This is the accepted manuscript made available via CHORUS. The article has been published as:

Extracting non-Gaussian features in gravitational wave observation data using self-supervised learning

Yu-Chiung Lin and Albert K. H. Kong

Phys. Rev. D **111**, 063520 — Published 7 March 2025

DOI: 10.1103/PhysRevD.111.063520

Extracting non-Gaussian Features in Gravitational Wave Observation Data Using Self-Supervised Learning

Yu-Chiung Lin* and Albert K. H. Kong†
Institute of Astronomy, National Tsing Hua University, Hsinchu 30013, Taiwan

We propose a self-supervised learning model to denoise gravitational wave (GW) signals in the time series strain data without relying on waveform information. Denoising GW data is a crucial intermediate process for machine-learning-based data analysis techniques, as it can simplify the model for downstream tasks such as detections and parameter estimations. We use the blind-spot neural network and train it with whitened strain data with GW signals injected as both input data and target. Under the assumption of a Gaussian noise model, our model successfully denoises over 90% of GW signals from binary black hole (BBH) and neutron star-black hole (NSBH) mergers when the signal signal-to-noise ratio (SNR) is larger than 13 and 16 and denoises over 50% of binary neutron star (BNS) signals when the SNR is larger than 23, with an overlap greater than 0.2, over which the denoised signal can have visually distinguishable patterns. Our model also denoises over 90% of BBH signals in H1, L1, and V1 data detected in the O1, O2, and O3 observation runs when the network SNR is larger than 11, 12, and 20, respectively. We also test the model's potential to extract glitch features, loud inspiral compact binary coalescence signals a few seconds before the merger, and unseen core-collapse supernova (CCSN) signals during training.

I. INTRODUCTION

The research on gravitational wave (GW) astronomy has been booming since the first detection of the GW signal from a binary black hole (BBH) coalescence (GW150914) [1] made by the advanced Laser Interferometer Gravitational-Wave Observatory (LIGO) [2] in 2015. In addition, in 2017, the LIGO-Virgo Collaboration [2, 3] detected the first neutron star binary coalescence (GW170817) [4] with electromagnetic follow-up observations [5–10], which is the first multi-messenger event containing GW emissions in history. In the first three observation runs from 2015 to 2020, the LIGO and Virgo Collaboration have reported 93 confidently-detected GW events from compact binary coalescence (CBC) events [11–14], including GWs from two binary neutron star (BNS) merger events [4, 15] and three possible candidates from neutron star-black hole binaries (NSBH) [14]. Moreover, in the first half of the fourth observation run (O4a), 81 significant GW candidates from CBC events were uploaded to the Gravitational-Wave Candidate Event Database [16] (GraceDB) and public alerts were sent out. On the other hand, the Kamioka Gravitational Wave Detector (KAGRA) [17, 18] in Japan started the joint observation with the GEO600 detector [19, 20] in 2020 (O3GK) for two weeks but did not find any events [21, 22]. With the addition of the KAGRA detector with the LIGO and Virgo forming the LIGO-Virgo-KAGRA (LVK) network, it is expected to detect more GW candidates from CBCs in O4b. Also, searches for GW signals from different sources such as subsolar-mass binaries [23–29], bursts [30–35] and continuous waves from spinning neutron stars [36–40] and low-mass X-ray binaries [41] are ongoing.

Most of the current GW search pipelines [42–48] are based on the matched-filtering search method, which uses a set of known GW waveforms as templates to correlate the strain data and search for the maximum response [49, 50]. The matched-filtering method is optimal for searching the GW signal if the background noise is Gaussian [49]. However, as the type of source and mass range increase, more GW templates that span wide frequency ranges are needed, which will require many computational resources during the search [51]. To tackle this challenge, the machine learning (ML) technique is a prominent alternative to the matched-filtering method for its short inference time, which is suitable for low-latency searches. So far the literature has explored the possibility of ML approaches to GW detection [52–63], source and glitch classification [64–66], sky localization [67–69], parameter estimation [52, 56, 61, 70, 71], and denoising [72–82].

GW signal denoising is essential for enhancing the signal-to-noise ratio (SNR) and providing robust astrophysical parameter estimation [73]. Current works have used dictionary learning [72], supervised learning [73], and denoising auto-encoder (DAE) [74, 76, 79–81] to recover the burst and the BBH waveforms from noisy strain data. However, these works rely on the clean waveform as the reference, which makes them effective only for denoising well-modeled waveforms such as CBC signals and a specific type of burst signals. On the other hand, a neural network model was proposed to remove glitches overlapping with the signals by training a model to reconstruct the gated portion of strain data with injection [78]. This method does not use clean signals as references but can only reconstruct CBC signals with large SNR in a gated portion of a few hundred milliseconds before the merger. Meanwhile, a total-variation method was employed to remove Gaussian noise in the strain data without relying on information from clean signals [83, 84]. This method suc-

* yuchiung.lin@mx.nthu.edu.tw
† akong@gapp.nthu.edu.tw

cessfully denoised loud core-collapse supernova (CCSN) signals and BBH signals. However, achieving an optimal denoising result requires searching for a regularization term using an additional artificial neural network (ANN) model.

We consider ML-based denoising as a feature-extraction task since it can remove noise and emphasize the important GW features in the strain data. Such a feature extractor can serve as an intermediate process for downstream tasks such as detection, parameter estimation, and glitch classification. It also simplifies the complexity of ML models that perform these tasks. We aim to build a general-purpose feature extractor in the ML-based GW processing pipeline, which can extract not only CBC signals, but also other interesting non-Gaussian or non-stationary features, such as unmodeled burst signals, or glitches that are yet fully understood, from the strain data. One can directly use our model to extract features for downstream tasks or fine-tune the specific tasks to focus on certain features for which the task is needed. In recent years, the self-supervised learning (SSL) technique has become a popular research field in computer vision [85–89] and natural language processing [90, 91]. By training with specific pre-text tasks and supervisory loss functions, the SSL model can learn to extract essential representations in the data without the need for human annotations [92, 93].

The SSL denoiser predicts the clean data point from the surrounding noisy data and is currently an active research field in computer vision due to the difficulty of obtaining clean ground-truth images in reality [85, 94–96]. With a suitable noise model as the supervisory loss function, the SSL denoiser can be trained using the original noisy image as both the input and the target. Training of the SSL denoiser can be seen as extracting the essential features in the noisy data under the given noise assumption. When using the noisy image as both the input and the target, traditional neural network models can only learn the identity since the pixels in the input image are directly mapped to the original image. Interestingly, in their work, Krull et al. [88] provided a remedy by proposing the NOISE2VOID training technique based on the blind-spot neural network where the receptive field of the network does not include the center pixel. Consequently, the model is forced to predict the center clean pixel using the surrounding noisy data. The model can be trained with only noisy images, assuming the corruption is zero-mean and independent between pixels. However, they created the blind spot by applying the masking scheme to the loss function, reducing the training efficiency. Laine et al. [94] proposed a blind-spot structure that combines multiple branches in which the receptive field is restricted in different directions. The restriction in the receptive field can be realized by shifting the output of a normal convolution so that only pixels synthesized before the current pixel are allowed in the receptive field [97–99].

This work uses the SSL denoiser to extract GW features in the strain data without providing clean waveforms as references. Assuming the Gaussian noise background with zero means, we use the blind-spot method proposed in Ref. [94] to create the blind spot in the WaveNet model [100] and train the model using the whitened strain data with CBC signals injected as both input and target. Our goal is to build a general-purpose feature extractor in the machine learning-based GW processing pipeline, which can extract not only astronomical features, such as GW signals from different types of GW sources like CBCs, CCSNe, continuous waves from spinning neutron stars, but also terrestrial features, such as anomalous glitches. The extracted features can then be used for downstream tasks such as detection, parameter estimation, glitch classification, and sky localization.

The rest of this paper is organized as follows: In Section II, we give an overview of the self-supervised learning denoiser and the blind-spot neural network and introduce the method used in this work, including the noise model, the neural network architecture, the process of generating training and testing data, and the detail of the training process. In Section III, we present the model performance results on the mock data and the actual events. This section also discusses the potential for extracting glitch features, CBC inspiral signals, and CCSN signals that have not been added to the training dataset. Finally, we conclude our work in Section IV.

II. SELF-SUPERVISED LEARNING GRAVITATIONAL WAVE DENOISER

A. The Gaussian Noise Model

The detector strain data $\mathbf{d}(t)$ can be seen as a combination of detector noise $\mathbf{n}(t)$ and the gravitational wave signals $\mathbf{h}(t)$ such that

$$\mathbf{d}(t) = \mathbf{n}(t) + \mathbf{h}(t). \tag{1}$$

If the noise is stationary and Gaussian, then we can model the noise of a single detector using the likelihood function [50]:

$$p(\mathbf{d}|\mathbf{h}) = \exp\left(-\frac{1}{2}\left[(\mathbf{d}-\mathbf{h}|\mathbf{d}-\mathbf{h}) + \int \ln(S_n(f))df\right]\right), \tag{2}$$

where $S_n(f)$ is the noise power spectral density estimation for the detector, and $(\mathbf{a}|\mathbf{b})$ is the noise-weighted inner product

$$(\mathbf{a}|\mathbf{b}) = 2\int_0^\infty \frac{\tilde{a}(f)\tilde{b}^*(f) + \tilde{a}^*(f)\tilde{b}(f)}{S_n(f)}df. \tag{3}$$

With the Gaussian noise likelihood described in Equation (2), we can train a model to predict the clean signal

$\hat{\mathbf{h}}$ from the noisy data $\mathbf{d}$, using the loss function

$$\begin{aligned}\mathcal{L} &= -\log p(\mathbf{d}|\hat{\mathbf{h}}) \\ &= \frac{1}{2}\left[(\mathbf{r}|\mathbf{r}) + \int \ln(S_n(f))df\right],\end{aligned} \tag{4}$$

where $\mathbf{r} = \mathbf{d} - \hat{\mathbf{h}}$ is the residual between the noisy data and the predicted clean signal.

Usually, the detector noise is several orders of magnitude greater than the GW signals, which makes it hard for the neural network model to extract GW features in such noisy data. To suppress the noise magnitude and emphasize the GW features, one has to whiten the time series data by dividing its Fourier coefficients by the noise amplitude spectral density estimation $\sqrt{S_n(f)}$ [50]. Equation (3) can therefore be seen as the correlation between two vector $\mathbf{a}_w$ and $\mathbf{b}_w$ whitened with the noise power spectral density $S_n(f)$:

$$(\mathbf{a}|\mathbf{b}) = 2\int_0^\infty \left(\tilde{a}_w(f)\tilde{b}_w^*(f) + \tilde{a}_w^*(f)\tilde{b}_w(f)\right) df, \tag{5}$$

where $\tilde{a}_w(f)$ and $\tilde{b}_w(f)$ are the Fourier transform of $\mathbf{a}$ and $\mathbf{b}$, respectively. In addition, we assume that $S(n) = 1$ for whitened strain, and then the second term in Equation (4) becomes zero. Therefore, the loss function can be simplified to

$$\mathcal{L} = 2\int_0^\infty |\tilde{r}_w(f)|^2\, df, \tag{6}$$

where $\tilde{r}_w(f) = \tilde{d}_w(f) - \tilde{h}_w(f)$ is the whitened residual between the noisy data and the predicted clean signal. Note that the loss function (6) is the mean squared error in the Fourier space.

In reality, the detector data is usually neither stationary nor Gaussian, and the cleaned signal $\hat{\mathbf{h}}(t)$ may contain non-stationary or non-Gaussian noises such as lines or glitches. However, we can also expect that the model will be able to probe the GW signals from different sources, such as CCSN and continuous waves from spinning neutron stars. Also, there is a possibility to have overlapping signals in the strain data. Our model can help remove the noise so that the other signal separation model can better split the overlapped signals. Therefore, our model can be seen as a general-purpose feature extractor for GW data and simplify the complexity of the model performing downstream tasks such as classification, parameter estimation, and signal separation.

B. The Blind-Spot WaveNet Model

Since we use the whitened strain data as input and output, the ordinary auto-encoder models will learn the identity, as they can see the entire input in the first place. Therefore, we use the blind-spot neural network whose receptive field does not include the center pixel. Because the center pixel is masked, the model cannot access all data simultaneously and has to make a new prediction for the masked part [88]. This neural network is trained self-supervised by predicting the output value based on the surrounding context without requiring an additional reference target. Laine et al. [94] proposed a method to implement the blind-spot neural network under the current deep learning framework by using a neural network with the receptive field restricted in different directions and augmenting the input data to have different rotation angles. The restriction on the receptive field can be easily done by shifting the output of a normal convolution [97–99]. An additional one-pixel offset is applied to create the blind spot. Therefore, the neural network can acquire the context information from the surrounding pixels except for the center pixel.

In this work, we apply the blind-spot technique in Reference [94] on the WaveNet-based model [100], and the structure of the model we use is shown in Figure 1. WaveNet is based on the causal convolution technique [100], which is accomplished by right-shifting the output of the standard convolution layer so that the receptive field is restricted to past data, and this is the same method used for constructing the blind-spot network in Reference [94]. Since the original WaveNet model only sees the left side of the data in the receptive field, we first create a time-reversed copy of the input data and stack them to the minibatch axis to make the model refer to more surrounding data. Thus, the batch size of the data is twice as large as the original input data. Then, we perform the causal convolution to obtain 1440 feature maps in which each pixel only correlates to one side of its surroundings. The data then pass through 12 convolution stacks; each stack has dilated causal convolution layers with residual connections [101], each with kernel size 3, and the dilation rate is cyclically switched in $1, 2, 4$. In each dilated causal convolution layer, we first perform the 1×1 convolution to downsize the number of feature maps to 360 before the causal convolution to save computational resources. We then apply a tanh and a sigmoid activation separately on the data and multiply them together. After the multiplication, we perform another 1×1 convolution on the data to increase the number of feature maps back to 1440. Next, we add the input data to the output of the convolution and pass the data to the next layer. We finally aggregate the last 1×1 convolution output in each layer to obtain the skip connections.

After applying an Elu activation function [102] on the skip connections, we shift the data by one pixel to the right to create the blind spot. We then split the minibatch axis into two pieces, undo the reverse operation on the second piece, and stack them to the feature axis. Therefore, the data now has the same batch size as the original input data, but the number of feature maps is twice as large as the original skip connections. Finally, we pass the data to three 1×1 convolution layers with the number of feature maps 2880, 1440, and 1, each followed by an Elu activation except for the last layer. The

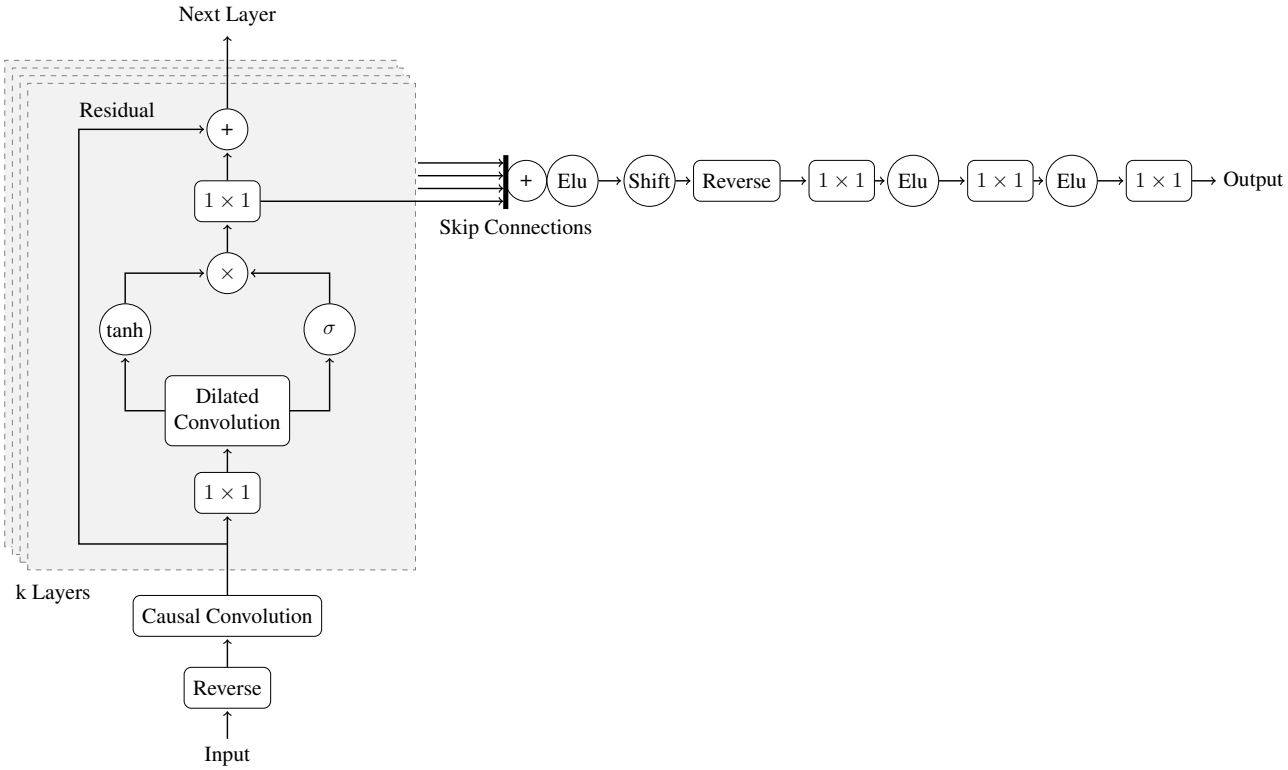


FIG. 1. Overview of the blind-spot WaveNet model used in this work.

parameter name	distribution
BBH approximant	SEOBNRv4T, IMRPhenomD
NSBH approximant	TaylorT4, IMRPhenomBHNS
BNS approximant	TaylorT4
$m_{\rm BH}$	uniform on $[3, 75]\ M_\odot$
$m_{\rm NS}$	uniform on $[1.4, 3]\ M_\odot$
mass ratio q	uniform on $[1, 10]$
right ascension	uniform on $[0, 2\pi]$
declination	uniform on $[-\pi/2, \pi/2]$
polarization angle ψ	uniform on $[0, 2\pi]$
$\rho_{\rm opt}$ for BBH	$[5, 25]$
$\rho_{\rm opt}$ for NSBH and BNS	$[10, 35]$

TABLE I. The parameters of GW signals used in the training dataset.

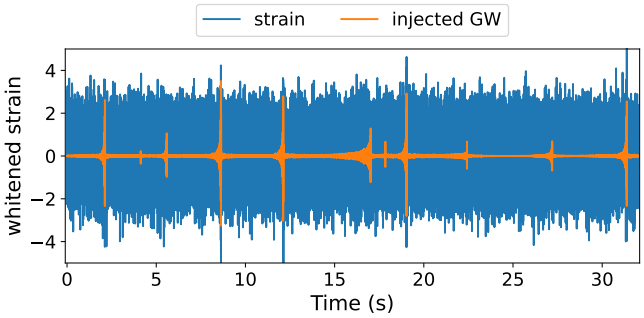


FIG. 2. An example of a 32-second segment in the training dataset. The blue curve is the whitened strain data, and the orange curve indicates the injected GW signals.

output of the last 1x1 convolution is the cleaned data. In this work, we do not rigorously tune the model's hyper-parameters, leaving space for optimization in the future.

C. Data Preparation

We use the public LIGO Hanford and Livingston data during O2 and O3 runs available on the Gravitational Wave Open Science Center (GWOSC) [103] as the background noise for the training dataset. The background noises in different observation periods can increase the diversity of features to be picked up by the denoiser model. We use the strain data having the quality flag `CBC_CAT3` = 100 with a sampling rate of 4096 Hz. We randomly chose 120k data segments of 32 seconds from these strain data as our training dataset. To prevent the bias of the test described in Section III B, we excluded the data segments containing known GW signals listed in References [11–14]. Note that in the actual application, there is no need to avoid these signals since our model does not need cleaned references for training since these signals are also the non-Gaussian features that we want our model to pick up.

Furthermore, to help the model recognize the feature

of GW signals, we randomly inject 8 to 10 simulated GW signals from compact binary coalescence into each segment, using the PyCBC [104] package with the parameters listed in Table I. The time interval between each merger event in a segment is randomly chosen from 3 to 5 seconds. We use the `SEOBNRv4T` [105] and the `IMRPhenomD` [106, 107] approximant for BBH signal generation, the `TaylorT4` [108] and the `IMRPhenomBHNS` [109, 110] approximant for NSBH, and the `TaylorT4` approximant for BNS, respectively. The binary is assumed to be spinless and arbitrarily located in the sky with an arbitrary polarization angle ψ. We choose the component masses range in $3M_\odot$ to $75M_\odot$ for black hole masses $m_{\rm BH}$, and $1.4M_\odot$ to $3M_\odot$ for neutron star masses $m_{\rm NS}$. The mass ratio $q = m_1/m_2$ of the component masses in the binary is constrained to be less than 10. We set the starting frequency of the simulated signal to be 15 Hz. While the difference of the waveform among different approximants in the parameter space we considered above is nearly negligible in the inspiral stage, their behavior can be different in the merger-ringdown stage, especially for binaries with high mass ratios ($q > 8$) [105]. Therefore, we believe using different approximants can help the model recognize more merger-ringdown features and enhance its ability to extract BBH features from actual data. When injecting the GW signal $h(t)$ into the strain data, we scale its amplitude so that the optimal SNR, $\rho_{\rm opt}$, defined by [49]

$$\rho^2_{\rm opt} = 4\int_0^\infty \frac{\left|\tilde{h}(f)\right|^2}{S_n(f)} df, \tag{7}$$

lies within the range of 5 to 25 for BBH signals and 10 to 35 for NSBH and BNS signals. The PSD $S_n(f)$ is estimated using the 128 seconds of data without injection around the chosen segment. After injection, we whiten the strain data using the same PSD. Figure 2 shows an example of a 32-second segment in the training dataset. The blue curve is the whitened strain data, and the orange curve indicates the injected GW signals. Finally, we save each segment in individual files. Note that we do not preserve individual simulated GW signals since we do not need them as training labels.

For the test dataset used in Section III A, we randomly pick 512 segments of 2 seconds from the strain data described above and inject GW signals into them with $\rho_{\rm opt}$ ranging from 5 to 25 in steps of 0.5 for BBH, and 8 to 35 in steps of 0.5 for NSBH and BNS respectively. Unlike the training dataset, we only inject one GW signal for each sample, which peaks at around a third quarter of the segment and saves the GW signal alone for the performance evaluation.

To maximize the model's ability for feature extraction, besides GW signals from CBCs, one may also inject other types of GW signals, such as GWs from CCSNe or continuous waves from rotating neutron stars, into the training data. In this work, we only inject GW signals from CBCs due to the difficulty of generating the simulated signals from other sources.

D. Training

We implement our model using the Tensorflow [111] package, the ADAMax optimizer [112] with default parameters, an initial learning rate of 0.0003, and the size of minibatch 8. Each minibatch consists of 2-second segments randomly cropped from the training dataset described in Section II D. We train our model using 0.5M mini-batches and ramp down the learning rate to 10^{-6} using the cosine schedule during the last 20% of training. Before feeding into the model, we standardize the data to have zero mean and unit variance. Although standardization may alter the scale of the Gaussian noise spectrum $S(f)$, the noise spectrum is constant for whitened data and does not contribute to the gradient update. Moreover, standardization can stabilize the training process and make the model more easily converge. When calculating the loss function, we first apply a Tukey window with $\alpha = 0.25$ to the residual $r(t)$ before the Fourier transforms to avoid spectral leakage. In this work, we sum up all the power ranging from 0 Hz to the Nyquist frequency of the data, which is 2048 Hz.

During the training, we monitored the loss function to check the convergence without using a validation dataset since there are no proper metrics to monitor whether the model is overfitting on extracting non-Gaussian features. Instead, we directly evaluate the model performance by running the test described in Section III after training.

We trained our model on the cluster with eight NVIDIA™ RTX 3080 GPUs in the Center of Informatics and Computation in Astronomy (CICA). The training process takes about 120 hours.

III. RESULTS

A. Quantitative Performance on Mock Data

Following the common metric used in GW denoising literature, we use the overlap $\mathcal{O}(h, s)$ to quantify the performance of the model of reconstructing the clean signal, defined by [12]

$$\mathcal{O}(h, s) = \frac{(h|s)}{\sqrt{(h|h)(s|s)}}, \tag{8}$$

where s is the cleaned signal and h is the injected GW template. We calculate the overlap in two ways: the first is to calculate the overlap using the full 2-second data segment to see the overall match between the cleaned signal and the injected GW signal. The second is to calculate the overlap using the data segment of 0.25s around the peak of the injected GW signal to see how well the model can recover the waveform around the merger event for detection and parameter estimation. The evaluation

is done on a personal computer with an NVIDIA™ RTX 3060 GPU. Since our model can extract non-Gaussian and non-stationary features, we expect the overlaps to fluctuate significantly in each dataset due to these non-signal features.

We present the overlap distribution on the BBH, NSBH, and BNS datasets in Figure 3a, 3b, and 3c, respectively. The area in each SNR bin indicates the overlap distribution of 512 samples in the SNR dataset, and the dot presents the mean of these values. The black error bars represent the 90% confidence interval of the distribution. The left panel in Figure 3 is the overlap calculated using the full 2-second data segment, and the right panel is the overlap calculated using the data segment of 0.25s around the peak of the injected GW signal.

