlinear structures within complex and long-duration signals, including events like binary neutron star (BNS) mergers (with durations ranging from minutes to hours)[36–42], continuous waves (spanning from hours to years)[43], and extreme mass-ratio inspirals (EMRI) (spanning from a few months to years)[34]. As detector sensitivity im-

[*] TianQin Research Center for Gravitational Physics, Sun Yat-sen University, 2 Daxue Rd., Zhuhai 519082, China; Also at School of Physics and Astronomy, Sun Yat-sen University, 2 Daxue Rd., Zhuhai 519082, China; SUPA, School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, United Kingdom; Max Planck Institute for Gravitational Physics (Albert Einstein Institute), D-14476 Potsdam, Germany; zhangxt57@mail2.sysu.edu.cn

[†] Astroparticule et Cosmologie, CNRS, Université Paris Cité, F-75013 Paris, France

[‡] Department of Physics and Astronomy, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

[§] SUPA, School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, United Kingdom

[¶] TianQin Research Center for Gravitational Physics, Sun Yat-sen University, 2 Daxue Rd., Zhuhai 519082, China; Also at School of Physics and Astronomy, Sun Yat-sen University, 2 Daxue Rd., Zhuhai 519082, China; Email:huyiming@mail.sysu.edu.cn

proves in the future, the increasing number of events to be analyzed and challenges such as signal overlap impose distinct requirements on data analysis pipelines, such as high computational cost. In these scenarios, machine learning algorithms demonstrate significant advantages in terms of processing efficiency and generalisation.

We propose to develop a search algorithm using convolutional neural network (CNN) for the search of GWs from sBBHs by TianQin. Moreover, once CNN and other neural networks complete training, their computational efficiency is very high. If we can quickly analyze the properties of a strong sBBH from a space-borne detector, it can provide early warning to other detectors. Of note, there are researches [7, 13] in which experts estimate the source parameters of an sBBH from the actual point, presuming the data will be ready from the search phase. However, without any prior knowledge from the source, the task becomes markedly difficult. Our assured pipeline can offer useful input for Bayesian inference.

We initiated the training of CNN using time-series data, data on the frequency domain, and data on the time-frequency spectrum. However, the sheer volume of data points within a single signal from a sBBH—spanning a duration of a few months—posed a challenge for CNN training in terms of GPU memory and computational time. Consequently, we employed incremental principal component analysis (IPCA) to reduce the number of data points, facilitating the construction of a search pipeline tailored for the detection of sBBH signals.

We have tried three different data representations for our signal: time domain, frequency domain, and time-frequency spectrum. The experiments that we have performed demonstrated that the frequency domain data was the most suitable for our purposes. We decided to work with the data compressed by IPCA model. For this reason we tested different representations of frequency domain data and we have found out that for the data processed with TianQin response it works best with the amplitude part of the data due to specific modulations in this response. Subsequent to this, we employed the trained IPCA model to compress the amplitude of detector data in the frequency domain. Upon projecting from the IPCA model, we trained a detection CNN model as well as a point parameter estimation CNN model. This serves as a demonstration of the compression- plus-search pipeline methodology for the long-lasting GW search of sBBHs.

The structure of this paper is as follows: In Section II, we introduce sBBH early inspiral GW and simulations of observations of GWs from sBBHs by TianQin. In section III, we explain the main techniques in data analysis that we used, including the principle component analysis and convolution neural network. In Section IV, we introduce searching strategies for GW from sBBH, including detection stage and point parameter estimation stage. In Section V, we present the results. In Section VI we summarize what we have achieved and discuss future work.

II. GW FROM STELLAR-MASS BINARY BINARY BLACK HOLE INSPIRAL

A. Stellar-mass binary black early inspiral

The evolution of an sBBH system comprises three distinct stages: the inspiral, merger, and ringdown phases. During each phase, the emitted gravitational waves can be characterized by various theories, including the Post-Newtonian (PN) approximation, which governs the behavior of gravitational waves at different levels of accuracy. Specifically, space borne detectors expect to capture gravitational waves radiated in the early inspiral stage of sBBH evolution, for which the 2.5PN (post-Newtonian) approximation can be accurate enough [44].

Despite that various formation channels, such as isolated binary evolution, dynamic formation, and primordial black holes, have attempted to explain the origin of sBBHs, no conclusion has been reached so far[45]. These models have made different predictions about the mass ratio, spin, and orbital parameters of the binary black holes. The validation of these predictions can be inferred via gravitational wave observations. While ground-based gravitational wave observations have completed three effective rounds of detection, more observational data are still needed to determine the model that best fits reality [46]. To mitigate the introduction of any preferred systematic bias in physical parameters, we incorporate a uniform distribution of sBBH populations throughout the Universe in our simulation.

For more streamlined data analysis, we can employ `IMRPhenomD` [47] to swiftly generate gravitational wave polarizations, taking into account only the dominant quadrupole moment. `IMRPhenomD` includes whole evolution process of sBBH (considering 3.5 PN approximation) and can easily be used to build up our pipeline [12]. In the context of the inspiral phase of the sBBH signal, this implies that the frequency of the system satisfies $f < f_{\text{LSO}}$, where f_{LSO} represents the frequency of the last stable orbit. Moreover, we use h_{22} by `IMRPhenomD` to characterize the original signal, expressed as:

$$\tilde{h}_{22}(f) = A(f)\exp\left(-i\Psi(f)\right), \qquad (1)$$

Here, f denotes the frequency evolution range, The representation of the GW amplitude is $A(f) = \frac{\mathcal{M}^{5/6}}{\pi^{2/3}D_L}f^{-7/6}$, where $\mathcal{M} = \eta^{3/5}M$ denotes the chirp mass. The symmetric mass ratio here is $\eta = \frac{m_1 m_2}{(m_1+m_2)^2}$ and the total mass is $M = m_1 + m_2$. Additionally, $\Psi(f)$ signifies the GW phase, as elaborated on in a specific set of physical parameters. These parameters include the masses (m_1, m_2) of the binary system and the spins (χ_1, χ_2) of each component within the binary, as detailed in the paper [47].

B. TianQin response

TianQin, a space-borne detector with sensitivity in the frequency range of 10^{-4} to 1 Hz, enabling the observation of gravitational waves originating from early inspirals of sBBHs. Through time delay interferometry (TDI) technology[48–50], we can integrate the collected data into three quasi-independent AET channels, thus effectively suppressing laser phase noise. It is common to choose these representation for TDI because the noise correlation matrix of these three combinations is diagonal. The noise budget can be described by the following power spectral density (PSD)[51]:

$$S_{A,E}(f) = \frac{2}{L^2} \sin^2 f_c \left[(\cos f_c + 2) S_p(f) \right.$$
$$\left. + 2(\cos(2f_c) + 2\cos f_c + 3) \frac{S_a(f)}{(2\pi f)^4} \right], \quad (2)$$

$$S_T(f) = 32 \sin^2 f_c \sin^2 \frac{f_c}{2} (4\sin^2 f_c S_a + S_p), \quad (3)$$

$$f_c = \frac{2\pi f L}{c}, \quad (4)$$

$$S_p = \frac{f_c^2}{L^2} \left(\frac{10^{-4}\text{Hz}}{f} \right), \quad (5)$$

$$S_a = \left(1 + \frac{10^{-4}}{f} \right) \frac{L^2}{f_c^2 c^4}. \quad (6)$$

Where $L = \sqrt{3} \times 10^8$ m is the arm length

When the gravitational wave signal from a sBBH system passes through TianQin the received signal will be modulated,which can be expressed by the transfer function related to each laser links among three satellites [52]:

$$\tilde{h}(f) = \sum_l \sum_m \mathcal{T}^{A,E,T}(f, t_{lm}(f)) \tilde{h}_{lm}(f). \quad (7)$$

The transfer function, denoted as $(\mathcal{T}^{A,E,T}(f, t_{lm}(f)))$, is dependent on a set of parameters, including $(t_{\text{ref}}, \phi_{\text{ref}}, \iota, \psi, \lambda, \beta)$. The reference time and phase are denoted by $(t_{\text{ref}}, \phi_{\text{ref}})$; the inclination and polarization angle are indicated by (ι) and (ψ) respectively. Furthermore, the ecliptic longitude and latitude in the solar-system barycenter (SSB) are represented by (λ, β).

