

# Identification of Stochastic Gravitational Wave Backgrounds from Cosmic String Using Machine Learning

Xianghe Ma,<sup>1,2</sup> Borui Wang,<sup>3</sup> Nan Yang,<sup>4</sup> Jin Li,<sup>1,2,\*</sup> Brendan McCane,<sup>5</sup>  
Mengfei Sun,<sup>1,2</sup> Jie Wu,<sup>1,2</sup> Minghui Zhang,<sup>6</sup> and Yan Meng<sup>1,7,†</sup>

<sup>1</sup>*Department of Physics, Chongqing University, Chongqing 401331, P.R. China*

<sup>2</sup>*Chongqing Key Laboratory for Strongly Coupled Physics,  
Chongqing University, Chongqing 401331, P.R. China*

<sup>3</sup>*Department of Earth and Sciences, Southern University of Science and Technology, Shenzhen 518055, P.R. China*

<sup>4</sup>*Department of Electronical Information Science and Technology, Xingtai University, Xingtai 054001, China*

<sup>5</sup>*School of Computing, University of Otago, Otago 9016, New Zealand*

<sup>6</sup>*Department of Physics, Southern University of Science and Technology, Shenzhen 518055, P.R. China*

<sup>7</sup>*Department of Physics, Xingtai University, Xingtai 054001, China*

(Dated: February 25, 2025)

Cosmic strings play a crucial role in enhancing our understanding of the fundamental structure and evolution of the universe, unifying our knowledge of cosmology, and potentially unveiling new physical laws and phenomena. The advent and operation of space-based detectors provide an important opportunity for detecting stochastic gravitational wave backgrounds (SGWB) generated by cosmic strings. However, the intricate nature of SGWB poses a formidable challenge in distinguishing its signal from the complex noise by some traditional methods. Therefore, we attempt to identify SGWB based on machine learning. Our findings show that the joint detection of LISA and Taiji significantly outperforms individual detectors, and even in the presence of numerous low signal-to-noise ratio (SNR) signals, the identification accuracy remains exceptionally high with 95%. Although our discussion is based solely on simulated data, the relevant methods can provide data-driven analytical capabilities for future observations of SGWB.

## I. INTRODUCTION

The LIGO/Virgo collaboration made a groundbreaking announcement with the first direct detection of gravitational waves, produced by the merger of binary black holes (BBHs) [1]. This discovery not only confirmed Einstein's general theory of relativity but also marked the dawn of gravitational wave astronomy. With continuous advancements in detector sensitivity, the number of observed gravitational wave events has steadily increased. Since the first detection in 2015 (GW150914), ground-based detectors have successfully recorded over 100 events, primarily in the frequency range of tens to hundreds of hertz. However, a significant gap still exists in the detection of low-frequency gravitational waves. To fill this gap, space-based gravitational wave detectors like LISA [2], the Taiji program [3], and the TianQin project [4] are being developed, offering new opportunities for low-frequency gravitational wave detection.

Space-based detectors are not only capable of detecting a wide range of compact binary coalescence (CBC) events [5], but they also serve as powerful tools for investigating stochastic gravitational wave backgrounds (SGWB) in the millihertz frequency band. The SGWB consists of a superposition of numerous independent gravitational wave signals from various directions in the sky, originating from a variety of sources across the universe. These

sources are classified into cosmological and astrophysical categories, spanning different epochs in the history of the universe.

The cosmological sources of the SGWB include primordial processes in the early universe, such as inflation [6] [7], cosmic strings [8–11], and phase transitions [12–15]. This paper primarily focuses on the SGWB generated by cosmic strings. The study of SGWB is crucial as it provides unique insights into the early universe, fundamental physics, and astrophysical populations and processes. Detecting and characterizing the SGWB can help validate inflation models, explore the physics of the early universe, and illuminate the population and evolution of compact objects over cosmic time [16–18].

Due to the inherent noise characteristics of gravitational wave detectors [19] [20], particularly the prevalent foreground noise in space-based detectors, identifying gravitational wave signals from the data is a significant challenge. Traditional signal identification methods, such as Bayesian inference [21] [22] and matched filtering [23] [24], have limitations when applied to the detection of SGWB. For example, Bayesian inference requires sufficient prior information, which is difficult to obtain for SGWB, and it involves considerable offline processing time. Matched filtering relies on correlating the received signal with known target templates; however, the stochastic gravitational wave background lacks defined waveform templates. As an alternative, we employ machine learning techniques to identify SGWB, offering a promising approach that complements traditional methods.

Machine learning has made significant progress in vari-

\* cqjinli1983@cqu.edu.cn

† 200221007@xttc.edu.cn

ous gravitational wave (GW) data analysis tasks, including GW detection [25–28], parameter estimation [29] [30], glitch classification [31–33], noise reduction [34] [35], and signal extraction [36–39]. In Ref. [40], Zhoujian Cao et al. applied machine learning techniques to identify SGWB within a general model and achieved promising results. In contrast, this study focuses on the SGWB generated by cosmic strings, considering a broader signal-to-noise ratio (SNR) range than the work in [40], and utilizes multi-detector joint observations with foreground noise.

This paper is organized as follows: Section 2 describes the data simulation process, including the simulation of SGWB signals, the sensitivity curves for LISA and Taiji, and the foreground noise. Section 3 provides a brief introduction to the machine learning network used in this study and discusses some key hyperparameters. Section 4 presents the results and analysis, including outcomes with detector noise and foreground noise. Finally, Section 5 concludes the paper and offers some remarks.

## II. DATA SIMULATION

Our simulated data are time-domain sequences consisting of six groups. The six groups refer to different kinds of noise ( $n_i(t)$ ,  $i=1,2,3,4,5,6$ ). In each group, the samples can be divided into two categories: one containing the SGWB from cosmic strings with noise, i.e.,  $h_{SGWB} + n_i(t)$ ; and the other only containing noise i.e.,  $n_i(t)$ .  $n_{i=1,2,3}(t)$  stands for noise, respectively:

- $n_1(t)$  is the instrumental noise from LISA detector,
- $n_2(t)$  is the instrumental noise from Taiji detector,
- $n_3(t)$  is the instrumental noise from LISA and Taiji detectors.

The third group  $h_{SGWB} + n_3(t)$  combines the data from the first two groups to simulate the effect of joint observation by LISA and Taiji. The remaining three groups are based on the joint observation instrumental noise with the additional foreground noises:

- $n_4(t)$  is  $n_3(t)$  plus the foreground noise from double white dwarfs,
- $n_5(t)$  is  $n_3(t)$  plus gravitational wave background noise generated by binary black holes (BBH) and binary neutron stars (BNS) observed by LIGO and Virgo,
- $n_6(t)$  is consist of  $n_4(t) + n_5(t)$ .

### A. The stochastic gravitational wave background signals generated by cosmic strings

We adopt the model proposed in [41–43] to describe the gravitational waves generated by cosmic strings, employing highly accurate analytical approximation formulas for

their representation [44][45]. The gravitational wave signals are influenced significantly by the cosmic string tension  $G\mu$ , which characterizes the size of the loops. The value of  $G\mu$  ranges from  $10^{-9}$  to  $10^{-7}$ , and we adopt a uniform distribution for sampling within each order of magnitude. We introduce a free parameter  $\alpha$  to denote the string loop size and define the total power emitted by the cosmic strings as  $\Gamma = 50$ . The value of  $\alpha$  ranges from  $10^{-3}$  to  $10^0$ , which are also sampled with a uniform distribution.

For the SGWB from cosmic strings, the contributions from cosmic string loops can be categorized into three periods:

- loops formed and decayed during the radiation era,
- loops formed during the radiation era and decayed during the matter era,
- loops formed during the matter era.