As we expected, the mean overlap improves when SNR increases. Also, there is a large scattering of overlaps in all datasets, including signals with high optimal SNR, and the maximum mean overlap values for BBH and NSBH datasets are only around 0.8 for the 2-second data, which agrees with our expectation that non-Gaussian features from noises affect the overlap calculation. The maximum mean overlap values for BBH and NSBH datasets improve to 0.9 when using the 0.25s segment around the GW peak. On the other hand, the low mean overlap for 2-second data may be due to the inability of the model to recover the low-amplitude waveform at the early inspiral stage or the remaining high-frequency noise in the denoised signal, which downgrades the overlap, as shown in Section III B.

The mean overlaps of the BNS dataset are lower than the BBH and NSBH datasets and do not improve significantly when using 0.25 second data segments. We believe this is because the BNS waveform has a much longer duration and wider frequency range and, therefore, can have a larger optimal SNR with a smaller amplitude around the merger event, which the model cannot cover well. In addition, the BNS waveform has higher frequencies around the merger and, therefore, can be more easily affected by the high-frequency noise remaining in the denoised signal.

Also, in Section III B, we empirically find that in the 0.25s window, the denoised signal can have a distinguishable structure that can possibly be recognized by downstream detection models when the overlap is more significant than 0.2. With this overlap value as the threshold, our model can denoise over 90% of the BBH signals in the mock dataset when $\rho_{\rm opt} > 13$, 16 for the NSBH signals, and can denoise over 50% of the BNS signals when $\rho_{\rm opt} > 23$.

To further investigate the large scattering in the overlap distributions, we consider the model performance against the signal frequency. The frequency evolution of the inspiral waveform has the form [113]

$$f_{\rm GW}^{-8/3}(t) = \frac{(8\pi)^{8/3}}{5}\left(\frac{G\mathcal{M}_c}{c^3}\right)^{5/3}(t_c - t) + \text{higher order corrections}, \tag{9}$$

where t_c is the coalescence time, G is the gravitational constant, c is the speed of light, and $\mathcal{M}_c$ is the chirp mass

$$\mathcal{M}_c = \frac{(m_1 m_2)^{3/5}}{(m_1+m_2)^{1/5}}. \tag{10}$$

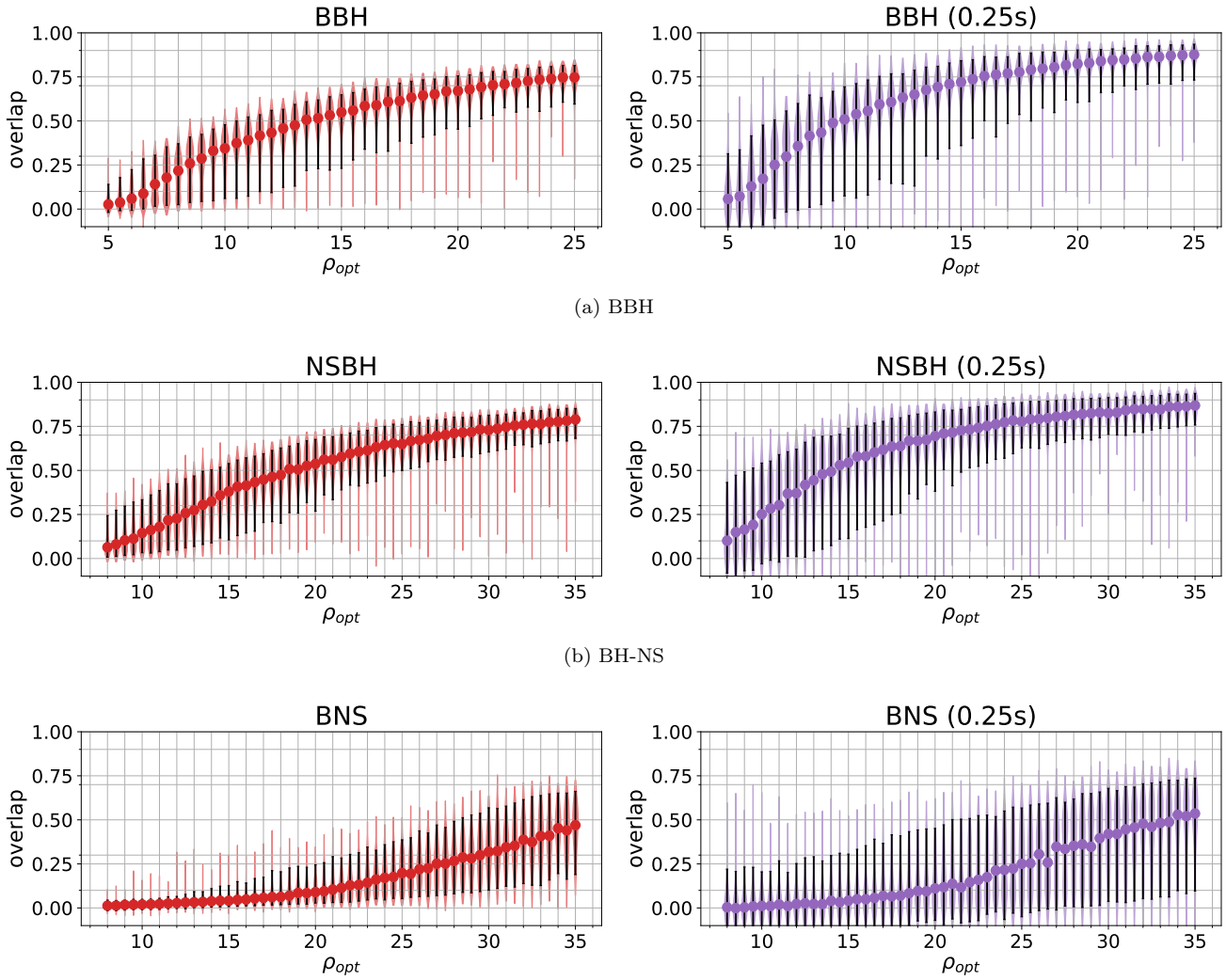
From Equation (9), we can see that a binary system with a larger chirp mass will have a lower frequency at a given time t. Therefore, we compare the overlap value with the chip mass in Figures 4, 5 and 6 respectively. We choose the samples with $\rho_{\rm opt} = 5, 10, 15, 20$ for the BBH dataset, $\rho_{\rm opt} = 15, 20, 25, 30$ for the NSBH dataset and $\rho_{\rm opt} = 20, 25, 30, 35$ for the BNS dataset. The red dots are the overlap results using the full 2-second data segment, and the purple dots are the overlap results using the data segment of 0.25s around the peak of the injected GW signal. At high $\rho_{\rm opt}$, our model performs well to recover GW signals with large chirp mass, which has lower frequencies around the merger event. On the other hand, our model performs poorly when recovering GW signals with small chirp mass, which has higher frequencies around the merger event. In addition, there are a few outliers in these figures, which may be due to the presence of glitches in the data.

Figures 3, 4, 5 and 6 indicate that our model performs not well on low-amplitude signals, and GW signals with small chirp mass tend to have low amplitude while the optimal SNR is unchanged. This can be due to the difficulty of distinguishing the low-amplitude feature from the noise background using Gaussian likelihood in the frequency domain, as our model does not know the shape of the GW waveform at the training stage. To improve the performance on low-amplitude signals, we may have to use different noise models or apply noise subtraction models, such as the DeepClean [76, 82] to remove loud line noise before whitening to enhance the significance of the signal in the strain data. We leave this as future work.

The left panel in Figure 3 shows that our model performs better in extracting GW features around the merger event. In the future, we will investigate which overlap values can be used as the detection and parameter estimation threshold. However, although our model cannot recover the full part of GW waveforms, it can still be used as the feature extractor for detection and parameter estimation.

B. Performance on Real Events

In this section, we test our model with the confidently detected events in the Gravitational Wave Transient Catalog (GWTC) reported by the LVK Collaboration during O1, O2, and O3 observation runs [11–14]. We use the 4096 seconds 4KHz strain files containing the merger events available in GWOSC and choose the 32-second segments, each with the event time at the center. Each segment is whitened using the PSD estimated from the



(a) BBH

(b) BH-NS

(c) BNS

FIG. 3. The overlap result of (a) the BBH test dataset, (b) the NSBH test dataset, and (c) the BNS test dataset. The violin plots are the overlap distribution of all 512 samples in different $\rho_{\rm opt}$ bins. The dot in each bin is the mean of the overlap distribution. The black error bars represent the 90% confidence interval of the distribution. The left panel is the overlap calculated using the full 2-second data segment, and the right panel is the overlap calculated using the data segment of 0.25s around the peak of the injected GW signal.

64 seconds of data 16 seconds before the event time. For BNS events GW170817 and GW190425, we use the 64-second data, which is 128 seconds before the event time, to estimate the PSD. We slice the center 2-second data of the segment as the model input. There is a loud glitch in the L1 data before the event GW170817, and we use the method described in Reference [114] to mitigate the glitch. To quantify the model performance, we generate the waveform templates using the approximant `SEOBNRv4T` for BBHs, `IMRPhenomBHNS` for BH-NSs, and `TaylorT4` for BNSs, respectively. We use the best-fit parameters of the waveforms provided on the GWOSC. To reconstruct the event waveforms, we perform the vanilla matched-filtering and then shift the template's time and phase to align with the strain data.

To see how the denoised results behave, we plot the denoised result of the 2-second H1 and L1 data around the event GW150914 in Figure 7a, and in Figure 7b we zoom the result of each detector data to the 0.25s window around the peak of the reconstructed waveform which we use to calculate overlap. The blue curves are the whitened strain data, the red curves are the denoised results, and the cyan curves are the reconstructed waveforms. The overlap is 0.82 for the H1 data and 0.84 for the L1 data. Figure 7 shows that our model can effectively remove the Gaussian noise and recover the GW features around the loud merger event, which can be used as a feature extractor for detection and parameter estimation tasks. However, we can see that the denoised results still contain some high-frequency noises, which

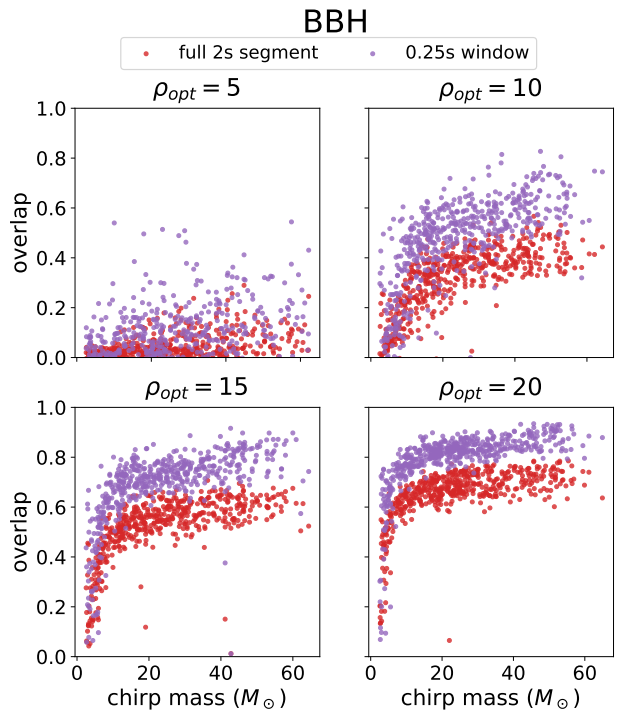


FIG. 4. The overlap versus chirp mass of the BBH test dataset with $\rho_{\rm opt} = 5, 10, 15, 20$, respectively. The red dots are the overlap results using the full 2-second data segment, and the purple dots are the overlap results using the data segment of 0.25s around the peak of the injected GW signal. Our model performs better in recovering GW signals with large chirp masses and around the merger event.

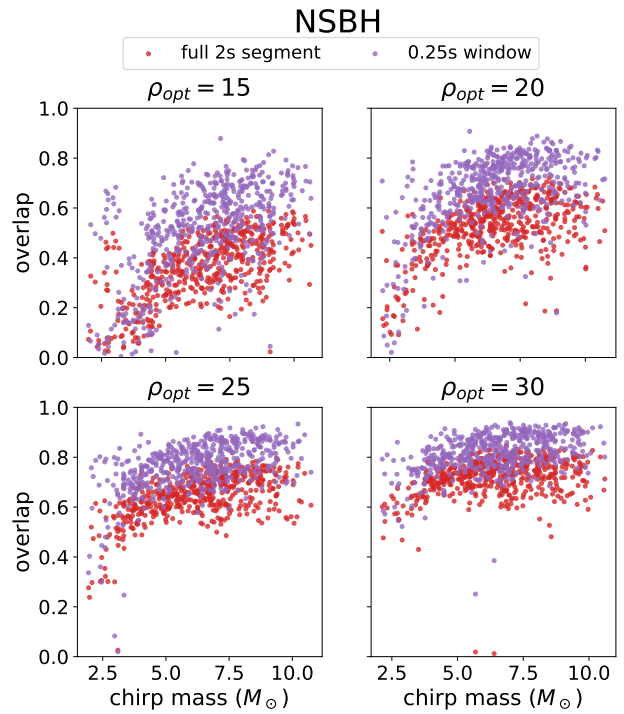


FIG. 5. The overlap versus chirp mass of the NSBH test dataset with $\rho_{\rm opt} = 15, 20, 25, 30$, respectively. Our model still performs better in recovering GW signals with large chirp masses and around the merger event.

decrease the overlap. In addition, the mismatches between the denoised results and the reconstructed waveforms at the early inspiral stage and the ring-down stage also downgrade the overlap. Our denoised results tend to follow the trend of low-frequency data fluctuation since our model does not know GW waveforms. The same phenomenon can be observed in the denoised results of event GW151012, as shown in Figure 8, where the denoised result of H1 data peaks earlier than the reconstructed waveform, which may bias the parameter estimation results.

On the other hand, our model cannot preserve the GW features of the BNS event GW170817, as shown in Figure 9. The overlap is 0.03 for the H1 data and 0.0 for the L1 and V1 data. We can see that the amplitudes of the reconstructed waveforms are insignificant in all detector time series data and may be recognized by the model as parts of Gaussian noise. As described in Section III A, using different noise models or subtracting line noises may help improve the model's ability to extract the GW features for BNS waveforms. We present the denoised results of the remaining confidently-detected GW events during O1, O2, and O3 runs in Appendix A. In these plots, we can see that when the overlap value is larger than 0.2, most of the denoised signal can have

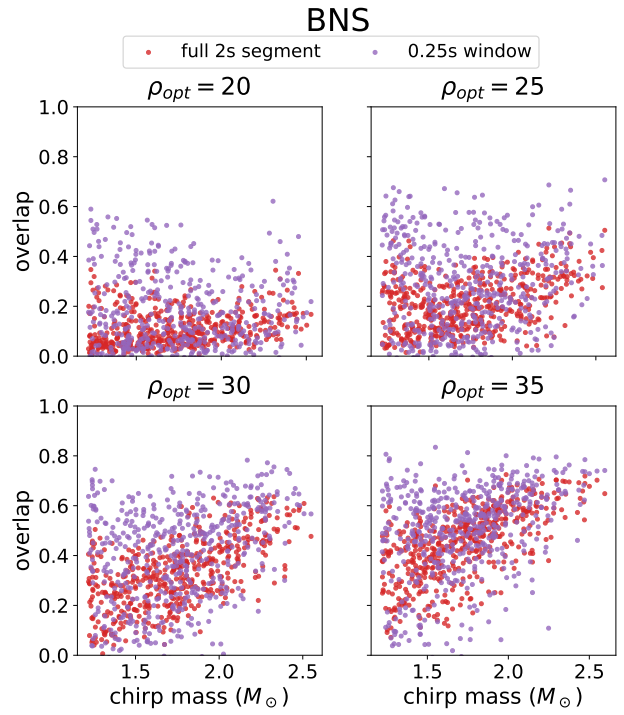
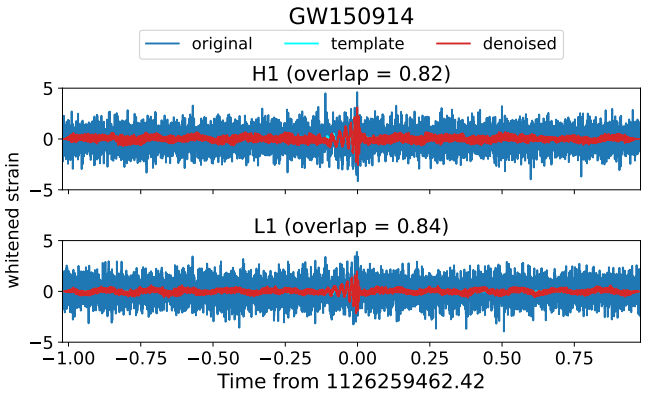
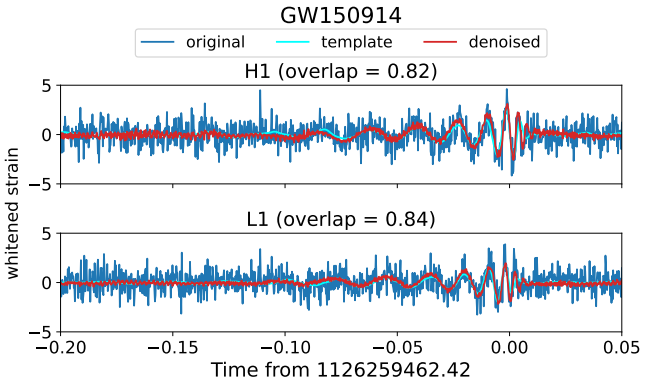


FIG. 6. The overlap versus chirp mass of the BNS test dataset with $\rho_{\rm opt} = 20, 25, 30, 35$, respectively. Again, our model performs better in recovering GW signals with large chirp masses, but there is no significant improvement in overlap when using the data segment around the merger event.



(a)



(b)

FIG. 7. (a) The denoised results of 2-second H1 and L1 data around the event GW150914. The blue curves are the whitened strain data, the red curves are the denoised results, and the cyan curves are the reconstructed waveforms. (b) The 0.25s windows around the peak of the reconstructed waveform, which we use to calculate overlaps.

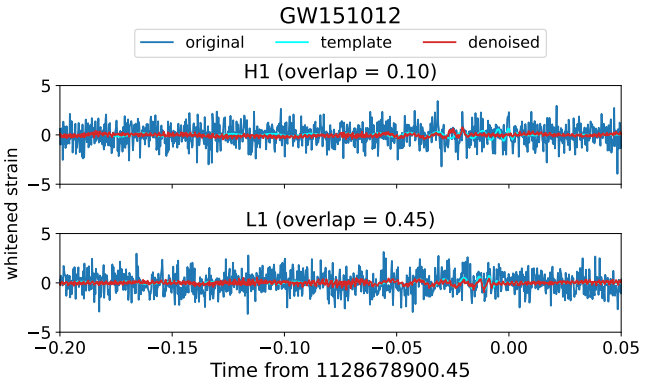


FIG. 8. The 0.25s windows of the denoised results of H1 and L1 data around GW151012.

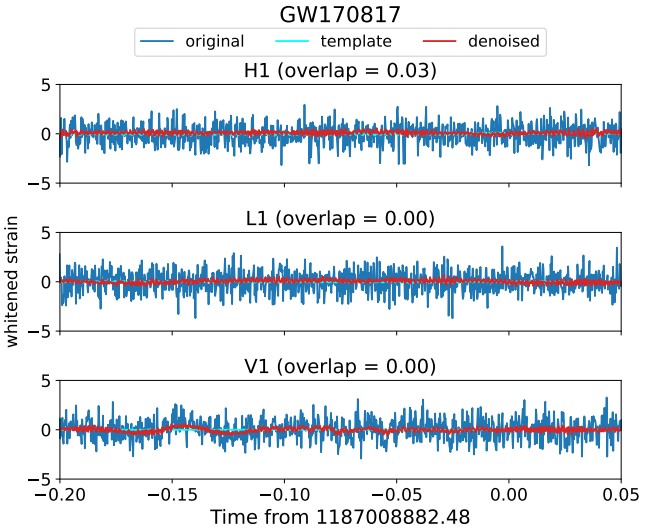


FIG. 9. The 0.25s windows of the denoised results of H1, L1 and V1 data around GW170817.

visually distinguishable structures that can possibly be recognized by the ML-based detection model after rescaling. However, these structures may not be adequate for parameter estimation, and more investigation will be needed.

We plot the overall result as the scatter plots of overlaps against the network SNR and the chirp mass of the source in Figure 10. The color of the dots represents the overlap value between 0 and 1. The overlaps are calculated using the 0.25s segment around the peak of the reconstructed waveform. If we see the denoised signal with overlap larger than 0.2 as successfully denoised, for LIGO H1 and L1 data, our model can denoise about 67.07% and 76.19% of the total events and can denoise over 90% of BBH events when the network SNR is larger than 11 and 12, respectively. On the other hand, for the Virgo data, our model can only denoise 10.87% of the total events and denoise over 90% of BBH events when the network SNR is larger than 20. This may be because the Virgo detector is less sensitive than the LIGO detectors, and the GW features in the Virgo time series data are not significant enough to be captured by our model. In addition, from Figure 10, we can see that our model predictions mostly have high overlaps on the LIGO H1 and L1 event data with high chirp masses or network SNRs and have low overlaps on the events with low chirp masses or network SNRs, which is in agreement with the mock data results in Section III A. The blue dot in the top left of each plot is the GW170817 BNS event, which our model fails to denoise in all detector data, although having the highest network SNR. This also agrees with the mock data results, which show that our model performs poorly on BNS signals.

C. Testing on Transient Noises

We also tested our model with the strain data containing glitches, as we expect there to be some glitches in the background data we used. These non-Gaussian features

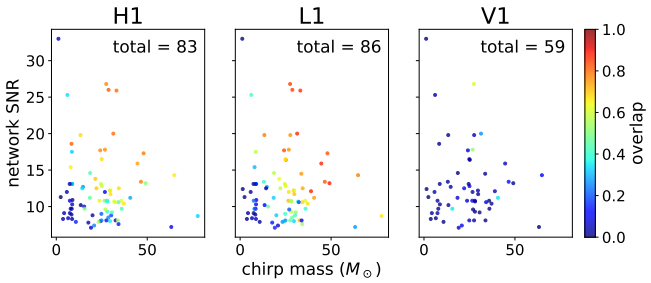


FIG. 10. The scatter plot of the overlap versus the network SNR and the chirp mass of the events. The color of the dots represents the value of overlaps. The overlaps are calculated using the 0.25s data segment around the peak of the reconstructed waveform.

can be probed when training the model with the Gaussian noise likelihood. We tested our model with three types of low-frequency glitches in LIGO O3b data, each with a different morphology in the spectrogram: Blip, Scattered Light, and Fast Scattering. To find the timestamp of these glitches, we refer to the public classification results of H1 and L1 O3b data from the Gravity Spy project [115–119], which is the machine learning project to classify the glitches found by the Omicron pipeline [120]. For each type of glitch, we randomly chose three glitch events with their SNR lying between 7.5 and 15 and their ML confidence larger than 0.95.

We present the denoised results of 2-second H1 and L1 data segments containing Blip, Scattered Light, and Fast Scattering glitches in Figure 11 and 12, respectively. Each figure contains three rows of plots corresponding to Blip, Scattered Light, and Fast Scattering glitches. In each row, there are denoised results of three strain segments. For each segment, we present the original strain time series (blue curve) and the denoised time series (red curve) in the top panel, and in the bottom panel, we plot the spectrograms of the original strain data on the left side, and the denoised data on the right side, respectively.

Our model can extract the structure of Blip glitches from time series data and emphasize their morphologies in the spectrogram for both H1 and L1 data. However, for Scattered Light and Fast Scattering glitches with low SNR, the extracted structures are insignificant in time series data, and some of their morphologies are also slightly altered in the spectrogram, especially in L1 data, which may affect the accuracy and confidence of glitch classification tasks. The removal of low SNR glitches may be due to their low amplitude in strain data; therefore, our model does not pick up their non-Gaussianities. In addition, Fast Scattering glitches are more frequent during the O3 run [121, 122], whereas our training dataset contains mostly O2 data.

From the perspective of GW detection, although our model also extracts glitch features, most of the glitches have very different structural characteristics from CBC signals, which we believe can be easily distinguished by a simple classifier model after rescaling the result. However, our model will not extract the GW signal alone when the signal overlaps with a glitch. The method to separate GW signals from glitches is left for future work.

D. Early Detection

Since the input data during training may contain only inspiral signals, it is interesting to see whether our model can extract GW inspiral features when the signal is loud. Therefore, we performed three tests. We randomly chose a 16-second data segment from H1 and L1 O3b strain data in each test. For the first test, we injected a BBH signal similar to the event GW170608 with component masses $m_1 = 11.0\,M_\odot$ and $m_2 = 7.6\,M_\odot$, and for the second test, we injected a BNS signal with component masses $m_1 = 1.7\,M_\odot$ and $m_2 = 1.5\,M_\odot$. Finally, for the third test, we injected an NSBH signal similar to the event GW200115_042309 with component masses $m_1 = 5.9\,M_\odot$ and $m_2 = 1.44\,M_\odot$. The signals were generated with the optimized orientation of the H1 detector. During the test, we adjust their optimal SNR from 20 to 50 in step 5. We then slice and whiten the data into several 2-second time windows with 1-second overlap and feed them into our model. We calculate the overlap between the denoised result and the injected GW signal in each window.

We plot the test results in Figure 13, Figure 14, and Figure 15. In each plot, the blue, orange, green, red, purple, brown, and pink dots correspond to the overlap results of each time window with injections of SNR $20, 25, 30, 35, 40, 45$, and 50, respectively. The gray curve is a normalized waveform to indicate the shape and position of the injected waveform in the strain data. The x-axis is the time distance of the center of time windows from the merger time. When the optimal SNR of the signal is larger than 50, our model can extract GW features with overlap larger than 0.2 in both detector data 2 seconds before the merger for BBH signals, 6 seconds for BNS signals, and 4 seconds before the merger for NSBH signals, respectively. Although the result of the BBH signal is inadequate for low-latency early-warning systems since we are using a 2-second window and need additional 1-second data at both sides for whitening, our model can extract inspiral features of NSBH and BNS events for other early-warning detection models when the SNR is high.

