Fusion of Multi-Instrument Time-Energy Data with Self-Attention for Enhanced GRB Detection

Xiaotian Wang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: wangxiaotian@st.xatu.edu.cn

Long Ma
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: malong@xatu.edu.cn

***Abstract*—Gamma-ray burst (GRB) detection is crucial for triggering rapid follow-up observations and enabling subsequent multi-wavelength studies, yet traditional methods are limited by their difficulty in modeling long-range temporal dependencies in noisy, non-stationary data. In this work, we propose GRBNet, a self-attention–based neural network for GRB detection from same-source multi-detector time-tagged event (TTE) sequences. GRBNet leverages multi-head self-attention to adaptively aggregate discriminative evidence over the full observation window, explicitly capturing long-term dependencies and multi-pulse structures, and integrates complementary responses from multiple detectors under a unified tokenization scheme to improve robustness against background drift and low SNR in individual instruments. Experiments on real Fermi/GBM observations show strong detection performance (recall=1.00, precision=1.00) under a stringent triggering threshold 0.99, indicating reliable performance for weak and morphologically complex GRB events.**

Keywords-GRB; Self-Attention; Style; Multi-Detector Fusion; Deep Learning

I. INTRODUCTION

Gamma-ray bursts (GRBs) [1] are among the most energetic transient phenomena in the Universe, characterized by a wide range of durations—from milliseconds to several hundreds of seconds—as well as highly complex temporal profiles and rapidly evolving energy spectra. Reliable detection of GRBs is not only a prerequisite for triggering subsequent multi-wavelength follow-up observations and rapid localization, but also a crucial starting point for investigating their physical origins and radiation mechanisms. However, under realistic observational conditions, GRB signals are often embedded in strong and non-stationary backgrounds, including diffuse instrumental background, cosmic-ray–induced events, intrinsic detector noise, and variations related to the satellite orbital environment. Moreover, due to differences in detector response functions, effective areas, incident angles, and energy channel configurations, different detectors record the same event with inconsistent amplitudes, signal-to-noise ratios, and temporal structures. These factors collectively make the stable identification of transient signals under strong noise, non-stationary backgrounds, and diverse burst morphologies a central challenge in GRB detection.

Traditional GRB detection methods [2–3] have been extensively used in high-energy astrophysics missions for decades. These approaches typically rely on manually defined time windows, predefined energy-band combinations, and fixed triggering thresholds, implicitly assuming that the background can be approximated by a simple local model. As a consequence, they operate primarily on aggregated statistical quantities rather than directly modeling fine-grained time-tagged event (TTE) sequences, which limits their ability to capture long-range temporal dependencies and complex global burst morphology.

Moreover, conventional detection pipelines usually treat observations from different instruments independently or combine them through heuristic logic, without explicitly modeling cross-detector consistency and complementary response patterns. This restricts their robustness under strong noise, background variability, or low signal-to-noise conditions.

While computationally efficient, such methods are fundamentally constrained by their reliance on local statistical assumptions and lack of unified multi-instrument modeling.

To address these challenges, we propose a method for GRB detection, which is characterized by two key components. (i) we build GRBNet, a self-attention–based [4] neural network designed for multi-instrument time-tagged event (TTE) modeling, which directly takes TTE data from multiple observing instruments as input. (ii) we propose a multi-instrument data fusion strategy that jointly models observations from different detectors observing the same physical event. This fusion mechanism enables the network to exploit cross-detector consistency and complementary response patterns, thereby improving robustness and discriminability.

Tests on real GRB observational data show that GRBNet achieves a detection accuracy of 1.0 with no false alarm even under a stringent triggering threshold 0.99. This result indicates that the proposed approach can maintain high reliability under strict triggering conditions, highlighting both the effectiveness of self-attention in modeling long-sequence patterns and the robustness gains brought by multi-instrument joint inputs.

II. FRAMEWORK

The overall workflow is straightforward. Specifically, we first preprocess the TTE event sequences recorded by multiple GRB detectors into a sequence of tokens [4], which are then fed into GRBNet to produce a binary classification probability corresponding to “GRB” versus “non-GRB” (Figure 1.). Under our experimental setting, each sample consists of discretized multi-energy-band photon count sequences from three detectors within a fixed time window.

