Using deep learning to predict matched signal-to-noise ratio of gravitational waves

CunLiang Ma,[1] Sen Wang,[1] Wei Wang,[2] and Zhoujian Cao[3,4,*]

[1]*School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou, 341000, China*

[2]*School of Computer Science, Fudan University, Shanghai 201203, China*

[3]*Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China*

[4]*School of Fundamental Physics and Mathematical Sciences, Hangzhou Institute for Advanced Study, UCAS, Hangzhou 310024, China*

(Received 2 October 2023; accepted 10 January 2024; published 5 February 2024)

The existing matched filtering method for gravitational wave (GW) search relies on a template bank. The computational efficiency of this method scales with the sizes of the templates within the bank. Higher-order modes and eccentricity will play an important role when third-generation detectors operate in the future. In this case, traditional GW search methods will hit computational limits. To speed up the computational efficiency of GW searches, we propose the utilization of a deep learning (DL) model bank as a substitute for the template bank. This model bank predicts the latent templates embedded in the strain data. Combining an envelope extraction network and an astrophysical origin discrimination network, we realize a novel GW search framework. The framework can predict the GW signal's matched filtering signal-to-noise ratio (SNR). Unlike the end-to-end DL-based GW search method, our statistical SNR holds stronger physical interpretability than the p_{score} metric. Moreover, the intermediate results generated by our approach, including the predicted template, offer valuable assistance in subsequent GW data processing tasks such as parameter estimation and source localization. Compared to the traditional matched filtering method, the proposed method can realize real-time analysis.

DOI: 10.1103/PhysRevD.109.043009

I. INTRODUCTION

The direct detection of gravitational waves (GWs) holds immense value for both validating and refining physical theories, as well as advancing our understanding of cosmology. Managed by the LIGO-VIRGO-KAGRA (LVK) Collaboration, the second-generation ground-based GW detectors have successfully identified more than 90 confident events originating from compact binary coalescences (CBCs) throughout their first, second, and third observing runs [1–4]. These data have effectively substantiated general relativity [5] and facilitated the measurement of the Hubble constant [6]. Additionally, these data have been employed to explore evidence of cosmic strings [7], continuous GWs [8], and stochastic GW backgrounds [9]. For differences in the detectable frequency bands, space-based GW detectors [10–12] and third-generation (3G) ground-based detectors [13,14] will substantially enhance the diversity and quantity of astrophysical sources associated with GWs. Consequently, the GW search pipeline must continually improve to meet the forthcoming challenges.

The LVK organization employs PyCBC [15], GstLAL [16], MBTA [17], SPIIR [18], and CWB [19] for the searches of GW signals. The CWB uses an unmodeled method, whereas the other four pipelines are all based on the template waveforms and the matched filtering (MF) technique. Although MF offers the benefit of being physically interpretable, it suffers from the drawback of having low computational efficiency due to the large amount of the template waveforms. When the higher-order modes and orbit eccentricity play an important role in the 3G era, current template-bank-based algorithms will hit computational limits.

To speed up the MF-based GW search method, in this work, a deep learning (DL) model bank is proposed to substitute for the waveform bank. The templates are predicted by the DL model bank rather than contained in a waveform bank. Notably, the outputs of the proposed method correspond to matched filtering signal-to-noise ratios, and the advantage of the proposed method is the accelerated computational speed.

DL methods for GW detection have been extensively investigated. Back in 2018, George *et al.* [20] and Gabbard *et al.* [21] independently showcased the potential of a DL-based algorithm for GW detection. Thus far, the

*Corresponding author: zjcao@amt.ac.cn

DL-based approaches for GW detection have undergone significant development [22–34]. Furthermore, there also have been various endeavors to utilize deep learning for other aspects of GW data processing, such as localizing the astrophysical origin [35,36] and estimating the parameters of GW sources [37–40].

Like computer vision [41] and nature language processing [42] tasks, DL-based GW detection can often exhibit a black-box nature. This characteristic makes the end-to-end DL models for GW detection less suitable for making statistically significant claims about gravitational wave detections [43]. One contribution of our work is that we propose a novel scheme for GW search with feedback. Instead of directly performing the end-to-end GW detection task, we use DL models for template prediction, and these predicted templates are then integrated into the conventional MF (matched filtering)–based method.

The results show that most of the confident events in GWTC-1, GWTC-2, GWTC-2.1, and GWTC-3 can be identified. Notably, the proposed method is physically interpretable, because other than the p_{score} in the previous DL-based detection method [20,21], the event significance is measured by the signal-to-noise ratio, and the significance of each coincident trigger can be estimated by calculating the false-alarm rate (FAR).

Most of the time slices only contain background noise. Almost all the DL-based GW detection methods have a disadvantage in using a time-sliding method with a small time-sliding step [28]. To address this drawback, we use the coalescence-time information predicted by the envelope extraction model [31] to align the data analysis window. When the coalescence time is obtained, the potential templates will be predicted by the models in the denoising model bank.

The potential templates derived from the denoising model cannot be directly employed for MF, due to the possibility that the denoised output might not accurately resemble the binary black hole (BBH) GW shape. To address this concern, we propose the astrophysical origin identification models to determine whether the denoised waveform corresponds to the BBH GW shape. The potential templates are selected by these proposed identification models.

The temporal extent of the GW signal within the sensitivity band of the LIGO-VIRGO-KAGRA network is influenced by the masses of the binary system. For BBH systems, this duration ranges from a fraction of a second for higher masses to a few seconds for lower masses. Recognizing this variability, we organize the denoising model bank to account for specific subparameter spaces. These subspaces are defined by dividing the parameter range based on the binary system's masses. Consequently, each model in the denoising model bank is focused on templates within a particular subparameter space. The input duration for denoising models is then adjusted according to the corresponding subparameter space.

This paper is organized as follows: The proposed framework and the key stages in the framework are introduced in Sec. II. In Sec. III, we describe the training datasets and training schemes for the neural networks in the proposed framework. In Sec. IV, we describe the results of the stages tested by the test dataset, the test results against the real confident events reported by LIGO-VIRGO-KAGRA Collaboration, and our model performance against half a month of detected data by the Hanford detector. The conclusion and discussion are given in Sec. V.

II. DEEP LEARNING FRAMEWORK FOR MATCHED SIGNAL-TO-NOISE RATIO PREDICTION OF GRAVITATIONAL WAVES

To alleviate the computational overhead issue of the MF-based method, we propose a mutistep processing method as shown in Fig. 1. The whole framework consists of four key stages: significant time prediction (STP), preliminary templates prediction (PTP), template selection (TS), and matched filtering (MF). The details of the four stages will be introduced in the following four subsections. For reference convenience, we call the newly proposed framework MSNRnet.

A. The significant time prediction stage

The output of the STP stage is a time set $S_t = \{t_1, t_2, t_3, \ldots\}$, where t_i is the significant time. These significant times correspond to the coalescence time predicted by the envelope prediction network. If the strain data only contain background noise, the time set vanishes: $S_t = \varnothing$. We feed the whitened strain $\boldsymbol{s}_X$ (the subscript $X \in H, L$ denoting the Hanford or Livingston interferometer) to the envelope extraction network. The relation between $\boldsymbol{s}_X$ and the output of the envelope prediction network ($\hat{\boldsymbol{h}}_X^{env}$) can be denoted as

$$\hat{\boldsymbol{h}}_X^{env} = EnvNet(\boldsymbol{s}_X|\boldsymbol{W}_{env}), \tag{1}$$

where $EnvNet$ represents a parametrized system which is proposed in [31], and $\boldsymbol{W}_{env}$ represents the trainable weights of the envelope extraction model. If $\max(\hat{\boldsymbol{h}}_X^{env}) > 0.5$, then the coalescence time of interferometer X can be predicted by

$$t_X = \Delta t \times \arg\max_n \hat{h}_X^{env}(n), \tag{2}$$

where Δt is the sampling period, $\hat{h}_X^{env}(n)$ represents the predicted envelope amplitude at time $n \times \Delta t$, and $\arg\max_n$ means the index n which maximizes the envelope. In this work, we employ the envelope extraction model trained in [31] for direct envelope prediction. The envelope extraction method proposed in [31] was used to test the coalescence time of different detectors (Hanford and Livingston).

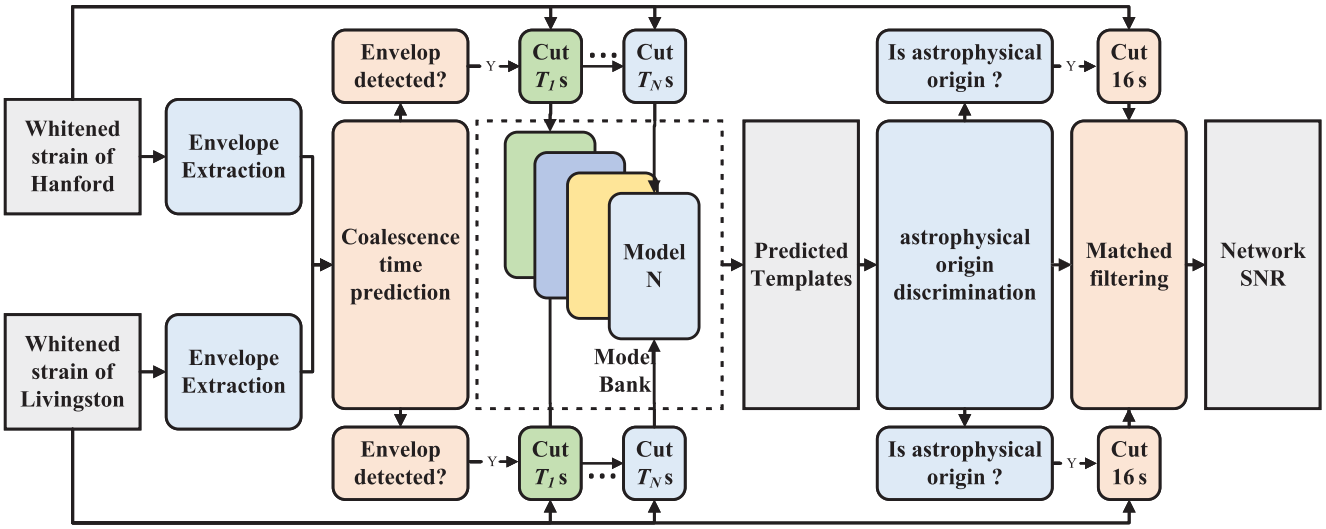

FIG. 1. The diagram of the proposed GW detection framework MSNRnet.

By contrast, the network is used to quickly find out the potential time segments that may contain GWs in the current work. In the STP stage, coincidence between the two detectors is not required. In cases where signals are strong in one detector but relatively weak in the other, the envelope extraction network may successfully extract the envelope from only one detector. In such instances, significant timing information will be predicted from just one detector. The algorithm in [31] used a scheme that forced coincidence, resulting in the omission of many signals in the work's results.

B. The preliminary templates prediction stage

In the case $S_t \neq \varnothing$, the PTP stage will be triggered. The outputs of the PTP stage are two template sets $S_{PT}^H = \{\hat{\boldsymbol{h}}_H^1, \hat{\boldsymbol{h}}_H^2, \hat{\boldsymbol{h}}_H^3, \ldots\}$ and $S_{PT}^L = \{\hat{\boldsymbol{h}}_L^1, \hat{\boldsymbol{h}}_L^2, \hat{\boldsymbol{h}}_L^3, \ldots\}$, where $\hat{\boldsymbol{h}}_X^i$ denotes the ith preliminary predicted template waveform. In the PTP stage, a DL model bank is used. The DL model bank is a set of DL models for denoising. The DL model bank can be denoted as

$$S_{\mathrm{DM}} = \{\mathrm{DenoiseModel}_i | i \in [1, N_M]\}, \tag{3}$$

where N_M is the number of the denoising models. Suppose the input of $\mathrm{DenoiseModel}_i$ is $\boldsymbol{s}_X^i \in \mathbb{R}^{1\times M_i}$, where $M_i = T_i \times 4096$, T_i is the time duration of the input strain of $\mathrm{DenoiseMode}_i$. The output can be denoted as

$$\hat{\boldsymbol{h}}_X^i = \mathrm{DenoiseModel}_i(\boldsymbol{s}_X^i | \boldsymbol{W}_{Mod_i}), \tag{4}$$

where $\hat{\boldsymbol{h}}_X^i$ is the denoised strain. $\boldsymbol{W}_{Mod_i}$ denotes the trainable weights of $\mathrm{DenoiseModel}_i$, which can be optimized by

$$\boldsymbol{W}_{Mod_i} = \arg\min_{\boldsymbol{W}_{Mod_i}} \frac{1}{N_T} \sum \|\mathrm{DenoiseModel}_i (\boldsymbol{s}_X^{i,k} | \boldsymbol{W}_{Mod_i}) - \boldsymbol{h}_X^{i,k}\|^2, \tag{5}$$

where $\boldsymbol{h}_X^{i,k}$ denotes the whitened signal buried in the kth whitened strain sample.

In contrast to recent approaches in GW denoising, like WaveNet [44], LSTM [45], and WaveFormer [46], our methodology employs a multitude of models. Each denoising model only focuses on a subset of the parameter space. The source parameter of the BBH is divided by the mass range of the binaries. The time duration of the GW signals within the sensitivity band of the detector network varies along with the BBH masses of the system. For the PyCBC pipeline, the candidate and background events are divided into three search classes based on template length [47,48]. Motivated by this, the input length of the denoising model varies with the concerned parameter range. The subparameter spaces and the input strain lengths of the denoising models are listed in Table I.

