Deep Learning for Glitch Classification in Gravitational Waves: a Comparison of CNN-BiLSTM and CNN-SLP Models

Federico Pivotto, Fabrizio Genilotti, Leonardo Egidati

University of Padua

Abstract

In many fields of physics the gravitational waves are studied to deepen the understanding of the fundamentals of physics and of many astronomic processes. The detection of gravitational waves can be achieved by laser-interferometric detectors, which are sensitive to changes in distance under the scale of atomic nuclei. Although the instruments are isolated from non-astrophysical noise, the detectors are still susceptible to instrumental and environmental noise, which can cause gravitational waves to be affected by phenomena called ”glitches” that hinder research. Building on the Gravity Spy project, which aims to classify glitches affecting gravitational wave detectors into morphological families, this paper proposes a deep learning technique to make a contribution to the accurate classification of these glitch phenomena. In this work, the Gravity Spy dataset is presented and used for the classification task. The dataset consists of time-frequency images of glitches, organized in 22 classes based on glitch morphology. In particular, for each gravitational wave, the dataset includes a set of four images captured with four different time windows. In this work we compare two deep learning network designs for glitch classification. Both approaches consists in an initial set of three CNNs, each processing images associated to a specific time window to extract feature vectors. On the second part, which performs the classification task, the first approach uses a Single-layer Perceptron (SLP) network, while the second approach uses a Bidirectional LSTM (BiLSTM) network. The proposals achieve high performance score while converging quickly, reducing training time and allowing room for future improvements.

Introduction

In physics, in particular in astrophysics, gravitational waves are essential for observing and understanding cosmic phenomena that are difficult to observe otherwise, such as the merge of two black holes. In recent years, advanced instruments like the Laser Interferometer Gravitational-Wave Observatory (LIGO) have been developed for the detection of these waves, enhancing the ability to study the universe. These instruments have the capability to sense distance variations at the sub-atomic level, enabling precise measurements of gravitational wave characteristics. However, their exceptional sensitivity also makes them susceptible to non-astrophysical noise, specifically from instrumental and environmental sources. The random noise can sometimes interfere with the detection of gravitational waves, leading to phenomenon known as a ”glitch”. To address this issue, researchers aim to recognize the types of glitches in order to study their origin and develop glitch removal methods.

In the last years, the advancements of artificial intelligence have enabled to tackle the problem using deep learning techniques. Specifically, one of the leading projects addressing this issue is the Gravity Spy project, which employs machine learning and deep learning to accurately classify glitches in gravitational waves into morphological families. In relation to the Gravity Spy project (Bahaadini et al. 2018), S. Bahaadini et al. have proposed a deep learning approach for the accurate classification of glitches in gravitational waves. The proposed method consists in an ensemble model composed by a set of three CNNs and a SVM, each classifying a gravitational wave spectrogram (captured within a specific time window) affected by glitch, with the final decision being made using a majority voting technique. This model has achieved a test accuracy of 98.21%.

Building on this work, this paper proposes and compares two deep learning approaches for assessing the glitch classification task: the CNN-BiLSTM model and the CNN-SLP model. This work aims to contribute to future advancements in the Gravity Spy project for the accurate identification of glitches.

Gravity Spy Dataset

The Gravity Spy dataset (Bahaadini et al. 2018) is composed by time-frequency spectrogram images of gravitational waves affected by random noise, causing the presence of glitches. Each gravitational wave is associated with a set of four time-frequency images, corresponding to different time window lengths: $\pm 0.25s$, $\pm 0.5s$, $\pm 1s$, and $\pm 2s$, for simplicity named View 1, View 2, View 3 and View 4.

The first version of this dataset includes 7881 instances with labels from 22 classes, associated with morphological families. Future versions might merge, add or remove families according to new studies of the glitches nature on gravitational waves. The dataset is divided in four portions, with each portion corresponding to a specific time window.

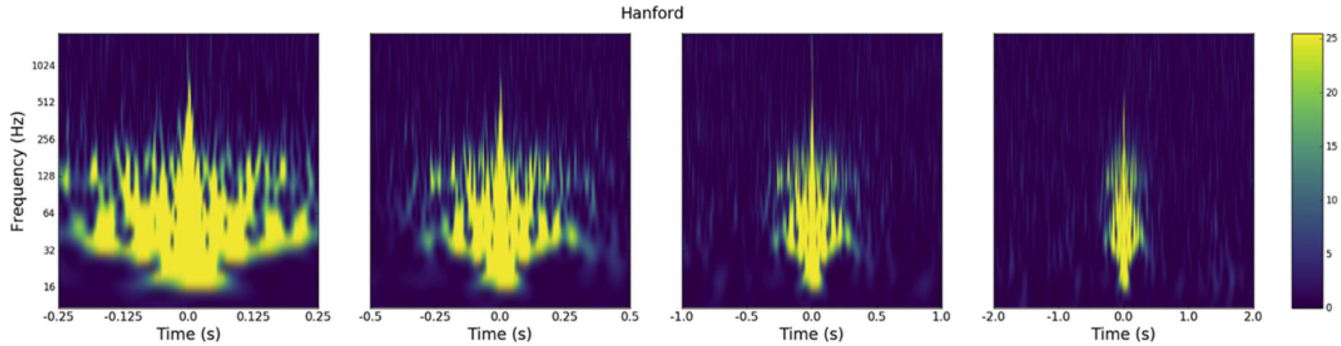

Figure 1: Time-frequency images of gravitational waves of the Hanford morphological class (Bahaadini et al. 2018)

