CONVWAVE: Searching for Gravitational Waves with Fully Convolutional Neural Nets

Timothy Gebhard[1,*] tgebhard@tue.mpg.de
Niki Kilbertus[1,2,*] nkilbertus@tue.mpg.de
Giambattista Parascandolo[1,3] gparascandolo@tue.mpg.de
Ian Harry[4] ian.harry@aei.mpg.de
Bernhard Schölkopf[1] bs@tue.mpg.de

[1] MPI for Intelligent Systems [2] Engineering Department, University of Cambridge
[3] Max Planck ETH Center for Learning Systems [4] MPI for Gravitational Physics
[*] Equal contribution

Abstract

The first detection of gravitational waves (GWs) from a binary black hole merger in 2015 was a milestone in modern physics, and just recently awarded with the Nobel Prize. However, despite the unparalleled sensitivity of the LIGO detectors, there still exist challenges in the analysis of the recorded data. We apply CONVWAVE, a dilated, fully convolutional neural net directly on the time series strain data to identify simulated GW signals from black hole mergers in real, non-Gaussian background measurements from the LIGO detectors. CONVWAVE performs well on simulated signals with masses and distances chosen from ranges that contain the estimated parameters of all previously detected real events. It efficiently runs on strain data of arbitrary length from any number of detectors in real time. Through our proposed evaluation approach, it has the potential to develop into a complementary trigger generator in the existing LIGO search pipeline.

1 Introduction

One of the most fascinating predictions of Albert Einstein's theory of general relativity was that its underlying equations allow wave-like solutions (Einstein, 1915, 1916). It was shown that accelerating masses are expected to lose energy by emitting *gravitational waves* (GWs) which propagate as perturbations in the metric of spacetime, stretching and compressing space as they pass through it.

After Hulse & Taylor (1975); Taylor et al. (1979) found indirect experimental evidence for GWs in the 1970s, in 2015 for the first time the Laser Interferometer Gravitational-Wave Observatory (LIGO) directly observed the characteristic *chirp* signature of two black holes approaching and finally merging after circling around each other in a binary black hole (BBH) system 1.3 billion lightyears from Earth (Abbott et al., 2016a). This outstanding achievement was recently awarded the Nobel Prize in Physics 2017. Yet even for violent events such as black hole mergers, the relative length changes on Earth due to gravitational waves are only of $\mathcal{O}(10^{-21})$; the equivalent of measuring the Earth–Sun distance to the precision of the size of a water molecule.

Although the LIGO detectors perform the most precise measurements attained by humankind to this day, many challenges and open problems remain in the analysis of the data. The task is to identify astrophysical signals in the non-Gaussian, non-stationary detector noise from *strain* data, the scalar time series measuring the relative length difference between the two arms of a giant laser interferometer. Non-Gaussian transients, also called *glitches*, stem from, e.g., instabilities in the

Workshop on Deep Learning for Physical Sciences (DLPS 2017), NIPS 2017, Long Beach, CA, USA.

power grid, thermal excitations of the mirror suspensions, or scattered laser light.[1] At the current sensitivity, glitches occur several orders of magnitude more frequently than real GW signals (e.g. Abbott et al., 2016c). For more details about the LIGO instruments, see also Aasi et al. (2015).

Current search pipelines for compact binary coalescences at LIGO are based on *matched filtering* (Allen, 2005; Allen et al., 2012), which by construction is optimal for stationary Gaussian noise. Through various signal consistency tests, the pipeline can also deal with non-Gaussian and non-stationary noise. However, in matched filtering one needs to first process data from each detector individually and afterwards combine the results, which does not scale well in the number of detectors. Increasing this number is crucial for both the sensitivity and the ability to localize the source of astrophysical events. Hence, machine learning can yield two potential improvements: Firstly, it may scale better computationally. Secondly, it can flexibly learn the structure of the specific glitches in each detector, making it easier to distinguish them from actual signals.

1.1 Related work

Most previous applications of machine learning for the detection task use *trigger data* from the matched filtering pipeline as input (Adams et al., 2013; Baker et al., 2015; Hodge, 2014; Kapadia et al., 2017). A trigger is a collection of numbers that describe a potential event, e.g., the time of merger, parameters of the matched template, the signal-to-noise ratio (SNR) or manually tuned test statistics to distinguish glitches from signals. There is also a related body of work on a slightly different task, namely glitch classification and denoising (Powell et al., 2015, 2016; Mukund et al., 2016; Torres-Forné et al., 2016; Zevin et al., 2017; George et al., 2017).

George & Huerta (2016) and Gabbard et al. (2017) are closest to our work, as they also apply 1D-CNNs to detect GWs in strain data. Using real LIGO noise, taken from advanced LIGO's first observing run, George & Huerta (2017) have shown, for the first time, that CNNs can detect real GW signals and estimate their parameters, in particular GW150914, after training on noise from GW151226.[2] They apply a sliding window individually to each detector for prediction on longer recordings than the fixed size input. Our architecture differs in that we do not use fully connected layers, thereby avoiding repeated computation of the same convolutions, and stack recordings from different detectors to a single, multi-channel input. Our present work focuses on how to label and evaluate detections for continuous output, and not on improving the detection sensitivity.

