

Article

Physics-Guided Dynamic Sparse Attention Network for Gravitational Wave Detection Across Ground and Space-Based Observatories

Tiancong Zhang ^{1,2}  and Wei Bian ^{1,2,*}

¹ Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences (UCAS), Hangzhou 310024, China; zhangtiancong23@mailsucas.ac.cn

² University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: bianwei@ucas.ac.cn

Abstract

Ground-based and space-based gravitational wave (GW) detectors cover complementary frequency bands, laying the foundation for future multi-band collaborative observations. Detecting weak signals within non-stationary noise remains challenging. To address this, we propose a Physics-Guided Dynamic Sparse Attention (PGDSA) framework. The framework introduces a differentiable wavelet layer to explicitly embed sensitive frequency bands and time–frequency priors while utilizing intra-block Top-K sparse attention for efficient long-range temporal modeling. Training is performed on space-based simulation data with joint optimization for signal detection and waveform reconstruction. We then evaluate detection performance and zero-shot transfer capability on ground-based data. Experimental results show that PGDSA achieves an ROC-AUC of 0.886 on the Kaggle G2Net private leaderboard. On GWOSC O3 real data, the model yields high confidence scores for confirmed binary black hole events. On LISA simulation data, the framework achieves detection rates exceeding 99% for multiple signal types (SNR = 50, FAR = 1%) with waveform reconstruction Overlap comparable to baseline methods. These results demonstrate that PGDSA enables unified modeling across both space-based and ground-based scenarios.

Keywords: gravitational wave detection; sparse attention; differentiable wavelet transform; multi-task learning; cross-platform transfer; LISA



Academic Editors: Phivos Mylonas, Katia Lida Kermanidis and Manolis Maragoudakis

Received: 29 December 2025

Revised: 8 February 2026

Accepted: 11 February 2026

Published: 15 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

1.1. Research Background and Significance

Gravitational waves (GWs) are ripples in spacetime predicted by Einstein’s general theory of relativity. On 14 September 2015, the Laser Interferometer Gravitational-Wave Observatory (LIGO) achieved the first direct detection of gravitational waves produced by a binary black hole (BBH) merger, GW150914, marking the birth of gravitational wave astronomy [1]. Since then, the ground-based detector network led by LIGO, Virgo, and KAGRA has reported a large number of compact binary coalescence (CBC) events during the O1–O3 observing runs. The GWTC-3 catalog contains 90 events together with their corresponding parameter estimation results [2]. Meanwhile, the Gravitational Wave Open Science Center (GWOSC) has continuously released observational data, significantly lowering the barrier for method reproducibility and cross-method comparisons [3]. With the commencement

and ongoing operation of the fourth observing run (O4), both the scale of event candidates and the demand for low-latency analyses have further increased, imposing higher requirements on scalable and reproducible automated analysis methods [4].

Unlike ground-based detectors, which are primarily sensitive to frequencies operating from tens of hertz to kilohertz, space-based gravitational wave detectors target low-frequency signals in the millihertz (mHz) band. Missions such as the Laser Interferometer Space Antenna (LISA) and China's Taiji and TianQin projects are expected to be launched or deployed in the 2030s, focusing on sources including massive black hole binaries (MBHBs), extreme-mass-ratio inspirals (EMRIs), galactic binary white dwarfs (BWDs), and the stochastic gravitational wave background (SGWB) [5,6]. Compared to the short-duration (second-scale) merger signals observed by ground-based detectors, space-based data are characterized by a combination of long-duration signals, weak amplitudes, and complex noise properties. Signals may persist for months or even years, resulting in massive time series data volumes, while the noise often exhibits non-stationary and non-Gaussian features [7].

Furthermore, ground-based and space-based detectors are naturally complementary in terms of frequency coverage and observational phases. Space-based detectors can observe the low-frequency inspiral stage of certain sources, such as stellar-mass binary black holes, at earlier evolutionary phases, while ground-based detectors capture the subsequent merger and ringdown phases at higher frequencies. Such space-ground (multi-band) observations are expected to enhance early-warning capabilities, improve parameter estimation accuracy, and enable consistency tests of source physics, thereby promoting a unified understanding of source evolution across frequency bands rather than isolated single-band events [4,6]. From a methodological perspective, developing unified modeling frameworks that can accommodate the data characteristics of both space-based and ground-based detectors is therefore an important research direction in anticipation of future observational paradigms.

Within this context, gravitational wave data analysis faces two closely coupled core tasks; the first is reliable signal detection in the presence of strong noise backgrounds, and the second is waveform reconstruction (denoising or extraction) based on detected signals to support subsequent physical interpretation, such as parameter estimation and astrophysical inference [8,9]. Traditional methods, including template-based CBC searches (e.g., PyCBC and GstLAL) and unmodeled burst searches (e.g., cWB) [10–12], are grounded in rigorous physical principles but often incur high computational costs when dealing with long-duration signals, weak amplitudes, and complex noise environments. To address the distinct challenges of long time series in space-based detectors and multi-channel non-Gaussian noise in ground-based detectors, this work proposes a **Physics-Guided Dynamic Sparse Attention (PGDSA)** framework. The framework jointly optimizes detection and waveform reconstruction tasks on space-based simulated datasets where noise-free reference waveforms are available, and subsequently evaluates detection performance and cross-domain transferability on ground-based datasets (G2Net and GWOSC O3). This approach aims to provide methodological insights for gravitational wave data analysis in the era of space-ground joint observations.

1.2. Review of Existing Methods

Existing gravitational wave data analysis approaches can be broadly categorized into four classes [9]: template-based matched filtering methods, template-free Bayesian inference methods, deep learning-based methods (including CNNs, Transformers, and generative models), and hybrid approaches that combine physical priors with deep learning. Table 1 summarizes the representative methods along with their key characteristics and limitations.

Table 1. Comparison of existing gravitational wave detection and reconstruction methods.

Method	Algorithms	Tasks	Advantages	Limitations
Template matching	PyCBC/GstLAL	Detection + param. est.	Theoretically optimal, high sensitivity	High cost, template dependence
Bayesian	BayesWave/cWB	Detection + reconstruction	Template-free, flexible	Expensive, short signals
CNN-based	ResNet/Inception	Detection	Mature, low cost	Limited interpretability
Transformer	Zhao et al. [13]	Detection + reconstruction	Long-range modeling	$O(N^2)$ complexity
Generative	WaveNet/cVAE	Extraction	High-quality reconstruction	Not jointly optimized
Physics hybrid	Wavelet + CNN	Detection	Time–freq priors	Simple fusion
This work	PGDSA	Detection + reconstr.	Physics-guided, multi-task, unified	–

1.2.1. Template-Based and Bayesian Methods

Matched filtering is the standard technique for gravitational wave detection, achieving an optimal SNR by cross-correlating data with precomputed waveform templates [14]. PyCBC and GstLAL have been successfully deployed in official LIGO/Virgo pipelines [10,11]. However, computational costs grow exponentially with parameter space dimensionality. Template-free methods such as cWB [12] and BayesWave [15] address this limitation but remain computationally intensive for long-duration signals.

1.2.2. Machine Learning Approaches

Deep learning has achieved notable developments in gravitational wave detection, including CNNs [16,17], Bayesian neural networks [18], random convolutional kernel methods for space-based signals [19], and attention-based models [9]. Zhao et al. proposed a unified framework for LISA sources achieving joint detection and waveform reconstruction [13]. More recently, Transformer-based architectures have been applied to gravitational wave data denoising [20] and waveform generation [21], while score-based diffusion models have shown promise for parameter estimation under non-Gaussian noise conditions [22]. However, these recent advances primarily target individual tasks (denoising, generation, or parameter estimation) rather than unified detection and reconstruction. More broadly, existing methods face challenges in physical interpretability, cross-distribution generalization, and long-sequence modeling ($O(N^2)$ complexity for standard attention). Most approaches are optimized for a single detector platform, leaving unified cross-platform modeling largely unexplored.

1.2.3. Hybrid Methods

Physics-inspired approaches combine time–frequency analysis with deep learning to balance interpretability and efficiency [9,23–25]. However, physical priors are often incorporated only at the preprocessing level rather than systematically embedded within architectures, and fusion strategies typically rely on simple concatenation without adaptive weighting.

1.3. Limitations and Challenges

Based on the above review, four core challenges can be identified: (1) **Physical interpretability vs. data-driven learning**: Existing methods lack unified frameworks that systematically integrate physical priors (e.g., time–frequency structure, detector sensitivity

bands) with deep representation learning [9,26]; (2) **Computational efficiency**: Standard self-attention scales as $O(N^2)$, making it unsuitable for long time series in space-based settings [23,27,28]; (3) **Cross-platform adaptability**: Ground-based and space-based detectors differ substantially in frequency bands, signal types, and noise characteristics, yet unified architectures remain largely unexplored [6,7]; (4) **Multi-task optimization**: Joint detection and waveform reconstruction can provide synergistic gains, but stable task synergy across platforms still needs to be achieved [13].

1.4. Research Objectives and Main Contributions

This study proposes the PGDSA framework to address limitations in physical interpretability, cross-platform adaptability, and task fragmentation. The main contributions are (1) **differentiable wavelet transforms** as learnable time–frequency modules with detector band constraints; (2) **block-wise sparse attention** combining Top-K sparsification to reduce $O(N^2)$ complexity; (3) **gated cross-modal fusion** to adaptively integrate physics-guided and data-driven features; (4) **multi-task learning** for joint detection and waveform reconstruction; (5) comprehensive evaluation on G2Net, GWOSC O3 (zero-shot), and LISA datasets, demonstrating cross-platform applicability.

2. Methods

2.1. Overall Architecture Design

This section provides a detailed description of the proposed PGDSA model, including its mathematical formulation and implementation details. As illustrated in Figure 1, the model adopts a **dual-stream encoding architecture** within a multi-task learning framework. It consists of a physics-inspired time–frequency branch and a neural network branch, a gated fusion mechanism, and two parallel output heads for **signal detection** and **waveform reconstruction (denoising/extraction)**.