The U-Net-like model for the envelope extraction task is quite effective [31]. We have also adopted the U-Net-like model for denoising, and its structure is shown in Fig. 2. We use the same model structure for all the denoising

TABLE I. The subparameters and input strain lengths of denoising models. The unit of the BBH's mass is $M_\odot$.

	m_1			
m_2	(5,10]	(10,20]	(20,40]	(40,80]
(5,10]	2.00 s	1.75 s	1.50 s	1.25 s
(10,20]	⋯	1.50 s	1.00 s	0.75 s
(20,40]	⋯	⋯	0.75 s	0.50 s
(40,80]	⋯	⋯	⋯	0.25 s

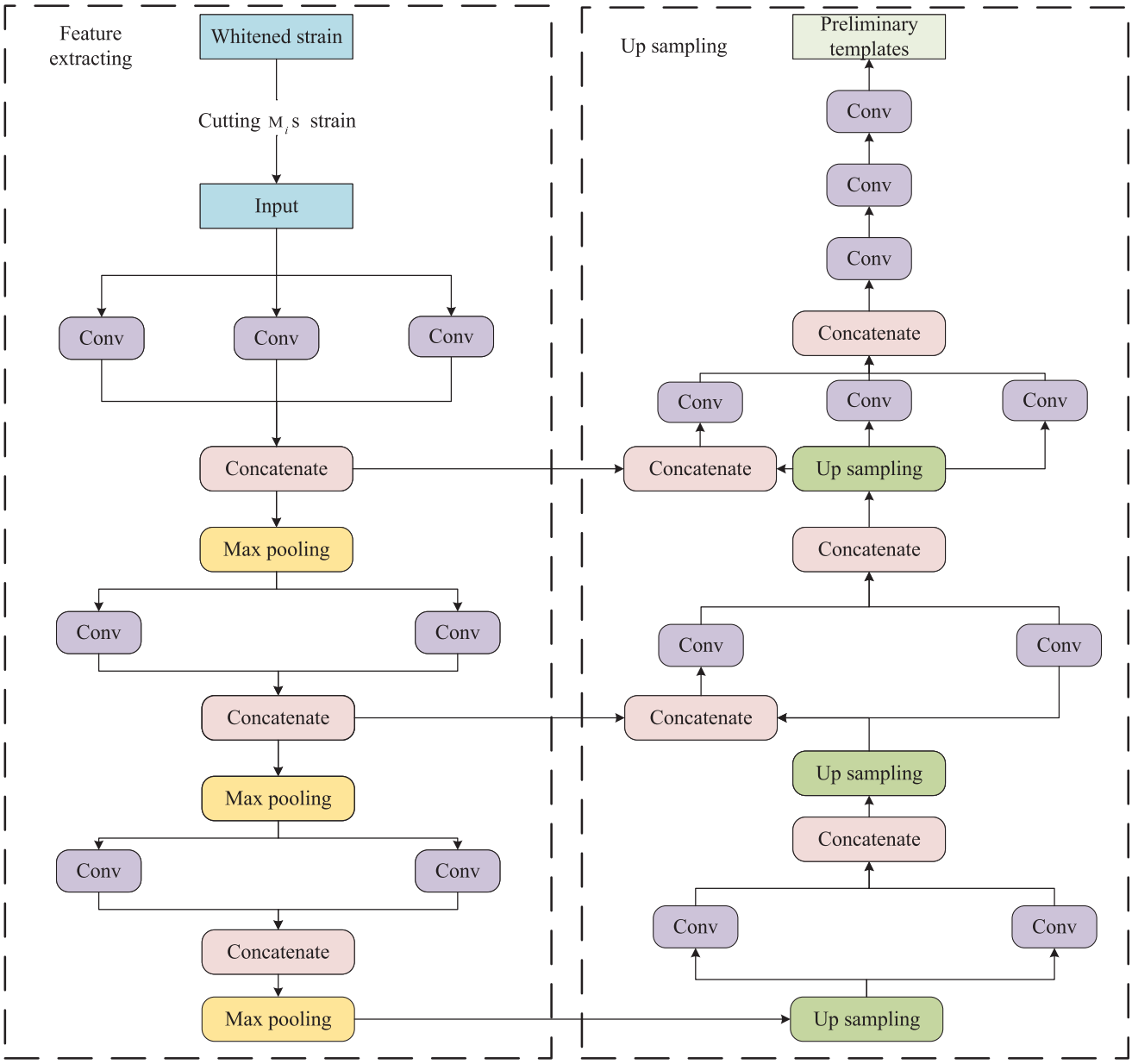

FIG. 2. Network structure of the denoising models.

models in the model bank. But we let the input shape of the denoising models vary with respect to the parameters of BBH. The different denoising models in the model bank are in charge of different parameter spaces. They are trained with different data with different input shapes.

C. The template selection stage

The astrophysical origin discrimination networks are used in this stage. The inputs of this stage are the preliminary templates $S^X_{PTP} = \{\hat{\boldsymbol{h}}^1_X, \ldots, \hat{\boldsymbol{h}}^{N_M}_X\}$, and the outputs of this stage are the selected templates that will be used in the matched filtering stage. Suppose the selected templates comprise a set S^X_{Tem}. If a preliminary template generated by the denoising model does not have a GW waveform shape, the matched filtering results will not reflect the GW's SNR information. Therefore, before the matched filtering stage, we select the templates according to the feature of the GW waveform shape. The STP stage does not mandate the coincidence phase. During the TS stage, we employ the Hanford and Livingston coincidence strategy. The predicted templates from both the Hanford and Livingston interferometers, possessing shapes of astrophysical origin, will be selected. However, the coincidence strategy has a disadvantage, in that it may result in missing some signals. Instances like GW170818, which were strong in one detector but relatively weak in the other, would be overlooked in the current scheme due to the requirement for coincidence.

We constructed astronomical origin discrimination networks, $\{AstroDisNet_i\}$, $i = 1, \ldots, N_M$, to do the selection work. An astrophysical origin discrimination network is a classification model which can distinguish the astronomically original waveforms from artificial ones. Plugging $\hat{\boldsymbol{h}}_X^i$ (the output of the ith denoising model of interferometer X) into $AstroDisNet_i$, we obtain the output $\boldsymbol{o}_{astro,X}^i$, where i denotes the ith astrophysical discrimination network and X denotes the interferometer:

$$\boldsymbol{o}_{\mathrm{astro},X}^i = AstroDisNet_i(\hat{\boldsymbol{h}}_X^i | \boldsymbol{W}_{\mathrm{astro}_i}), \tag{6}$$

where $\boldsymbol{W}_{\mathrm{astro}_i}$ denotes the trainable weights and is optimized by

$$\boldsymbol{W}_{\mathrm{astro}_i} = \arg\min_{\boldsymbol{W}_{\mathrm{astro}_i}} \frac{1}{N_{a_i}} \sum_k \mathrm{CrossEntropy}(\boldsymbol{y}^k, \boldsymbol{o}_{\mathrm{astro}}^{i,k}) \tag{7}$$

when the kth sample admits an astrophysical origin, $y^k = 1$; otherwise, $y^k = 0$. N_{a_i} is the number of samples. In this work, we use a convolutional network to utilize $AstroDisNet_i$. Every network only focuses on the waveforms corresponding to a subparameter space. When both $o_{astro,H}^i$ and $o_{astro,L}^i$ are greater than a threshold o^{th}, and the peak time difference of $\hat{\boldsymbol{h}}_H^i$ and $\hat{\boldsymbol{h}}_L^i$ is less than 15 ms, then $\hat{\boldsymbol{h}}_H^i$ and $\hat{\boldsymbol{h}}_L^i$ will be put into S_{Tem}^H and S_{Tem}^L, respectively. Consequently, the number of templates in S_{Tem}^H and S_{Tem}^L is the same.

We adopt the GW detection model designed by Gabbard *et al.* [21] as the structure of the astrophysical origin discrimination network [28,29,49,50]. For reference convenience, we list the network structure in Table II. Because the time length of the input varies with the focused parameter space, the input shape of the networks is different. We apply the Elu activation function after every layer except the last dense layer. For each convolutional layer, no padding is applied. A soft-max activation function is used after the last dense layer.

D. The matched filtering stage

The SNR information can further quantify the confidence level of the signal. While the template selection phase can exclude most false triggers, we believe it cannot definitively determine whether an event is real. As a result, we continue to analyze SNR information through matched filtering. If the set of the selected template $S_{Tem} \neq \varnothing$, the matched filtering stage will be triggered. The output of this stage is two SNR sets, $S_{snr}^H = \{SNR_1^H, SNR_2^H, SNR_3^H, \ldots\}$ and $S_{snr}^L = \{SNR_1^L, SNR_2^L, SNR_3^L, \ldots\}$. The number of SNRs is equal to the number of templates in the sets S_{Tem}^H and S_{Tem}^L. Here, we detail the method for obtaining S_{snr}^X (interferometer X). For the template $h_i^X(t) \in S_{Tem}$, the output of the matched filtering can be written as

$$\rho_{i,X}(t) = \frac{\langle \boldsymbol{d} | \boldsymbol{h}_i^X \rangle(t)}{\sqrt{\langle \boldsymbol{h}_i^X | \boldsymbol{h}_i^X \rangle}}, \tag{8}$$

where $d(t)$ is the strain. $\langle \boldsymbol{d} | \boldsymbol{h}_i^X \rangle(t)$ denotes the time-dependent inner product of $d(t)$ and $h_i^X(t)$. In this work, $\boldsymbol{d}$ and $\boldsymbol{h}_i^X$ are all whitened by the amplitude spectrum density (ASD) of the background noise, so

$$\langle \boldsymbol{d} | \boldsymbol{h}_i \rangle(t) = 4\mathcal{R}e \int_0^\infty \tilde{d}(f) \tilde{h}_i^{X^*}(f) e^{-2\pi j f t} df \tag{9}$$

and

$$\langle \boldsymbol{h}_i | \boldsymbol{h}_i \rangle = 4 \int_0^\infty \tilde{h}_i^X(f) \tilde{h}_i^{X^*}(f) df, \tag{10}$$

where $\tilde{d}(f)$, $\tilde{h}_i^X(f)$ are the Fourier transform results of $d(t)$ and $h_i^X(t)$, and $\tilde{h}_i^{X^*}(f)$ is the complex conjugate of $\tilde{h}_i^X(f)$. Then, the network SNR can be calculated by

$$SNR = \max_i \sqrt{\rho_{i,H}^2 + \rho_{i,L}^2}\,. \tag{11}$$

TABLE II. The structure of the astronomical origin discrimination network.

Layer	Layer type	Number of neurons	Filter size	Max pool size	Dropout	Activation function
1	Convolution	32	32	Not applicable	0	Elu
2	Convolution	32	32	8	0	Elu
3	Convolution	64	16	Not applicable	0	Elu
4	Convolution	64	16	4	0	Elu
5	Convolution	128	8	Not applicable	0	Elu
6	Convolution	128	8	4	0	Elu
7	Dense	64	Not applicable	Not applicable	0.5	Elu
8	Dense	32	Not applicable	Not applicable	0	Elu
9	Dense	2	Not applicable	Not applicable	0	Softmax

III. TRAINING METHODS

In the previous section, we have introduced the proposed GW detection framework. In this section, we will detail the training dataset and the training schemes.

A. Training dataset for the denoising model

We construct ten training datasets corresponding to the ten subparameter spaces shown in Table I. For each subparameter space, 10^5 samples are generated for training. We take the GPS time randomly from 1238163456 to 1238806528. For each GPS time, we take the corresponding LIGO data as the background noise. Each sample of the training dataset contains one background noise time series $n(t)$ and one GW signal time series $h(t)$ originating from a BBH system. Both $n(t)$ and $h(t)$ were whitened by the noise ASD, which is estimated by the Weltch method. The time durations of $n(t)$ and $h(t)$ are all 16 s, and the sampling rate is 4096 Hz.

The waveforms are generated with the IMRPhenomD model for the cases where at least one of the black hole's masses is less than $10M_\odot$. Otherwise, we use the SEOBNRv4 model to generate the GW waveforms. In the waveform simulation, we sample the dimensionless spin randomly in (0,0.998). The polarization angle and coalescence phase are sampled randomly in the interval $(0, 2\pi)$. The declination and the right ascension are sampled uniformly over the sphere. We set the cosine of the inclination parameter sampled uniformly. Because no precession and higher modes are included, this simulation parameter configuration does not influence the signal shape.

B. Training scheme for the denoising model

The training scheme for the denoising model is improved compared to our previous work [31]. In [31], the noise and the signal were summed up before training according to the fixed SNR for each data sample. In the current work, for each batch, the noise and the signal were synthesized at the training stage as

$$\boldsymbol{s} = \lambda \times h(t_b^h : t_e^h) + n(t_b^n : t_e^n), \tag{12}$$

where $t_1 : t_2$ means from t_1 to t_2. λ is used to control the SNR of the training sample. $t_e^h - t_b^h = t_e^n - t_b^n$ are used to set the input time duration of the modeled waveform and noise. The background noise is randomly sampled from the whole 16 s length data for every batch of training. The coalescence time is set within 60% to 95% of the time window. In the training stage, the λ is changed to fit a SNR range. The SNR ranges considered include [(17,20), (14,17), (10,14), (5,10), (5,10), (10,14), (14,17)]. The Adam algorithm is applied to optimize the model parameters, the learning rate is set to 1.5×10^{-5}, and the batch number value is 64, for a total of 30 epochs trained.

C. Data for training the astrophysical origin discrimination networks

As detailed in the previous section, the preliminary templates obtained by the denoising models are afterwards selected by the astrophysical origin discrimination network. The discrimination networks are deep convolutional neural networks designed for classification. We construct ten astrophysical discrimination networks corresponding to ten denoising networks.

For each discrimination network, two classes of data are used for training. One is the positive class, and the other is the negative class. The positive class is composed of the data with an astrophysical origin. The negative class is composed of the data without an astrophysical origin. In order to distinguish whether the waveform has an astrophysical origin or not, we employ automatic labeling for the positive class and manual labeling for the negative class.

Labeling the data with a positive astrophysical origin is straightforward. Feeding the denoising model with the strain $s(t) = h(t) + n(t)$, we then obtain the output $\hat{h}(t)$. When the overlap of $h(t)$ and $\hat{h}(t)$ is bigger than a threshold, we label $\hat{h}(t)$ as a sample with an astrophysical origin. In this paper, we define the threshold as 0.8.

Labeling the data without an astrophysical origin is challenging. We use three methods to generate the negative data samples. The three methods are described as follows.