For each harmonic of the waveform, assuming it satisfies the stationary phase approximation and the shifted uniform asymptotic, its time-frequency relationship [52] can be expressed as

$$t_{lm}(f) = t_{\text{ref}} - \frac{1}{2\pi} \frac{d\Phi_{lm}(f)}{df}. \quad (8)$$

The aforementioned extrinsic parameters are defined in the solar system barycentric coordinate system. By combining each harmonic $h_{lm}(f)$ with its corresponding transfer function $\mathcal{T}(f, t_{lm}(f))$, we can simulate the received signal samples of each TDI channel. We use
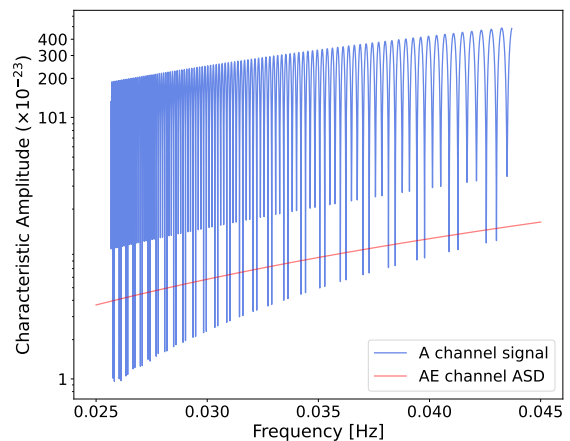
(a)
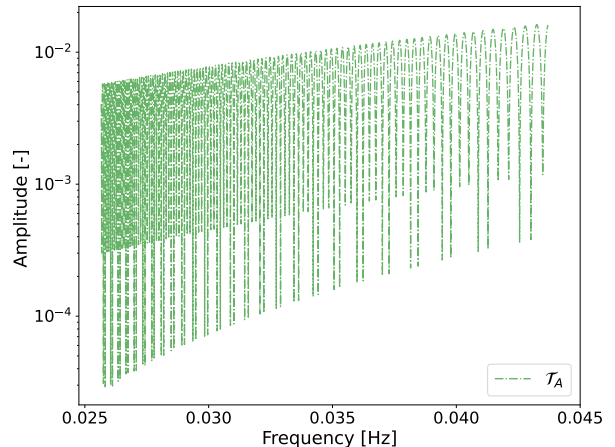
(b)

FIG. 1. In Subfigure (a), the 'A' channel signal from a GW190521-like binary black hole system is depicted as a blue curve, while the amplitude spectral density (ASD) of the 'AE' channel is shown in red. The data reflects an observation period of three months at a sample rate of 0.25 Hz, with each channel containing 972001 data points. The binary system under observation comprises two black holes of approximately 85 $M_\odot$ and 66 $M_\odot$. This GW signal's SNR is 50, and its sky location is given by $(\lambda = 5.6$ Rad, $\beta = -4.7$ Rad). Subfigure (b) presents the amplitude of the respective dimensionless transfer function in the 'A' channel. The oscillation magnitude of the resultant signal is derived from this transfer function, which is governed by Equation Eq. (9).

IMRPhenomD to describe the original waveform (only considering the 22 mode, i.e., $l = 2, m = 2$), which is expressed as:

$$\tilde{h}_{A,E,T} = \mathcal{T}^{A,E,T}(f, t_{22}(f)) \tilde{h}_{22}(f). \quad (9)$$

Using a GW190521-like sBBH system as an example, as shown in Figure 1(a), this signal exhibits multiple cycles within the sensitive band of TianQin. The prominent oscillations are a result of detector modulations, primarily influenced by the transfer function.

At the same time, although we know that the detector data is composed of three channels (A, E, T), as the signal is effectively suppressed in the T channel, we only use the data from the A and E channels in subsequent analyses.

III. METHODS

A. Principle component analysis

Taking into account the duty cycle of TianQin (operating for three months, followed by another three months off), the dimensionality of a single source in the time domain for low-frequency sBBH signals is (3, 1944000). Here, "3" refers to the three channels of data from the TDI AEI channel, while "1944000" accounts for the number of data points obtained over a 3-month duration at a sampling rate of 0.25 Hz. The use of a CNN for processing such lengthy signals comes with its own challenges, primarily due to the large size of the signals. This poses difficulties in terms of GPU memory and training time. To tackle these issues, attempts have been made to reduce the signal size by truncating signal in each channel at the frequency domain, thus lowering the number of data points.

In this context, we employ principal component analysis (PCA) to compress the data and extract initial features. Using matrix decomposition, high-dimensional data M_{mn} can be represented as a set of lower-dimensional bases U_{mk} and the values of each basis V_{kn} [53], where m represents the number of data sample, n is the initial data dimension, and k is the data dimension obtained through matrix transformation, with $n > k$

$$M_{mn} \approx U_{mk}V_{kn} \tag{10}$$

As k approaches n, the product of the transformation matrix and projection obtained by PCA is increasingly close to the matrix formed by the original data. As the new, lower-dimensional basis ($u^i \in U_{mk}$) is chosen by selecting the eigenvalues of the covariance matrix of the data in descending order, based on their associated variance, it showcases the importance of these eigenvalues in the data representation. In order to minimize variance loss, a larger number of components may be required. This, in turn, results in building a larger lower-dimensionality matrix U_{mk}. Preserving more information from the original data will reduce the variance loss, thus enabling a more accurate representation.

This method establishes the conversion matrix and realizes the dimension reduction of high-dimensional data by finding the maximum variance direction of the data to determine the first projection base, requiring the second largest variance direction to be orthogonal to the previous one, and so on, to obtain the decomposition and dimension reduction direction of the data.

Limited by the GPU memory, we cannot read a large amount of simulated signals at once. Therefore, we use IPCA [54] as an alternative to the actual PCA computation process. This IPCA method is also based on the concept of low-rank approximation, seeking a projection space similar to PCA, but only needing to read a batch of data at a time, instead of loading all the data into memory at once. We extracted the IPCA model based on the third-party software `cuML` [55].

B. Convolutional neural network for searching GW signals

CNNs are a type of deep learning algorithms that use convolutional kernels to capture and learn data features [56]. Convolutional kernels are capable of learning to capture features at different abstraction levels. Consequently, the combination of multiple kernels allows the extraction of higher-level features, aiding the network in learning more complex patterns. A typical CNN consists of convolutional layers, pooling layers and/or or fully connected layers. Currently, CNNs are extensively employed in the field of Gravitational Wave astronomy. They are utilized for various purposes, such as detecting GW signals [19, 20, 22, 57] and classifying glitches [58, 59]. The convolution kernels of these networks play a crucial role in characterizing the distinctive features of the signals and glitches.

1. Signal detection

In the task of signal detection, detector data can be modeled as the data that either contains a signal or not. Using a CNN for signal detection means classifying the detector dataset, which is essentially calculating the probability that the data contains a signal. The specific mathematical model can be represented as:

$$y_{\text{pred}} = \mathcal{P}_{\text{CNN}}^{\text{Detection}}(x), \tag{11}$$
$$x = g(d_f).$$
$$d_f = \begin{cases} s_f(\theta) + n_f \\ n_f. \end{cases}$$

where x is the input to a CNN, which also can be a suitable representation g of the detector output data d. In our case, we take the compressed amplitude of the simulated data d_f in the frequency domain as the input to CNNs. y_{pred} is the probability whether this input contains a GW signal or not.

This detection task is equivalent to using a CNN for classification. In this scenario, the data is categorized into two classes with distinct labels. In our case the detector data containing signals is assigned the label 1, while the detector data with pure noise is labeled as 0.

We can train the CNN based on cross-entropy. The formula for binary cross-entropy is as follows:

$$\text{Loss}_{\text{bce}} = -\frac{1}{N} \sum_{i=1}^{i=N} \big[y_i \cdot \log(y_{\text{pred}}^i) +$$
$$(1 - y_i) \cdot \log^i (1 - y_{\text{pred}}^i) \big] \qquad (12)$$

where y_i represents the input label and y_{pred} is the corresponding output probability of the CNN.

2. Point estimation

Following our successful predictions regarding the presence of a signal in the data, we can proceed to evaluate the parameters of the GW signal which is present in the data. Consequently, we employ the similar CNN architecture that has proven effective in the detection task, repurposing its structure to estimate the physical parameters of the signal — a task that is commonly referred to as 'regression' in machine learning applications. This approach will yield point estimates of the physical parameters The specific mathematical model can be represented as:

$$\theta_{\text{pred}} = \mathcal{P}_{\text{CNN}}^{\text{Estimation}}(x), \qquad (13)$$
$$x = g(d_f(\theta_{\text{actual}})).$$
$$d_f(\theta) = s_f(\theta) + n_f.$$

In this formula, $d_f(\theta)$ is a sample of signal plus noise in the frequency domain, and x_i is the representation of the data sample. There is a certain function mapping relationship between the input and physical parameters. A well-trained CNN model can learn this mapping f_{CNN}.