A. Tokenization

To meet the input requirements of the proposed model, we first convert the raw Fermi/GBM TTE photon event data into a time–energy sequence representation (Figure 2.). We then obtain the final token sequence by applying a simple dimension rearrangement to this representation. The same preprocessing procedure is applied to every TTE sample in a consistent and reproducible manner.

Along the time dimension, we construct a fixed 120 s window referenced to the trigger time, spanning from 10 s before the GRB trigger to 110 s after it. This design ensures that each sample contains at least 10 s of pre-trigger background while still covering the main emission phase following the trigger. Within this window, photon events are discretized using a 64 ms time bin, yielding a time series of length T=1875. As a result, all samples share a unified temporal length, which facilitates subsequent positional encoding and efficient mini-batch training.

Along the energy dimension, we map the original detector channels into nine universal energy bands to obtain a stable, physically meaningful spectral partition. Specifically, for each photon event, we calculate its central energy from its minimum and maximum boundaries, and then assign photon events to one of the predefined energy bands: 25–50, 50–100, 50–300, 100–300, 100–500, 100–900, 300–500, 300–900, and 500–900 keV. The use of these nine “universal” bands serves two purposes. First, the majority of photon events—and thus most GRB-relevant information—are concentrated within these bands. Second, a multi-band representation mitigates discrepancies in the original channel configurations arising from different detectors or observational conditions, improving cross-sample comparability while preserving the time-varying characteristics of GRB emission across energy ranges.

After processing in both time and energy, we construct a multi-band light-curve matrix. Finally, to stabilize training and prevent the network from being dominated by scale differences among energy bands, we independently standardize each band within the matrix by subtracting its mean and dividing by its standard deviation. This per-band standardization encourages the model to focus on relative temporal variations rather than absolute count levels, which is more suitable for learning transient structures and temporal evolution patterns in GRB detection. When the counts in a band are nearly constant within the window, leading to a zero standard deviation, we set the

normalized values to zero to avoid numerical issues. After these steps, the raw TTE event data are uniformly transformed into a fixed-length, fixed-band, and scale-consistent time–energy sequence representation. Then A simple dimension rearrangement to these time–energy sequence representations produces the corresponding tokens, where each token simultaneously encodes observational evidence across energy bands and across detectors. This tokenization enables the model to learn two key types of dependencies: cross-band/cross-detector coupling within the same time step and temporal evolution patterns across different time steps. Overall, this preprocessing pipeline provides a stable and reliable data foundation for training GRBNet.