The Architecture of Physics-Guided Dynamic Sparse Attention (PGDSA) Model

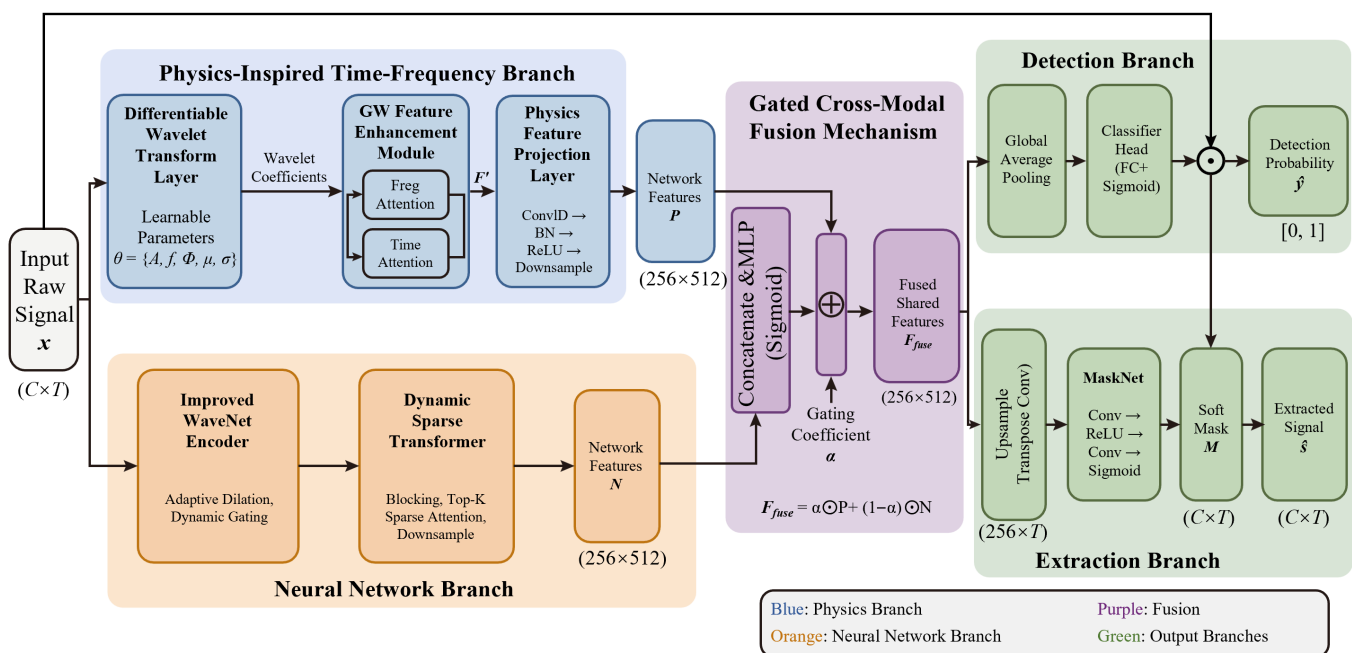


Figure 1. Overall architecture of PGDSA. The model contains a physics-inspired time–frequency branch (top) and a neural network branch (bottom). After gated fusion, it outputs the detection probability (detection head) and the reconstructed signal (waveform reconstruction/extraction head).

From a mathematical perspective, the architecture can be viewed as an end-to-end multi-task functional mapping. Let the input be $x \in \mathbb{R}^{C \times T}$, where C denotes the number of detector channels (for ground-based data, $C = 3$, corresponding to Hanford, Livingston, and Virgo; for space-based data, $C = 1$), and T denotes the number of time steps per channel. The model jointly outputs the detection probability $\hat{y}$ and the extracted signal $\hat{s}$:

$$F_{\text{fuse}} = \alpha \odot \Phi(F_p) + (1 - \alpha) \odot \Psi(F_n) \quad (1)$$

$$\hat{y} = \sigma(\text{Classifier}(\text{Pool}(F_{\text{fuse}}))) \quad (2)$$

$$\hat{s} = x \odot \sigma(\text{MaskNet}(\text{Upsample}(F_{\text{fuse}}))) \quad (3)$$

Here, $F_p = \Gamma(x)$ denotes features extracted by the physics-inspired time–frequency branch, and $F_n = \Omega(x)$ denotes features encoded by the neural network branch. Φ and Ψ are feature projection functions, α is an adaptive gating coefficient, $\odot$ denotes the Hadamard (element-wise) product, and σ is the Sigmoid activation. Pool denotes global average pooling, MaskNet denotes a mask generation network, and Upsample denotes an upsampling operation. This multi-task design enables the model to simultaneously perform signal detection and waveform reconstruction (denoising/extraction).

The following subsections describe the design principles and implementations of the key modules in detail.

2.2. Physics-Inspired Time–Frequency Branch

As illustrated in Figure 2, the physics-inspired time–frequency branch aims to exploit the time–frequency characteristics of gravitational wave signals and extract physically meaningful representations from raw data. In this work, “physics-inspired” refers to using the time–frequency structure of signals within the detector’s sensitive band as an inductive bias at the feature extraction stage. Specifically, we construct a general time–frequency analysis branch centered on a differentiable wavelet transform.

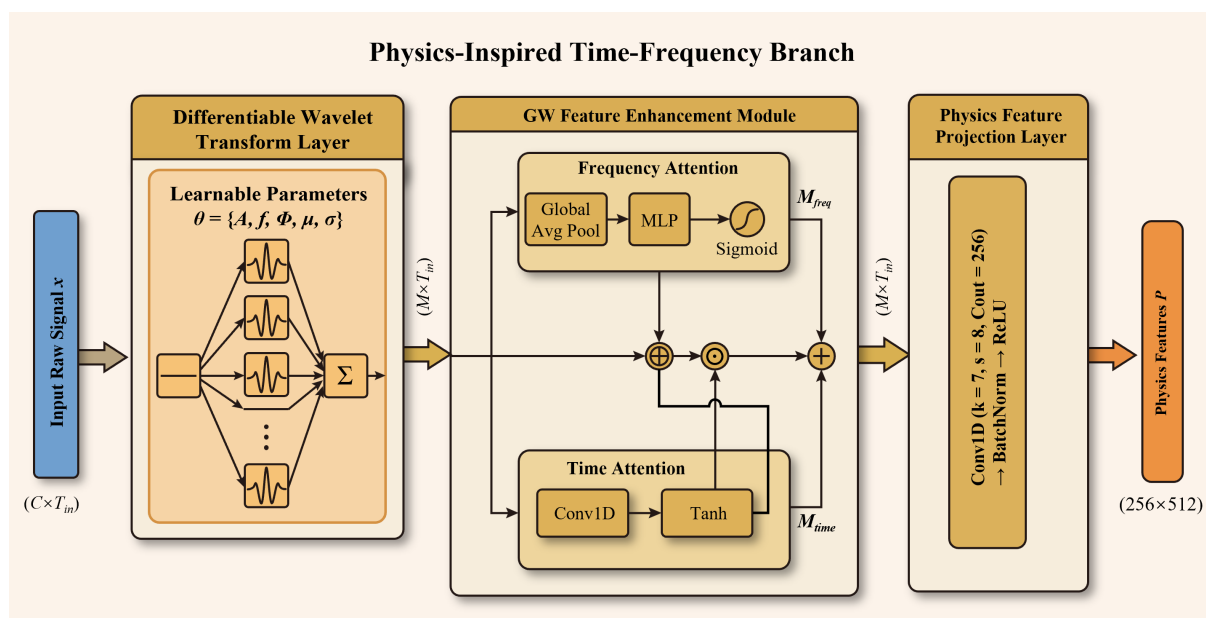


Figure 2. Schematic diagram of the physics-inspired time–frequency branch. The differentiable wavelet layer uses learnable Morlet kernels with detector band constraints (for G2Net, $f_\theta \in [20, 500]$ Hz; for LISA, the range is shifted to the mHz band). The enhanced time–frequency map is projected by a Conv1D ($k = 7$, stride $s = 8$, $C_{\text{out}} = 256$) to reduce the temporal length (e.g., $T = 4096 \rightarrow 512$) and form the physical feature tensor $P \in \mathbb{R}^{256 \times 512}$.

This branch contains three core modules: a differentiable wavelet transform layer, a gravitational wave feature enhancement module, and a physical feature projection layer. The wavelet analysis adapts to different signal types, tracking frequency evolution for chirp-like signals, acting as narrow-band filters for quasi-monochromatic BWD signals, and capturing power spectrum properties for the SGWB.

2.2.1. Differentiable Wavelet Transform Layer

We design a differentiable wavelet transform layer with learnable Morlet wavelet kernels [23,24]:

$$\psi_{\theta}(t) = A_{\theta} \cdot e^{j(2\pi f_{\theta}t + \phi_{\theta})} \cdot e^{-\frac{(t-\mu_{\theta})^2}{2\sigma_{\theta}^2}} \quad (4)$$

with learnable parameters $\theta = \{A, f, \phi, \mu, \sigma\}$ controlling amplitude, center frequency, phase, temporal center, and time–frequency resolution, respectively. The frequency parameter f_{θ} is constrained to match detector sensitivity bands: [20, 500] Hz for G2Net (ground-based) and $[10^{-4}, 0.05]$ Hz for LISA (space-based). This differentiable design enables end-to-end optimization, transforming physical priors into adaptive, task-dependent constraints.

2.2.2. Gravitational Wave Feature Enhancement Module

This module employs dual attention to amplify gravitational wave signatures while suppressing noise. For wavelet output $F \in \mathbb{R}^{M \times T}$, frequency-domain attention M_{freq} highlights important frequency components via an MLP with compression ratio $r = 4$, while time-domain attention M_{time} captures temporal correlations via Conv1D ($k = 5$). Enhanced features are computed as $F' = F \odot (M_{\text{freq}} \oplus M_{\text{time}}) + F$, where residual connections preserve original information.

2.2.3. Physical Feature Projection Layer

The physical feature projection layer compresses the enhanced time–frequency representation into a compact feature tensor to facilitate subsequent fusion with neural features. This layer is implemented using a one-dimensional convolution:

$$P = \text{ReLU}(\text{BatchNorm}(\text{Conv1D}(F', C_{\text{out}} = 256, k = 7, s = 8))) \quad (5)$$

where Conv1D uses kernel size 7, stride 8, and output channel number 256. BatchNorm denotes batch normalization, and ReLU is the activation function.

With this design, every eight time steps in the time–frequency map are compressed into one feature step, reducing the sequence length from $T = 4096$ to 512 while increasing the channel dimension to 256, yielding the physical feature representation $P \in \mathbb{R}^{256 \times 512}$.

2.3. Neural Network Branch

As shown in Figure 3, the neural network branch adopts a purely data-driven strategy to extract representations directly from the raw signals. It consists of two core components: an improved WaveNet encoder and a dynamic sparse Transformer.

