



# Nonlinear Noise Cleaning in Gravitational-Wave Detectors With Convolutional Neural Networks

Hang Yu<sup>1\*</sup> and Rana X. Adhikari<sup>2</sup>

<sup>1</sup> TAPIR, Walter Burke Institute for Theoretical Physics, MC 350-17, California Institute of Technology, Pasadena, CA, United States, <sup>2</sup> LIGO Laboratory, MC 100-36, California Institute of Technology, Pasadena, CA, United States

## OPEN ACCESS

### Edited by:

Mia Liu,  
Purdue University, United States

### Reviewed by:

Gabriel Nathan Perdue,  
Fermilab Accelerator Complex,  
United States

Jonathan Andrew Miller,  
Onto Innovation, United States

Dylan Rankin,  
Massachusetts Institute of  
Technology, United States

### \*Correspondence:

Hang Yu  
hangyu@caltech.edu

### Specialty section:

This article was submitted to  
Big Data and AI in High Energy  
Physics,  
a section of the journal  
Frontiers in Artificial Intelligence

**Received:** 09 November 2021

**Accepted:** 08 February 2022

**Published:** 17 March 2022

### Citation:

Yu H and Adhikari RX (2022) Nonlinear  
Noise Cleaning in Gravitational-Wave  
Detectors With Convolutional Neural  
Networks.  
*Front. Artif. Intell.* 5:811563.  
doi: 10.3389/frai.2022.811563

Currently, the sub-60 Hz sensitivity of gravitational-wave (GW) detectors like Advanced LIGO (aLIGO) is limited by the control noises from auxiliary degrees of freedom which nonlinearly couple to the main GW readout. One promising way to tackle this challenge is to perform nonlinear noise mitigation using convolutional neural networks (CNNs), which we examine in detail in this study. In many cases, the noise coupling is bilinear and can be viewed as a few fast channels' outputs modulated by some slow channels. We show that we can utilize this knowledge of the physical system and adopt an explicit "slow × fast" structure in the design of the CNN to enhance its performance of noise subtraction. We then examine the requirements in the signal-to-noise ratio (SNR) in both the target channel (i.e., the main GW readout) and in the auxiliary sensors in order to reduce the noise by at least a factor of a few. In the case of limited SNR in the target channel, we further demonstrate that the CNN can still reach a good performance if we use curriculum learning techniques, which in reality can be achieved by combining data from quiet times and those from periods with active noise injections.

**Keywords:** gravitational-wave detectors, Advanced LIGO, noise regression, machine learning, neural networks

## 1. INTRODUCTION

Since September 14, 2015 (Abbott et al., 2016), gravitational-wave (GW) observatories including Advanced LIGO (aLIGO; LIGO Scientific Collaboration, 2015), Advanced Virgo (Acernese et al., 2015), and KAGRA (Kagra Collaboration, 2019) have achieved great success with dozens of GW event detected so far (Abbott et al., 2019a, 2020).

While the high-frequency ( $\gtrsim 60$  Hz) part of aLIGO's sensitivity is steadily approaching its designed target especially with the implementation of quantum squeezing (Tse and et al., 2019), there is nonetheless a big gap between the current and the designed sensitivity at lower frequencies (Martyunov et al., 2016; Buikema et al., 2020).

If we can remove the excess contamination in the sub-60 Hz band, it would greatly promote a wide array of science cases including the early warning of binary neutron star mergers (Cannon et al., 2012; Abbott et al., 2019b; Chu et al., 2020; Sachdev et al., 2020; Yu et al., 2021), the detection of intermediate-mass black holes (Mandel et al., 2008; Graff et al., 2015; Veitch et al., 2015; LIGO Scientific Collaboration and Virgo Collaboration, 2020), the constraining of eccentricities of binary black holes and hence their formation channels (Abbott et al., 2019; Romero-Shaw et al., 2019), and many more. Refer also the discussions in, e.g., Yu et al. (2018) and references therein.

The limiting factors to the current sensitivity below 60 Hz are "technical noises" due to environmental perturbations and control noises of auxiliary degrees of freedom. Unlike

fundamental noises due to quantum and thermal fluctuations in the main GW readout channel that cannot be mitigated without instrumental upgrades, the technical noises can in principle be removed *via* regression techniques as their source fluctuations are also witnessed and recorded by hundreds of auxiliary sensors (i.e., sensors that do not detect GW signals) employed by aLIGO. In fact, the linear component of technical noises has successfully been removed in aLIGO (refer to, e.g., Davis et al., 2019; Driggers et al., 2019).

The remaining challenge of the regression problem is to tackle noises that couple to the main GW readout *nonlinearly*. In fact, many noise sources in aLIGO couple in a bilinear way to the GW readout. It happens naturally when the coupling coefficient of an auxiliary channel is modulated by some slow motion ( $\lesssim 1$  Hz) in the interferometer. The modulation destroys the linear coherence between the auxiliary channel and the GW readout, forbidding the use of a standard linear regression technique like the one employed in Driggers et al. (2019). Moreover, it is typically challenging to reconstruct the modulation directly because of the pollution from large ambient motion and cross-couplings from complicated control feedback loops stabilizing aLIGO at below 1 Hz.

Fortunately, machine learning (ML) techniques, especially the use of convolutional neural networks (CNNs; Lecun et al., 2015), offer an attractive potential solution to the nonlinear noise regression problem (refer to, e.g., Ormiston et al., 2020; Vajente et al., 2020; Mogushi et al., 2021; Yu et al., 2021). By inputting to a CNN sufficiently many auxiliary witnesses that contain all the information about the noise coupling, and utilizing properly designed network structure and training strategies, we can let the algorithm figure out the coupling mechanism behind the noise even it involves nonlinearity and complicated blending of different sensors. Furthermore, after the training process, the subsequent predicting of the contamination using CNNs is highly efficient computationally, allowing the noise cleaning to be performed in real time (i.e., online). This would be especially beneficial for searches that require low latency, such as the early warning of binary neutron star mergers (Baltus et al., 2021; Yu et al., 2021). Other successful usage of ML techniques in GW astronomy include the identification of various GW events (Bayley et al., 2020; Chan et al., 2020; Dreissigacker and Prix, 2020; Huerta et al., 2020; Krastev, 2020; Schäfer et al., 2020; Wong et al., 2020; Beheshtipour and Papa, 2021; Chang et al., 2021; Chatterjee et al., 2021; López et al., 2021; Marianer et al., 2021; Mishra et al., 2021; Saiz-Pérez et al., 2021; Wei and Huerta, 2021; Yan et al., 2021), source parameter estimations (Gabbard et al., 2019; Chatterjee et al., 2020; Chua and Vallisneri, 2020; Green et al., 2020; Talbot and Thrane, 2020; Álvares et al., 2021; D’Emilio et al., 2021; Krastev et al., 2021; Williams et al., 2021; Xia et al., 2021), and detector characterization (Biswas et al., 2020; Colgan et al., 2020; Cuoco et al., 2020; Essick et al., 2020; Torres-Forné et al., 2020; Mogushi, 2021; Sankarapandian and Kulis, 2021; Soni et al., 2021; Zhan et al., 2021). Besides GW astronomy, the usage of CNNs has led to breakthroughs in a variety of topics related to time-series forecasting and classification (e.g., Refs. Ahmed et al., 2010; Fawaz et al., 2019; Pavlyshenko, 2019; Lim and Zohren, 2021 and references therein), as well as

image denoising (e.g., Refs. Pravin and Ojha, 2020; Ilesanmi and Ilesanmi, 2021; Zhou et al., 2021 and references therein).

In this study, we explore in detail how we could use CNNs to potentially mitigate the angular noise in aLIGO, which is the limiting noise source of the current sensitivity in the 30 Hz band and is also a classical example of bilinear coupling in aLIGO. Our implementation utilizes the code `Keras` (Chollet et al., 2015), a deep learning application programming interface written in `Python`, running on top of the ML platform `TensorFlow` (Abadi et al., 2015).

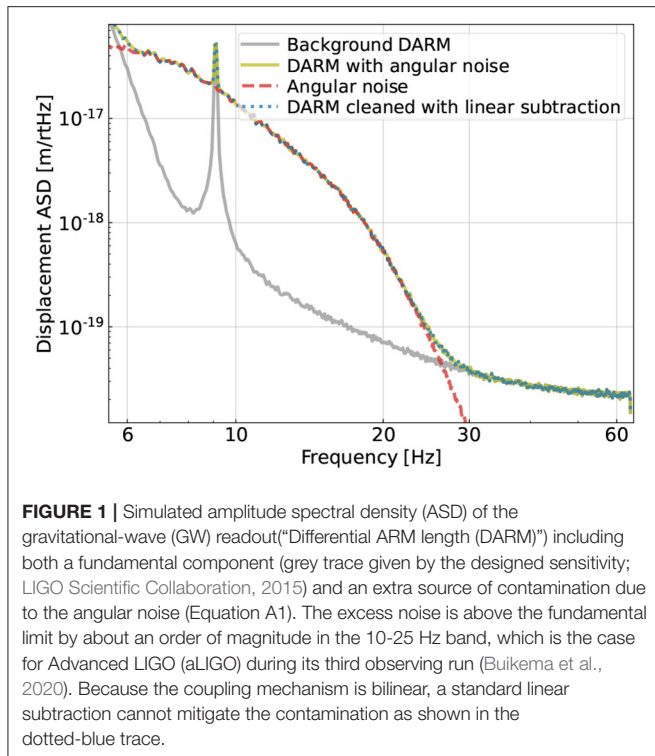
The rest of the article is organized as follows. In Section 2, we describe the coupling mechanism behind the angular noise and how we generate mock data so that our study can be carried out in a controlled way. The mathematical details will be supplemented in the **Appendix 1**. We then explore in Section 3 different CNN structures that can be employed to tackle the nonlinear noise regression problem, with a focus on a general deep-filtering structure and one inspired by the physics of the interferometer. This is followed by Section 4 in which we compare the performance of the two structures. In Section 5, we assess the CNN performance as a function of the signal-to-noise ratio, or SNR, in both the target channel (GW readout) and the input witnesses. We then demonstrate in Section 6 the effects of curriculum learning (CL), which may help the convergence of a CNN when its SNR is low in the target. Finally, we conclude and discuss the implications for future research in Section 7.

## 2. SIMULATING THE ANGULAR NOISE IN THE LIGO SYSTEM

In this study, we will focus on using CNNs to mitigate noise due to the angular control system, which is one of the major noise sources limiting the sub-30 Hz sensitivity of aLIGO currently (Buikema et al., 2020). To make our study controlled, we will use simulated time series for various channels with characteristics similar to the real aLIGO system. Throughout this study, we will use a fixed sampling rate of 128 Hz for all the time series.