We use the Mean Squared Error (MSE) to train the difference between the input chirp mass value θ_{actual} and the CNN output estimate θ_{pred}, thereby constructing a one-to-one mapping relationship between the predicted physical parameters and the corresponding input data. The formula for the mean squared error is as follows:

$$\text{Loss}_{\text{mse}} = \frac{1}{N} \sum_{i=1}^{N} (\theta_{\text{actual}}^i - \theta_{\text{pred}}^i) \qquad (14)$$

where N is the number of the used data, θ_{actual} is the actual physical parameter set of the corresponding data $d(\theta)$, and θ_{pred} is the point estimation of the physical parameter set.

3. CNN Architecture

The final CNN structure used in signal detection is shown in Table I. The final output layer of the CNN is augmented with the softmax function to normalize the output, enabling its interpretation as the probability of whether the data contains a GW signal. The other layers utilize the 'ReLU' activation function to achieve non-linear mapping from the input to the output. The 'ReLU' function yields zero for any negative input while returning positive values without any changes, thereby introducing sparsity into the model and potentially improving computational efficiency in the signal detection task.

TABLE I. The architecture of the CNN. The number of training parameters is 497466. The stride for each layer is 1.

	Layers	Neutrons number	Kernel size	Output size
1	Input	(2×480)	. . .	. . .
2	Convolution	64	(1×5)	(64×96)
3	BatchNorm	64	. . .	(64×96)
4	Convolution	128	(1×5)	(128×92)
5	Convolution	256	(1×5)	(256×88)
6	Convolution	64	(1×3)	(64×86)
7	Convolution	16	(1×3)	(16×84)
8	Convolution	8	(1×3)	(8×82)
9	Flatten	. . .		656
10	Dense	. . .		64
11	Dropout	$p = 0.5$	. . .	
12	Dense	. . .		32
13	Output	. . .		2

In the task of point estimation, we employ a CNN architecture similar to that used for signal detection. However, all 'ReLU' activation functions are replaced with the 'Tanh' function, and we remove any activation function from the output layer. This is done because for point estimation we noticed that 'ReLU' function did not yield sufficiently good results therefore we replaced with a smooth function. The 'Tanh' function maps real-valued numbers to a range between -1 and 1, essentially compressing the inputs into a narrower range, which can facilitate convergence during training in our case.

IV. IMPLEMENTATION

To obtain a well-trained machine learning model, the representation of data and labels are crucial. Initially, the data from each output channel is subjected to a whitening operation, where it is divided by its respective noise's PSD, as demonstrated in Equation Eq. (2) . Following this process, the data is then fed into the CNN model. However, facing the situation where sBBH has many individual data points, this directly challenges the size of the GPU memory and the training duration. Therefore, we have employed the IPCA method to compress the data, obtain a new representation of observation data, and then use the newly compressed features to train a CNN to differentiate whether the original observation data contains a sBBH signals. In cases where signals are included, we further train the similar CNN structure to estimate the chirp mass of the candidate signal.

A. Data simulation

We strive to explore technical possibilities within a baseline distribution of sBBH populations. Specifically, we consider simulating signals uniformly distributed in the co-moving volume, shown in Table II. We have utilized `IMRPhenomD` waveform to generate signals in the frequency domain [47], with parameters range set as shown in the Table II, and analyze them buried in Gaussian noise, assuming the TianQin TDI AET PSD Eq. (2) [51].

Subsequently, we conduct pre-processing on the simulated detector data, including whitening operations.
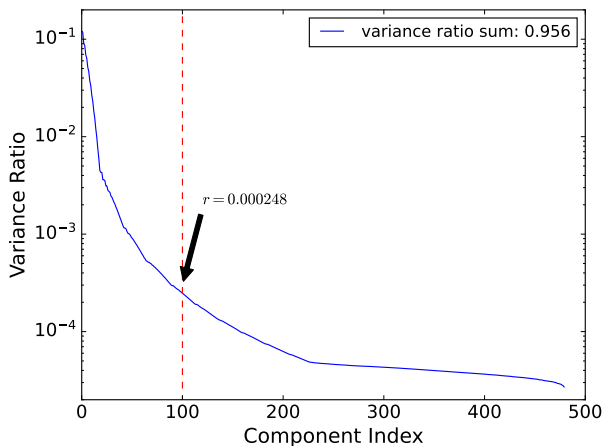
FIG. 2. The variance ratio of each component is arranged in descending order. The cumulative variance of k components across the entire dataset amounts to 95.6%. The 100th component has a variance ratio of 0.000248, which is less than 10^{-3}. The cumulative variance ratio of the preceding 100 components is 93.2%, indicating that the following components contribute less to the feature representation.

B. Compression

We adopt IPCA to extract the principal components from the signals, thereby reducing the dimensionality. We tested on different data representations, including time-domain signals, real and imaginary parts of frequency-domain signals, amplitude and phase of frequency-domain signals, etc. In our evaluations of various representations of responded signals, it has been observed that IPCA encounters challenges in extracting features from signals exhibiting substantial fluctuations. Consequently, we have chosen to exclusively derive the IPCA model from the amplitude of frequency-domain signals. This yields the corresponding IPCA$_i$, where 'i' notes the detector output channel, designated as either A or E. We used the IPCA method from the `cuML` [55] third-party package to load signals into GPU memory in batches, thereby extracting the principal components of

signals from a single channel after applying the detector response. These principal components represent the main characteristics of the signal amplitude. In the Figure 5, an example is depicted within the considered frequency band. This illustration provides a comparison between the original signal amplitude and the reconstructed signal amplitude, based on estimated IPCA components, together with their residual. We can see that boundaries at the signal's cutoff frequencies have higher residuals which means greater loss resulting from the compression process. This probably happens due to the sharp edges of the original signal.

Taking the extraction process of the amplitude IPCA model from the TDI A channel signal as an example, the size of the amplitude of a single signal in the A channel is (1,972001) obtained over a 3-month duration at a sampling rate of 0.25 Hz. By sampling within the astrophysical parameter range, we generate 4800000 simulated signals and a set of amplitudes, with the dimensionality is (4800000 x 1 x 972001). The variance ratio, denoted as r_i, signifies the variance σ_i of the i-th basis in relation to the total variance Σ of all signals, given by

$$r_i = \frac{\sigma_i}{\Sigma}. \tag{15}$$

These ratios are arranged in descending order, reflecting the efficacy of the extraction process. A lower index indicates a greater variance direction for the basis. As more basis components are considered, the cumulative variance ratio increases, thereby minimally affecting the compression loss.

In our case, fitting factor is a better measure to the performance of the component extraction. We obtain the projection value of a GW signal amplitude in a channel, based on the corresponding IPCA principle components or eigenvectors. By making use of the amplitude projection values (equivalent to the coefficients of the principal components), we can reconstruct its amplitude of the original signal and then measure the similarity between the reconstruction and the original signal amplitude by calculating the fitting factor. The formula for calculating the fitting factor is shown below.

$$\mathrm{FF} = \frac{\langle s(\theta)|s'(\theta)\rangle}{\sqrt{\langle s(\theta)|s(\theta)\rangle}\sqrt{\langle s'(\theta)|s'(\theta)\rangle}}. \tag{16}$$

where $s(\theta)$ is the original responded signal, $s'(\theta)$ is the reconstruction of IPCA model, $\langle\cdot\rangle$ is inner product between two signals

$$\langle s(\theta)|s'(\theta)\rangle = 4\sum_{f_{\min}}^{f_{\max}} \frac{|s(\theta)|}{\sqrt{S_n(f)}}\frac{|s'(\theta)|}{\sqrt{S_n(f)}}df. \tag{17}$$

where $f_{\min} = 0.02$ Hz and $f_{\max} = 0.1$ Hz.