As the first method, we generate the negative samples $n(t)$ with background noises. Feeding the denoising model with the strain $n(t)$, we then get the output $\hat{h}(t)$. We find that some outputs $\hat{h}(t)$ can mimic true waveforms very well. We thus drop these mimickers. We generate 10,000 samples for each model through this method. We denote the dataset composed of these samples as the negative dataset I.

As the second method, we feed the denoising model with the strain $s(t) = h(t) + n(t)$ and obtain the output $\hat{h}(t)$. When the overlap between $h(t)$ and $\hat{h}(t)$ is less than 0.8, we label the $\hat{h}(t)$ as negative class. We generate 10,000 samples for each model through this method. We denote the dataset composed of these samples as the negative dataset II. We believe that the labeling method warrants further investigation in the future. The quality of the training data significantly impacts the classification performance of the discriminator. Although the size of the discriminator's training data is extensive, we manually labeled only 100,000 data points. The objective of classifying unreliable templates as negative is to enhance the differentiation between template and nontemplate shapes. We believe that an overlap threshold of 0.8 might not be the optimal choice. Conducting extensive experiments is necessary to confirm better data annotation methods.

As the third method, we generate the negative samples from selected background noises that trigger the envelope extraction model. The background noises from the first half-month of August 2017 that trigger the envelope extraction

TABLE III. Training datasets for APOD_MODEL_I, APOD_MODEL_II, and APOD_MODEL_III.

Name	Negative dataset I	Negative dataset II	Negative dataset III	Samples in dataset
APOD_MODEL_I	10000	⋯	⋯	20000
APOD_MODEL_II	10000	10000	⋯	40000
APOD_MODEL_III	10000	10000	9080	58160

model are used. We then feed these strains to the denoising models. We select the output as the negative samples. There are 9080 samples generated for each denoising model by this method. We denote the dataset composed of these samples as the negative dataset III.

We train three kinds of astrophysical origin discrimination models, denoted as APOD_MODEL_I, APOD_MODEL_II, and APOD_MODEL_III, respectively. The only difference among these three types of models is the selection of negative samples in the training dataset. The number of positive samples is equal to the number of negative samples. The detailed settings of the training datasets for the three models are shown in Table III.

To label the data without astrophysical origin shape (shape represents whether there are data features of astrophysical origin), as shown in Fig. 3. We employ this program to annotate the data through mouse clicks. The user can judge whether there is a gravitational wave shape through the details of the waveform. The labeled data are automatically saved in an npy format file that can be used for training.

We think the labeling method should be refined in the future. The labeling strategy can influence the performance of the discrimination network. If the discrimination network tends toward wrong judgment of signal-like shapes' waveforms as noise, the true positive probability of the framework will be low. Otherwise, if it tends toward wrong judgment of noise-like shapes as signals, the false positive probability of the framework will be high.

D. Training scheme for the astrophysical origin discrimination model

In order to achieve the purpose of identifying the astrophysical origin signal, we use the astrophysical origin discrimination model described in the previous section to determine whether it is an astrophysical origin waveform. The loss function is the binary cross-entropy, which is used to evaluate the deviation between the predicted values and the actual values in the training set. The gradient descent strategy is Adam. The learning rate is set to 1.5×10^{-5}, and the batch number value is set as 16.

IV. PERFORMANCE OF MSNRnet FRAMEWORK

We have established the MSNRnet framework for GW searches. The framework takes the whitened strain as its input and provides a set with SNRs as output.

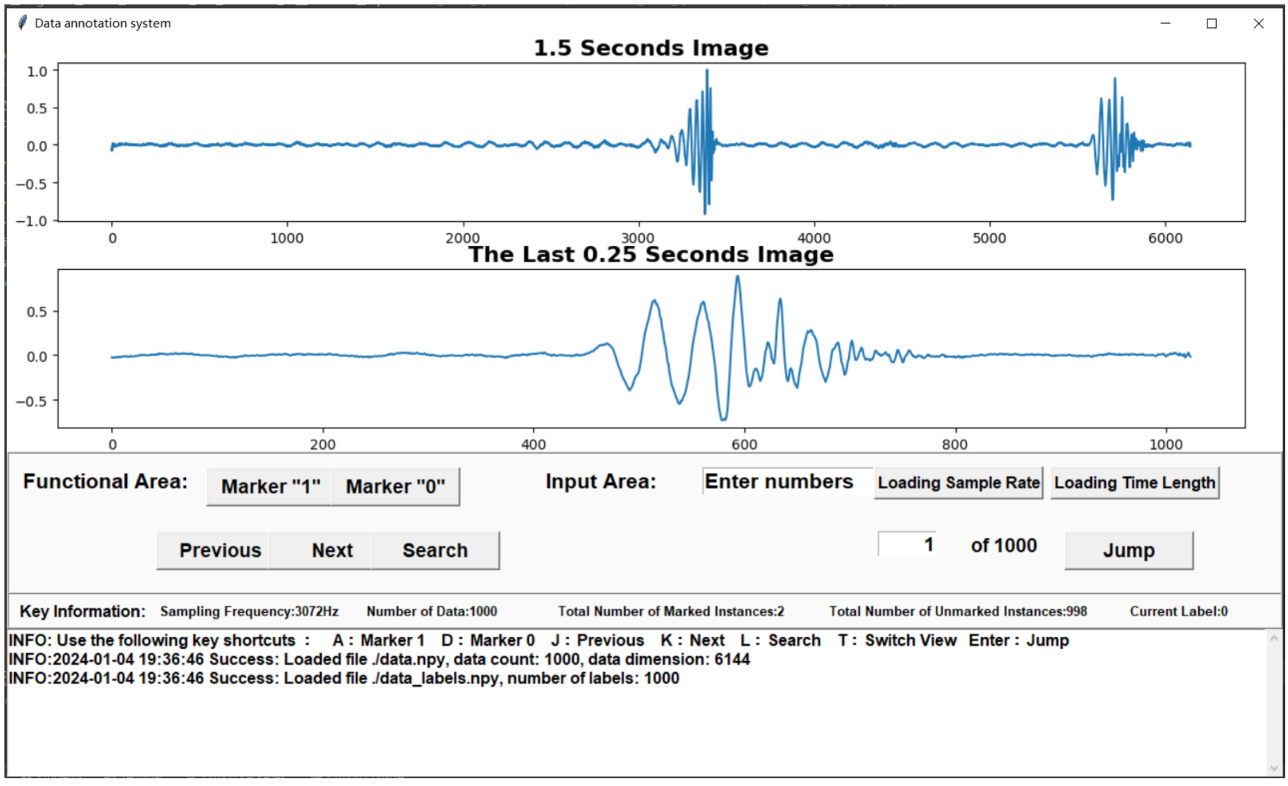

FIG. 3. The graphical interface of the data annotation system.

This framework comprises an envelope extraction model, along with ten denoising models and ten models for discriminating astrophysical origins. We have tested the performance of both the denoising models and the astrophysical discrimination models individually, as well as the overall MSNRnet framework. Both the test dataset and real strain data of Hanford and Livingston interferometers are used. The results convincingly showcase the effectiveness of the newly proposed framework MSNRnet.

A. Results based on the test dataset

In order to test the performance of the MSNRnet framework, we generate ten test datasets. The GPS time of the background noise is sampled from 1238904832 to 1239023616, which is different from the GPS time sampling employed in the training dataset.

1. Performance of the denoising models

In this subsection, we investigate the performance of the denoising model based on the test dataset. We randomly select 16 samples from the test dataset and feed them to the corresponding denoising models. The whitened strain, the buried signal, and the output of the denoising model of each sample are shown in Fig. 4. The input and output lengths of the denoising model vary in accordance with the model's associated parameter space. For a BBH with relatively large masses, the duration of the denoised output is less than two seconds. In such cases, the denoised output's residual portion is zero-padded to match a two-second length. From Fig. 4, we can see that in most scenarios, the denoising model effectively recovers the buried signal.

In order to measure the consistency of the denoised result and the buried signal, the overlap between them is calculated. The overlap between the denoising model's output $\hat{\boldsymbol{h}}$ and the buried signal $\boldsymbol{h}$ can be calculated by

$$o(\boldsymbol{h},\hat{\boldsymbol{h}}) = \frac{\langle \boldsymbol{h},\hat{\boldsymbol{h}}\rangle}{\sqrt{\langle \boldsymbol{h},\boldsymbol{h}\rangle\langle \hat{\boldsymbol{h}},\hat{\boldsymbol{h}}\rangle}}. \tag{13}$$

The signal's duration within the sensitivity band of the detection network varies with the masses of the system. In the case of BBH systems, this duration spans from fractions of a second to several seconds. In contrast to the pioneer works [44,45] for DL-based GW denoising, we take the variation of the duration with the source parameters into consideration, and the detailed design has been discussed in the previous section. Essentially, we use the parameter division strategy. Different denoising models are in charge of different subparameter spaces and are trained with different datasets falling within the corresponding subparameter space. To illustrate the advantage of this parameter division strategy, we establish a denoising model trained with a full parameter dataset. Since a two-second-length whitened signal adequately capturing the characteristics of the signal originates from binary masses between $5M_\odot$ and $80M_\odot$, we set the input duration of the full parameter denoising model to 2 s. The model structure of the full parameter denoising model is the same as the model illustrated previously.

For each sample within every subparameter test dataset, we feed it into two distinct denoising models: the denoising model specific to the corresponding subparameter and the full-parameter denoising model. Subsequently, we obtain the respective outputs. Next, we calculate the overlap between the output and the buried GW signal. The outcome is presented in Fig. 5. The orange color denotes the subparameter denoising model's result, and the blue color denotes the full-parameter denoising model's result.

Figure 5 also shows that the average overlap decreases when the mass of the BBH becomes smaller. This fact can be understood as follows: For the low-mass binary system case, the duration of the signal will be long. The waveform features of these systems admit long-range dependencies. The U-Net employed in the current study is based on convolutional neural networks (CNNs). However, CNN-based models exhibit a limitation in modeling long-range dependencies. Convolution operators are restricted to local receptive fields. Only after traversing multiple convolutional layers does perceiving long-range dependencies become possible.

Differently from CNNs, the WaveNet tackles the limitation of local receptive fields by employing deeper models and using dilated convolutional layers [44]. In order to avoid the huge neuron number of a denoising model which is too computationally expensive to us, we insist on using CNN rather than WaveNet in the current work. Since the denoising models constitute just one component of our comprehensive framework MSNRnet, it is straightforward to replace WaveNet [44], LSTM [45], Waveformer [46], and other possibilities with CNN within the framework MSNRnet. Moreover, the well-behaved parameter division strategy, as shown in Fig. 5, can also be applied to these possible denoising models, including WaveNet, LSTM, Waveformer, and others.

2. Performance of the matched SNR predictions

We thoroughly examine the precision of matched SNR predictions by our framework MSNRnet in this subsection. We feed samples in the test set into the denoising model and then compute the matched filtering SNRs using the denoising output as a waveform template. A comparison between the SNRs calculated by our framework and the actual SNRs is illustrated in Fig. 6. For most cases, the SNRs predicted by our framework MSNRnet are highly consistent with the actual SNRs. Several panels contain low-SNR events that are predicted with significantly higher SNRs. We guess that the predicted template may contain features of the noise. The results in Fig. 6 and Table V further support the opinion that the proposed method potentially introduces a bias toward

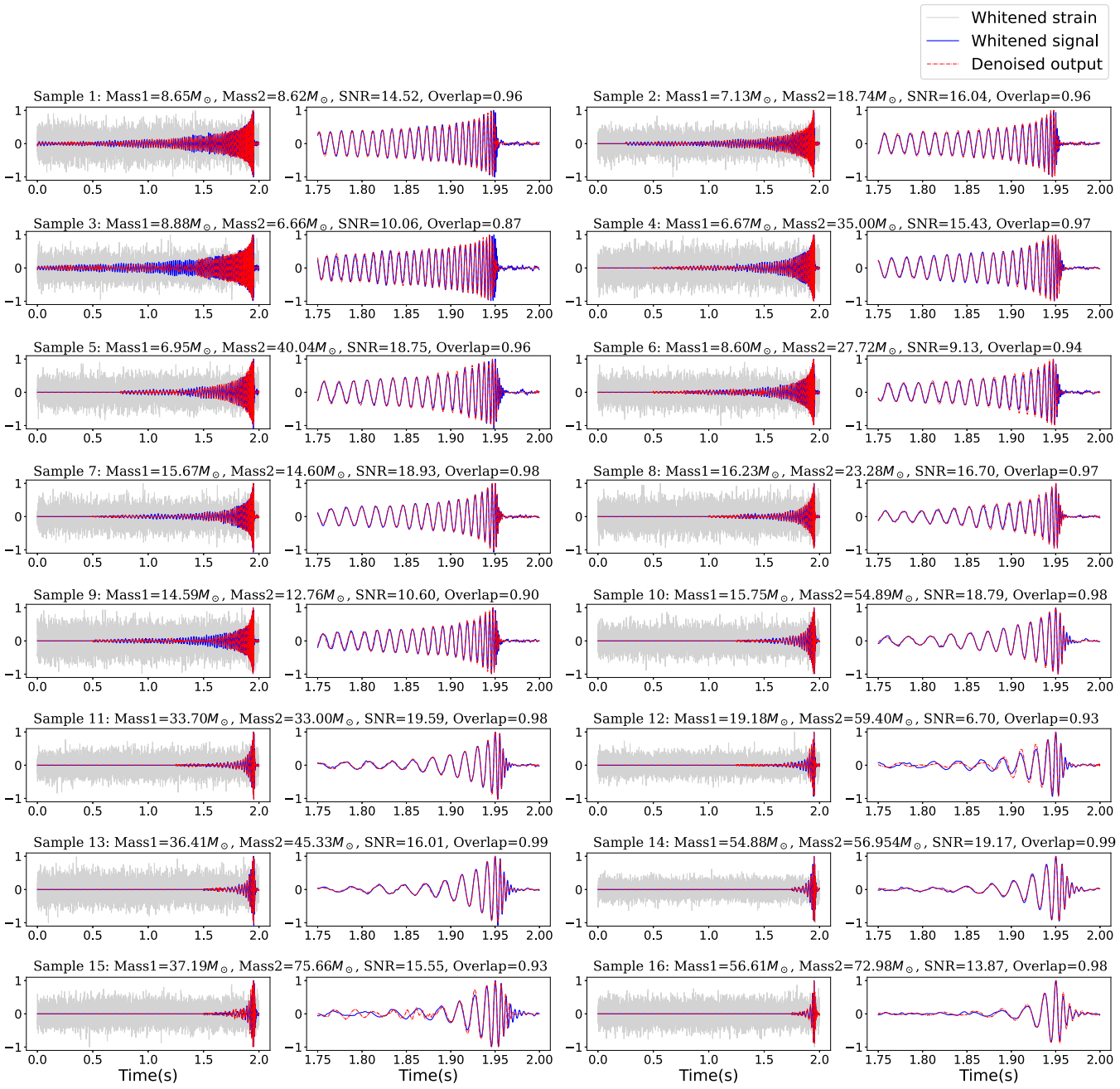

FIG. 4. The whitened strain, the whitened signal, and the denoised output of 16 randomly selected samples in the test dataset. The left subplot of each sample shows the whole 2 s time duration. The right subplot of each sample is the enlargement of the merger part including a 0.25 s time duration. The right subplot includes only the whitened signal and the denoised output for clear comparison. Note that we use the knowledge of the parameters of the injected signal and the corresponding denoising model.

higher SNRs during the derivation of the "template" from the data. Correspondingly, we show the distribution of the relative error of the predicted SNR in Fig. 7. The relative error is defined as

$$\text{relative\_error} \equiv \frac{SNR_{pred} - SNR_{\text{true}}}{|SNR_{\text{true}}|}, \tag{14}$$

where SNR_{pred} is the SNR predicted by the proposed framework and SNR_{true} is the matched-filter SNR of the strain and the buried signal. We can see that the relative error is less than 10% for most samples.