In **Figure 1** we show a typical plot of the amplitude spectral density (ASD) of our simulated GW readout including both the fundamental noise (grey trace; which cannot be further reduced by offline regression) and the simulated angular noise (red-dashed trace). Note that the coupling of the angular noise to the main GW readout is bilinear (which we will describe in detail shortly). Consequently, a standard linear regression fails to mitigate its contamination (blue-dotted trace), which thus motivates us to investigate regression strategies utilizing ML. In this Section, we sketch out our simulation of the angular noise (red trace in **Figure 1**) and present the details in **Appendix 1**.

At the current power level of aLIGO, the angular control noise couples to the main GW readout (i.e., the monitor of the Differential ARM length, or “DARM”) *via* a geometrical effect. If the beam spot is not at the center of the rotational pivot of an aLIGO test mass (which is also a mirror), then an angular motion will be converted to a length fluctuation bilinearly as Barsotti et al.



(2010) and Yu (2019).

$$\delta x^{\text{(mir)}}(t) = x_{\text{spot}}^{\text{(mir)}}(t)\theta^{\text{(mir)}}(t). \quad (1)$$

Here,  $\theta^{\text{(mir)}}(t)$  corresponds to a fast ( $\gtrsim 10$  Hz) angular perturbation to the mirror, which is further induced by the sensing noise in the angular control system being fed back to the mirror. The  $x_{\text{spot}}^{\text{(mir)}}$ , on the other hand, corresponds to a slow ( $\lesssim 1$  Hz) motion of the beam spot on the test mass induced by the seismic motion.

In reality, the mirrors are not controlled locally (in the “mirror basis”) but instead on a radiation-torque basis. This choice is to tackle the Sidles-Sigg effect (Sidles and Sigg, 2006; Hirose et al., 2010; Dooley et al., 2013): as the alignment changes, it creates a radiation torque feeding back to the alignment. In other words, the alignments of the input and end test masses are no longer independent but coupled together *via* the radiation torque. Depending on the sign of the feedback, the radiation torque either hardens (i.e., makes it stiffer) or softens (makes it less stiff) the restoring torque of pendulums suspending the test masses. This allows us to decompose the alignments of the two test masses into a hard mode and a soft mode. The angular control is then performed in terms of this radiation pressure basis.

In aLIGO, the tolerance on the residual hard-mode motion is more stringent, and therefore, it has a greater control bandwidth (with a unity-gain frequency UGF  $\simeq 3$  Hz) than the soft mode (with a UGF  $\lesssim 1$  Hz). As a result, only the hard-mode control feeds back a significant amount of its sensing noise in the 10 – 30 Hz band while the soft mode has negligible motion at high

frequency. Consequently, in our simulation, we only simulate  $\theta^{\text{(h)}}$  for the high-frequency angular motion and set  $\theta^{\text{(s)}} = 0$ , where we have used the superscript “h” (“s”) to denote the motion in the hard (soft) mode. Approximately, the spectral shape of  $\theta^{\text{(h)}}$  corresponds to the red trace in **Figure 1** (refer to also, Martynov et al., 2016; Yu, 2019; Buikema et al., 2020).

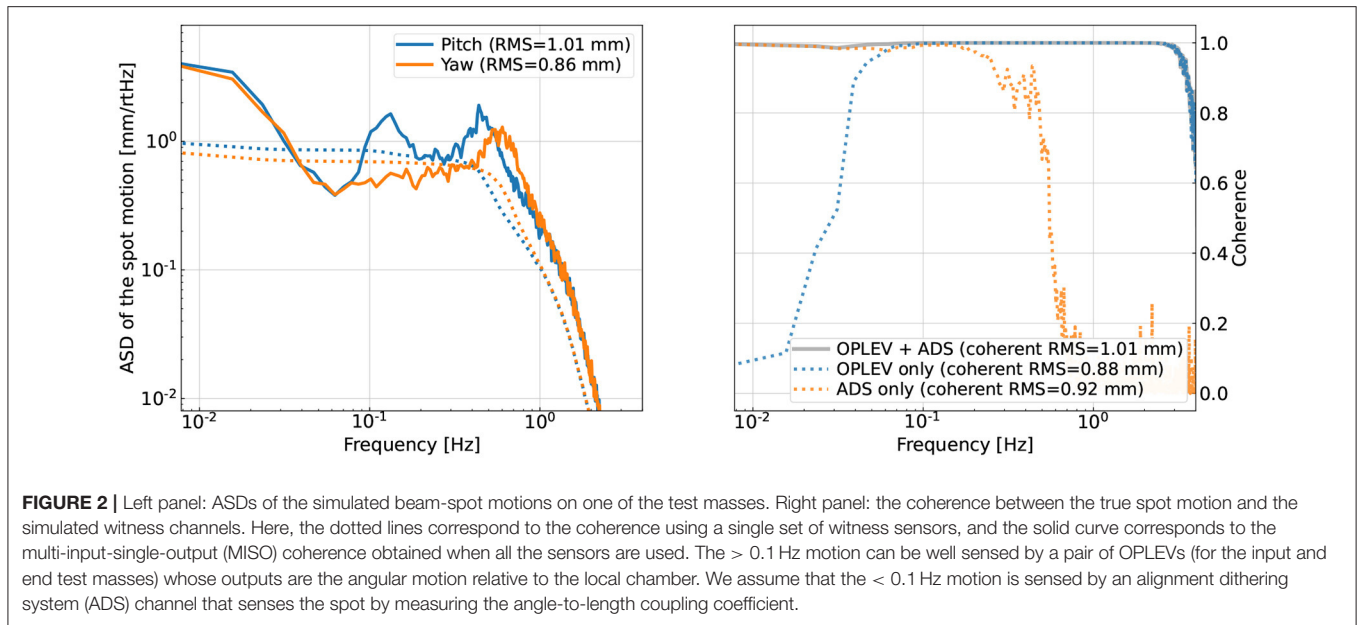
After introducing the high-frequency angular motion  $\theta(t)$ , we now describe how we simulate the low-frequency spot motion  $x_{\text{spot}}(t)$ . Here, we simulate directly the motion of each mirror.<sup>1</sup> For an ideal suspension, the mirror should only have a significant amount of angular motion in pitch because at the suspension point, the seismic motion is mostly longitudinal and the suspension should only couple it into pitch but not yaw. However, in reality, the suspension is not perfectly balanced; the imperfections in control loops also introduce cross-couplings between different degrees of freedom. As a result, the total root-mean-squared (RMS) motions in pitch and yaw can be comparable. Therefore, in our simulation, we assume they have similar RMS values of  $\simeq 1$  mm. We nonetheless give them different spectral shapes as shown in the left panel of **Figure 2** as a challenge mimicking the real aLIGO system for our CNN to tackle.

Unlike the high-frequency angular fluctuations that are directly readout in the control sensors, in aLIGO there are no sensors that directly probe the low-frequency spot motions over the entire band of interest. Instead, we would have to mix different sensors’ outputs in a frequency-dependent way in order to reconstruct the true spot motion. In fact, the reconstruction of  $x_{\text{spot}}(t)$  is the most challenging component in mitigating the angular noise in the real aLIGO system currently.

In our simulation, we assume the spot motion can be reconstructed by two sets of sensors. One set senses the spot motion *via* a modulation-demodulation technique. Following LIGO convention, we will refer to these sensors as “ADS” sensors (with ADS standing for the alignment dithering system; Buikema et al., 2020). They probe the very low frequency ( $< 0.1$  Hz) portion of the spot motion.

The other set of sensors are known as “optical levers,” and we will refer to them as OPLEVs following the LIGO convention (Black et al., 2010). They sense the spot motion in the 0.1 – 3 Hz range. Consequently, we would need to combine both ADS and OPLEV sensors to reconstruct the spot motion, as shown in the right panel of **Figure 2** where the coherence between each sensor and the simulated true spot motion is shown. We use the color orange and blue to, respectively, represent the coherence with an ADS sensor and a pair of OPLEVs (as the spot motion on a mirror depends on the angular

<sup>1</sup>Because the hard-mode and soft-mode control loops have different shapes, the spot motion in the hard mode and the soft mode in principle will have different values (similar to the case of high-frequency angular motion). However, subtleties arise because the spot motion governing the length fluctuation in Equation (1) is measured with respect to a local rotational pivot, whereas the angular control is performed relative to an input beam (i.e., a global quantity). At low-frequencies ( $< 1$  Hz), the two reference points drift independently. Our current simulation does not capture this effect and for simplicity, we treat the spot motion on each mirror as an independent.



motion of both mirrors forming the cavity). The multi-input-single-output (MISO) coherence using all the sensors is shown in the grey trace. In the legend, we also quote the RMS value of spot motion that is coherent with the specific set of sensors (the true spot motion has an RMS of 1 mm here). Note that the coherence is related to the SNR at each frequency bin as

$$\text{SNR}^2 = \frac{\text{Coherence}}{1 - \text{Coherence}}. \quad (2)$$

In total, we would need to include at least 16 sensors (8 ADSs + 8 OPLEVs) in order to sense spot motions on the 4 test masses in both pitch and yaw over the entire band of interest. Together with 4 fast channels corresponding to the 4 hard mode feedback (as there are 2 ARM cavities and each cavity can move in both pitch and yaw), we include 20 auxiliary channels in total in our simulation of the angular noise.

We note that in reality, more channels may be needed for the spot motion. In the right panel of **Figure 2**, while the simulated OPLEVs have comparable sensitivity to the real ones, we are nonetheless being optimistic about the SNR of the ADS sensors. As we will see later in Section 5.2, without sufficient SNR in the  $< 0.1$  Hz band the subtraction performance will be significantly limited. Therefore, in order to achieve successful noise regression in reality, we would need either more low-frequency channels with a more complicated CNN structure or more accurate sensing schemes for the  $\lesssim 0.1$  Hz spot motion.