As we analyze a multi-band GW event with a binary coalescence time of approximately 1 year and a duration of 3 months, the frequency range of most signals should

TABLE II. Physical parameters and their meanings: When considering sBBH as observed by space-borne detectors, it's worth noting that during the early inspiral stage, which spans several months, the time to coalescence is approximately set at one year, ranging from 9 months to 15 months.

| Notation	Meaning	Distribution	Units
m_1	primary mass	log uniform [5,100]	$M_\odot$
m_2	secondary mass	log uniform [5,m_1]	$M_\odot$
χ_{1z}	dimensionless aligned-spin magnitude of component 1	uniform [-1,1)	-
χ_{2z}	dimensionless aligned-spin magnitude of component 2	uniform [-1,1)	-
λ	the ecliptic longitude of the source	uniform $[0, 2\pi]$	Radian
β_S	the ecliptic latitude of the source	$\sin\beta = (\text{uniform}[-1,1])$	Radian
ι	inclination	$\cos\iota = (\text{uniform}[-1,1])$	Radian
ψ	polarization	uniform $[0, \pi]$	Radian
Φ_c	merger phase	uniform $[0, 2\pi]$	Radian
t_c	merger time at the reference time $t = 0$ s	uniform [23760000,39312000]	Seconds
D_L	luminosity distance to source	re-scaled inversely by the SNR, the reference distance is 50 Mpc.	Mpc
f_0	initial frequency at the reference time $t = 0$ s	calculated by Equation Eq. (18)	Hz
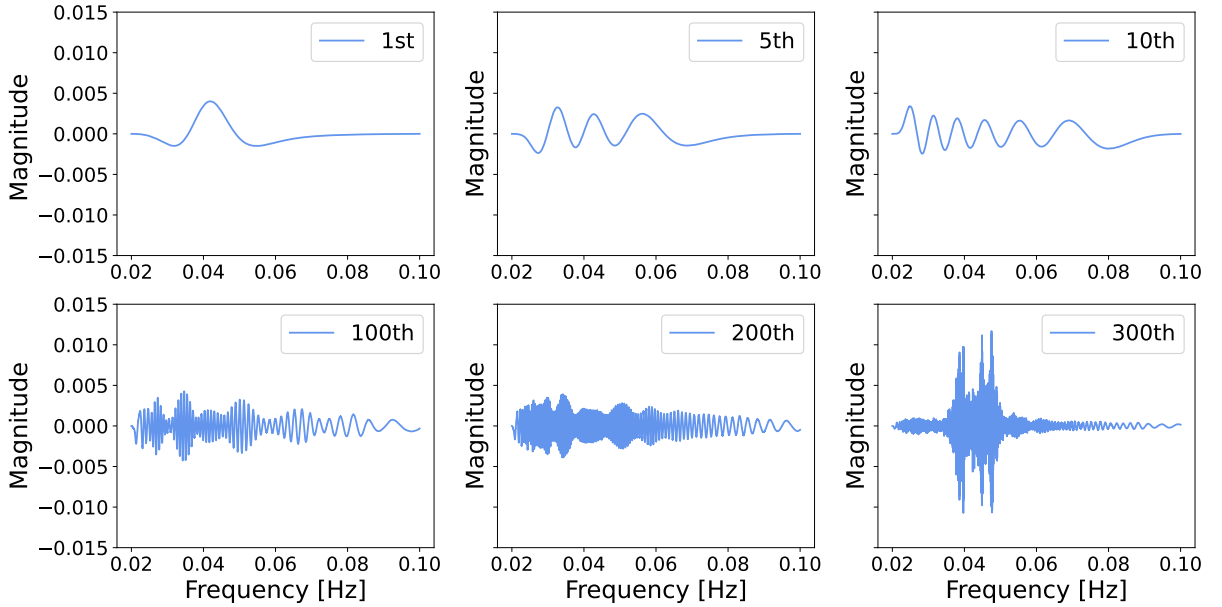
FIG. 3. Each eigenvector in the IPCA serves as a basis. The preceding few bases can effectively capture the primary signal evolution, which can be described by the Post-Newtonian approximation 1, depending on the chirp mass. Higher-order bases can capture multiple oscillation features, primarily originating from the response of the TianQin detector.

be within the band from 0.02 Hz to 0.1 Hz, as calculated by the instantaneous frequency formula at the time t before the merger time t_c :

$$f = \left(\frac{5}{256}\right)^{3/8} \mathcal{M}^{-5/8} (t_c - t)^{-3/8}. \quad (18)$$

Here, $\mathcal{M}$ represents the chirp mass of the binary system. Therefore, we truncated the amplitude of the frequency-domain signal, reducing the size of each responded channel's signal from 972001 to 622080.

Finally, we trained the IPCA model using 4800000 simulated signals, yielding a transformation matrix of (622080, 480). The cumulative variance ratio of these 480 components amounts to 0.958, indicating that data compressed through IPCA retains most of the variances from the amplitude of GW signals, as shown in Figure 2. Each component is a basis vector visually represented in Figure 3. From these 480 bases, it's apparent that all are focused on the high fluctuations in relatively lower frequency band, rather than on the high frequency. Using these IPCA model, the simulated data from a single detector (2 x 9720001) can be compressed into a matrix of size (2,480). This process can also be referred to as the data pre-processing module.

To demonstrate the compression loss, we examine the fitting factor across various astrophysical parameters. Lower fitting factors predominantly correlate with high oscillations in the signal amplitude, derived from Tian-
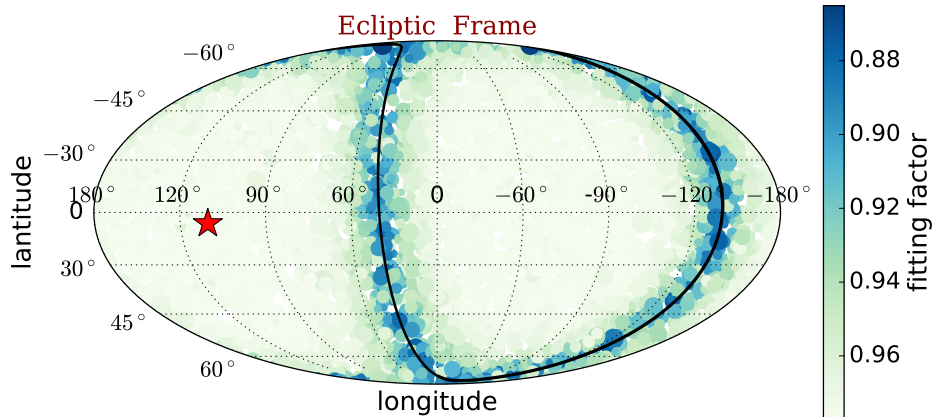
FIG. 4. A sky map showing the fitting factors for gravitational wave signals from sources distributed across the sky . In the map, the red star symbolizes the position of the DWD system J0806, which aligns with the direction of the TianQin constellation. The black solid curve delineates the orbit of TianQin. Darker blue points indicating smaller fitting factors, suggesting a lower similarity between the reconstruction and the original signal. Additionally, the size of each point corresponds to the chirp mass of the source. It's worth noting that sources located on TianQin's orbit tend to demonstrate lower fitting factors.
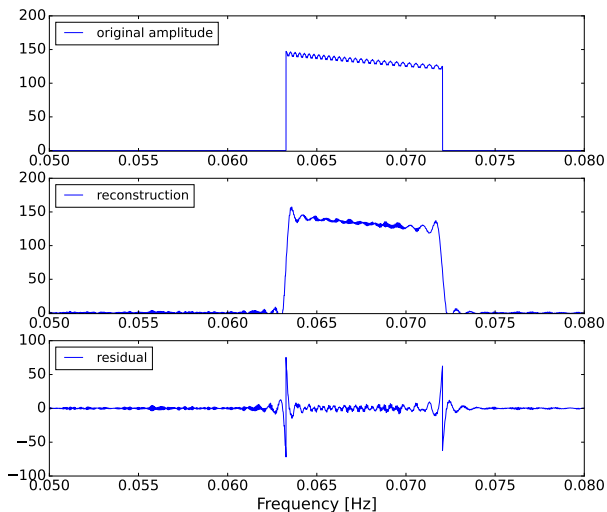
FIG. 5. The original amplitude from a random signal in 'A' channel with SNR 50 and the reconstruction from IPCA.The system in question has a chirp mass and symmetric mass ratio of approximately 8.3 $M_\odot$ and 0.18 respectively, and coalesces in a time-span of 28,923,738 seconds. Additional source parameters are referenced in Table II, including ($m_1 = 18.05, m_2 = 5.4 \cos(\iota) = 0.1, \psi = 2.74, \phi_c = 2.64, \chi_{1z} = -0.36, \chi_{2z} = 0.59$, and sky position at $\lambda = 2.07, \beta = -0.63$.)

