# Using deep learning to denoise and detect gravitational waves

CunLiang Ma,[1,2] ShuoQiu Li,[1] Zhoujian Cao,[3,4,*] and Mingzhen Jia[1,2]

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

[2]*Jiangxi Province Key Laboratory of Multidimensional Intelligent Perception and Control, Ganzhou 341000, 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 17 June 2024; accepted 16 July 2024; published 6 September 2024)

We have upgraded the MSNRnet framework to MSNRnet-2 by refining the training strategy, drawing inspiration from generative adversarial networks for data generation. The astrophysical discrimination network enforces constraints on the denoising output, ensuring that the "signal + noise" case conforms to an astrophysical-origin shape, while limiting the denoised output of the "noise" case to a nonastrophysical origin shape. This improvement has led to enhanced denoising and astrophysical discrimination performance in most scenarios. For the analysis of confident events, we observed a 1.3% increase in the detection rate. Remarkably, no false triggers were generated during the analysis of the data jointly processed by the Hanford and Livingston interferometers in O3b (about five months of data). These findings suggest that the denoising-discrimination-matched filtering framework holds promising potential for gravitational wave searches in the future.

DOI: 10.1103/PhysRevD.110.063010

## I. INTRODUCTION

From 2015 to 2020, the advanced LIGO and Virgo interferometers have successfully identified more than 90 confident gravitational wave (GW) events originating from compact binary coalescences (CBCs) through the first, second, and third observing runs [1–4]. In 2023, four teams of pulsar timing array (PTA) groups announced the evidence for vastly-low-frequency (nanohertz) gravitational waves [5–8]. Thanks to these initial detections, numerous noteworthy achievements have already been attained in both astrophysics and fundamental physics. Recently, the Laser Interferometer Space Antenna (LISA) [9] that focuses on low-frequency GW searches was successfully adopted by European Space Agency (ESA). In the future, third-generation (3G) ground-based detectors, such as ET [10] and CE [11], will offer remarkable advancements in sensitivity [12], increasing it by an order of magnitude, and significantly broadening the bandwidth, extending both towards lower and higher frequencies. These detectors hold exceptional potential for groundbreaking discoveries in astrophysics, cosmology, and fundamental physics.

LIGO–Virgo–KAGRA (LVK) Collaboration employs matched filtering (MF) method [13–16] for the search of GW signals. Under the assumption of stationary Gaussian background noise, the MF-based method is optimal in the Neyman–Pearson senses. In the 3G era, when higher-order modes and orbit eccentricity become significant, the MF algorithm relying on a template bank will encounter computational limits. To solve the problem of computational efficiency, recently, many works have investigated the deep learning (DL) based methods for GW search [17,18]. Research has shown that convolutional neural network (CNN) [19–26], dilated CNN [27,28], the ensemble of multiple deep networks [29–31], recurrent neural network (RNN) [32], and other deep learning model structures can be applied to GW search. Different data preprocessing methods were also investigated in the previous works [33,34], such as adding a matching perception layer and removing high-frequency components after wavelet transform. Apart from the compact binary coalescence, some works investigate the search of the GW emitted by core-collapse supernova [35,36], stochastic backgrounds [37], and rotating compact objects [38–40]. In addition to the GW search, DL has also been studied at multiple aspects of gravitational wave data processing, including glitch classification [41–44], denoising [45–47], source localization [48–51], and parameter estimation [52–56]. Besides the data processing of second-generation ground-based GW detectors, many works investigate the utilization of DL to other forms of GW detectors such as 3G ground-based detectors [57,58], space-based

*Contact author: zjcao@amt.ac.cn

 © 2024 American Physical Society

detectors [59,60], pulsar timing array [61,62], and decihertz GW signal processing [63].

Most DL-based GW searches rely on end-to-end classification methods, which are prone to saturation and pose challenges in distinguishing low signal-to-noise ratio (SNR) events. The MF-based GW search method produces an output that represents the fluctuation of the matched signal-to-noise ratio with respect to time, which offers valuable insights for subsequent endeavors such as source localization [64]. Meanwhile, end-to-end classification methods suffer from the problem of physical interpretability. To solve these problems, in the previous work we propose a different framework MSNRnet [65] for GW search, and we use denoising output as a substitute for the template bank of the MF-based method. The denoised output cannot be immediately employed for subsequent matched filtering. Only after verifying that it possesses the GW shape we can proceed to match it with the original strain. Within the MSNRnet framework, we initially utilize the envelope extraction network [34] to predict the significant time that needs to be analyzed. Subsequently, we employ ten denoising networks to generate preliminary templates. Following this, we engage the astrophysical origin discrimination networks to select the templates. Ultimately, the selected templates are used in matched filtering. The framework's output provides an explanation regarding the matched filtering signal-to-noise ratio.

In a previous work [65], we optimize GW denoising networks and astrophysical origin discrimination networks separately. We call such a model as MSNRnet. The performance of the networks rely on the training set. Three kinds of astrophysical origin discrimination networks are found using different datasets. In this work, we introduce a novel training approach which trains the denoising model and the astrophysical discrimination model simultaneously.

We find that the denoising model of the original MSNRnet tends to 'create' short timescale GW signal from the pure noise. This fact leads to the judgment of noise as gravitational waves. Drawing inspiration from the generative adversarial network (GAN) [66–68], we refine MSNRnet to avoid this kind of creation. In the original MSNRnet, the denoising networks were optimized solely based on the mean square error, and only the signal + noise scenario was used for training. Additionally, the denoising networks and astrophysical origin discrimination networks were optimized independently from each other.

Similar to the GAN framework, in the current work, we simultaneously optimize both the GW denoising networks and the astrophysical origin discrimination networks. To optimize the denoising models, we adopt two objectives. First, we aim to denoise the signal + noise case to signal and the noise case to zero. Second, we aim to denoise the signal + noise case to astrophysical origin shape and the noise case to nonastrophysical origin shape. The second objective is inspired by GANs, where the discriminator aims to determine whether the data belongs to the training set or not and the generator aims to generate data that triggers the discriminator.

Inspired by transfer learning and reinforcement learning with human feedback, after using a GAN-like method to train the denoising and astrophysical discrimination networks, we then fine-tune the 0.25 s timescale. This timescale corresponds to the parameter space $M_1 \in (40M_\odot, 80M_\odot)$ and $M_2 \in (40M_\odot, 80M_\odot)$. Correspondingly, the astrophysical discrimination model uses manually annotated data used in Method I of the original MSNRnet.

We explain our upgraded MSNRnet model in the next section. Then in Sec. III we compare the performance of the updated version and the original version of MSNRnet on the confident events and all the strain data of O3b (approximately five months of data). Remarkably, the updated framework did not generate any false alarms while scanning through all the noise present in the O3b dataset. At last we conclude the paper in Sec. IV with some discussions.

## II. UPGRADED MSNRNET MODEL

### A. Model structure of MSNRnet

In our framework, we focus on the data from Livingston and Hanford detectors, while excluding Virgo from the analysis. The reason for this selection is that our framework encounters difficulties in predicting the template of events with very low SNR. Considering that Virgo's sensitivity is relatively inferior to Hanford's and Livingston's, most confident events detected by Virgo exhibit an SNR of less than 6. Regrettably, our framework cannot accurately forecast the template at such a low SNR. Therefore, we opted to exclude Virgo data from our primary analysis. However, we believe that the data collected by Virgo could still be valuable for subsequent tasks, such as source localization.

The MSNRnet is a multistep processing framework as shown in Fig. 1. The entire framework comprises four critical stages: significant time prediction (STP), preliminary templates prediction (PTP), template selection (TS), and matched filtering. First, we put the whitened strain to envelope extraction network which is proposed in [34], and we get the envelope output that has coalescence time information. Second, if the coalescence times of Hanford and/or Livingston interferometers are successfully predicted, we will put the whitened strain for each interferometer to $N$ denoising models and get $2N$ denoised outputs. Third, the $2N$ denoised outputs may have GW shape or not, and $N$ astrophysical origin discrimination networks (DiscriminationModel) are used to distinguish whether the denoised outputs have astrophysical origin or

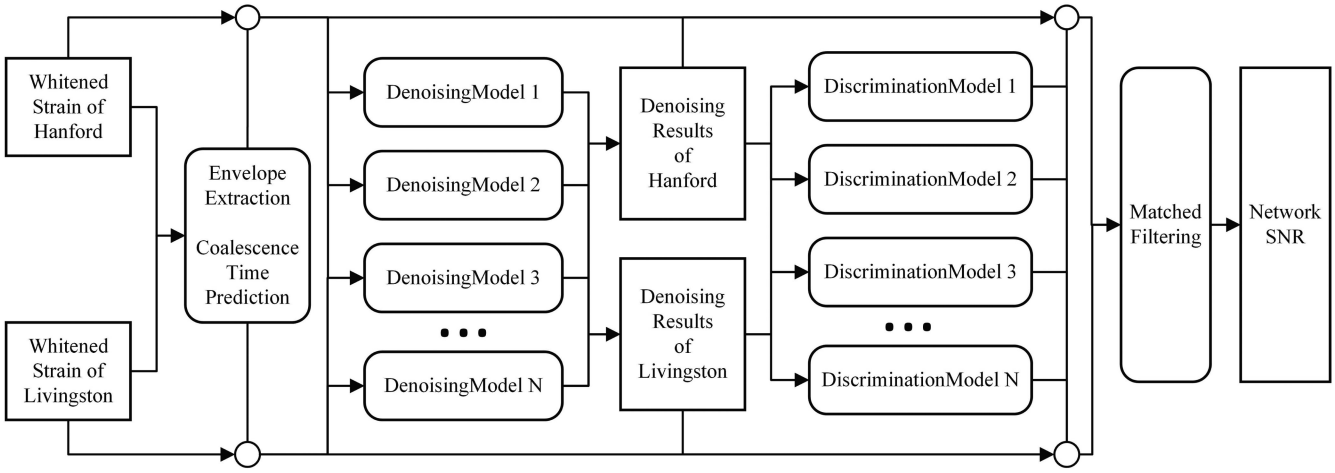


FIG. 1. The diagram of the MSNRnet. The circles in the diagram represent the switch operation of data flow.

not. Fourth, if the DiscriminationModel$_i$ verifies that the outputs from both the Hanford and Livingston interferometers of the DenoisingModel$_i$ exhibit waveforms of astrophysical origin, we will proceed with matched filtering using the outputs from the DenoisingModel$_i$ and the original strain.