### 3. GENERAL VS. SPECIFIC CNN STRUCTURES

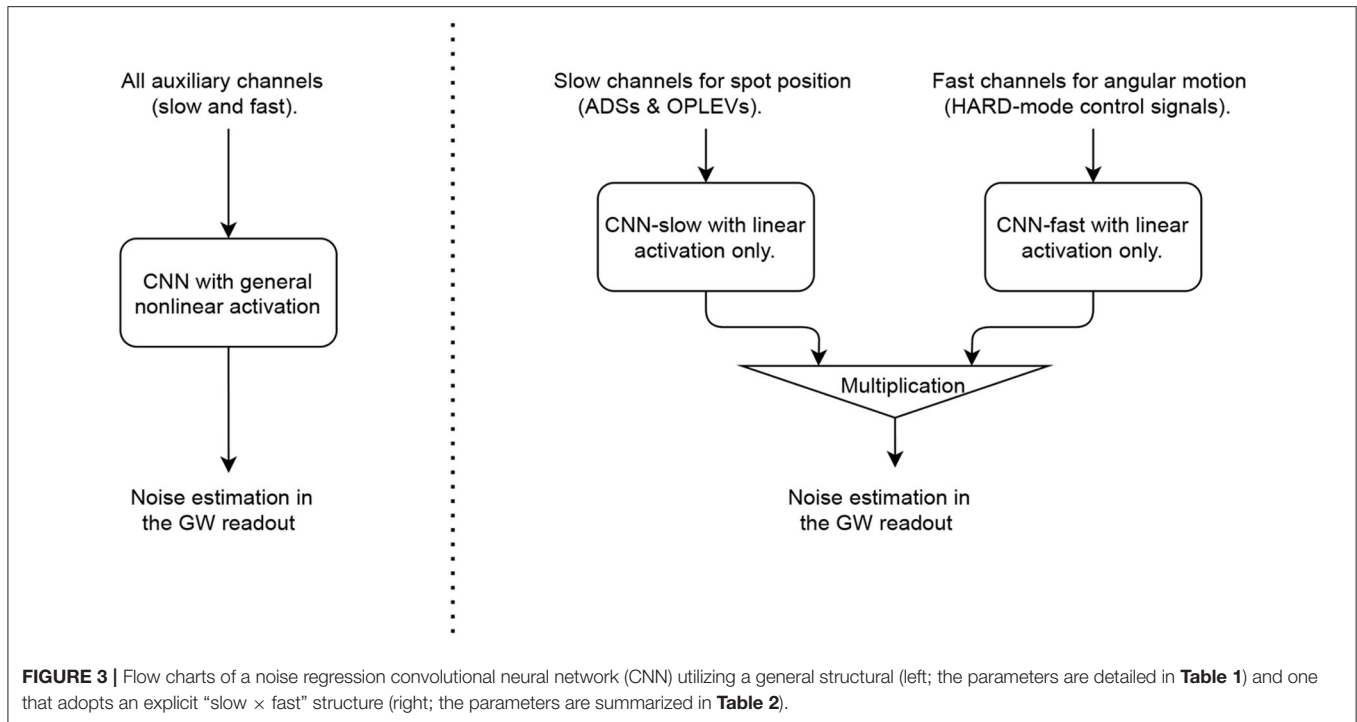
Having described our noise simulation, we now discuss the mitigation of the noise using CNNs. In particular, we discuss the

process by which we explore different CNN structures in order to optimize the regression performance.

One option is to use a general deep-filtering structure as illustrated in the left part of **Figure 3**. Such a CNN should contain sufficiently many convolutional layers together with densely connected layers, and at least some of the layers should involve a nonlinear activation function such as ReLU or ELU at each layer's output. The convolutional layers would then behave similarly to finite-impulse-response filters, enabling the frequency-dependent blending of different auxiliary sensors' time series (i.e., "input" to the CNN). Moreover, with sufficiently many layers with nonlinear activations, the nonlinearity involved in the coupling mechanism can be represented in a series-expansion sense. Such a deep-filtering CNN has the advantage of being straightforward and general. It requires a small amount of prior knowledge and, thus, is particularly suitable to handle noise sources with unknown couplings.

On the other hand, for the bilinearly coupled angular noise described in Section 2, we nonetheless know how the noise propagates to the GW readout. The challenging part of its mitigation is the reconstruction of the spot motion on each test mass. In this scenario, we can in fact utilize our knowledge of the coupling mechanism and design a more specific CNN structure to tackle this problem as shown in the right part of **Figure 3**. In particular, we divide the auxiliary channels into a slow set [ADSs and OPLEVs for the spot motion  $x_{\text{spot}}(t)$ ] and a fast set [HARD mode control signals for the angular motion  $\theta^{(h)}(t)$ ]. Each set goes through a CNN (which will be referred to as "CNN-slow" and "CNN-fast," respectively) which requires only linear activation. These CNNs act effectively as FIR filters to convert linearly (yet in frequency-dependent ways) the auxiliary witnesses outputs in digital counts into "super-sensors" that monitor the instantaneous spot positions and alignments of the LIGO test masses. Once the CNNs have





**TABLE 1** | Network using a general structure.

	Layer	Output dimension	Kernel size	Activation
General structure	Conv1D	32	1,024	Linear
	Dropout	–	–	rate = $8 \times 10^{-6}$
	Conv1D	128	32	Linear
	Conv1D	16	8	Linear
	Conv1D	128	16	Linear
	Dropout	–	–	rate = $8 \times 10^{-6}$
	Dense	128	–	ELU
	Dense	64	–	ELU
	Dense	32	–	ELU
	Dense	8	–	ELU
	Dense	1	–	Linear

All the 20 auxiliary channels (16 slow channels and 4 fast ones) are directly input to the first layer and the nonlinearity is achieved by a series of Exponential Linear Unit (ELU) activation in the hidden layers. In this network, there are about 8,63,000 trainable parameters in total.

**TABLE 2** | Network explicitly utilizing the “slow × fast structure.”

	Layer	Output dimension	Kernel size	Activation
slow path	Conv1D	32	1,024	Linear
	Dropout	–	–	rate = $1 \times 10^{-8}$
	Conv1D	128	32	Linear
	Conv1D	16	8	Linear
	Conv1D	128	16	Linear
	Dropout	–	–	rate = $1 \times 10^{-8}$
	Dense	128	–	Linear
	Dense	64	–	Linear
	Dense	32	–	Linear
	Dense	4	–	Linear
fast path	Conv1D	16	8	Linear
	Dense	64	–	Linear
	Dense	32	–	Linear
	Dense	4	–	Linear
after the multiply layer	Dense	32	–	Linear
	Dense	8	–	Linear
	Dense	4	–	Linear
	Dense	1	–	Linear

We input the 16 slow channels to the slow path and use a series of linear operations (signal filtering and blending) to form 4 super-sensors that effectively monitor the spot position on each test mass. These super-sensors’ output is then multiplied with the 4 angular witnesses input to the fast path by a multiply layer to eventually become a length signal. Note that all the activations are linear as the only nonlinearity is explicitly incorporated by the multiplied layer. This network has about 7,36,000 trainable parameters in total.

constructed the super-sensors, the nonlinearity required by this problem is given by Equation (1) which is made explicit via a multiplication layer. For the rest of the Section, we will then demonstrate how the specific structure may improve the mitigation performance relative to the general one. In aLIGO, the coupling mechanisms behind many noise sources are in fact established (refer to, e.g., Martynov et al., 2016 and Buikema et al., 2020 and references therein; refer to also Section 7). This Section, thus, serves further as a demonstration of the benefits of incorporating our knowledge on the instrument when considering noise regression problems.

To compare the performance of the two structures, we consider CNNs with hyper-parameters listed in **Tables 1, 2**, respectively for the general and the “slow×fast” structures. Note

here the input is a 3-dimension array. The first dimension corresponds to the batch dimension (i.e., different realization of the simulated data). The second dimension corresponds to the number of input channels. This is the dimension that is densely connected to allow the CNNs to recognize the correct combinations of different channels as required by, e.g., the underlying cavity geometry. It starts as 20 for the general structure, 16 for CNN-slow, 4 for CNN-fast, and its shape changes according to the “output dimension” column in **Tables 1, 2** in subsequent layers. Note that for the final layer it should be 1 corresponding to the main GW readout (i.e., the target of the training). Finally, the final dimension corresponds to the temporal dimension and it is the axis along which the convolution is performed to achieve frequency-dependent filtering.

In LIGO, the spot position is not directly sensed but it has to be reconstructed from different sets of sensors. Therefore, most of the trainable parameters are in the CNN-slow path which contains 4 convolutional layers followed by 4 densely connected layers. In the first convolutional layer, we further choose a kernel size of 1,024 (corresponding to 8 s of data) such that the CNN can achieve low- and high-passing at  $\sim 0.1$  Hz (**Figure 2**). Other hyperparameters are tuned empirically at the current stage. It is possible to enhance the network’s performance further with more educated and exhaustive tuning, though we defer this optimization to future studies.

The fast angular motions, on the other hand, are directly sensed in aLIGO. Therefore, the CNN-fast path requires only a small number of trainable parameters. In fact, if the fast sensors are properly preconditioned to physical angles  $\theta^{(h)}(t)$  (by filtering them with the aLIGO suspension model), they can be directly fed to the multiplication layer. Here, we nonetheless leave a small CNN-fast network in the path so that the preconditioning does not need to be exact.

To make fair comparisons, we intentionally keep the CNN using the general structure (**Table 1**) to be similar to CNN-slow with comparable numbers of trainable parameters. We empirically choose ELU activations in the hidden, densely connected layers to achieve the nonlinear operation in Equation (1).<sup>2</sup>

We employ a custom loss function  $\mathcal{L}$  based on the band-limited PSD of the GW readout as

$$\mathcal{L} = k \int_{f_1}^{f_2} \frac{S_n(f)}{S_n^{(0)}(f)} \left( \frac{f}{f_1} \right)^\alpha df, \quad (3)$$

where  $S_n$  is the PSD of the GW readout (i.e., the target channel) after the subtraction it is normalized by  $S_n^{(0)}$ , the original PSD before noise subtraction (which is fixed during the training). We have additionally included a power-law weighting with an index  $\alpha$ ; we empirically set  $\alpha = -1$  in our case. The loss is computed over the frequency band  $[f_1, f_2]$ , and we tune the overall gain  $k$  such that the initial loss is about order unity.

<sup>2</sup>In our case, we empirically find that having nonlinear activations in the convolutional layers does not significantly enhance the subtraction performance but makes the convergence slower. Therefore, we apply only the linear activation to the convolutional layers.

Note that Equation (3) does not constrain the DC value of the GW readout. While such a DC offset in the GW readout does not directly affect the sensitivity to astrophysical sources, we nonetheless choose to avoid introducing any DC offsets during noise regression. This is achieved by adding the regular mean-square error together with Equation (3) to form our final loss function used during the training process. The relative weights are tuned such that the mean-square error contributes about 10% to the total loss initially.

Because the RMS value of the spot motion contains physical information, we do not normalize each channel by its variance [which may be different for different datasets as the RMS of  $x_{\text{spot}}(t)$  varies over time]. Instead, we want each channel to be always normalized by the same constant. For this purpose, we calibrate the ADS channels to be in [mm], OPLEVs in [mrad], and the main GW readout (“DARM”) in [fm]. For the HARD-mode error signals, we calibrate them to [pm] and further multiply them by the open-loop gain to precondition them to be proportional to the physical  $\theta^{(h)}(t)$ .

To train the CNNs to recognize the underlying physical coupling mechanism in Equations (1) and (A1), we simulate 1,536 s of data for the main GW readout channel (target) and another 20 time series of the same length for the auxiliary witness channels (inputs to the CNNs). Out of the 20 auxiliary channels, 16 are used for the construction of the spot positions, and 4 are used for the fast angular perturbations. All the channels are sampled at the same rate of 128 Hz. During the training, we further divide the time series into 8 batches, each containing 192 s of data. Besides the training data, we additionally prepare 256 s of data for validation and 256 s for testing. We train each CNN until the loss on the validation data plateaus.

## 4. COMPARING THE EXTRAPOLATION PERFORMANCE

After describing the two network structures in the previous section, we now compare their performance.