E. Test on CCSN waveforms

Our previous results show that our model performs better on short-duration signals. Therefore, our model on CCSN waveforms is worth testing, although we did not put them in the training dataset. We test our model using the publicly available GW waveforms from Vartanyan's simulations [123]. The simulations contain 11

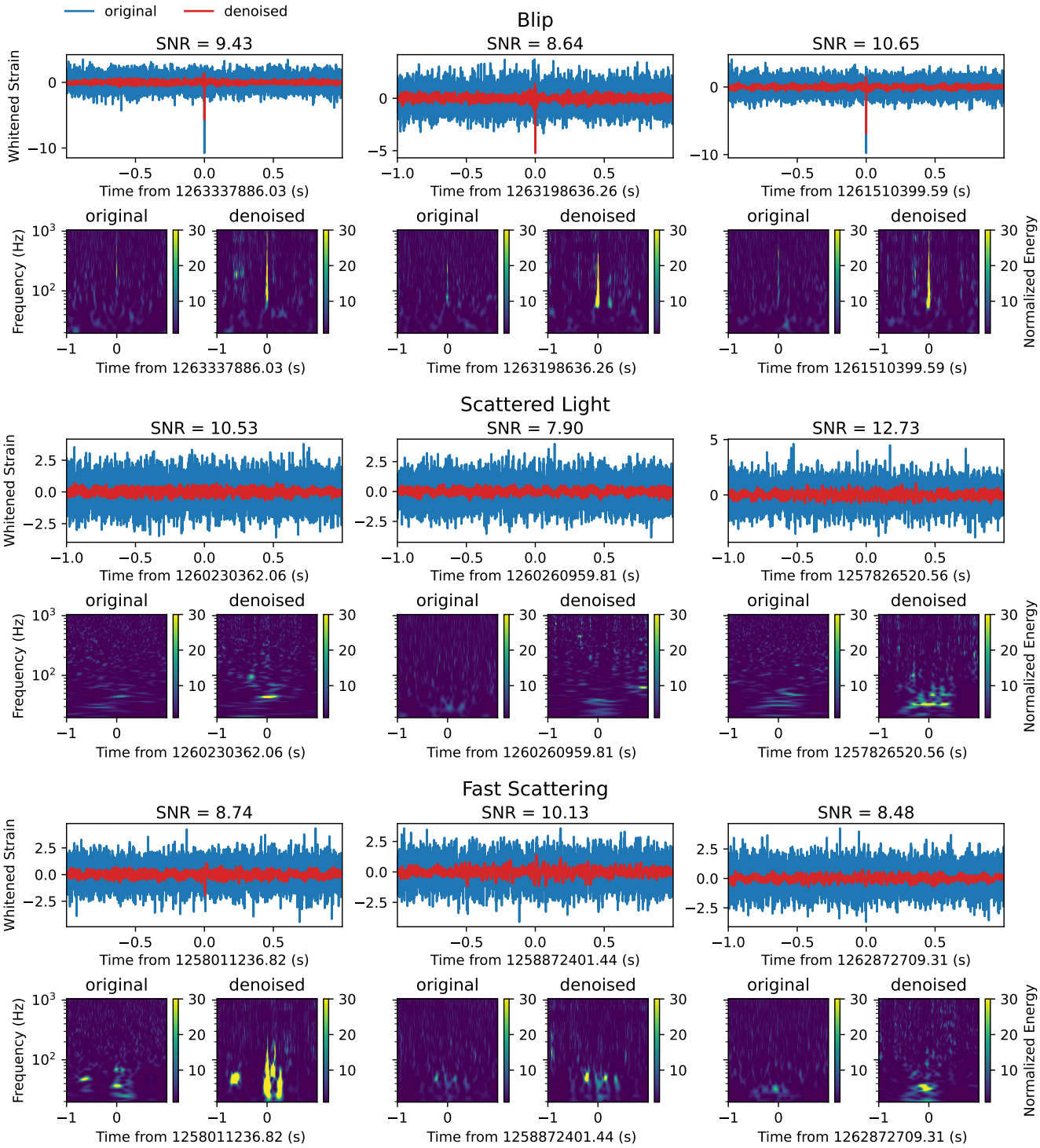


FIG. 11. The denoised results of the 2-second H1 O3b data containing three types of low-frequency glitches: Blip (top row), Scattered Light (middle row), and Fast Scattering (bottom row). The strain segments are randomly chosen according to the glitch timestamps in the Gravity Spy O3b ML classification results with ML confidence larger than 0.95 and glitch SNR between 7.5 and 15. The glitch SNRs are calculated by the Omicron pipeline. Each panel contains denoised results of three strain segments. The upper part of the denoised result of each segment is the time series data, the blue curve is the whitened strain data, and the red curve is the denoised results. The two plots below each time series plot are the spectrogram of the original strain data (left) and the denoised data (right). Our model can extract low-frequency glitches and emphasize their morphology in the spectrogram.

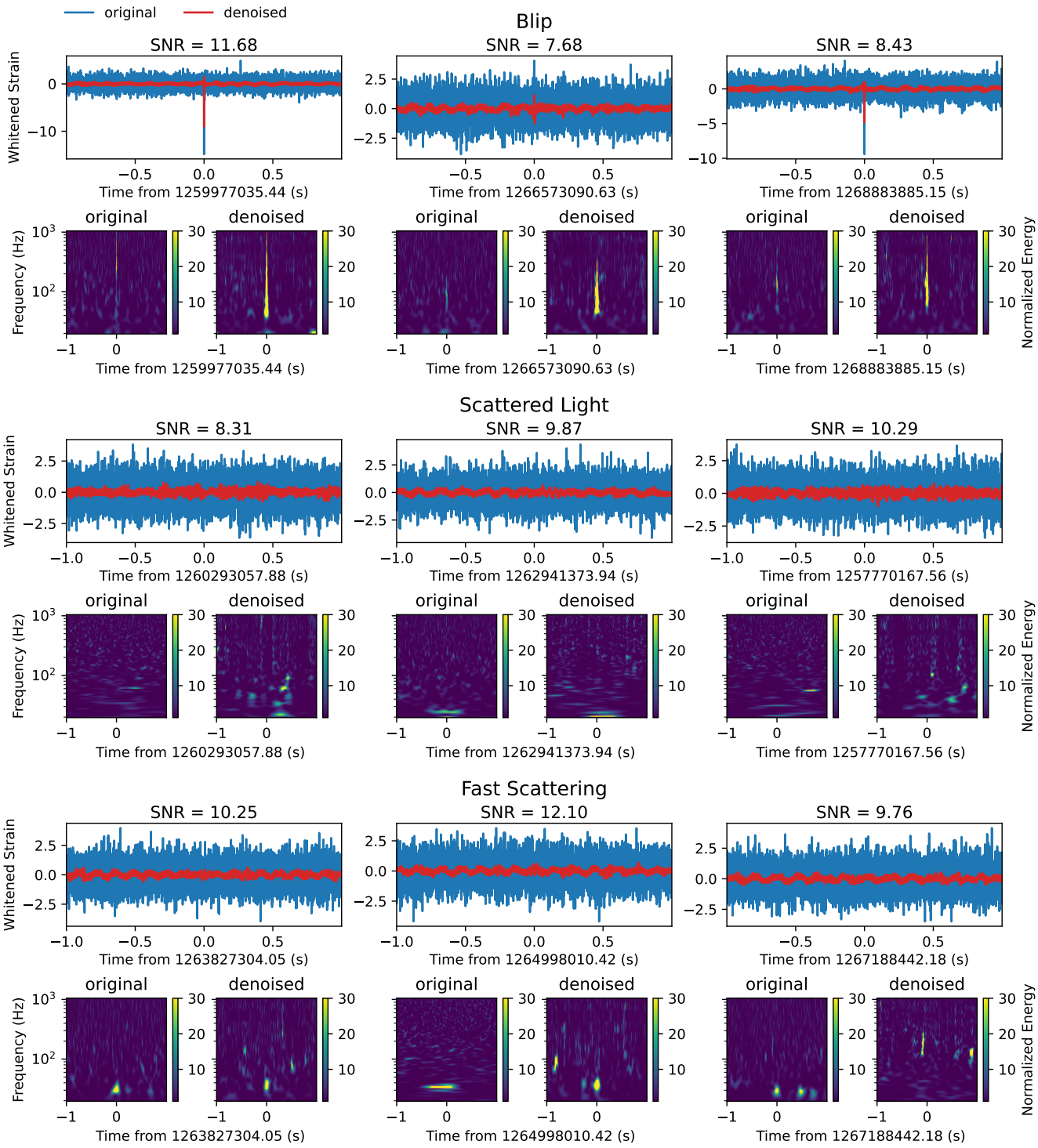


FIG. 12. The denoised results of the 2-second L1 O3b data containing three types of low-frequency glitches: Blip (top row), Scattered Light (middle row), and Fast Scattering (bottom row). Again, the strain segments are randomly chosen according to the glitch timestamps in the Gravity Spy O3b ML classification results with ML confidence larger than 0.95 and glitch SNR between 7.5 and 15. The conventions in plots are the same as those in Figure 11.

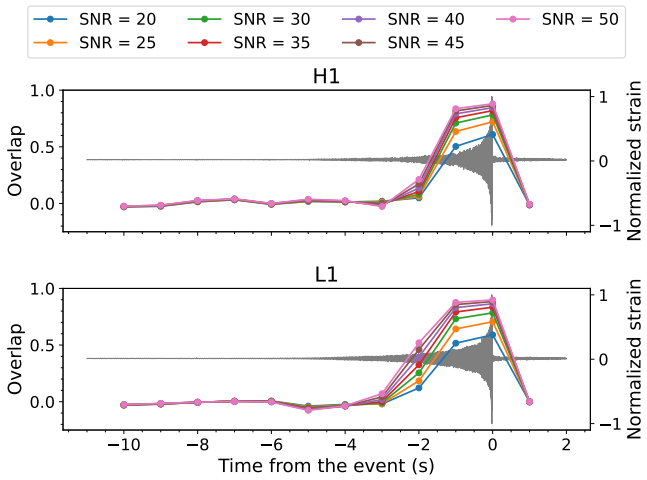


FIG. 13. The overlap result of the time windows containing a BBH signal with component masses $m_1 = 11.0\,M_\odot$ and $m_2 = 7.6\,M_\odot$. The gray curve is a normalized waveform to indicate the signal's shape and position in the strain data.

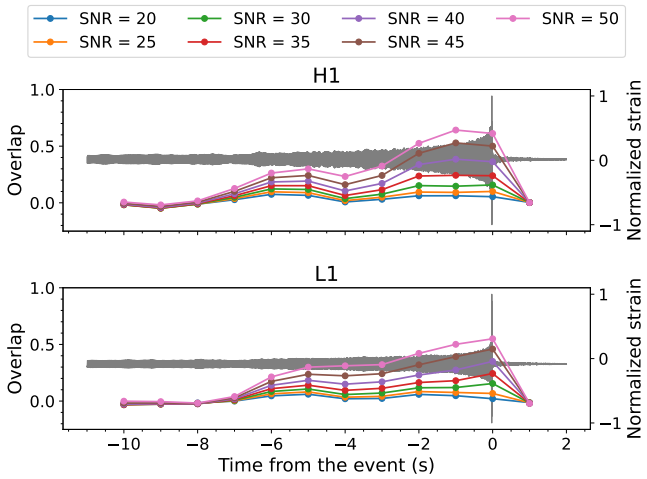


FIG. 14. The overlap result of the time windows containing a BNS signal with component masses $m_1 = 1.7\,M_\odot$ and $m_2 = 1.5\,M_\odot$. The gray curve is a normalized waveform to indicate the signal's shape and position in the strain data.

neutrino-driven, non-rotating progenitor models, including four $9M_\odot$ models simulated on different computing clusters and models with progenitor masses 9.25, 9.5, 11, 12.25, 14, 15.01, and $23M_\odot$. All progenitor models are employed with 12 neutrino energy groups and the SFHo equation of state [123, 124]. Among these models, the $12.25M_\odot$ and $14M_\odot$ progenitors failed to explode. The waveforms are observed along the positive x axis in the simulation frame. Also, note that the waveforms we used only come from the matter terms of the simulation and do not contain signals from neutrino memory [125].

The original waveform data are sampled at various sampling rates. To match the sampling rate we used in this work, we first applied a Butterworth low-pass filter at 2048 Hz to prevent aliasing, then downsampled the data to the sampling rate of 4096 Hz. We also used a high-pass filter at 15 Hz to remove the low-frequency components in the data; otherwise, the data can have a large offset from 0, impacting the model performance. We randomly chose 4-second data segments from H1 and L1 O3b strain data to perform the test. We injected a CCSN signal located at the optimal direction of each detector, with its starting time at the 1.25 second of the data. We then whitened the data and kept the center 2-second segment as the model input. During the test, we use the waveforms from CCSN models described above and adjust their $\rho_{\rm opt}$ from 10 to 40. We calculate the overlap between model predictions and injected GW signals. Although, in reality, it is unlikely to search for CCSN signals using matched filtering, the matched filter SNR can help us scale the waveform better so that the statistics of the strain data will not deviate too much from the training dataset.

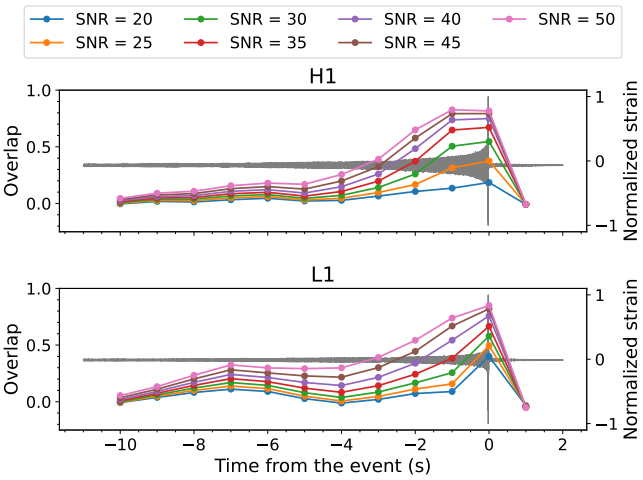


FIG. 15. The overlap result of the time windows containing an NSBH signal with component masses $m_1 = 5.9\,M_\odot$ and $m_2 = 1.44\,M_\odot$. The gray curve is a normalized waveform to indicate the signal's shape and position in the strain data.

We present the test result in Figure 17. In each plot, the blue, light blue, orange, light orange, green, light green, red, pink, violet, light purple, and purple triangle dots correspond to the overlap results of progenitor models $9a$ $M_\odot$, $9b$ $M_\odot$, $9_{\rm BW}$ $M_\odot$, $9_{\rm TH}$ $M_\odot$, 9.25 $M_\odot$, 9.5 $M_\odot$, 11 $M_\odot$, 12.25 $M_\odot$, 14 $M_\odot$, 15.01 $M_\odot$, and 23 $M_\odot$, respectively. Our model can partially extract features of all waveform models when $\rho_{\rm opt} \geq 25$ except for the 15.01 $M_\odot$ model in the H1 detector and 11 $M_\odot$ model in both detectors. The poor performance of these progenitor models is due to their longer waveform duration and, therefore, smaller amplitudes when scaling to the same SNR.

The test result shows that our model can extract short-duration CCSN signals even if it does not see them during the training stage. In the future, we will inject more CCSN waveforms into the training dataset to improve the

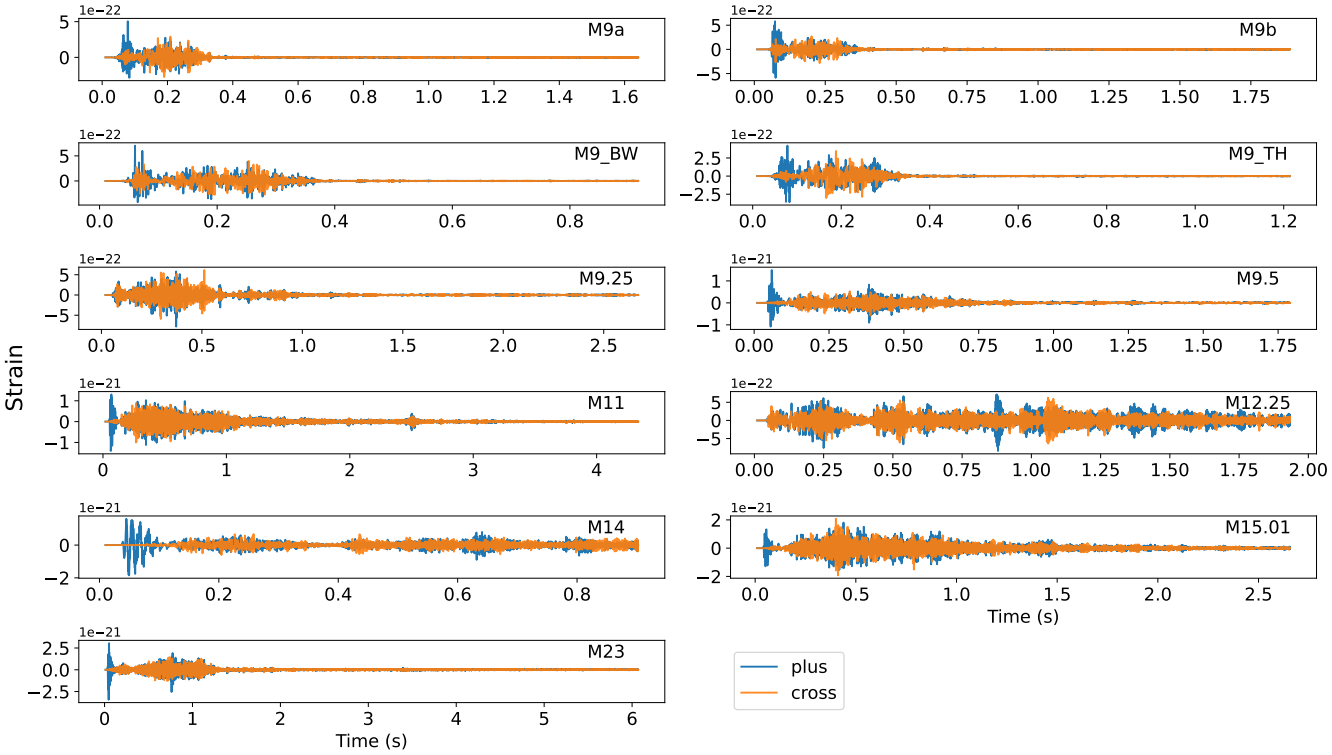


FIG. 16. The plus (blue curves) and cross (yellow curves) polarizations of CCSN waveforms for testing our model in this work. The waveforms are generated from 3D simulations in Ref. [123] with various progenitor mass models. We applied an anti-aliasing filter before downsampling to 4096 Hz and a low-pass filter at 15 Hz to remove low-frequency components making large deviations from 0. The waveforms are observed along the positive x axis of the simulation frame, and we scale the waveforms so that the sources are placed at a distance of 1kpc.

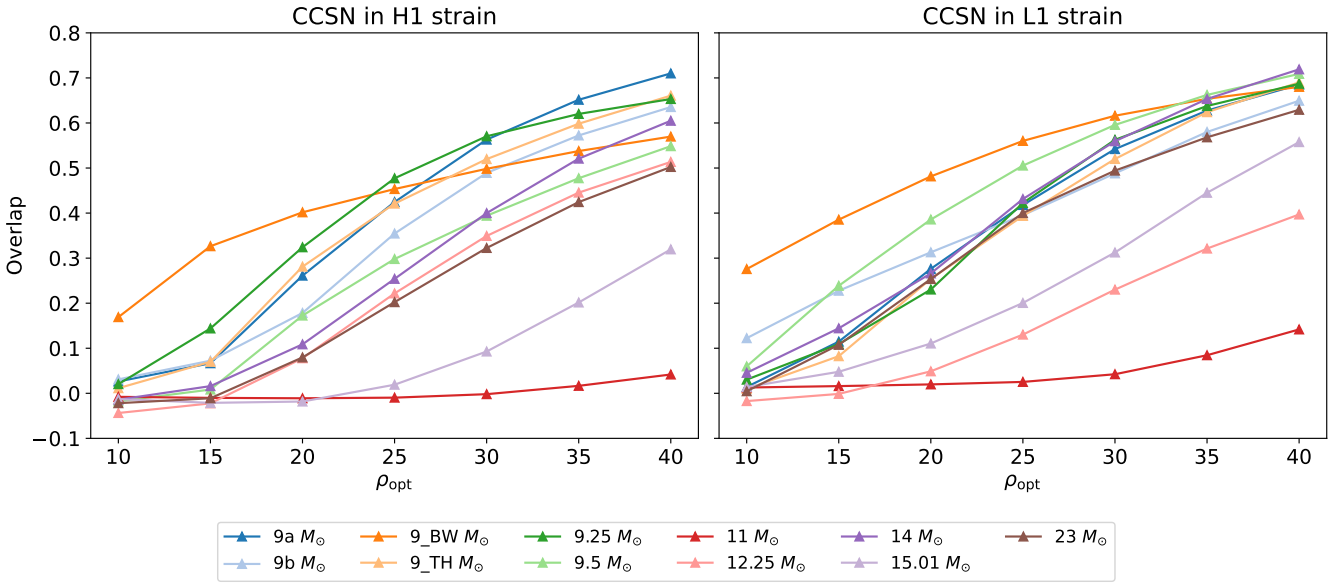


FIG. 17. The performance of our model on different CCSN models at different optimal SNR.

model performance and robustness against CCSN signals.

IV. CONCLUSION

We developed a self-supervised learning denoiser model to extract GW features from the time series data without explicitly providing cleaned GW waveforms during training. We trained our model using the Gaussian noise likelihood assumption. By setting the minimum overlap to have distinguishable patterns that the ML-based detection model can recognize to be 0.2, our model can denoise over 90% BBH and NSBH signals in the 0.25s window when the signal SNR is larger than 13 and 16, respectively. Our model can denoise over 50% of the signals for BNS signals when SNR is larger than 23.

We also tested our model on the actual events detected by LIGO and Virgo during O1, O2, and O3 observation runs. Our model can denoise about 67% and 76% of the events to have overlaps exceeding 0.2 in the H1 and L1 data, respectively. For the Virgo data, our model can only denoise about 11% of the events to have overlap exceeding 0.2 due to the large noise background in the data. In addition, our model can denoise over 90% of BBH signals in H1, L1, and V1 event data when the network SNR is larger than 11, 12, and 20, respectively. We also tested our model on the strain data containing glitches and found that our model can effectively reveal the glitch structures that can be easily distinguished from CBC signals. In addition, our model shows the potential to extract features and allow detection models to send early warning alerts for BNS and NSBH events when the SNR is high. Furthermore, our model can partially extract loud CCSN features even if not added to the training dataset.

Our work shows the potential of using self-supervised learning to extract GW features from time series data by training with injected strain data alone. In the future, we plan to use different noise models and apply subtraction techniques to remove loud line noise beforehand to enhance the model's performance in extracting low-amplitude features. We also plan to perform quantitative investigations on the relation between overlap and downstream tasks. We will also investigate the possibility of using our model to extract GW features from other types of GW signals, such as GWs from other CCSN models or continuous waves from spinning neutron stars.

ACKNOWLEDGMENTS

This research has made use of data or software obtained from the Gravitational Wave Open Science Center (gw-openscience.org), a service of LIGO Laboratory, the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA. LIGO Laboratory and Advanced LIGO are funded by the United States National Science Foundation (NSF) as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. The construction and operation of KAGRA are funded by Ministry of Education, Culture, Sports, Science and Technology (MEXT), and Japan Society for the Promotion of Science (JSPS), National Research Foundation (NRF) and Ministry of Science and ICT (MSIT) in Korea, Academia Sinica (AS) and the National Science and Technology Council (NSTC) in Taiwan.

We acknowledge the support from NSTC through grants 112-2811-M-007-041 and 112-2112-M-007-042. This work used high-performance computing facilities operated by the Center for Informatics and Computation in Astronomy (CICA) at National Tsing Hua University. This equipment was funded by the Ministry of Education of Taiwan, NSTC, and National Tsing Hua University.

We thank Suroji Saha, En-Tzu Lin, and Chun-Hsiang Chan for suggestions and discussions during the writing of this paper.

[1] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Observation of gravitational waves from a binary black hole merger, Phys. Rev. Lett. **116**, 061102 (2016).

[2] J. Aasi *et al.* (LIGO Scientific), Advanced LIGO, Class. Quant. Grav. **32**, 074001 (2015), arXiv:1411.4547 [gr-qc].

[3] F. Acernese *et al.*, Advanced virgo: a second-generation interferometric gravitational wave detector, Classical and Quantum Gravity **32**, 024001 (2014).

[4] B. P. Abbott *et al.* (LIGO Scientific, Virgo), GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, Phys. Rev. Lett. **119**, 161101 (2017), arXiv:1710.05832 [gr-qc].

[5] E. Troja *et al.*, The X-ray counterpart to the gravitational wave event GW 170817, Nature **551**, 71 (2017), arXiv:1710.05433 [astro-ph.HE].

[6] P. A. Evans *et al.*, Swift and NuSTAR observations of GW170817: detection of a blue kilonova, Science **358**, 1565 (2017), arXiv:1710.05437 [astro-ph.HE].

[7] G. Hallinan *et al.*, A radio counterpart to a neutron star merger, Science **358**, 1579 (2017), arXiv:1710.05435

[astro-ph.HE].