For loops formed and decayed in the radiation region, the form of stochastic gravitational wave background is given by

$$\Omega_{GW}^r(f) = \frac{128}{9} \pi A_r \Omega_r \frac{Gu}{\varepsilon_r} \left[ \left( \frac{f(1+\varepsilon_r)}{\frac{B_r \Omega_m}{\Omega_r} + f} \right)^{\frac{3}{2}} - 1 \right], \quad (1)$$

here

$$\varepsilon_r = \frac{\alpha}{\Gamma G\mu}, \quad A_r = \frac{\tilde{c}}{\sqrt{2}} F \frac{v_r}{\xi_r^3}, \quad B_r = \frac{2H_0 \Omega_r^{\frac{1}{2}}}{\nu_r \Gamma G\mu}, \quad (2)$$

The label  $r$  indicates the radiation era and in these equations  $v_r = 0.662$ ,  $\xi_r = 0.271$ ,  $\nu_r = 1/2$ ,  $F = 0.1$  and  $\tilde{c}$  is a phenomenological parameter which can be set as  $\tilde{c} = 0.23 \pm 0.04$  [46]. In our work, the evolution of the universe is assumed to follow a standard  $\Lambda$ CDM model, with its underlying parameters are  $H_0 = 100h \text{ km}/(\text{s} \cdot \text{Mpc})$ ,  $h = 0.678$ ,  $\Omega_r = 8.397 \times 10^{-5}$ ,  $\Omega_m = 0.308$ . For loops formed in the radiation region and decayed in the matter region, their contribution to SGWB has the following form

$$\begin{aligned} \Omega_{GW}^{rm}(f) = & 32\sqrt{3}\pi (\Omega_m \Omega_r)^{\frac{3}{4}} H_0 \frac{A_r}{\Gamma} \frac{(1+\varepsilon_r)^{\frac{3}{2}}}{f^{\frac{1}{2}} \varepsilon_r} \\ & \times \left\{ \frac{\left(\frac{\Omega_m}{\Omega_r}\right)^{\frac{1}{4}}}{\left(B_m \left(\frac{\Omega_m}{\Omega_r}\right)^{\frac{1}{2}} + f\right)^{\frac{1}{2}}} \left[ 2 + \frac{f}{B_m \left(\frac{\Omega_m}{\Omega_r}\right)^{\frac{1}{2}} + f} \right] \right. \\ & \left. - \frac{1}{B_m + f} \left[ 2 + \frac{f}{B_m + f} \right] \right\}, \quad (3) \end{aligned}$$

here

$$B_r = \frac{2H_0 \Omega_m^{\frac{1}{2}}}{\nu_m \Gamma G\mu}, \quad (4)$$

where  $\nu_m = 2/3$ . The contribution of loops generated in the matter period to the SGWB generation by cosmic strings is given by

$$\Omega_{GW}^m(f) = 54\pi H_0 \Omega_m^{\frac{3}{2}} \frac{A_m}{\Gamma} \frac{\varepsilon_m + 1}{\varepsilon_m} \frac{B_m}{f} \times \left\{ \frac{2B_m + f}{B_m(B_m + f)} - \frac{1}{f} \frac{2\varepsilon_m + 1}{\varepsilon_m(\varepsilon_m + 1)} + \frac{2}{f} \log \left( \frac{\varepsilon_m + 1}{\varepsilon_m} \frac{B_m}{B_m + f} \right) \right\}, \quad (5)$$

here

$$\varepsilon_m = \frac{\alpha}{\Gamma G u}, A_m = \frac{\tilde{c}}{\sqrt{2}} F \frac{v_m}{\xi_m^3}. \quad (6)$$

The label  $m$  indicates the matter era and in these equations  $v_m = 0.583$ ,  $\xi_m = 0.625$ . Therefore, the SGWB generated by cosmic strings can be well approximated as

$$\Omega_{GW}(f) = \Omega_{GW}^r(f) + \Omega_{GW}^{rm}(f) + \Omega_{GW}^m(f). \quad (7)$$

For the dimensionless energy density in GWs  $\Omega_{GW}$ , we need to convert to the frequency domain strain  $h(f)$  [47], given by

$$H_0^2 \Omega_{GW}(f) = \frac{8\pi^2}{3} f^4 |h(f)|^2. \quad (8)$$

We perform a Fourier transform and add random phases to convert the frequency domain signal to a time domain signal.

### B. Simulation of instrument noise

Leveraging the triangular geometry of LISA-like detectors, time-delay interferometry (TDI) techniques can be employed to combine phase differences with varying time delays, effectively canceling laser frequency noise [48, 49]. For simplicity, we assume that the instrument noise consists of two primary components: test mass acceleration noise and optical path length fluctuations. These noise sources are considered identical for each spacecraft. Given that the arm lengths are equal, the LISA instrument effectively forms an equilateral triangle [50].

For computational convenience, we adopt approximations of the gravitational wave (GW) response in the A, E, and T channels, as provided in [51]:

$$R_A^i(f) = R_E^i(f) = \frac{9}{20} |W^i(f)|^2 \left[ 1 + \left( \frac{f}{4f_i/3} \right)^2 \right]^{-1}, \quad (9)$$

and

$$R_T^i(f) \simeq \frac{1}{4032} \left( \frac{f}{f_i} \right)^6 |W^i(f)|^2 \left[ 1 + \frac{5}{16128} \left( \frac{f}{f_i} \right)^8 \right]^{-1}, \quad (10)$$

where  $i = LISA, Taiji$ ,  $W^i(f) = 1 - e^{-2if/f_i}$ , and for the LISA-like detector,  $f_i = c/(2\pi L_i)$ , with  $L_{LISA} = 2.5 \times 10^6$  km, and  $L_{Taiji} = 3 \times 10^6$  km.

We adopt the noise model outlined in the LISA Science Requirements Document [19, 52] and assume that the same model applies to the LISA-like detector, Taiji. This model presumes that the noise in all channels remains constant and identical [49], specifically accounting for two primary noise sources: acceleration noise and optical path disturbance noise, which are explicitly defined as follows [2, 53]:

$$N_{acc}^i(f) = \frac{N_a^i}{(2\pi f)^4} \left( 1 + \left( \frac{f_1}{f} \right)^2 \right) = \frac{\left( \sqrt{(\delta a_i)^2 / L_i} \right)}{(2\pi f)^4} \left( 1 + \left( \frac{f_1}{f} \right)^2 \right), \quad (11)$$

$$N_{op}^i(f) = N_o^i \left( \sqrt{(\delta x_i)^2 / L_i} \right)^2, \quad (12)$$

where

$$\sqrt{(\delta a_{LISA})^2} = 3 \times 10^{-15} m/s^2, \quad (13)$$

$$\sqrt{(\delta x_{LISA})^2} = 1.5 \times 10^{-11} m, \quad (14)$$

and [54]

$$\sqrt{(\delta a_{Taiji})^2} = 3 \times 10^{-15} m/s^2, \quad (15)$$

$$\sqrt{(\delta x_{Taiji})^2} = 8 \times 10^{-12} m, \quad (16)$$

where  $i = LISA, Taiji$ ,  $f_1 = 0.4$  mHz. These noise models can be transformed into interferometer noise through

$$N_X^i(f) = [4N_{op}^i(f) + 8[1 + \cos^2(f/f_i)] N_{acc}^i(f)] |W^i(f)|^2, \quad (17)$$

$$N_{XY}^i(f) = -[2N_{op}^i(f) + 8N_{acc}^i(f)] \cos(f/f_i) |W^i(f)|^2. \quad (18)$$

The noise models for the power spectral densities in the AET channels are obtained by diagonalizing the covariance matrix of the XYZ channels. The resulting diagonal entries are given by:

$$N_A^i(f) = N_E^i(f) = N_X^i(f) - N_{XY}^i(f), \quad (19)$$

$$N_T^i(f) = N_X^i(f) + 2N_{XY}^i(f). \quad (20)$$

The noise spectral density for different channels can be derived from the noise power spectral density and the corresponding response function, expressed as:

$$S_j^i(f) = \frac{N_j^i(f)}{R_j^i(f)}, \quad (21)$$

where  $j = A, E, T$ . Based on the noise spectral density, the total equivalent energy density for a single LISA-like detector can be formulated as follows [51]:

$$\Omega^i = \frac{4\pi^2 f^3}{3H_0^2} \left( \sum_{j=A,E,T} \frac{R_j^i(f)}{N_j^i(f)} \right)^{-1}. \quad (22)$$

### C. Simulation of foreground noise

In actual observations, the data consists not only of gravitational wave signals generated by cosmic strings but also of astrophysical foreground noise. This foreground noise primarily includes two components: double white dwarf (DWD) and inspiral binary black holes (BBH)/binary neutron stars (BNS) based on the observations of LIGO and Virgo [55–57].