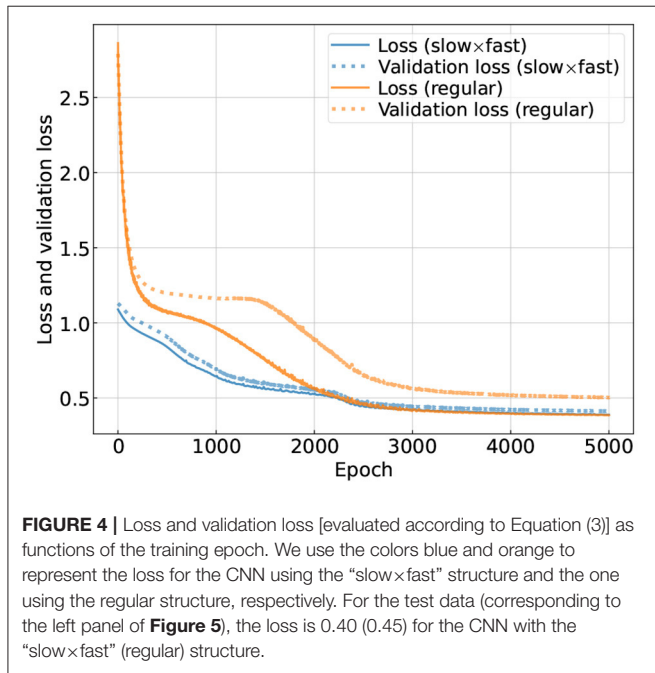
One interesting comparison is the convergence speed which we examine in **Figure 4** by showing the loss and the validation loss as functions of the training epoch. Because the “slow×fast” CNN employs our knowledge of the noise coupling mechanism, it naturally converges faster than the regular CNN assuming little prior knowledge. Eventually the training loss for the two CNNs plateaus at similar values, yet for the specific one, the validation loss follows the loss more closely than the regular one, indicating that the specific structure is less subject to over-fitting.

More importantly, we can examine the subtraction results on the testing data in **Figure 5**. Here, the brown trace is the ASD including both the fundamental noise and the bilinear noise. The ASD of the fundamental noise is also shown separately in the grey trace; in the ideal case, a CNN should use the information stored in the auxiliary channels to reduce the brown ASD to the grey one. The ASDs of residual time series cleaned with our CNNs are shown in the dotted traces. Here, we use the color orange and blue to respectively represent the results obtained from the general (**Table 1**) and the “slow×fast” (**Table 2**) CNNs.

One question we want to address is how well each CNN extrapolates. Particularly, one quantity we extrapolate is frequency. While during the training we would like to focus on the frequency band where the noise contamination is the highest to get the best mitigation there, we also want to avoid injecting excess noise outside this band during the subtraction process. To test this point, we, thus, set the training band (i.e., the band where we compute the loss function in Equation 3) to be  $[f_1, f_2]=[10, 20]$  Hz, corresponding to the two red-vertical lines in Figure 5. As shown in the figure, inside the  $[10, 20]$  Hz

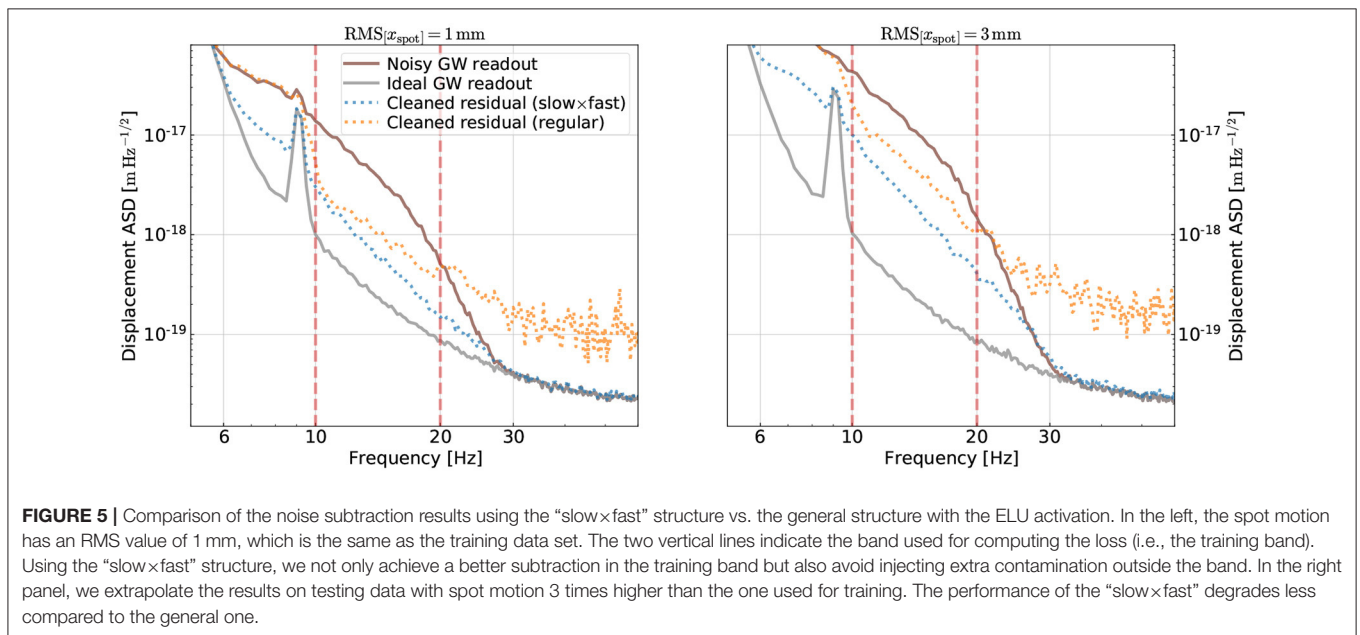
band both the general and the “slow×fast” CNNs have decent and comparable subtraction performance (about a factor of 10 at 10 Hz; the “slow×fast” CNN has a slightly better performance at other frequencies). The losses are 0.45 and 0.40 for the general and the “slow×fast” CNNs, respectively. Outside the training band, on the other hand, the “slow×fast” CNN significantly outperforms the general one. Whereas, the general CNN starts to inject excess noise into the GW readout almost immediately outside the training band, the one adopting the “slow×fast” structure continues to reduce the noise for the entire band from 6 Hz to 30 Hz where the angular noise is above the fundamental limit. Above 30 Hz, the “slow×fast” CNN also avoids making the noise worse than the fundamental limit. We, thus, see that by utilizing our knowledge of physical systems, we can achieve good extrapolation properties with respect to frequency.

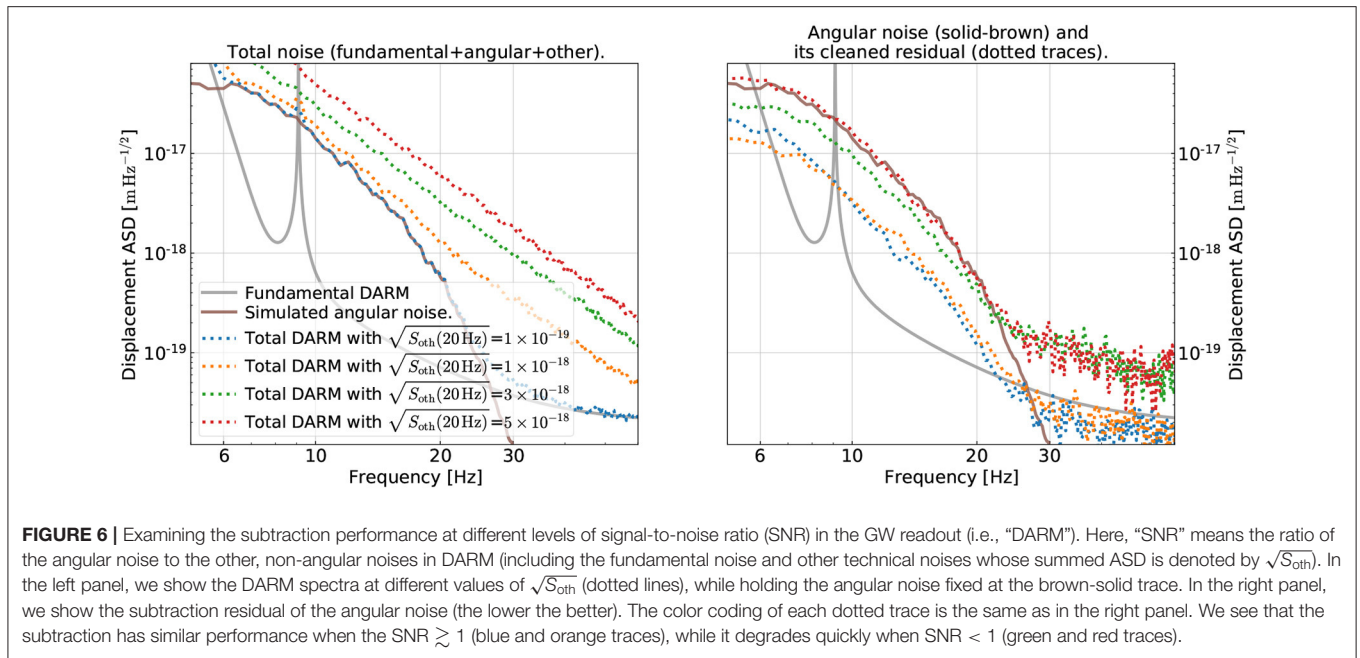
Meanwhile, we would also like to assess how the CNNs perform at different values of RMS of the spot motion. This is because in the real aLIGO system the spot motion is induced by the  $< 1$  Hz seismic motion which varies over time, and the CNN’s performance would, thus, need to be robust against the variation in the spot motion. We address this point by simulating an additional 256 s of data with an RMS spot motion of 3 mm and then testing our CNNs on this dataset with 3 times higher spot motion than the training dataset. The result is summarized in the right panel in Figure 5. While the fractional noise reduction at 10 Hz degrades for both CNNs, the degradation is less if one uses the “slow×fast” structure than the general structure.



### 5. SNR REQUIREMENTS

In the section above, we have considered a simple case where the total GW readout contains only a fundamental component and all the excess contamination is due to the angular noise. In reality, there are many other types of excess noise in the low-frequency part of the aLIGO sensitivity band whose information





**FIGURE 6** | Examining the subtraction performance at different levels of signal-to-noise ratio (SNR) in the GW readout (i.e., “DARM”). Here, “SNR” means the ratio of the angular noise to the other, non-angular noises in DARM (including the fundamental noise and other technical noises whose summed ASD is denoted by  $\sqrt{S_{\text{oth}}}$ ). In the left panel, we show the DARM spectra at different values of  $\sqrt{S_{\text{oth}}}$  (dotted lines), while holding the angular noise fixed at the brown-solid trace. In the right panel, we show the subtraction residual of the angular noise (the lower the better). The color coding of each dotted trace is the same as in the right panel. We see that the subtraction has similar performance when the SNR  $\gtrsim 1$  (blue and orange traces), while it degrades quickly when SNR  $< 1$  (green and red traces).

is not captured in the auxiliary channels we input to the CNN. Effectively, the presence of other contamination reduces the angular noise’s SNR in the GW readout (i.e., the target channel; note, here, we treat the angular noise as the signal because it is what we want to remove by the CNN). We, thus, explore how the CNN’s performance varies with respect to the GW readout in Section 5.1.