[8] B. P. Abbott *et al.* (LIGO Scientific, Virgo, Fermi GBM, INTEGRAL, IceCube, AstroSat Cadmium Zinc Telluride Imager Team, IPN, Insight-Hxmt, ANTARES, Swift, AGILE Team, 1M2H Team, Dark Energy Camera GW-EM, DES, DLT40, GRAWITA, Fermi-LAT, ATCA, ASKAP, Las Cumbres Observatory Group, OzGrav, DWF (Deeper Wider Faster Program), AST3, CAASTRO, VINROUGE, MASTER, J-GEM, GROWTH, JAGWAR, CaltechNRAO, TTU-NRAO, NuSTAR, Pan-STARRS, MAXI Team, TZAC Consortium, KU, Nordic Optical Telescope, ePESSTO, GROND, Texas Tech University, SALT Group, TOROS, BOOTES, MWA, CALET, IKI-GW Follow-up, H.E.S.S., LOFAR, LWA, HAWC, Pierre Auger, ALMA, Euro VLBI Team, Pi of Sky, Chandra Team at McGill University, DFN, ATLAS Telescopes, High Time Resolution Universe Survey, RIMAS, RATIR, SKA South Africa/MeerKAT), Multi-messenger Observations of a Binary Neutron Star Merger, Astrophys. J. Lett. **848**, L12 (2017), arXiv:1710.05833 [astro-ph.HE].

[9] B. P. Abbott *et al.* (LIGO Scientific, Virgo, Fermi-GBM, INTEGRAL), Gravitational Waves and Gamma-rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A, Astrophys. J. Lett. **848**, L13 (2017), arXiv:1710.05834 [astro-ph.HE].

[10] K. P. Mooley *et al.*, A mildly relativistic wide-angle outflow in the neutron star merger GW170817, Nature **554**, 207 (2018), arXiv:1711.11573 [astro-ph.HE].

[11] B. P. Abbott *et al.* (LIGO Scientific, Virgo), GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs, Phys. Rev. **X9**, 031040 (2019), arXiv:1811.12907 [astro-ph.HE].

[12] R. Abbott *et al.* (LIGO Scientific, Virgo), GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run, Phys. Rev. X **11**, 021053 (2021), arXiv:2010.14527 [gr-qc].

[13] R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), GWTC-2.1: Deep Extended Catalog of Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run, arXiv e-prints , arXiv:2108.01045 (2021), arXiv:2108.01045 [gr-qc].

[14] R. Abbott *et al.* (KAGRA, VIRGO, LIGO Scientific), 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), arXiv:2111.03606 [gr-qc].

[15] B. P. Abbott *et al.* (LIGO Scientific, Virgo), GW190425: Observation of a Compact Binary Coalescence with Total Mass $\sim 3.4 M_{\odot}$, Astrophys. J. Lett. **892**, L3 (2020), arXiv:2001.01761 [astro-ph.HE].

[16] Https://gracedb.ligo.org/superevents/public/O4/.

[17] T. Akutsu *et al.*, Kagra: 2.5 generation interferometric gravitational wave detector, Nature Astronomy **3**, 35 (2019).

[18] T. Akutsu *et al.*, Overview of KAGRA: KAGRA science, Progress of Theoretical and Experimental Physics **2021**, 05A103 (2020), https://academic.oup.com/ptep/article-pdf/2021/5/05A103/37953064/ptaa120.pdf.

[19] K. L. Dooley *et al.*, GEO 600 and the GEO-HF upgrade program: successes and challenges, Class. Quant. Grav. **33**, 075009 (2016), arXiv:1510.00317 [physics.ins-det].

[20] J. Lough, E. Schreiber, F. Bergamin, H. Grote, M. Mehmet, H. Vahlbruch, *et al.*, First Demonstration of 6 dB Quantum Noise Reduction in a Kilometer Scale Gravitational Wave Observatory, Phys. Rev. Lett. **126**, 041102 (2021), arXiv:2005.10292 [physics.ins-det].

[21] H. Abe *et al.* (KAGRA), Performance of the KAGRA detector during the first joint observation with GEO 600 (O3GK), PTEP **2023**, 10A101 (2023), arXiv:2203.07011 [astro-ph.IM].

[22] R. Abbott *et al.* (KAGRA, VIRGO, LIGO Scientific), First joint observation by the underground gravitational-wave detector KAGRA with GEO 600, PTEP **2022**, 063F01 (2022), arXiv:2203.01270 [gr-qc].

[23] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Search for subsolar-mass ultracompact binaries in advanced ligo's first observing run, Phys. Rev. Lett. **121**, 231103 (2018).

[24] B. P. Abbott *et al.* (LIGO Scientific Collaboration and the Virgo Collaboration), Search for subsolar mass ultracompact binaries in advanced ligo's second observing run, Phys. Rev. Lett. **123**, 161102 (2019).

[25] A. H. Nitz and Y.-F. Wang, Search for gravitational waves from the coalescence of subsolar mass and eccentric compact binaries, The Astrophysical Journal **915**, 54 (2021).

[26] A. H. Nitz and Y.-F. Wang, Search for gravitational waves from the coalescence of subsolar-mass binaries in the first half of advanced ligo and virgo's third observing run, Phys. Rev. Lett. **127**, 151101 (2021).

[27] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), Search for Subsolar-Mass Binaries in the First Half of Advanced LIGO's and Advanced Virgo's Third Observing Run, Phys. Rev. Lett. **129**, 061104 (2022), arXiv:2109.12197 [astro-ph.CO].

[28] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), Search for subsolar-mass black hole binaries in the second part of Advanced LIGO's and Advanced Virgo's third observing run, arXiv e-prints (2022), arXiv:2212.01477 [astro-ph.HE].

[29] A. H. Nitz and Y.-F. Wang, Broad search for gravitational waves from subsolar-mass binaries through ligo and virgo's third observing run, Phys. Rev. D **106**, 023024 (2022).

[30] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Optically targeted search for gravitational waves emitted by core-collapse supernovae during the first and second observing runs of advanced LIGO and advanced Virgo, Phys. Rev. D **101**, 084002 (2020), arXiv:1908.03584 [astro-ph.HE].

[31] R. Abbott *et al.* (KAGRA, VIRGO, LIGO Scientific), All-sky search for short gravitational-wave bursts in the third Advanced LIGO and Advanced Virgo run, Phys. Rev. D **104**, 122004 (2021), arXiv:2107.03701 [gr-qc].

[32] R. Abbott *et al.* (KAGRA, VIRGO, LIGO Scientific), All-sky search for long-duration gravitational-wave bursts in the third Advanced LIGO and Advanced Virgo run, Phys. Rev. D **104**, 102001 (2021), arXiv:2107.13796 [gr-qc].

[33] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), Search for gravitational-wave transients associated with magnetar bursts in Advanced LIGO and Advanced

Virgo data from the third observing run, arXiv e-prints (2022), arXiv:2210.10931 [astro-ph.HE].

[34] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA, CHIME/FRB), Search for Gravitational Waves Associated with Fast Radio Bursts Detected by CHIME/FRB during the LIGO–Virgo Observing Run O3a, Astrophys. J. **955**, 155 (2023), arXiv:2203.12038 [astro-ph.HE].

[35] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), Search for Gravitational Waves Associated with Gamma-Ray Bursts Detected by Fermi and Swift during the LIGO–Virgo Run O3b, Astrophys. J. **928**, 186 (2022), arXiv:2111.03608 [astro-ph.HE].

[36] R. Abbott *et al.* (LIGO Scientific, Virgo, VIRGO), All-sky search in early O3 LIGO data for continuous gravitational-wave signals from unknown neutron stars in binary systems, Phys. Rev. D **103**, 064017 (2021), [Erratum: Phys.Rev.D 108, 069901 (2023)], arXiv:2012.12128 [gr-qc].

[37] R. Abbott *et al.* (KAGRA, VIRGO, LIGO Scientific), All-sky search for continuous gravitational waves from isolated neutron stars in the early O3 LIGO data, Phys. Rev. D **104**, 082004 (2021), arXiv:2107.00600 [gr-qc].

[38] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), Searches for Gravitational Waves from Known Pulsars at Two Harmonics in the Second and Third LIGO-Virgo Observing Runs, Astrophys. J. **935**, 1 (2022), arXiv:2111.13106 [astro-ph.HE].

[39] R. Abbott *et al.* (KAGRA, LIGO Scientific, VIRGO), Search for continuous gravitational wave emission from the Milky Way center in O3 LIGO-Virgo data, Phys. Rev. D **106**, 042003 (2022), arXiv:2204.04523 [astro-ph.HE].

[40] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), Search for continuous gravitational waves from 20 accreting millisecond x-ray pulsars in O3 LIGO data, Phys. Rev. D **105**, 022002 (2022), arXiv:2109.09255 [astro-ph.HE].

[41] R. Abbott *et al.* (LIGO Scientific, KAGRA, VIRGO), Model-based Cross-correlation Search for Gravitational Waves from the Low-mass X-Ray Binary Scorpius X-1 in LIGO O3 Data, Astrophys. J. Lett. **941**, L30 (2022), arXiv:2209.02863 [astro-ph.HE].

[42] S. Sachdev *et al.*, The GstLAL Search Analysis Methods for Compact Binary Mergers in Advanced LIGO's Second and Advanced Virgo's First Observing Runs, arXiv e-prints , arXiv:1901.08580 (2019), arXiv:1901.08580 [gr-qc].

[43] S. A. Usman *et al.*, The PyCBC search for gravitational waves from compact binary coalescence, Classical and Quantum Gravity **33**, 215004 (2016), arXiv:1508.02357 [gr-qc].

[44] S. Hooper, S. K. Chung, J. Luan, D. Blair, Y. Chen, and L. Wen, Summed Parallel Infinite Impulse Response (SPIIR) Filters For Low-Latency Gravitational Wave Detection, Phys. Rev. D **86**, 024012 (2012), arXiv:1108.3186 [gr-qc].

[45] Q. Chu, M. Kovalam, L. Wen, T. Slaven-Blair, J. Bosveld, Y. Chen, *et al.*, SPIIR online coherent pipeline to search for gravitational waves from compact binary coalescences, Phys. Rev. D **105**, 024023 (2022), arXiv:2011.06787 [gr-qc].

[46] F. Aubin *et al.*, The MBTA pipeline for detecting compact binary coalescences in the third LIGO–Virgo observing run, Class. Quant. Grav. **38**, 095004 (2021), arXiv:2012.11512 [gr-qc].

[47] L. Tsukada, P. Joshi, S. Adhicary, R. George, A. Guimaraes, C. Hanna, R. Magee, A. Zimmerman, P. Baral, A. Baylor, K. Cannon, S. Caudill, B. Cousins, J. D. E. Creighton, B. Ewing, H. Fong, P. Godwin, R. Harada, Y.-J. Huang, R. Huxford, J. Kennington, S. Kuwahara, A. K. Y. Li, D. Meacher, C. Messick, S. Morisaki, D. Mukherjee, W. Niu, A. Pace, C. Posnansky, A. Ray, S. Sachdev, S. Sakon, D. Singh, R. Tapia, T. Tsutsui, K. Ueno, A. Viets, L. Wade, and M. Wade, Improved ranking statistics of the gstlal inspiral search for compact binary coalescences, Phys. Rev. D **108**, 043004 (2023).

[48] C. Alléné *et al.*, The MBTA Pipeline for Detecting Compact Binary Coalescences in the Fourth LIGO-Virgo-KAGRA Observing Run, (2025), arXiv:2501.04598 [gr-qc].

[49] B. S. Sathyaprakash and B. F. Schutz, Physics, Astrophysics and Cosmology with Gravitational Waves, Living Rev. Rel. **12**, 2 (2009), arXiv:0903.0338 [gr-qc].

[50] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals, Class. Quant. Grav. **37**, 055002 (2020), arXiv:1908.11170 [gr-qc].

[51] B. J. Owen and B. S. Sathyaprakash, Matched filtering of gravitational waves from inspiraling compact binaries: Computational cost and template placement, Phys. Rev. D **60**, 022002 (1999).

[52] D. George and E. A. Huerta, Deep Neural Networks to Enable Real-time Multimessenger Astrophysics, Phys. Rev. **D97**, 044039 (2018), arXiv:1701.00008 [astro-ph.IM].

[53] D. George and E. A. Huerta, Deep Learning for Real-time Gravitational Wave Detection and Parameter Estimation: Results with Advanced LIGO Data, Phys. Lett. **B778**, 64 (2018), arXiv:1711.03121 [gr-qc].

[54] T. D. Gebhard, N. Kilbertus, I. Harry, and B. Schölkopf, Convolutional neural networks: a magic bullet for gravitational-wave detection? (2019) arXiv:1904.08693 [astro-ph.IM].

[55] H. Gabbard, M. Williams, F. Hayes, and C. Messenger, Matching matched filtering with deep networks for gravitational-wave astronomy, Phys. Rev. Lett. **120**, 141103 (2018), arXiv:1712.06041 [astro-ph.IM].

[56] J. a. D. Álvares, J. A. Font, F. F. Freitas, O. G. Freitas, A. P. Morais, S. Nunes, A. Onofre, and A. Torres-Forné, Gravitational-wave parameter inference using Deep Learning (2020) arXiv:2011.10425 [gr-qc].

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

[58] M. L. Chan, I. S. Heng, and C. Messenger, Detection and classification of supernova gravitational wave signals: A deep learning approach, Phys. Rev. D **102**, 043022 (2020), arXiv:1912.13517 [astro-ph.HE].

[59] V. Skliris, M. R. K. Norman, and P. J. Sutton, Real-Time Detection of Unmodelled Gravitational-Wave Transients Using Convolutional Neural Networks, arXiv e-prints , arXiv:2009.14611 (2020), arXiv:2009.14611 [astro-ph.IM].

[60] A. Iess, E. Cuoco, F. Morawski, and J. Powell, Core-Collapse Supernova Gravitational-Wave Search

and Deep Learning Classification, arXiv e-prints , arXiv:2001.00279 (2020), arXiv:2001.00279 [gr-qc].

[61] P. G. Krastev, K. Gill, V. A. Villar, and E. Berger, Detection and Parameter Estimation of Gravitational Waves from Binary Neutron-Star Mergers in Real LIGO Data using Deep Learning, Phys. Lett. B **815**, 136161 (2021), arXiv:2012.13101 [astro-ph.IM].

[62] Y.-C. Lin and J.-H. P. Wu, Detection of gravitational waves using Bayesian neural networks, Phys. Rev. D **103**, 063034 (2021), arXiv:2007.04176 [astro-ph.IM].

[63] A. Iess, E. Cuoco, F. Morawski, C. Nicolaou, and O. Lahav, LSTM and CNN application for core-collapse supernova search in gravitational wave real data, Astron. Astrophys. **669**, A42 (2023), arXiv:2301.09387 [astro-ph.IM].

[64] M. Llorens-Monteagudo, A. Torres-Forné, J. A. Font, and A. Marquina, Classification of gravitational-wave glitches via dictionary learning, Classical and Quantum Gravity **36**, 075005 (2019), arXiv:1811.03867 [astro-ph.IM].

[65] T. C. Abbott, E. Buffaz, N. Vieira, M. Cabero, D. Haggard, A. Mahabal, and J. McIver, GWSkyNet-Multi: A Machine-learning Multiclass Classifier for LIGO-Virgo Public Alerts, Astrophys. J. **927**, 232 (2022), arXiv:2111.04015 [astro-ph.IM].

[66] J. Yan, A. P. Leung, and D. C. Y. Hui, On improving the performance of glitch classification for gravitational wave detection by using Generative Adversarial Networks, Mon. Not. Roy. Astron. Soc. **515**, 4606 (2022), arXiv:2207.04001 [astro-ph.HE].

[67] C. Chatterjee, L. Wen, K. Vinsen, M. Kovalam, and A. Datta, Using Deep Learning to Localize Gravitational Wave Sources, Phys. Rev. D **100**, 103025 (2019), arXiv:1909.06367 [astro-ph.IM].

[68] A. Kolmus, G. Baltus, J. Janquart, T. van Laarhoven, S. Caudill, and T. Heskes, Fast sky localization of gravitational waves using deep learning seeded importance sampling, Phys. Rev. D **106**, 023032 (2022), arXiv:2111.00833 [gr-qc].

[69] C. Chatterjee, M. Kovalam, L. Wen, D. Beveridge, F. Diakogiannis, and K. Vinsen, Rapid Localization of Gravitational Wave Sources from Compact Binary Coalescences Using Deep Learning, Astrophys. J. **959**, 42 (2023), arXiv:2207.14522 [gr-qc].

[70] M. Dax, S. R. Green, J. Gair, M. Pürrer, J. Wildberger, J. H. Macke, A. Buonanno, and B. Schölkopf, Neural importance sampling for rapid and reliable gravitational-wave inference, Phys. Rev. Lett. **130**, 171403 (2023).

[71] K. W. K. Wong, M. Isi, and T. D. P. Edwards, Fast gravitational-wave parameter estimation without compromises, The Astrophysical Journal **958**, 129 (2023).

[72] A. Torres-Forné, A. Marquina, J. A. Font, and J. M. Ibáñez, Denoising of gravitational wave signals via dictionary learning algorithms, Phys. Rev. D **94**, 124040 (2016), arXiv:1612.01305 [astro-ph.IM].

[73] W. Wei and E. A. Huerta, Gravitational wave denoising of binary black hole mergers with deep learning, Physics Letters B **800**, 135081 (2020), arXiv:1901.00869 [gr-qc].

[74] H. Shen, D. George, E. A. Huerta, and Z. Zhao, Denoising Gravitational Waves with Enhanced Deep Recurrent Denoising Auto-Encoders, arXiv e-prints , arXiv:1903.03105 (2019), arXiv:1903.03105 [astro-ph.CO].

[75] G. Vajente, Y. Huang, M. Isi, J. C. Driggers, J. S. Kissel, M. J. Szczepańczyk, and S. Vitale, Machine-learning nonstationary noise out of gravitational-wave detectors, Phys. Rev. D **101**, 042003 (2020).

[76] R. Ormiston, T. Nguyen, M. Coughlin, R. X. Adhikari, and E. Katsavounidis, Noise Reduction in Gravitational-wave Data via Deep Learning, Phys. Rev. Res. **2**, 033066 (2020), arXiv:2005.06534 [astro-ph.IM].

[77] R. Essick, P. Godwin, C. Hanna, L. Blackburn, and E. Katsavounidis, idq: Statistical inference of non-gaussian noise with auxiliary degrees of freedom in gravitational-wave detectors, Machine Learning: Science and Technology **2**, 015004 (2020).

[78] K. Mogushi, R. Quitzow-James, M. Cavaglià, S. Kulkarni, and F. Hayes, NNETFIX: an artificial neural network-based denoising engine for gravitational-wave signals, Mach. Learn. Sci. Tech. **2**, 035018 (2021), arXiv:2101.04712 [gr-qc].

[79] C. Chatterjee, L. Wen, F. Diakogiannis, and K. Vinsen, Extraction of binary black hole gravitational wave signals from detector data using deep learning, Phys. Rev. D **104**, 064046 (2021), arXiv:2105.03073 [gr-qc].

[80] P. Bacon, A. Trovato, and M. Bejger, Denoising gravitational-wave signals from binary black holes with a dilated convolutional autoencoder, Mach. Learn. Sci. Tech. **4**, 035024 (2023), arXiv:2205.13513 [gr-qc].

[81] H. Wang, Y. Zhou, Z. Cao, Z. Guo, and Z. Ren, Waveformer: transformer-based denoising method for gravitational-wave data, Machine Learning: Science and Technology **5**, 015046 (2024).

[82] M. Saleem, A. Gunny, C.-J. Chou, L.-C. Yang, S.-W. Yeh, A. H. Y. Chen, R. Magee, W. Benoit, T. Nguyen, P. Fan, D. Chatterjee, E. Marx, E. Moreno, R. Omer, R. Raikman, D. Rankin, R. Sharma, M. Coughlin, P. Harris, and E. Katsavounidis, Demonstration of machine learning-assisted low-latency noise regression in gravitational wave detectors, Classical and Quantum Gravity **41**, 195024 (2024).

[83] A. Torres, A. Marquina, J. A. Font, and J. M. Ibáñez, Total-variation-based methods for gravitational wave denoising, Phys. Rev. D **90**, 084029 (2014), arXiv:1409.7888 [gr-qc].

[84] A. Torres-Forné, E. Cuoco, A. Marquina, J. A. Font, and J. M. Ibáñez, Total-variation methods for gravitational-wave denoising: performance tests on Advanced LIGO data, Phys. Rev. D **98**, 084013 (2018), arXiv:1806.07329 [astro-ph.SR].

[85] C. Doersch, A. Gupta, and A. A. Efros, Unsupervised Visual Representation Learning by Context Prediction, arXiv e-prints , arXiv:1505.05192 (2015), arXiv:1505.05192 [cs.CV].

[86] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, Context Encoders: Feature Learning by Inpainting, arXiv e-prints , arXiv:1604.07379 (2016), arXiv:1604.07379 [cs.CV].

[87] S. Gidaris, P. Singh, and N. Komodakis, Unsupervised Representation Learning by Predicting Image Rotations, arXiv e-prints , arXiv:1803.07728 (2018), arXiv:1803.07728 [cs.CV].

[88] A. Krull, T.-O. Buchholz, and F. Jug, Noise2Void - Learning Denoising from Single Noisy Images, arXiv e-prints , arXiv:1811.10980 (2018), arXiv:1811.10980 [cs.CV].

[89] X. Chen, S. Xie, and K. He, An Empirical Study of Training Self-Supervised Vision Transformers, arXiv

e-prints , arXiv:2104.02057 (2021), arXiv:2104.02057 [cs.CV].

[90] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, arXiv e-prints , arXiv:1810.04805 (2018), arXiv:1810.04805 [cs.CL].

[91] OpenAI, :, J. Achiam, *et al.*, GPT-4 Technical Report, arXiv e-prints , arXiv:2303.08774 (2023), arXiv:2303.08774 [cs.CL].

[92] R. Zhang, P. Isola, and A. A. Efros, Split-Brain Autoencoders: Unsupervised Learning by Cross-Channel Prediction, arXiv e-prints , arXiv:1611.09842 (2016), arXiv:1611.09842 [cs.CV].

[93] L. Jing and Y. Tian, Self-supervised Visual Feature Learning with Deep Neural Networks: A Survey, arXiv e-prints , arXiv:1902.06162 (2019), arXiv:1902.06162 [cs.CV].

[94] S. Laine, T. Karras, J. Lehtinen, and T. Aila, High-quality self-supervised deep image denoising, Advances in Neural Information Processing Systems **32** (2019).

[95] K. Kim and J. C. Ye, Noise2Score: Tweedie's Approach to Self-Supervised Image Denoising without Clean Images, arXiv e-prints , arXiv:2106.07009 (2021), arXiv:2106.07009 [eess.IV].

[96] Y. Xie, Z. Wang, and S. Ji, Noise2Same: Optimizing A Self-Supervised Bound for Image Denoising, arXiv e-prints , arXiv:2010.11971 (2020), arXiv:2010.11971 [cs.CV].

[97] A. van den Oord, N. Kalchbrenner, O. Vinyals, L. Espeholt, A. Graves, and K. Kavukcuoglu, Conditional Image Generation with PixelCNN Decoders, arXiv e-prints , arXiv:1606.05328 (2016), arXiv:1606.05328 [cs.CV].

[98] A. van den Oord, N. Kalchbrenner, and K. Kavukcuoglu, Pixel Recurrent Neural Networks, arXiv e-prints , arXiv:1601.06759 (2016), arXiv:1601.06759 [cs.CV].

[99] T. Salimans, A. Karpathy, X. Chen, and D. P. Kingma, PixelCNN++: Improving the PixelCNN with Discretized Logistic Mixture Likelihood and Other Modifications, arXiv e-prints , arXiv:1701.05517 (2017), arXiv:1701.05517 [cs.LG].

[100] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, WaveNet: A Generative Model for Raw Audio, arXiv e-prints , arXiv:1609.03499 (2016), arXiv:1609.03499 [cs.SD].

[101] K. He, X. Zhang, S. Ren, and J. Sun, Deep Residual Learning for Image Recognition, arXiv e-prints , arXiv:1512.03385 (2015), arXiv:1512.03385 [cs.CV].

[102] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, Fast and accurate deep network learning by exponential linear units (elus), arXiv preprint arXiv:1511.07289 (2015).

[103] R. Abbott *et al.*, Open data from the first and second observing runs of advanced ligo and advanced virgo, SoftwareX **13**, 100658 (2021).

[104] A. Nitz, I. Harry, D. Brown, C. M. Biwer, J. Willis, T. D. Canton, C. Capano, T. Dent, L. Pekowsky, G. S. C. Davies, S. De, M. Cabero, S. Wu, A. R. Williamson, D. Macleod, B. Machenschalk, F. Pannarale, P. Kumar, S. Reyes, dfinstad, S. Kumar, M. Tápai, L. Singer, veronica villa, S. Khan, S. Fairhurst, B. U. V. Gadre, P. Kumar, K. Chandra, and A. Nielsen, gwastro/pycbc: v2.2.1 release of PyCBC (2023).

[105] A. Bohe, L. Shao, A. Taracchini, A. Buonanno, S. Babak, I. W. Harry, *et al.*, Improved effective-one-body model of spinning, nonprecessing binary black holes for the era of gravitational-wave astrophysics with advanced detectors, Phys. Rev. **D95**, 044028 (2017), arXiv:1611.03703 [gr-qc].

[106] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé, Frequency-domain gravitational waves from nonprecessing black-hole binaries. I. New numerical waveforms and anatomy of the signal, Phys. Rev. D **93**, 044006 (2016).

[107] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era, Phys. Rev. D **93**, 044007 (2016).