The gravitational wave model for double white dwarfs is a modulated signal based on the LISA orbital motion. Its energy spectral density can be approximated using the broken power-law model proposed in Refs. [48, 56], which is given by:

$$\Omega_{DWD}(f) = \frac{A_1 \left( \frac{f}{f_i} \right)^{\alpha_1}}{1 + A_2 \left( \frac{f}{f_i} \right)^{\alpha_2}}, \quad (23)$$

where  $A_1 = 7.44 \times 10^{-14}$ ,  $A_2 = 2.96 \times 10^{-7}$ ,  $\alpha_1 = -1.98$ ,  $\alpha_2 = -2.6$ .

For the superposition of gravitational wave background produced by inspiralling BBHs/BNS observed by LIGO and Virgo, it is given by

$$\Omega_{astro}(f) = \Omega_{astro} \left( \frac{f}{f_*} \right)^{\alpha_{astro}}, \quad (24)$$

where  $f_* = 3mHz$ ,  $\Omega_{astro} = 4.44 \times 10^{-12}$  and  $\alpha_{astro} = 2/3$ .

In Figure 1, we present the sensitivity curves of the detectors, the foreground noise, and the relationship between the equivalent energy density of the SGWB. The equivalent energy density of the noise is converted into power spectral density (PSD), and Gaussian noise in the time domain is generated to simulate both instrumental and foreground noise. The equivalent energy density of the DWD signal is determined based on the arm length of the LISA detector.

The sensitivity curves of LISA and Taiji are quite similar; however, Taiji's longer arm length results in lower instrumental noise. In the joint observation of LISA and Taiji, SGWB signals produced by cosmic strings are assumed to be long-lasting. Therefore, it is unlikely that one detector would observe the signal while the other does not.

In Figure 2, we present example samples of the generated data. As shown, the SGWB behaves like noise in the time domain, which complicates the process of signal identification. In Figure 3, we illustrate their SNR

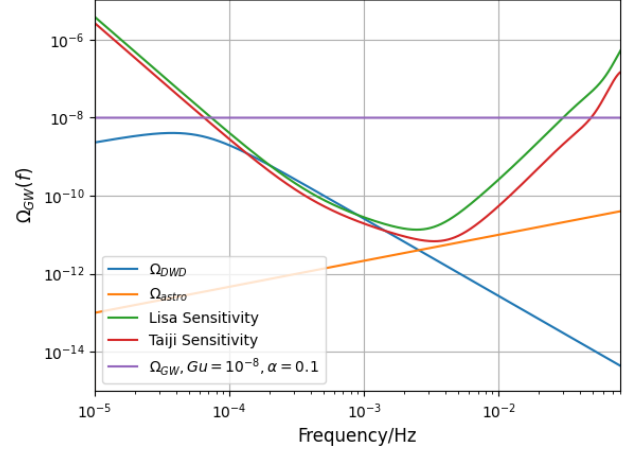


FIG. 1: The equivalent energy density of LISA, Taiji, foreground noise, SGWB.

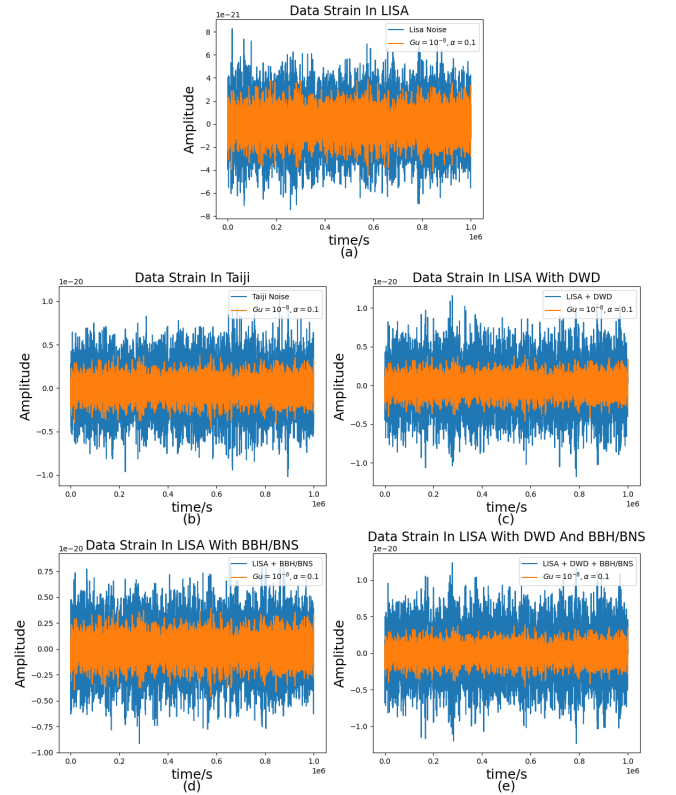


FIG. 2: Some data samples. (a) shows the instrumental noise of LISA and the SGWB in the time-domain data, while (b) illustrates the instrumental noise of Taiji. (c) shows the data with double white dwarf (DWD) noise, (d) shows the data with inspiral BBH/BNS based on the observations of LIGO and Virgo, and (e) shows the data with both types of foreground noise.

distributions. For the same data, the minimum SNR of LISA is 13, while Taiji achieves a minimum SNR of 19. Similarly, the maximum SNR for Taiji is 122.2, whereas LISA only reaches a maximum SNR of 83. This dif-

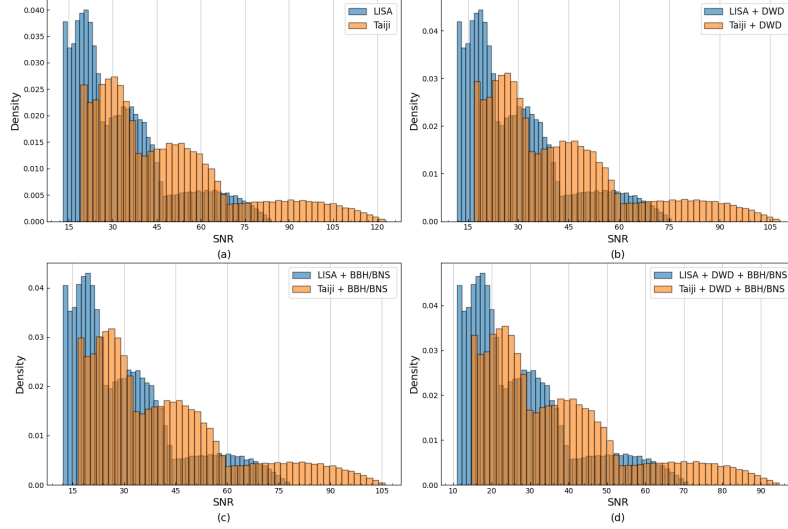


FIG. 3: The SNR distributions of our samples. The blue column represents the SNR for the LISA detector, while the orange column represents the SNR for the Taiji detector. (a) shows the SNR with only the detector’s intrinsic noise; (b) shows the data with DWD foreground noise on top of (a); (c) shows the data with inspiral BBH/BNS based on the observations of LIGO and Virgo foreground noise on top of (a); and (d) shows the data with both DWD and inspiral BBH/BNS based on the observations of LIGO and Virgo foreground noises.

ference arises because Taiji, with its longer arm lengths compared to LISA, has lower instrumental noise, resulting in a higher SNR. Additionally, the SNR distribution for LISA is more concentrated. However, the presence of foreground noise adds complexity and chaos to the data, further increasing the difficulty of signal identification. The SNR is calculated using the formula provided in [58].

$$\rho = 2 \left( \int_{f_{\min}}^{f_{\max}} \frac{|h(f)|^2}{S(f)} df \right)^{\frac{1}{2}}, \quad (25)$$

where  $S(f)$  is the PSD of the noise and  $h(f)$  is the signal.