Another thing we would like to explore is the SNR of ADS sensors for the very low frequency ( $\lesssim 0.1$  Hz) spot motion. The coherence level shown in the right panel of **Figure 2** is likely to be an optimistic estimation for the  $\lesssim 0.1$  Hz band because sensing the spot motion in this band is intrinsically challenging. Not only do the ADS sensors have high sensing noise as described in Section 2, but also the spot estimation could be biased by other contamination mechanisms.<sup>3</sup> Therefore, we also assess the CNN’s performance with respect to the SNR in the ADS sensors in Section 5.2.

### 5.1. In the GW Readout

To modify the SNR of the simulated angular noise in the main GW readout, we simulate the total length fluctuation  $\delta x_{\text{tot}}$  in the GW readout as

$$\delta x_{\text{tot}}(t) = \delta x_{\text{fun}}(t) + \delta x_{\text{ang}}(t) + \delta x_{\text{oth}}(t), \quad (4)$$

where  $\delta x_{\text{fun}}$  corresponds to the fundamental noise component (grey trace in **Figure 1**),  $\delta x_{\text{ang}}$  is the angular noise we want to

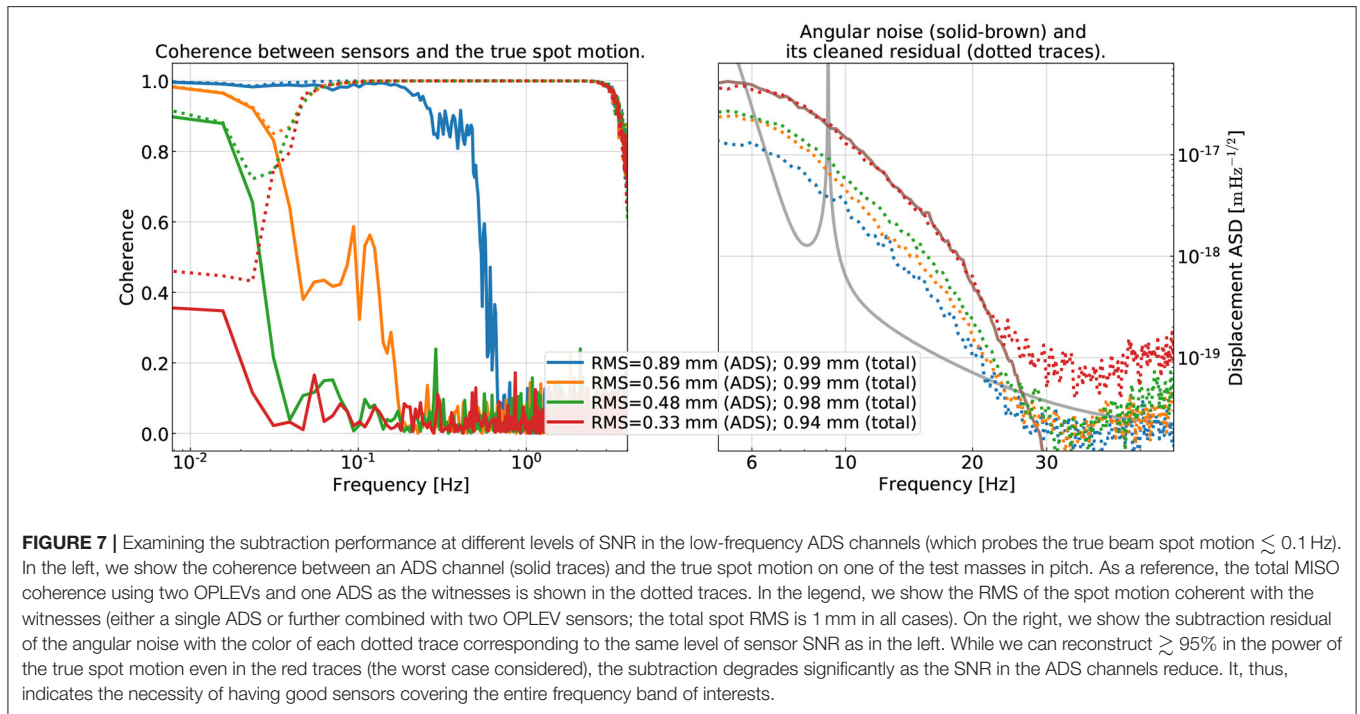
<sup>3</sup>For example, the power circulating in the arm cavity may be modulated by  $\theta^{(\text{mir})}(t)$  due to both the residual angular motion at  $< 1$  Hz and/or imperfections in the mirror coating. The power fluctuation could further lead to a longitudinal perturbation mitigated by radiation pressure in addition to the geometrical effect Equation (1). Thus, the effective spot position inferred at a given dithering frequency could be a biased estimation for another frequency.

subtract remove (red trace in **Figure 1**), and  $\delta x_{\text{oth}}$  are other types of contamination to DARM whose information is not contained in the auxiliary channels we input. For the other contamination, we further assume that it has an ASD given by  $\sqrt{S_{\text{oth}}(f)} \propto f^{-3}$ . This way, we can control the SNR  $[= \sqrt{S_{\text{ang}}(f)} / \sqrt{S_{\text{fun}}(f) + S_{\text{oth}}(f)}]$  of the bilinear noise by changing the overall magnitude of the other noise’s ASD.

In the left panel of **Figure 6**, we show the ASDs of the total DARM displacement, now including the contribution from other contamination simulated according to Equation (4). We use dotted lines with different colors to indicate different levels of the other noise while the angular noise is held fixed as shown in the solid-brown trace in this section. For each level of SNR, we regenerate 2,048 s of data (1,536 s for training, 256 s for validation, and 256 second for testing). We then train our CNN on the new data and check how much it could mitigate the bilinearly coupled angular noise (solid-brown trace in the left panel of **Figure 6**). Here, we focus on the CNN with the “slow×fast” structure (**Table 2**) as it has a better performance than the general one (Section 3). The auxiliary channels are still assumed to have good sensitivity as shown in the right panel of **Figure 2**. The training strategy follows the one outlined in Section 3 except for that we now evaluate the loss, Equation (3), over a broader band of  $[f_1, f_2] = [8, 40]$  Hz.

In the right panel, we show the ASDs of the residual time series after noise mitigation by our CNN. For presentation purposes, we show specifically the residual angular noise component in each dotted line. This is obtained by subtracting from the residual [i.e., the difference between the original  $\delta x_{\text{tot}}(t)$  and the one predicted by our CNN] further the  $\delta x_{\text{fun}}(t)$  and  $\delta x_{\text{oth}}(t)$  components when generating the plot, which is possible as we are dealing with the simulated data. Different colors correspond to different levels of SNR and the correspondence is the same as in the left panel.





**FIGURE 7** | Examining the subtraction performance at different levels of SNR in the low-frequency ADS channels (which probes the true beam spot motion  $\lesssim 0.1$  Hz). In the left, we show the coherence between an ADS channel (solid traces) and the true spot motion on one of the test masses in pitch. As a reference, the total MISO coherence using two OPLEVs and one ADS as the witnesses is shown in the dotted traces. In the legend, we show the RMS of the spot motion coherent with the witnesses (either a single ADS or further combined with two OPLEV sensors; the total spot RMS is 1 mm in all cases). On the right, we show the subtraction residual of the angular noise with the color of each dotted trace corresponding to the same level of sensor SNR as in the left. While we can reconstruct  $\gtrsim 95\%$  in the power of the true spot motion even in the red traces (the worst case considered), the subtraction degrades significantly as the SNR in the ADS channels reduce. It, thus, indicates the necessity of having good sensors covering the entire frequency band of interests.

As shown in the plot, when the angular noise has an SNR  $\gtrsim 1$  at each frequency bin in the GW readout, the CNN can achieve decent and consistent subtraction performance as shown by the blue and orange traces. At 10 Hz, the angular noise could be reduced by about an order of magnitude. However, when the SNR drops below unity as shown in the green trace (with an SNR of about 0.5 at 10 Hz), the CNN could only marginally reduce the angular noise in the 10-20 Hz band, and it starts to inject excess contamination at higher frequencies.

Therefore, for the CNN to be able to recognize the correct noise coupling mechanism, it would typically need an SNR of unity in the target channel. Nonetheless, we note that it is not a *necessary* condition. As we will discuss later in Section 6, we may still be able to remove contaminations in DARM with a sub-unity SNR if we utilize active injection and CL.

### 5.2. In the Witness Sensors

So far we have assumed the ADS sensors have good SNR of the true spot motion below 0.1 Hz as indicated in the left panel of **Figure 2**. This may be an optimistic assumption in reality as discussed at the beginning of Section 5. We, thus, explore here how does the SNR in the ADS channels (i.e., inputs or witnesses) affects the performance of the CNN.

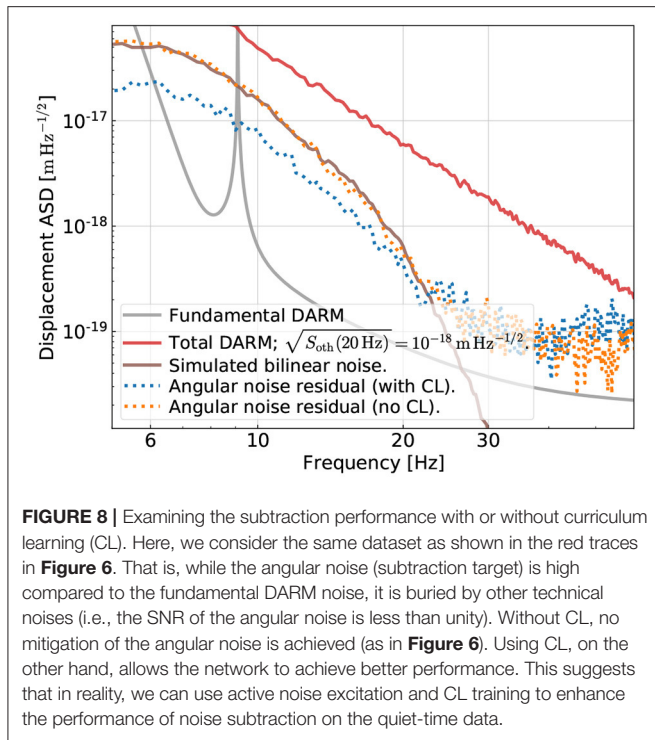
For this purpose, we vary the sensing noise in each ADS channel to modify its coherence with the true spot motion. This is indicated in the left panel of **Figure 7**. We use the solid trace to represent the coherence between an ADS’s output and the true spot motion it senses, and the dotted trace the MISO coherence if we further include a pair of OPLEVs. In the legend, we quote the RMS value of the spot motion that is coherent with the sensors

(the true spot motion has an RMS of 1 mm). Note the coherence is related to the SNR *via* Equation (2).

