PROCEEDINGS OF SCIENCE

Deep Learning for Detecting Gravitational Waves from Compact Binary Coalescences and Its Visualization by Grad-CAM

Seiya Sasaoka,[a,*] **Yilun Hou,**[a] **Diego Dominguez,**[a] **Suyog Garg,**[b,c] **Naoki Koyama,**[d] **Yusuke Sakai,**[e] **Yuto Omae,**[f] **Kentaro Somiya**[a] **and Hirotaka Takahashi**[b,e]

[a] *Department of Physics, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8551, Japan*
[b] *Institute of Cosmic Ray Research (ICRR), The University of Tokyo, 5-1-5 Kashiwa-no-Ha, Kashiwa City, Chiba 277-8582, Japan*
[c] *Department of Physics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*
[d] *Graduate School of Science and Technology, Niigata University, 8050 Ikarashi-2-no-cho, Nishi-ku, Niigata City, Niigata 950-2181, Japan*
[e] *Research Center for Space Science, Advanced Research Laboratories, Tokyo City University, 3-3-1 Ushikubo-Nishi, Tsuzuki-Ku, Yokohama, Kanagawa 224-8551, Japan*
[f] *Department of Industrial Engineering and Management, Nihon University, 1-2-1 Izumi-cho, Narashino, Chiba 275-8575, Japan*

E-mail: sasaoka@gw.phys.titech.ac.jp, elon@gw.phys.titech.ac.jp, diego@gw.phys.titech.ac.jp, gargsuyog@g.ecc.u-tokyo.ac.jp, f221004a@mail.cc.niigata-u.ac.jp, g2191402@tcu.ac.jp, oomae.yuuto@nihon-u.ac.jp, somiya@phys.titech.ac.jp, hirotaka@tcu.ac.jp

The field of gravitational wave astronomy has made remarkable progress in recent years, with 90 successful detections by Advanced LIGO and Advanced Virgo in three observing runs. The use of deep learning to analyze gravitational wave data is an active area of research with the potential to improve our ability to detect and study these signals. However, the inherent black-box nature of deep learning models poses challenges in interpreting their predictions. To address this, we applied gradient-weighted class activation mapping technique to visualize our 4-class classification model trained on signals from binary black hole mergers, neutron star-black hole mergers, binary neutron star mergers, and noise. The visualization allows us to gain insight into which part of the strain was most influential in the model's predictions. The visualized maps indicated that as the signal duration increased, the model prioritized data before the merger time.

38th International Cosmic Ray Conference (ICRC2023)
26 July - 3 August, 2023
Nagoya, Japan

ICRC2023
38th International Cosmic Ray Conference
The Astroparticle Physics Conference

*Speaker

PoS(ICRC2023)1498

1. Introduction

Since the first detection of gravitational waves (GWs) from a binary black hole (BBH) merger by Advanced LIGO in 2015 [1], 90 events from compact binary coalescences (CBCs) have been detected in three observing runs by Advanced LIGO and Advanced Virgo, which include two neutron star-black hole (NSBH) mergers and two binary newtron star (BNS) mergers [2]. Detecting GWs, electromagnetic waves and neutrinos from mergers of binary systems with neutron stars is crucial for elucidating the physical properties of neutron star interiors, encoded in their equation of state.

GWs from CBC sources are primarily detected using matched filtering technique, which uses analytical waveforms. In recent years, deep learning, especially convolutional neural network (CNN) has been increasingly applied to various tasks in the analysis of GWs, following the pioneering study by George and Huerta [3, 4], such as Nousi *et al.* [5], which outperformed the existing pipelines in BBH searches under certain conditions, and Qiu *et al.* [6], which demonstrated CNN for identifying all CBC sources.

One of the drawbacks of deep learning is that even if an accurate model is obtained, we do not know how it estimated the result. To overcome this, class activation mapping (CAM) technique [7] was proposed to help us to identify the region in the input CNN is looking at while making a prediction. Gradient-weighted class activation mapping (Grad-CAM) [8] is a generalization of CAM and is applicable to various types of CNNs.

In this study, we train a one-dimensional CNN to detect and classify GWs from CBC sources using whitened timeseries as an input. We then visualize the trained model using Grad-CAM to see which regions in the input affect the model's prediction.

The remainder of this paper is structured as follows. In Sec. 2, we describe our datasets, CNN model and Grad-CAM technique. The results and the visualization of the model are shown in Sec. 3. We conclude the paper in Sec. 4.