In the case of the original MSNRnet, we optimized the GW denoising networks and astrophysical origin discrimination networks independently. However, this approach made the performance of both networks reliant on the specific training dataset. Additionally, the astrophysical origin discrimination networks were trained using datasets that required manual annotation, introducing a level of human subjectivity. To address this, we present a new training method and denote the new model as MSNRnet-2 in this work which removes the dependency on manually annotated data.

Within the parameter range where $M_1$ and $M_2$ both fall between $5M_\odot$ and $10M_\odot$, the original MSNRnet [65] exhibits a commendably low false alarm probability. Given this performance, we do not utilize the updated method for this specific parameter space. Instead, we utilize Method II from the original MSNRnet for this specified range of parameters.

### B. The MSNRnet-2

Drawing inspiration from GAN, we introduce MSNRnet-2. GAN stands as one of the most widely used generative algorithms. This algorithm finds extensive application in generating novel data points, accommodating various datasets such as images and sound signals. GAN has achieved remarkable success in tasks like generating realistic human faces [69], converting gray scale images to colored ones [70], translating text descriptions into realistic images [71], and numerous other applications. In terms of gravitational wave data analysis, some works investigate the utilization of GANs to glitch classification [43] and glitch and/or gravitational wave generation [72–74].

The GAN architecture is depicted in the top panel of Fig. 2. In the training phase, two types of models are employed: the generator and the discriminator. Let us assume that the generator and discriminator can be viewed as parametrized systems. Specifically, the generator can be denoted as $G(z, W_G)$ and the discriminator as $D(x, W_D)$, where $z$ and $x$ represent their respective inputs, and $W_G$ and $W_D$ mean the trainable variables for the generator and discriminator, respectively. The generator learns the distribution of samples within the training dataset, while the objective of the discriminator is to differentiate between the samples from the training dataset and the ones generated by the generator. Loss function of discriminator $L^D$ and generator $L^G$ can be written as:

$$L^D = -\mathbb{E}_{x\sim p_x(x)}[\log D(x)] - \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))], \quad (1)$$

$$L^G = -\mathbb{E}_{z\sim p_z(z)} \log(D(G(z))). \quad (2)$$

Here the notation $\mathbb{E}_{x\sim p(x)}[f(x)]$ represents the expected value of $f(x)$ when $x$ is drawn from the distribution $p(x)$.

Returning to the MSNRnet framework, we can draw parallels between its components and those of the GAN framework: the denoising network is analogous to the generator in GAN, and the astrophysical origin discrimination network (DiscriminationModel) corresponds to the discriminator in GAN. The proposed framework for training the two models is shown in the bottom panel of Fig. 2. Equation (1) and Eq. (2) cannot be used to optimize the two models, because the two models have no adversarial relationship. The objective of the DenoisingModel is to produce signal-like patterns for signal + noise instances, and generate non-signal-like patterns for cases that consist of noise alone. Meanwhile, the goal of the

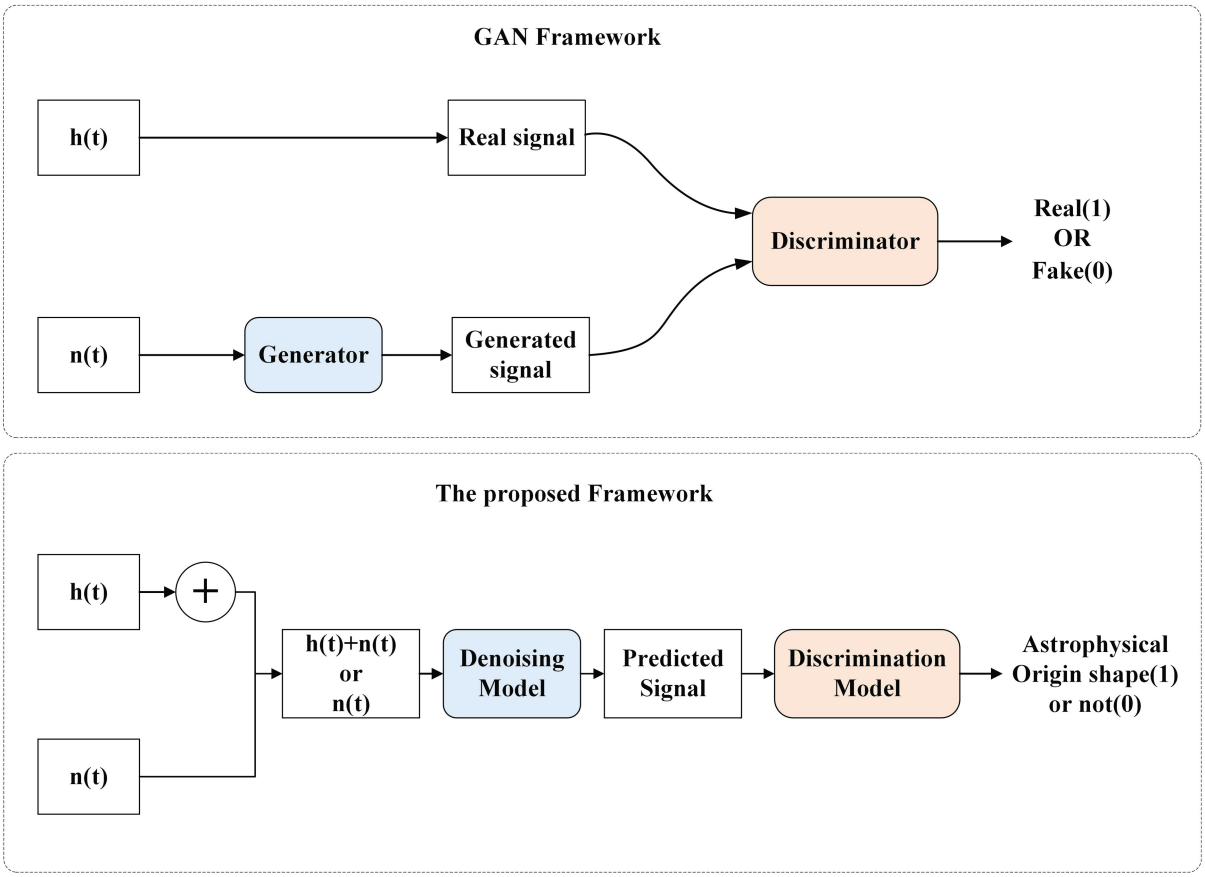


FIG. 2. The training diagram of the generator and discriminator in GAN framework (top panel) and DenoisingModel and DiscriminationModel in the proposed framework (bottom panel).

DiscriminationModel is to differentiate between signal-like and non-signal-like patterns. Suppose the DenoisingModel is denoted as $GD(s, W_{GD})$ and the DiscriminationModel is denoted as $DD(x, W_{DD})$, where $s$ denotes the signal + noise or noise, and $W_{GD}$ and $W_{DD}$ are all trainable variables. Based on the above analysis, we obtain the loss function of the two networks:

$$L^{\mathrm{TWO}} = -\alpha \mathbb{E}_{s\sim p(n+h)}[\log\left(DD(GD(s))\right)] - \mathbb{E}_{s\sim p(n)}[\log\left(1 - DD(GD(s))\right)], \tag{3}$$

where $\alpha$ can be adjusted to control the relative importance of distinguishing the denoising output, weighing the significance of signal + noise against noise. In addition, another object of the denoising network is to restore the buried signals waveform and suppress the amplitude of noise. Like the original MSNRnet and other works about GW denoising, we also use the mean square error (MSE) loss of the denoising output:

$$L^{\mathrm{MSE}} = -\mathbb{E}_{s\sim p(n+h)}[\|GD(s) - h\|^2] + \mathbb{E}_{s\sim p(n)}[\|GD(s)\|^2]. \tag{4}$$

The strain $s$ in the first term on the right side of Eq. (4) can be written as $s = n + h$, where $n$ denotes the noise and $h$ denotes the signal. Combine Eq. (3) and Eq. (4) we get the loss function of the overall DenoisingModel and DiscriminationModel,

$$L^{\mathrm{ALL}} = L^{\mathrm{TWO}} + L^{\mathrm{MSE}}. \tag{5}$$

### C. Training strategy of MSNRnet-2

Compared to the original MSNRnet, we augment the noise samples in the training dataset. In the data processing pipeline, the denoising step is conducted after successful envelope extraction. In the previous work the results show that if the envelope of the noise sample is successfully extracted, the denoising output will have high probability of having GW shape. To alleviate this problem, in the revised version of MSNRnet we augment the noise samples in the training dataset with noise that passed the envelope extraction network successfully. All the noise samples of the Hanford interferometer in August 2017 that triggered the envelope extraction model (18,558 samples) are used to augment the "noise" sample in training dataset. All the

TABLE I. The binary black hole (BBH) parameter regions and the corresponding $\alpha$ values. The unit of the black holes' mass is $M_\odot$. Since $M_1 \geq M_2$ only half of the table makes sense. For the mass range $M_{1,2} \in (5, 10]$ Method II of the original MSNRnet is used, so there is no $\alpha$ parameter involved.

| | $M_1$ | | | |
|---|---|---|---|---|
| $M_2$ | (5, 10] | (10, 20] | (20, 40] | (40, 80] |
| (5, 10] | ⋯ | 0.1 | 1.0 | 0.3 |
| (10, 20] | | 0.4 | 0.2 | 0.4 |
| (20, 40] | | | 0.5 | 0.1 |
| (40, 80] | | | | 0.8 |

"noise" and "signal" samples for training previous MSNRnet are also used in this work.