We repeat the training process now on data with noisy ADS sensors and the subtraction result is summarized in the right panel of **Figure 7**. The color coding of each curve has the same meaning as in the left panel. We see that as the ADS sensors’ sensitivity decreases, the noise subtraction performance degrades significantly. Take the green trace as an example. Even the MISO coherence is greater than at least 0.7 over the entire band of interest and the RMS of the spot motion coherent with the sensors (ADS + OPLEVs) almost matches the RMS of the true value, the amount of noise the CNN can subtract reduces by a factor of 2 compared to the blue trace. Our study, thus, suggests that the high-accuracy reconstruction of the beam spot on each test mass over the entire  $< 1$  Hz band is crucial for the success of the mitigation of the angular noise. In fact, spot reconstruction should be a topic deserving dedicated studies in its own right.

## 6. CURRICULUM LEARNING

As discussed in the previous section, in order for the CNN to be able to remove the angular noise, we would need to have good SNRs in both the target (the main GW readout) and the witness sensors (the ADS sensors). On the other hand, the difference exists between the two scenarios. If the SNR is low in the witnesses (Section 5.2), it means that we do not have enough information to recover the noise coupling and the only solution is to incorporate more sensors to recover the missing information. If the SNR is low only in the target (Section 5.1), then the fact that the CNN does not achieve a good subtraction is simply due to it not converging to the right physics during the training process.



All the necessary information to predict the angular noise is still available. In this case, we show in this section that we can still have decent mitigation if we utilize CL techniques (George and Huerta, 2018a,b).

The key idea behind CL is the following. We can first train the CNN on datasets with a high SNR in the target channel, which can be produced in reality with active injections. A high-SNR dataset is a simple task to be tackled and it helps the CNN to converge to the right physics initially. We can then gradually increase the task's difficulty by incorporating data with lower excitation levels, and eventually the quiet-time data (i.e., data without any active injections). Since the CNN is guided through the training process, it stays close to the true physics and can thus predict the desired target even using quiet-time data. This process is especially suitable to be combined with a physically-inspired CNN structure (such as the “slow×fast” one) that has a good extrapolation property.

We demonstrate this point, here, by first training our CNN on data corresponding to the blue traces in **Figure 6**.<sup>4</sup> After its convergence, we then incorporate into the training set the data corresponding to the orange traces (as different “batches”) and continue training the CNN obtained from the first step until its validation loss plateaus. In the last step, we further include data corresponding to the red trace (with the lowest SNR in the GW readout) and retrain the CNN from the previous step to convergence.

<sup>4</sup>In reality,  $\sqrt{S_{\text{0th}}}$  is approximately fixed and to create high-SNR data we should instead inject excess motion to, e.g.,  $\theta^{(h)}$ . Nonetheless, since what matters most is the SNR of the bilinear noise in the GW readout, here, we simply reuse the existing data.

After the CL training, we test the resultant CNN on the 256-s testing data for the red trace in **Figure 6** and the result is presented in **Figure 8** in the dotted-blue trace. As a comparison, we also showed the subtraction result without CL in the dotted-orange trace (using the CNN obtained in Section 5.1). Whereas the CNN could not mitigate any of the angular noise without CL, we see that with the help of CL the CNN can reduce the noise by a factor of 2 at around 10 Hz. It, thus, demonstrates that CL could be a useful training strategy to help the CNN remove noise that has a low SNR in the GW readout (yet it can still be high compared to the fundamental limit and, thus, needs to be tackled).

## 7. SUMMARY AND DISCUSSION

In this study, we explored how we may use CNNs to improve the sub-60 Hz sensitivity of aLIGO. Here, we focused specifically on the bilinearly-coupled angular noise which is one of the limiting noise sources in the 10-30 Hz band. Using simulated data with characteristics similar to the real aLIGO system (Section 2), we explored various factors affecting a CNN's performance. One of the most critical factors is to utilize our knowledge in the design of CNN structures (Section 3). Specifically, a “slow×fast” structure is suitable to mitigate the angular noise as it incorporates the nonlinearity involved in the problem in an explicit way, which leads to good extrapolation properties (Section 4). We further explored the SNR requirements in both the target (GW readout; Section 5.1) and the witness sensors (ADS sensors; Section 5.2) for the CNN to converge. To overcome the lack of SNR in the witnesses, it would require improving the sensing technology and/or including more sensors to recover the information. On the other hand, when the SNR is only low in the target, we demonstrated in Section 6 that CL training can be used to facilitate the CNN's convergence.

While in this study, we focused on the angular noise that couples to DARM according to Equation (1), we note that the coupling mechanisms of many other noises in LIGO share the same bilinear structure and can be modeled as a fast channel modulated by a slow one (i.e., the product of the two). For example, the signal-recycling cavity's length fluctuation (a fast channel) couples to DARM with a slow modulation due to low-frequency variations in the DARM offset. Other noise like light scattering, when expanded into a Taylor series, will also have terms similar to Equation (1) serving as the lowest-order nonlinear terms. Therefore, the “slow×fast” CNN structure we present in Section 3 will have broad applications in LIGO noise mitigation beyond just the angular noise. We plan to explore this point more in future studies.

We note that the CL result shown in **Figure 8**, while having an improved performance compared to the case without CL, has not yet reached a level comparable to the case where it converges to the right physics (e.g., the blue and orange traces in the right panel of **Figure 6**). It, thus, indicates that the detailed CL steps have rooms for further optimization, which we defer to be explored by future studies.

Throughout the analysis, we have focused on using simulated data. As we have mentioned in the main text, mitigating the angular noise in the real aLIGO system is more challenging. This is because there are many other noise sources in the same frequency band (including ones that are not yet captured by the current noise budgeting; Martynov et al., 2016; Buikema et al., 2020), limiting the SNR of the angular noise in the GW readout. Moreover, to reconstruct  $x_{\text{spot}}(t)$ , it is likely requiring more sensors than just the ADSs and OPLEVs. We also ignored potential transients (“glitches”) in both the GW readout and witness sensors, which is another crucial application of ML in GW astrophysics in its own right (refer to, e.g., Cuoco et al. (2020) and references therein). On the real data, our CNN has not yet achieved a significant broadband reduction of nonlinearly coupled noises, yet it shows promising signs such as removing sidebands around dithering lines that are used to create the ADS signals (Yu and Adhikari, 2021). We plan to investigate this further in future studies, especially combining it with a fine-tuned CNN hyper-parameter set and an optimized CL training strategy. Meanwhile, LIGO has released a 3-h data stretch in its second observing run including major auxiliary channels<sup>5</sup>. We would, thus, like to encourage interested readers to use either the CNN structures we proposed in this study or original CNN structures to help the further improvements of aLIGO’s sensitivity.

<sup>5</sup>The data can be downloaded at <https://www.gw-openscience.org/auxiliary/GW170814/>.

## REFERENCES

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org.
- Abbott, B. P., Abbott, R., Abbott, T. D., Abernathy, M. R., Acernese, F., Ackley, K., et al. (2016). Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.* 116, 061102. doi: 10.1103/PhysRevLett.116.061102
- Abbott, B. P., Abbott, R., Abbott, T. D., Abernathy, M. R., Acernese, F., Ackley, K., et al. (2019). Search for eccentric binary black hole mergers with advanced LIGO and advanced virgo during their first and second observing runs. *Astrophys. J.* 883, 149. doi: 10.3847/1538-4357/ab3c2d
- Abbott, B. P., Abbott, R., Abbott, T. D., Abraham, S., Acernese, F., Ackley, K., et al. (2019a). GWTC-1: a gravitational-wave transient catalog of compact binary mergers observed by LIGO and virgo during the first and second observing runs. *Phys. Rev. X* 9, 031040. doi: 10.1103/PhysRevX.9.031040
- Abbott, B. P., Abbott, R., Abbott, T. D., Abraham, S., Acernese, F., Ackley, K., et al. (2019b). Low-latency gravitational-wave alerts for multimessenger astronomy during the second advanced LIGO and virgo observing run. *Apj* 875, 161. doi: 10.3847/1538-4357/ab0e8f
- Abbott, R., Abbott, T. D., Abraham, S., Acernese, F., Ackley, K., Adams, A., et al. (2020). GWTC-2: compact binary coalescences observed by LIGO and virgo during the first half of the third observing run. *arXiv e-prints*, arXiv:2010.14527.
- Acernese, F., Agathos, M., Agatsuma, K., Aisa, D., Allemandou, N., Allocca, A., et al. (2015). Advanced Virgo: a second-generation interferometric gravitational wave detector. *Class. Quant. Grav.* 32, 024001. doi: 10.1088/0264-9381/32/2/024001
- Ahmed, N. K., Atiya, A. F., Gayar, N. E., and El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometr. Rev.* 29, 594–621. doi: 10.1080/07474938.2010.481556
- Álvares, J. D., Font, J. A., Freitas, F. F., Freitas, O. G., Morais, A. P., Nunes, S., et al. (2021). Exploring gravitational-wave detection and parameter

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

HY performed the training and testing of the CNNs. RA supervised the research. Both authors contributed to the article and approved the submitted version.

## FUNDING

HY is supported by the Sherman Fairchild Foundation. RA is supported by NSF grant no. PHY-1764464. The authors gratefully acknowledge the computational resources provided by the LIGO Laboratory and supported by NSF grant nos. PHY-0757058 and PHY-0823459.

## ACKNOWLEDGMENTS

We thank Gabriele Vajente and Szabolcs Marka for helpful discussions and comments during the preparation of this manuscript, and we acknowledge feedback from fellow participants in the LIGO Machine Learning Algorithms Call.