After checking the "noise + signal" samples, we move to pure "noise" samples. We feed the background noise from the ten test sets into the denoising models and compute the matched SNR of the background noise using the denoising model's output as a waveform template. Each of the test sets encompasses 10,000 samples. The distributions of matched-filtering SNRs for the denoising output are depicted in Fig. 8.

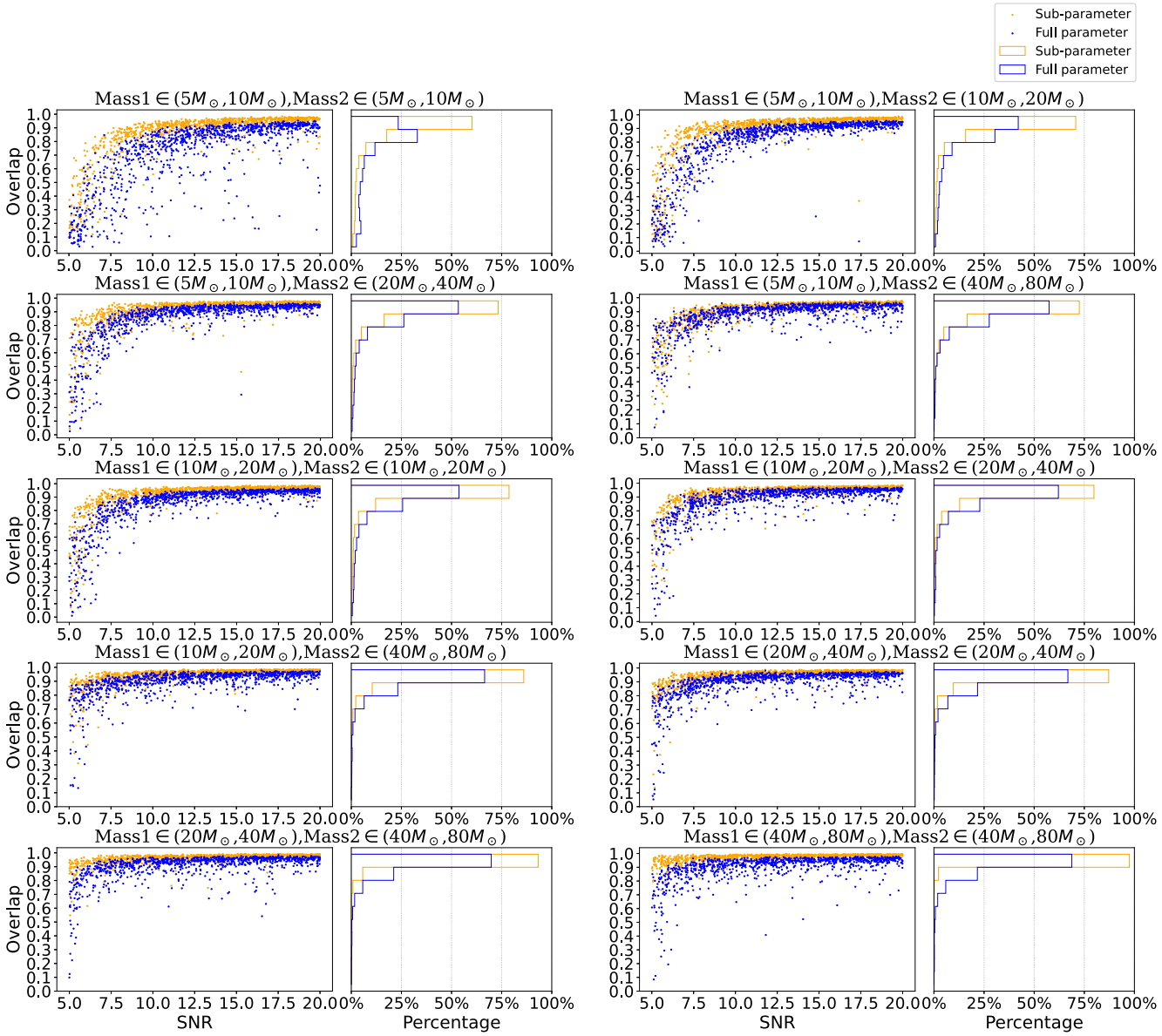

FIG. 5. The overlap between the outputs of the denoising models and the buried GW signals. The denoising models trained by the subparameter space training data (orange) and full parameter training data (blue) are compared. In order to avoid mutual coverage of a large number of points, only 1000 points of each model are shown in the scatter plot. All of the 10,000 samples are used to plot the histogram in the right panel of each subplot. Note that we use the knowledge of the parameters of the injected signal and the corresponding denoising model.

From Fig. 8, we can see that most predicted SNRs by models for larger than $40M_\odot$ are less than 6. But a significant amount of predicted SNRs by models for smaller than $40M_\odot$ fall within the range (10,12). This fact indicates that the denoising model may result in a significant number of fake signals. This is the major motivation for us to introduce the discrimination networks to judge if the denoising output admits an astrophysical origin or not. In the subsequent subsection, we will delve into the performance of the astrophysical origin discrimination networks in the test set.

3. *Performance of the astrophysical origin discrimination networks*

From the last subsection, we have learned that the denoising output may result in a significant number of fake signals. To solve this problem, we introduce an astrophysical origin discrimination network. As described in Sec. III C, we have constructed three kinds of astrophysical discrimination networks (APOD_MODEL_I, APOD_MODEL_II, and APOD_MODEL_III) for each subparameter space based on the training datasets. In this subsection, we will investigate

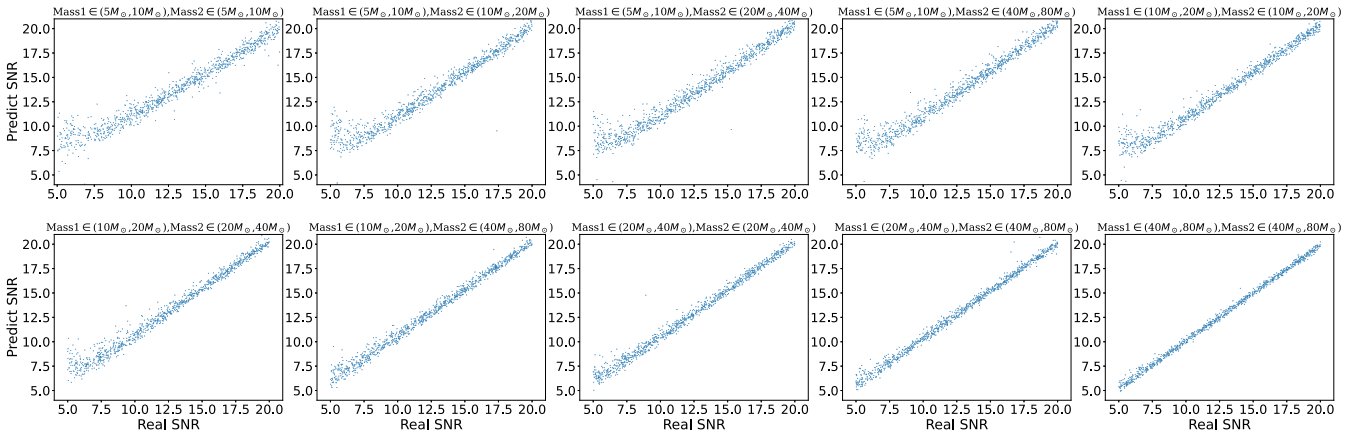

FIG. 6. The SNRs of the "noise + signal" samples. Vertical coordinates indicate the SNRs calculated by the denoising output. Horizontal coordinates indicate the actual SNRs calculated by the injected signal. To avoid mutual coverage of a large number of points, each subfigure only shows a randomly selected set of 1000 points. Note that we use the knowledge of the parameters of the injected signal and the corresponding denoising model.

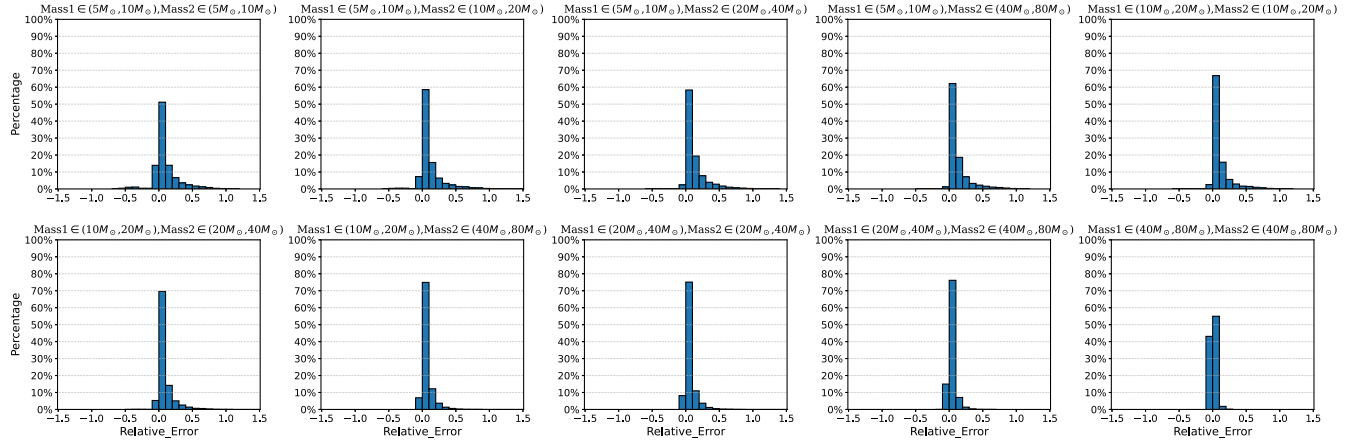

FIG. 7. The distribution of the relative error range of the predicted SNR. This figure corresponds to Fig. 6. The bin size 0.1 is used in this figure. Note that we use the knowledge of the parameters of the injected signal and the corresponding denoising model.

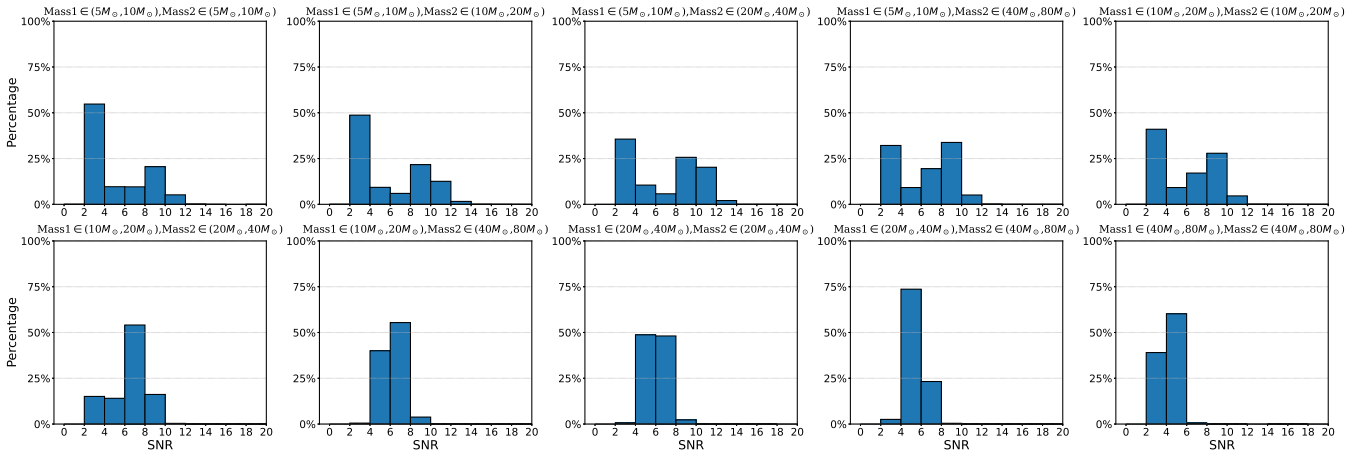

FIG. 8. Distribution of the predicted matched-filtering SNR for pure "noise" samples and the corresponding "noise."

TABLE IV. The percentage of samples classified correctly by the astrophysical discrimination network. I, II, and III correspond to APOD_MODEL_I, APOD_MODEL_II, and APOD_MODEL_III respectively.