Direct optimization of Eq. (5) poses significant challenges. We employ a three-step approach to progressively optimize the equation. First, we train the denoising model using MSE loss only:

$$W_{GD} = \arg\min_{W_{GD}} \mathbb{E}_{s\sim p(n+h)}[\|GD(s) - h\|^2] + \mathbb{E}_{s\sim p(n)}[\|GD(s)\|^2]. \tag{6}$$

Second, we set the weight of the denoising model nontrainable and train the DiscriminationModel only:

$$W_{DD} = \arg\min_{W_{DD}} -\alpha\mathbb{E}_{s\sim p(n+h)}[\log(DD(GD(s)))] - \mathbb{E}_{s\sim p(n)}[\log(1 - DD(GD(s)))]. \tag{7}$$

The parameter $\alpha > 0$ can control the relative importance of signal and noise cases. If $\alpha < 1$, then the model is inclined to reduce the false alarm probability. If $\alpha > 1$, then the model's output is inclined to have a signal shape. In the original MSNRnet, we find that the denoising model tends to denoise the noise into the GW shape, and we also find that different parameter spaces perform differently. So, in this paper, the $\alpha$ is chosen according to the parameter space via experience results. Through extensive experimental validation, we chose the $\alpha$ according to the parameter space which is shown in Table I. After the second step, preliminary DenoisingModels and DiscriminationModels are found. Third, for each parameter space, optimize both DenoisingModel and DiscriminationModel together:

$$W_{GD}, W_{DD} = \arg\min_{W_{GD}, W_{DD}} L^{ALL}. \tag{8}$$

In the first two steps, we utilize the Adam method for optimization. The batch size was set to 64 and the learning rate was set to $1 \times 10^{-5}$. We train 35 epochs and introduce a differentiated learning rate strategy similar to GAN during the third training step. We set the learning rate to $1 \times 10^{-5}$ for optimizing $W_{GD}$ and $1 \times 10^{-6}$ for optimizing $W_{DD}$. This can ensure the stability of training. For this stage, we apply the Adam algorithm to train for five epochs.

## III. PERFORMANCE OF THE UPGRADED MSNRNET AGAINST THE REAL DATA

Given the impressively low false alarm rate for the mass range $(5M_\odot, 10M_\odot)$ exhibited by the original MSNRnet, for this parameter space, we use the models trained in the previous work [65]. For other mass ranges, we train both denoising networks and astrophysical discrimination networks using the aforementioned methodology. Subsequently, we rigorously evaluate both types of networks using a test dataset, confident events, and approximately five months of O3b data. The GPS time of the background noise of the test dataset is sampled from 1238904832 to 1239023616, which is different from the training dataset (from 1238163456 to 1238806528). We believe that maintaining distinct GPS times for the training and testing datasets serves to guarantee that the background noises in both sets are different. This variation in noise allows for a more impartial assessment of the model's performance. For each black hole mass range parameter space, 10,000 samples are generated to test the performance of the models.

### A. Performance of the denoising networks

We evaluate each denoising network using a signal + noise case. Specifically, we feed samples from the corresponding test dataset into the denoising network and obtain the output. Then, we analyze the overlap between the denoised output and the original buried signal, which can be calculated by

$$o(h, \hat{h}) = \frac{\langle h, \hat{h}\rangle}{\sqrt{\langle h, h\rangle\langle \hat{h}, \hat{h}\rangle}}, \tag{9}$$

where $h$ denotes the buried signal and $\hat{h}$ represents the denoising output. The overlap distributions for all denoising networks on their respective test datasets are presented in Fig. 3, including a comparison with the denoising output generated by the original MSNRnet. For most cases, there is a slight improvement of the denoising networks performance in MSNRnet-2. As observed in Fig. 3(c), for the case where $M_1 \in (5M_\odot, 10M_\odot)$ and $M_2 \in (40M_\odot, 80M_\odot)$, there is a notable increase in the percentage of overlap within the range of (0.9, 1.0), rising from 72% to 78%. However, MSNRnet-2's performance slightly diminishes in the parameter case of $M_1 \in (10M_\odot, 20M_\odot)$ and $M_2 \in (10M_\odot, 20M_\odot)$. Additionally, Fig. 3(i) reveals that the denoising networks of both MSNRnet and MSNRnet-2 exhibit identical performance for the parameter case of $M_1 \in (40M_\odot, 80M_\odot)$ and $M_2 \in (40M_\odot, 80M_\odot)$. The

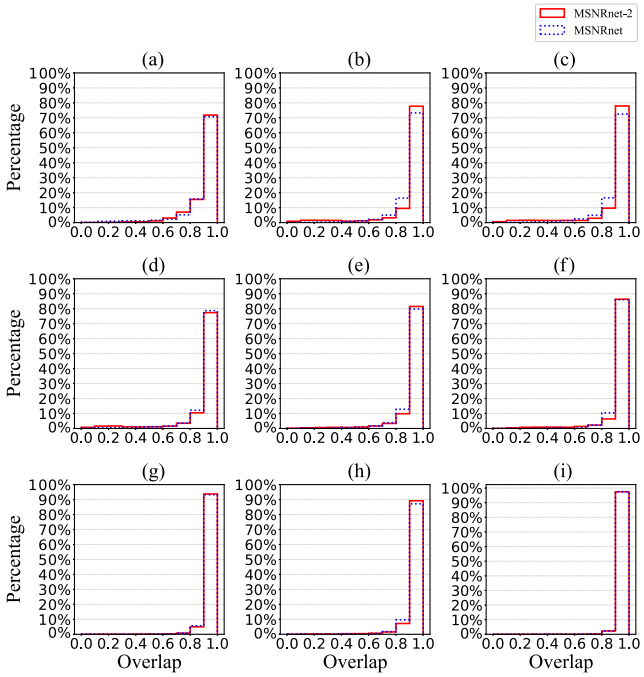


FIG. 3. The overlap of the denoised output and the buried signal in test dataset. (a) $M_1 \in (5M_\odot, 10M_\odot)$ and $M_2 \in (10M_\odot, 20M_\odot)$, (b) $M_1 \in (5M_\odot, 10M_\odot)$ and $M_2 \in (20M_\odot, 40M_\odot)$, (c) $M_1 \in (5M_\odot, 10M_\odot)$ and $M_2 \in (40M_\odot, 80M_\odot)$, (d) $M_1 \in (10M_\odot, 20M_\odot)$ and $M_2 \in (10M_\odot, 20M_\odot)$, (e) $M_1 \in (10M_\odot, 20M_\odot)$ and $M_2 \in (20M_\odot, 40M_\odot)$, (f) $M_1 \in (10M_\odot, 20M_\odot)$ and $M_2 \in (40M_\odot, 80M_\odot)$, (g) $M_1 \in (20M_\odot, 40M_\odot)$ and $M_2 \in (20M_\odot, 40M_\odot)$, (h) $M_1 \in (20M_\odot, 40M_\odot)$ and $M_2 \in (40M_\odot, 80M_\odot)$, (i) $M_1 \in (40M_\odot, 80M_\odot)$ and $M_2 \in (40M_\odot, 80M_\odot)$.

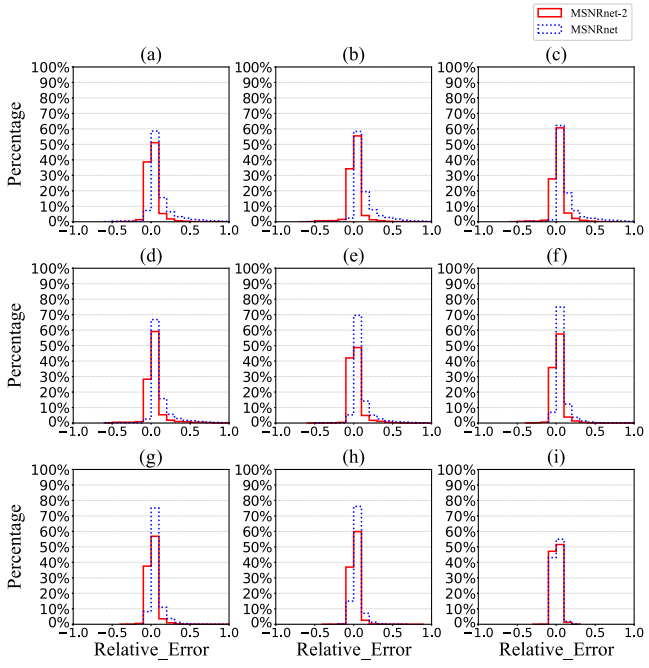


FIG. 4. The statistic results of relative error of predicted SNR for different parameter spaces. The relative error is calculated by $\frac{\mathrm{SNR_{predict}}-\mathrm{SNR_{true}}}{\mathrm{SNR_{true}}}$. (a) $M_1 \in (5M_\odot, 10M_\odot)$ and $M_2 \in (10M_\odot, 20M_\odot)$, (b) $M_1 \in (5M_\odot, 10M_\odot)$ and $M_2 \in (20M_\odot, 40M_\odot)$, (c) $M_1 \in (5M_\odot, 10M_\odot)$ and $M_2 \in (40M_\odot, 80M_\odot)$, (d) $M_1 \in (10M_\odot, 20M_\odot)$ and $M_2 \in (10M_\odot, 20M_\odot)$, (e) $M_1 \in (10M_\odot, 20M_\odot)$ and $M_2 \in (20M_\odot, 40M_\odot)$, (f) $M_1 \in (10M_\odot, 20M_\odot)$ and $M_2 \in (40M_\odot, 80M_\odot)$, (g) $M_1 \in (20M_\odot, 40M_\odot)$ and $M_2 \in (20M_\odot, 40M_\odot)$, (h) $M_1 \in (20M_\odot, 40M_\odot)$ and $M_2 \in (40M_\odot, 80M_\odot)$, (i) $M_1 \in (40M_\odot, 80M_\odot)$ and $M_2 \in (40M_\odot, 80M_\odot)$.

denoising result of signal + noise case illustrates the denoising network's proficiency in noise reduction.