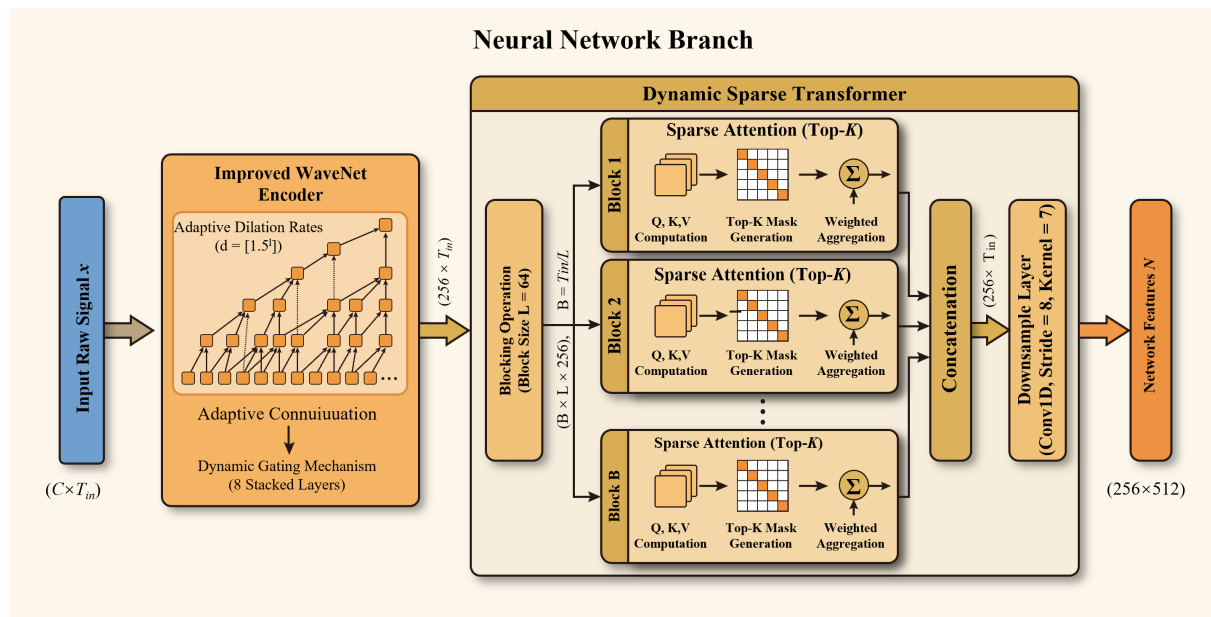


Figure 3. Schematic diagram of the neural network branch. The neural branch contains an improved WaveNet encoder (8 layers with adaptive dilation and dynamic gating, threshold $\tau = 0.6$) followed by a block sparse Transformer. The sparse attention is computed within blocks of length $L = 64$ and retains the Top-K entries ($\approx 30\%$ of L^2) per block. A downsampling step aligns the output to $N \in \mathbb{R}^{256 \times 512}$ for fusion with the physics branch.

2.3.1. Improved WaveNet Encoder

WaveNet was originally developed for audio generation, and its dilated convolutional structure is naturally suited for capturing multi-scale temporal patterns [29]. In this work, we introduce three key improvements over the standard WaveNet: **adaptive dilation rates**, a **dynamic gating mechanism**, and **learnable residual scaling**.

The dynamic sparse gating mechanism is defined as follows:

$$G(x) = \text{Sigmoid}(W_g x + b_g) \quad (6)$$

$$M = \mathbb{I}(G(x) > \tau) \quad (7)$$

where $\tau = 0.6$ is an empirical threshold and $\mathbb{I}(\cdot)$ denotes the indicator function. To avoid the non-differentiability introduced by binarization, we adopt a **Straight-Through Estimator (STE)** during backpropagation; hard gating with M is used in the forward pass, while the gradient is approximated by the derivative of $\text{Softplus}(G(x) - \tau)$ as a smooth surrogate, enabling stable optimization of the gating parameters.

A single layer of the improved WaveNet is defined as follows:

$$z = \tanh(W_{f,k} *_{d_l} x) \odot \sigma(W_{g,k} *_{d_l} x) \quad (8)$$

$$y = x + \alpha_l \cdot (M \odot z) \quad (9)$$

where $*_{d_l}$ denotes a dilated convolution with dilation rate d_l , $W_{f,k}$ and $W_{g,k}$ are the convolution kernels for the filter and gate, respectively, α_l is a learnable residual scaling factor, and M is the dynamic sparse mask.

The dilation rates are designed adaptively as follows:

$$d_l = \lfloor 1.5^l \rfloor, \quad l = 1, 2, \dots, 8 \quad (10)$$

This design yields receptive fields ranging from a few time steps to several hundred time steps, enabling the encoder to capture both short-range and long-range temporal dependencies.

The complete improved WaveNet encoder consists of eight cascaded layers. Each detector channel time series $x_c \in \mathbb{R}^T$ is encoded separately, producing an intermediate feature representation $W_c \in \mathbb{R}^{256 \times T}$. The per-channel representations are then aggregated via element-wise summation to form the unified representation $W \in \mathbb{R}^{256 \times T}$.

2.3.2. Dynamic Sparse Transformer

The self-attention mechanism in standard Transformers has a computational complexity of $O(N^2)$, which becomes prohibitively expensive for long sequences. We therefore adopt a **dynamic sparse Transformer** that reduces computational cost via an adaptive sparsification strategy. The central idea is to process the input sequence in blocks and apply sparse attention within each block. The sparse attention computation is given by the following:

$$X_b = \text{Partition}(X, L = 64) \quad (11)$$

$$Q_b, K_b, V_b = X_b W_Q, X_b W_K, X_b W_V \quad (12)$$

$$S_b = \frac{Q_b K_b^T}{\sqrt{d}} + B \in \mathbb{R}^{L \times L} \quad (13)$$

$$M_b = \text{TopK}(S_b, k = 0.3L^2) \quad (14)$$

$$A_b = \text{Softmax}(S_b + (1 - M_b) \cdot (-10^9)) \quad (15)$$

$$O_b = A_b V_b \quad (16)$$

where the Partition operation splits the sequence into T/L blocks, each of length L . W_Q , W_K , and W_V are learnable projection matrices. B is a relative positional bias matrix, defined as $B_{ij} = \frac{|i-j|}{L} \cdot w_b$. TopK performs element-wise global Top-K selection within each block on the attention score matrix S_b , where $K = \lfloor 0.3 \times L^2 \rfloor$ is a fixed sparsity ratio retaining approximately 30% of the elements. M_b is the resulting sparse mask, and O_b is the output feature.

The dynamic sparse Transformer first computes sparse self-attention within each block, then concatenates block-level outputs and applies a one-dimensional convolutional downsampling step to align the temporal dimension with that of the physics branch. A detailed derivation of the time and space complexity is provided in Section 4.1.1.

The dynamic sparse Transformer takes WaveNet features $W \in \mathbb{R}^{256 \times T}$ as input and outputs $N \in \mathbb{R}^{256 \times 512}$ after block sparse attention and downsampling.

2.4. Gated Cross-Modal Fusion

The gated fusion module adaptively balances physics-guided and data-driven features [30]. Unlike fixed additive or multiplicative fusion, we learn an input-dependent weighting coefficient α that dynamically adjusts branch contributions: a larger α for high-SNR signals consistent with physical templates, a smaller α for complex or anomalous conditions. This enables **synergistic gains** where the full model outperforms simple combinations of individual components.

Let the physical features be $P \in \mathbb{R}^{256 \times 512}$ and the neural features be $N \in \mathbb{R}^{256 \times 512}$. The fusion process consists of three steps: feature concatenation, gate coefficient estimation, and dynamic fusion.

First, during feature concatenation, the two feature maps are concatenated along the channel dimension:

$$F_{\text{concat}} = [P; N] \in \mathbb{R}^{512 \times 512} \quad (17)$$

Second, during gate coefficient estimation, an adaptive, position-wise fusion weight is generated using a multi-layer perceptron (MLP). For simplicity of notation, the concatenated features are flattened, transformed via a linear mapping, and reshaped back to match the shape of P and N :

$$\alpha = \sigma(\text{MLP}(F_{\text{concat}})) \in \mathbb{R}^{256 \times 512} \quad (18)$$

where the MLP contains a single hidden layer and σ denotes the Sigmoid activation. The gating coefficient α has the same dimensionality as P and N , enabling **position-wise adaptive fusion** in which different time steps and feature channels can have distinct fusion weights.

Finally, dynamic fusion is performed via element-wise weighted combination:

$$F_{\text{fuse}} = \alpha \odot P + (1 - \alpha) \odot N \in \mathbb{R}^{256 \times 512} \quad (19)$$

where $\odot$ denotes the Hadamard (element-wise) product. This design allows the model to flexibly adjust the relative contributions of the physics and neural branches across feature positions.

With this gating mechanism, the model assigns higher weights to the physics branch when encountering high-SNR samples whose morphologies closely match physical templates, and relies more on data-driven features when facing complex noise characteristics or unexpected waveform variations.

From the perspective of signal type adaptability, the gating mechanism plays an additional critical role. Different gravitational wave source classes exhibit different degrees of compatibility with physical priors. For chirp-like signals such as BBH, MBHB, and EMRI, the physics branch effectively captures time–frequency evolution patterns, and α typically takes larger values. For quasi-monochromatic BWD signals or the SGWB, the wavelet-based physics branch primarily acts as a band selection filter, contributing less directly, while the neural network branch complements it by learning statistical signal properties. This adaptive, signal type-aware behavior is one of the key mechanisms enabling PGDSA to handle multiple gravitational wave sources within a unified framework.

2.5. Multi-Task Output Modules

To jointly optimize signal detection and waveform reconstruction (denoising/extraction), the PGDSA model adopts a multi-task learning framework with two parallel output heads: a detection head and a waveform reconstruction (extraction) head. This design is inspired by mask-based signal separation approaches in speech enhancement, where shared representations enable synergistic learning across tasks.

2.5.1. Detection Head

The detection head maps the fused features to a gravitational wave event detection probability through feature aggregation and classification. During feature aggregation, global average pooling is applied along the temporal dimension:

$$f = \frac{1}{T'} \sum_{t=1}^{T'} F_{\text{fuse}}[:, t] \in \mathbb{R}^{256} \quad (20)$$

where $T' = 512$ is the temporal length of the fused feature representation.

During classification, a fully connected layer followed by a Sigmoid activation outputs the detection probability:

$$\hat{y} = \sigma(W_c \cdot f + b_c) \in [0, 1] \quad (21)$$

where $W_c \in \mathbb{R}^{1 \times 256}$ is the classification weight matrix, b_c is the bias term, and σ denotes the Sigmoid function.

2.5.2. Waveform Reconstruction (Extraction) Head

The waveform reconstruction (extraction) head aims to recover gravitational wave signals from noisy observations using a **mask-based approach**. The core idea is to learn a mask matrix with the same dimensionality as the input, which separates the signal and noise via element-wise multiplication.

First, the fused features are upsampled to recover the original temporal resolution:

$$F_{\text{up}} = \text{Upsample}(F_{\text{fuse}}) \in \mathbb{R}^{256 \times T} \quad (22)$$

where the Upsample operation is implemented using transposed convolutions, restoring the temporal dimension from T' to T . Next, a mask generation network (MaskNet) produces the mask matrix:

$$M = \sigma(\text{Conv1D}(\text{ReLU}(\text{Conv1D}(F_{\text{up}})))) \in \mathbb{R}^{C \times T} \quad (23)$$

where C is the number of input channels ($C = 3$ for ground-based data and $C = 1$ for space-based data). The Sigmoid activation constrains the mask values to the range $[0, 1]$.

Finally, the extracted signal is obtained via element-wise multiplication:

$$\hat{s} = x \odot M \quad (24)$$

where x denotes the original input signal and $\hat{s}$ is the estimated gravitational wave signal. This masking mechanism enables the model to adaptively retain or suppress signal components across time and frequency, thereby achieving effective noise suppression and signal recovery.