		Positive class			Negative class		
Mass1 ($M_\odot$)	Mass2 ($M_\odot$)	I	II	III	I	II	III
[5, 10]	[5, 10]	98.70%	92.50%	93.30%	69.70%	95.90%	93.30%
[5, 10]	[10, 20]	98.90%	95.60%	94.80%	64.90%	85.70%	86.60%
[5, 10]	[20, 40]	95.40%	90.60%	92.30%	68.50%	91.10%	90.70%
[5, 10]	[40, 80]	93.40%	88.30%	90.70%	50.80%	90.50%	89.90%
[10, 20]	[10, 20]	97.70%	92.00%	94.70%	72.90%	94.00%	92.00%
[10, 20]	[20, 40]	97.30%	90.60%	91.00%	62.50%	94.40%	95.00%
[10, 20]	[40, 80]	97.50%	91.60%	90.60%	71.00%	92.70%	93.10%
[20, 40]	[20, 40]	97.90%	94.20%	94.00%	71.30%	96.10%	96.20%
[20, 40]	[40, 80]	95.90%	91.80%	93.00%	83.10%	95.50%	96.70%
[40, 80]	[40, 80]	95.60%	90.40%	89.40%	85.50%	96.60%	97.50%

the performance of the astrophysical origin discrimination networks.

We randomly chose 20,000 samples from the test dataset which is constructed though the same procedure as the training dataset to test the astrophysical discrimination network. Corresponding to the ten groups of subparameter space, each subparameter space has 2000 samples (1000 positive and 1000 negative samples). The performances of the astrophysical discrimination network on these samples are shown in Table IV. APOD_MODEL_I, APOD_MODEL_II, and APOD_MODEL_III show similar behavior. As we expected, if the true alarm rate is better, then the false alarm rate is worse. Roughly, APOD_MODEL_I admits a higher true alarm rate than APOD_MODEL_II and APOD_MODEL_III.

As Table IV shows, the discrimination network can really recognize fake waveforms. But it may also mistake some true signals as fake ones. Hopefully, this situation can be eliminated if only the signal is strong. To check this fact, we feed the denoising outputs into astrophysical origin discrimination networks and investigate the corresponding classification outcomes. A statistical analysis of the classification results is shown in Fig. 9. We can see that in the low-SNR range case, APOD_MODEL_I and APOD_MODEL_II outperform APOD_MODEL_III. In the high-SNR case, the three kinds of astrophysical origin discrimination models perform similarly. Even for APOD_MODEL_III, a significant majority of the denoising outputs can be correctly classified when the buried signal's $SNR \geq 8$.

4. Performance of the combination of PTP, TS, and MF stages

In this subsection, we will investigate the performance of the combination of PTP, TS, and MF stages through the test dataset. The test set is composed of samples with relatively short durations. Noting that the "noise" samples often fail to trigger the STP stage, we do not investigate this specific stage in this section. The comprehensive framework including STP will undergo testing using the confident events and half-month detected strain data detailed in Secs. IV B and IV C.

Similarly to various other classification tasks, we use receiver operator characteristic (ROC) curves to evaluate the performance of the search methods. The ROC curve reflects the relationship between the true alarm probability (TAP) and false alarm probability (FAP). The TAP and FAP are defined as follows:

$$TAP := \frac{TP}{TP + FN}, \tag{15}$$

$$FAP := \frac{FP}{FP + TN}. \tag{16}$$

In this context, TP represents the number of true positive samples, FP signifies the number of false positive samples, TN corresponds to the number of true negative samples, and FN denotes the number of false negative samples.

In the proposed framework, two comparable metrics are used. One is the output of the astrophysical discrimination network ($\boldsymbol{o}_{\text{astro}}$), and the other is the output of the matched filtering process (SNR). For simplification, we set the threshold of $\boldsymbol{o}_{\text{astro}}$ to 0.5. In the case of $\boldsymbol{o}_{\text{astro}} < 0.5$, the proposed framework categorizes the strain as noise. Conversely, when $\boldsymbol{o}_{\text{astro}} > 0.5$, the MF stage will be further conducted, and the confidence of the strain-containing signal is evaluated using SNR. When the SNR surpasses a predefined threshold, a GW trigger will generate. It sweeps the SNR threshold from 5 to 20, recording the TAP and FAP for each threshold value to produce the ROC curve of the proposed framework. It is important to acknowledge that due to the fixed threshold value of $\boldsymbol{o}_{\text{astro}}$, certain "noise + signal" samples might not surpass this threshold, thereby preventing TAP from reaching 100% in such cases.

Figure 10 shows the ROC curves of the combination of PTP, TS, and MF stages. For comparison, the results of the combination of PTP and TS stages and that of the

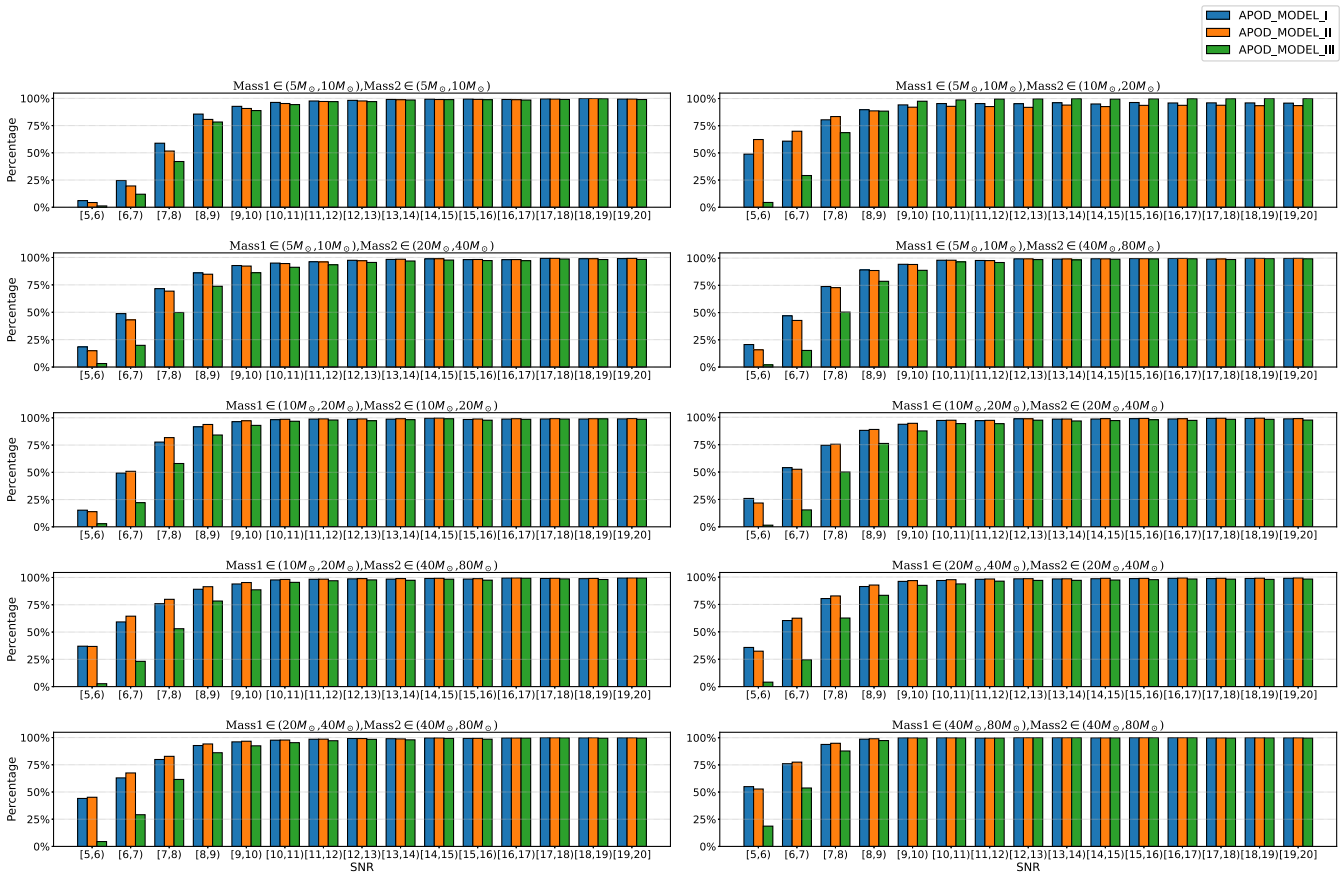

FIG. 9. The percentage of correctly classified results by the astrophysical discrimination network for the denoised results of "signal + noise" input. Note that we use the knowledge of the parameters of the injected signal and the corresponding denoising model.

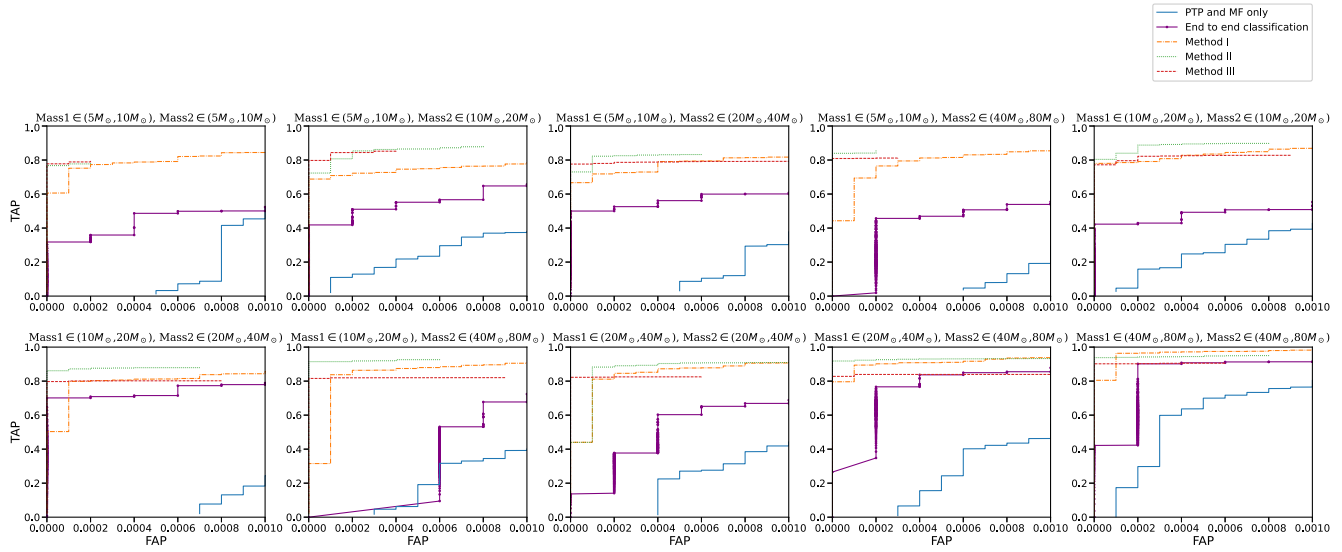

FIG. 10. The ROC curves of the combination of PTP, TS, and MF stages; the combination of PTP and TS stages and the end-to-end DL method. Methods I, II, and III correspond to the three different astrophysical origin discrimination models. Note that we use the knowledge of the parameters of the injected signal and the corresponding denoising model.

end-to-end DL method are also shown together. Each plot corresponds to a specific subparameter space. The classification model takes the whitened strain as input and generates an output in the form of P_{score}, which falls within the range of 0 to 1.

We conduct three methods (Methods I, II, and III) based on the PTP, TS, and MF stages. The disparity among these three methods lies in the choice of astrophysical origin discrimination networks during the TS stage. Specifically, we employ APOD_MODEL_I for Method I, APOD_MODEL_II for Method II, and APOD_MODEL_III for Method III.

For real data, most of the strain's GW signals are often indiscernible. Real data are consequently assumed to consist mainly of background noise. Despite a relatively low FAP, numerous false triggers can still emerge for a searching method, especially in scenarios involving long-duration detection. Therefore, we only consider the low FAP range here. Figure 10 displays the FAP range below 10^{-3}.

From Fig. 10, we can see clearly the improvement of the combination of the PTP, TS, and MF stages. Specifically, when FAP $= 10^{-4}$, for the combination of PTP, TS, and MF stages, the TAPs are consistently around 0.8. On the contrary, for the combination without the TS stage, the TAPs are almost 0. This is consistent with Fig. 8 and proves once again the importance of the astrophysical origin discrimination model. The output of the denoising models may result in fake signals, which result in a large SNR.

Interestingly, the FAP of the end-to-end DL method can be effectively brought down to the magnitude of 10^{-4}, as shown in Fig. 10. The results in [30] and [28] all show that the FAP of such a model can hardly reach 10^{-4}. The difference presented here is that we have considered the subparameter space division strategy. This suggests that an end-to-end DL method aided by a subparameter space division strategy has the potential to significantly enhance search performance. Within the context of a subparameter GW search, the ability to optimize the duration of the time window for data analysis becomes feasible. This shows the potential of the subparameter end-to-end classification model for GW search, which we believe is a promising avenue for further investigation in the future.

B. Results for confident events

In this subsection, we investigate the detection performance of the proposed MSNRnet framework on the confident events listed in the GWTC-1, GWTC-2, GWTC-2.1, and GWTC-3 catalogs. We first investigate the denoising performance related to the confident events. Subsequently, those events that work well with both Hanford and Livingston interferometers are studied by all four stages (STP, PTP, TS, and MF) of the proposed framework.

Here, we analyze the denoising outcomes in relation to strains containing confident events. To begin, we extract the envelope (in the STP stage detailed in Sec. II) of the whitened strain. Subsequently, we cut T_i seconds whitened strain to the ith denoising model and get the output. Note that the envelope's peak time corresponds to the final 0.05 s of the T_i seconds data. We select 16 samples of the confident events from the Hanford interferometer. Our investigation encompasses a wide range of masses for the detected binary black hole (BBH) mergers.

The comparison between the denoising outputs and the optimal templates (sampled from the posterior distribution provided by GWTC-2.1 and GWTC-3) is depicted in Fig. 11. Figure 11 illustrates that our denoising model effectively retrieves the signal from the confident events. Intriguingly, our denoising model successfully reconstructs the GW170608 signal detected by the Hanford interferometer. The overlap between the denoising output and the template is notably high, measuring 0.95 at a two-second timescale and reaching an impressive 0.99 at a 0.25-second timescale. Notably, the denoising output of GW170608 outperforms the WaveNet denoising result, which achieved an overlap of 0.73 according to previous findings [44].