Qin's response. This response is represented by transfer functions that relies on 6-dimensional parameters, including the sky location. In Figure 4, the sources in specific locations (like the source localization in Figure 1(b)), related to TianQin's orbit, have a tendency to display high oscillation. This behavior subsequently leads to a lower fitting factor, exemplified by the dark blue points.

C. Input data preparation

For both signal detection and point parameter estimation, the input data fed to the CNN comprises amplitude projections from the IPCA. The data from channels A and E undergo transformation using the respective IPCA components to obtain the projection values for each sample. The neural network performs better when the input data is normalized, shown in Figure 6. We normalized the input samples by the mean and standard deviation from the training samples. Visually identifying a gravitational wave signal submerged in the detector noise using the naked eye presents a significant challenge.
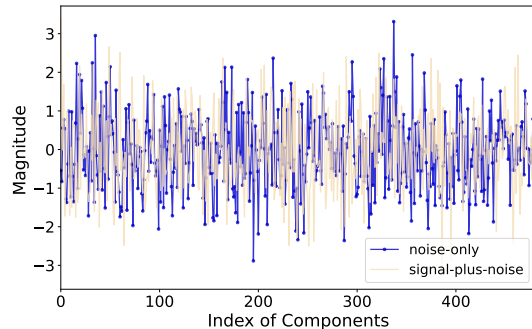

FIG. 6. The projected values from two different types of input samples in the 'A' channel have been normalized. The SNR is set to 50. The associated masses of the bodies are $32.4M_\odot$ and $8.62M_\odot$, and they occupy a sky position marked by $\lambda = 0.17$ Rad and $\beta = -5.17$ Rad. Their spins are quantified as -0.65 and -0.36, while the coalescence time is set as 30020992 seconds. A similar scenario is observed with projection values from data on the 'E' channel.

For the signal detection task, the training data are derived from the detector channel data and consist of

two distinct types: signal-plus-noise samples and pure noise samples. We utilize the amplitude projection values as inputs to the CNN. Within these input samples, we assign a label of 1 if a signal is present in the original data; conversely, if the sample does not contain a signal, it receives a label of 0.

In the point parameter estimation task, the input data for the CNN consists of amplitude projections, but exclusively from signal-plus-noise samples. These samples are obtained from simulated detector data containing signals, where the data is represented as $d(\theta) = s(\theta) + n$. To facilitate proper training, we employ normalized physical parameters as labels, as defined by the following equation:

$$\theta_{\text{norm}} = \frac{(\theta_{\text{origin}} - \theta_{\text{min}})}{(\theta_{\text{max}} - \theta_{\text{min}})} \qquad (19)$$

Here, θ_{origin} represents the original parameter value, and θ_{norm} serves as the normalized parameter value, which is utilized as the input label for the point estimation CNN model. During the training process for point parameter estimation, we conducted experiments with multidimensional parameter labels for the CNN. In comparison to using single labels for chirp mass, the utilization of multidimensional labels offers a wealth of information for training, thereby enhancing the effectiveness of CNN training. By experiments, we determined that 6-dimensional labels ($\mathcal{M}_{\text{norm}}$, $t_{c,\text{norm}}$, $\chi_{1z,\text{norm}}$, $\chi_{2z,\text{norm}}$, η_{norm} and $f_{0,\text{norm}}$) are the most optimal choice.

D. Training

A concise overview of the training settings used in each task is presented in Table III, and we will delve into a more detailed explanation later. The training strategy for both tasks is the same, therefore, let's consider the training process in task 1 (detection), as an example. The trained network is aimed at distinguishing the amplitude projections between 'signal-plus-noise' samples and 'noise-only' samples. It can then make judgments regarding the presence of a signal based on the projection values obtained from IPCA.

We start by training a CNN model with the training datasets of the fixed SNR of 50. We employed batch input with a size of 1000 to feed training samples into the network. Subsequently, we computed the binary cross-entropy loss using Equation 12 for each epoch. The model's hyperparameters were then updated using the Adam optimization algorithm. This entire training process spans multiple epochs. Following each training epoch, we assess the model's performance using a validation dataset to track the training progress and mitigate overfitting issues through comparison of the validation loss with the training loss, such as depicted in Figure 7. In response to the validation outcomes, we fine-tune the model's hyperparameters, including adjusting the learning rate and modifying the CNN architecture, to enhance its performance. Once the model has reached convergence, we decrease the fixed SNR 50 for generating new training samples. We decided to retain the working architecture for SNR values between 30 and 50. We have chosen 30 as the lower bound for the SNR because the network did not perform well for the lower SNR values.
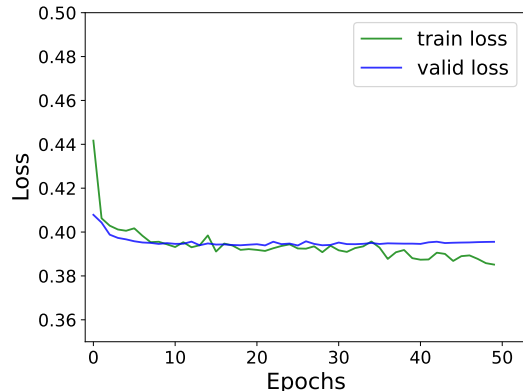

FIG. 7. The binary cross-entropy loss was calculated using fixed SNR 50 samples over 50 epochs. The best model is at 26th epoch.
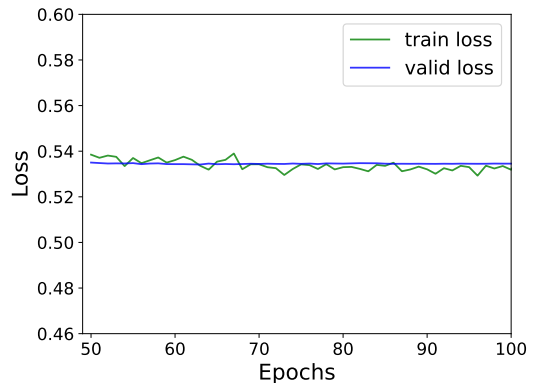

FIG. 8. The binary cross-entropy loss was calculated using samples with a uniformly sampled SNR ranging from 30 to 50 over 30 epochs. The best model is at 63th epoch.

The fine-tuned architecture of the CNN is detailed in Table I, consisting of a total of 6 convolution layers, 1 batch normalization layer, 2 fully connected layers and 1 dropout layer. We configured the Adam optimizer [60] with the following parameters: a learning rate of $\gamma = 10^{-4}$, momentum settings of $\beta_1 = 0.9$ and $\beta_2 = 0.999$, along with a decay factor of $\epsilon = 10^{-8}$. Additional training settings are provided in Table III. In the accompanying Figure 6, we observe that the difference in projection values between a 'signal-plus-noise' sample and a 'noise-only' sample is minimal, with the projection values exhibiting variations in different eigenvectors across a broad range. The introduction of a batch normalization layer after the first convolution layer plays

TABLE III. The training configuration. The total dataset will be divided, with 10% allocated for validation data and the remaining 90% used for training.

	setting	task 1 : signal detection	task 2 : point estimation
data samples		two categories: signal-plus-noise samples and noise-only samples	signal-plus-noise samples
	N_{total}	12000000	12000000
	labels	signals-plus-noise: 1, pure-noise : 0	θ_{norm}
best epoch / maximum epoch number		63/100	86/100
batch size		1000	1000
learning rate	$\gamma = 0.0001$		
optimizer	Adam		
training time		~ 50 hours	~ 50 hours
output model		detection CNN model trained with uniform SNR [30, 50] samples	point estimation CNN model trained with the uniform SNR [30, 50] samples

a crucial role in data normalization, aiding the model in achieving quicker convergence. Furthermore, we employed a learning rate reduction strategy, whereby we reduced the learning rate by a factor of 0.35 after 20 epochs in the absence of a minimum 0.005 improvement in the validation loss during that period.

In the final experiment, we utilize the top-performing CNN with $10,000,000$ SNR 50 training samples as the pre-training model, employing $12,000,000$ training samples with uniformly sampled SNR ranging from 30 to 50 to continue training the model in total 300 epochs. This process yields the converged detection CNN model. The loss curves for the uniformly sampled SNR data are shown in Figure 8, The validation loss aligns closely with the training loss and keeps stable.