2. Method

Our CNN model is trained to classify strains at three detector LIGO H1, LIGO L1 and Virgo into four class: noise, BBH, NSBH and BNS. This section describes the data and the model used in this analysis, and briefly explains the CAM technique.

2.1 Data set

To train and test our model, we use non-spinning CBC waveforms and inject them into noise obtained from O3 real data of Advanced LIGO and Advanced Virgo, which are available at Gravitational Wave Open Science Center [9]. To generate BBH samples, we use `SEOBNRv4` approximant [10] with component masses uniformly sampled between 5 and 80 $M_\odot$. `SpinTaylorT4` approximant [11] is used for both BNS and NSBH samples, with NSBH component masses ranging between 1 and 2 $M_\odot$ for NS, and 5 and 35 $M_\odot$ for BH. For BNS waveforms, component masses are uniformly sampled between 1 and 2 $M_\odot$. These waveforms and noise are sampled with a sampling rate of 4096 Hz, and four seconds of data with the merger time uniformly located in the last 0.1 seconds are used. Location of the source is randomly selected from all sky and gravitational wave amplitude is computed, considering the antenna pattern functions and the delays in arrival time of detectors.

For noise samples and background noise for signal samples, O3 real data from GPS time 1238163456 to 1239879680 at LIGO H1, LIGO L1 and Virgo was used for training, validation and test set. Signal samples are rescaled in order that the matched filter signal-to-noise ratio (SNR) with three detectors network is between 8 and 24. Each signal sample and noise sample is whitened in frequency domain. In total, we obtain 96000 training samples, 96000 validation samples and 544000 test samples.

2.2 Model

Our model is one-dimensional ResNet-54 whose input is a whitened timeseries, which was used in Nousi *et al.* [5], but we modified it to use two fully connected layers instead of convolutional layers in the last of the network so that we can apply CAM technique. Deep adaptive input normalization layer [12] is also used in this analysis as used in Nousi *et al.* [5]. For training the models, cross entropy is used as the loss function and Adam optimizer with a learning rate of 10^{-3} is used to optimize the weights. In the training, we adopt curriculum learning as a strategy to enhance the model and accelerate the training by starting from inputting high SNR samples and gradually adding lower SNR samples. We trained the model for 300 epochs with a mini-batch size of 1024.

2.3 Grad-CAM

After training the model, we apply Grad-CAM to visualize the regions in the input that influenced the model's prediction. The CAM value is computed as a linear sum of feature maps, which are the output of the last of the convolutional layer, and weighted parameters. In Grad-CAM, the gradients of the predicted score of the class of interest with respect to the feature maps are used as weighted parameters w_c^k. The ReLU function is then applied to extract only features that have a positive influence on the predictions. The resulting map of class c is expressed as

$$w_k^c = \sum_{i,j} \frac{\partial Y^c}{\partial A_{ij}^k}, \tag{1}$$

$$M^c = \mathrm{ReLU}\left(\sum_k w_k^c A^k\right). \tag{2}$$

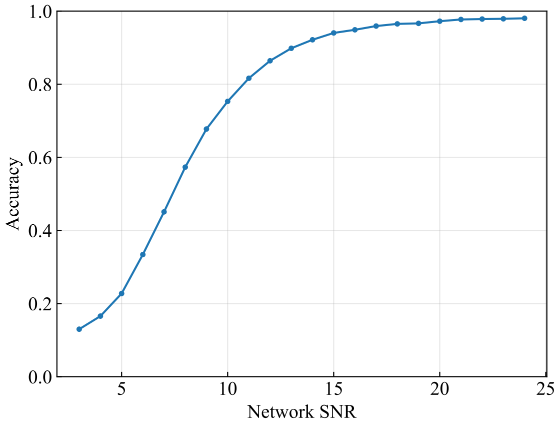
3. Results

This section describes the classification performance of our model and discusses the model's decision-making process using Grad-CAM values.