[108] A. Buonanno, B. R. Iyer, E. Ochsner, Y. Pan, and B. S. Sathyaprakash, Comparison of post-Newtonian templates for compact binary inspiral signals in gravitational-wave detectors, Phys. Rev. D **80**, 084043 (2009).

[109] F. Pannarale, E. Berti, K. Kyutoku, B. D. Lackey, and M. Shibata, Aligned spin neutron star-black hole mergers: A gravitational waveform amplitude model, Phys. Rev. D **92**, 084050 (2015).

[110] T. Dietrich, A. Samajdar, S. Khan, N. K. Johnson-McDaniel, R. Dudi, and W. Tichy, Improving the NRTidal model for binary neutron star systems, Phys. Rev. D **100**, 044003 (2019).

[111] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems (2015), software available from tensorflow.org.

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

[113] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), The basic physics of the binary black hole merger GW150914, Annalen der Physik **529**, 1600209 (2017).

[114] C. Pankow *et al.*, Mitigation of the instrumental noise transient in gravitational-wave data surrounding GW170817, Phys. Rev. D **98**, 084016 (2018), arXiv:1808.03619 [gr-qc].

[115] J. Glanzer, S. Banagiri, S. B. Coughlin, S. Soni, M. Zevin, C. P. L. Berry, O. Patane, S. Bahaadini, N. Rohani, K. Crowston, V. Kalogera, C. Østerlund, L. Trouille, and A. Katsaggelos, Data quality up to the third observing run of advanced LIGO: Gravity Spy glitch classifications, Classical and Quantum Gravity **40**, 065004 (2023).

[116] M. Zevin, S. Coughlin, S. Bahaadini, E. Besler, N. Rohani, S. Allen, M. Cabero, K. Crowston, A. K. Katsaggelos, S. L. Larson, T. K. Lee, C. Lintott, T. B. Littenberg, A. Lundgren, C. Østerlund, J. R. Smith, L. Trouille, and V. Kalogera, Gravity Spy: integrating advanced LIGO detector characterization, machine

learning, and citizen science, Classical and Quantum Gravity **34**, 064003 (2017).
[117] S. Bahaadini, V. Noroozi, N. Rohani, S. Coughlin, M. Zevin, J. Smith, V. Kalogera, and A. Katsaggelos, Machine learning for Gravity Spy: Glitch classification and dataset, Information Sciences **444**, 172 (2018).
[118] S. Coughlin, S. Bahaadini, N. Rohani, M. Zevin, O. Patane, M. Harandi, C. Jackson, V. Noroozi, S. Allen, J. Areeda, M. Coughlin, P. Ruiz, C. P. L. Berry, K. Crowston, A. K. Katsaggelos, A. Lundgren, C. Østerlund, J. R. Smith, L. Trouille, and V. Kalogera, Classifying the unknown: Discovering novel gravitational-wave detector glitches using similarity learning, Phys. Rev. D **99**, 082002 (2019).
[119] S. Soni, C. P. L. Berry, S. B. Coughlin, M. Harandi, C. B. Jackson, K. Crowston, C. Østerlund, O. Patane, A. K. Katsaggelos, L. Trouille, V.-G. Baranowski, W. F. Domainko, K. Kaminski, M. A. L. Rodriguez, U. Marciniak, P. Nauta, G. Niklasch, R. R. Rote, B. Téglás, C. Unsworth, and C. Zhang, Discovering features in gravitational-wave data through detector characterization, citizen science and machine learning, Classical and Quantum Gravity **38**, 195016 (2021).
[120] F. Robinet, N. Arnaud, N. Leroy, A. Lundgren, D. Macleod, and J. McIver, Omicron: A tool to characterize transient noise in gravitational-wave detectors, SoftwareX **12**, 100620 (2020).
[121] S. Joshua, aLIGO LLO Logbook 44803, https://alog.ligo-la.caltech.edu/aLOG/index.php?callRep=44803 (2019).
[122] C. Austin, Daytime Scatter at LLO, https://dcc.ligo.org/LIGO-G1900753/public (2019), Tech. Rep. G1900753 LSC.
[123] D. Vartanyan, A. Burrows, T. Wang, M. S. B. Coleman, and C. J. White, Gravitational-wave signature of core-collapse supernovae, Phys. Rev. D **107**, 103015 (2023).
[124] A. W. Steiner, M. Hempel, and T. Fischer, Core-collapse supernova equations of state based on neutron star observations, The Astrophysical Journal **774**, 17 (2013).
[125] Gravitational waves from core-collapse supernovae.

Appendix A: Denoised Results of other Events

In this appendix, we present the rest of the denoised results of the confidently detected events on the Gravitational Wave Transient Catalog (GWTC) reported by the LVK Collaboration during O1, O2, and O3 observation runs, in addition to the events GW150914, GW151012, and GW170817 described in Section III B. Although the input data are the 2-second segments, for each event, we focus on the 0.25s data segment around the peak of the reconstructed waveform, which is used for calculating the overlap. For O1 and O2 events, we use the name listed on GWTC. Since starting from O3, GW candidates have been uploaded to the Gravitational-wave candidate event database (GraceDB). In addition to the GWTC name, we attached the GraceDB name as a reference for O3 events.

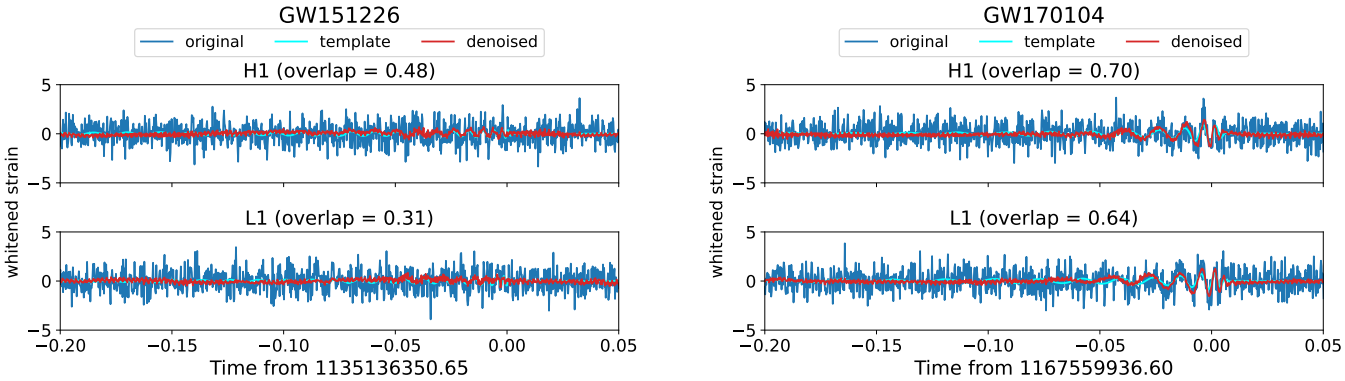
The denoised results of the event catalog are presented from Figures A1 to A43. Among all confidently detected events, there are two BNS events: GW170817 and GW190425 (S190425z), an NSBH event: GW191219_163120 (S191219ax), and the event GW200210_092254 (S200210ba) that can be either a BBH or an NSBH event [11–14]. Apart from GW170817 which have shown in Section III B, the denoised result of GW190425, GW191219_163120 and GW200210_092254 are shown in Figure A8a, A32b and A38a, respectively. Unfortunately, our model cannot denoise these GW signals due to their low amplitudes.

Other events in the catalog are BBH events. Our model can reveal the structure of most of the loud signals in the H1 and L1 strain data. Moreover, for events GW170608, GW190412 (S190412m), GW190708_232457 (S190708ap), GW190814 (S190814bv), GW191204_171526 (S191204r) and GW191216_213338 (S191216ap), our model can extract the inspiral signal more than 0.1s before the merger event, as shown in Figures A2a, A6a, A16b, A21a, A31a and A32a, respectively. Also, although our model cannot extract the feature in most of the V1 data due to the large noise background, it can still extract the feature for significant events GW200129p_065458 (S200129m) and GW200311_115853 (S200311bg) in Figures A35b and A43a. Furthermore, although the signal of GW190925_232845 (S190925ad) is not loud, our model can still extract some of the inspiral features in the V1 data, as shown in Figure A25a.

Since we only use vanilla matched-filtering to reconstruct waveform templates, sometimes the reconstructed waveform cannot accurately align with the strain data, which may be due to the poor estimation of the PSD. For example, for the event GW200225_060421 (S200225q) in Figure A41a, we cannot reconstruct the waveform correctly in the L1 data, but our model can extract the GW features around the event, which leads to the very low overlap value.

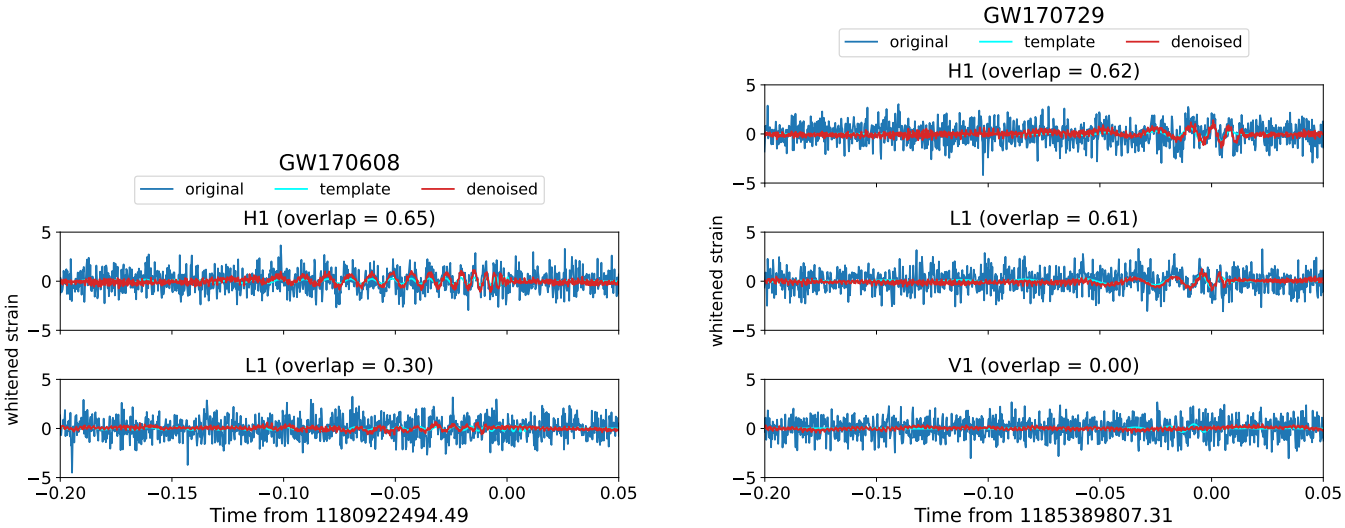
Finally, in the event GW200220_124850 (S200220aw), there are some noise fluctuations around the event in the H1 data, as shown in Figure A40a. Our model can ex-

tract GW features in the H1 data along with the fluctuation. Since our model does not intend to separate GW signals from glitches, it also extracts the feature of noise fluctuation, which may affect the result of the following parameter estimation or localization tasks.



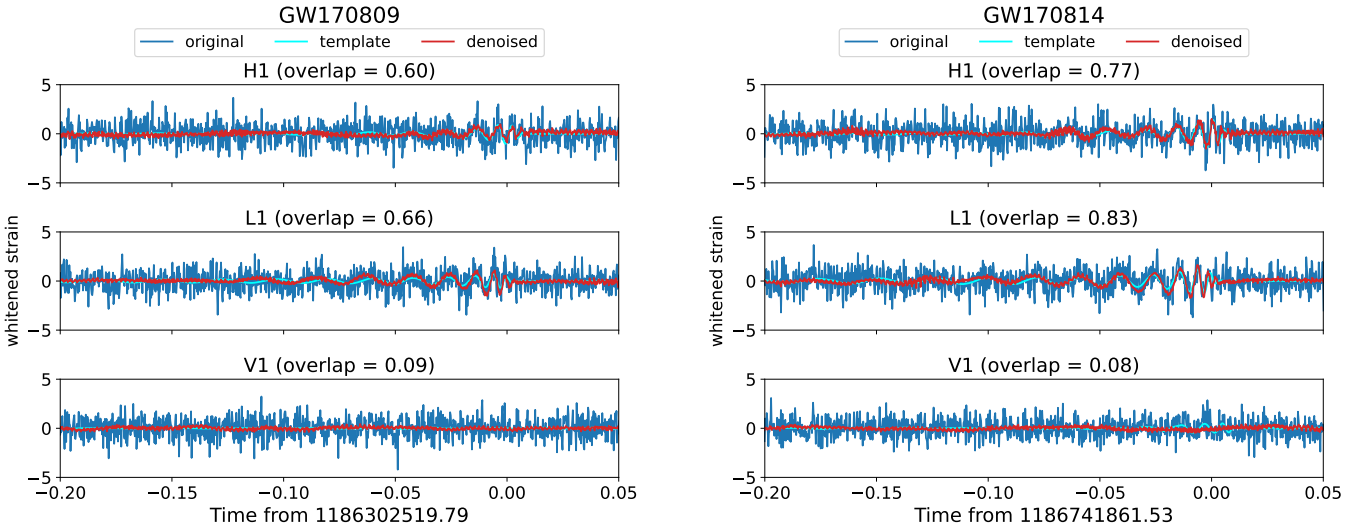
(a) (b)

FIG. A1. The 0.25s windows of the denoised results of the H1 and L1 data around GW151226 and GW170104.



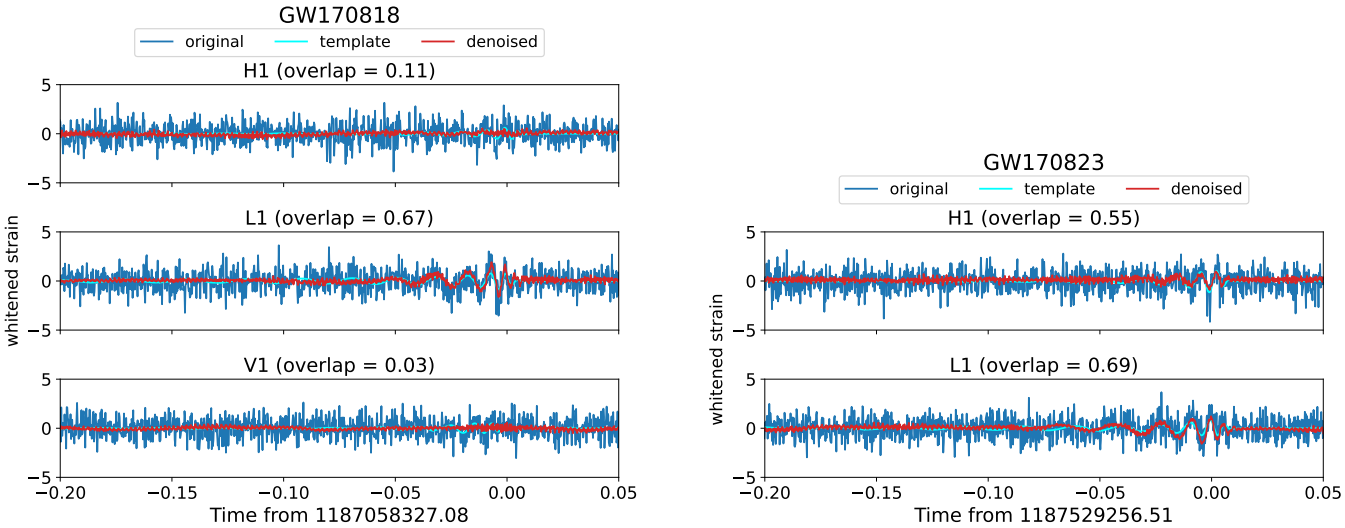
(a) (b)

FIG. A2. The 0.25s windows of the denoised results of the H1, L1 and V1 data around GW170608 and GW170729.



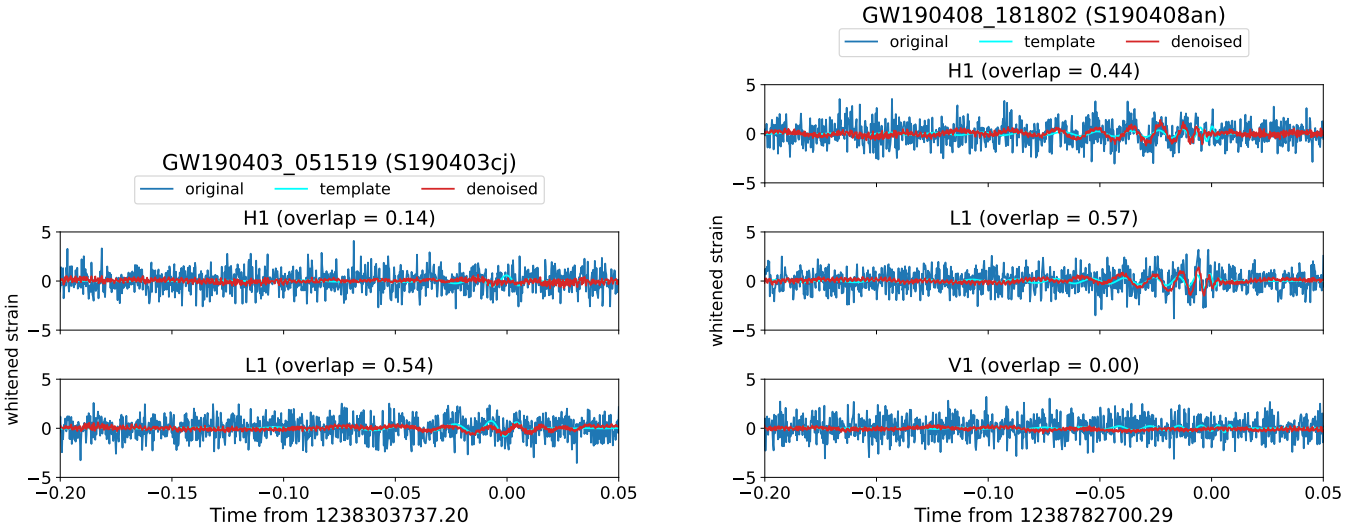
(a) (b)

FIG. A3. The 0.25s windows of the denoised results of the H1, L1 and V1 data around GW170809 and GW170814.



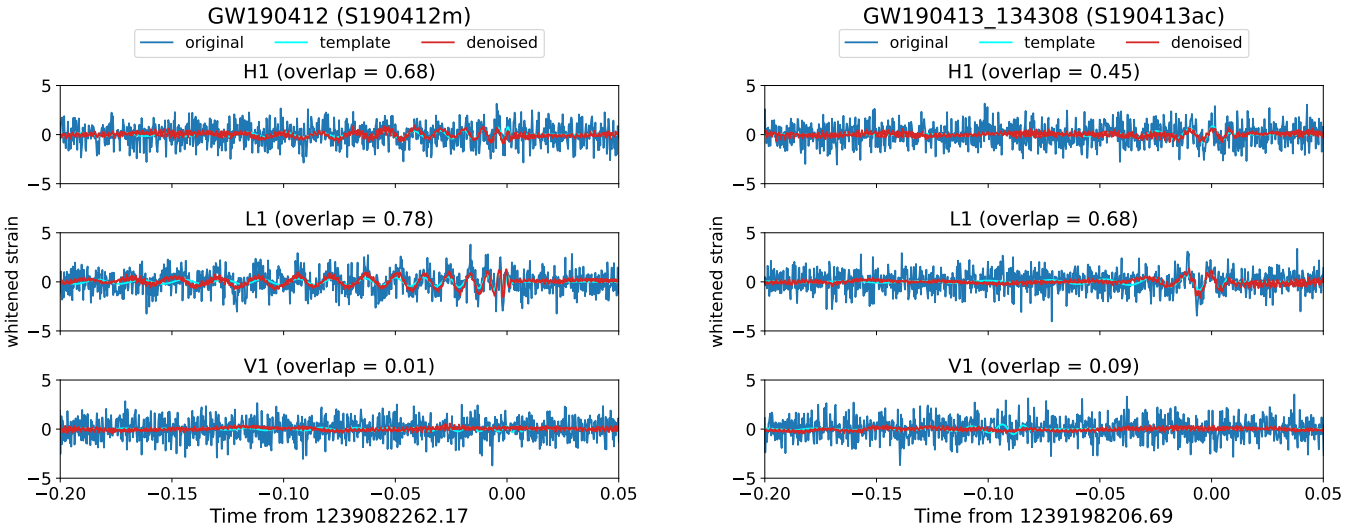
(a) (b)

FIG. A4. The 0.25s windows of the denoised results of the H1, L1 and V1 data around GW170818 and GW170823.



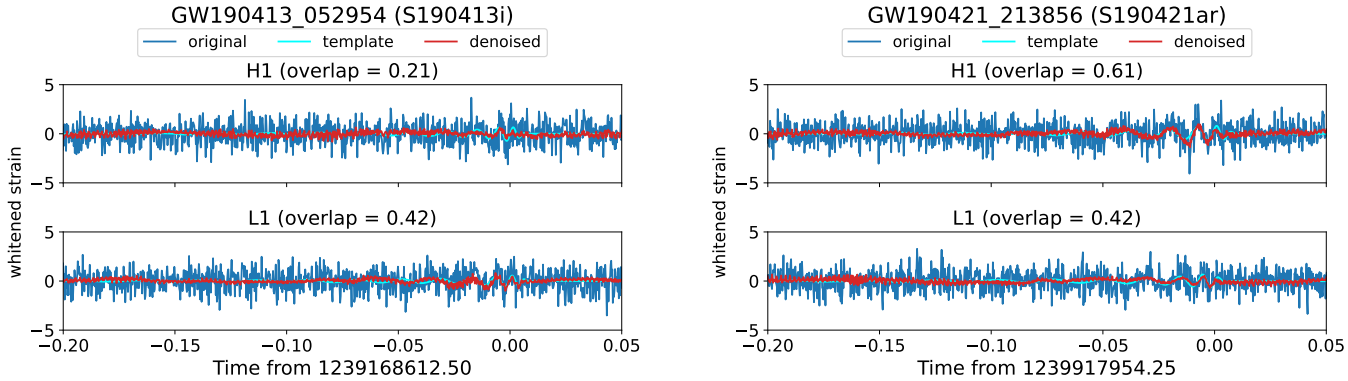
(a) (b)

FIG. A5. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S190403cj and S190408an.



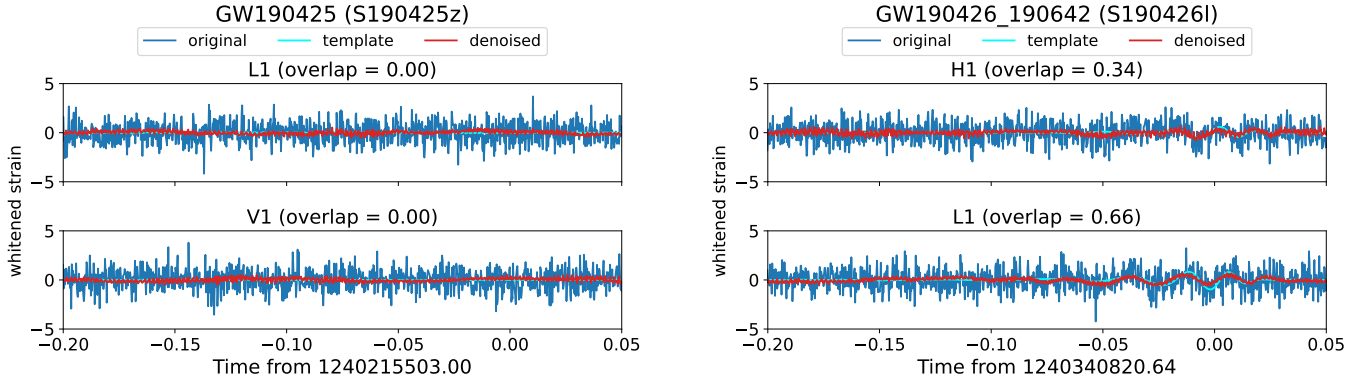
(a) (b)

FIG. A6. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S190412m and S190413ac.



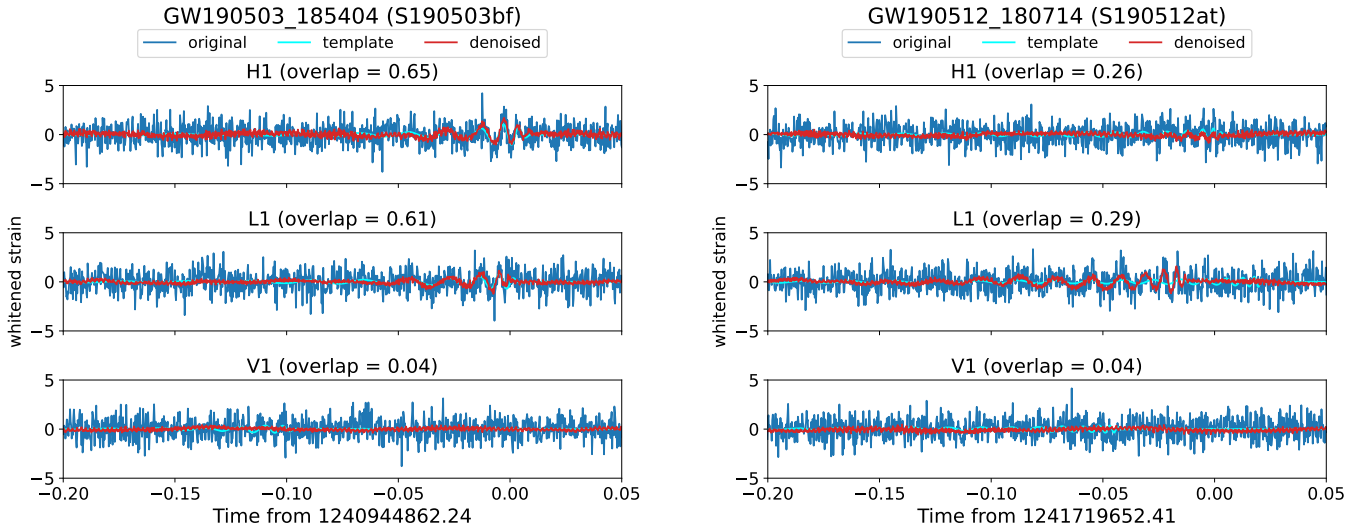
(a) (b)

FIG. A7. The 0.25s windows of the denoised results of the H1 and L1 data around S190413i and S190421ar.