In task 2 (point estimation of the physical parameter), the training strategies employed are identical to those utilized in task 1. This was decided to ensure the convolution kernels in both tasks of signal detection and point estimation would capture the same area of receptive field pertinent to sBBH GW signals. However, a smoother latent space was maintained in point estimation as compared to signal detection, due to the different activation functions in use.

V. RESULTS

A. Detection

We use the receiver operator characteristics (ROC) to describe the performance of the trained CNN in detecting sBBH signals. When we set a detection statistical threshold $y_{threshold}$, if the statistic returned by the CNN (i.e., the probability that the data when the GW signal is present) exceeds the threshold $y_{threshold}$, it can be recognized as the data contains a signal; otherwise, it is recognized as a noise sample. We computed both the false alarm probability (FAP) and true alarm probability (TAP) across various detection thresholds, constructing the ROC curve. The FAP provides a measure of the number of noise samples inaccurately identified as GW

TABLE IV. When FAP is set to 10% , the corresponding TAP for each distinct SNR testing data set are individually displayed refs to the Figure. 9

FAP = 10%	SNR			
	50	40	30	20
TAP	90.9%	61.8%	27.4%	14.7%

signals. Simultaneously, the TAP, offers a measure of the number of signal-plus-noise samples correctly identified as GW signals.

To evaluate the performance of the detection CNN model, we assemble test data with a fixed SNR by rescaling the luminosity distance. Meanwhile, other source parameters remain consistent with the distribution found in the training data, as indicated in Table II. Test samples are generated at fixed SNR of 20, 30, 40, and 50. We choose a reference value for false alarm probability to be 10% because given the long duration of the signals, from several months to years, coupled with the data segment's length of three months, it's acceptable for the false alarm probability to hit one false alarm per 100 tests. As such, every test dataset includes 2500 pure noise samples alongside an equivalent number of signal-plus-noise samples.

Figure 9 showcases the detection ability of the CNN model, which has been trained with SNR uniformly distributed within the 30 to 50 range, to differentiate signals based on their varying SNR. It is observed that an increase in the SNR corresponds to an augmentation in the TAP, with the specifics delineated in Table V A. The final trained detection CNN model can effectively detect sBBH sources with an SNR of 50 or higher, achieving an area under the curve (AUC) of 0.961. Additionally, it demonstrates that sources with fixed SNRs of 40, 30, and 20 achieved corresponding area under the curves (AUCs) values of 0.847, 0.669, and 0.557, respectively.
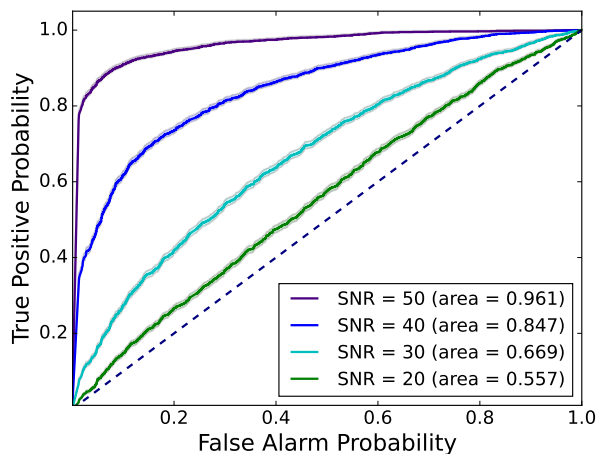
FIG. 9. The ROC curves are displayed using testing data at different fixed SNR. Each curve is represented by a distinct color corresponding to a specific SNR. The AUC quantifies the volume beneath the ROC curve, with a higher AUC indicating a higher detection probability. Additionally, we've indicated the 1-σ confidence intervals with shaded regions.

B.   Mass estimation

In order to estimate the value of the chirp mass by the point parameter estimation CNN model trained with a uniformly sampled SNR ranging from 30 to 50, we took 2000 test samples with a fixed SNR of 50 which had the same parameter distribution as the training samples. As shown in Figure 10, most of the chirp mass predictions are tightly distributed along the diagonal, indicating that there is consistency between the predicted values and actual values of the source's chirp mass. In this test dataset, 90% of the samples have an absolute error of 2.48 $M_\odot$ and a relative error of 0.12. This means that our model can predict the chirp mass of SNR 50 signals in Tian-Qin runs. Leaving aside those sources with larger errors, the significant variance observed in most sources yielding accurate predictions can largely be ascribed to added detector noise.

We also studied the characteristics of signals with larger prediction errors for chirp mass (absolute error exceeding 10 solar masses). We found that some signals have a smaller fitting factor, indicating that more information was lost after the IPCA projection. Signals characterized by a larger symmetric mass ratio or larger effective spins will exhibit a reduced number of data points in frequency domain within the same duration. This particular property introduces challenges for CNNs as they strive to effectively learn these signal features and make accurate predictions of the chirp mass. From the current results, the point estimation CNN can still quickly provide chirp mass predictions for most sBBH signals of SNR 50 with a good accuracy.
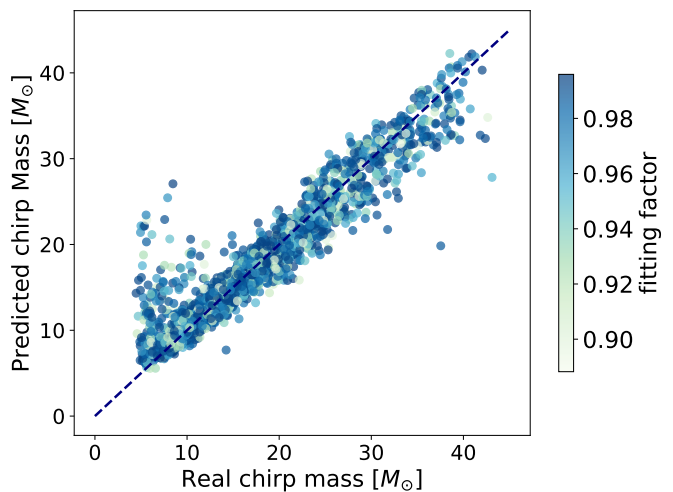

FIG. 10. The scatter plot illustrates the estimation of chirp mass of the fixed SNR 50 test signals compared to the actual values. The horizontal axis represents the actual chirp mass of the test samples, while the vertical axis displays the predictions obtained from the point estimation CNN model trained with uniformly sampled SNR samples. The error is calculated by subtracting the predicted chirp mass from the actual chirp mass: error = $\mathcal{M}_{\text{actual}} - \mathcal{M}_{\text{pred}}$.

VI.   CONCLUSION AND DISCUSSION

In this study, we effectively compressed the data through low rank matrix approximation, and extracted the main features of the frequency signal amplitude, thereby building the IPCA model. By projecting the simulated detector data (including signal plus noise samples and pure noise samples) through IPCA, we classify the compressed data into two categories by a trained CNN and show the sensitivity via the ROC curve. Simultaneously, in terms of the point estimation of parameters, we utilized the CNN model to estimate chirp mass, marginalising over all other parameters, for the detector data containing sBBH signals with a SNR of 50, and estimated its measurement error (90% confidence interval) to be 2.48 solar masses.

Regarding sBBH search performance, we have also recognized the potential for further enhancing our search model's sensitivity. Given the challenges of analyzing prolonged GW signals and the information loss incurred during the IPCA data compression process, the priority lies in improving the compression algorithm and feature extraction rather than further enhancing the existing neural network architecture. To be specific, the current IPCA model demonstrates certain limitations in reconstructing complex signals, there remains a need to investigate the influence of wave source parameter distribution and sampling methods on IPCA performance. Therefore, in future research, we plan to explore feature extraction methods aimed at minimizing information loss while maintaining high compression efficiency.

These may include using feedforward neural networks, as proposed in [61].

Additionally, some might suggest utilizing a classic architecture of CNN. We attempted to use the *resnet-18* model (with approximately 10^7 hyperparameters), but it exhibited susceptibility to overfitting and required a significantly larger number of training samples and training time. Specifically, our *resnet-18* experiment demanded around 10^7 training samples and approximately 300 hours for 100 training epochs, which was more than three times the computational time of our current model.