We also evaluate the performance of the SNR prediction within the revised framework. Specifically, we process the strains from the test dataset through the corresponding denoising network and obtained the outputs. Subsequently, we compute the matched filtering between the denoised outputs and the respective strains. Following this, we determine the relative error of the predicted SNR. The relative error can be calculated by

$$\mathrm{SNR_{error}} = \frac{\mathrm{SNR_{predict}} - \mathrm{SNR_{true}}}{\mathrm{SNR_{true}}}, \tag{10}$$

where $\mathrm{SNR_{predict}}$ represents the SNR predicted by our framework, and $\mathrm{SNR_{true}}$ denotes the SNR calculated using the matched filter applied to the original strain and buried signal. For comparison, we also calculate the relative error for the original MSNRnet. The distributions of relative errors for SNR prediction in both frameworks are presented in Fig. 4. The updated framework has significantly enhanced the accuracy of SNR prediction. For the case (a) $M_1 \in (5M_\odot, 10M_\odot)$ and $M_2 \in (10M_\odot, 20M_\odot)$ the percentage of relative error in range $(-0.1, 0.1)$ is improved from 65% to 90%. In previous work, there was a noticeable right deviation phenomenon observed in the SNR prediction results, where in most cases, the predicted values were higher than the actual ones. In the updated version, this problem has been largely resolved. This serves as further evidence that the new work has successfully enhanced the precision of SNR estimations. In Sec. III C, we will showcase the improved accuracy of the revised framework for SNR prediction on confident events.

### B. Performance of the astrophysical discrimination networks

The astrophysical discrimination networks aim to categorize the denoising output of "signal + noise" as belonging to the positive class, while classifying the denoising output of pure "noise" as the negative class. In this section, we conduct experiments to investigate these two aspects.

#### *1. The performance of astrophysical discrimination networks to denoising output of the* signal + noise *case*

We feed the denoising outputs from each denoising model for the "signal + noise" instances from the

FIG. 5. The percentage of correctly classified results by the astrophysical discrimination network for the denoised results of "signal + noise" input.

corresponding test dataset into their respective astrophysical origin discrimination networks to obtain the classification results. Ten thousand samples are used for each parameter space. Figure 5 depicts the variation in the proportion of correct classifications with respect to the SNR interval of buried signals. We can see that both the MSNRnet and MSNRnet-2 can work effectively for "signal + noise." The results indicate that for the case SNR $> 8$, most of the denoising results have a signal shape. For the low SNR case, such as SNR $\in (5, 6)$, most of the denoising output have a noise shape. This indicates that further refinement is necessary for the denoising model, particularly in handling low SNR strain data.

### 2. The performance of astrophysical discrimination networks to denoising output of the noise-only case

In the preceding subsection, we examine the performance of astrophysical discrimination networks for the "signal + noise" scenario. This subsection focuses on exploring the reaction of these networks to denoised outputs from pure "noise" cases. Since both denoising and discrimination processes occur after successfully extracting envelope information, our investigation here solely concentrates on noise that activates the envelope extraction network. We acquire 10,000 noise samples from the O3b Hanford interferometer that triggered the envelope extraction network. Subsequently, we process these samples through denoising networks and forward the denoised outputs to their corresponding astrophysical discrimination networks. We analyze the number of false triggers, which are detailed in Table II. Our new method significantly decreases the frequency of false triggers. As an example, within the subparameter space of $M_1 \in (20M_\odot, 40M_\odot)$ and

TABLE II. The number of false triggers of the astrophysical discrimination networks on the noise that trigger the envelope extraction network in O3b.

| $M_1(M_\odot)$ | $M_2(M_\odot)$ | MSNRnet | MSNRnet-2 |
|---|---|---|---|
| [5, 10] | [10, 20] | 131 | 87 |
| [5, 10] | [20, 40] | 301 | 89 |
| [5, 10] | [40, 80] | 290 | 79 |
| [10, 20] | [10, 20] | 279 | 38 |
| [10, 20] | [20, 40] | 632 | 91 |
| [10, 20] | [40, 80] | 2398 | 58 |
| [20, 40] | [20, 40] | 1094 | 355 |
| [20, 40] | [40, 80] | 2310 | 157 |
| [40, 80] | [40, 80] | 1057 | 1622 |

$M_2 \in (40M_\odot, 80M_\odot)$, the number of false triggers has dropped significantly from 2310 to 157. However, a remaining challenge lies in the black hole's mass parameter space of $M_1 \in (40M_\odot, 80M_\odot)$ and $M_2 \in (40M_\odot, 80M_\odot)$, where both MSNRnet and MSNRnet-2 still produce a high number of false triggers. It is plausible that background noise is more apt to imitate short-duration GW waveforms. In the lower mass ranges, where the signal duration is longer, false triggers are significantly reduced. Conversely, in the higher mass ranges where the signal duration is shorter, the reduction in false triggers is less pronounced.

### C. Performance of the GW search framework on confident events

In the preceding subsection, we discover that numerous false triggers are generated from the Hanford interferometer specifically within the parameter space defined by $M_1 \in (40M_\odot, 80M_\odot)$ and $M_2 \in (40M_\odot, 80M_\odot)$. Upon inspection of the denoising output, we observe that certain false triggers lack a distinct gravitational wave shape. Consequently, we refine the astrophysical discrimination network for this specific parameter space using manually labeled data curated from previous research. It is important to clarify that we focus our fine-tuning efforts solely on this one parameter space.

The GW search performance of the overall framework has been investigated against the confident events listed in the GWTC-1, GWTC-2, GWTC-2.1, and GWTC-3 catalogs, focusing on those where the Hanford and Livingston interferometers are operating simultaneously. For each confident event, we whiten the data encompassing it and extract a 16-second segment for a more indepth analysis. We then employ an envelope extraction model to scan each of these segments using a two-second sliding time step, identifying key timestamps. Once these timestamps are determined, we proceed with denoising and astrophysical discrimination processes. If the waveform activates the astrophysical discrimination network, we perform matched filtering to compute SNR. We further verify the event by ensuring that the time variation of peak SNR between Hanford and Livingston interferometers is less than 15 milliseconds. It is important to note that these findings are derived from a blind search, without prior knowledge of the parameter space or signal length. Table III showcases the network SNR for confident events as predicted by MSNRnet-2. For comparison, we have also provided the offline analysis results from four MF-based pipelines: GstLAL, MBTA, PyCBC, and PyCBC_BBH, alongside the original MSNRnet's outcomes. We have also calculated the detection rates for all

TABLE III. SNRs of GW events reported by LVK during O1, O2, and O3 observing runs are listed, focusing on events that performed well with both Hanford and Livingston interferometers. The table presents SNR results obtained from GstLAL, MBTA, PyCBC, PyCBC_BBH, and the previous MSNRnet, which utilized APOD_MODEL_II as the astrophysical origin discrimination model. Herein, we compare the SNRs predicted by the MSNRnet-2 framework against these previously reported results. The SNR information detected by MBTA and PyCBC_BBH during the first observing run has not been made public; therefore, we represent these cases with blanks in the table. In the table, we denote events that were not detected by the corresponding pipeline with three dots. For some events, the FAR (false alarm rate) fails to meet the criteria of being less than two per year. In such instances, we highlight the SNR value in bold. The last line displays the detection rate for each pipeline. For the four MF-based pipelines, the detection rate determined by the FAR criteria is exhibited in black, while the detection rate determined by the $p_{\text{astro}}$ criteria is showcased in bold. The detection rate means the ratio of successfully detected events and the total events number listed in the table.

| Name | GstLAL | MBTA | PyCBC | PyCBC_BBH | MSNRnet | MSNRnet-2 | Time differences of MSNRnet-2(s) |
|---|---|---|---|---|---|---|---|
| GW150914 | 24.4 | | 23.6 | | 24.5 | 24.3 | 0.00439 |
| GW151012 | 10.0 | | 9.5 | | 10.6 | 11.6 | 0.00317 |
| GW151226 | 13.1 | | 13.1 | | 13.7 | 11.9 | 0.00293 |
| GW170104 | 13.0 | | 13.0 | | 14.4 | 13.5 | 0.00562 |
| GW170608 | 14.9 | | 15.4 | | 16.9 | 15.8 | 0.00024 |
| GW170729 | 10.8 | | 9.8 | | 11.6 | 10.7 | 0.00586 |
| GW170809 | 12.4 | | 12.2 | | 13.4 | 13 | 0.01196 |
| GW170814 | 15.9 | | 16.3 | | 17.1 | 16.4 | 0.0105 |
| GW170818 | 11.3 | | ... | | ... | ... | ... |
| GW170823 | 11.5 | | 11.1 | | 12.7 | 11.5 | 0.00513 |
| GW190403-051519 | ... | ... | **16.3** | **8.0** | ... | ... | ... |
| GW190408-181802 | 14.7 | 14.4 | 13.1 | 13.7 | 15.6 | 14.8 | 0.00049 |