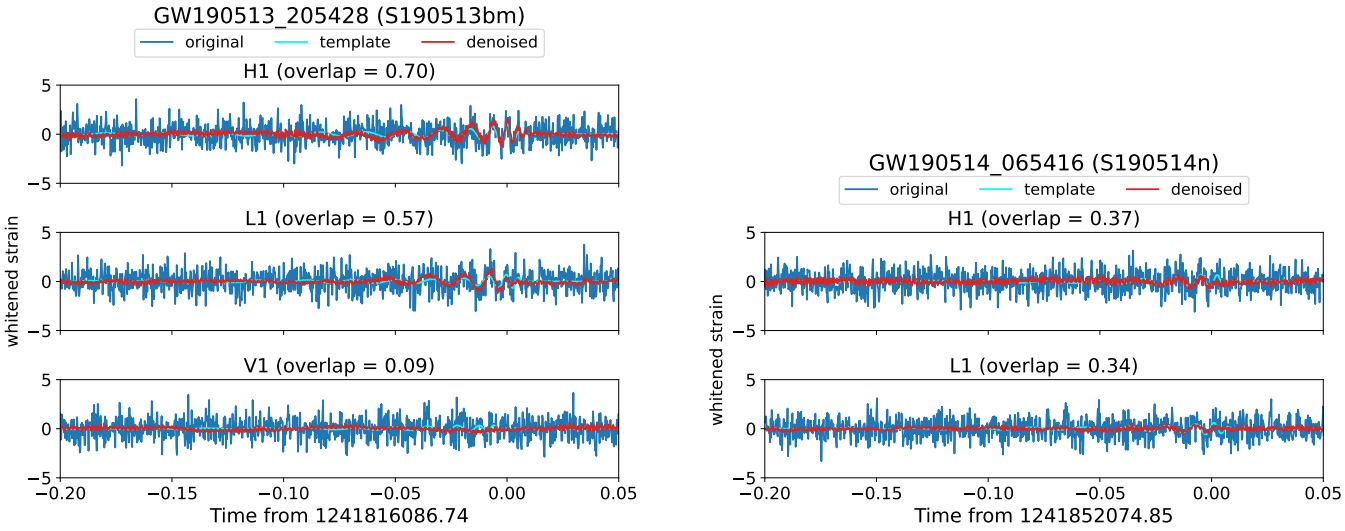
(a) (b)

FIG. A8. The 0.25s windows of the denoised results of the H1 and L1 data around S190425z and S190426l.



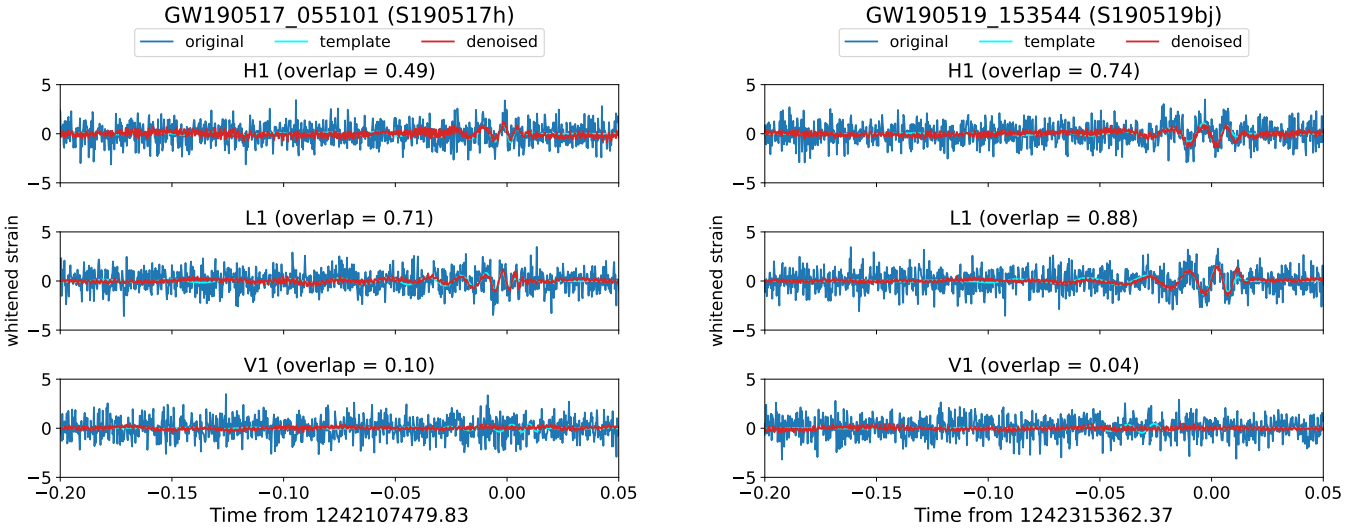
(a) (b)

FIG. A9. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S190503bf and S190512at.



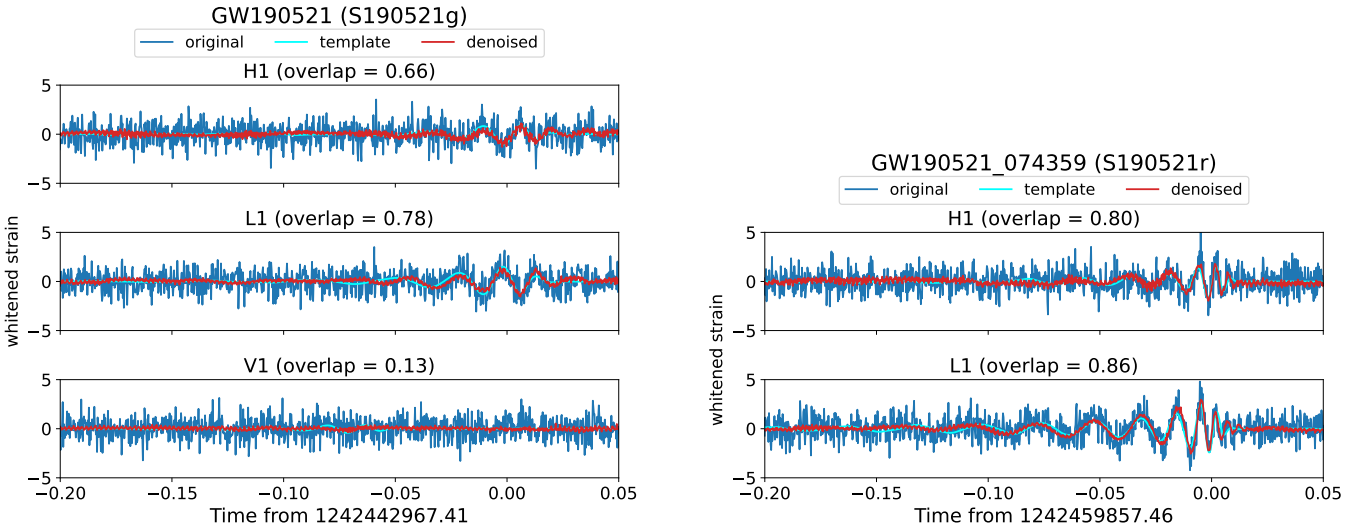
(a) (b)

FIG. A10. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S190513bm and S190514n.



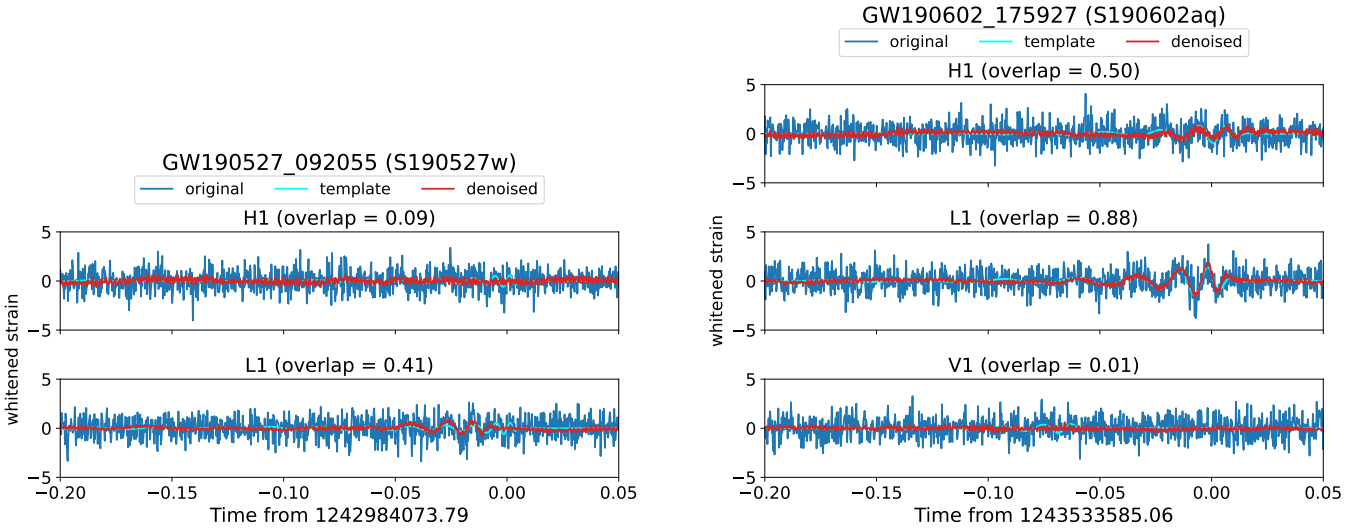
(a) (b)

FIG. A11. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S190517h and S190519bj.



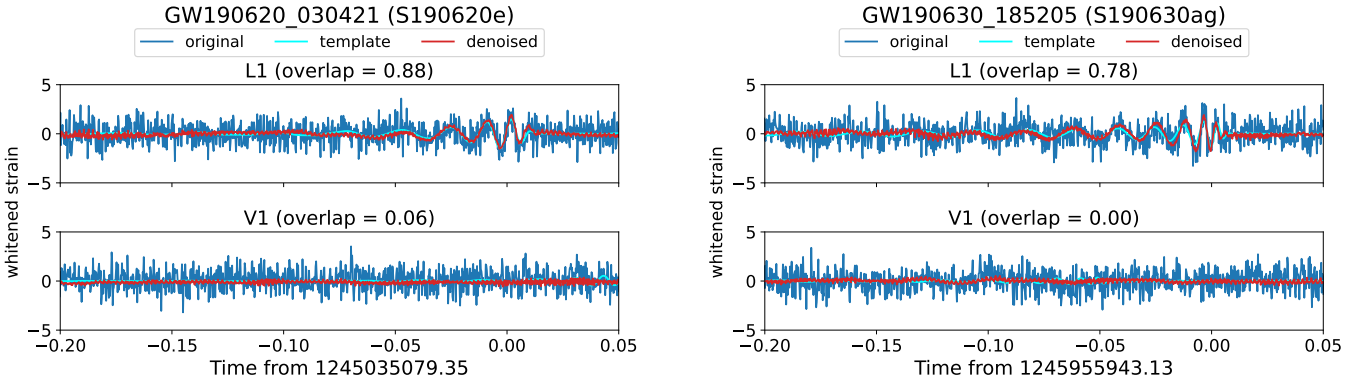
(a) (b)

FIG. A12. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S190521g and S190521r.



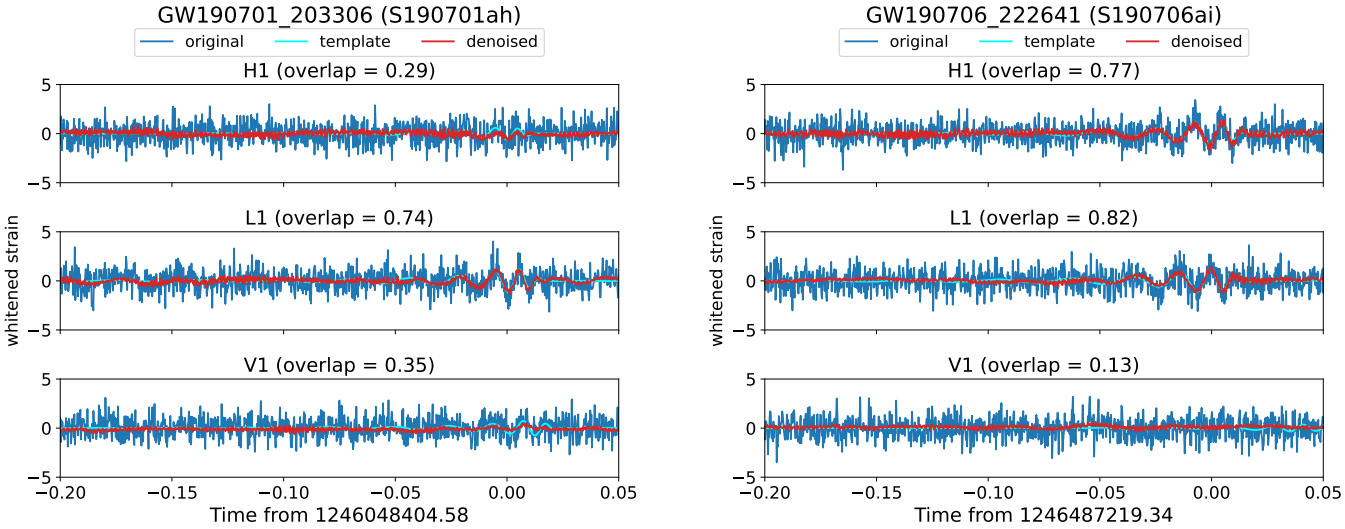
(a) (b)

FIG. A13. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S190527w and S190602aq.



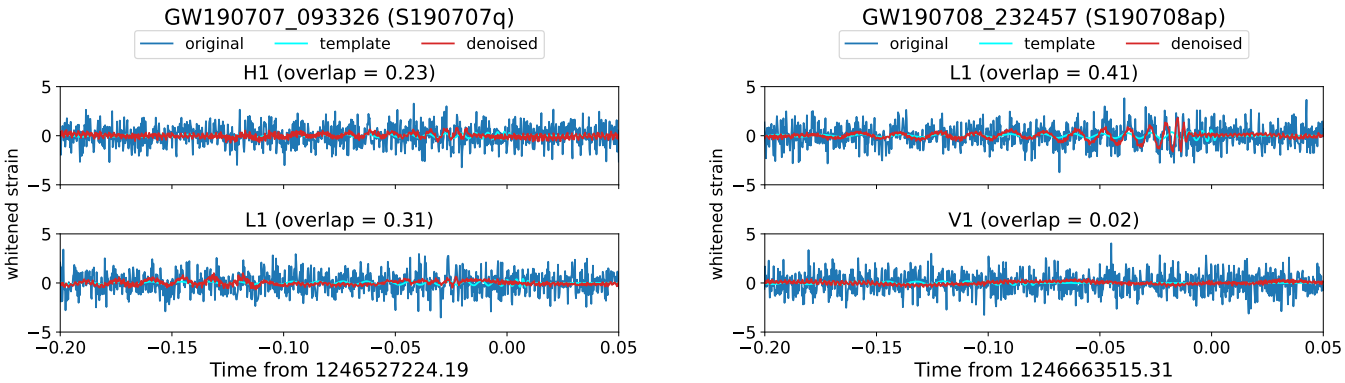
(a) (b)

FIG. A14. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S190620e and S190630ag.



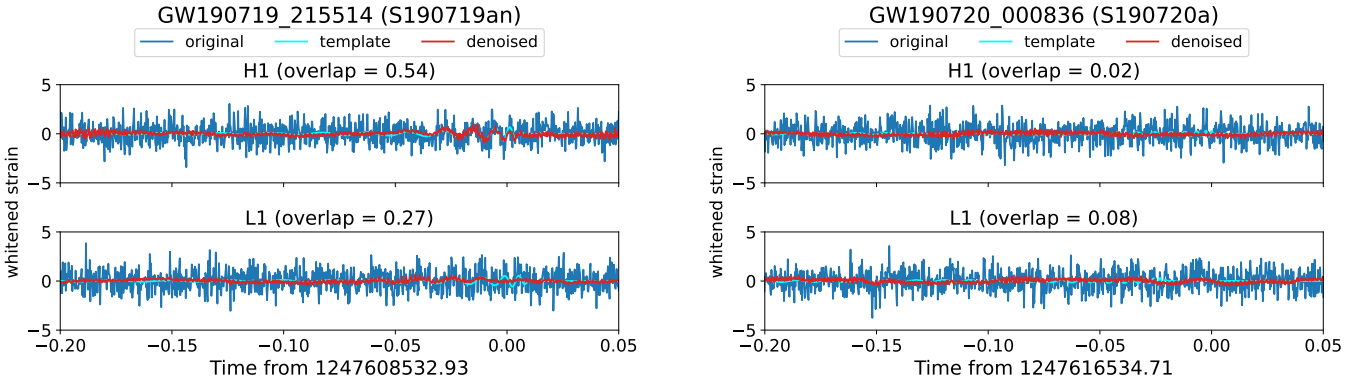
(a) (b)

FIG. A15. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S190701ah and S190706ai.



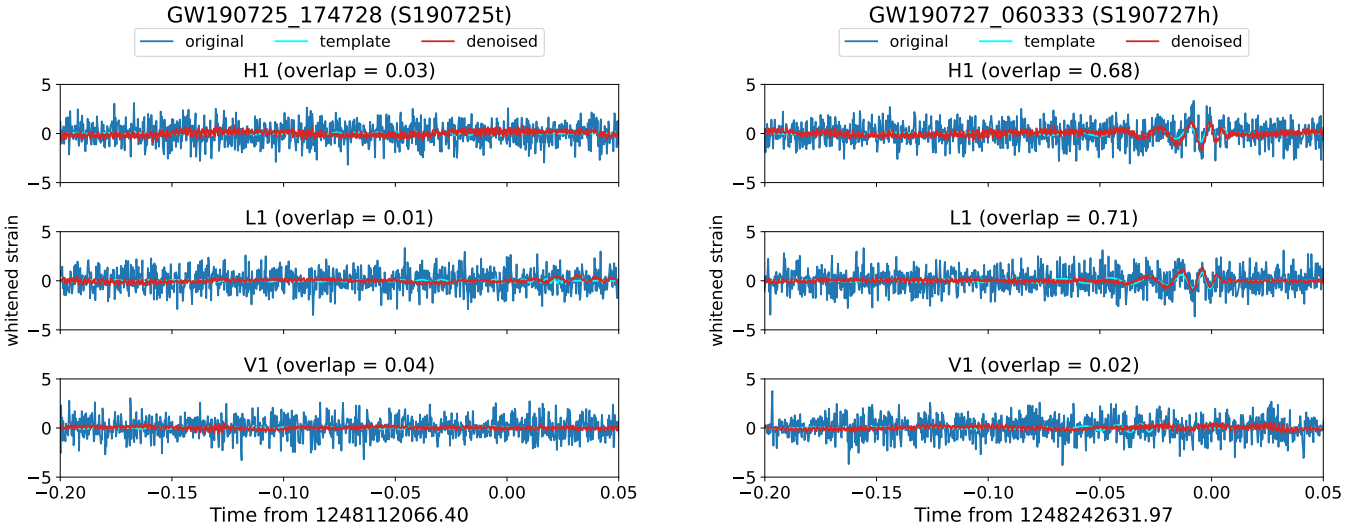
(a) (b)

FIG. A16. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S190707q and S190708ap.



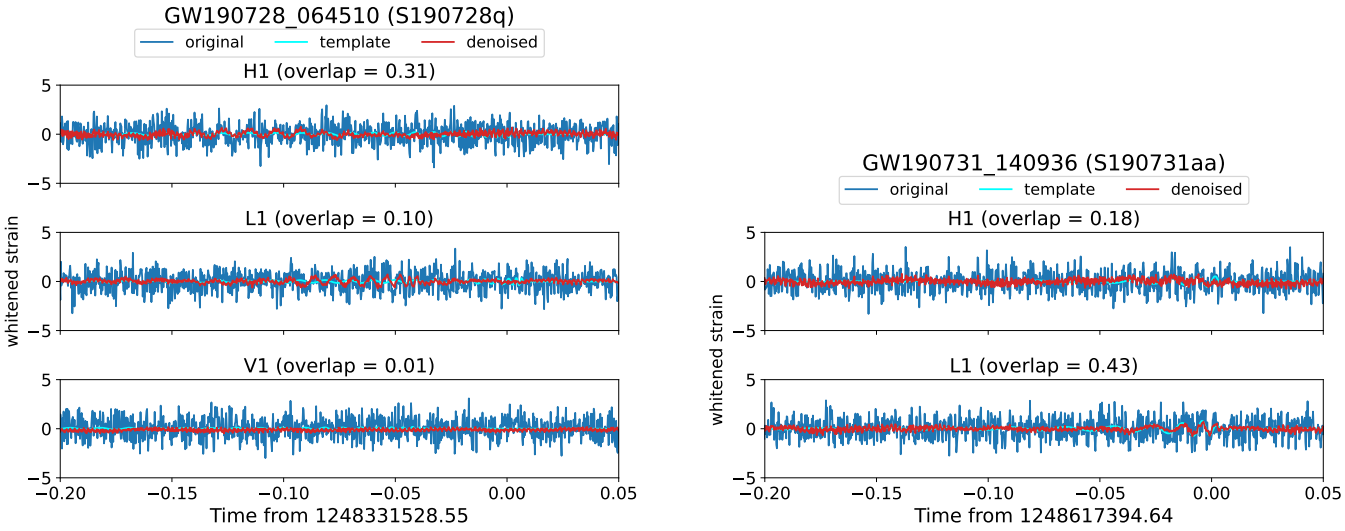
(a) (b)

FIG. A17. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S190719an and S190720a.



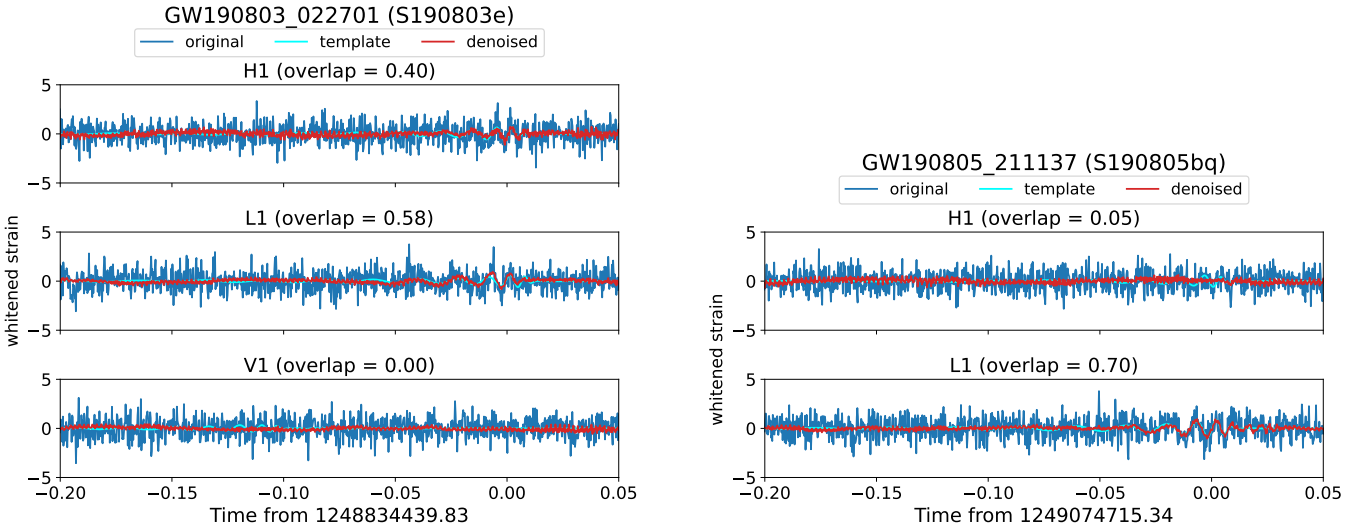
(a) (b)

FIG. A18. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S190725t and S190727h.