Looking at the analysis of sBBH search results using the TianQin detector, according to existing theoretical analyses [16], confirming the presence of signals in data from a single space-borne detector is a formidable challenge. The challenge arises primarily from the immense number of templates required for a coherent search, exceeding 10^{31} [16], which makes confirming the presence of signals within data from a single space-borne detector particularly difficult. Despite the absence of phase information in our study and the information loss during IPCA, the neural networks can still effectively detect some sBBH signals at a 10% false alarm probability. If a sBBH candidate is proximate to our observatory, our detection CNN model is highly likely to identify it with relatively little computational burden. For future sBBH signal detection, we consider the joint observation by multiple detectors, such as TianQin and LISA, with the hope of lowering the SNR threshold, providing richer information for the neural network, and effectively detecting sBBH signals with lower SNR. Additionally, we may explore the use of more complex neural network architectures to enhance sensitivity.

In terms of point estimation of chirp mass, the current CNN results demonstrate the ability of neural network algorithms to capture the characteristics of sBBH signals, although achieving accurate prediction for all physical parameters has proved challenging in our experiments. This is largely due to our reliance on amplitude in the frequency domain. One possible improvement could involve expanding the information input into the neural network, including data and label representations. Furthermore, optimizing the neural network architecture used, such as considering the application of autoencoders, and normalizing flow, can contribute to more precise parameter estimation.

In conclusion, this strategy of combining data compression and search presents a promising new approach for the detection of long-lived sBBH signals. Intending to refine this method further in the future, our search for early inspiral phases of sBBH will directly contribute to a better understanding of the formation channels of sBBH.

ACKNOWLEDGMENTS

[1] B. P. Abbott et al. (LIGO Scientific, Virgo), 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), arXiv:1811.12907 [astro-ph.HE].

[2] R. Abbott et al. (LIGO Scientific, Virgo), 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), arXiv:2010.14527 [gr-qc].

[3] The LIGO Scientific Collaboration, the Virgo Collaboration, the KAGRA Collaboration, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, N. Adhikari, R. X. Adhikari, and et al., GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run, arXiv e-prints , arXiv:2111.03606 (2021), arXiv:2111.03606 [gr-qc].

[4] J. Luo et al. (TianQin), TianQin: a space-borne gravitational wave detector, Class. Quant. Grav. **33**, 035010 (2016), arXiv:1512.02076 [astro-ph.IM].

[5] P. Amaro-Seoane, H. Audley, S. Babak, J. Baker, E. Barausse, P. Bender, E. Berti, P. Binetruy, M. Born, D. Bortoluzzi, J. Camp, C. Caprini, V. Cardoso, M. Colpi, J. Conklin, N. Cornish, C. Cutler, K. Danzmann, R. Dolesi, L. Ferraioli, V. Ferroni, E. Fitzsimons, J. Gair, L. Gesa Bote, D. Giardini, F. Gibert, C. Grimani, H. Halloin, G. Heinzel, T. Hertog, M. Hewitson, K. Holley-Bockelmann, D. Hollington, M. Hueller, H. Inchauspe, P. Jetzer, N. Karnesis, C. Killow, A. Klein, B. Klipstein, N. Korsakova, S. L. Larson, J. Livas, I. Lloro, N. Man, D. Mance, J. Martino, I. Mateos, K. McKenzie, S. T. McWilliams, C. Miller, G. Mueller, G. Nardini, G. Nelemans, M. Nofrarias, A. Petiteau, P. Pivato, E. Plagnol, E. Porter, J. Reiche, D. Robertson, N. Robertson, E. Rossi, G. Russano, B. Schutz, A. Sesana, D. Shoemaker, J. Slutsky, C. F. Sopuerta, T. Sumner, N. Tamanini, I. Thorpe, M. Troebs, M. Vallisneri, A. Vecchio, D. Vetrugno, S. Vitale, M. Volonteri, G. Wanner, H. Ward, P. Wass, W. Weber, J. Ziemer, and P. Zweifel, Laser Interferometer Space Antenna, arXiv e-prints , arXiv:1702.00786 (2017), arXiv:1702.00786 [astro-ph.IM].

[6] S. Marsat, J. G. Baker, and T. Dal Canton, Exploring the Bayesian parameter estimation of binary black holes with LISA, Phys. Rev. D **103**, 083011 (2021), arXiv:2003.00357 [gr-qc].

[7] M. C. Digman and N. J. Cornish, Parameter estimation for stellar-origin black hole mergers in LISA, Phys. Rev. D **108**, 023022 (2023), arXiv:2212.04600 [gr-qc].

[8] A. Toubiana, S. Babak, S. Marsat, and S. Ossokine, Detectability and parameter estimation of GWTC-3 events with LISA, Phys. Rev. D **106**, 104034 (2022), arXiv:2206.12439 [gr-qc].

[9] S. Marsat, J. G. Baker, and T. D. Canton, Exploring the Bayesian parameter estimation of binary black holes with LISA, Phys. Rev. D **103**, 083011 (2021), arXiv:2003.00357 [gr-qc].

[10] L.-G. Zhu, L.-H. Xie, Y.-M. Hu, S. Liu, E.-K. Li, N. R. Napolitano, B.-T. Tang, J.-D. Zhang, and J. Mei, Constraining the Hubble constant to a precision of about 1% using multi-band dark standard siren detections, Science China Physics, Mechanics, and Astronomy **65**, 259811 (2022), arXiv:2110.05224 [astro-ph.CO].

[11] A. Sesana, Prospects for Multiband Gravitational-Wave Astronomy after GW150914, Phys. Rev. Lett. **116**, 231102 (2016), arXiv:1602.06951 [gr-qc].

[12] S. Liu, Y.-M. Hu, J.-d. Zhang, and J. Mei, Science with the TianQin observatory: Preliminary results on stellar-mass binary black holes, Phys. Rev. D **101**, 103027 (2020), arXiv:2004.14242 [astro-ph.HE].

[13] X. Lyu, E.-K. Li, and Y.-M. Hu, Parameter estimation of stellar mass binary black holes in the network of TianQin and LISA, Phys. Rev. D **108**, 083023 (2023), arXiv:2307.12244 [gr-qc].

[14] H. Wang, I. Harry, A. Nitz, and Y.-M. Hu, The Challenge of Eccentricity when Observing Stellar-mass Binary Black Holes with Space-Based Gravitational Wave Detectors, arXiv e-prints , arXiv:2304.10340 (2023), arXiv:2304.10340 [astro-ph.HE].

[15] S.-J. Huang, Y.-M. Hu, X. Chen, J.-d. Zhang, E.-K. Li, Z. Gao, and X.-Y. Lin, Measuring the Hubble constant using strongly lensed gravitational wave signals, JCAP **08**, 003, arXiv:2304.10435 [astro-ph.CO].

[16] C. J. Moore, D. Gerosa, and A. Klein, Are stellar-mass black-hole binaries too quiet for LISA?, Mon. Not. Roy. Astron. Soc. **488**, L94 (2019), arXiv:1905.11998 [astro-ph.HE].

[17] B. J. Owen, Search templates for gravitational waves from inspiraling binaries: Choice of template spacing, Phys. Rev. D **53**, 6749 (1996), arXiv:gr-qc/9511032.

[18] D. Bandopadhyay and C. J. Moore, LISA stellar-mass black hole searches with semi-coherent and particle-swarm methods, arXiv e-prints , arXiv:2305.18048 (2023), arXiv:2305.18048 [gr-qc].

[19] D. George and E. A. Huerta, Deep Neural Networks to Enable Real-time Multimessenger Astrophysics, Phys. Rev. D **97**, 044039 (2018), arXiv:1701.00008 [astro-ph.IM].

[20] 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), arXiv:1712.06041 [astro-ph.IM].

[21] M. B. Schäfer et al., First machine learning gravitational-wave search mock data challenge, Phys. Rev. D **107**, 023021 (2023), arXiv:2209.11146 [astro-ph.IM].

[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), arXiv:1711.03121 [gr-qc].

[23] 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), arXiv:1904.08693 [astro-ph.IM].

[24] 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, Physics Letters B **815**, 136161 (2021), arXiv:2012.13101 [astro-ph.IM].

[25] E. A. Huerta et al., Accelerated, scalable and reproducible AI-driven gravitational wave detection, Nature Astron. **5**, 1062 (2021), arXiv:2012.08545 [gr-qc].

[26] A. J. K. Chua and M. Vallisneri, Learning Bayesian posteriors with neural networks for gravitational-wave inference, Phys. Rev. Lett. **124**, 041102 (2020), arXiv:1909.05966 [gr-qc].