*(Table continued)*

TABLE III. *(Continued)*

| Name | GstLAL | MBTA | PyCBC | PyCBC_BBH | MSNRnet | MSNRnet-2 | Time differences of MSNRnet-2(s) |
|---|---|---|---|---|---|---|---|
| GW190412 | 19.0 | 18.2 | 17.4 | 17.9 | 20.0 | 18.7 | 0.00586 |
| GW190413-052954 | … | … | … | 8.5 | 9.1 | … | … |
| GW190413-134308 | … | 10.3 | … | 8.9 | 9.9 | 9.9 | 0.00903 |
| GW190421-213856 | 10.5 | 9.7 | **10.1** | 10.1 | 11.2 | 10.2 | 0.00146 |
| GW190426-190642 | … | … | … | **9.6** | … | … | … |
| GW190503-185404 | 12.0 | 12.8 | 12.2 | 12.2 | 14.9 | 11.8 | 0.0061 |
| GW190512-180714 | 12.2 | 11.7 | 12.4 | 12.4 | … | 12.1 | 0.00732 |
| GW190513-205428 | 12.3 | 13.0 | … | 11.8 | 14.1 | 12.9 | 0.00098 |
| GW190514-065416 | … | … | … | **8.4** | … | … | … |
| GW190517-055101 | 10.8 | 11.3 | 10.4 | 10.3 | 11.1 | 10.6 | 0.00366 |
| GW190519-153544 | 12.4 | 13.7 | 13.2 | 13.2 | 15.1 | 14.7 | 0.00732 |
| GW190521 | 13.3 | 13.0 | 13.7 | 13.6 | … | … | … |
| GW190521-074359 | 24.4 | 22.2 | 24.0 | 24.0 | 24.4 | 24.3 | 0.00122 |
| GW190527-092055 | 8.7 | … | … | … | … | … | … |
| GW190602-175927 | 12.3 | 12.6 | 11.9 | 11.9 | 12.8 | 11.9 | 0.01294 |
| GW190701-203306 | 11.7 | **11.3** | 11.9 | 11.7 | 10.7 | 10.2 | 0.01392 |
| GW190706-222641 | 12.5 | 11.9 | 11.7 | 12.6 | 13.6 | 12.7 | 0.00537 |
| GW190707-093326 | 13.2 | 12.6 | 13.0 | 13.0 | 14.2 | 12.9 | 0.00732 |
| GW190719-215514 | … | … | … | 8.0 | 10.1 | 8.7 | 0.00269 |
| GW190720-000836 | 11.5 | 11.6 | 10.6 | 11.4 | … | … | … |
| GW190725-174728 | … | **9.8** | 9.1 | 8.8 | … | … | … |
| GW190727-060333 | 12.1 | 12 | 11.4 | 11.1 | 12.6 | 12.1 | 0.00537 |
| GW190728-064510 | 13.4 | 13.1 | 13.0 | 13 | 14.9 | 13.9 | 0.00195 |
| GW190731-140936 | 8.5 | **9.1** | … | 7.8 | 8.6 | 8.5 | 0.0105 |
| GW190803-022701 | 9.1 | **9.0** | … | 8.7 | 9.6 | 9.1 | 0.00195 |
| GW190805-211137 | … | … | … | 8.3 | … | … | … |
| GW190814 | 22.2 | 20.4 | 19.5 | … | 23.4 | 23.9 | 0.00928 |
| GW190828-063405 | 16.3 | 15.2 | 13.9 | 15.9 | 17.3 | 16.5 | 0.00317 |
| GW190828-065509 | 11.1 | 10.8 | 10.5 | 10.5 | 11.3 | … | … |
| GW190915-235702 | 13.0 | 12.7 | 13.0 | 13.1 | 13.6 | 13.2 | 0.00977 |
| GW190916-200658 | … | **8.2** | … | **7.9** | 9.2 | … | … |
| GW190917-114630 | 9.5 | … | … | … | … | … | … |
| GW190924-021846 | 13.0 | 11.9 | 12.4 | 12.5 | … | … | … |
| GW190926-050336 | 9.0 | … | … | … | 9.0 | 8.5 | 0.00342 |
| GW190929-012149 | 10.1 | **10.3** | … | … | … | … | … |
| GW190930-133541 | 10.1 | 10.0 | 9.8 | 10 | … | … | … |
| GW191103-012549 | … | … | **9.3** | 9.3 | … | … | … |
| GW191105-143521 | … | 10.7 | 9.8 | 9.8 | … | 10.3 | 0.00488 |
| GW191109-010717 | 15.8 | 15.2 | 13.2 | 14.4 | 15.8 | 15.8 | 0.00293 |
| GW191113-071753 | … | **9.2** | … | … | … | … | … |
| GW191126-115259 | … | … | … | **8.5** | … | … | … |
| GW191127-050227 | 10.3 | 9.8 | … | **8.7** | 9.5 | 8.9 | 0.00659 |
| GW191129-134029 | 13.3 | 12.7 | 12.9 | 12.9 | … | … | … |
| GW191204-110529 | … | … | … | **8.9** | 10.5 | … | … |
| GW191204-171526 | 15.6 | 17.1 | 16.9 | 16.9 | 17.3 | 16.7 | 0.00171 |
| GW191215-223052 | 10.9 | 10.8 | 10.3 | 10.2 | 11.3 | 10.7 | 0 |

*(Table continued)*

TABLE III. *(Continued)*

| Name | GstLAL | MBTA | PyCBC | PyCBC_BBH | MSNRnet | MSNRnet-2 | Time differences of MSNRnet-2(s) |
|---|---|---|---|---|---|---|---|
| GW191219-163120 | ··· | ··· | **8.9** | ··· | ··· | ··· | ··· |
| GW191222-033537 | 12.0 | 10.8 | 11.5 | 11.5 | 11.8 | 11.9 | 0.00366 |
| GW191230-180458 | 10.3 | ··· | ··· | 9.9 | 10.5 | 9.9 | 0.00171 |
| GW200115-042309 | 11.5 | 11.2 | 10.8 | ··· | ··· | ··· | ··· |
| GW200128-022011 | 10.1 | **9.4** | 9.8 | 9.9 | 10.1 | 9.8 | 0.00317 |
| GW200129-065458 | 26.5 | ··· | 16.3 | 16.2 | 26.9 | 26.1 | 0.00854 |
| GW200202-154313 | 11.3 | ··· | ··· | 10.8 | ··· | ··· | ··· |
| GW200208-130117 | 10.7 | 10.4 | 9.6 | 10.8 | ··· | 9.3 | 0.00122 |
| GW200208-222617 | ··· | ··· | ··· | **7.9** | ··· | ··· | ··· |
| GW200209-085452 | 10.0 | **9.7** | ··· | 9.2 | ··· | 9.3 | 0.00049 |
| GW200210-092254 | 9.5 | ··· | **8.9** | **8.9** | ··· | ··· | ··· |
| GW200216-220804 | 9.4 | ··· | ··· | **8.7** | 8.8 | 8.7 | 0.01221 |
| GW200219-094415 | 10.7 | 10.6 | 9.9 | 10 | 11.6 | 10.4 | 0.00684 |
| GW200220-061928 | ··· | ··· | ··· | **7.5** | ··· | ··· | ··· |
| GW200220-124850 | ··· | **8.2** | ··· | ··· | ··· | 7.8 | 0.00342 |
| GW200224-222234 | 18.9 | 19.0 | 19.2 | 18.6 | 19.6 | 18.0 | 0.00659 |
| GW200225-060421 | 12.9 | 12.5 | 12.3 | 12.3 | 14.5 | 13.3 | 0.00684 |
| GW200306-093714 | ··· | **8.5** | ··· | ··· | ··· | ··· | ··· |
| GW200308-173609 | ··· | ··· | ··· | **8** | ··· | ··· | ··· |
| GW200311-115853 | 17.7 | 16.5 | 17 | 17.4 | 16.7 | 16.1 | 0.00977 |
| GW200316-215756 | 10.1 | ··· | 9.3 | 9.3 | ··· | ··· | ··· |
| GW200322-091133 | ··· | **9.0** | ··· | **9.6** | ··· | ··· | ··· |
| Detection rate (%) | 73.8(**73.8**) | 51.4(**68.6**) | 57.5(**63.8**) | 65.7(**85.7**) | 60.0 | 61.3 | |

pipelines, which are clearly outlined in the last row of Table III. With a detection rate of 61.3%, MSNRnet-2 marginally outperforms the original MSNRnet, which had a detection rate of 60%. Furthermore, our findings indicate that MSNRnet, MSNRnet-2, and other MF-based pipelines exhibit comparable detection rates. GW search pipelines, such as PyCBC, GstLAL, and MBTA, which are based on matched filtering, operate in two distinct modes: "online" and "offline". The low-latency or "online" mode facilitates the swift detection of gravitational wave events, thereby enabling prompt communication to the external scientific community for electromagnetic follow-up observations. Conversely, the "offline" analysis takes place over extended timescales compared to the low-latency distribution of strain data. The confident events in GWTC-1, GWTC-2, GWTC-2.1, and GWTC-3 catalogs are obtained by the "offline" mode of these GW search pipelines. These results shown in Table III do not demonstrate MSNRnet-2's capability to perform either online or offline gravitational wave search tasks. They also necessitate long-term noise analysis outcomes. In Sec. III C of this section, we will utilize the proposed MSNRnet-2 to scan all data in O3b collected by the Hanford and Livingston interferometers.

We also conduct a comparison between the SNRs predicted by our proposed framework and those generated by traditional MF-based pipelines, as illustrated in Fig. 6. The SNRs from the traditional MF-based pipelines are derived by averaging multiple pipeline predictions. Figure 6 reveals that the majority of SNRs predicted by MSNRnet are marginally higher than the average predictions from traditional MF-based pipelines. Interestingly, the SNRs forecasted by MSNRnet-2 are scattered on both sides of the average value obtained from traditional MF-based pipelines.

To delve deeper into the disparities between our proposed approach and traditional matched filtering, we scrutinize the relative errors between the two methods. The results are exhibited in Fig. 7. This figure demonstrates that the proportion of SNRs with a relative error in range $(-0.1, 0.1)$ has escalated from 60% to 90%. This significant increase underscores the enhanced performance of MSNRnet-2 compared to MSNRnet.