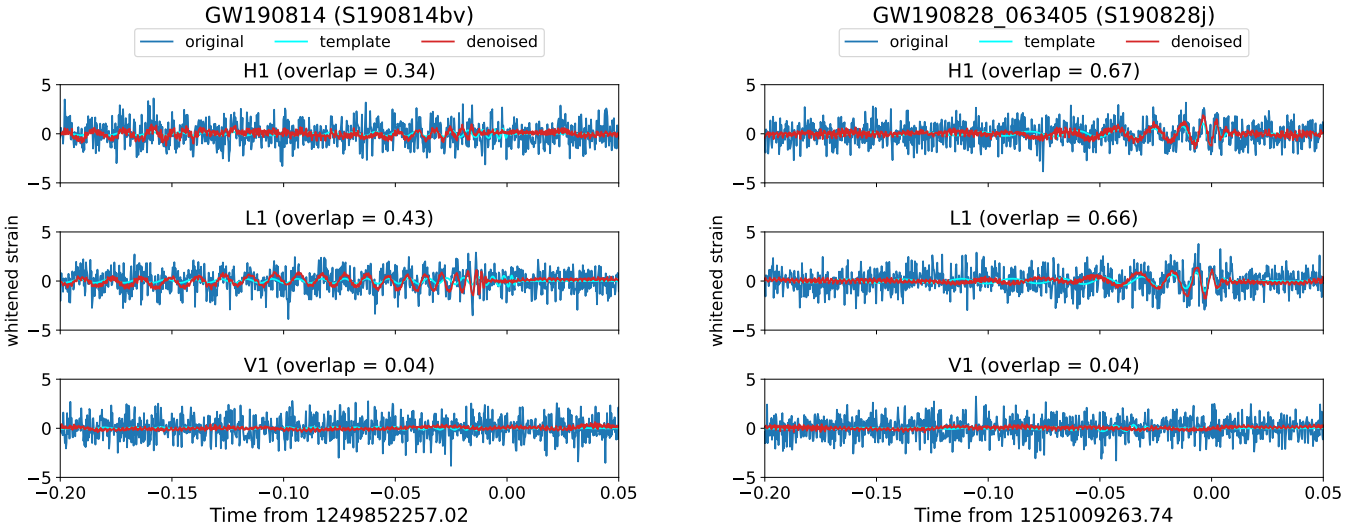
(a) (b)

FIG. A19. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S190728q and S190731aa.



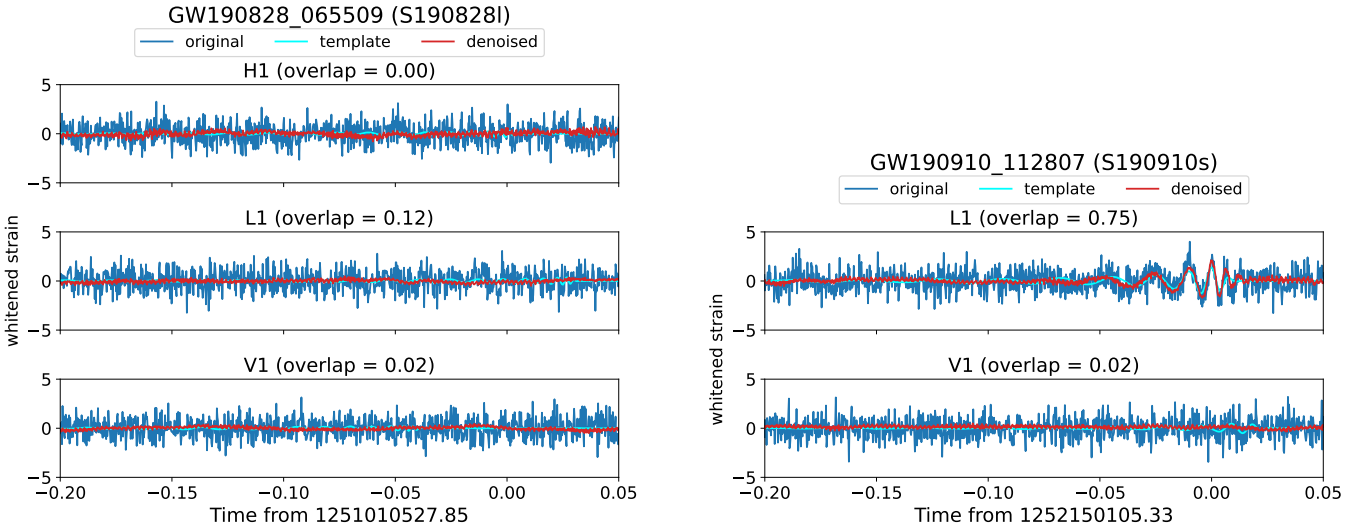
(a) (b)

FIG. A20. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S190803e and S190805bq.



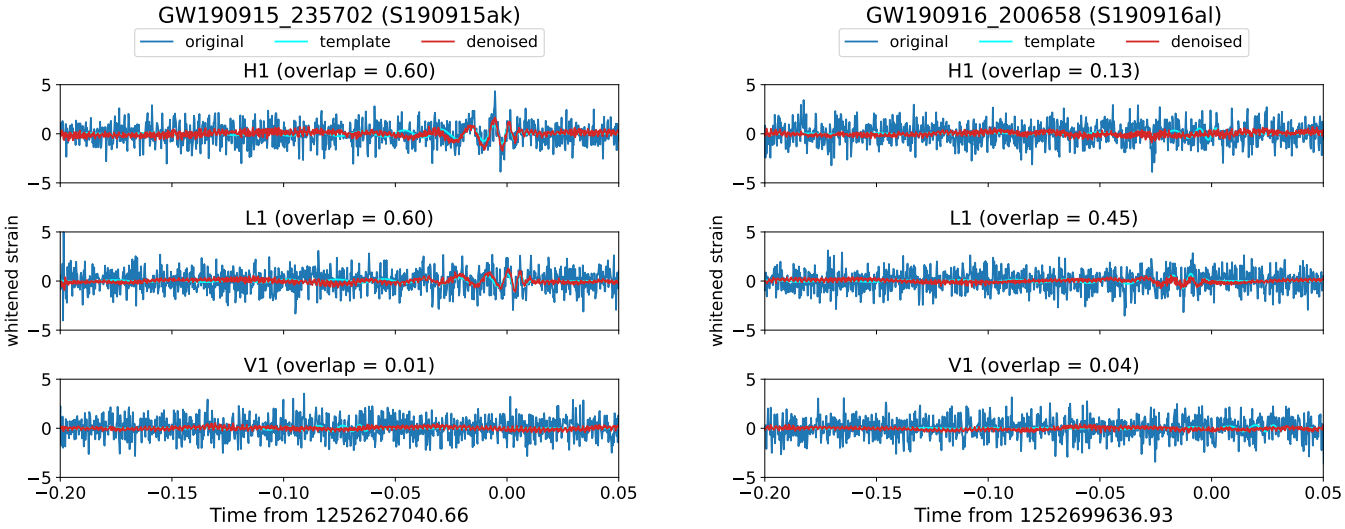
(a) (b)

FIG. A21. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S190814bv and S190828j.



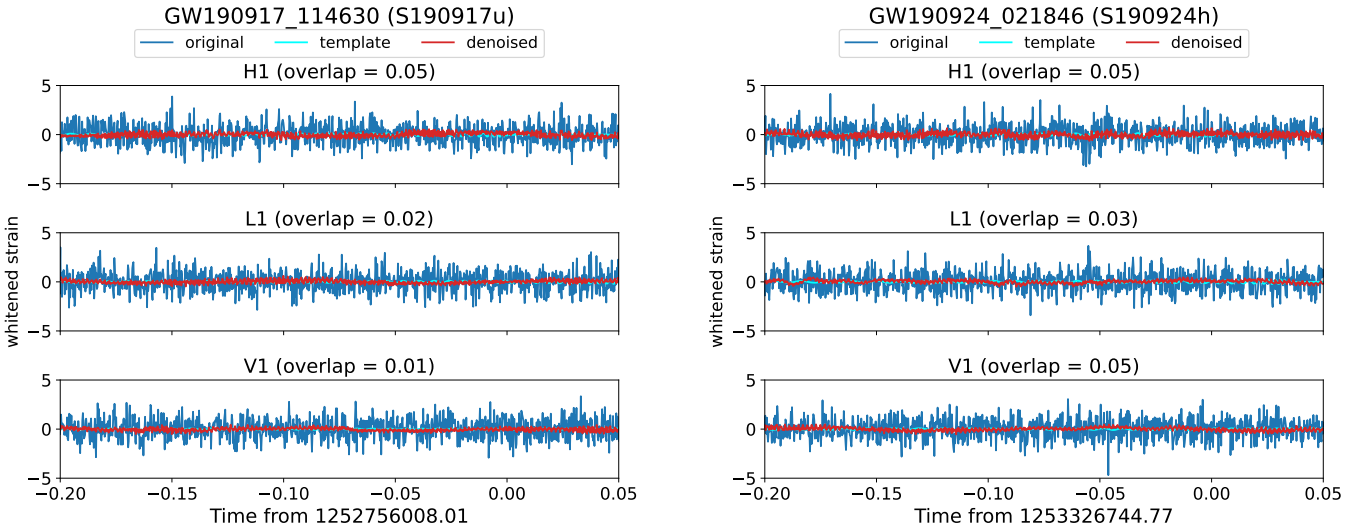
(a) (b)

FIG. A22. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S190828l and S190910s.



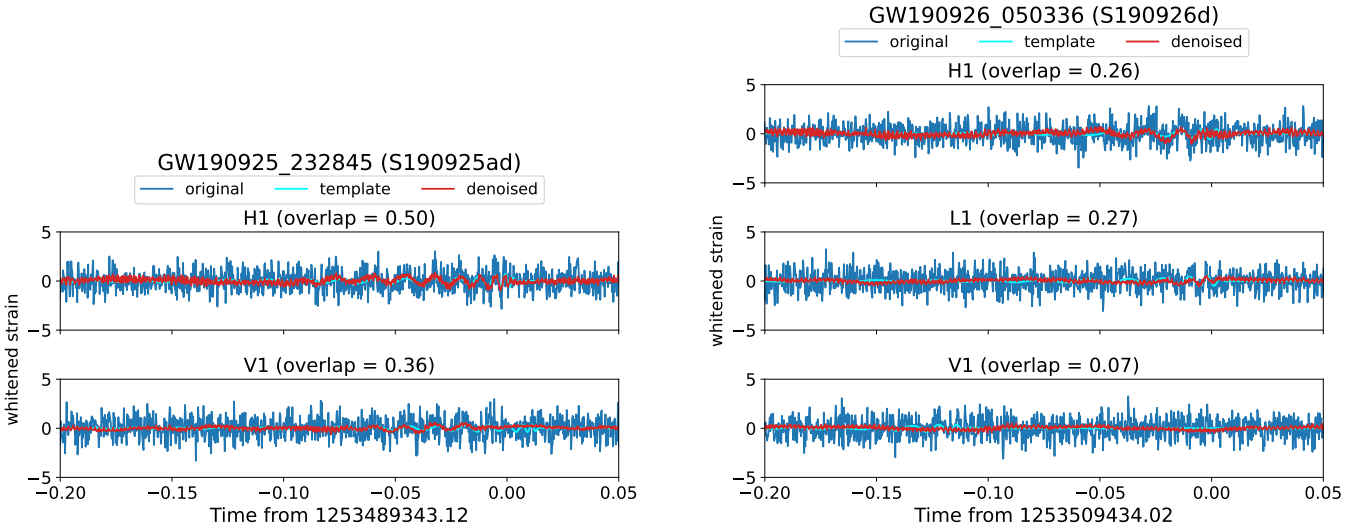
(a) (b)

FIG. A23. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S190915ak and S190916al.



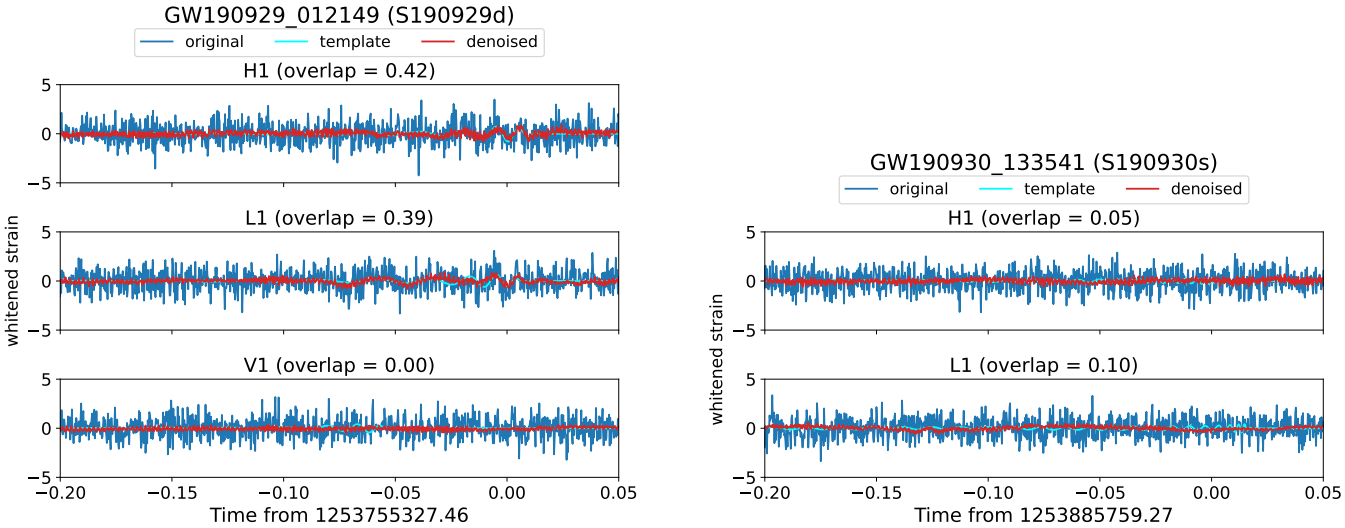
(a) (b)

FIG. A24. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S190917u and S190924h.



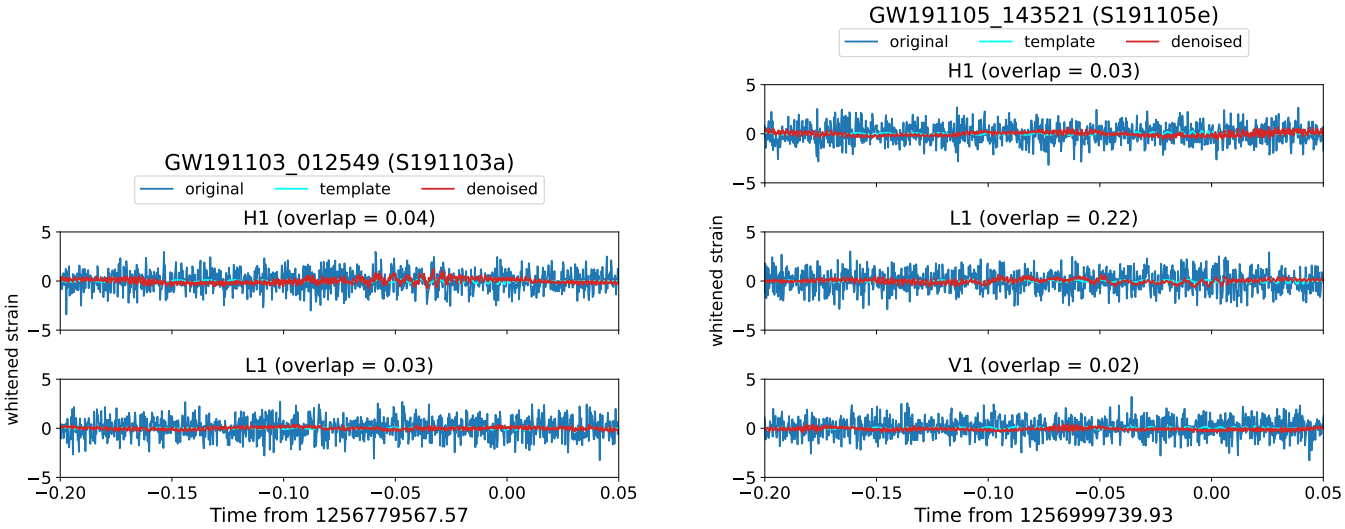
(a) (b)

FIG. A25. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S190925ad and S190926d.



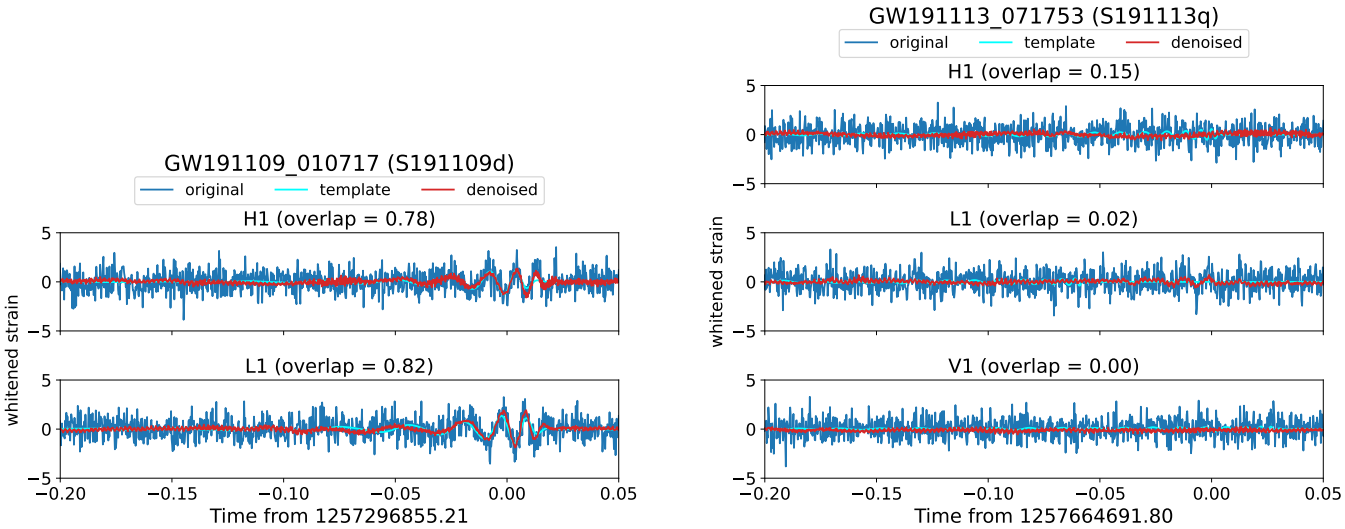
(a) (b)

FIG. A26. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S190929d and S190930s.



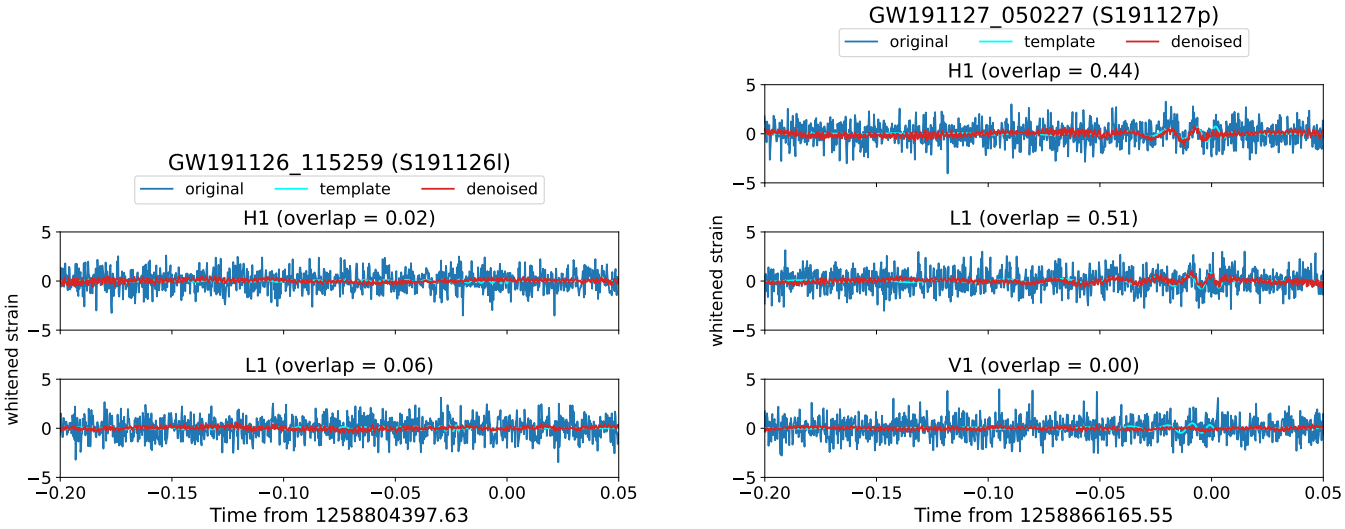
(a) (b)

FIG. A27. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S191103a and S191105e.



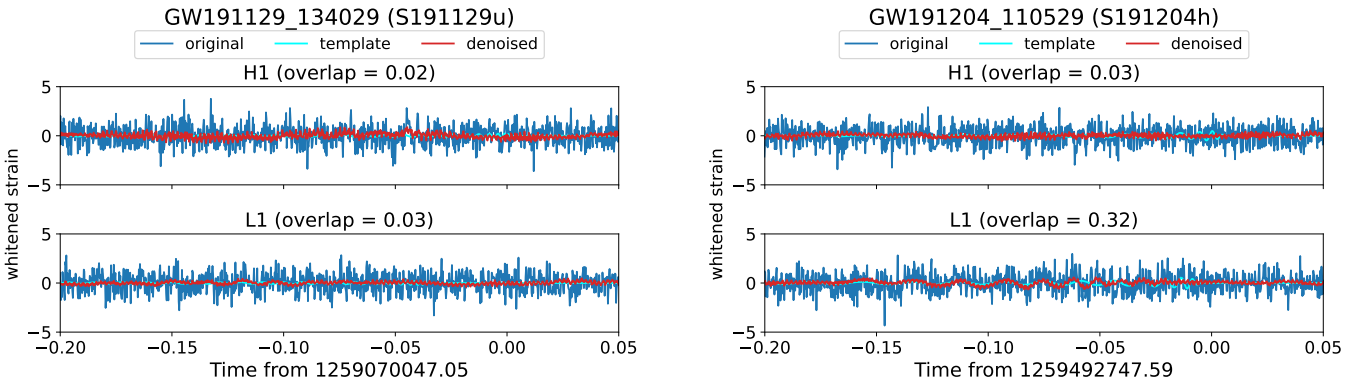
(a) (b)

FIG. A28. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S191109d and S191113q.



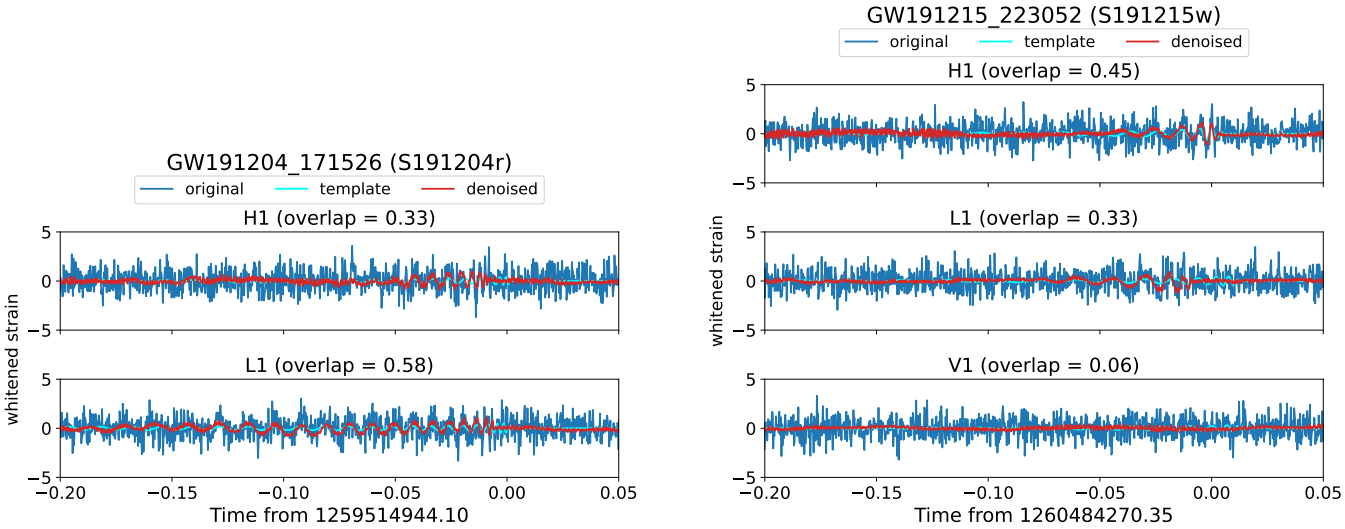
(a) (b)

FIG. A29. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S191126l and S191127p.



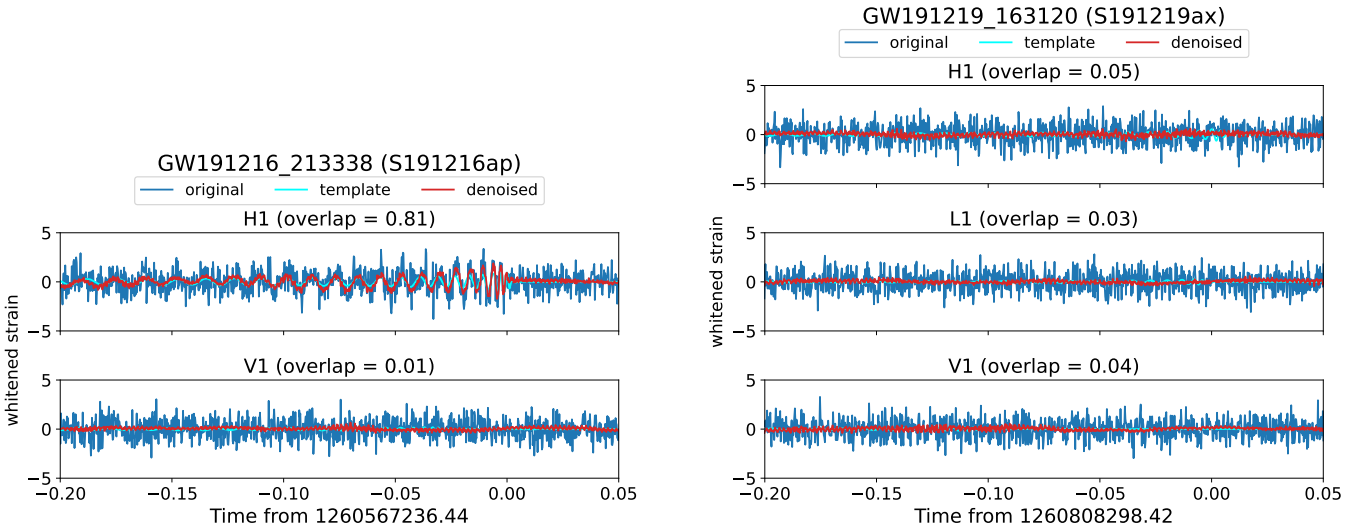
(a) (b)

FIG. A30. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S191129u and S191204h.