2.6. Multi-Task Loss Function

PGDSA adopts a multi-task learning strategy to jointly optimize the objectives of detection and extraction. The overall loss is defined as follows:

$$\mathcal{L} = \mathcal{L}_{\text{detect}} + \lambda \mathcal{L}_{\text{extract}} \quad (25)$$

where λ is a task-balancing coefficient. In this study, we perform joint training for detection + waveform reconstruction (denoising/extraction) on space-based LISA simulations where noise-free ground-truth waveform templates (reference templates) are available, and set $\lambda = 1.0$. For the ground-based G2Net experiments and the real GWOSC O3 case studies, we evaluate detection only (the reconstruction loss is not computed during training, which can be viewed as $\lambda = 0$).

Detection loss. We use the binary cross-entropy loss

$$\mathcal{L}_{\text{detect}} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (26)$$

where y_i is the ground-truth label and $\hat{y}_i$ is the predicted probability.

Extraction loss. We adopt the scale-invariant signal-to-distortion ratio (SI-SDR), a standard objective in speech enhancement and source separation [13].

For signal-present samples ($y = 1$), the extraction loss is defined as follows:

$$\mathcal{L}_{\text{extract}}^{(\text{sig})} = -\text{SI-SDR}(\hat{s}, s) \quad (27)$$

where the SI-SDR is given by

$$\text{SI-SDR} = 10 \log_{10} \frac{\|s_{\text{target}}\|^2}{\|e_{\text{noise}}\|^2} \quad (28)$$

$$s_{\text{target}} = \frac{\langle \hat{s}, s \rangle}{\|s\|^2 + \epsilon} s, \quad e_{\text{noise}} = \hat{s} - s_{\text{target}} \quad (29)$$

Here, s denotes the target signal (the whitened gravitational wave template), $\hat{s}$ denotes the extracted signal produced by the model, and $\langle \cdot, \cdot \rangle$ denotes the inner product. By projecting $\hat{s}$ onto s , the SI-SDR removes scale ambiguity and focuses on waveform shape fidelity.

For noise-only samples ($y = 0$), the SI-SDR is mathematically undefined because $s = 0$. We therefore do not compute the SI-SDR, and instead impose an energy constraint that encourages the extracted output to be close to zero:

$$\mathcal{L}_{\text{extract}}^{(\text{noise})} = \|\hat{s}\|_2^2 \quad (30)$$

Accordingly, in the LISA multi-task training, we use the following unified formulation:

$$\mathcal{L}_{\text{extract}} = y \cdot \mathcal{L}_{\text{extract}}^{(\text{sig})} + (1 - y) \cdot \mathcal{L}_{\text{extract}}^{(\text{noise})} \quad (31)$$

2.7. Model Training and Optimization

We optimize the model using the Adam optimizer with an initial learning rate of 0.001 and a cosine annealing schedule. To mitigate overfitting, we incorporate regularization techniques including Dropout with a drop rate of 0.1, L2 weight decay with coefficient 1×10^{-5} , and a learning rate warmup strategy.

With the multi-task learning scheme, the detection and waveform reconstruction tasks share feature representations, enabling knowledge transfer and mutual benefits. On the space-based LISA dataset where reconstruction supervision is available, joint training improves detection performance compared to training the detection task alone while simultaneously providing a high-quality waveform reconstruction capability.

2.8. Key Dimensional Transformations and Temporal Processing Flow

The data flow in PGDSA consists of multiple processing stages with the following dimensional transformations: For a three-channel raw input signal of shape 3×4096 , the wavelet transform first extracts time–frequency features along the temporal dimension for each channel and aggregates them across channels to produce a combined time–frequency representation of shape 32×4096 . After feature enhancement, the time–frequency resolution remains unchanged, and the physical feature projection reduces it to 256×512 .

In the neural network branch, the WaveNet encoder maps the raw input into a feature representation of shape 256×4096 . After processing by the sparse Transformer, it is similarly reduced to 256×512 . The two branch features are then fused via gated weighted fusion to form a shared representation.

In the multi-task output stage, the shared features are routed to two heads in parallel. The detection head applies global pooling ($256 \times 512 \rightarrow 256$) followed by a linear classifier to output the detection probability. The extraction head applies transposed convolution upsampling ($256 \times 512 \rightarrow 256 \times 4096$) and a mask generation network to produce a mask matrix with the same dimensionality as the input (3×4096), and finally computes the extracted signal via element-wise multiplication.

In terms of temporal processing, the model is designed to be highly parallel, and the entire inference pipeline can be divided into four stages: (1) **Parallel encoding**, where the

physics branch and the neural network branch process the input simultaneously. (2) **Feature projection**, which performs dimensionality reduction and temporal alignment. (3) **Fusion**, which applies gated weighted fusion to obtain the shared representation. (4) **Multi-task outputs**, where the detection and extraction heads run in parallel. This parallel design improves computational efficiency in practice.

3. Experimental Design

3.1. Datasets and Evaluation Protocol

3.1.1. Dataset Details

To validate the cross-platform applicability of PGDSA and its potential for transfer to real observational data, we employ three datasets: a ground-based simulated gravitational wave dataset (G2Net), a ground-based real observational dataset (GWOSC O3), and a space-based simulated gravitational wave dataset (LISA).

(1) Ground-based simulated dataset (G2Net)

The G2Net Gravitational Wave Detection Challenge dataset [31] is a widely used benchmark for gravitational wave signal detection. It contains time series data from three detector channels (LIGO Hanford, LIGO Livingston, and Virgo). The original sampling rate is 4096 Hz and is downsampled to 2048 Hz. Each sample spans 2 s, corresponding to 4096 sampling points per channel. The dataset includes approximately 560,000 training samples and 226,000 test samples. Roughly half of the samples contain simulated stellar-mass BBH coalescence signals, while the other half are pure noise. The BBH parameter space covers broad ranges of component masses and spins, with a diverse distribution of signal-to-noise ratios (SNRs). The noise model is constructed to mimic real detector characteristics and includes both Gaussian background noise and non-Gaussian transient disturbances (glitches), making the task challenging. During training, each channel is preprocessed by applying a bandpass filter in the [20, 500] Hz range and whitening the data using the power spectral density estimated via Welch's method, thereby converting the colored detector noise into approximately white Gaussian noise and enhancing signal visibility within the detector-sensitive frequency band.

It should be noted that G2Net provides only binary labels (signal present vs. absent) and does not provide clean waveform templates. Therefore, waveform reconstruction metrics such as overlap cannot be computed. In this work, we evaluate **detection performance only** on G2Net, using it as a benchmark platform to assess the classification capability and noise robustness of PGDSA in the ground-based setting. We train the model using the official G2Net training set (with provided labels). Since the labels of the official test set are hidden, the reported ROC-AUC for G2Net is obtained from the **Kaggle leaderboard** by submitting predicted probabilities for the test set.

(2) Ground-based real observational dataset (GWOSC O3)

To evaluate PGDSA on real observational data, we obtain strain data from GWOSC for the third observing run (O3) [3]. The O3 run spans from April 2019 to March 2020, during which the LIGO–Virgo Collaboration confirmed 90 compact binary coalescence events that are included in the GWTC-3 transient catalog [2].

We construct a validation set by selecting representative BBH merger events from GWTC-3 using the following criteria: (1) the network SNR covers a medium-to-high range (approximately 10–50); (2) the source class is BBH; (3) at least two detectors are simultaneously operating with good data quality. To match the model input format (sampling rate 2048 Hz, 4096 points per channel, corresponding to 2 s), we extract strain data in a 2 s window centered on the merger time (1 s before and 1 s after coalescence). Data preprocessing follows the standard GWOSC pipeline; a bandpass filter is applied in the range [20, 500] Hz (consistent with the preprocessing applied during G2Net training), and the strain data are whitened using

the estimated power spectral density (PSD) computed from adjacent off-source segments. Data retrieval and preprocessing are implemented using the GWpy toolkit [32]. Note that GWOSC O3 data are used **only for testing**, not for model training.

(3) Space-based simulated dataset (LISA)

To validate PGDSA in the space-based detection setting and enable fair comparisons with existing baseline methods, we construct a LISA simulated dataset by strictly following the data generation procedure described in [13]. All data are generated with a unified sampling rate of 0.1 Hz. Each sample contains 16,000 points, corresponding to a duration of 160,000 s (approximately 44.4 h). The dataset covers four major gravitational wave source classes:

- **EMRI (Extreme-Mass-Ratio Inspiral):** Produced by a low-mass compact object inspiraling into a massive black hole. Waveforms are generated using the AAK (Augmented Analytic Kludge) model [33], with parameter ranges including central black hole mass $M \in [10^5, 10^7] M_\odot$, spin $a \in [10^{-3}, 0.99]$, eccentricity $e_0 \in [10^{-3}, 0.5]$, and inclination $\cos \iota \in [-1, 1]$.
- **MBHB (Massive Black Hole Binary):** Produced by the coalescence of two massive black holes during galaxy mergers. Waveforms are generated using the SEOBNRv4 model [34], with parameter ranges including total mass $M_{\text{tot}} \in [10^6, 10^8] M_\odot$ (log-uniform), mass ratio $q \in [0.01, 1]$, and spins $s_{1,2}^z \in [-0.99, 0.99]$.
- **BWD (Binary White Dwarfs):** The most common compact binaries in the Milky Way, producing quasi-monochromatic signals. Waveforms are generated following the procedure described in [13].
- **SGWB (Stochastic Gravitational Wave Background):** A random signal formed by the superposition of numerous unresolved sources. We adopt a power-law spectrum model

$$h^2 \Omega_{\text{GW}}(f) = 10^\alpha \left(\frac{f}{f_*} \right)^{n_t} \quad (32)$$

where $n_t = 2/3$ corresponds to a background generated by compact binary coalescences, $f_* = 10^{-3}$ Hz is the reference frequency, and the amplitude parameter α is set to -11.35 , -11.55 , and -11.75 for testing [13].

Figure 4 provides representative signal examples of the four LISA source classes.

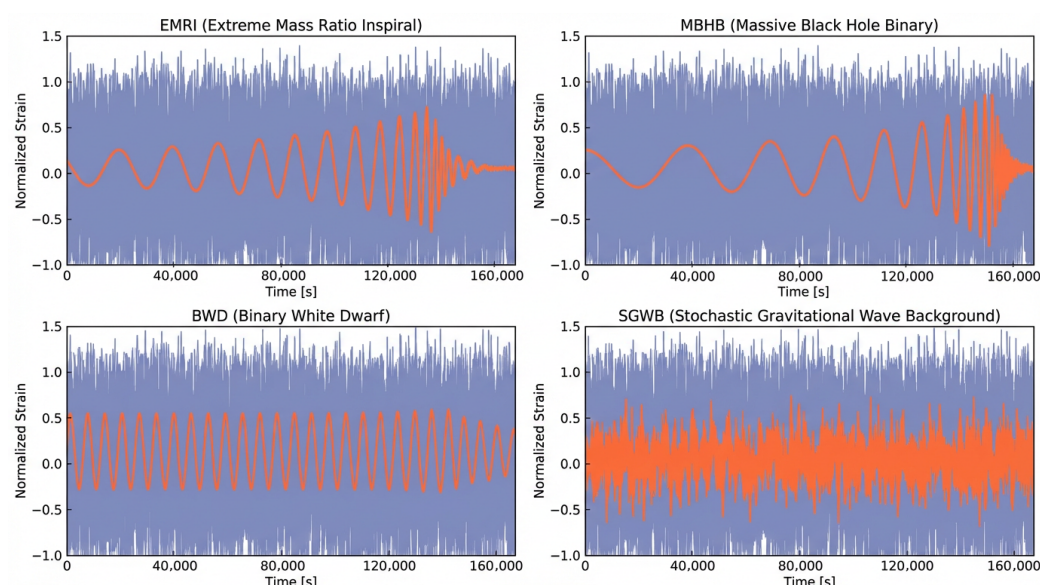


Figure 4. Representative signal examples of the four LISA source classes, illustrating their differences in time-domain and frequency-domain morphologies (blue: noise background; orange: gravitational wave signal; examples are shown in whitened form).