1.2 Contributions

We propose an approach for real-time tagging of multi-detector strain data recordings to detect GWs using a deep, dilated, fully convolutional neural net.[3] Our approach, nicknamed ConvWave, has the following properties: a) It directly processes arbitrary length inputs during training and testing without computational overhead, giving outputs with full time resolution, b) the model inherently handles any number of detectors (no individual application for each detector), c) our method adapts to unseen events and detector characteristics, d) the net can potentially be developed into a real-time, complementary trigger generator for the current search pipeline. To this end, we propose a labeling procedure and detection metrics for continuous predictions. In experiments with simulated signals and real detector recordings from the LIGO detectors, we demonstrate the applicability of ConvWave.

2 Methods

2.1 Data

Our data are in principle created by taking a stretch of real detector noise as background, and adding a simulated GW signal to it ("injection"). For the background noise, we use 4096 s chunks of strain data recorded by the LIGO detectors around the events GW150914 and GW151226 and down-sample them to 2048 Hz. Simulated waveforms are generated with the

[1] For more information about glitches, see e.g. the *GravitySpy* "Field Guide": `https://gravityspy.org`.

[2] The used data are available at the *LIGO Open Science Center* (LOSC): `https://losc.ligo.org/events`. See also Vallisneri et al. (2015).

[3] All code is publicly available at: `https://github.com/nikikilbertus/convwave`.

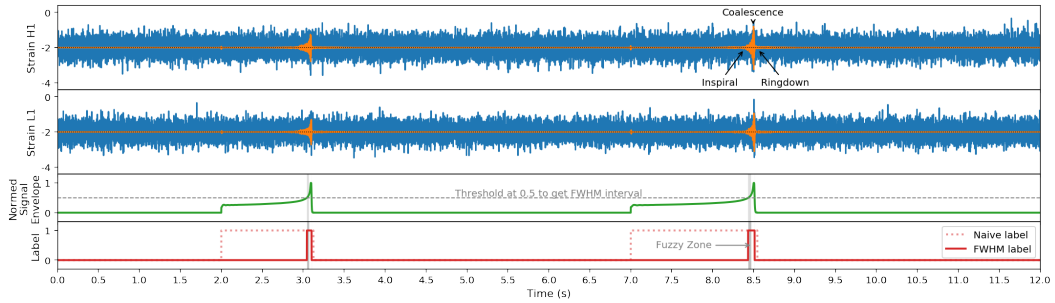

Figure 1: We show a training example with the corresponding labels including the fuzzy zones.

`pycbc.waveform.waveform.get_td_waveform` function of the PyCBC software package (Canton et al., 2014; Usman et al., 2016; Nitz et al., 2017). In our experiments, we use a particular BBH waveform approximant tuned to numerical relativity, `approximant='SEOBNRv4'` (Bohé et al., 2017). We chose the minimum frequency `f_lower=15`. Both masses and the distance are drawn uniformly at random between 2 to 50 solar masses and 100 Mpc to 1700 Mpc respectively. For now, the antenna pattern function for H1/L1 was ignored for simplicity.

We repeatedly select 12 s of recording at random from both the Hanford and Livingston data and inject up to two waveforms, each cropped to a random length uniformly between 1 s to 3 s. Further, for each waveform we introduce a small random difference in injection time between the detectors, consistent with their distance. Different injections do not overlap within individual detectors. We whiten the results using the PSD obtained from the corresponding detector measurements, and apply a band pass from 42 Hz to 800 Hz. In total, we generate $5 \cdot 4096$ such examples of length $d := 12\,\mathrm{s} \cdot 2048\,\mathrm{Hz} = 24\,576$ for both training and testing. The waveforms used for training and testing are sampled from the same parameter space, but do not overlap. We also generate a *label vector* of the same length, marking the parts of the time series containing an injection.

The *naïve label vector* $y \in \{0,1\}^d$ consists of ones during injections, and zeros everywhere else. We use these labels in our first experiments. However, real black hole merger waveforms do not have such a well-defined beginning. Furthermore, in practice it suffices for a detection to tag only the salient *coalescence* around the peak of the waveform, see Figure 1. There is no natural definition of the coalescence period, hence sharp on/off transitions are both artificial and overly demanding. We address this problem by a) defining the coalescence period as the *full width at half maximum* (FWHM) of the waveform envelope computed via a Hilbert transform, and b) adding *fuzzy zones* around the coalescence during training, which are parts of the (predicted) label that are ignored when computing the loss (faintest 20% of the FWHM interval). This helps to avoid the net being "distracted" by negative feedback from the arbitrary edges. Figure 1 also illustrates this procedure.