### III. DEEP LEARNING

We employed a Residual Shrinkage Network (ResNet) for this learning task, incorporating Dropout and regularization to mitigate overfitting [59–61]. The ResNet is a deep learning model that combines residual networks with soft thresholding, aiming to enhance robustness against high-noise data. It is particularly well-suited for processing time-domain signal data. Figure 4 illustrates the streamlined workflow of the network we utilized. The architecture of our network is characterized by the following key features: residual connections [62–64], soft thresholding [65–67], regularization [68, 69], Dropout [70, 71], and multi-scale feature extraction [72].

Each residual block includes a skip connection [73, 74], which directly adds the input to the output. Skip connections help alleviate the vanishing gradient problem and enable the network to more effectively learn identity mappings. Soft thresholding is integrated into each residual

block, where an adaptive threshold is computed based on the global average pooling of the mean absolute value of the input features. This operation effectively filters out noise, significantly enhancing signal recognition performance, particularly in low signal-to-noise ratio (SNR) scenarios.

Our loss function is Focal Loss [75][76], a weighted cross-entropy loss function designed to enhance the model’s emphasis on samples that pose greater classification challenges, which is

$$L_{Focal} = -\frac{1}{N} \sum_{i=1}^N \alpha^{(i)} \cdot \left(1 - p_t^{(i)}\right)^{\gamma} \cdot \log \left(p_t^{(i)}\right) \quad (26)$$

where,  $\alpha^{(i)}$  represents the weighting coefficient for each sample, and  $p_t^{(i)}$  denotes the predicted probability for the sample, expressed as follows:

$$\alpha^{(i)} = y_i \alpha^* + (1 - y_i) (1 - \alpha^*), \quad (27)$$

and

$$p_t^{(i)} = y_i \hat{y}_i + (1 - y_i) (1 - \hat{y}_i), \quad (28)$$

where,  $y_i$  represents the ground truth label,  $\hat{y}_i$  denotes the predicted probability for the positive class,  $\alpha^*$  and  $\gamma$  are hyperparameters, and  $N$  indicates the total number of samples.

To mitigate the complexity of training and enhance overall efficiency, we employed a curriculum-based learning approach [77–79], which emphasizes a learning sequence from simple to complex. Specifically, this method involves adjusting the order of training samples based on

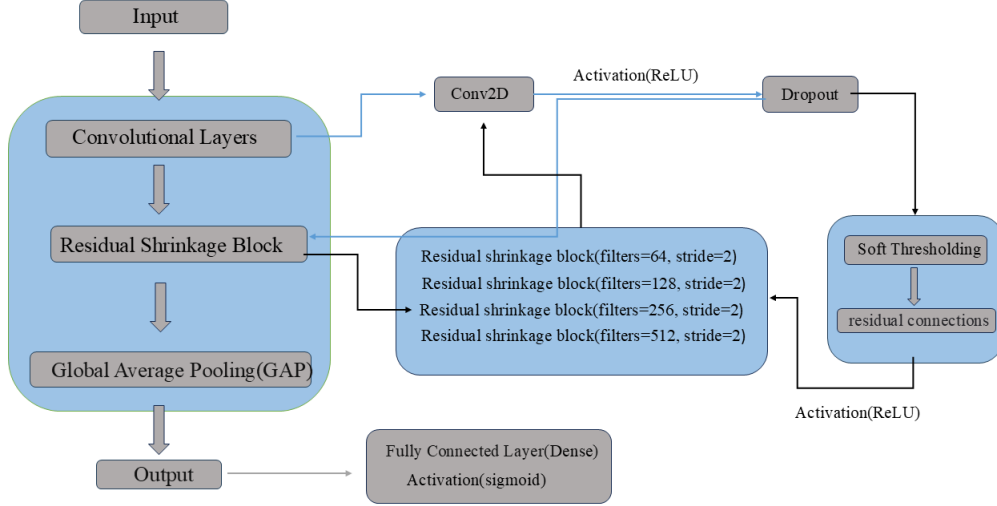


FIG. 4: The model architecture Schematic representation of our machine learning model. We employed a ResNet architecture and incorporated Dropout and regularization at appropriate stages. An initial convolutional layer is added before the residual blocks to activate the input data and extract preliminary features. The residual block consists of four stacked Residual Shrinkage Blocks (RSBs), each containing two convolutional layers. Dropout is applied between the two convolutional layers, followed by soft-thresholding to suppress noise and highlight important features. A global average pooling layer is then used to compress the high-dimensional features into low-dimensional vectors, facilitating subsequent classification. Finally, a fully connected Dense layer with a sigmoid activation function outputs the binary classification probabilities.

their difficulty, starting with simpler or more easily understood samples and gradually introducing more complex ones. This approach helps the model better learn intricate patterns, thereby improving both training efficiency and performance.

In our study, the dataset consists of 200,000 samples, with an equal number of positive and negative samples. The positive samples are sorted by SNR in descending order. Starting from the highest SNR, every 20,000 positive samples, along with an equal number of negative samples, are selected together to form a subset, resulting in five subsets. Each subset thus contains 20,000 positive samples and 20,000 negative samples. For each subset, 20% of the samples are randomly selected as the validation set, while the remaining 80% are used for training.

Correspondingly, the training process is divided into five stages. In the first stage, only the subset with the highest SNR is used for training. In the second stage, the subset with a slightly lower SNR is added to the training set. This progression continues through the third and fourth stages. Finally, in the fifth stage, the entire training dataset is used for training [78, 79]. This progressive training approach allows the model to initially learn fundamental features from high-SNR samples, gradually adapting to the complexities introduced by low-SNR samples, thereby enhancing overall learning performance. Table 1 presents some of the hyperparameters used in the network.

TABLE I: The hyperparameters of the network.

optimizer	Adam
learning rate	0.0003/0.0003/0.0003/0.0001/0.0001
loss	Focal Loss( $\alpha^*=0.25/0.25/0.2/0.2/0.2$ )
loss	Focal Loss( $\gamma=1/2/2/2/2$ )
epochs	10/20/20/30/30
batch size	64/64/64/32/32

#### IV. RESULTS AND ANALYSIS

An early stopping strategy was employed during the training process, which involved terminating the current phase of training and moving to the next phase when the loss showed minimal change over five consecutive epochs. In the first phase, the loss for all three models decreased rapidly to a low value. As each new phase introduced additional samples with lower SNR for training, a noticeable increase in loss occurred at the beginning of each phase. The curriculum learning strategy, which leverages high-SNR samples to establish a robust initial framework, ensures that the loss variations in subsequent phases remain moderate.

We analyzed the performance of individual detectors and joint detection in identifying the stochastic gravitational wave background separately. The results show that joint detection outperforms individual detectors, with Taiji achieving better results than LISA. When foreground noise is included, the detection accuracy decreases for both individual detectors and joint detection.

However, joint detection still demonstrates superior performance compared to individual detectors. Additionally, in the cases with foreground noise, we analyzed how detection accuracy varies with SNR.

### A. Machine learning performance with instrumental noise

The ROC curve and confusion matrix are commonly used to evaluate the performance of machine learning models. In this section, we focus solely on the data containing the SGWB and instrumental noise. As shown in Figure 5, the ROC curve for the LISA detector has the smallest area, with an AUC value of 0.9453. In contrast, machine learning with the Taiji detector and the joint detection of both detectors achieved AUC values exceeding 0.98, demonstrating excellent detection performance.