In order to comprehensively evaluate the effectiveness of our MSNRnet framework on real data, we feed the whitened strains of the confident events into the framework encompassing the STP, PTP, TS, and MF stages. For the TS stage, we employ three distinct models: APOD_MODEL_I, APOD_MODEL_II, and APOD_MODEL_III, corresponding to Methods I, II, and III, respectively, as described in the above subsections.

Table V shows the SNRs predicted by our MSNRnet framework (Methods I, II, and III) alongside those generated by five widely used GW search pipelines (cWB, GstLAL, MBTA, PyCBC, and PyCBC_BBH). Since GWTC-1 does not include results of MBTA or PyCBC_BBH, the ten O1 and O2 BBH events' network SNRs predicted by the two pipelines are missing.

Once the strains from both the Hanford and Livingston interferometers within the same subparameter space (denoising and astrophysical origin discrimination) successfully pass the STP, PTP, and TS stages, the MF stage is subsequently engaged to compute the network SNR of that subparameter space. For each method, the outcome takes the form of a network SNR set.

Every SNR in the set corresponds to the SNR associated with a distinct subparameter space. The final network SNR is calculated by

$$NetworkSNR = \max_i NetworkSNR_i, \tag{17}$$

where $NetworkSNR_i$ is the predicted network SNR, and

$$NetworkSNR_i = \sqrt{SNR_{i,H}^2 + SNR_{i,L}^2}, \tag{18}$$

where $SNR_{i,H}$ and $SNR_{i,L}$ are the predicted SNR values for the ith subparameter spaces in Hanford and Livingston, respectively.

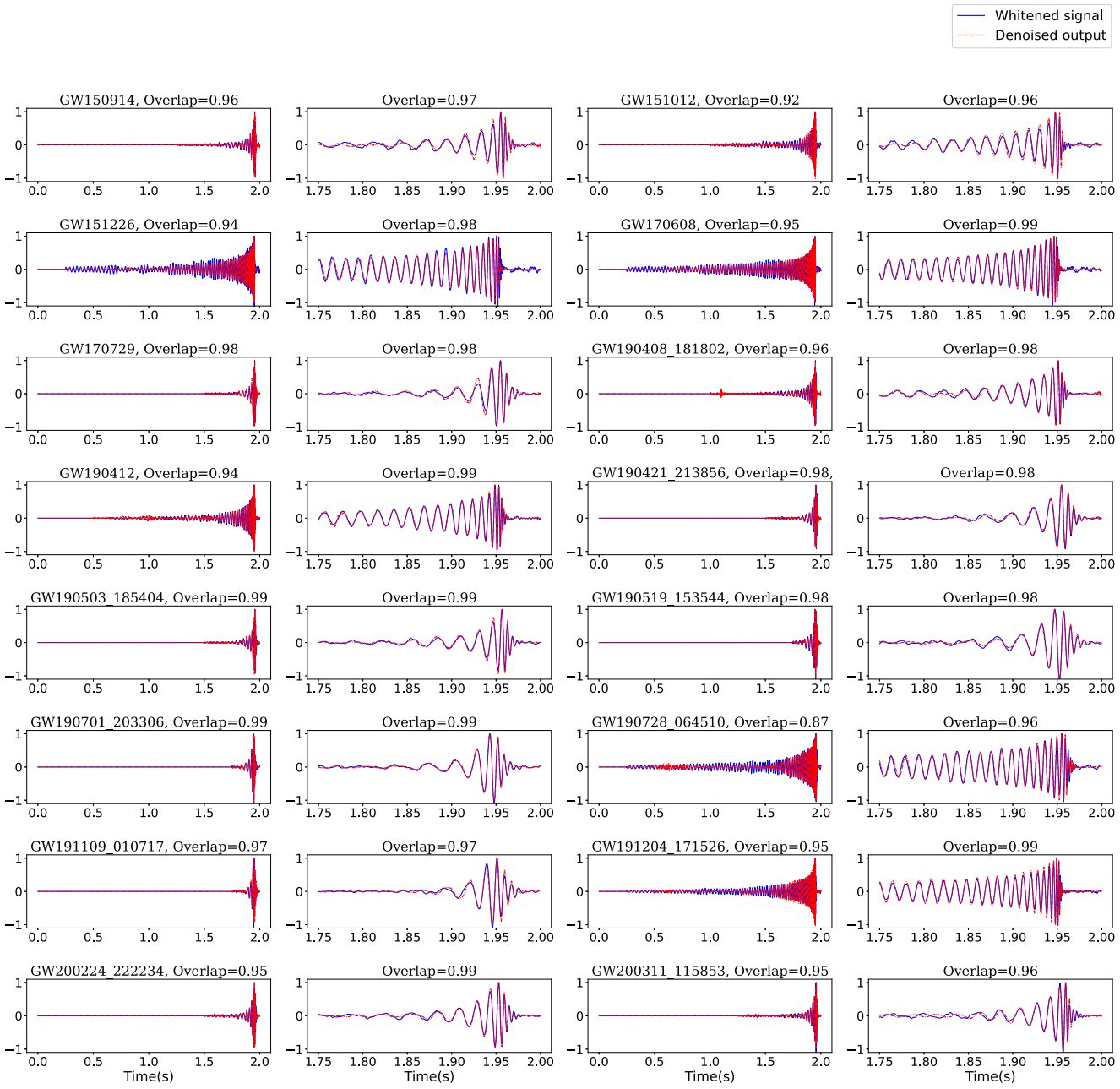

FIG. 11. Comparison between the denoised waveform from real data and the expected waveform from LVK reports.

In O1 and O2, nine confident events are successfully identified using both Method I and Method II. However, owing to the weakness of the signal at the Hanford interferometer, all three methods fail to detect GW170818. Method III also fails to successfully detect GW151012, GW151226, and GW170823. Method I achieves a success rate of approximately 71% in event detection, followed by Method II at 60%, and Method III at 41%. Method I has a higher correct detection probability than MBTA and PyCBC. Note that this does not mean that the superior of Method I to MBTA and PyCBC pipelines, because the detection benchmarks of our proposed framework and traditional MF-based pipelines are different. Conventional MF-based pipelines employ p_{astro} as their detection benchmark, while we utilize the output of the astrophysical origin discrimination network, $\boldsymbol{o}_{\text{astro}}$, for our detection benchmark.

In the case of Method II, we also examine the merge-time differences in denoised strains between the Hanford and Livingston interferometers. The merge-time difference is defined by

$$diff t = |\arg\max_t \hat{h}_L(t) - \arg\max_t \hat{h}_H(t)|, \qquad (19)$$

where $\hat{h}_H(t)$ is the denoised strain of Hanford, and $\hat{h}_L(t)$ is the denoised strain of Livingston. The results (presented in the final column of Table V) show that the merge-time

TABLE V. SNRs of the GW events reported by LVK in O1, O2, and O3 [1–3]. Events that work well with both Hanford and Livingston interferometers are listed. The SNRs predicted by the pipelines cWB, GstLAL, MBTA, PyCBC, and PyCBC_BBH are shown. Here we compare SNRs predicted by our MSNRnet framework to these reported results. Because three kinds of astrophysical origin discrimination networks are constructed, we use three GW search methods (Method I, Method II, and Method III). Methods I, II, and III use APOD_MODEL_I, APOD_MODEL_II, and APOD_MODEL_III for the astrophysical origin discrimination network, respectively. Note that some events' FAR or p_{astro} are not in the confident event range, and in this case, the SNR value is not shown. The time differences between the peak values of the denoised strain (Method II) of Hanford and Livingston interferometers are also shown. Note that these results are obtained through a blind search (the corresponding parameter space or signal length not being known).

Name	cWB	GstLAL	MBTA	PyCBC	PyCBC BBH	Method I	Method II	Method III	Time differences of Method II(s)
GW150914	25.2	24.4	···	23.6	···	24.5	24.5	24.4	0.00439
GW151012	···	10.0	···	9.5	···	13.4	10.6	···	0.00293
GW151226	11.39	13.1	···	13.1	···	14.8	13.7	···	0.00269
GW170104	13.0	13.0	···	13.0	···	14.4	14.4	14.2	0.00024
GW170608	14.1	14.9	···	15.4	···	17.0	16.9	16.9	0.00024
GW170729	10.2	10.8	···	9.8	···	10.9	11.6	10.8	0.00586
GW170809	···	12.4	···	12.2	···	14.7	13.4	13.4	0.01220
GW170814	17.2	15.9	···	16.3	···	17.1	17.1	17.1	0.01220
GW170818	···	11.3	···	···	···	···	···	···	···
GW170823	10.8	11.5	···	11.1	···	12.7	12.7	···	0.00537
GW190403-051519	···	···	···	16.3	8.0	···	···	···	···
GW190408-181802	14.8	14.7	14.4	13.1	13.7	15.6	15.6	15.6	0.00659
GW190412	19.7	19.0	18.2	17.4	17.9	20.0	20.0	19.5	0.00122
GW190413-052954	···	···	···	···	8.5	9.1	9.1	···	0.00049
GW190413-134308	···	···	10.3	···	8.9	10.3	9.9	···	0.00098
GW190421-213856	9.3	10.5	9.7	10.1	10.1	11.2	11.2	10.2	0.00171
GW190426-190642	···	···	···	···	9.6	···	···	···	···
GW190503-185404	11.5	12.0	12.8	12.2	12.2	14.9	14.9	12.4	0.00610
GW190512-180714	10.7	12.2	11.7	12.4	12.4	13.4	···	···	···
GW190513-205428	···	12.3	13.0	···	11.8	14.1	14.1	13.0	0.0000
GW190514-065416	···	···	···	···	8.4	9.1	···	···	···
GW190517-055101	10.7	10.8	11.3	10.4	10.3	10.9	11.1	10.9	0.00342
GW190519-153544	14.0	12.4	13.7	13.2	13.2	14.8	15.1	14.8	0.00269
GW190521	14.4	13.3	13.0	13.7	13.6	···	···	···	···
GW190521-074359	24.7	24.4	22.2	24.0	24.0	25.0	24.4	24.4	0.00098
GW190527-092055	···	8.7	···	···	···	8.8	···	···	···
GW190602-175927	11.1	12.3	12.6	11.9	11.9	12.0	12.8	11.9	0.01293
GW190701-203306	10.2	11.7	11.3	11.9	11.7	10.7	10.7	10.4	0.00513
GW190706-222641	12.7	12.5	11.9	11.7	12.6	13.6	13.6	12.7	0.00488
GW190707-093326	···	13.2	12.6	13.0	13.0	14.9	14.2	14.2	0.00708
GW190719-215514	···	···	···	···	8.0	10.1	10.1	···	0.00269
GW190720-000836	···	11.5	11.6	10.6	11.4	···	···	···	···
GW190725-174728	···	···	9.8	9.1	8.8	12.2	···	···	···
GW190727-060333	11.4	12.1	12.0	11.4	11.1	13.6	12.6	12.0	0.00537
GW190728-064510	···	13.4	13.1	13.0	13.0	14.9	14.9	14.9	0.00220
GW190731-140936	···	8.5	9.1	···	7.8	9.5	8.6	···	0.01050
GW190803-022701	···	9.1	9.0	···	8.7	10.1	9.6	···	0.00220
GW190805-211137	···	···	···	···	8.3	8.8	···	···	···
GW190814	···	22.2	20.4	19.5	···	23.7	23.4	23.4	0.00806
GW190828-063405	16.6	16.3	15.2	13.9	15.9	17.3	17.3	17.3	0.00317
GW190828-065509	···	11.1	10.8	10.5	10.5	12.3	11.3	···	0.00024
GW190915-235702	12.3	13.0	12.7	13.0	13.1	14.1	13.6	13.6	0.00342
GW190916-200658	···	···	8.2	···	7.9	9.2	9.2	···	0.00513
GW190917-114630	···	9.5	···	···	···	···	···	···	···
GW190924-021846	···	13.0	11.9	12.4	12.5	···	···	···	···
GW190926-050336	···	9.0	···	···	···	9.0	9.0	···	0.00342
GW190929-012149	···	10.1	10.3	···	···	···	···	···	···
GW190930-133541	···	10.1	10.0	9.8	10.0	···	···	···	···

(Table continued)

TABLE V. *(Continued)*

Name	cWB	GstLAL	MBTA	PyCBC	PyCBC BBH	Method I	Method II	Method III	Time differences of Method II(s)
GW191103-012549	···	···	···	9.3	9.3	···	···	···	···
GW191105-143521	···	···	10.7	9.8	9.8	11.9	···	···	···
GW191109-010717	15.6	15.8	15.2	13.2	14.4	16.6	15.8	15.8	0.00293
GW191113-071753	···	···	9.2	···	···	···	···	···	···
GW191126-115259	···	···	···	···	8.5	···	···	···	···
GW191127-050227	···	10.3	9.8	···	8.7	9.5	9.5	···	0.00708
GW191129-134029	···	13.3	12.7	12.9	12.9	···	···	···	···
GW191204-110529	···	···	···	···	8.9	10.5	10.5	···	0.00244
GW191204-171526	17.1	15.6	17.1	16.9	16.9	17.3	17.3	17.3	0.00171
GW191215-223052	9.8	10.9	10.8	10.3	10.2	10.8	11.3	···	0.00635
GW191219-163120	···	···	···	8.9	···	···	···	···	···
GW191222-033537	11.1	12.0	10.8	11.5	11.5	14.0	11.8	11.8	0.00366
GW191230-180458	10.3	10.3	···	···	9.9	12.1	10.5	9.9	0.00195
GW200115-042309	···	11.5	11.2	10.8	···	···	···	···	···
GW200128-022011	8.8	10.1	9.4	9.8	9.9	10.1	10.1	9.8	0.00317
GW200129-065458	···	26.5	···	16.3	16.2	27.3	26.9	26.5	0.00269
GW200202-154313	···	11.3	···	···	10.8	···	···	···	···
GW200208-130117	···	10.7	10.4	9.6	10.8	10.6	···	···	···
GW200208-222617	···	···	···	···	7.9	···	···	···	···
GW200209-085452	···	10.0	9.7	···	9.2	9.5	···	···	···
GW200210-092254	···	9.5	···	8.9	8.9	···	···	···	···
GW200216-220804	···	9.4	···	···	8.7	8.8	8.8	8.8	0.01221
GW200219-094415	9.7	10.7	10.6	9.9	10.0	12.7	11.6	···	0.00122
GW200220-061928	···	···	···	···	7.5	···	···	···	···
GW200220-124850	···	···	8.2	···	···	10.2	···	···	···
GW200224-222234	18.8	18.9	19.0	19.2	18.6	19.6	19.6	18.3	0.00098
GW200225-060421	13.1	12.9	12.5	12.3	12.3	14.5	14.5	14.5	0.00684
GW200306-093714	···	···	8.5	···	···	···	···	···	···
GW200308-173609	···	···	···	···	8.0	···	···	···	···
GW200311-115853	16.2	17.7	16.5	17.0	17.4	18.4	16.7	16.6	0.00342
GW200316-215756	···	10.1	···	9.3	9.3	···	···	···	···
GW200322-091133	···	···	9.0	···	9.6	···	···	···	···
Detection (%)	40.0	73.8	60.0	62.5	73.8	71.3	60.0	41.3	

differences of all identified event mergers are within 13 ms. Similarly to the MF-based approach, the time differences across multiple interferometers can be employed to further validate the results.