[27] H. Gabbard, C. Messenger, I. S. Heng, F. Tonolini, and R. Murray-Smith, Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy, Nature Phys. **18**, 112 (2022), arXiv:1909.06296 [astro-ph.IM].

[28] M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf, Real-Time Gravitational Wave Science with Neural Posterior Estimation, Phys. Rev. Lett. **127**, 241103 (2021), arXiv:2106.12594 [gr-qc].

[29] M. Dax, S. R. Green, J. Gair, M. Pürrer, J. Wildberger, J. H. Macke, A. Buonanno, and B. Schölkopf, Neural Importance Sampling for Rapid and Reliable Gravitational-Wave Inference, Phys. Rev. Lett. **130**, 171403 (2023), arXiv:2210.05686 [gr-qc].

[30] U. Bhardwaj, J. Alvey, B. K. Miller, S. Nissanke, and C. Weniger, Peregrine: Sequential simulation-based inference for gravitational wave signals, arXiv e-prints , arXiv:2304.02035 (2023), arXiv:2304.02035 [gr-qc].

[31] J. Langendorff, A. Kolmus, J. Janquart, and C. Van Den Broeck, Normalizing Flows as an Avenue to Studying Overlapping Gravitational Wave Signals, Phys. Rev. Lett. **130**, 171402 (2023), arXiv:2211.15097 [gr-qc].

[32] L. Dinh, J. Sohl-Dickstein, and S. Bengio, Density estimation using Real NVP, arXiv e-prints , arXiv:1605.08803 (2016), arXiv:1605.08803 [cs.LG].

[33] W.-H. Ruan, H. Wang, C. Liu, and Z.-K. Guo, Rapid search for massive black hole binary coalescences using deep learning, Phys. Lett. B **841**, 137904 (2023), arXiv:2111.14546 [astro-ph.IM].

[34] X.-T. Zhang, C. Messenger, N. Korsakova, M. L. Chan, Y.-M. Hu, and J.-d. Zhang, Detecting gravitational waves from extreme mass ratio inspirals using convolutional neural networks, Phys. Rev. D **105**, 123027 (2022), arXiv:2202.07158 [astro-ph.HE].

[35] T. Zhao, R. Lyu, Z. Ren, H. Wang, and Z. Cao, Space-based gravitational wave signal detection and extraction with deep neural network, arXiv e-prints , arXiv:2207.07414 (2022), arXiv:2207.07414 [gr-qc].

[36] P. G. Krastev, Real-Time Detection of Gravitational Waves from Binary Neutron Stars using Artificial Neural Networks, Phys. Lett. B **803**, 135330 (2020), arXiv:1908.03151 [astro-ph.IM].

[37] A. L. Miller et al., How effective is machine learning to detect long transient gravitational waves from neutron stars in a real search?, Phys. Rev. D **100**, 062005 (2019),

arXiv:1909.02262 [astro-ph.IM].

[38] 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. (Beijing) **15**, 24602 (2020).

[39] 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), arXiv:2006.01509 [astro-ph.HE].

[40] H. Yu, R. X. Adhikari, R. Magee, S. Sachdev, and Y. Chen, Early warning of coalescing neutron-star and neutron-star-black-hole binaries from the nonstationary noise background using neural networks, Phys. Rev. D **104**, 062004 (2021), arXiv:2104.09438 [gr-qc].

[41] 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), arXiv:2205.04750 [gr-qc].

[42] J. a. Aveiro, F. F. Freitas, M. Ferreira, A. Onofre, C. Providência, G. Gonçalves, and J. A. Font, Identification of binary neutron star mergers in gravitational-wave data using object-detection machine learning models, Phys. Rev. D **106**, 084059 (2022), arXiv:2207.00591 [astro-ph.IM].

[43] J. Bayley, C. Messenger, and G. Woan, Rapid parameter estimation for an all-sky continuous gravitational wave search using conditional varitational auto-encoders, Phys. Rev. D **106**, 083022 (2022), arXiv:2209.02031 [astro-ph.IM].

[44] A. Mangiagli, A. Klein, A. Sesana, E. Barausse, and M. Colpi, Post-Newtonian phase accuracy requirements for stellar black hole binaries with LISA, Phys. Rev. D **99**, 064056 (2019), arXiv:1811.01805 [gr-qc].

[45] M. Zevin, S. S. Bavera, C. P. L. Berry, V. Kalogera, T. Fragos, P. Marchant, C. L. Rodriguez, F. Antonini, D. E. Holz, and C. Pankow, One Channel to Rule Them All? Constraining the Origins of Binary Black Holes Using Multiple Formation Pathways, Astrophys. J. **910**, 152 (2021), arXiv:2011.10057 [astro-ph.HE].

[46] R. Abbott et al. (KAGRA, VIRGO, LIGO Scientific), Population of Merging Compact Binaries Inferred Using Gravitational Waves through GWTC-3, Phys. Rev. X **13**, 011048 (2023), arXiv:2111.03634 [astro-ph.HE].

[47] P. Ajith et al., Inspiral-merger-ringdown waveforms for black-hole binaries with non-precessing spins, Phys. Rev. Lett. **106**, 241101 (2011), arXiv:0909.2867 [gr-qc].

[48] M.-Y. Zhou, X.-C. Hu, B. Ye, S. Hu, D.-D. Zhu, X. Zhang, W. Su, and Y. Wang, Orbital effects on time delay interferometry for TianQin, Phys. Rev. D **103**, 103026 (2021), arXiv:2102.10291 [astro-ph.IM].

[49] M. Tinto, J. W. Armstrong, and F. B. Estabrook, Discriminating a gravitational-wave background from instrumental noise using time-delay interferometry, Classical and Quantum Gravity **18**, 4081 (2001).

[50] M. Tinto and J. W. Armstrong, Cancellation of laser noise in an unequal-arm interferometer detector of gravitational radiation, Phys. Rev. D **59**, 102003 (1999).

[51] Z.-C. Liang, Y.-M. Hu, Y. Jiang, J. Cheng, J.-d. Zhang, and J. Mei, Science with the TianQin Observatory: Preliminary results on stochastic gravitational-wave background, Phys. Rev. D **105**, 022001 (2022), arXiv:2107.08643 [astro-ph.CO].

[52] S. Marsat and J. G. Baker, Fourier-domain modulations and delays of gravitational-wave signals, arXiv e-prints , arXiv:1806.10734 (2018), arXiv:1806.10734 [gr-qc].

[53] J. Shlens, A Tutorial on Principal Component Analysis, arXiv e-prints , arXiv:1404.1100 (2014), arXiv:1404.1100 [cs.LG].

[54] D. Ross, J. Lim, R. Lin, and M. Yang, Incremental learning for robust visual tracking, International Journal of Computer Vision **77**, 125 (2008).

[55] S. Raschka, J. Patterson, and C. Nolet, Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence, arXiv preprint arXiv:2002.04803 (2020).

[56] I. J. Goodfellow, Y. Bengio, and A. Courville, Deep Learning (MIT Press, Cambridge, MA, USA, 2016) http://www.deeplearningbook.org.

[57] C. Verma, A. Reza, G. Gaur, D. Krishnaswamy, and S. Caudill, Can Convolution Neural Networks Be Used for Detection of Gravitational Waves from Precessing Black Hole Systems?, arXiv e-prints , arXiv:2206.12673 (2022), arXiv:2206.12673 [gr-qc].

[58] D. George, H. Shen, and E. A. Huerta, Glitch Classification and Clustering for LIGO with Deep Transfer Learning, in NiPS Summer School 2017 (2017) arXiv:1711.07468 [astro-ph.IM].

[59] S. Alvarez-Lopez, A. Liyanage, J. Ding, R. Ng, and J. McIver, GSpyNetTree: A signal-vs-glitch classifier for gravitational-wave event candidates, arXiv e-prints , arXiv:2304.09977 (2023), arXiv:2304.09977 [gr-qc].

[60] D. P. Kingma and J. Ba, Adam: A Method for Stochastic Optimization, arXiv e-prints , arXiv:1412.6980 (2014), arXiv:1412.6980 [cs.LG].

[61] O. Gramaxo Freitas, J. Calderón Bustillo, J. A. Font, S. Nunes, A. Onofre, and A. Torres-Forné, Comparison of neural network architectures for feature extraction from binary black hole merger waveforms, arXiv e-prints , arXiv:2307.16668 (2023), arXiv:2307.16668 [gr-qc].