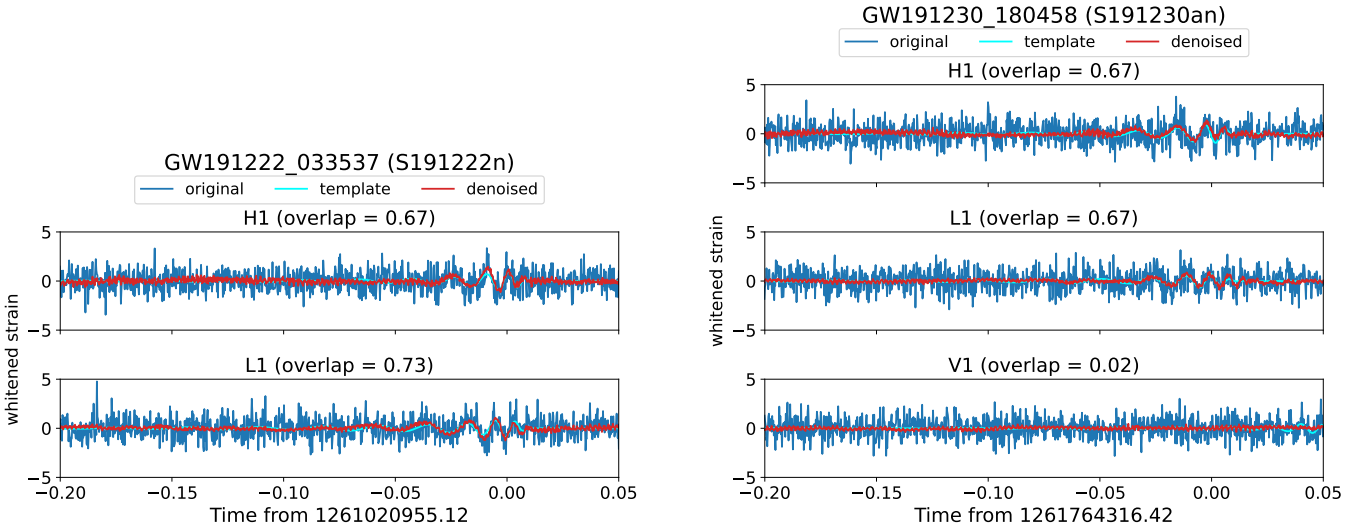
(a) (b)

FIG. A31. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S191204r and S191215w.



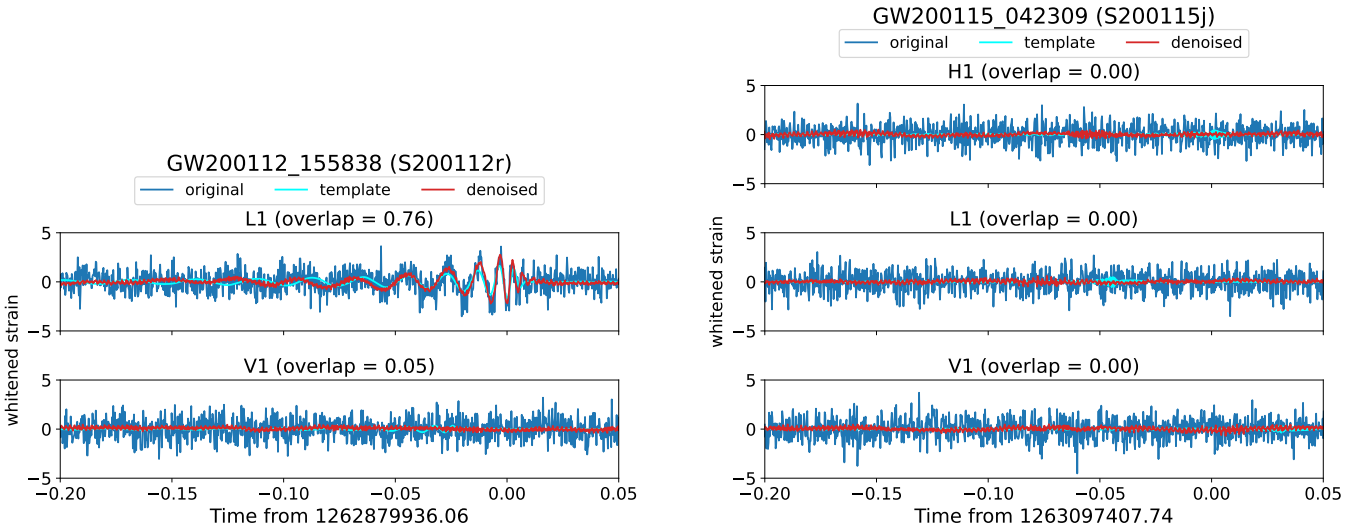
(a) (b)

FIG. A32. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S191216ap and S191219ax.



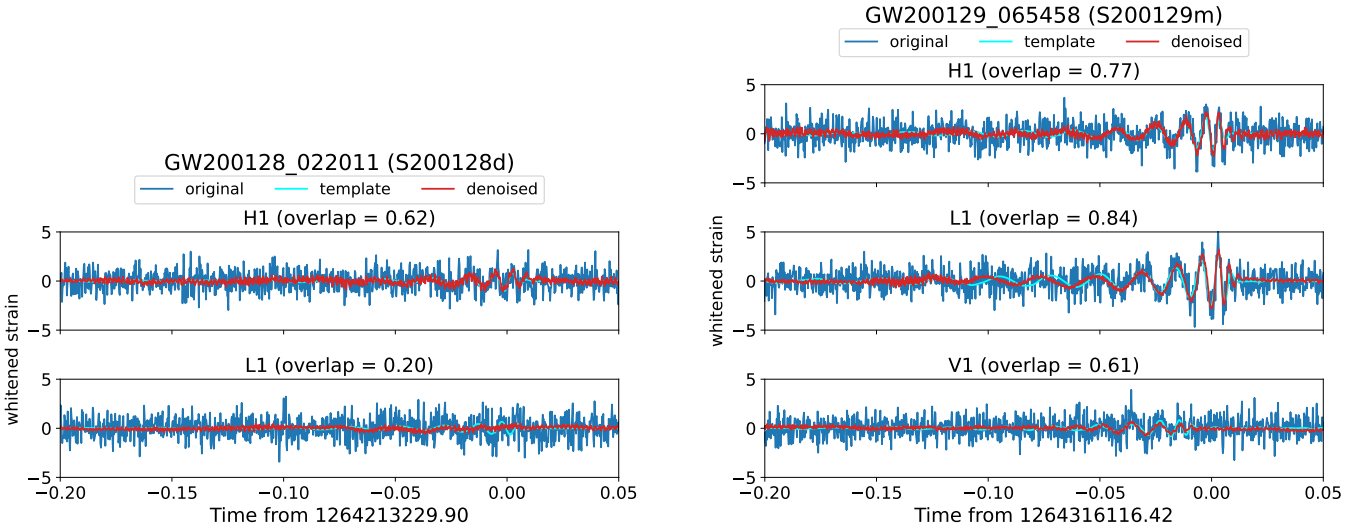
(a) (b)

FIG. A33. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S191222n and S191230an.



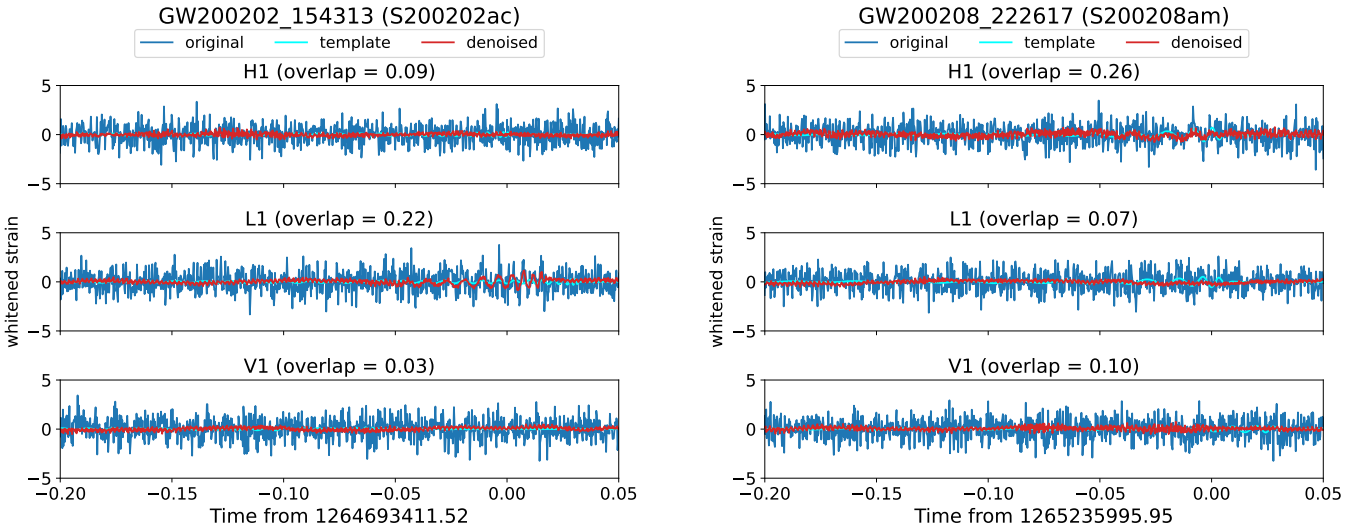
(a) (b)

FIG. A34. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S200112r and S200115j.



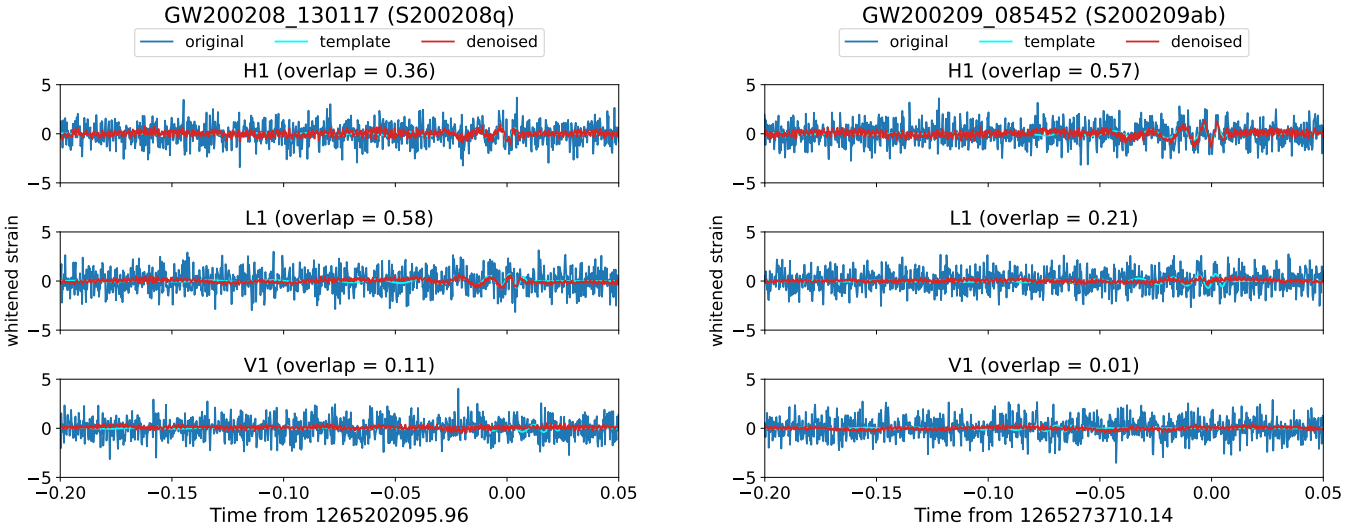
(a) (b)

FIG. A35. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S200128d and S200129m.



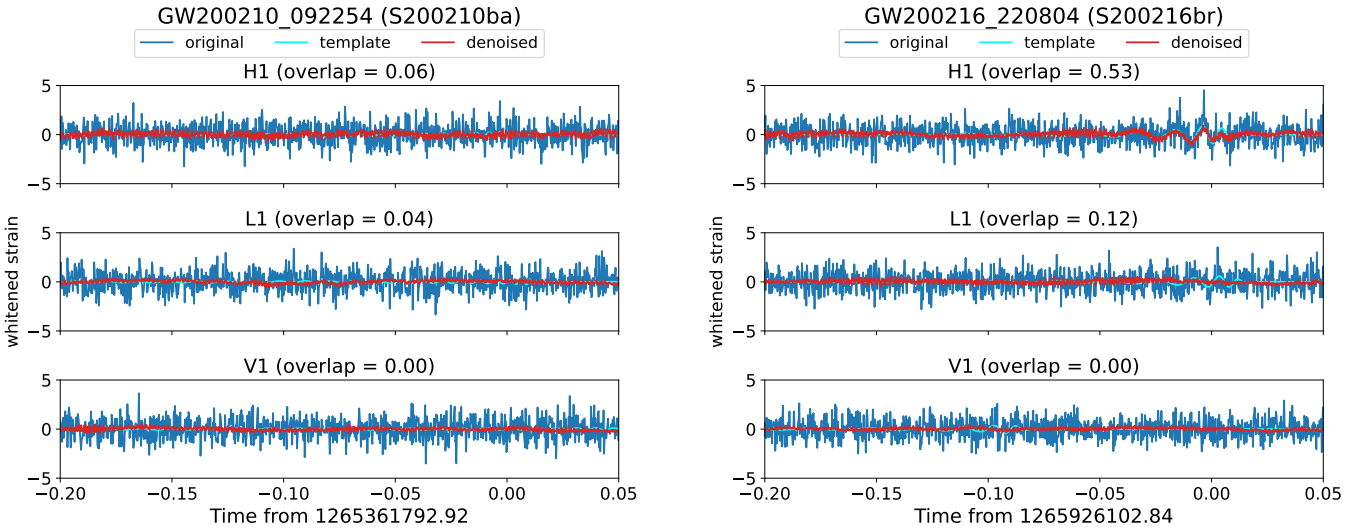
(a) (b)

FIG. A36. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S200202ac and S200208am.



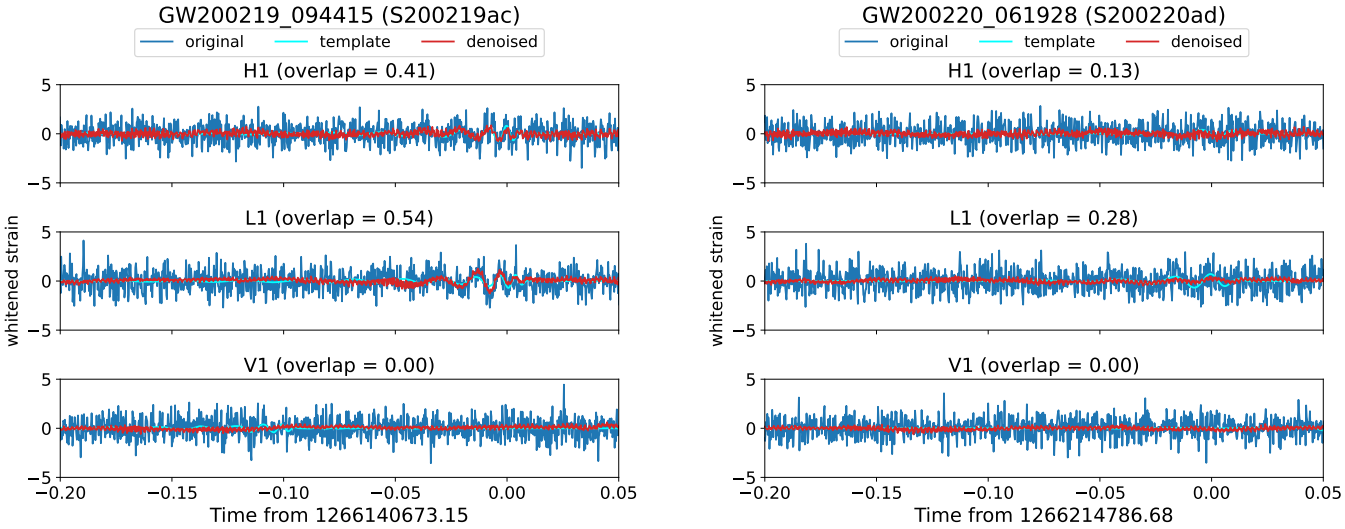
(a) (b)

FIG. A37. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S200208q and S200209ab.



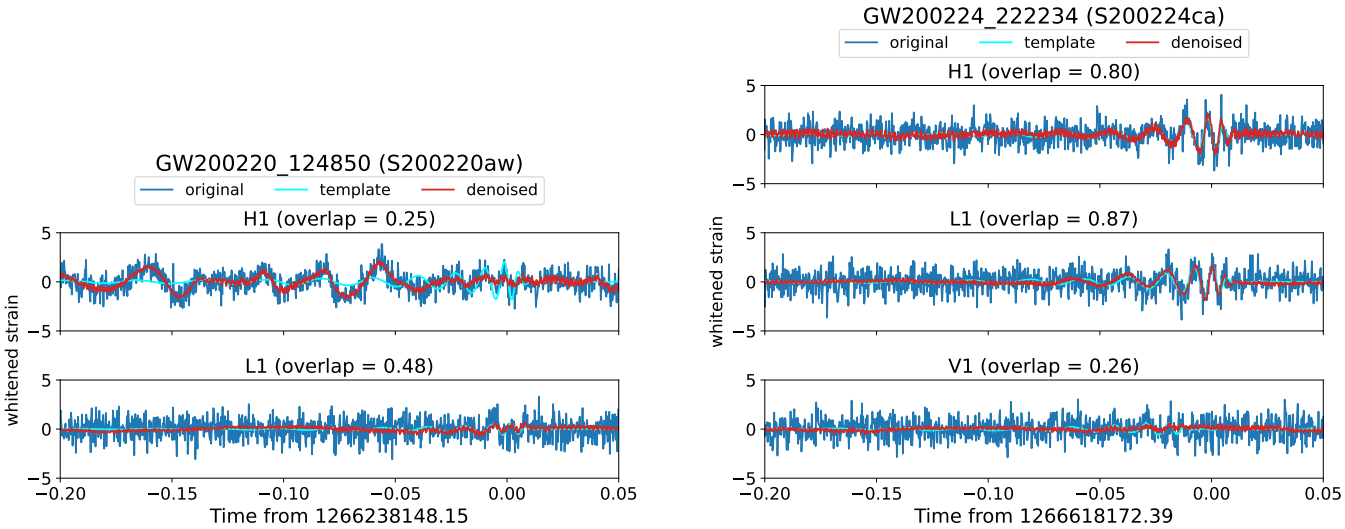
(a) (b)

FIG. A38. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S200210ba and S200216br.



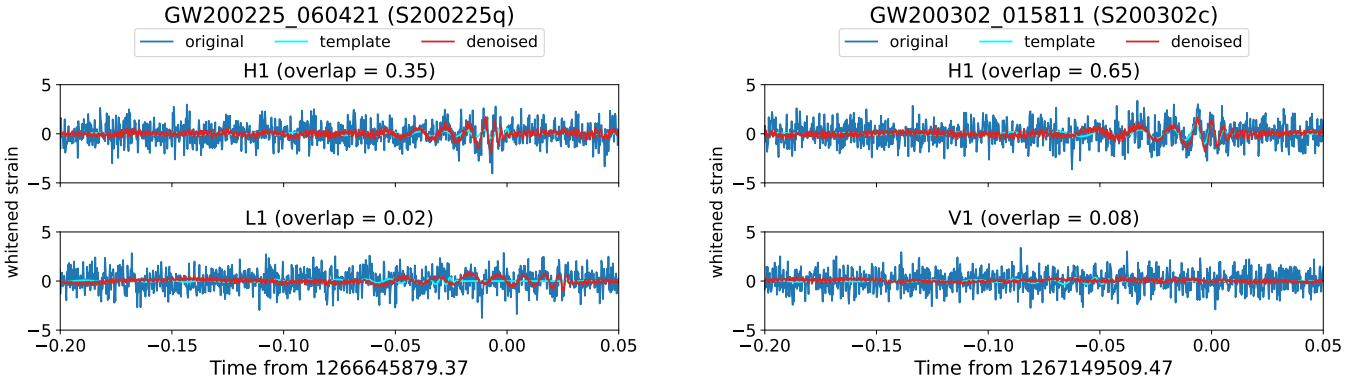
(a) (b)

FIG. A39. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S200219ac and S200220ad.



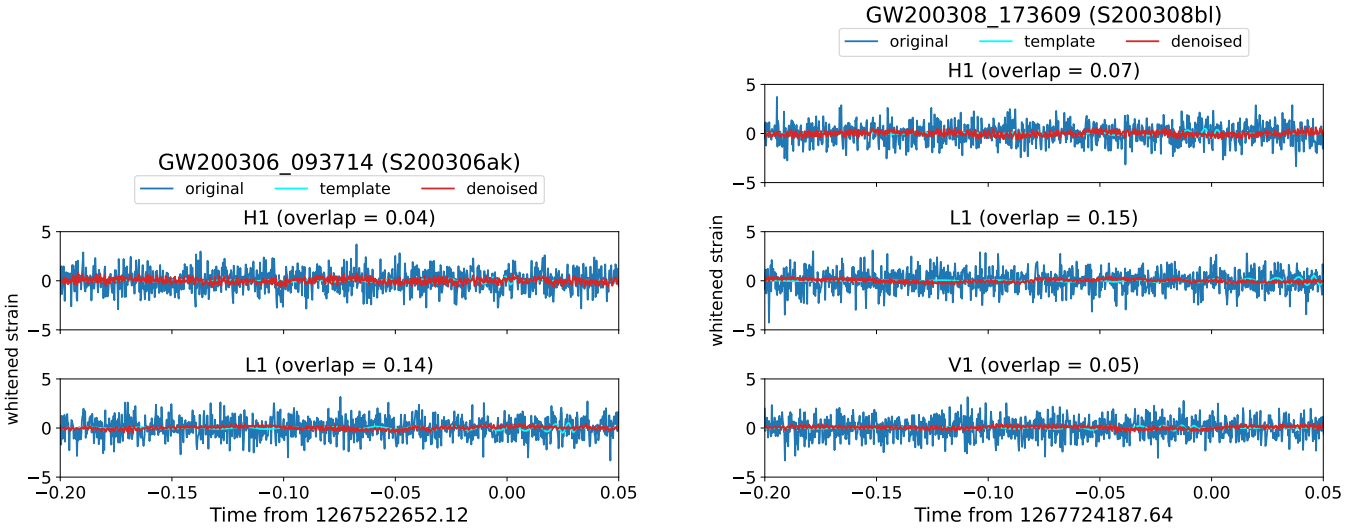
(a) (b)

FIG. A40. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S200220aw and S200224ca.



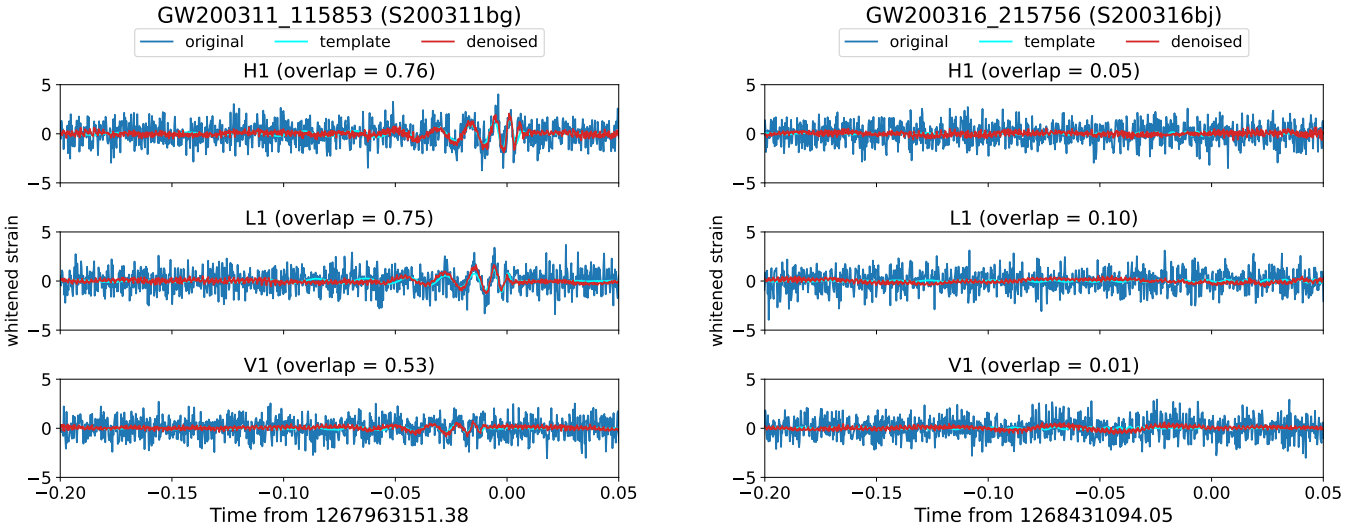
(a) (b)

FIG. A41. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S200115j and S200128d.



(a) (b)

FIG. A42. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S200306ak and S200308bl.



(a) (b)

FIG. A43. The 0.25s windows of the denoised results of the H1, L1 and V1 data around S200311bg and S200316bj.