Gaussian noise is generated using the LISA sensitivity curve [7]. The corresponding PSD is

$$S_n(f) = \frac{1}{L^2 \mathcal{R}(f)} \left(P_{\text{OMS}} + 2(1 + \cos^2(f/f_*)) \frac{P_{\text{acc}}}{(2\pi f)^4} \right) \quad (33)$$

where $L = 2.5 \times 10^9$ m is the arm length, $f_* = 19.09$ mHz, and P_{OMS} and P_{acc} denote the optical metrology noise and the acceleration noise, respectively.

3.1.2. Evaluation Metrics

To comprehensively evaluate model performance, we adopt the following metrics:

Detection metrics:

(1) **ROC-AUC (Receiver Operating Characteristic Area Under the Curve)**, measuring overall classification performance:

$$\text{AUC} = \int_0^1 \text{TPR}(x) d\text{FPR}(x) \quad (34)$$

(2) **True-positive rate (TPR) at a fixed false-alarm rate (FAR)**, with the FAR being set to 1% in this work:

$$\text{TPR}_{\text{@FAR}=1\%} = \text{TPR}|_{\text{FPR}=0.01} \quad (35)$$

Waveform reconstruction (extraction) metrics (space-based data):

(3) **Overlap**, measuring the match between the extracted signal and the target waveform [13]:

$$\mathcal{O}(h, s) = (\hat{h}|\hat{s}) \quad (36)$$

where the inner product is defined by

$$(h|s) = 2 \int_{f_{\min}}^{f_{\max}} \frac{\tilde{h}^*(f)\tilde{s}(f) + \tilde{h}(f)\tilde{s}^*(f)}{S_n(f)} df \quad (37)$$

The optimal matched-filter SNR is defined as follows:

$$\text{SNR} = (s|s)^{1/2} \quad (38)$$

3.1.3. Experimental Environment

All experiments are conducted in a unified hardware and software environment. The hardware setup consists of an NVIDIA RTX PRO 6000 GPU and an AMD EPYC 7413 64-core CPU. The software environment includes PyTorch 2.1.0, CUDA 11.8, and Python 3.10.

3.2. Experimental Settings and Implementation Details

3.2.1. Model Configuration

We implement multiple variants of PGDSA for comparative experiments. Separate models are trained independently for ground-based and space-based scenarios; one model is trained on G2Net data for ground-based detection, and another is trained on LISA simulated data for space-based detection and waveform reconstruction. The model configurations are as follows:

Ground-based configuration (G2Net): The input shape is 3×4096 (three-detector time series). The model uses the dual-stream encoding architecture with 32 learnable Morlet wavelet kernels shared across all channels (frequency constraint range [20, 500] Hz), an eight-head dynamic sparse attention mechanism, and gated cross-modal fusion.

Space-based configuration (LISA): The input shape is set to $1 \times 16,000$ (single-channel LISA response); the frequency constraint range of the differentiable wavelet layer is adjusted

to $[10^{-4}, 0.05]$ Hz to match the LISA sensitive band; and the block size L is set to 100 to handle longer sequences.

3.2.2. Training Strategy

We use the Adam optimizer with an initial learning rate of 0.001. The learning rate scheduler follows cosine annealing with a minimum learning rate of 1×10^{-6} , and a warmup is applied for the first 10 epochs. The batch size is 256 (reduced to 128 for the standard Transformer baseline due to memory constraints). Training runs for 100 epochs with early stopping.

Data augmentation includes Gaussian noise injection (SNR range 15–30 dB), random time shifts (± 100 time steps), random signal flipping (probability 0.5), and random detector channel masking (probability 0.1). Regularization includes Dropout (rate 0.1), weight decay (1×10^{-5}), and gradient clipping (max norm 1.0).

3.3. Baseline Methods for Comparison

We select representative baseline methods with the scope of comparison determined by the evaluation constraints and data characteristics of each dataset.

Ground-based data (G2Net). The G2Net competition dataset [31] was designed as a large-scale benchmark for end-to-end deep learning approaches, providing only binary labels (signal present/absent) without access to clean waveform templates or underlying signal parameters. This design inherently limits the applicability of traditional matched filtering methods [10,11], which require extensive template bank searches with computational costs scaling from minutes to hours per sample, rendering systematic evaluation across 560,000+ samples computationally prohibitive. Similarly, Bayesian inference methods such as BayesWave [15] require iterative MCMC sampling, which is impractical at this scale. Accordingly, our G2Net evaluation focuses on deep learning baselines that are directly applicable to this benchmark setting, including ResNet-50 [35], Transformer [28], and CNN–wavelet hybrid architectures [25].

Space-based data (LISA). For LISA simulations, we adopt the evaluation protocol established by Zhao et al. [13], employing their self-attention-based neural network as the primary baseline (denoted as “Baseline” in Section 5.1.3). This enables direct performance comparison under identical data generation procedures, signal parameter distributions, and evaluation metrics. Traditional MCMC-based Bayesian methods, while theoretically applicable, require computational times in the order of hours per signal [13], compared to subsecond inference for deep learning approaches—a factor of approximately 10^5 improvement that motivates the development of neural network-based analysis pipelines for future large-scale observations.

4. Model Complexity and Performance Analysis

4.1. Computational Complexity Analysis

4.1.1. Time Complexity

We analyze the time complexity of the core components in PGDSA. Let the input sequence length be T , the feature dimension be d , the block size be L , and the sparsity ratio of the attention matrix be s .

For standard self-attention, the computational complexity is $O(T^2d)$, since attention weights must be computed for all pairs of positions. The proposed dynamic sparse attention reduces computation via block partitioning and a Top-K sparsification strategy. After partitioning a length- T sequence into T/L blocks, the dominant upper bound for attention score computation becomes $O((T/L) \cdot L^2d) = O(TLd)$.

Table 2 reports the normalized compute time of the sparse attention module in the PGDSA relative to standard Transformer attention under different sequence lengths.

Table 2. Normalized compute time of the attention module under different sequence lengths.

Sequence Length	Normalized Time (Standard = 1)
1024	0.34
2048	0.23
4096	0.15
8192	0.11

As the sequence length increases, the relative compute cost of sparse attention decreases compared with standard self-attention, making it more suitable for long time series inputs.

4.1.2. Space Complexity

In terms of space complexity, a standard Transformer needs to store the full attention matrix, resulting in a memory complexity of $O(T^2)$. Under the block-wise computation setting, the peak attention-related workspace can be reduced to $O(L^2)$ by computing blocks sequentially and releasing intermediate buffers immediately. Empirically, measurements indicate that, for sequences of length 4096, the GPU memory usage associated with attention matrices can be reduced by approximately 64%.

4.1.3. Model Size

The complete PGDSA model under the ground-based configuration (Section 3.2.1) contains approximately 20.7 million trainable parameters, which is smaller than the ResNet-50 baseline (~ 25.6 M). Notably, the physics-inspired time–frequency branch contributes only approximately 0.06M parameters ($< 0.3\%$ of total), consisting of 160 wavelet kernel parameters, a lightweight dual-attention enhancement module, and a single convolutional projection layer. This demonstrates that incorporating physical priors via a differentiable wavelet layer is highly parameter-efficient. The majority of the parameters reside in the WaveNet encoder and the sparse Transformer; however, the dynamic sparse attention mechanism reduces FLOPs and peak memory rather than parameter count, since the Q/K/V projection matrices are the same size regardless of the sparsity ratio.

4.2. Effectiveness Analysis of the Model Design

4.2.1. Role of the Physics-Inspired Time–Frequency Branch

The differentiable wavelet transform is the core component of the physics-inspired time–frequency branch, motivated by the time–frequency characteristics of gravitational wave signals. Wavelet analysis is naturally suited for non-stationary chirp signals, as it can capture patterns of frequency evolution over time at multiple scales.

Ablation results (Section 5.2) indicate that this design improves detection performance. In addition, the wavelet parameter analysis in Section 5.4.3 shows that the learned wavelet kernels' center frequencies converge to a range close to the sensitive band of LIGO (20–350 Hz), which is consistent with the motivation of using **learnable physical priors**.

4.2.2. Representational Capacity of Sparse Attention

The dynamic sparse attention mechanism retains approximately 30% of the most important attention connections via Top-K selection. Experimental results suggest that this sparsification strategy substantially reduces computation while only marginally affecting performance. Compared with 100% dense attention, 30% sparsity yields only

a slight decrease in AUC (approximately 0.002), while the computational cost can be significantly reduced.

4.2.3. Adaptivity and Synergistic Gains from the Gated Fusion Mechanism

The gated fusion mechanism is critical to the synergistic gains of PGDSA. Traditional additive fusion $f = f_p + f_n$ or multiplicative fusion $f = f_p \odot f_n$ uses fixed fusion rules and cannot adapt to signal characteristics. In contrast, gated fusion $f = \alpha \odot f_p + (1 - \alpha) \odot f_n$ achieves input-adaptive dynamic fusion through a learnable weight α .

This adaptivity provides two advantages: First, **complementarity**: when the physics branch is more reliable for certain samples, the model automatically increases α ; when the data-driven branch is more robust, the model decreases α . Second, synergistic gains: the gating mechanism learns *when to trust physics and when to trust data*.

5. Results and Discussion

This section presents a detailed analysis of the experimental results, compares PGDSA with existing methods, and discusses the contributions and limitations of key components.

5.1. Main Performance Comparisons

5.1.1. Ground-Based Gravitational Wave Detection Performance (G2Net)

Table 3 compares PGDSA with several representative deep learning methods on the G2Net test set. The official evaluation metric of the G2Net challenge is ROC-AUC [31].

Table 3. Detection performance comparison on the G2Net dataset.

Model	AUC
ResNet-50 [35]	0.880
Transformer [28]	0.880
CNN + wavelet [25]	0.870
PGDSA (ours)	0.886

5.1.2. Validation on Real Events from GWOSC O3

To evaluate PGDSA on real observational data, we perform **zero-shot testing** on BBH events selected from the GWTC-3 catalog. The model is trained on simulated G2Net data and directly applied to GWOSC O3 strain data without any fine-tuning or retraining. Table 4 reports detection results for a set of representative events.