Method II shows that about 40% of a successfully detected event's loudest SNR is actually obtained by a predicted template that comes from the "correct" denoising model (from a denoising network trained on signals with masses the same as the injection). We think that the loudest SNR denoising model cannot be used for the mass range parameter estimation. However, due to the generalization ability of neural networks, the denoised shape may be used for parameter estimation in the future.

C. Results for GW search based on continued half-month data

The previous experiments indicate the effectiveness of the proposed framework for GW search on test datasets and confident events in GWTC-1, GWTC-2.0, GWTC-2.1, and GWTC-3. Nevertheless, these previous findings may not adequately capture the framework's performance in long-time-duration detection scenarios. In this subsection, we check the framework's ability to detect strains spanning a half-month duration in August 2017.

The data were used in the previous research [28,31]. Because a section of the first half of the August 2017 strain was employed for training the astrophysical origin discrimination network, we only test the second half of the strain, and the GPS times of the strain fall between 1186683883 and 1187733597. To prevent redundancy from earlier sections, our focus here is solely on examining the "noise" background's response. Consequently, the confident events that occurred in August 2017 are not subjected to investigation within this subsection.

Table VI presents the experiment results of false triggers of the half-month strain. We investigated the variation of the FAR concerning the SNR threshold through the

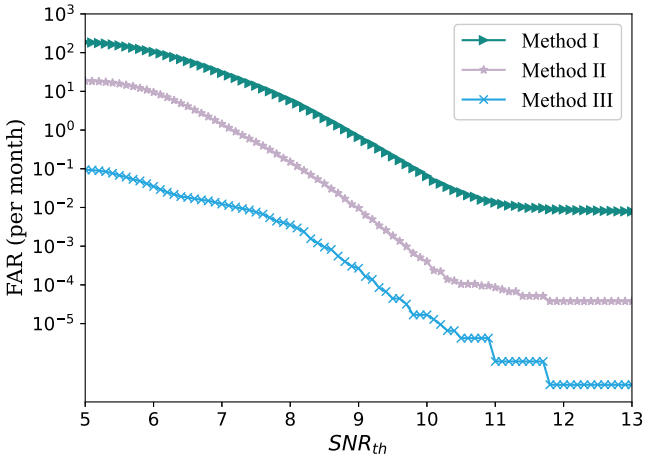

FIG. 12. The variation of SNR_{th} to the FAR of Methods I, II, and III.

half-month data analysis. Because the sliding window step of the STP stage is two seconds, the FAR can be roughly predicted by $\frac{FAP}{2}$ per second. The FAP of interferometer X (FAP_X) varies with the SNR threshold (SNR_{th}), and FAP_X can be estimated by

$$FAP_X(SNR_{th}) = \frac{N(SNR > SNR_{th})}{N_{\text{det}}}, \tag{20}$$

where N_{det} denotes the detection number, which can be evaluated by $\frac{T}{2}$, where T denotes the whole time duration (in this work, we analyzed about nine days). Considering that we use the coincidence strategy in the TS stage and supposing $FAP_H = FAP_L$, the FAR of the detector network containing two interferometers (Hanford and Livingston) can be calculated by

$$\begin{aligned} FAR(SNR_{th}) &= \frac{FAP(SNR_{th})}{2} \\ &= \alpha \times \frac{FAP_X^2(SNR_{th})}{2}, \end{aligned} \tag{21}$$

where α denotes the probability of peak values of $\hat{\boldsymbol{h}}_H$ and $\hat{\boldsymbol{h}}_L$ below 15 ms. We suppose that for the background-noise-only case, the peak values difference of $\hat{\boldsymbol{h}}_H$ and $\hat{\boldsymbol{h}}_L$ is uniformly distributed. Because most of the false triggers are generated by the analysis with a time window greater than 0.5 s in length, we set $\alpha \approx \frac{0.015}{0.5} = 0.03$. We calculate the variation of SNR_{th} to the FAR of Methods I, II, and III. The relationships are plotted in Fig. 12. From the figure, it is evident that the upper limits of the FARs for Methods I, II, and III are about 185, 18, and 0.1 per month, respectively. When considering the SNR threshold, the lower limits of the FARs for Methods I, II, and III are about 0.1, 10^{-3}, and 10^{-6} per month.

Certain false triggers generated by our proposed framework warrant further investigation. We plot the information of the two false triggers generated by Method II in Fig. 13. For easier comparison, the whitened strain is rescaled by 1/50. These two false triggers are characterized by GPS times approximately close to 1187014846 and 1187612152. We checked and found that the GPS times do not overlap with any of the subthreshold triggers in the LIGO-Virgo Catalogues or Open Gravitational Wave Catalogues. Denoised results of the two false triggers exhibit chirplike shapes. The SNRs calculated through

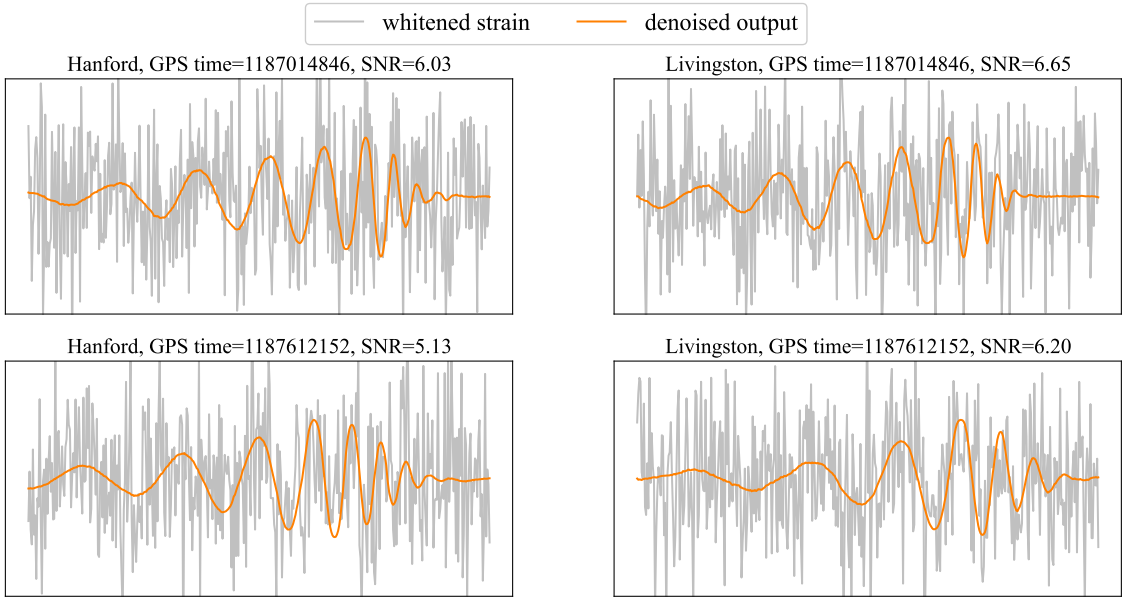

FIG. 13. Whitened strain and denoised outputs near the two false triggers of Method II.

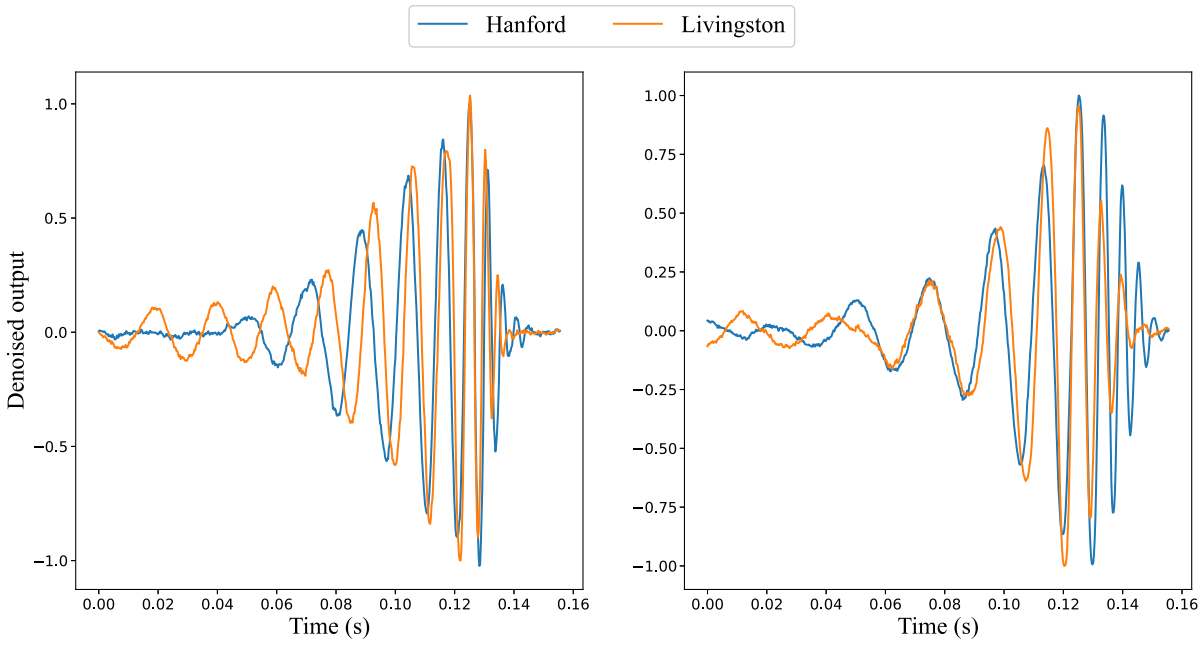

FIG. 14. Comparation of the two false triggers' denoised output of Hanford and Livingston, where the waveform of Livingston has been flipped and time-shifted due to the detector alignment relation between Hanford and Livingston. Left: The denoised outputs near GPS time 1187014846. Right: The denoised outputs near GPS time 1187612152.

matched filtering between the denoised results and the whitened strain are all above 5.

Displayed in Fig. 14 is a comparison between the denoising outputs from Hanford and Livingston for the two false triggers. The two signals denoised by the strain of Hanford and Livingston are shifted to align with the time corresponding to the maximum value. Interestingly, the denoised shapes of Hanford and Livingston around GPS time 1187014846 appear remarkably similar. We calculated the overlap between the denoised output of the two interferometers near GPS time 1187014846 and yielded 88.9%. We believe that the strain near the GPS time can be further investigated in the future. This indicates the potential of our proposed framework for conducting more in-depth investigations into the archived data of GWTC-1, GWTC-2, and GWTC-3.

The computational costs of the proposed framework can be evaluated based on the runtime of each stage involved in analyzing half a month's worth of data. We calculated the running time of each stage using a workstation equipped with a consumer-grade NVIDIA 3060 GPU, and the results are displayed in Table VII. From the table, it is evident that the combined running time of all stages for analyzing half a month's data is under 14 hours. Moreover, the operational speed of the proposed framework can be enhanced by leveraging multiple GPU platforms.

TABLE VI. The number of false triggers for the proposed framework detected by Hanford only (H), and that detected by Hanford and tested by Livingston (H + L), for Methods I, II, and III. The abbreviations H, L, T, and S denote Hanford, Livingston, time-difference testing, and network SNR threshold testing. The network SNR threshold is set to 8.

		Method I				Method II				Method III			
Mass1 ($M_\odot$)	Mass2 ($M_\odot$)	H	$H+$L	H+L+T	H+L+T+S	H	$H+$L	H+L+T	H+L+T+S	H	$H+$L	H+L+T	H+L+T+S
[5, 10]	[5, 10]	87	1	0	0	3	0	0	0	2	0	0	0
[5, 10]	[10, 20]	383	9	0	0	135	0	0	0	29	0	0	0
[5, 10]	[20, 40]	1196	19	0	0	284	1	0	0	44	0	0	0
[5, 10]	[40, 80]	3265	32	2	2	300	0	0	0	49	0	0	0
[10, 20]	[10, 20]	717	5	0	0	280	0	0	0	27	0	0	0
[10, 20]	[20, 40]	2710	39	1	1	646	0	0	0	34	0	0	0
[10, 20]	[40, 80]	5259	55	7	7	2290	6	1	1	92	0	0	0
[20, 40]	[20, 40]	3823	41	4	4	1197	2	0	0	45	0	0	0
[20, 40]	[40, 80]	5185	60	15	11	2238	7	0	0	129	0	0	0
[40, 80]	[40, 80]	4375	135	65	3	1061	2	1	1	246	1	0	0

TABLE VII. Execution time.

Stages	Time
STP	<7.02 h
PTP	<3.60 h
TS	<3.20 h
MF	<2.80 s

V. CONCLUSION AND DISCUSSION

In this work, we introduce a novel framework to predict the matched filtering SNR without a theoretical template bank. This framework employs an envelope extraction network alongside denoising networks and astrophysical origin discrimination networks. Notably, we have developed and trained ten denoising networks and ten astrophysical discrimination networks within this proposed framework. The test results of both the test set and the real LIGO data demonstrate the effectiveness of the proposed framework.