Network Architectures

The network consists of two main parts. The first part employs three parallel CNNs, each processing the same time-frequency image captured with different time windows and extracting its feature vectors. The second part is responsible for the classification task. It proposes two possible approaches: a Bidirectional LSTM (BiLSTM) to capture dependencies between feature vectors from different time-windows, and a Single Layer Perceptron (SLP) to capture dependencies by aggregating the features of all the time-windows.

Table 1: Original network and parallel CNN architecture comparison

Original network	**Parallel CNN**
Input 280x340	Input 280x340
5x5 Convolutional layer (16) with reg.	5x5 Conv layer (16) with BatchNorm
2x2 Maxpooling, Drop-out (0.5)	2x2 Maxpooling
5x5 Convolutional layer (32) with reg.	5x5 Conv layer (32) with BatchNorm
2x2 Maxpooling, Drop-out (0.5)	2x2 Maxpooling
5x5 Convolutional layer (64) with reg.	5x5 Conv layer (64) with BatchNorm
2x2 Maxpooling, Drop-out (0.5)	2x2 Maxpooling
5x5 Convolutional layer (64) with reg.	5x5 Conv layer (64) with BatchNorm
2x2 Maxpooling, Drop-out (0.5)	2x2 Maxpooling
Fully connected (256), Drop-out (0.5)	Fully connected (256) with BatchNorm
	Drop-out (0.5)
Softmax (22)	

Parallel CNNs

The main structure of the network defined by S. Bahaadini et al. for the Gravity Spy project (Bahaadini et al. 2018) was preserved. The architecture consists of four 2D convolutional layers with 2D max-pooling and two fully-connected layers. As shown in the Table 1, in each CNN batch normalization is applied after every convolutional layer to enhance the training stability and accelerate convergence. In order to simplify the training phase, the weight decay factor applied is ignored with respect to the original network. Additionally, a dropout layer is preserved only at the end of the first fully-connected layer to further improve the training convergence and avoid overfitting.

BiLSTM and SLP

Once the feature vectors have been extracted from the gravitational wave images, two classification approaches are proposed for identifying glitches: Bilateral LSTM (BiLSTM) and Single Layer Perceptron (SLP). The BiLSTM approach considers individual feature vectors to extract dependencies across different time windows. In this specific case, where feature vectors lack a specific order, the bidirectional nature of BiLSTM allows to capture relationships between feature vectors without emphasizing their sequence. To maintain a streamlined network architecture, only a single BiLSTM layer is used, along with the classification layer.
In contrast, SLP approach concatenates all feature vectors into a single input, extracting information from the aggregation of feature vectors. However, unlike BiLSTM, the SLP does not account for dependencies between feature vectors from different time windows.

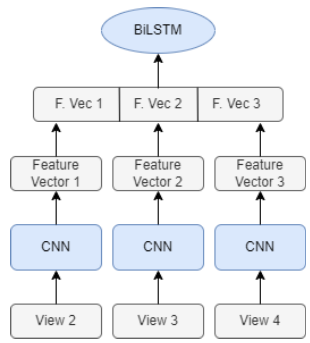

Figure 2: Feature vectors are queued

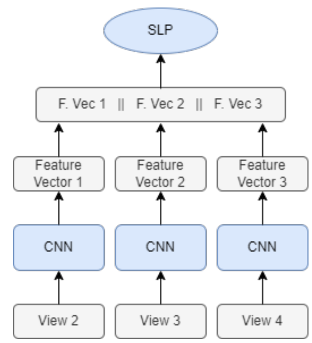

Figure 3: Feature vectors are concatenated

Figure 4: Parallel CNNs with BiLSTM / SLP

Learning Setup

Training and Validation

The training phase consists of two main steps. In the first step, each of the four CNNs is individually trained on a specific portion of dataset, corresponding to glitch images of a specific time window. Afterwards, the trained CNNs are used to extract feature vectors from the dataset, which are then reorganized into four new portions, forming a new dataset of feature vectors. In the second step, this extracted dataset of feature vectors is used to train both the BiLSTM and SLP networks.

Table 2: CNN models test accuracies.

CNN	Test accuracy [%]
View 1	93.90
View 2	95.32
View 3	95.82
View 4	95.65

Table 3: BiLSTM and SLP models test accuracies.