Table 4. Example detection scores for on-source windows of real BBH events in GWOSC O3.

Event	Network SNR	Detection Score	Event Characteristics
GW190412	19.1	0.987	Asymmetric mass ratio
GW190521	14.7	0.952	Intermediate-mass BH component
GW190828_063405	16.3	0.971	Typical BBH
GW191109_010717	15.6	0.963	Typical BBH

These results indicate that PGDSA assigns high detection scores to the on-source windows of confirmed GWTC-3 BBH events, and the scores exhibit an overall positive association with network SNR.

5.1.3. Space-Based Gravitational Wave Detection Performance (LISA)

Table 5 reports the TPR at a fixed FAR of 1% under different SNR conditions.

Table 5. Detection performance on the LISA dataset (TPR at FAR = 1%).

Signal type	Method	SNR = 30	SNR = 40	SNR = 50
EMRI	Baseline	98.20%	99.70%	99.71%
	PGDSA (Ours)	98.56%	99.78%	99.82%
MBHB	Baseline	99.99%	99.999%	99.999%
	PGDSA (Ours)	99.996%	99.999%	99.999%
BWD	Baseline	99.37%	99.97%	99.98%
	PGDSA (Ours)	99.52%	99.98%	99.99%
SGWB *	Baseline	95.05%	99.97%	100.00%
	PGDSA (Ours)	95.89%	99.98%	100.00%

* For SGWB, the three columns correspond to amplitude parameters $\alpha = -11.75, -11.55$, and -11.35 , respectively.

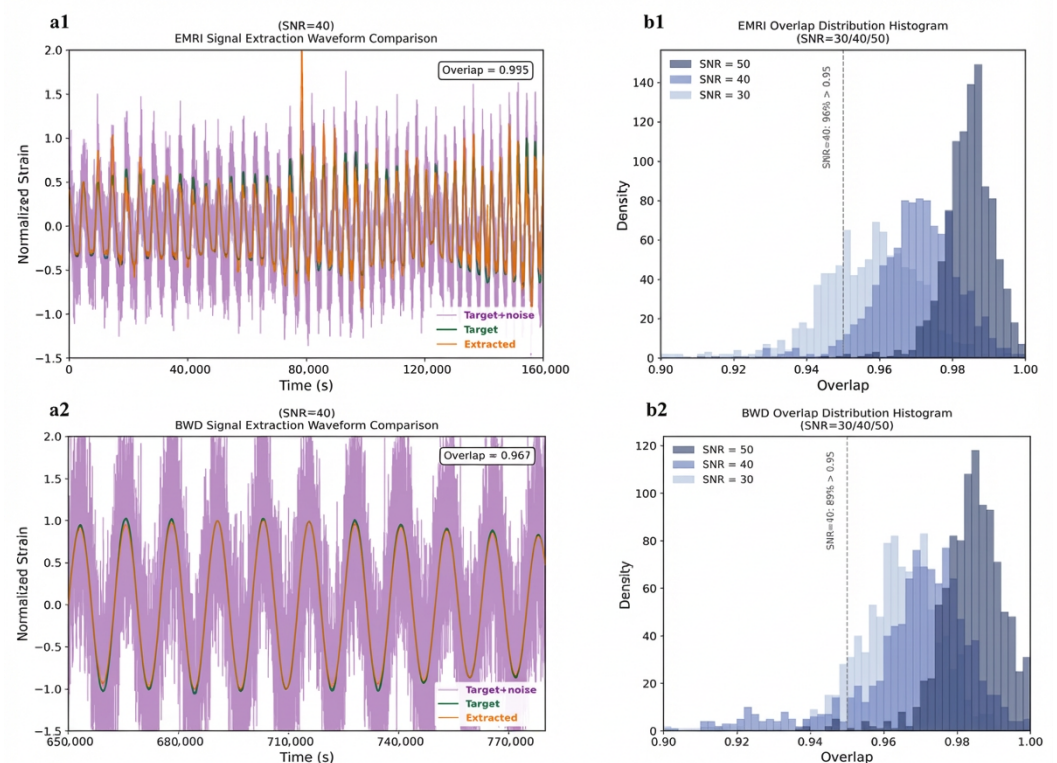
5.1.4. Space-Based Waveform Reconstruction Performance

Table 6 reports waveform reconstruction results at SNR = 50 for different signal types.

Table 6. Waveform reconstruction performance on the LISA dataset (SNR = 50).

Signal Type	Method	Samples with Overlap > 0.95	Typical Overlap
EMRI	Baseline	92%	~0.96
	PGDSA (Ours)	93%	~0.96
MBHB	Baseline	100%	>0.99
	PGDSA (Ours)	100%	>0.99
BWD	Baseline	95%	~0.98
	PGDSA (Ours)	95%	~0.98

Figure 5 shows representative extraction examples produced by PGDSA for different signal types.

**Figure 5.** Cont.

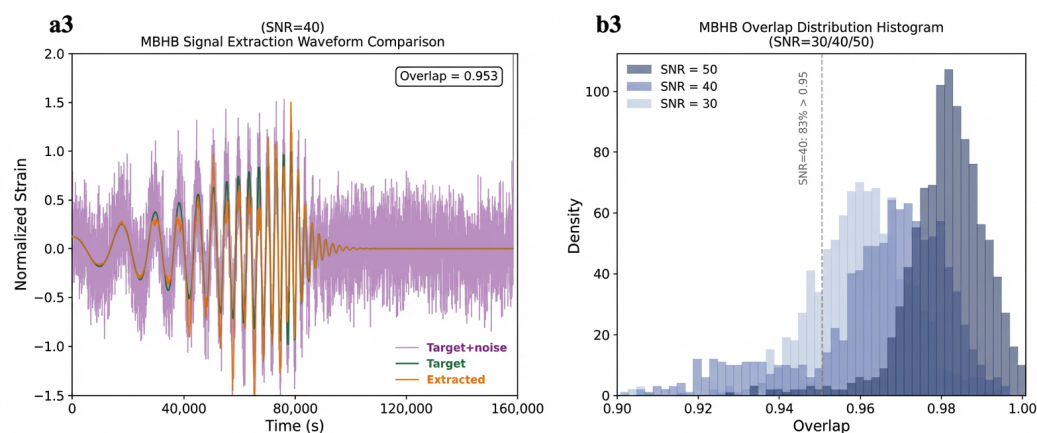


Figure 5. Waveform reconstruction examples and Overlap distribution for EMRI, BWD, and MBHB signals (from top to bottom). **Left column (a1–a3):** waveform comparison of representative samples at SNR = 40, selected from the 15th percentile of the test set Overlap distribution to illustrate challenging reconstruction cases (purple: noisy input; orange: reconstructed signal; green: target template). The displayed Overlap values indicate the reconstruction quality of each individual sample. **Right column (b1–b3):** Overlap histograms computed over the full test set at SNR = 30, 40, and 50.

5.2. Ablation Results

5.2.1. Contributions of Major Components

Table 7 reports the ablation results on the G2Net dataset.

Table 7. Ablation results on the G2Net dataset.

Model ID	Physics Module	Sparse Attention	Gated Fusion	AUC
M1	✓	×	×	0.862
M2	×	✓	×	0.871
M3	✓	✓	×	0.874
M4	✓	✓	✓	0.886

The ablation results show that the physics module, sparse attention, and gated fusion each improve detection performance to some extent, and that combining components typically yields better results than using any single component alone.

5.2.2. Fine-Grained Ablation Analysis

Effect of learnable wavelet parameters. Learnable wavelet parameters outperform fixed wavelet parameters, indicating that end-to-end learning allows the wavelet kernels to adaptively match dataset-specific signal characteristics.

Effect of attention sparsity. Approximately 30% sparsity provides a favorable balance; detection accuracy is largely preserved while computation and memory consumption are reduced.

Fusion mechanism comparison. Gated fusion shows clear advantages over additive and multiplicative fusion, especially on low-SNR samples.

5.3. Noise Robustness Analysis

Results show that PGDSA is overall no worse than baseline deep learning methods across SNR levels. In particular, PGDSA exhibits certain advantages in the low-SNR regime, suggesting that the physics-inspired time–frequency branch may help extract weak signal features. To further evaluate robustness under challenging conditions, Figure 6 presents waveform reconstruction results at SNR levels of 15, 20, and 30. Even at SNR = 20, PGDSA

achieves Overlap values exceeding 0.90 for all three signal types (EMRI: 0.939, BWD: 0.924, MBHB: 0.907), confirming the framework's reconstruction capability in low-SNR regimes.