- inference using deep learning methods. *Class. Quant. Gravity* 38, 155010. doi: 10.1088/1361-6382/ac0455
- Baltus, G., Janquart, J., Lopez, M., Reza, A., Caudill, S., and Cudell, J.-R. (2021). Convolutional neural networks for the detection of the early inspiral of a gravitational-wave signal. *arXiv e-prints*, arXiv:2104.00594. doi: 10.1109/CBMI50038.2021.9461919
- Barsotti, L., Evans, M., and Fritschel, P. (2010). Alignment sensing and control in advanced LIGO. *Class. Quant. Gravity* 27, 084026. doi: 10.1088/0264-9381/27/8/084026
- Bayley, J., Messenger, C., and Woan, G. (2020). Robust machine learning algorithm to search for continuous gravitational waves. *Phys. Rev. D* 102, 083024. doi: 10.1103/PhysRevD.102.083024
- Beheshtipour, B., and Papa, M. A. (2021). Deep learning for clustering of continuous gravitational wave candidates. II. Identification of low-SNR candidates. *Phys. Rev. D* 103, 064027. doi: 10.1103/PhysRevD.103.064027
- Biswas, A., McIver, J., and Mahabal, A. (2020). New methods to assess and improve LIGO detector duty cycle. *Class. Quant. Gravity* 37, 175008. doi: 10.1088/1361-6382/ab8650
- Black, E. D., Chelermongsak, T., DeSalvo, R., Korth, Z., Barton, M., Cook, D., et al. (2010). Advanced-LIGO optical levers design requirements. *LIGO Document T0900174*.
- Buikema, A., Cahillane, C., Mansell, G. L., Blair, C. D., Abbott, R., Adams, C., et al. (2020). Sensitivity and performance of the advanced LIGO detectors in the third observing run. *Phys. Rev. D* 102, 062003. doi: 10.1103/PhysRevD.102.062003
- Cannon, K., Cariou, R., Chapman, A., Crispin-Ortuzar, M., Fotopoulos, N., Frei, M., et al. (2012). Toward early-warning detection of gravitational waves from compact binary coalescence. *Apj* 748, 136. doi: 10.1088/0004-637X/748/2/136
- Chan, M. L., Heng, I. S., and Messenger, C. (2020). Detection and classification of supernova gravitational wave signals: a deep learning approach. *Phys. Rev. D* 102, 043022. doi: 10.1103/PhysRevD.102.043022



- Chang, S.-W., Onken, C. A., Wolf, C., Luvaul, L., Möller, A., Scalzo, R., et al. (2021). SkyMapper optical follow-up of gravitational wave triggers: alert science data pipeline and LIGO/Virgo O3 run. *PASA* 38:e024. doi: 10.1017/pasa.2021.17
- Chatterjee, C., Wen, L., Diakogiannis, F., and Vinsen, K. (2021). Extraction of binary black hole gravitational wave signals from detector data using deep learning. *Phys. Rev. D* 104, 064046. doi: 10.1103/PhysRevD.104.064046
- Chatterjee, D., Ghosh, S., Brady, P. R., Kapadia, S. J., Miller, A. L., Nissanke, S., et al. (2020). A machine learning-based source property inference for compact binary mergers. *Apj* 896, 54. doi: 10.3847/1538-4357/ab8dbe
- Chollet, F., et al. (2015). *Keras*. Available online at: [https://keras.io/getting\\_started/faq/#how-should-i-cite-keras](https://keras.io/getting_started/faq/#how-should-i-cite-keras)
- Chu, Q., Kovalam, M., Wen, L., Slaven-Blair, T., Bosveld, J., Chen, Y., et al. (2020). The SPIIR online coherent pipeline to search for gravitational waves from compact binary coalescences. *arXiv e-prints*, arXiv:2011.06787.
- Chua, A. J. K., and Vallisneri, M. (2020). Learning bayesian posteriors with neural networks for gravitational-wave inference. *Phys. Rev. Lett.* 124, 041102. doi: 10.1103/PhysRevLett.124.041102
- Colgan, R. E., Corley, K. R., Lau, Y., Bartos, I., Wright, J. N., Márka, Z., et al. (2020). Efficient gravitational-wave glitch identification from environmental data through machine learning. *Phys. Rev. D* 101, 102003. doi: 10.1103/PhysRevD.101.102003
- Cuoco, E., Powell, J., Cavaglia, M., Ackley, K., Beijer, M., Chatterjee, C., et al. (2020). Enhancing gravitational-wave science with machine learning. *arXiv e-prints*, arXiv:2005.03745. doi: 10.1088/2632-2153/abb93a
- Davis, D., Massinger, T., Lundgren, A., Driggers, J. C., Urban, A. L., and Nuttall, L. (2019). Improving the sensitivity of advanced LIGO using noise subtraction. *Class. Quant. Gravity* 36, 055011. doi: 10.1088/1361-6382/ab01c5
- D'Emilio, V., Green, R., and Raymond, V. (2021). Density estimation with Gaussian processes for gravitational wave posteriors. *Mon. Notices R. Astron. Soc.* 508, 2090–2097. doi: 10.1093/mnras/stab2623
- Dooley, K. L., Barsotti, L., Adhikari, R. X., Evans, M., Fricke, T. T., Fritschel, P., et al. (2013). Angular control of optical cavities in a radiation-pressure-dominated regime: the enhanced LIGO case. *J. Opt. Soc. Am. A* 30, 2618. doi: 10.1364/JOSAA.30.002618
- Dreissigacker, C., and Prix, R. (2020). Deep-learning continuous gravitational waves: multiple detectors and realistic noise. *Phys. Rev. D* 102, 022005. doi: 10.1103/PhysRevD.102.022005
- Driggers, J. C., Vitale, S., Lundgren, A. P., Evans, M., Kawabe, K., Dwyer, S. E., et al. (2019). Improving astrophysical parameter estimation via offline noise subtraction for Advanced LIGO. *Phys. Rev. D* 99, 042001. doi: 10.1103/PhysRevD.99.042001
- Essick, R., Godwin, P., Hanna, C., Blackburn, L., and Katsavounidis, E. (2020). iDQ: statistical inference of non-gaussian noise with auxiliary degrees of freedom in gravitational-wave detectors. *arXiv e-prints*, arXiv:2005.12761. doi: 10.1088/2632-2153/abab5f
- Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L., and Muller, P.-A. (2019). Deep learning for time series classification: a review. *Data Min. Knowl. Discov.* 33, 917–963. doi: 10.1007/s10618-019-00619-1
- Gabbard, H., Messenger, C., Heng, I. S., Tonolini, F., and Murray-Smith, R. (2019). Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy. *arXiv e-prints*, arXiv:1909.06296.
- George, D., and Huerta, E. A. (2018a). Deep learning for real-time gravitational wave detection and parameter estimation: results with advanced LIGO data. *Phys. Lett. B* 778, 64–70. doi: 10.1016/j.physletb.2017.12.053
- George, D., and Huerta, E. A. (2018b). Deep neural networks to enable real-time multimessenger astrophysics. *Phys. Rev. D* 97, 044039. doi: 10.1103/PhysRevD.97.044039
- Graff, P. B., Buonanno, A., and Sathyaprakash, B. S. (2015). Missing Link: bayesian detection and measurement of intermediate-mass black-hole binaries. *Phys. Rev. D* 92, 022002. doi: 10.1103/PhysRevD.92.022002
- Green, S. R., Simpson, C., and Gair, J. (2020). Gravitational-wave parameter estimation with autoregressive neural network flows. *Phys. Rev. D* 102, 104057. doi: 10.1103/PhysRevD.102.104057
- Hirose, E., Kawabe, K., Sigg, D., Adhikari, R., and Saulson, P. R. (2010). Angular instability due to radiation pressure in the LIGO gravitational-wave detector. *Appl. Opt.* 49:3474. doi: 10.1364/AO.49.003474
- Huerta, E. A., Khan, A., Huang, X., Tian, M., Levental, M., Chard, R., et al. (2020). Confluence of artificial intelligence and high performance computing for accelerated, scalable and reproducible gravitational wave detection. *arXiv e-prints*, arXiv:2012.08545. doi: 10.21203/rs.3.rs-138409/v1
- Ilesanmi, A. E., and Ilesanmi, T. O. (2021). Methods for image denoising using convolutional neural network: a review. *Complex Intell. Syst.* 7, 1–20. doi: 10.1007/s40747-021-00428-4
- Kagra Collaboration (2019). KAGRA: 2.5 generation interferometric gravitational wave detector. *Nat. Astron.* 3, 35–40. doi: 10.1038/s41550-018-0658-y
- Krastev, P. G. (2020). Real-time detection of gravitational waves from binary neutron stars using artificial neural networks. *Phys. Lett. B* 803:135330. doi: 10.1016/j.physletb.2020.135330
- Krastev, P. G., Gill, K., Villar, V. A., and Berger, E. (2021). Detection and parameter estimation of gravitational waves from binary neutron-star mergers in real LIGO data using deep learning. *Phys. Lett. B* 815, 136161. doi: 10.1016/j.physletb.2021.136161
- Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature.14539
- LIGO Scientific Collaboration and Virgo Collaboration (2020). GW190521: a binary black hole merger with a total mass of 150  $M_{\odot}$ . *Phys. Rev. Lett.* 125, 101102. doi: 10.1103/PhysRevLett.125.101102
- LIGO Scientific Collaboration. (2015). Advanced LIGO. *Class. Quant. Gravity* 32, 074001. doi: 10.1088/0264-9381/32/7/074001
- Lim, B., and Zohren, S. (2021). Time-series forecasting with deep learning: a survey. *Philos. Trans. R. Soc. A* 379, 20200209. doi: 10.1098/rsta.2020.0209
- López, M., Di Palma, I., Drago, M., Cerdá-Durán, P., and Ricci, F. (2021). Deep learning for core-collapse supernova detection. *Phys. Rev. D* 103, 063011. doi: 10.1103/PhysRevD.103.063011
- Mandel, I., Brown, D. A., Gair, J. R., and Miller, M. C. (2008). Rates and characteristics of intermediate mass ratio inspirals detectable by advanced ligo. *Apj* 681, 1431–1447. doi: 10.1086/588246
- Marianer, T., Poznanski, D., and Prochaska, J. X. (2021). A semisupervised machine learning search for never-seen gravitational-wave sources. *Mon. Notices R. Astron. Soc.* 500, 5408–5419. doi: 10.1093/mnras/staa3550
- Martynov, D. V., Hall, E. D., Abbott, B. P., Abbott, R., Abbott, T. D., Adams, C., et al. (2016). Sensitivity of the Advanced LIGO detectors at the beginning of gravitational wave astronomy. *Phys. Rev. D* 93, 112004. doi: 10.1103/PhysRevD.93.112004
- Mishra, T., O'Brien, B., Gayathri, V., Szczepańczyk, M., Bhaumik, S., Bartos, I., et al. (2021). Optimization of model independent gravitational wave search for binary black hole mergers using machine learning. *Phys. Rev. D* 104, 023014. doi: 10.1103/PhysRevD.104.023014
- Mogushi, K. (2021). Reduction of transient noise artifacts in gravitational-wave data using deep learning. *arXiv preprint arXiv:2105.10522*.
- Mogushi, K., Quitzow-James, R., Cavaglia, M., Kulkarni, S., and Hayes, F. (2021). NNETFIX: An artificial neural network-based denoising engine for gravitational-wave signals. *arXiv e-prints* arXiv:2101.04712. doi: 10.1088/2632-2153/abea69
- Ormiston, R., Nguyen, T., Coughlin, M., Adhikari, R. X., and Katsavounidis, E. (2020). Noise reduction in gravitational-wave data via deep learning. *Phys. Rev. Res.* 2, 033066. doi: 10.1103/PhysRevResearch.2.033066
- Pavlyshenko, B. M. (2019). Machine-learning models for sales time series forecasting. *Data* 4, 15. doi: 10.3390/data4010015
- Pravin, C., and Ojha, V. (2020). “A novel ecg signal denoising filter selection algorithm based on conventional neural networks,” in *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)* (Miami, FL: IEEE), 1094–1100.
- Romero-Shaw, I. M., Lasky, P. D., and Thrane, E. (2019). Searching for eccentricity: signatures of dynamical formation in the first gravitational-wave transient catalogue of LIGO and Virgo. *Mon. Notices R. Astron. Soc.* 490, 5210–5216. doi: 10.1093/mnras/stz2996
- Sachdev, S., Magee, R., Hanna, C., Cannon, K., Singer, L., SK, J. R., et al. (2020). An early-warning system for electromagnetic follow-up of gravitational-wave events. *Apj* 905, L25. doi: 10.3847/2041-8213/abc753
- Saiz-Pérez, A., Torres-Forné, A., and Font, J. A. (2021). Classification of the core-collapse supernova explosion mechanism with learned dictionaries. *arXiv e-prints*, arXiv:2110.12941.
- Sankarapandian, S., and Kulis, B. (2021).  $\beta$ -Annealed variational autoencoder for glitches. *arXiv e-prints*, arXiv:2107.10667.