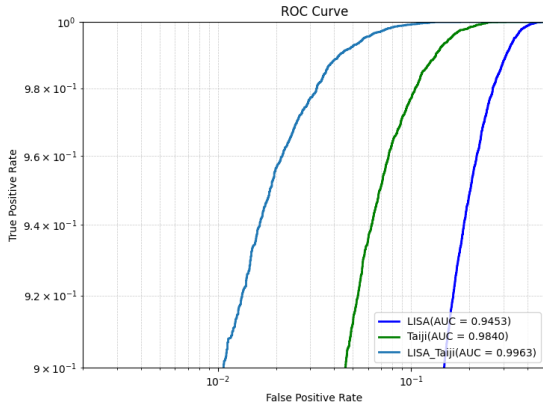


FIG. 5: The ROC curves for LISA, Taiji, and the joint detection of both detectors under instrument noise.

As shown in Figure 6, the majority of misclassifications involve incorrectly identifying signal-containing samples as noise-only samples, with this type of error occurring more than twice as frequently as the other type of misclassification. This is primarily due to the fact that the physical behavior of the SGWB closely resembles that of noise. The detection accuracies for LISA, Taiji, and joint detection are 87.60%, 94.11%, and 97.31%, respectively. It is clear that joint detection outperforms the individual detectors in SGWB identification.

As shown in Figure 7, for LISA, a significant increase in loss is observed in the third phase; for Taiji, this increase becomes evident in the fourth phase; and for joint detection, the increase is only noticeable in the fifth phase. Furthermore, the final loss for joint detection is considerably lower than that for the individual detectors.

By analyzing the ROC curves, confusion matrices, and loss reduction plots, we conclude that joint detection outperforms individual detectors in identifying and detecting SGWB signals. Consequently, joint detection is likely

to be a more effective and feasible approach for SGWB signal detection.

### B. Machine learning performance with foreground noise

In our previous analysis, we concluded that joint detection provides higher accuracy compared to using a single detector. Therefore, the analysis in this section is based on joint detection. We further investigated the impact of SNR ranges on machine learning performance, considering the DWD and inspiraling BBH/BNS foreground noise based on the observations of LIGO and Virgo. Specifically, we use several datasets with different SNR ranges: 200,000 samples (SNR range: 11 to 71), 180,000 samples (SNR range: 13.5 to 71), 160,000 samples (SNR range: 15.8 to 71), 140,000 samples (SNR range: 18 to 71), and 120,000 samples (SNR range: 20 to 71). Since the SNR for the same positive sample can be correspondingly obtained in both the Taiji and LISA detectors, we refer to the SNR values under the LISA detector for joint detection throughout this analysis, in order to avoid any confusion.

As shown in Figures 8 and 9, when the recognition accuracy exceeds 95%, the corresponding AUC of the ROC curve consistently remains above 0.99. Figure 8 underscores the importance of selecting appropriate SNR ranges for evaluating model performance, and shows how AUC values vary across different SNR intervals. Notably, within the SNR range of 15.8 to 71, the AUC values consistently exceed 0.99, indicating superior model performance in this range. Figure 9 further illustrates the recognition accuracy across different SNR ranges. Starting from the lowest SNR range, the recognition accuracies are 90.07%, 93.36%, 95.76%, 97.45%, and 99.00%, respectively. These results demonstrate that an SNR range of 15.8 to 71 provides a solid foundation for feature learning, thereby enhancing both recognition performance and generalization capability.

However, this does not imply that all samples with SNRs greater than 15.8 can be classified correctly. While high-SNR samples provide a solid foundation for feature learning and classification, they may still be misclassified due to other factors, such as noise complexity or subtle variations in signal characteristics. Furthermore, samples with SNRs below 15.8 pose a greater challenge to the model, as their signal features become more indistinguishable from noise, resulting in a higher misclassification rate.

These results indicate that within the SNR range of 15.8–71, even with the presence of foreground noise, machine learning with joint detection can still effectively identify SGWB signals in the data. As shown in Figure 10, after adding the two types of foreground noise, only a small fraction of the samples remain unrecognized. For parameter ranges where  $Gu \geq 2 \times 10^8$  and  $\alpha \geq 0.1$ , the samples can be correctly classified by our machine learn-



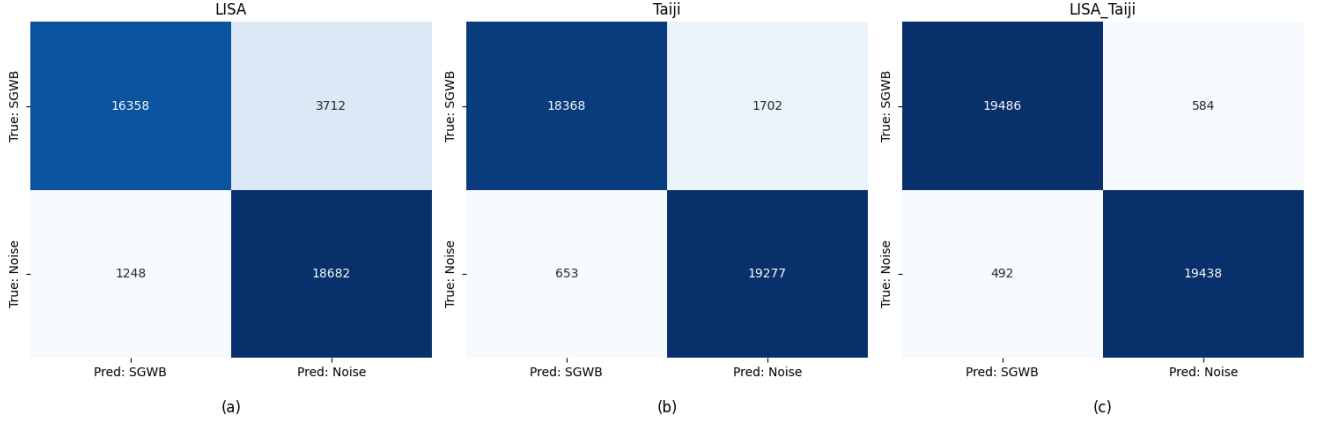


FIG. 6: The confusion matrices to the results of LISA, Taiji, and joint detection, with detection accuracies of 87.60%, 94.11%, and 97.31%, respectively.

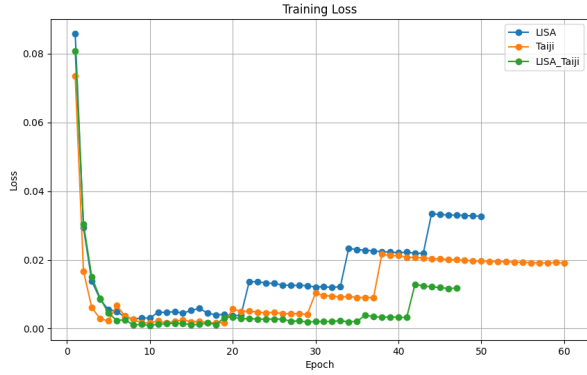


FIG. 7: The loss variations with epochs for the LISA, Taiji, and joint detection models. At lower SNRs, the loss for joint detection is consistently smaller than that of the individual detectors. The significant increase in loss at the beginning of a new phase occurs due to adding the lower SNR samples

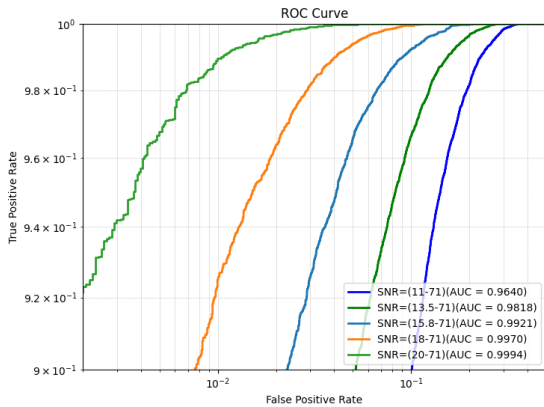


FIG. 8: ROC curve for different SNR.

ing model. This suggests that SGWB signals from cosmic

strings with these parameter values may be detectable using machine learning in future space-based gravitational wave joint detection efforts.