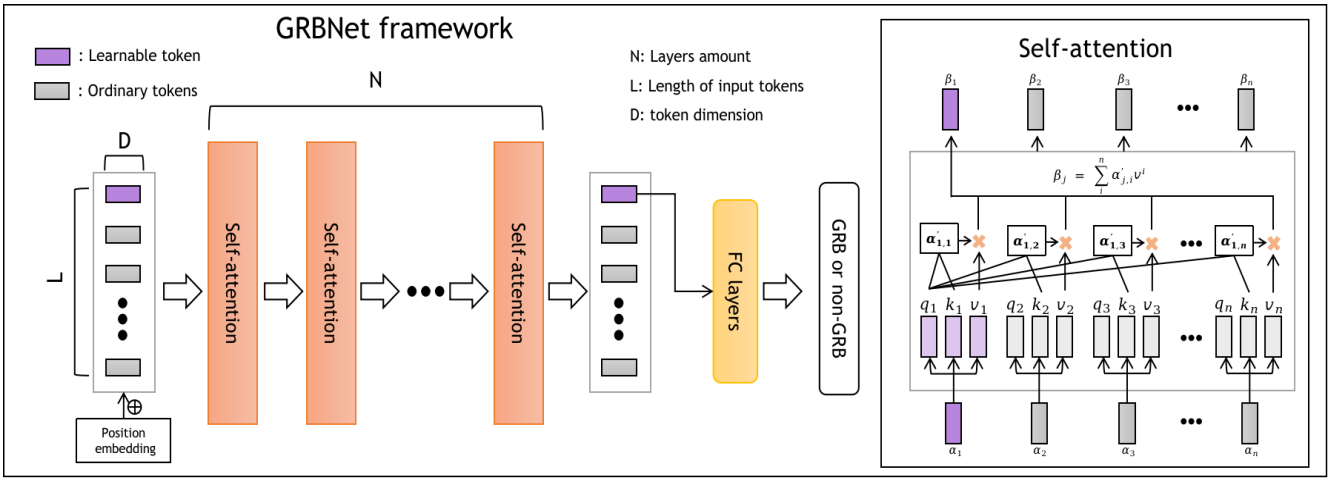

Figure 1. GRBNet Framework. After applying positional embeddings, the input tokens are fed into a self-attention – based feature extraction module to obtain high-level feature tokens. The high-level feature token corresponding to the learnable token is then passed to a fully connected layer for prediction.

B. GRBNet

GRBNet is a self-attention–based network designed for GRB detection. It takes as input the time-series count-feature tokens constructed from multiple instruments observing the same physical signal under a unified energy-band definition, and outputs a binary classification probability for "GRB" versus "non-GRB."

Specifically, we first prepend a learnable token to the input token sequence as a global information aggregation node. The resulting sequence, combined with learnable positional embeddings, is then fed into a feature extraction module composed of eight self-attention layers with four attention heads, where token-wise interactions are performed to exchange, fuse, and extract discriminative representations, yielding higher-level feature tokens. Finally, within the higher-level feature tokens, the token corresponding to the learnable aggregation token is passed to a classification head consisting of three fully connected layers with dimensions 64, 16, 2 respectively, producing a two-dimensional probability vector. The presence of a GRB event is then determined according to the maximum-probability criterion.

With this design, GRBNet enables an end-to-end learnable decision process: it can automatically localize informative temporal segments across the full observation window, integrate evidence from multiple detectors, and model long-range dependencies, making it well suited for GRB detection under real conditions with complex burst morphologies and non-stationary backgrounds.

C. Loss Function

During the training of deep learning models, the loss function is used to quantify the discrepancy between the model's predictions and the ground-truth labels, and it serves as the core objective that guides parameter updates. By minimizing the loss, the model can gradually adjust its parameters so that its predictions increasingly approach the true distribution.

In this work, GRB detection is formulated as a binary classification problem. The model takes as input a multi-energy-band time series acquired

from multiple instruments observing the same physical signal, and GRBNet outputs a two-dimensional probability vector corresponding to the predicted probabilities of "GRB" and "non-GRB." Because cross-entropy directly measures the difference between the predicted probability distribution and the true label distribution, and is both mathematically concise and well-behaved in terms of gradients—with broad empirical validation in classification settings—we adopt cross-entropy [8] as the loss function in this study. Its definition is given as follows:

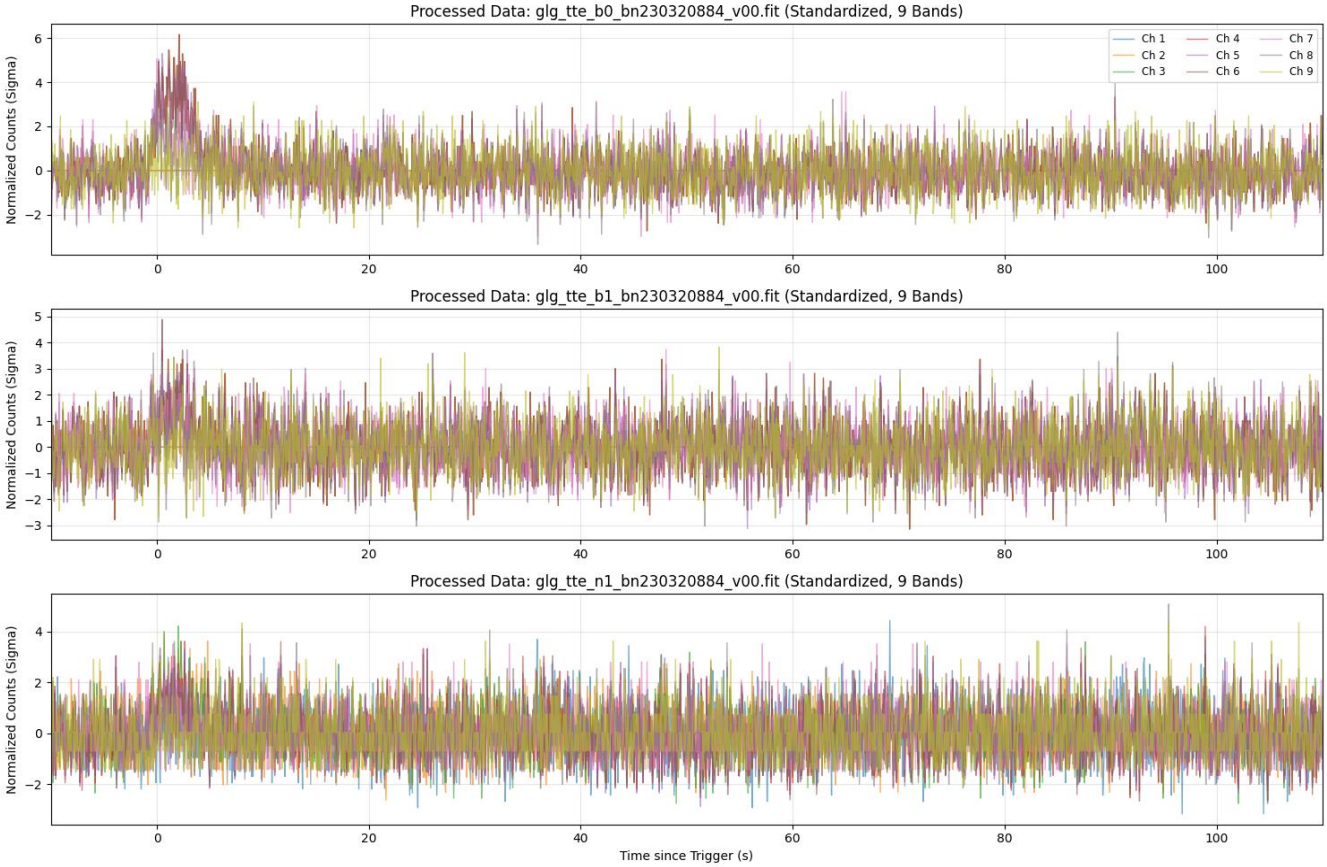

Figure 2. Time–energy sequence representation. This figure illustrates the time – energy sequence data after preprocessing. From top to bottom, the three sequences correspond to the processed observations from the b0, b1, and n1 detectors, respectively.

$$\mathcal{L}_{CE} = -\sum_{k\in\{0,1\}} y_k \log(\hat{y}_k) \qquad (1)$$

Where, $\hat{y}_k$ denotes the predicted probability assigned by the model to class k, and y_k is the corresponding ground-truth label in one-hot form (i.e., y_k=1 when the sample belongs to class k, and y_k =0 otherwise). The physical interpretation of this loss is intuitive: when the model assigns a high probability to the true class, the loss becomes small; conversely, when the model is biased toward an incorrect class, the loss increases substantially. By minimizing this loss, GRBNet progressively learns discriminative features that distinguish GRB signals from background noise, thereby improving overall detection performance.

D. Training Details

We employed the data augmentation methods described in [9], generating a total of 381,000 training samples.

All experiments in this study were conducted in a CUDA-enabled PyTorch environment [10]. Both training and inference were accelerated using a single GPU [11], specifically an NVIDIA TITAN

RTX. This hardware platform provides sufficient memory capacity and computational throughput, allowing the model to perform stable end-to-end training and performance evaluation under long time-series inputs.