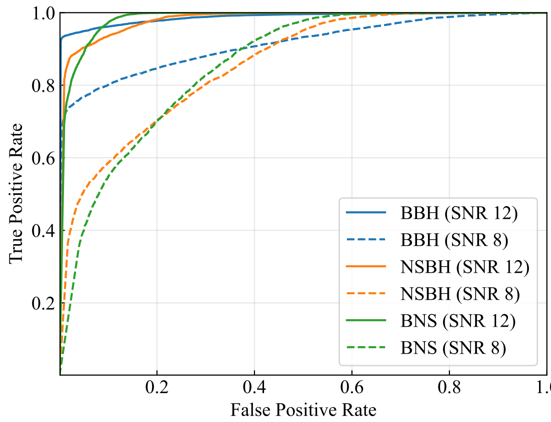
3.1 Classification performance

Figure 1 shows the performance of our trained model for the test set. We plotted accuracy over network SNR and receiver operating characteristic (ROC) curve for each signal. The classification accuracy exceeds 90% for samples with SNR above 13. The ROC curves show that when the false positive rate is set to a sufficiently small value, the true positive rate is high for BBH, NSBH and BNS, in that order, as expected from their ampltitudes.

PoS(ICRC2023)1498



(a) Accuracy vs. SNR curve.



(b) ROC curve of the three signals at SNR 12 and 8.

Figure 1: Classification performane of our model.

3.2 Grad-CAM visualization

Figure 2 shows test samples of noise, BBH, NSBH and BNS and corresponding CAM values produced by Grad-CAM. The SNR of each signal sample is set to 15 and all of them are correctly classified by our model. For BBH sample, the CAM takes large values around the merger time, which means that the model's prediction that the sample is BBH is based on the input around the merger. On the other hand, for NSBH sample, CAM takes the highest values before the merger. The CAM values of the BNS sample is lower than BBH and NSBH, and the time at which the CAM takes its maximum is even earlier than in the case of the NSBH sample. From these signal samples, it is said that as the signal gets longer, the data that the model considers important comes before the merger time.

In the case of the noise sample, the CAM values are almost constant, unlike the signal samples. This means that the model did not find any important parts in the input and predicted that this sample does not contain any signals.

4. Conclusions

In this study, we trained a one-dimensional CNN to detect and classify GWs from CBC sources using whitened time-series as input. We then applied the Grad-CAM technique to visualize the regions in the input that influenced the model's prediction.

The classification performance of our model on the test set was promising, achieving more than 90% accuracy with SNR over 13.

The visualization via Grad-CAM shed light on the CNN's decision-making process. For BBH signals, the input data around the merger time played a crucial role in predicting BBH events. Conversely, for NSBH signals, the model emphasized data preceding the merger, suggesting the importance of the inspiral phase for this class. For BNS signals, multiple peaks appeared during the inspiral phase. These results suggest that as the duration of the signal increased, the model placed greater emphasis on the data preceding the merger time.