With this SNR range (15.8–71), we evaluated the performance of machine learning with three types of foreground noise: only DWD noise, only inspiral BBH/BNS noise based on LIGO and Virgo observations, and a mixture of both noise types. The total number of samples was reduced from 200,000 to 160,000, with positive and negative samples equally distributed. Figures 11 and 12 show that the AUC of the ROC curves for all three cases exceeds 0.99, and the accuracy remains above 95%. The best results are from inspiraling BBH/BNS noise based on LIGO and Virgo observations with 99.07% accuracy. For the case with only DWD foreground noise, the accuracy reaches 97.56%. Even with the simultaneous addition of both types of foreground noise, the model still achieves an accuracy of 95.76%.

In summary, the results demonstrate that restricting the SNR range (e.g., selecting samples with SNRs between 15.8 and 71) significantly enhances the model's overall performance. This underscores the critical role of high-SNR samples in improving model accuracy, while highlighting the necessity of future efforts to refine algorithms and models to better handle low-SNR samples in complex signal-noise environments.

## V. SUMMARY AND DISCUSSION

In our research, the computational configuration consists of an NVIDIA RTX 3090 GPU and an Intel(R) Xeon(R) Silver 4214R CPU (2.40 GHz, 48-thread processor). We employed full analytical approximations to simulate the SGWB generated by cosmic strings and modeled the instrumental noise of both the LISA and Taiji detectors based on their sensitivity curves. By leveraging residual shrinkage networks and curriculum learning strategies, we developed a robust model and



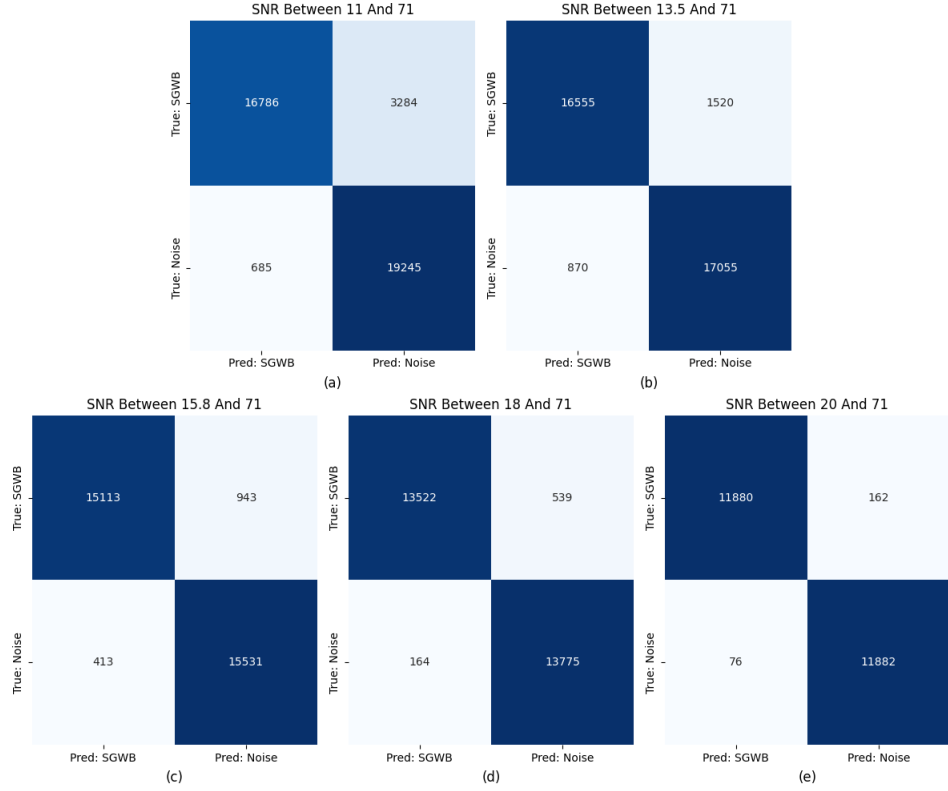


FIG. 9: Confusion matrix for different SNR. (a) shows a SNR between 11 and 71 with 90.07% accuracy. (b) shows a SNR between 13.5 and 71 with 93.36% accuracy. (c) shows a SNR between 15.8 and 71 with 95.76% accuracy. (d) shows a SNR between 18 and 71 with 97.45% accuracy. (e) shows a SNR between 20 and 71 with 97.45% accuracy.

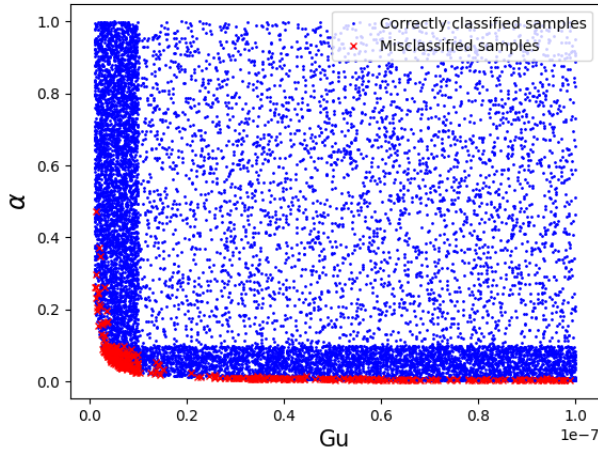


FIG. 10: The classification in the parameter space. Blue dots represent the samples correctly classified by the model, while red crosses represent the misclassified samples, which corresponds to Figure 9(c).

demonstrated that multi-detector joint observations significantly enhance the detection of SGWB signals.

In traditional machine learning and deep learning training, training samples are typically shuffled, and the

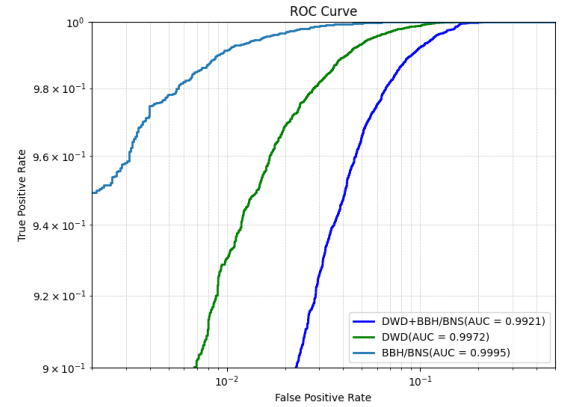


FIG. 11: ROC curve with different foreground noises in the joint detection. The analysis includes the simultaneous injection of the aforementioned two types of foreground noise, using samples with an SNR range of 15.8 to 71.

model encounters samples of varying difficulty throughout the process, with these difficulties being uniformly distributed. In contrast, curriculum learning mimics human learning by gradually guiding the model through the learning process according to the difficulty of the samples, enabling it to acquire knowledge more efficiently.

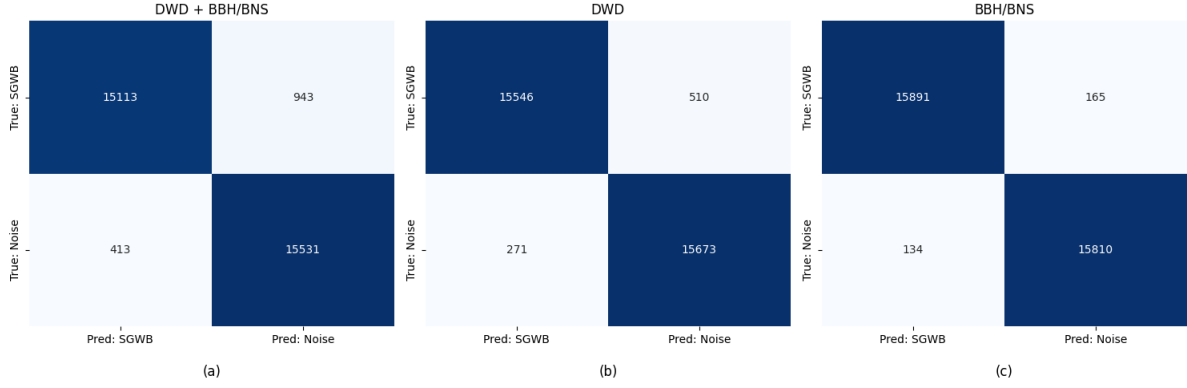


FIG. 12: Confusion matrix for different foreground noise. The sample selection is consistent with Figure 11. (a) shows data with DWD foreground noise and inspiral BBH/BNS noise based on LIGO and Virgo observations. (b) shows data with DWD foreground noise. (c) shows data with inspiral BBH/BNS noise based on LIGO and Virgo observations.