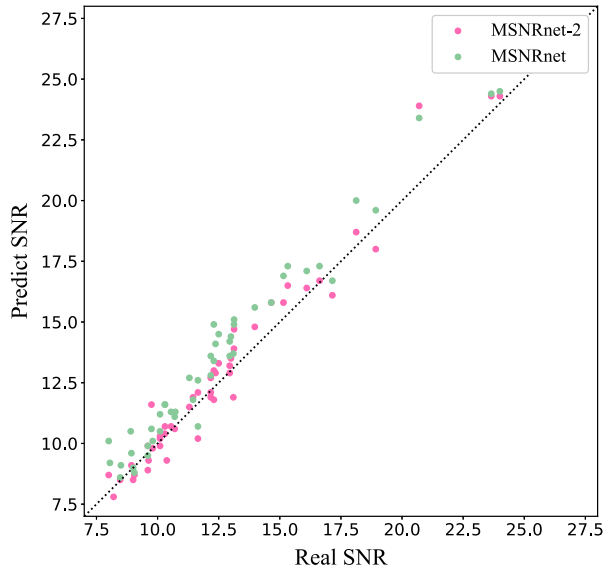


FIG. 6. The relationship between the SNR of the confident events predicted by traditional MF-based pipelines and the proposed framework. The results of MSNRnet and MSNRnet-2 are all shown. The real SNR is set to the average SNR value of the MF-base pipelines.

### D. Performance of the GW search framework on O3b data

In this subsection, we aim to assess the efficacy of our proposed framework by applying it to real strain data containing only noise. To this end, we utilize our framework to scan the entirety of the noisy strain data associated with O3b. This dataset specifically encompasses recordings obtained between 1 November 2019 at 15:00 Universal Time Coordicated (UTC) and 27 March 2020 at 17:00 UTC.

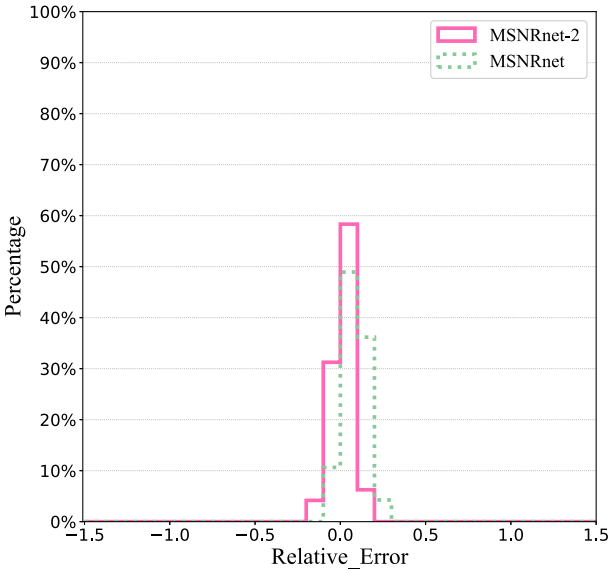


FIG. 7. The relative error of predicted SNR for the confident events.

Note that we only analyze the time when both Hanford and Livingston interferometers were normally working.

We utilize a moving time strategy to scan the data, with a time step of two seconds within the envelope extraction model's four-second operation window. Every segment that activates the envelope extraction model undergoes further analysis via denoising and astrophysical discrimination models. A false trigger is only generated if both the Hanford and Livingston interferometers are simultaneously triggered. Our proposed framework outputs a time series of SNRs, allowing for further verification through the peak SNR time difference between the two interferometers and the peak network SNR value. We have set the time difference threshold to 15 milliseconds and the network SNR threshold to eight. Table IV lists the number of false triggers for MSNRnet and MSNRnet-2 using various judgment strategies. Notably, in the "H + L + T + S" scenario, MSNRnet generates 16 false triggers, whereas MSNRnet-2 produces no false triggers. This remarkable result demonstrates the effectiveness of our proposed framework in processing noisy data over extended periods without generating false triggers.

Since the analysis of O3b background data does not yield any false triggers, a direct calculation of the FAR is not feasible. In our revised paper, we have adopted a time-shift approach to determine the FAR. Specifically, we shift in time the triggers obtained from one detector (after they have passed through the astrophysical discrimination model) relative to those from the second detector. Given that Hanford and Livingston cowork at O3b for roughly three months, each time shift generates an additional three months of triggers. Utilizing this method, we have generated approximately ten years of background data and corresponding triggers. Subsequently, we employ a coincident test to identify confident events, which in this context are false alarm events. This test is based on three criteria, which align with the standards for true alarms: Firstly, the time difference between the peak SNR at Hanford and Livingston must be less than 15 ms. Secondly, the predicted templates for Hanford and Livingston must share the same parameter range. Thirdly, the network SNR must exceed eight. Employing this method, we have generated seven false triggers from a ten-year data span. Based on this, we deduce that the FAR is approximately 0.7 per year.

### E. Comparison with the end-to-end classification framework for GW search

Recently, the AResGW [75] code emerged victorious in the first Machine-Learning Gravitational-Wave Mock Data Challenge (MLGWSC-1) [26] specifically for the real O3a noise scenario. AResGW stands as an end-to-end classification framework tailored for gravitational wave searches. Its architecture boasts deep adaptive input normalization for input preprocessing, coupled with a 54-layer deep residual network for feature extraction. Both the AResGWs code

TABLE IV. Number of false triggers generated while scanning noise in O3b data: "H + L" indicates triggers activated by both Hanford and Livingston. "H + L + T" signifies triggers from "H + L" that are further validated based on trigger time differences. "H + L + T + S" represents triggers from "H + L + T" that undergo additional verification through the network SNR threshold.

| | | MSNRnet | | | MSNRnet-2 | | |
|---|---|---|---|---|---|---|---|
| Mass1($M_\odot$) | Mass2($M_\odot$) | H + L | H + L + T | H + L + T + S | H + L | H + L + T | H + L + T + S |
| [5, 10] | [5, 10] | 0 | 0 | 0 | 0 | 0 | 0 |
| [5, 10] | [10, 20] | 2 | 1 | 1 | 1 | 0 | 0 |
| [5, 10] | [20, 40] | 8 | 1 | 0 | 0 | 0 | 0 |
| [5, 10] | [40, 80] | 3 | 0 | 0 | 4 | 1 | 0 |
| [10, 20] | [10, 20] | 6 | 0 | 0 | 1 | 0 | 0 |
| [10, 20] | [20, 40] | 9 | 1 | 1 | 2 | 0 | 0 |
| [10, 20] | [40, 80] | 49 | 9 | 7 | 0 | 0 | 0 |
| [20, 40] | [20, 40] | 20 | 3 | 2 | 14 | 0 | 0 |
| [20, 40] | [40, 80] | 50 | 11 | 3 | 0 | 0 | 0 |
| [40, 80] | [40, 80] | 17 | 7 | 2 | 11 | 4 | 0 |

and its trained weights are accessible through the public via [76].

Nonetheless, until now, no comprehensive evaluation has been conducted on all confident events using AResGW. In this subsection, we undertake a comparative analysis between AResGW and MSNRnet-2. Since the sampling frequency of the test dataset in MLGWSC-1 is 2048 Hz, whereas MSNRnet-2 requires an input sampling frequency of 4096 Hz, it is not possible to evaluate MSNRnet-2 using the MLGWSC-1 test dataset. We employ actual data encompassing confident events along with genuine noise (O3b) to conduct a comparative analysis of the two methods.

For the confident events analysis, we initially downsample the strain data of confident events, recorded when both Hanford and Livingston interferometers were operational, to 2048 Hz using the downsample program provided in the MLGWSC-1 code package. Subsequently, we process the 2048 Hz strain data of confident events through the improved_d4_model of AResGW to obtain the $p_{\text{score}}$ output. To evaluate its performance, we compare the number of false triggers generated at the same detection rate (61.3%) during the O3b observing run.

By scanning the $p_{\text{score}}$ output of confident events, we determine a threshold that matched the detection rate of MSNRnet-2 (61.3%). This threshold is found to be a $p_{\text{score}}$ of 0.9967. We then employ AResGW to analyze the entire O3b noise dataset, generating a false trigger whenever the $p_{\text{score}}$ exceeded 0.9967. To our analysis, false triggers occurring within 0.3 seconds of each other were considered as a single event.

Remarkably, our results reveal that 1177 false triggers are generated by AResGW, significantly exceeding the 16 and zero false triggers produced by MSNRnet and MSNRnet-2, respectively. These findings underscore the superiority of the denoising-discrimination-matched filtering framework compared to the end-to-end classification framework.

### F. Sensitive distance of the GW search framework

Existing studies exploring machine learning based GW detection algorithms favor traditional machine learning measures like receiver operator characteristic (ROC) curves as an evaluation metric [20,29,77]. This approach poses challenges in direct comparisons, as these metrics are inherently reliant on the parameter distributions of the tested events. In recent years, numerous studies have adopted the variation of sensitive distance with respect to the FAR as a key metric for evaluating the efficacy of GW search methodologies [25,26,78,79]. In this subsection, we compute the sensitive distance under the condition of zero false triggers for O3b data, which corresponds to a FAR of 0.7 per year.

The variation of sensitive volume with respect to the FAR can be determined by:

$$V(\mathcal{F}) = \int dx d\Lambda \epsilon(\mathcal{F}; x, \Lambda)\phi(x, \Lambda), \tag{11}$$

where $x$ represents the source location, $\Lambda$ stands for the source parameters, $\epsilon(\mathcal{F}; x, \Lambda)$ signifies the search efficiency at a given FAR $\mathcal{F}$, and $\phi(x, \Lambda)$ denotes the distribution of $x$ and $\Lambda$.