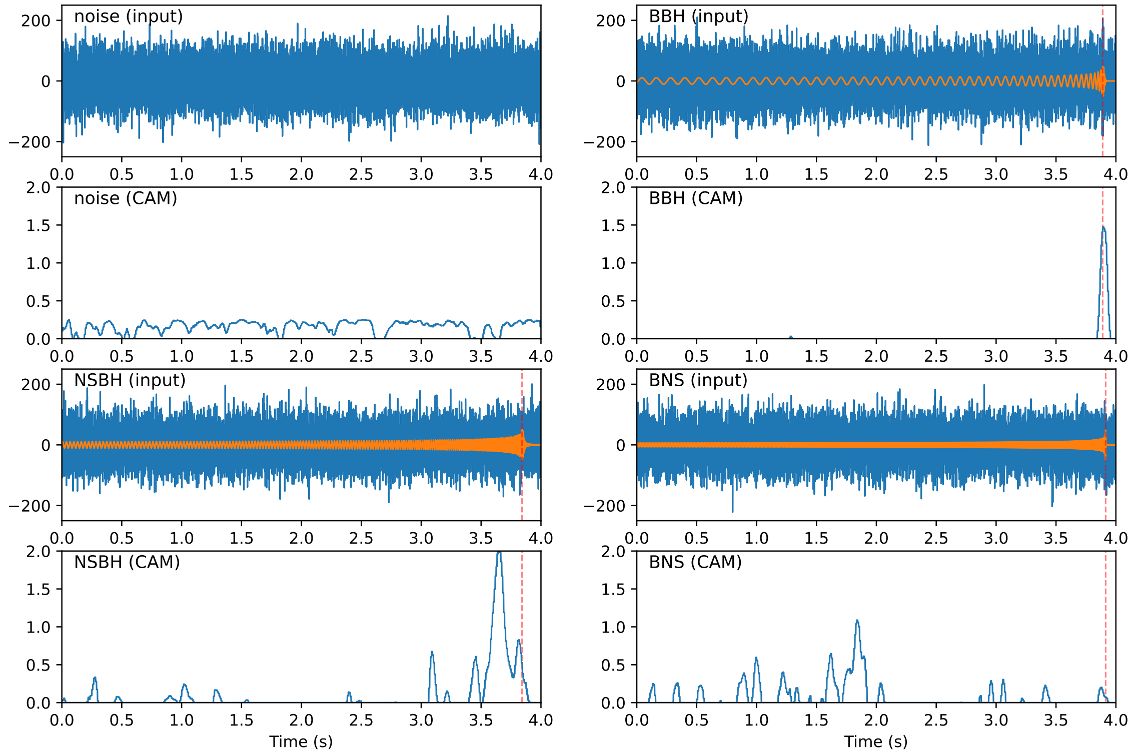


Figure 2: Test samples of noise, BBH, NSBH and BNS and Grad-CAM visualizations. The red dashed line in each signal sample shows the merger time. The SNR of the signal samples is 15.

One of the future works would be to consider training a model on spinning CBC sources and see how the Grad-CAM features would be changed from those plotted in this paper. We would also like to train a two-dimensional CNN to compare the regions used for the prediction with those used by the one-dimensional one.

Acknowledgments

This research was supported in part by JSPS Grant-in-Aid for Scientific Research [No. 22H01228 (K. Somiya), and Nos. 19H01901, 23H01176 and 23H04520 (H. Takahashi)]. This research was also supported by the Joint Research Program of the Institute for Cosmic Ray Research, University of Tokyo and Tokyo City University Prioritized Studies. This research has made use of data or software obtained from the Gravitational Wave Open Science Center (gwosc.org), a service of the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation, as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. KAGRA is supported by Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan Society for the Promotion of Science (JSPS) in Japan; National Research Foundation (NRF) and Ministry of Science and ICT (MSIT) in Korea; Academia Sinica (AS) and National Science and Technology Council (NSTC) in Taiwan.

References

[1] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *Observation of Gravitational Waves from a Binary Black Hole Merger*, *Phys. Rev. Lett.* **116**, 061101 (2016) [gr-qc/1602.03837].

[2] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), *GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run*, arXiv:2111.03606 (2021).

[3] D. George and E. A. Huerta, *Deep neural networks to enable real-time multimessenger astrophysics*, *Phys. Rev. D* **97**, 044039 (2018) [astro-ph/1701.00008].

[4] D. George and E. A. Huerta, *Deep Learning for real-time gravitational wave detection and parameter estimation: Results with Advanced LIGO data*, *Phys. Lett. B* **778**, 64 (2018) [gr-qc/1711.03121].

[5] P. Nousi, A. E. Koloniari, N. Passalis, P. Iosif, N. Stergioulas and A. Tefas, *Deep Residual Networks for Gravitational Wave Detection*, *Phys. Rev. D* **108**, 024022 (2023) [gr-qc/2211.01520].

[6] R. Qiu, P. G. Krastev, K. Gill and E. Berger, *Deep learning detection and classification of gravitational waves from neutron star-black hole mergers*, *Phys. Lett. B* **840**, 137850 (2023) [astro-ph/2210.15888].

[7] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva and A. Torralba, *Learning Deep Features for Discriminative Localization*, arXiv:1512.04150 (2015).

[8] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, *Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization*, in *2017 IEEE International Conference on Computer Vision (ICCV)* , pp. 618–626 [cs/1610.02391].

[9] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration and KAGRA Collaboration), *Open data from the third observing run of LIGO, Virgo, KAGRA and GEO*, arXiv:2302.03676 (2023).

[10] A. Bohé *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* **95**, 044028 (2017) [gr-qc/1611.03703].

[11] R. Sturani, S. Fischetti, L. Cadonati, G. M. Guidi, J. Healy, D. Shoemaker and A. Viceré, *Complete phenomenological gravitational waveforms from spinning coalescing binaries*, *J. Phys.: Conf. Ser.* **243**, 012007 (2010) [gr-qc/1005.0551].

[12] N. Passalis, A. Tefas, J. Kanniainen, M. Gabbouj and A. Iosifidis, *Deep Adaptive Input Normalization for Time Series Forecasting*, in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 9, pp. 3760-3765, Sept. 2020 [q-fin/1902.07892].