During training, at each iteration the model loads a mini-batch of 128 samples, performs a forward pass to obtain predictions, and computes the loss for the current batch. It then carries out backpropagation [12] to obtain parameter gradients, after which the network parameters are updated using the AdamW optimizer [13] with an initial learning rate of 1×10^{-4}, driving the model toward convergence. Meanwhile, we record the loss throughout training and visualize its evolution in real time, which enables intuitive monitoring of training stability—such as oscillations or abnormal divergence—and provides practical guidance for subsequent hyperparameter tuning and convergence analysis.

III. Result Analysis

All test results reported in this work are based on real Fermi observational data. Specifically, the evaluation set consists of 250 real GRB samples and 250 pure-background samples.

In the GRB detection task, for samples that are truly GRBs, if the model successfully identifies them and outputs "GRB," the prediction is counted as a true positive (TP); if the model misclassifies a true GRB as background, it is counted as a false negative (FN). Correspondingly, for samples that are truly background, a correct prediction of background is counted as a true negative (TN), whereas a misclassification as "GRB" is counted as a false positive (FP). Based on these basic quantities, statistical metrics can be further constructed to evaluate detection performance.

Specifically, there are two most critical performance metrics, recall and precision. Recall measures the model's ability to capture true burst events and is defined as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

Where Recall indicates, among all true GRBs, how many are successfully detected by the model. A high recall implies a low miss rate, which is particularly important for transient astronomy because missing a real burst often directly results in losing opportunities for subsequent multi-wavelength follow-up observations. In contrast to recall, precision measures the reliability of the model's "triggered" results and is defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

Where Precision describes, among all samples predicted as GRBs, how many are truly bursts. If the model frequently produces erroneous triggers on background data, the number of FP will increase significantly, leading to a lower precision; when precision approaches 1, it indicates that the model produces almost no false alarms. Recall emphasizes "not missing bursts," whereas precision emphasizes "not misclassifying background," and together they form the key trade-off between detection capability and credibility, serving as essential metrics for evaluating the performance of GRBNet.

Table I shows that GRBNet achieves perfect detection results on the test data, with all 250 GRB samples correctly identified and no FPs or FNs, corresponding to a recall and precision of 1.0. This indicates that the model is able to simultaneously maximize burst detection capability and suppress false triggers under the adopted decision threshold.

TABLE I. TEST RESULT STATISTICS

TP	TN	FP	FN	Recall	Precision
250	250	0	0	1.00	1.00

Table II shows the performance of the model in terms of recall and precision with increasing numbers of observation devices. As the number of devices fused increases, both recall and precision improve steadily. With just one device, the model achieves a recall of 0.91 and precision of 0.95. Upon integrating data from two devices, both metrics improve further to 0.98 and 0.97, respectively. The performance continues to

enhance as more devices are incorporated, reaching perfect scores (1.00 recall and 1.00 precision) when three devices are used. These results clearly demonstrate that the fusion of multiple observation devices significantly enhances the model's ability to detect GRBs, especially in terms of both sensitivity and accuracy, highlighting the importance of multi-instrument data integration in improving the detection performance.

TABLE II. RESULTS WITH VARYING DEVICE AMOUNT

Device amount	Recall	Precision
1	0.91	0.95
2	0.98	0.97
3	1.00	1.00

Table III shows the comparison between ResNet and GRBNet with respect to recall and precision when input data from three observation devices are fused. The results indicate that GRBNet outperforms ResNet in both recall and precision. Specifically, GRBNet achieves a recall of 1.00 and a precision of 1.00, whereas ResNet achieves a recall of 0.93 and precision of 0.91. This demonstrates that GRBNet, with its self-attention-based architecture, is better equipped to capture long-range temporal dependencies and complex burst structures, leading to improved detection performance. In contrast, ResNet, which primarily relies on convolutional operations, may struggle to model global temporal patterns and long-range dependencies, which are crucial for accurate GRB detection. These results highlight the effectiveness of GRBNet in handling complex multi-pulse structures and weak bursts, further validating the advantages of self-attention mechanisms for GRB detection.