When injections are conducted uniformly across a volume up to a predefined maximum distance $(d_{\text{max}})$, Eq. (11) can be approximated as follows:

$$V(\mathcal{F}) \approx V(d_{\text{min}}) + [V(d_{\text{max}}) - V(d_{\text{min}})]\frac{N_{I,\mathcal{F}}}{N_I}, \tag{12}$$

where $V(d_{\text{max}}) = \frac{4}{3}\pi d_{\text{max}}^3$ is the volume of a sphere, $N_{I,\mathcal{F}}$ represents the number of injections that have been successfully detected at a given FAR of $\mathcal{F}$, and $N_I$ is the total number of injections.

Here, we provide a detailed description of the methodology for generating samples in order to calculate the

TABLE V. The parameter ranges and the corresponding sensitive distance at FAR $\approx 0.7$ per year.

| $M_1(M_\odot)$ | $M_2(M_\odot)$ | Sensitive distance(Mpc) |
|---|---|---|
| [5, 10] | [10, 20] | 555.300 |
| [5, 10] | [20, 40] | 731.598 |
| [5, 10] | [40, 80] | 857.753 |
| [10, 20] | [10, 20] | 602.886 |
| [10,20] | [20, 40] | 1156.696 |
| [10, 20] | [40, 80] | 1511.607 |
| [20, 40] | [20, 40] | 1821.274 |
| [20, 40] | [40, 80] | 2405.055 |
| [40, 80] | [40, 80] | 2764.233 |

sensitive volume. We use IMRPhenomD as the approximate model for the $h_+$ and $h_\times$ calculation.

Assuming the center of the earth is situated at the origin of the Cartesian coordinate system, its coordinates are designated as (0, 0, 0). Initially, we generate the location of the gravitational wave source, denoted as (x, y, z), by uniformly sampling within a cubic space defined by the range $x \in (-d_{\max}, d_{\max})$, $y \in (-d_{\max}, d_{\max})$, and $z \in (-d_{\max}, d_{\max})$. If the sampled location falls within the specified distance range, where the Euclidean distance $\sqrt{x^2+y^2+z^2}$ lies between $[d_{\min}, d_{\max}]$, we proceed to accept this position and subsequently convert it to spherical coordinates for further simulation of the corresponding waveform. By implementing these steps, we ensure that the distance and direction parameters are sampled uniformly across a volumetric range, spanning from $d_{\min}$ to $d_{\max}$, thereby guaranteeing a consistent distribution within the specified distance interval. In this subsection we set $d_{\min} = 530$ Mpc and $d_{\max} = 6834$ Mpc. Regarding other parameters, such as the BBH masses and spins, their distributions are identical to those generated in the training data.

For both the Hanford and Livingston interferometers, we extracted noise strain data spanning 1,600,000 seconds from the O3b stage. Afterwards, within each mass parameter space, we generated 100,000 sample signals and introduced them into the Hanford and Livingston interferometers. It is noteworthy that each signal, corresponding to a specific parameter range, is injected only one time, resulting in a cumulative total of 100,000 injections for each parameter range, so that approximately every 16 seconds, an injection is performed. The maximum detectable signal length in the parameter space is approximately two seconds, therefore injecting every 16 seconds guarantees that two consecutive signals do not interfere with each other.

The sensitive distance for the full parameter space where $M_1, M_2 \in (5M_\odot, 80M_\odot)$ is 2253.301 Mpc. In addition we have already calculated the sensitive distance for each parameter ranges. The results are displayed in Table V.

## IV. CONCLUSION AND DISCUSSION

In this study, we upgrade the framework MSNRnet to MSNRnet-2. We comprehensively evaluate the performance of our proposed method in comparison to traditional MF-based pipelines and our previous model, MSNRnet. The results obtained through rigorous experimentation provide compelling evidence of the superiority of our new approach.

Initially, we compare the overlap distribution and predicted SNRs for "signal + noise" case generated by MSNRnet-2, MSNRnet, and the traditional MF-based pipelines. The results demonstrate the superiority of the denoising step in MSNRnet-2. The confident event analysis shows that the proportion of SNRs with a relative error in range $(-0.1, 0.1)$ has increased significantly, jumping from 60% (MSNRnet) to 90% (MSNRnet-2). This substantial improvement indicates the enhanced precision of MSNRnet-2 compared to its predecessor, MSNRnet. We use the MSNRnet-2 to analyze five months of data (all the data of O3b that Hanford and Livingston interferometers cowork), and surprisingly in this case zero false triggers are generated. The denoising discrimination and matched filtering framework holds promising potential for both online and offline gravitational wave searches in the future.

## 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. This work was supported in part by the National Key Research and Development Program of China Grant No. 2021YFC2203001, in part by the NSFC (No. 11920101003 and No. 12021003) and the Natural Science Foundation of Jiangxi (No. 20224BAB211012).

[1] Benjamin 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] Richard 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] Richard 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] Gabriella Agazie, Akash Anumarlapudi, Anne M. Archibald, Zaven Arzoumanian, Paul T. Baker, Bence Bécsy, Laura Blecha, Adam Brazier, Paul R. Brook, Sarah Burke-Spolaor *et al.*, The nanograv 15 yr data set: Evidence for a gravitational-wave background, Astrophys. J. Lett. **951**, L8 (2023).

[6] J Antoniadis, P. Arumugam, S. Arumugam, S. Babak, M. Bagchi, A-S Bak Nielsen, C. G. Bassa, A. Bathula, A. Berthereau, M. Bonetti *et al.*, The second data release from the european pulsar timing array-III, search for gravitational wave signals, Astron. Astrophys. **678**, A50 (2023).

[7] Heng Xu, Siyuan Chen, Yanjun Guo, Jinchen Jiang, Bojun Wang, Jiangwei Xu, Zihan Xue, R. Nicolas Caballero, Jianping Yuan, Yonghua Xu *et al.*, Searching for the nano-hertz stochastic gravitational wave background with the chinese pulsar timing array data release I, Res. Astron. Astrophys. **23**, 075024 (2023).

[8] Daniel J. Reardon, Andrew Zic, Ryan M. Shannon, George B. Hobbs, Matthew Bailes, Valentina Di Marco, Agastya Kapur, Axl F. Rogers, Eric Thrane, Jacob Askew *et al.*, Search for an isotropic gravitational-wave background with the parkes pulsar timing array, Astrophys. J. Lett. **951**, L6 (2023).

[9] Pierre Auclair, David Bacon, Tessa Baker, Tiago Barreiro, Nicola Bartolo, Enis Belgacem, Nicola Bellomo, Ido Ben-Dayan, Daniele Bertacca, Marc Besancon *et al.*, Cosmology with the laser interferometer space antenna, Living Rev. Relativity **26**, 5 (2023).

[10] Einstein Telescope *et al.*, Science with the Einstein telescope: A comparison of different designs, J. Cosmol. Astropart. Phys. 07 (2023) 068.

[11] Varun Srivastava, Derek Davis, Kevin Kuns, Philippe Landry, Stefan Ballmer, Matthew Evans, Evan D. Hall, Jocelyn Read, and B. S. Sathyaprakash, Science-driven tunable design of cosmic explorer detectors, Astrophys. J. **931**, 22 (2022).

[12] Teng Zhang, Huan Yang, Denis Martynov, Patricia Schmidt, and Haixing Miao, Gravitational-wave detector for postmerger neutron stars: Beyond the quantum loss limit of the fabry-perot-michelson interferometer, Phys. Rev. X **13**, 021019 (2023).

[13] 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).

[14] 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).

[15] 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).

[16] 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).

[17] Tianyu Zhao, Ruijun Shi, Yue Zhou, Zhoujian Cao, and Zhixiang Ren, Dawning of a new era in gravitational wave data analysis: Unveiling cosmic mysteries via artificial intelligence–a systematic review, arXiv:2311.15585.

[18] Elena Cuoco, Jade Powell, Marco Cavaglià, Kendall Ackley, Michał Bejger, Chayan Chatterjee, Michael Coughlin, Scott Coughlin, Paul Easter, Reed Essick *et al.*, Enhancing gravitational-wave science with machine learning, Mach. Learn. **2**, 011002 (2020).

[19] 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).

[20] 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).

[21] Plamen G. Krastev, Kiranjyot Gill, V. Ashley Villar, and Edo Berger, Detection and parameter estimation of gravitational waves from binary neutron-star mergers in real LIGO data using deep learning, Phys. Lett. B **815**, 136161 (2021).

[22] XiLong Fan, Jin Li, Xin Li, YuanHong Zhong, and JunWei Cao, Applying deep neural networks to the detection and space parameter estimation of compact binary coalescence with a network of gravitational wave detectors, Sci. China Phys., Mech. Astron. **62**, 1 (2019).

[23] 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).

[24] Xiang-Ru Li, Wo-Liang Yu, Xi-Long Fan, and G. Jogesh Babu, Some optimizations on detecting gravitational wave using convolutional neural network, Front. Phys. **15**, 1 (2020).