- Schäfer, M. B., Ohme, F., and Nitz, A. H. (2020). Detection of gravitational-wave signals from binary neutron star mergers using machine learning. *Phys. Rev. D* 102, 063015. doi: 10.1103/PhysRevD.102.063015
- Sidles, J. A., and Sigg, D. (2006). Optical torques in suspended fabry perot interferometers. *Phys. Lett. A* 354, 167–172. doi: 10.1016/j.physleta.2006.01.051
- Soni, S., Berry, C. P. L., Coughlin, S. B., Harandi, M., Jackson, C. B., Crowston, K., et al. (2021). Discovering features in gravitational-wave data through detector characterization, citizen science and machine learning. *Class. Quant. Gravity* 38, 195016. doi: 10.1088/1361-6382/ac1ccb
- Talbot, C., and Thrane, E. (2020). Fast, flexible, and accurate evaluation of gravitational-wave Malmquist bias with machine learning. *arXiv e-prints*, arXiv:2012.01317.
- Torres-Forné, A., Cuoco, E., Font, J. A., and Marquina, A. (2020). Application of dictionary learning to denoise LIGO's blip noise transients. *Phys. Rev. D* 102, 023011. doi: 10.1103/PhysRevD.102.023011
- Tse, M., and et al. (2019). Quantum-enhanced Advanced LIGO detectors in the era of gravitational-wave astronomy. *Phys. Rev. Lett.* 123, 231107. doi: 10.1103/PhysRevLett.123.231107
- Vajente, G., Huang, Y., Isi, M., Driggers, J. C., Kissel, J. S., Szczepańczyk, M. J., et al. (2020). Machine-learning nonstationary noise out of gravitational-wave detectors. *Phys. Rev. D* 101, 042003. doi: 10.1103/PhysRevD.101.042003
- Veitch, J., Pürrer, M., and Mandel, I. (2015). Measuring intermediate-mass black-hole binaries with advanced gravitational wave detectors. *Phys. Rev. Lett.* 115, 141101. doi: 10.1103/PhysRevLett.115.141101
- Wei, W., and Huerta, E. A. (2021). Deep learning for gravitational wave forecasting of neutron star mergers. *Phys. Lett. B* 816, 136185. doi: 10.1016/j.physletb.2021.136185
- Williams, M. J., Veitch, J., and Messenger, C. (2021). Nested sampling with normalizing flows for gravitational-wave inference. *Phys. Rev. D* 103, 103006. doi: 10.1103/PhysRevD.103.103006
- Wong, K. W. K., Ng, K. K. Y., and Berti, E. (2020). Gravitational-wave signal-to-noise interpolation via neural networks. *arXiv e-prints*, arXiv:2007.10350.
- Xia, H., Shao, L., Zhao, J., and Cao, Z. (2021). Improved deep learning techniques in gravitational-wave data analysis. *Phys. Rev. D* 103, 024040. doi: 10.1103/PhysRevD.103.024040
- Yan, J., Avagyan, M., Colgan, R. E., Veske, D., Bartos, I., Wright, J., et al. (2021). Generalized approach to matched filtering using neural networks. *arXiv e-prints* arXiv:2104.03961. doi: 10.1103/PhysRevD.105.043006
- Yu, H. (2019). *Astrophysical signatures of neutron stars in compact binaries and experimental improvements on gravitational-wave detectors* (Ph.D. thesis). MIT.
- Yu, H., and Adhikari, R. (2021). Subtracting bilinear noise using machine-learning neural networks. *LIGO Document G2100738*.
- Yu, H., Adhikari, R. X., Magee, R., Sachdev, S., and Chen, Y. (2021). Early warning of coalescing neutron-star and neutron-star-black-hole binaries from the nonstationary noise background using neural networks. *Phys. Rev. D* 104, 062004. doi: 10.1103/PhysRevD.104.062004
- Yu, H., Martynov, D., Vitale, S., Evans, M., Shoemaker, D., Barr, B., et al. (2018). Prospects for detecting gravitational waves at 5 hz with ground-based detectors. *Phys. Rev. Lett.* 120, 141102. doi: 10.1103/PhysRevLett.120.141102
- Zhan, C., Jia, M., Ma, C., Lu, Z., and Lin, W. (2021). The response of the convolutional neural network to the transient noise in gravitational wave detection. *arXiv e-prints*, arXiv:2103.03557.
- Zhou, H., Feng, X., Dong, Z., Liu, C., and Liang, W. (2021). Application of denoising cnn for noise suppression and weak signal extraction of lunar penetrating radar data. *Remote Sens.* 13, 779. doi: 10.3390/rs13040779

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Yu and Adhikari. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## A. DETAILS OF NOISE SIMULATION

In this Appendix, we supplement details of our simulation of the angular noise presented in Section 2.

Note that Equation (1) describes the local length fluctuation of a single mirror [and hence a superscript “(mir)” is added to each quantity]. A physical interest is the length fluctuation of an ARM cavity formed by an input test mass and an end test mass. In other words, we want to measure the relative distance between the two test masses (which are also mirrors) as the GW signal is readout *via* the differential-ARM (DARM) length. Therefore, the angular noise affects the cavity length as

$$\delta x(t) = \begin{bmatrix} x_{\text{spot}}^{(e)}(t) & x_{\text{spot}}^{(i)}(t) \end{bmatrix} \begin{bmatrix} \theta^{(e)}(t) \\ \theta^{(i)}(t) \end{bmatrix}, \quad (\text{A1})$$

where the superscript “(e)” [“(i)”] stands for the end (input) test mass. Its contamination to DARM is further obtained by first incoherently simulating each cavity’s length according to Equation (A1) and then taking the difference.

As we described in Section 2, the real angular control is performed on a radiation pressure basis (Sidles and Sigg, 2006; Yu, 2019), which can be understood as the following. Mathematically, we note that the radiation torque depends on the spot motions on the test masses, which further relate to the alignment of the test masses *via* the cavity geometry. For pitch motion (yaw motion is similar), this is given by Sidles and Sigg (2006), Barsotti et al. (2010), and Yu (2019).

$$\begin{bmatrix} x_{\text{spot}}^{(e)} \\ x_{\text{spot}}^{(i)} \end{bmatrix} = \frac{L}{g^{(e)}g^{(i)} - 1} \begin{bmatrix} g^{(i)} & 1 \\ 1 & g^{(e)} \end{bmatrix} \begin{bmatrix} \theta^{(e)} \\ \theta^{(i)} \end{bmatrix} \equiv \mathbf{M}_{\text{a2s}} \begin{bmatrix} \theta^{(e)} \\ \theta^{(i)} \end{bmatrix}, \quad (\text{A2})$$

where  $L$  is the length of the ARM cavity, and  $g^{(e)} = 1 - L/R^{(e)}$  [and similarly  $g^{(i)} = 1 - L/R^{(i)}$ ] with  $R^{(e)}$  [ $R^{(i)}$ ] the radius of curvature of the end (input) test mass. For aLIGO, we have  $L = 3,995$  m,  $g^{(e)} = -0.779$ , and  $g^{(i)} = -1.065$  (Barsotti et al., 2010). The hard (soft) mode then corresponds to the eigenvector of the  $\mathbf{M}_{\text{a2s}}$  matrix associated with a positive (negative) eigenvalue. The conversion between the mirror basis ( $[\theta^{(e)}, \theta^{(i)}]^T$ ) and the radiation-torque basis ( $[\theta^{(h)}, \theta^{(s)}]^T$ ) is given by

$$\begin{bmatrix} \theta^{(h)} \\ \theta^{(s)} \end{bmatrix} = \mathbf{M}_{\text{m2r}} \begin{bmatrix} \theta^{(e)} \\ \theta^{(i)} \end{bmatrix} = \begin{bmatrix} 0.756 & -0.655 \\ 0.655 & 0.756 \end{bmatrix} \begin{bmatrix} \theta^{(e)} \\ \theta^{(i)} \end{bmatrix}, \quad (\text{A3})$$

where in the second equality, we have plugged in numerical values for the aLIGO interferometer. In our simulation, we set  $\theta^{(s)} = 0$  and simulate only  $\theta^{(h)}$  as only the hard mode has a significant control bandwidth and feeds back its sensing noise in the 10–30 Hz band. Above the UGF, the physical  $\theta^{(h)}$  is the product between a white sensing noise with a typical ASD of  $\sim 5 \times 10^{-14}$  rad Hz $^{-1/2}$  and the open-loop gain of the angular control loop (Martynov et al., 2016; Yu, 2019; Buikema et al., 2020). Once we have  $\theta^{(h)}$ , it can then be converted back to the mirror basis by inverting Equation (A3).

To sense the low-frequency spot position motion  $x_{\text{spot}}(t)$ , we simulate two sets of sensors similar to the realistic aLIGO system. One set corresponds to the ADS sensors. Specifically, we intentionally excite each mirror in angle at a known frequency and then demodulate the length readout (i.e., DARM output) at the same frequency. From Equation (1) we see that the demodulated signal is directly proportional to the spot motion on the test mass. However, this technique has a limited SNR because only a weak excitation is allowed in order to avoid saturation of actuators and sensors as well as up-conversion of the low-frequency noise into the sensitivity ( $> 10$  Hz) band. Consequently, the ADS sensors are sensitive to only the very low frequency ( $< 0.1$  Hz) portion of the spot motion.

To probe the spot position in the 0.1–3 Hz band, we employ the OPLEV sensors. An OPLEV probes a mirror’s alignment locally, which then allows us to infer the spot motion using Equation (A2). However, the reference point drifts slowly, and thus, an OPLEV can only probe spot motion in the  $\gtrsim 0.1$  Hz band. One of the main challenges to the CNN in our study is, thus, to recognize the frequency-dependent blending of the ADS and OPLEV sensors according to the right panel of **Figure 2**.

Finally, we point out that the channels we simulate are in fact based on real LIGO auxiliary channels. The fast angular motion  $\theta^{(h)}(t)$  are contained in channels like L1:ASC-CHARD\_P\_IN1\_DQ. The slow channels are designed to mimic auxiliary channels like L1:ASC-ADS\_PIT4\_DOFF\_OUT\_DQ for the ADS sensors, and L1:SUS-ITMX\_L3\_OPLEV\_PIT\_OUT\_DQ for the OPLEVs.