For the sake of simplicity in our analysis, we opt for a relatively compact neural network structure for denoising, consisting of just 11 layers. This denoising network can be executed and trained on a NVIDIA GeForce RTX 3060 Laptop GPU (6 GB). We demonstrate the effectiveness of a neural network with a relatively small scale for denoising.

We are of the opinion that more complex network architectures, such as deeper networks like WaveNet [44], wider networks like WaveFormer [46], and sequential modeled networks like CNN-LSTM [45], can be seamlessly integrated into our framework to enhance the overall gravitational wave search performance.

The denoising output serves as an intermediary outcome within our approach. This intermediary result encompasses crucial details like arrival time delays, signal amplitudes, and phases. Such information can be harnessed to enhance various other facets of gravitational wave data processing, including tasks like localization and parameter estimation.

Given the notably low phase recovery error exhibited by the end-to-end denoising model, our proposed method holds promise for effectively tackling the phase-related challenges in the search for Extreme Mass Ratio Inspiral (EMRI) GW signals [32] in the future. Additionally, this method exhibits potential utility in detecting GW signals stemming from binaries with quantifiable eccentricities.

ACKNOWLEDGMENTS

This research has made use of data and web tools obtained from the gravitational-wave Open Science Center, a service of LIGO Laboratory, the LIGO Scientific Collaboration, and the Virgo Collaboration. We thank Xinquan Chen for his help with the data annotation. This work was supported in part by the National Key Research and Development Program of China Grant No. 2021YFC2203001, in part by the NSFC (Grants No. 11920101003 and No. 12021003) and the Natural Science Foundation of Jiangxi (Grant No. 20224BAB211012). Z. Cao was supported by the CAS Project for Young Scientists in Basic Research YSBR-006.

[1] B. P. Abbott, Richard Abbott, TDea Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, Christoph Affeldt *et al.*, 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).

[2] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. X. Adhikari, V. B. Adya, Christoph Affeldt *et al.*, 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).

[3] R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, N. Adhikari, R. X. Adhikari, V. B. Adya, C. Affeldt, D. Agarwal *et al.*, GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run, Phys. Rev. X **13**, 041039 (2023).

[4] R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, N. Adhikari, R. X. Adhikari, V. B. Adya, C. Affeldt, D. Agarwal *et al.*, Population of merging compact binaries inferred using gravitational waves through GWTC-3, Phys. Rev. X **13**, 011048 (2023).

[5] Benjamin P Abbott, R. Abbott, T. D. Abbott, Fausto Acernese, K. Ackley, C. Adams, T. Adams, Paolo Addesso, Rana X Adhikari, Vaishali B. Adya *et al.*, Tests of general relativity with GW170817, Phys. Rev. Lett. **123**, 011102 (2019).

[6] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya *et al.*, A gravitational-wave standard siren measurement of the Hubble constant, Nature (London) **551**, 85 (2017).

[7] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt *et al.*, Constraints on cosmic strings using data from the third Advanced LIGO–Virgo observing run, Phys. Rev. Lett. **126**, 241102 (2021).

[8] R. Abbott, H. Abe, F. Acernese, K. Ackley, N. Adhikari, R. X. Adhikari, V. K. Adkins, V. B. Adya, C. Affeldt, D. Agarwal *et al.*, All-sky search for continuous gravitational

waves from isolated neutron stars using Advanced LIGO and Advanced Virgo O3 data, Phys. Rev. D **106**, 102008 (2022).

[9] Alba Romero-Rodriguez, Mario Martinez, Oriol Pujolas, Mairi Sakellariadou, and Ville Vaskonen, Search for a scalar induced stochastic gravitational wave background in the third LIGO-Virgo observing run, Phys. Rev. Lett. **128**, 051301 (2022).

[10] Quentin Baghi, Natalia Korsakova, Jacob Slutsky, Eleonora Castelli, Nikolaos Karnesis, and Jean-Baptiste Bayle, Detection and characterization of instrumental transients in LISA Pathfinder and their projection to LISA, Phys. Rev. D **105**, 042002 (2022).

[11] Zheng-Cheng Liang, Yi-Ming Hu, Yun Jiang, Jun Cheng, Jian-dong Zhang, Jianwei Mei, Science with the tianqin observatory: Preliminary results on stochastic gravitational-wave background, Phys. Rev. D **105**, 022001 (2022).

[12] Ziren Luo, Yan Wang, Yueliang Wu, Wenrui Hu, and Gang Jin, The taiji program: A concise overview, Prog. Theor. Exp. Phys. **2021**, 05A108 (2021).

[13] Michele Maggiore, Chris Van Den Broeck, Nicola Bartolo, Enis Belgacem, Daniele Bertacca, Marie Anne Bizouard, Marica Branchesi, Sebastien Clesse, Stefano Foffa, Juan García-Bellido *et al.*, Science case for the Einstein telescope, J. Cosmol. Astropart. Phys. 03 (2020) 050.

[14] Evan D Hall, Cosmic explorer: A next-generation ground-based gravitational-wave observatory, Galaxies **10**, 90 (2022).

[15] Christopher Michael Biwer, Collin D. Capano, Soumi De, Miriam Cabero, Duncan A. Brown, Alexander H. Nitz, and Vivien Raymond, Pycbc inference: A Python-based parameter estimation toolkit for compact binary coalescence signals, Publ. Astron. Soc. Pac. **131**, 024503 (2019).

[16] Kipp Cannon, Sarah Caudill, Chiwai Chan, Bryce Cousins, Jolien D. E. Creighton, Becca Ewing, Heather Fong, Patrick Godwin, Chad Hanna, Shaun Hooper *et al.*, Gstlal: A software framework for gravitational wave discovery, SoftwareX **14**, 100680 (2021).

[17] Florian Aubin, Francesco Brighenti, Roberto Chierici, Dimitri Estevez, Giuseppe Greco, Gianluca Maria Guidi, Vincent Juste, Frédérique Marion, Benoit Mours, Elisa Nitoglia *et al.*, The MBTA pipeline for detecting compact binary coalescences in the third LIGO-Virgo observing run, Classical Quantum Gravity **38**, 095004 (2021).

[18] Qi Chu, Manoj Kovalam, Linqing Wen, Teresa Slaven-Blair, Joel Bosveld, Yanbei Chen, Patrick Clearwater, Alex Codoreanu, Zhihui Du, Xiangyu Guo *et al.*, Spiir online coherent pipeline to search for gravitational waves from compact binary coalescences, Phys. Rev. D **105**, 024023 (2022).

[19] Pablo J Barneo, Alejandro Torres-Forné, José A Font, Marco Drago, Jordi Portell, and Antonio Marquina, Implementation of the regularized rudin-osher-fatemi denoising method in the coherent wave burst pipeline for gravitational-wave data analysis, Phys. Rev. D **106**, 022002 (2022).

[20] Daniel George and Eliu Antonio Huerta, Deep learning for real-time gravitational wave detection and parameter estimation: Results with Advanced LIGO data, Phys. Lett. B **778**, 64 (2018).

[21] Hunter Gabbard, Michael Williams, Fergus Hayes, and Chris Messenger, Matching matched filtering with deep networks for gravitational-wave astronomy, Phys. Rev. Lett. **120**, 141103 (2018).

[22] Man Leong Chan, Ik Siong Heng, and Chris Messenger, Detection and classification of supernova gravitational wave signals: A deep learning approach, Phys. Rev. D **102**, 043022 (2020).

[23] Grégory Baltus, Justin Janquart, Melissa Lopez, Amit Reza, Sarah Caudill, and Jean-René Cudell, Convolutional neural networks for the detection of the early inspiral of a gravitational-wave signal, Phys. Rev. D **103**, 102003 (2021).

[24] Banafsheh Beheshtipour and Maria Alessandra Papa, Deep learning for clustering of continuous gravitational wave candidates. II. Identification of low-SNR candidates, Phys. Rev. D **103**, 064027 (2021).

[25] Rich Ormiston, Tri Nguyen, Michael Coughlin, Rana X Adhikari, and Erik Katsavounidis, Noise reduction in gravitational-wave data via deep learning, Phys. Rev. Res. **2**, 033066 (2020).

[26] Chayan Chatterjee, Linqing Wen, Kevin Vinsen, Manoj Kovalam, and Amitava Datta, Using deep learning to localize gravitational wave sources, Phys. Rev. D **100**, 103025 (2019).

[27] He Wang, Shichao Wu, Zhoujian Cao, Xiaolin Liu, and Jian-Yang Zhu, Gravitational-wave signal recognition of LIGO data by deep learning, Phys. Rev. D **101**, 104003 (2020).

[28] CunLiang Ma, Wei Wang, He Wang, and Zhoujian Cao, Ensemble of deep convolutional neural networks for real-time gravitational wave signal recognition, Phys. Rev. D **105**, 083013 (2022).

[29] Marlin B. Schäfer and Alexander H. Nitz, From one to many: A deep learning coincident gravitational-wave search, Phys. Rev. D **105**, 043003 (2022).

[30] Marlin B Schäfer, Ondřej Zelenka, Alexander H Nitz, Frank Ohme, and Bernd Brügmann, Training strategies for deep learning gravitational-wave searches, Phys. Rev. D **105**, 043002 (2022).

[31] Cunliang Ma, Wei Wang, He Wang, and Zhoujian Cao, Artificial intelligence model for gravitational wave search based on the waveform envelope, Phys. Rev. D **107**, 063029 (2023).

[32] Xue-Ting Zhang, Chris Messenger, Natalia Korsakova, Man Leong Chan, Yi-Ming Hu, and Jian-dong Zhang, Detecting gravitational waves from extreme mass ratio inspirals using convolutional neural networks, Phys. Rev. D **105**, 123027 (2022).

[33] Christoph Dreissigacker, Rahul Sharma, Chris Messenger, Ruining Zhao, and Reinhard Prix, Deep-learning continuous gravitational waves, Phys. Rev. D **100**, 044009 (2019).

[34] M. López, I. Di Palma, M. Drago, P. Cerdá-Durán, and F. Ricci, Deep learning for core-collapse supernova detection, Phys. Rev. D **103**, 063011 (2021).

[35] Seiya Sasaoka, Yilun Hou, Kentaro Somiya, and Hirotaka Takahashi, Localization of gravitational waves using machine learning, Phys. Rev. D **105**, 103030 (2022).

[36] Chayan Chatterjee, Linqing Wen, Kevin Vinsen, Manoj Kovalam, and Amitava Datta, Using deep learning to

localize gravitational wave sources, Phys. Rev. D **100**, 103025 (2019).
[37] Maximilian Dax, Stephen R. Green, Jonathan Gair, Jakob H. Macke, Alessandra Buonanno, and Bernhard Schölkopf, Real-time gravitational wave science with neural posterior estimation, Phys. Rev. Lett. **127**, 241103 (2021).
[38] He Wang, Zhoujian Cao, Yue Zhou, Zong-Kuan Guo, and Zhixiang Ren, Sampling with prior knowledge for high-dimensional gravitational wave data analysis, Big Data Mining and Anal. **5**, 53 (2021).
[39] Jurriaan Langendorff, Alex Kolmus, Justin Janquart, and Chris Van Den Broeck, Normalizing flows as an avenue to studying overlapping gravitational wave signals, Phys. Rev. Lett. **130**, 171402 (2023).
[40] Hunter Gabbard, Chris Messenger, Ik Siong Heng, Francesco Tonolini, and Roderick Murray-Smith, Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy, Nat. Phys. **18**, 112 (2022).
[41] Huilan Luo, Pei Wang, Hongkun Chen, and Min Xu, Object detection method based on shallow feature fusion and semantic information enhancement, IEEE Sens. J. **21**, 21839 (2021).
[42] Shuxin Yang, Suxin Tong, Guixiang Zhu, Jie Cao, Youquan Wang, Zhengfa Xue, Hongliang Sun, and Yu Wen, Mve-flk: A multi-task legal judgment prediction via multi-view encoder fusing legal keywords, Knowledge-Based Systems **239**, 107960 (2022).
[43] Timothy D. Gebhard, Niki Kilbertus, Ian Harry, and Bernhard Schölkopf, Convolutional neural networks: A magic bullet for gravitational-wave detection?, Phys. Rev. D **100**, 063015 (2019).
[44] Wei Wei and E. A. Huerta, Gravitational wave denoising of binary black hole mergers with deep learning, Phys. Lett. B **800**, 135081 (2020).
[45] Chayan Chatterjee, Linqing Wen, Foivos Diakogiannis, and Kevin Vinsen, Extraction of binary black hole gravitational wave signals from detector data using deep learning, Phys. Rev. D **104**, 064046 (2021).
[46] Zhixiang Ren, He Wang, Yue Zhou, Zong-Kuan Guo, and Zhoujian Cao, Intelligent noise suppression for gravitational wave observational data, arXiv:2212.14283.
[47] Alexander H. Nitz, Thomas Dent, Tito Dal Canton, Stephen Fairhurst, and Duncan A. Brown, Detecting binary compact-object mergers with gravitational waves: Understanding and improving the sensitivity of the PyCBC search, Astrophys. J. **849**, 118 (2017).
[48] Samantha A. Usman, Alexander H. Nitz, Ian W. Harry, Christopher M. Biwer, Duncan A. Brown, Miriam Cabero, Collin D. Capano, Tito Dal Canton, Thomas Dent, Stephen Fairhurst *et al.*, The PyCBC search for gravitational waves from compact binary coalescence, Classical Quantum Gravity **33**, 215004 (2016).
[49] Hua-Mei Luo, Wenbin Lin, Zu-Cheng Chen, and Qing-Guo Huang, Extraction of gravitational wave signals with optimized convolutional neural network, Front. Phys. **15**, 1 (2020).
[50] Heming Xia, Lijing Shao, Junjie Zhao, and Zhoujian Cao, Improved deep learning techniques in gravitational-wave data analysis, Phys. Rev. D **103**, 024040 (2021).