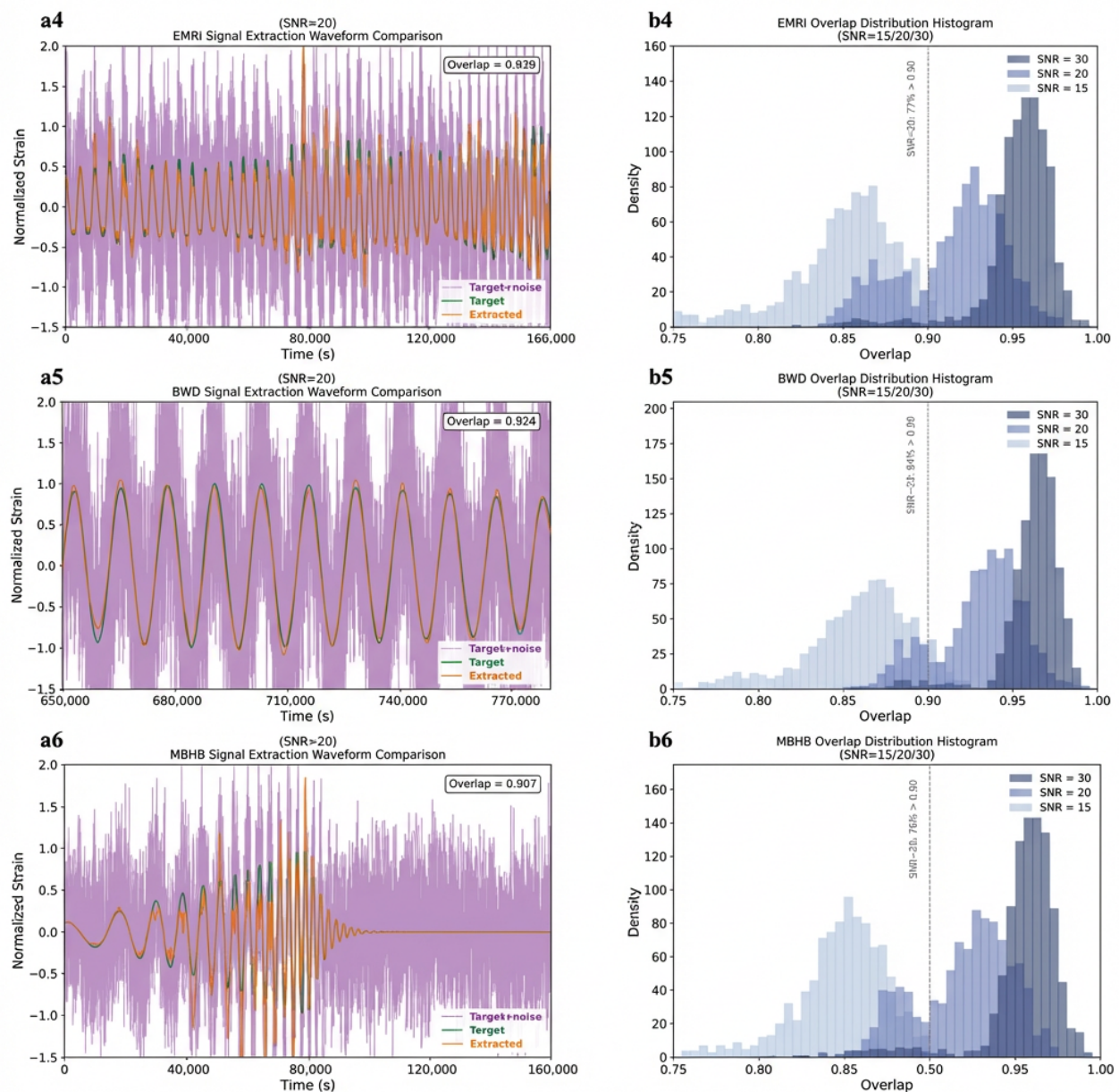


Figure 6. Waveform reconstruction examples and Overlap distribution under low-SNR conditions. **Left column (a4–a6):** representative samples at SNR = 20 for EMRI (Overlap = 0.939), BWD (Overlap = 0.924), and MBHB (Overlap = 0.907), selected from the 15th percentile of the test set Overlap distribution. **Right column (b4–b6):** Overlap histograms at SNR = 15, 20, and 30, demonstrating that PGDSA maintains meaningful reconstruction quality under substantially more challenging noise conditions.

5.4. Visualization Analysis and Interpretation

5.4.1. An Interpretability Example for MBHB Signals

Figure 7 uses an MBHB merger signal as an example to visualize how the attention weights and the gating coefficient α evolve when chirp structures appear.

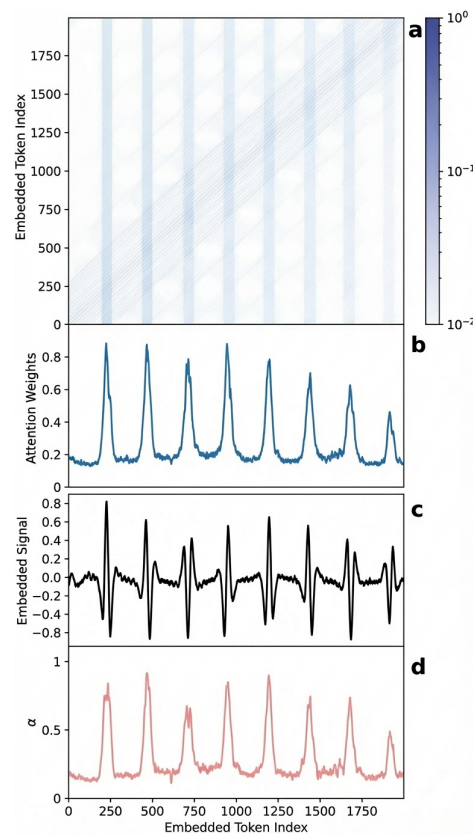


Figure 7. Interpretability example for an MBHB signal: variations of attention weights and the gating coefficient α during the chirp stage. (a) Sparse attention weight matrix from the last attention block (averaged over heads; log-scale color bar indicates normalized weight magnitude). (b) Mean attention weight along the temporal axis, showing pronounced peaks at chirp phases. (c) Embedded gravitational wave signal. (d) Gating coefficient α along the temporal axis. This visualization is reported for a single MBHB sample. Attention weights are normalized to $[0, 1]$ for visualization, and the fusion gate $\alpha \in [0, 1]$ is the position-wise coefficient used in $F_{\text{fuse}} = \alpha \odot P + (1 - \alpha) \odot N$. The highlighted chirp stage corresponds to the time interval where the waveform amplitude and instantaneous frequency increase rapidly.

5.4.2. Analysis of the Gating Coefficient

Figure 8 illustrates example patterns of α in the time–frequency domain and its relationship with SNR.

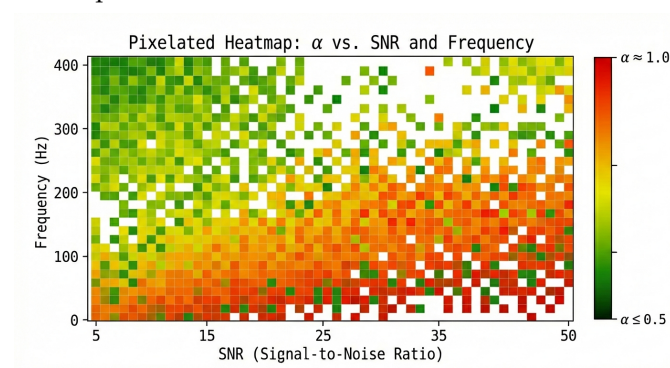


Figure 8. Time–frequency distribution of the gating coefficient α and its relation to SNR (ground-based example). The gate map α is the position-wise fusion weight ($[0, 1]$) and is visualized in the time–frequency domain after projecting the fused representation back to the corresponding bins (color bar indicates α). The SNR is computed following the dataset protocol (Section 3.1.1), and the right panel statistics summarize the relationship between sample-level SNR and the average α , showing higher reliance on the physics branch for higher-SNR and lower-frequency components.

The analysis reveals three salient properties: (1) **frequency dependence**; (2) **temporal dependence**; and (3) **SNR dependence**.

5.4.3. Analysis of Learnable Wavelet Parameters

We find that, for models trained on G2Net, the distribution of wavelet kernel center frequencies f_θ is consistent with the sensitive band of ground-based detectors, suggesting that the learnable wavelet layer can form a frequency band inductive bias through end-to-end training.

5.5. Discussion

This study explores the feasibility of combining learnable physical priors with deep learning under a multi-task framework to perform signal detection and extraction.

Role of multi-task learning. When extraction supervision is available, joint optimization of detection and extraction encourages shared representation learning.

Cross-platform adaptability. PGDSA adopts a unified architecture that can be applied to both ground-based and space-based data by adjusting only a small number of configurations.

Transfer potential to real data. The GWOSC O3 validation shows that a model trained only on simulated G2Net data assigns high detection scores to on-source windows of confirmed BBH events, demonstrating zero-shot transfer potential from simulations to real observations. The zero-shot evaluation setting was deliberately chosen to assess cross-domain generalization under realistic deployment conditions; with only approximately 90 confirmed events in GWTC-3, each providing roughly 2 s of on-source data, the available real data sample size is insufficient for training or fine-tuning deep neural networks. Investigating few-shot fine-tuning strategies and semi-supervised approaches that leverage unlabeled observational data represents a promising direction for future work.

Interpretation of performance gains. On the LISA benchmark, the baseline method [13] already achieves detection rates exceeding 99% for most signal types at moderate-to-high SNR, leaving limited room for absolute numerical improvement. Under such near-saturation conditions, even modest absolute gains correspond to meaningful relative error reductions. For example, at SNR = 30 the EMRI detection rate improves from 98.20% to 98.56%, reducing the miss rate from 1.80% to 1.44%—a relative error reduction of approximately 20%. For the most challenging SGWB signals ($\alpha = -11.75$), PGDSA improves detection from 95.05% to 95.89%. We emphasize that the primary contributions of this work extend beyond numerical gains on a single benchmark; the unified cross-platform architecture, physics-guided interpretability, sparse attention efficiency, and multi-task learning capability collectively represent a methodological advance that complements the quantitative improvements.

Computational efficiency. The dynamic sparse attention mechanism achieves substantial computational savings compared to standard full self-attention. At sequence length 4096, the attention module achieves a normalized compute time of 0.15 relative to dense attention (Table 2), corresponding to an approximately $6.7\times$ speedup, with progressively greater savings at longer sequence lengths ($9.1\times$ at length 8192). While the overall model inference is moderately slower than pure convolutional baselines due to the dual-branch architecture, this trade-off is justified by the improved detection accuracy and enhanced interpretability, particularly for off-line gravitational wave data analysis applications where accuracy takes priority over raw throughput.

Limitations. This work adopts a morphology-based physics-guided strategy which achieves a trade-off between computational feasibility and physical plausibility but does not incorporate first-principles constraints such as explicit general relativistic equations.

Additionally, the LISA experiments adopt a sampling rate of 0.1 Hz following the protocol of Zhao et al. [13], which limits the analyzable frequency band to below 50 mHz. While this covers the primary science targets of LISA (EMRI, MBHB inspiral, and BWD), extension to higher frequencies would require increased sampling rates and retraining.

Future work. Possible extensions include expanding the framework to parameter estimation; incorporating deeper physical constraints; adopting more expressive signal separation decoders for the extraction head; performing retrieval-level evaluation on larger real datasets; and improving multi-detector fusion strategies.

6. Conclusions

This paper proposes the Physics-Guided Dynamic Sparse Attention (PGDSA) framework for gravitational wave detection across ground-based and space-based observatories. The framework introduces differentiable wavelet transforms as learnable time–frequency analysis modules, employs block-wise Top-K sparse attention to reduce computational costs for long sequences, and uses gated fusion to adaptively integrate physics-guided and data-driven features.

Experimental results demonstrate that PGDSA achieves an ROC-AUC score of 0.886 on the G2Net ground-based dataset (Kaggle private leaderboard), produces high confidence scores for confirmed BBH events in GWOSC O3 real data under zero-shot testing conditions, and achieves detection rates exceeding 99% on LISA simulated data (SNR=50, FAR=1%) with waveform reconstruction Overlap comparable to baseline methods. Furthermore, PGDSA maintains robust reconstruction performance under challenging low-SNR conditions, with Overlap values exceeding 0.90 at SNR = 20 across all tested signal types, and the dynamic sparse attention mechanism reduces attention module compute time to 15% of standard Transformer attention at sequence length 4096.

These results indicate that the PGDSA framework enables unified cross-platform modeling for both space-based and ground-based scenarios, combining physics-guided interpretability with computational efficiency, thereby providing a practical reference for gravitational wave data analysis in the era of multi-band observations.

Author Contributions: Conceptualization, T.Z. and W.B.; methodology, T.Z.; software, T.Z.; validation, T.Z.; formal analysis, T.Z.; investigation, T.Z.; writing—original draft preparation, T.Z.; writing—review and editing, W.B.; supervision, W.B.; project administration, W.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets used in this study are publicly available. The G2Net dataset is available from the Kaggle G2Net Gravitational Wave Detection Challenge [31]: <https://www.kaggle.com/competitions/g2net-gravitational-wave-detection> (accessed on 1 February 2026). The space-based gravitational wave dataset used in this study was generated using code from the repository associated with Reference [13]. To reproduce these datasets, please follow the instructions provided in the repository documentation https://github.com/AI-HPC-Research-Team/space_signal_detection_1 (accessed on 1 February 2026). The source code for the PGDSA framework is available at <https://github.com/UCAS-ZTC/PGDSA-GW-Detection> (accessed on 1 February 2026); the complete implementation, including model architecture, training scripts, and evaluation protocols, will be released upon acceptance of the manuscript.