Model	Test accuracy [%]
BiLSTM	96.74
SLP	96.82

Training Configuration During training, the Gravity Spy dataset is split into 6093 train, 667 validation and 1121 test samples. The parallel CNNs, as well as the BiLSTM and the SLP networks, are all trained using the same hyperparameter configuration. Training is performed over 30 epochs, with a mini-batch size of 30 and Stochastic Gradient Descent (SGDM) as the optimizer. Moreover the learning rate is initially set to $1e-3$ and reduced by a factor of 0.1 every 15 epochs.

Results

CNNs Evaluation and Selection The four trained CNN models achieved test accuracies as reported in Table 2. Overall, the CNNs performed a test accuracy over 95.0%, except for the CNN trained on the glitch images with a time window of $\pm 0.25s$, which achieved a score of 93.9%. Due to the lower performance of this model, it was excluded from the design of the final model.

BiLSTM vs SLP The trained BiLSTM and SLP achieved test accuracies as reported in Table 3. As it can be seen, the SLP achieved the highest accuracy of 96.82%, slightly outperforming the BiLSTM, which scored 96.74%. The better performance of SLP may be attributed to its simplicity, allowing it to converge to higher accuracies in lower training time. The comparable performance of both models suggest an absence of overfitting. Furthermore, the confusion matrices of the models shown in Figure 5 and Figure 6 indicate that both models capture similar information across the classes, although each model performs better in some classes.

Conclusion

In conclusion, the glitch classification task on spectrograms of gravitational waves can be successfully solved with high accuracy. Even though the proposed models do not yet match the state-of-art performance, both the BiLSTM and SLP models achieve high test accuracy within a low number of training epochs. In particular, the SLP model achieved the highest test accuracy of 96.82%.

Future advancements could further improve performance of the two models, such as increasing the number of training epochs, adopting a more sophisticated learning rate schedule, calibrating the mini-batch size to enhance the stability and effectiveness of batch normalization, and conducting an additional training phase on the overall model to refine information extraction.

The implementation of our networks is available on GitHub (Pivotto, Genilotti, and Egidati 2024).

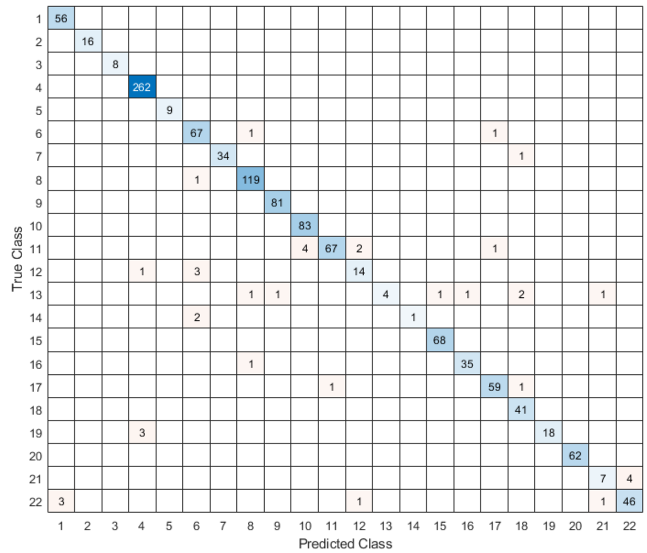

Figure 5: BiLSTM confusion matrix

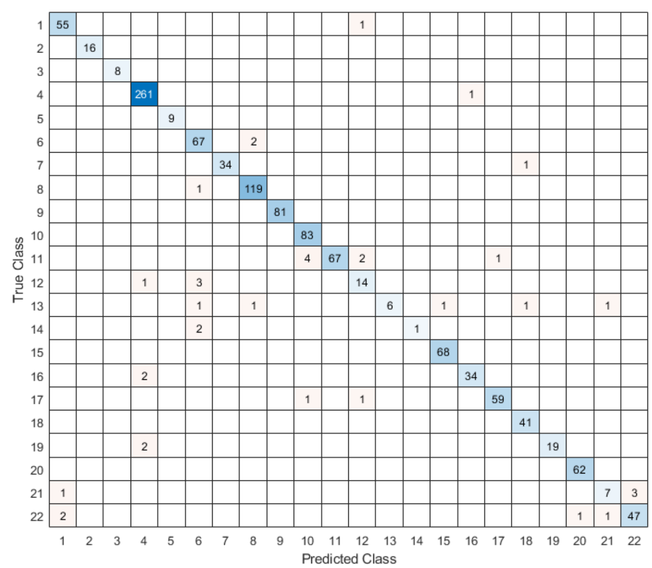

Figure 6: SLP confusion matrix

References

Bahaadini, S.; Noroozi, V.; Rohani, N.; Coughlin, S.; Zevin, M.; Smith, J.; Kalogera, V.; and Katsaggelos, A. 2018. Machine learning for Gravity Spy: Glitch classification and dataset. *Information Sciences*, 444: 172–186.

Pivotto, F.; Genilotti, F.; and Egidati, L. 2024. Gravity Glitch Net. https://github.com/Fabrifio/gravity-glitch-net.