TABLE III. RESULTS WITH DIFFERENT NETWORK

Device amount	Recall	Precision
ResNet-1D	0.93	0.91
GRBNet	1.00	1.00

IV. CONCLUSIONS

Experimental results demonstrate that the proposed Fusion of Multi-Instrument Time–Energy Data with Self-Attention framework achieves outstanding performance in GRB detection. Specifically, evaluation on a test set consisting of 250 real GRB observations and 250 pure background samples shows that, even under a stringent decision threshold of 0.99, the method attains a recall of 1.0 and a precision of 1.0. These results indicate that the proposed approach is capable of reliably detecting weak bursts and complex multi-pulse structures, while effectively suppressing false alarms under conditions of strong background noise and pronounced background variability.

Despite the promising performance achieved, the proposed method still has aspects that can be further improved. The current model takes input data constructed from a single fixed time window and a unified energy-band division. As a result, these inputs may not fully capture the relevant information contained in spectral evolution processes. In addition, the model has not yet been adapted to an operational online triggering system, and its real-time performance and robustness in continuous data streams still require further validation.

In future work, we will improve both the input representation and model architecture to better account for spectral evolution. Specifically, we plan to introduce physically motivated multi-scale attention mechanisms to capture both millisecond-scale spikes and longer-duration components, and to develop more flexible time–energy representations beyond a single fixed window and unified energy-band scheme. We also aim to integrate GRBNet into a real-time triggering pipeline to evaluate its performance in continuous data streams.

V. REFERENCES

[1] Gehrels, Neil, and Péter Mészáros. "Gamma-ray bursts." Science 337, no. 6097 (2012): 932-936.

[2] Dilillo, Giuseppe, Kes Ward, Idris A. Eckley, Paul Fearnhead, Riccardo Crupi, Yuri Evangelista, Andrea Vacchi, and Fabrizio Fiore. "Gamma-ray burst detection with Poisson-FOCuS and other trigger algorithms." The Astrophysical Journal 962, no. 2 (2024): 137.

[3] Li, X. Q., X. Y. Wen, Z. H. An, C. Cai, Z. Chang, G. Chen, C. Chen et al. "The technology for detection of gamma-ray burst with GECAM satellite." Radiation Detection Technology and Methods 6, no. 1 (2022): 12-25.

[4] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." Advances in neural information processing systems 30 (2017).

[5] Keles, Feyza Duman, Pruthuvi Mahesakya Wijewardena, and Chinmay Hegde. "On the computational complexity of self-attention." In International conference on algorithmic learning theory, pp. 597-619. PMLR, 2023.

[6] Zhao, Hengshuang, Jiaya Jia, and Vladlen Koltun. "Exploring self-attention for image recognition." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 10076-10085. 2020.

[7] Zhang, Han, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena. "Self-attention generative adversarial networks." In International conference on machine learning, pp. 7354-7363. PMLR, 2019.

[8] Mao, Anqi, Mehryar Mohri, and Yutao Zhong. "Cross-entropy loss functions: Theoretical analysis and applications." In International conference on Machine learning, pp. 23803-23828. pmlr, 2023.

[9] Zhang, Peng, Bing Li, Ren-zhou Gui, Shao-lin Xiong, Yu Wang, Yan-qiu Zhang, Chen-wei Wang et al. "Advancing Gamma-Ray Burst Identification through Transfer Learning with Convolutional Neural Networks." arXiv preprint arXiv:2408.13598 (2024).

[10] Paszke, Adam, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen et al. "Pytorch: An imperative style, high-performance deep learning library." Advances in neural information processing systems 32 (2019).

[11] Owens, John D., Mike Houston, David Luebke, Simon Green, John E. Stone, and James C. Phillips. "GPU computing." Proceedings of the IEEE 96, no. 5 (2008): 879-899.

[12] Cilimkovic, Mirza. "Neural networks and back propagation algorithm." Institute of Technology Blanchardstown, Blanchardstown Road North Dublin 15, no. 1 (2015): 18.

[13] Llugsi, Ricardo, Samira El Yacoubi, Allyx Fontaine, and Pablo Lupera. "Comparison between Adam, AdaMax and Adam W optimizers to implement a Weather Forecast based on Neural Networks for the Andean city of Quito." In 2021 IEEE Fifth Ecuador Technical Chapters Meeting (ETCM), pp. 1-6. IEEE, 2021.