Acknowledgments: The authors thank the LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration for making gravitational wave data publicly available through the Gravitational Wave Open Science Center (GWOSC). We also thank the organizers of the G2Net Kaggle competition for providing the benchmark dataset.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Abbott, B.P.; Abbott, R.; Abbott, T.D.; Abernathy, M.R.; Acernese, F.; Ackley, K.; Adams, C.; Adams, T.; Addesso, P.; LIGO Scientific Collaboration and Virgo Collaboration; et al. Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.* **2016**, *116*, 061102. <https://doi.org/10.1103/PhysRevLett.116.061102>.
- Abbott, R.; Abbott, T.; Acernese, F.; Ackley, K.; Adams, C.; Adhikari, N.; Adhikari, R.; Adya, V.; Affeldt, C.; LIGO Scientific Collaboration and Virgo Collaboration; et al. GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run. *Phys. Rev. X* **2023**, *13*, 041039. <https://doi.org/10.1103/PhysRevX.13.041039>.
- Abbott, R.; Abe, H.; Acernese, F.; Ackley, K.; Adhikari, S.; Adhikari, N.; Adhikari, R.X.; Adkins, V.K.; Adya, V.B.; Affeldt, C.; et al. Open Data from the Third Observing Run of LIGO, Virgo, KAGRA, and GEO. *Astrophys. J. Suppl. Ser.* **2023**, *267*, 29. <https://doi.org/10.3847/1538-4365/acdc9f>.
- Abbott, B.P.; Abbott, R.; Abbott, T.D.; Abraham, S.; Acernese, F.; Ackley, K.; Adams, C.; Adya, V.B.; Affeldt, C.; Agathos, M.; et al. Prospects for Observing and Localizing Gravitational-Wave Transients with Advanced LIGO, Advanced Virgo and KAGRA. *Living Rev. Relativ.* **2020**, *23*, 3. <https://doi.org/10.1007/s41114-020-00026-9>.
- Luo, Z.; Wang, Y.; Wu, Y.; Hu, W.; Jin, G. The Taiji Program: A Concise Overview. *Prog. Theor. Exp. Phys.* **2021**, *2021*, 05A108. <https://doi.org/10.1093/ptep/ptaa083>.
- Amaro-Seoane, P.; Audley, H.; Babak, S.; Baker, J.; Barausse, E.; Bender, P.; Berti, E.; Binetruy, P.; Born, M.; Bortoluzzi, D.; et al. Laser Interferometer Space Antenna. *arXiv* **2017**, arXiv:1702.00786. <https://doi.org/10.48550/arXiv.1702.00786>.
- Robson, T.; Cornish, N.J.; Liu, C. The Construction and Use of LISA Sensitivity Curves. *Class. Quantum Gravity* **2019**, *36*, 105011. <https://doi.org/10.1088/1361-6382/ab1101>.
- Schutz, B.F. Gravitational Wave Sources and Their Detectability. *Class. Quantum Gravity* **1989**, *6*, 1761–1780. <https://doi.org/10.1088/0264-9381/6/12/006>.
- Benedetto, V.; Berberi, L.; Capozziello, S.; Feoli, A. AI in Gravitational Wave Analysis: An Overview. *Appl. Sci.* **2023**, *13*, 9886. <https://doi.org/10.3390/app13179886>.
- Usman, S.A.; Nitz, A.H.; Harry, I.W.; Biwer, C.M.; Brown, D.A.; Cabero, M.; Capano, C.D.; Canton, T.D.; Dent, T.; Fairhurst, S.; et al. The PyCBC Search for Gravitational Waves from Compact Binary Coalescence. *Class. Quantum Gravity* **2016**, *33*, 215004. <https://doi.org/10.1088/0264-9381/33/21/215004>.
- Sachdev, S.; Caudill, S.; Fong, H.; Lo, R.K.L.; Messick, C.; Mukherjee, D.; Magee, R.; Tsukada, L.; Blackburn, K.; Brady, P.; et al. The GstLAL Search Analysis Methods for Compact Binary Mergers in Advanced LIGO's Second and Advanced Virgo's First Observing Runs. *arXiv* **2019**, arXiv:1901.08580. <https://doi.org/10.48550/arXiv.1901.08580>.
- Klimenko, S.; Yakushin, I.; Mercer, A.; Mitselmakher, G. A Coherent Method for Detection of Gravitational Wave Bursts. *Class. Quantum Gravity* **2008**, *25*, 114029. <https://doi.org/10.1088/0264-9381/25/11/114029>.
- Zhao, T.; Lyu, R.; Wang, H.; Cao, Z.; Ren, Z. Space-Based Gravitational Wave Signal Detection and Extraction with Deep Neural Network. *Commun. Phys.* **2023**, *6*, 212. <https://doi.org/10.1038/s42005-023-01334-6>.
- Sathyaprakash, B.S.; Schutz, B.F. Physics, Astrophysics and Cosmology with Gravitational Waves. *Living Rev. Relativ.* **2009**, *12*, 2. <https://doi.org/10.12942/lrr-2009-2>.
- Cornish, N.J.; Littenberg, T.B. BayesWave: Bayesian Inference for Gravitational Wave Bursts and Instrument Glitches. *Class. Quantum Gravity* **2015**, *32*, 135012. <https://doi.org/10.1088/0264-9381/32/13/135012>.
- George, D.; Huerta, E.A. Deep Learning for Real-Time Gravitational Wave Detection and Parameter Estimation: Results with Advanced LIGO Data. *Phys. Lett. B* **2018**, *778*, 64–70. <https://doi.org/10.1016/j.physletb.2017.12.053>.
- Gebhard, T.D.; Kilbertus, N.; Harry, I.; Schölkopf, B. Convolutional Neural Networks: A Magic Bullet for Gravitational-Wave Detection? *Phys. Rev. D* **2019**, *100*, 063015. <https://doi.org/10.1103/PhysRevD.100.063015>.
- Lin, Y.-C.; Wu, J.-H.P. Detection of Gravitational Waves Using Bayesian Neural Networks. *Phys. Rev. D* **2021**, *103*, 063034. <https://doi.org/10.1103/PhysRevD.103.063034>.
- Poghosyan, R.; Luo, Y. Random Convolutional Kernels for Space-Detector Based Gravitational Wave Signals. *Electronics* **2023**, *12*, 4360. <https://doi.org/10.3390/electronics12204360>.
- Wang, H.; Zhou, Y.; Cao, Z.; Guo, Z.-K.; Ren, Z. WaveFormer: Transformer-Based Denoising Method for Gravitational-Wave Data. *Mach. Learn. Sci. Technol.* **2024**, *5*, 015046. <https://doi.org/10.1088/2632-2153/ad2f54>.
- Shi, R.; Zhou, Y.; Zhao, T.; Cao, Z.; Ren, Z. Compact Binary Systems Waveform Generation with Generative Pre-trained Transformer. *arXiv* **2023**, arXiv:2310.20172. <https://doi.org/10.48550/arXiv.2310.20172>.
- Legin, R.; Isi, M.; Wong, K.W.K.; Hezaveh, Y.; Perreault-Levasseur, L. Gravitational-Wave Parameter Estimation in Non-Gaussian Noise Using Score-Based Likelihood Characterization. *arXiv* **2024**, arXiv:2410.19956. <https://doi.org/10.48550/arXiv.2410.19956>.
- Rhif, M.; Ben Abbes, A.; Farah, I.R.; Martínez, B.; Sang, Y. Wavelet Transform Application for/in Non-Stationary Time-Series Analysis: A Review. *Appl. Sci.* **2019**, *9*, 1345. <https://doi.org/10.3390/app9071345>.
- Kang, S.-K.; Yie, S.-Y.; Lee, J.-S. Noise2Noise Improved by Trainable Wavelet Coefficients for PET Denoising. *Electronics* **2021**, *10*, 1529. <https://doi.org/10.3390/electronics10131529>.

25. Cuoco, E.; Powell, J.; Cavaglià, M.; Ackley, K.; Beijer, M.; Chatterjee, C.; Coughlin, M.; Coughlin, S.; Easter, P.; Essick, R.; et al. Enhancing Gravitational-Wave Science with Machine Learning. *Mach. Learn. Sci. Technol.* **2021**, *2*, 011002. <https://doi.org/10.1088/2632-2153/abb93a>.
26. Scorzato, L. Reliability and Interpretability in Science and Deep Learning. *Minds Mach.* **2024**, *34*, 27. <https://doi.org/10.1007/s11023-024-09682-0>.
27. Tay, Y.; Dehghani, M.; Bahri, D.; Metzler, D. Efficient Transformers: A Survey. *ACM Comput. Surv.* **2022**, *55*, 109. <https://doi.org/10.1145/3530811>.
28. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In Proceedings of the Advances in Neural Information Processing Systems 30 (NeurIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008. <https://doi.org/10.48550/arXiv.1706.03762>.
29. Van Den Oord, A.; Dieleman, S.; Zen, H.; Simonyan, K.; Vinyals, O.; Graves, A.; Kalchbrenner, N.; Senior, A.; Kavukcuoglu, K. WaveNet: A Generative Model for Raw Audio. *arXiv* **2016**, arXiv:1609.03499. <https://doi.org/10.48550/arXiv.1609.03499>.
30. Zhang, Y.; Liu, B.; Yang, Y.; Yang, Q. Gated Cross-Attention for Universal Speaker Extraction: Toward Real-World Applications. *Electronics* **2024**, *13*, 2046. <https://doi.org/10.3390/electronics13112046>.
31. G2Net Gravitational Wave Detection Challenge. Kaggle Competition. Available online: <https://www.kaggle.com/competitions/g2net-gravitational-wave-detection> (accessed on 1 February 2026).
32. Macleod, D.M.; Areeda, J.S.; Coughlin, S.B.; Massinger, T.J.; Urban, A.L. GWpy: A Python Package for Gravitational-Wave Astrophysics. *SoftwareX* **2021**, *13*, 100657. <https://doi.org/10.1016/j.softx.2021.100657>.
33. Katz, M.L.; Chua, A.J.; Speri, L.; Warburton, N.; Hughes, S.A. Fast Extreme-Mass-Ratio-Inspiral Waveforms: New Tools for Millihertz Gravitational-Wave Data Analysis. *Phys. Rev. D* **2021**, *104*, 064047. <https://doi.org/10.1103/PhysRevD.104.064047>.
34. Bohé, A.; Shao, L.; Taracchini, A.; Buonanno, A.; Babak, S.; Harry, I.W.; Hinder, I.; Ossokine, S.; Pürrer, M.; Raymond, V.; 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. D* **2017**, *95*, 044028. <https://doi.org/10.1103/PhysRevD.95.044028>.
35. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016*; IEEE: New York, NY, USA; pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.