[25] 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).
[26] Marlin B. Schäfer, Ondřej Zelenka, Alexander H. Nitz, He Wang, Shichao Wu, Zong-Kuan Guo, Zhoujian Cao, Zhixiang Ren, Paraskevi Nousi, Nikolaos Stergioulas *et al.*, First machine learning gravitational-wave search mock data challenge, Phys. Rev. D **107**, 023021 (2023).
[27] 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).
[28] Wei Wei, Asad Khan, E. A. Huerta, Xiaobo Huang, and Minyang Tian, Deep learning ensemble for real-time gravitational wave detection of spinning binary black hole mergers, Phys. Lett. B **812**, 136029 (2021).
[29] EA Huerta, Asad Khan, Xiaobo Huang, Minyang Tian, Maksim Levental, Ryan Chard, Wei Wei, Maeve Heflin, Daniel S. Katz, Volodymyr Kindratenko *et al.*, Accelerated, scalable and reproducible ai-driven gravitational wave detection, Nat. Astron. **5**, 1062 (2021).
[30] 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).
[31] Seiya Sasaoka, Naoki Koyama, Diego Dominguez, Yusuke Sakai, Kentaro Somiya, Yuto Omae, and Hirotaka Takahashi, Comparative study of 1d and 2d convolutional neural network models with attribution analysis for gravitational wave detection from compact binary coalescences, Phys. Rev. D **109**, 043011 (2024).
[32] Yuewei Zhang, Haiguang Xu, Meilin Liu, Cheng Liu, Yuanyuan Zhao, and Jie Zhu, Deep learning model based on a bidirectional gated recurrent unit for the detection of gravitational wave signals, Phys. Rev. D **106**, 122002 (2022).
[33] 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).
[34] 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).
[35] 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).
[36] 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).
[37] Takahiro S. Yamamoto, Sachiko Kuroyanagi, and Guo-Chin Liu, Deep learning for intermittent gravitational wave signals, Phys. Rev. D **107**, 044032 (2023).
[38] Christoph Dreissigacker, Rahul Sharma, Chris Messenger, Ruining Zhao, and Reinhard Prix, Deep-learning continuous gravitational waves, Phys. Rev. D **100**, 044009 (2019).
[39] Banafsheh Beheshtipour and Maria Alessandra Papa, Deep learning for clustering of continuous gravitational wave candidates, Phys. Rev. D **101**, 064009 (2020).
[40] Christoph Dreissigacker and Reinhard Prix, Deep-learning continuous gravitational waves: Multiple detectors and realistic noise, Phys. Rev. D **102**, 022005 (2020).
[41] Robert E. Colgan, K. Rainer Corley, Yenson Lau, Imre Bartos, John N. Wright, Zsuzsa Márka, and Szabolcs Márka, Efficient gravitational-wave glitch identification from environmental data through machine learning, Phys. Rev. D **101**, 102003 (2020).
[42] Tiago Fernandes, Samuel Vieira, Antonio Onofre, Juan Calderón Bustillo, Alejandro Torres-Forné, and José A. Font, Convolutional neural networks for the classification of glitches in gravitational-wave data streams, Classical Quantum Gravity **40**, 195018 (2023).
[43] Jianqi Yan, Alex P. Leung, and C. Y. Hui, On improving the performance of glitch classification for gravitational wave detection by using generative adversarial networks, Mon. Not. R. Astron. Soc. **515**, 4606 (2022).
[44] Yunan Wu, Michael Zevin, Christopher P. L. Berry, Kevin Crowston, Carsten Østerlund, Zoheyr Doctor, Sharan Banagiri, Corey B. Jackson, Vicky Kalogera, and Aggelos K. Katsaggelos, Advancing glitch classification in gravity spy: Multi-view fusion with attention-based machine learning for Advanced LIGO's fourth observing run, arXiv:2401.12913.
[45] Wei Wei and E. A. Huerta, Gravitational wave denoising of binary black hole mergers with deep learning, Phys. Lett. B **800**, 135081 (2020).
[46] 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).
[47] He Wang, Yue Zhou, Zhoujian Cao, Zong-Kuan Guo, and Zhixiang Ren, WaveFormer: transformer-based denoising method for gravitational-wave data, Mach. Learn. **5**, 015046 (2024).
[48] 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).
[49] Seiya Sasaoka, Yilun Hou, Kentaro Somiya, and Hirotaka Takahashi, Localization of gravitational waves using machine learning, Phys. Rev. D **105**, 103030 (2022).
[50] Alex Kolmus, Grégory Baltus, Justin Janquart, Twan Van Laarhoven, Sarah Caudill, and Tom Heskes, Fast sky localization of gravitational waves using deep learning seeded importance sampling, Phys. Rev. D **106**, 023032 (2022).
[51] Chayan Chatterjee, Manoj Kovalam, Linqing Wen, Damon Beveridge, Foivos Diakogiannis, and Kevin Vinsen, Rapid localization of gravitational wave sources from compact binary coalescences using deep learning, Astrophys. J. **959**, 42 (2023).
[52] Stephen R. Green, Christine Simpson, and Jonathan Gair, Gravitational-wave parameter estimation with autoregressive neural network flows, Phys. Rev. D **102**, 104057 (2020).

[53] Han-Shiang Kuo and Feng-Li Lin, Conditional noise deep learning for parameter estimation of gravitational wave events, Phys. Rev. D **105**, 044016 (2022).
[54] 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).
[55] Alvin J. K. Chua and Michele Vallisneri, Learning Bayesian posteriors with neural networks for gravitational-wave inference, Phys. Rev. Lett. **124**, 041102 (2020).
[56] Jonas Wildberger, Maximilian Dax, Stephen R. Green, Jonathan Gair, Michael Pürrer, Jakob H. Macke, Alessandra Buonanno, and Bernhard Schölkopf, Adapting to noise distribution shifts in flow-based gravitational-wave inference, Phys. Rev. D **107**, 084046 (2023).
[57] Quirijn Meijer, Melissa Lopez, Daichi Tsuna, and Sarah Caudill, Gravitational-wave searches for cosmic string cusps in Einstein telescope data using deep learning, Phys. Rev. D **109**, 022006 (2024).
[58] Wathela Alhassan, Tomasz Bulik, and Mariusz Suchenek, Detection of Einstein telescope gravitational wave signals from binary black holes using deep learning, Mon. Not. R. Astron. Soc. **519**, 3843 (2022).
[59] Tianyu Zhao, Ruoxi Lyu, He Wang, Zhoujian Cao, and Zhixiang Ren, Space-based gravitational wave signal detection and extraction with deep neural network, Commun. Phys. **6**, 212 (2023).
[60] 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).
[61] MengNi Chen, YuanHong Zhong, Yi Feng, Di Li, and Jin Li, Machine learning for nanohertz gravitational wave detection and parameter estimation with pulsar timing array, Sci. China Phys. Mech. Astron. **63**, 129511 (2020).
[62] David Shih, Marat Freytsis, Stephen R. Taylor, Jeff A. Dror, and Nolan Smyth, Fast parameter inference on pulsar timing arrays with normalizing flows, Phys. Rev. Lett. **133**, 011402 (2024).
[63] Mengfei Sun, Jin Li, Shuo Cao, and Xiaolin Liu, Deep learning forecasts of cosmic acceleration parameters from deci-hertz interferometer gravitational-wave observatory, Astron. Astrophys. **682**, A177 (2024).
[64] Leo P. Singer and Larry R. Price, Rapid bayesian position reconstruction for gravitational-wave transients, Phys. Rev. D **93**, 024013 (2016).
[65] CunLiang Ma, Sen Wang, Wei Wang, and Zhoujian Cao, Using deep learning to predict matched signal-to-noise ratio of gravitational waves, Phys. Rev. D **109**, 043009 (2024).
[66] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, Generative adversarial nets, in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2014).
[67] Xudong Mao, Qing Li, Haoran Xie, Raymond Y. K. Lau, Zhen Wang, and Stephen Paul Smolley, Least squares generative adversarial networks, in *Proceedings of the IEEE International Conference on Computer Vision* (Institute of Electrical and Electronics Engineers(IEEE), 2017), pp. 2794–2802.
[68] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C. Courville, Improved training of wasserstein gans, in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2017).
[69] Shu Hu, Yuezun Li, and Siwei Lyu, Exposing gan-generated faces using inconsistent corneal specular highlights, in *Proceedings of the ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, New York, 2021), pp. 2500–2504.
[70] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros, Image-to-image translation with conditional adversarial networks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Institute of Electrical and Electronics Engineers(IEEE), 2017), pp. 1125–1134.
[71] Ming Tao, Hao Tang, Fei Wu, Xiao-Yuan Jing, Bing-Kun Bao, and Changsheng Xu, Df-gan: A simple and effective baseline for text-to-image synthesis, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Institute of Electrical and Electronics Engineers(IEEE), 2022), pp. 16515–16525.
[72] Melissa Lopez, Vincent Boudart, Kerwin Buijsman, Amit Reza, and Sarah Caudill, Simulating transient noise bursts in LIGO with generative adversarial networks, Phys. Rev. D **106**, 023027 (2022).
[73] Tom Dooney, Lyana Curier, Daniel Tan, Melissa Lopez, Chris Van Den Broeck, and Stefano Bromuri, cdvgan: One flexible model for multi-class gravitational wave signal and glitch generation, arXiv:2401.16356.
[74] Jade Powell, Ling Sun, Katinka Gereb, Paul D. Lasky, and Markus Dollmann, Generating transient noise artefacts in gravitational-wave detector data with generative adversarial networks, Classical Quantum Gravity **40**, 035006 (2023).
[75] Paraskevi Nousi, Alexandra E. Koloniari, Nikolaos Passalis, Panagiotis Iosif, Nikolaos Stergioulas, and Anastasios Tefas, Deep residual networks for gravitational wave detection, Phys. Rev. D **108**, 024022 (2023).
[76] https://github.com/vivinousi/gw-detection-deep-learning
[77] Heming Xia, Lijing Shao, Junjie Zhao, and Zhoujian Cao, Improved deep learning techniques in gravitational-wave data analysis, Phys. Rev. D **103**, 024040 (2021).
[78] Ethan Marx, William Benoit, Alec Gunny, Rafia Omer, Deep Chatterjee, Ricco C. Venterea, Lauren Wills, Muhammed Saleem, Eric Moreno, Ryan Raikman *et al.*, A machine-learning pipeline for real-time detection of gravitational waves from compact binary coalescences, arXiv:2403.18661.
[79] Marlin B. Schäfer and Alexander H. Nitz, From one to many: A deep learning coincident gravitational-wave search, Phys. Rev. D **105**, 043003 (2022).