These results suggest that machine learning holds great promise for the effective detection of SGWB signals by space-based gravitational wave detectors. Furthermore, combining multiple detectors could further improve the performance of machine learning models, enabling more efficient and accurate identification of SGWB signals.

Moreover, foreground noise presents a significant challenge for space-based detectors, as many of the received data streams are likely to be substantially affected by such noise. To address this, we incorporated foreground noise into our study, including contributions from DWD and inspiraling BBH/BNS, based on the observations of LIGO and Virgo. By analyzing the model's performance across different SNR ranges, we found that our machine learning model can effectively handle the challenges posed by foreground noise, providing a solid foundation for future SGWB signal detection efforts.

Certainly, the actual data from future detectors will be more complex than the data discussed in this paper, and the task of utilizing machine learning models for this analysis will become even more challenging. However, the discussions presented here offer a meaningful attempt and valuable reference for future space-based gravitational wave data analysis powered by machine learning.

## ACKNOWLEDGEMENTS

This work was supported by the National Key Research and Development Program of China (Grant No. 2021YFC2203004), the National Natural Science Foundation of China (Grant No. 12347101), the Natural Science Foundation of Chongqing (Grant No. CSTB2023NSCQ-MSX0103).

- 
- [1] B. P. Abbott, R. Abbott, T. Abbott, M. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, *et al.*, Observation of gravitational waves from a binary black hole merger, *Physical Review L* **116**, 061102 (2016).
  - [2] P. Amaro-Seoane, H. Audley, S. Babak, J. Baker, E. Barausse, P. Bender, E. Berti, P. Binetruy, M. Born, D. Bortoluzzi, *et al.*, Laser interferometer space antenna, *arXiv preprint arXiv:1702.00786* (2017).
  - [3] Z. Luo, Z. Guo, G. Jin, Y. Wu, and W. Hu, A brief analysis to taiji: Science and technology, *Results in Physics* **16**, 102918 (2020).
  - [4] J. Luo, L.-S. Chen, H.-Z. Duan, Y.-G. Gong, S. Hu, J. Ji, Q. Liu, J. Mei, V. Milyukov, M. Sazhin, *et al.*, Tianqin: a space-borne gravitational wave detector, *Classical and Quantum Gravity* **33**, 035010 (2016).
  - [5] M. Pitkin, S. Reid, S. Rowan, and J. Hough, Gravitational wave detection by interferometry (ground and space), *Living reviews in relativity* **14**, 1 (2011).
  - [6] M. S. Turner, Detectability of inflation-produced gravitational waves, *Physical Review D* **55**, R435 (1997).
  - [7] M. C. Guzzetti, N. Bartolo, M. Liguori, and S. Matarrese, Gravitational waves from inflation, *La Rivista del Nuovo Cimento* **39**, 399 (2016).
  - [8] S. Sarangi and S.-H. H. Tye, Cosmic string production towards the end of brane inflation, *Physics Letters B* **536**, 185 (2002).
  - [9] T. Damour and A. Vilenkin, Gravitational wave bursts from cosmic strings, *Physical Review Letters* **85**, 3761 (2000).
  - [10] M. J. Stott, T. Elghozi, and M. Sakellariadou, Gravitational wave bursts from cosmic string cusps and pseudo-cusps, *Physical Review D* **96**, 023533 (2017).
  - [11] X. Siemens, V. Mandic, and J. Creighton, Gravitational-wave stochastic background from cosmic strings, *Physical Review Letters* **98**, 111101 (2007).
  - [12] E. Witten, Cosmic separation of phases, *Physical Review D* **30**, 272 (1984).
  - [13] A. Kosowsky, M. S. Turner, and R. Watkins, Gravitational waves from first-order cosmological phase transitions, *Physical review letters* **69**, 2026 (1992).

- [14] P. B. Dev and A. Mazumdar, Probing the scale of new physics by advanced ligo/virgo, *Physical Review D* **93**, 104001 (2016).
- [15] B. Von Harling, A. Pomarol, O. Pujolas, and F. Rompineve, Peccei-quinn phase transition at ligo, *Journal of High Energy Physics* **2020**, 1 (2020).
- [16] C. Caprini and D. G. Figueroa, Cosmological backgrounds of gravitational waves, *Classical and Quantum Gravity* **35**, 163001 (2018).
- [17] S. Kuroyanagi, T. Chiba, and T. Takahashi, Probing the universe through the stochastic gravitational wave background, *Journal of Cosmology and Astroparticle Physics* **2018** (11), 038.
- [18] N. Christensen, Stochastic gravitational wave backgrounds, *Reports on Progress in Physics* **82**, 016903 (2018).
- [19] E. S. Phinney *et al.*, Lisa science requirements (2002).
- [20] C. Liu, W.-H. Ruan, and Z.-K. Guo, Confusion noise from galactic binaries for taiji, *Physical Review D* **107**, 064021 (2023).
- [21] G. E. Box and G. C. Tiao, *Bayesian inference in statistical analysis* (John Wiley & Sons, 2011).
- [22] U. Von Toussaint, Bayesian inference in physics, *Reviews of Modern Physics* **83**, 943 (2011).
- [23] G. Turin, An introduction to matched filters, *IRE transactions on Information theory* **6**, 311 (1960).
- [24] L. Stanković and D. Mandić, Convolutional neural networks demystified: A matched filtering perspective-based tutorial, *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **53**, 3614 (2023).
- [25] D. George and E. Huerta, Deep neural networks to enable real-time multimessenger astrophysics, *Physical Review D* **97**, 044039 (2018).
- [26] H. Gabbard, M. Williams, F. Hayes, and C. Messenger, Matching matched filtering with deep networks for gravitational-wave astronomy, *Physical review letters* **120**, 141103 (2018).
- [27] X.-T. Zhang, C. Messenger, N. Korsakova, M. L. Chan, Y.-M. Hu, and J.-d. Zhang, Detecting gravitational waves from extreme mass ratio inspirals using convolutional neural networks, *Physical Review D* **105**, 123027 (2022).
- [28] M. López, I. Di Palma, M. Drago, P. Cerdá-Durán, and F. Ricci, Deep learning for core-collapse supernova detection, *Physical Review D* **103**, 063011 (2021).
- [29] H. Gabbard, C. Messenger, I. S. Heng, F. Tonolini, and R. Murray-Smith, Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy, *Nature Physics* **18**, 112 (2022).
- [30] M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf, Real-time gravitational wave science with neural posterior estimation, *Physical review letters* **127**, 241103 (2021).
- [31] R. E. Colgan, K. R. Corley, Y. Lau, I. Bartos, J. N. Wright, Z. Márka, and S. Márka, Efficient gravitational-wave glitch identification from environmental data through machine learning, *Physical Review D* **101**, 102003 (2020).
- [32] M. Cavaglia, K. Staats, and T. Gill, Finding the origin of noise transients in ligo data with machine learning, *arXiv preprint arXiv:1812.05225* (2018).
- [33] M. Razzano and E. Cuoco, Image-based deep learning for classification of noise transients in gravitational wave detectors, *Classical and Quantum Gravity* **35**, 095016 (2018).
- [34] R. Ormiston, T. Nguyen, M. Coughlin, R. X. Adhikari, and E. Katsavounidis, Noise reduction in gravitational-wave data via deep learning, *Physical Review Research* **2**, 033066 (2020).
- [35] K. Mogushi, Reduction of transient noise artifacts in gravitational-wave data using deep learning, *arXiv preprint arXiv:2105.10522* (2021).
- [36] A. Torres-Forné, A. Marquina, J. A. Font, and J. M. Ibáñez, Denoising of gravitational wave signals via dictionary learning algorithms, *Physical Review D* **94**, 124040 (2016).
- [37] W. Wei and E. Huerta, Gravitational wave denoising of binary black hole mergers with deep learning, *Physics Letters B* **800**, 135081 (2020).
- [38] H. Shen, D. George, E. Huerta, and Z. Zhao, Denoising gravitational waves using deep learning with recurrent denoising autoencoders, *arxiv e-prints*, *arXiv preprint arXiv:1711.09919* (2017).
- [39] C. Chatterjee, L. Wen, F. Diakogiannis, and K. Vinsen, Extraction of binary black hole gravitational wave signals from detector data using deep learning, *Physical Review D* **104**, 064046 (2021).
- [40] T. Zhao, R. Lyu, H. Wang, Z. Cao, and Z. Ren, Space-based gravitational wave signal detection and extraction with deep neural network, *Communications Physics* **6**, 212 (2023).
- [41] B.-R. Wang and J. Li, Ability of lisa, taiji, and their networks to detect the stochastic gravitational wave background generated by cosmic strings, *Physical Review D* **109**, 063520 (2024).
- [42] B.-R. Wang, J. Li, and H. Wang, Probing the gravitational wave background from cosmic strings with the alternative lisa-taiji network, *The European Physical Journal C* **83**, 1010 (2023).
- [43] L. Sousa, P. P. Avelino, and G. S. Guedes, Full analytical approximation to the stochastic gravitational wave background generated by cosmic string networks, *Physical Review D* **101**, 103508 (2020).
- [44] J. J. Blanco-Pillado, K. D. Olum, and B. Shlaer, Number of cosmic string loops, *Physical Review D* **89**, 023512 (2014).
- [45] C. Ringeval, M. Sakellariadou, and F. R. Bouchet, Cosmological evolution of cosmic string loops, *Journal of Cosmology and Astroparticle Physics* **2007** (02), 023.
- [46] C. Martins and E. Shellard, Extending the velocity-dependent one-scale string evolution model, *Physical Review D* **65**, 043514 (2002).
- [47] C. J. Moore, R. H. Cole, and C. P. Berry, Gravitational-wave sensitivity curves, *Classical and Quantum Gravity* **32**, 015014 (2014).
- [48] G. Boileau, N. Christensen, C. Gowling, M. Hindmarsh, and R. Meyer, Prospects for lisa to detect a gravitational-wave background from first order phase transitions, *Journal of Cosmology and Astroparticle Physics* **2023** (02), 056.
- [49] G. Wang and W.-B. Han, Alternative lisa-taiji networks: Detectability of the isotropic stochastic gravitational wave background, *Physical Review D* **104**, 104015 (2021).
- [50] M. Vallisneri and C. R. Galley, Non-sky-averaged sensitivity curves for space-based gravitational-wave observatories, *Classical and Quantum Gravity* **29**, 124015 (2012).

- [51] T. L. Smith and R. R. Caldwell, Lisa for cosmologists: calculating the signal-to-noise ratio for stochastic and deterministic sources, *Physical Review D* **100**, 104055 (2019).
- [52] J. Baker, J. Bellovary, P. L. Bender, E. Berti, R. Caldwell, J. Camp, J. W. Conklin, N. Cornish, C. Cutler, R. DeRosa, *et al.*, The laser interferometer space antenna: Unveiling the millihertz gravitational wave sky, arXiv preprint arXiv:1907.06482 (2019).
- [53] W.-R. Hu and Y.-L. Wu, The taiji program in space for gravitational wave physics and the nature of gravity (2017).
- [54] W.-H. Ruan, C. Liu, Z.-K. Guo, Y.-L. Wu, and R.-G. Cai, The lisa–taiji network, *Nature Astronomy* **4**, 108 (2020).
- [55] Z.-C. Chen, F. Huang, and Q.-G. Huang, Stochastic gravitational-wave background from binary black holes and binary neutron stars and implications for lisa, *The Astrophysical Journal* **871**, 97 (2019).
- [56] G. Boileau, A. C. Jenkins, M. Sakellariadou, R. Meyer, and N. Christensen, Ability of lisa to detect a gravitational-wave background of cosmological origin: The cosmic string case, *Physical Review D* **105**, 023510 (2022).
- [57] G. Wang, W.-T. Ni, W.-B. Han, and C.-F. Qiao, Algorithm for time-delay interferometry numerical simulation and sensitivity investigation, *Physical Review D* **103**, 122006 (2021).
- [58] K. Haris, A. K. Mehta, S. Kumar, T. Venumadhav, and P. Ajith, Identifying strongly lensed gravitational wave signals from binary black hole mergers, arXiv preprint arXiv:1807.07062 (2018).
- [59] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, Deep residual shrinkage networks for fault diagnosis, *IEEE Transactions on Industrial Informatics* **16**, 4681 (2019).
- [60] Y. Li and H. Chen, Image recognition based on deep residual shrinkage network, in *2021 International Conference on Artificial Intelligence and Electromechanical Automation (AIEA)* (IEEE, 2021) pp. 334–337.
- [61] X. Liu and Z. Hou, Improved deep residual shrinkage network for noisy image recognition, in *2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)* (IEEE, 2024) pp. 1408–1411.
- [62] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in *Proceedings of the AAAI conference on artificial intelligence*, Vol. 31 (2017).
- [63] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016) pp. 770–778.
- [64] F. Yu, V. Koltun, and T. Funkhouser, Dilated residual networks, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017) pp. 472–480.
- [65] D. L. Donoho, De-noising by soft-thresholding, *IEEE transactions on information theory* **41**, 613 (1995).
- [66] J. Joy, S. Peter, and N. John, Denoising using soft thresholding, *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering* **2**, 1027 (2013).
- [67] K. Bredies and D. A. Lorenz, Linear convergence of iterative soft-thresholding, *Journal of Fourier Analysis and Applications* **14**, 813 (2008).
- [68] F. Girosi, M. Jones, and T. Poggio, Regularization theory and neural networks architectures, *Neural computation* **7**, 219 (1995).
- [69] Y. Tian and Y. Zhang, A comprehensive survey on regularization strategies in machine learning, *Information Fusion* **80**, 146 (2022).
- [70] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, *The journal of machine learning research* **15**, 1929 (2014).
- [71] V. Tinto, Dropout from higher education: A theoretical synthesis of recent research, *Review of educational research* **45**, 89 (1975).
- [72] M. Pauly, R. Keiser, and M. Gross, Multi-scale feature extraction on point-sampled surfaces, in *Computer graphics forum*, Vol. 22 (Wiley Online Library, 2003) pp. 281–289.
- [73] A. E. Orhan and X. Pitkow, Skip connections eliminate singularities, arXiv preprint arXiv:1701.09175 (2017).
- [74] H. Wang, P. Cao, J. Wang, and O. R. Zaiane, Uctransnet: rethinking the skip connections in u-net from a channel-wise perspective with transformer, in *Proceedings of the AAAI conference on artificial intelligence*, Vol. 36 (2022) pp. 2441–2449.
- [75] J. Mukhoti, V. Kulharia, A. Sanyal, S. Golodetz, P. Torr, and P. Dokania, Calibrating deep neural networks using focal loss, *Advances in Neural Information Processing Systems* **33**, 15288 (2020).
- [76] X. Li, W. Wang, L. Wu, S. Chen, X. Hu, J. Li, J. Tang, and J. Yang, Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection, *Advances in Neural Information Processing Systems* **33**, 21002 (2020).
- [77] M. Lisicki, A. Afkanpour, and G. W. Taylor, Evaluating curriculum learning strategies in neural combinatorial optimization, arXiv preprint arXiv:2011.06188 (2020).
- [78] G. Hacohen and D. Weinshall, On the power of curriculum learning in training deep networks, in *International conference on machine learning* (PMLR, 2019) pp. 2535–2544.
- [79] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, Curriculum learning, in *Proceedings of the 26th annual international conference on machine learning* (2009) pp. 41–48.