2.2 Training and Evaluation

Inspired by Google's *WaveNet* (van den Oord et al., 2016), we use a stack of dilated convolutions to get a large receptive field with a reasonable number of layers. We use 12 convolutional layers with exponentially increasing (1, 2, 4, …, 2048) dilations (kernel size 2; 64 kernels each), ELU activations, and batch-normalization, resulting in a receptive field of 4096 (2 s). To preserve the resolution from input to output we use zero-padding. One additional conv-layer with kernel size 1 at both the top and bottom of the stack adjusts the number of channels for the input from 2 to 128, and from 128 to 1 for the output. We train using *Adam* (Kingma & Ba, 2014) and mini-batches of size 16, reducing the initial learning rate of 0.0001 by a factor of $\sqrt{2}$ every 3 epochs. The weighted binary cross-entropy loss uses 0 as a weight for the fuzzy zones, and 1 everywhere else, effectively ignoring the prediction in the fuzzy zones. We implemented ConvWave in *PyTorch*.

Besides the loss, we report the accuracy $\mathrm{Acc} := {}^1\!/_d \cdot \sum_{i=1}^{d} \mathbb{1}\{y_i = \mathrm{round}(\hat{y}_i)\}$, with the prediction $\hat{y} \in [0,1]^d$ and the true label $y \in \{0,1\}^d$. For the following metrics, we post-process the prediction $\hat{y}$ by smoothing it using a rolling average with window size 20. We consider an injection as *detected*, iff $\mathrm{round}(\hat{y})$ is one for more than a fraction of $\theta_{\mathrm{det}} \in [0,1]$ of the FWHM, giving true positives (TP) and false negatives (FN). Further, we define a false positive (FP), or *false alarm*,

Table 1: We report test loss and accuracy using naïve labels on the five datasets (characterized by median, minimum and maximum SNR) with and without curriculum learning (CL), and the injection detection evaluation using the FWHM characterization of coalescences for $\theta_{\text{det}} = 0.1, \theta_{\text{fa}} = 0.9$. Results differ by $< 3\%$ points between the two datasets; we therefore only report for GW151226.

	SNR H1	SNR L1	Naïve Label		Naïve + CL		Coalescence Detection (FWHM)			
Distance (Mpc)	Median $^{\text{Max}}_{\text{Min}}$	Median $^{\text{Max}}_{\text{Min}}$	Loss	Acc.	Loss	Acc.	Loss	Acc.	Det. Ratio	FA Ratio
100–300	$6.53\,^{28.47}_{0.59}$	$6.38\,^{26.64}_{0.59}$	0.112	96.6%	—	—	0.015	99.8%	97.4%	0.4%
250–500	$3.48\,^{11.58}_{0.31}$	$3.43\,^{10.78}_{0.32}$	0.151	95.4%	0.145	95.6%	0.016	99.8%	93.8%	1.0%
400–800	$2.28\,^{7.42}_{0.21}$	$2.23\,^{7.01}_{0.20}$	0.190	94.2%	0.178	94.4%	0.016	99.8%	89.3%	0.7%
700–1200	$1.40\,^{4.15}_{0.13}$	$1.37\,^{3.77}_{0.14}$	0.226	92.6%	0.220	92.8%	0.018	99.7%	75.2%	1.5%
1000–1700	$0.97\,^{2.93}_{0.10}$	$0.95\,^{2.76}_{0.09}$	0.260	91.2%	0.252	91.4%	0.021	99.7%	64.1%	1.6%

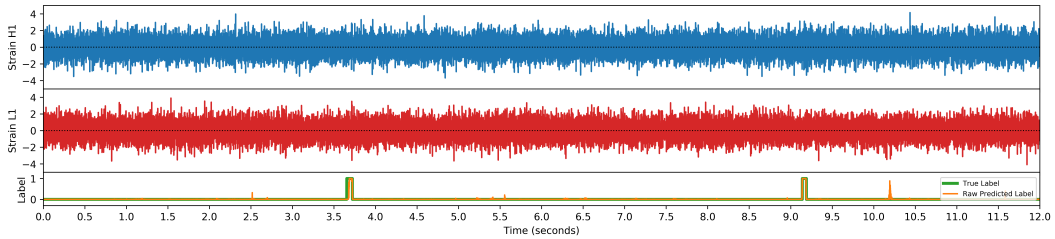

Figure 2: Example FWHM predictions from the 1000–1700 Mpc test set (GW150914).

as a series of ones in round$(\hat{y})$ disagreeing with y on a fraction of more than $\theta_{\text{fa}} \in [0, 1]$.[4] For a meaningful evaluation, we also need a measure for the "faintness" of an event. One natural choice is the peak signal-to-noise ratio (SNR), the ratio of the peak amplitude of the (whitened) waveform to the standard deviation of the (whitened) noise.[5]

3 Experiments and Results

First, we split the overall distance range 100 Mpc to 1700 Mpc into five subranges corresponding to five levels of "difficulty".[6] In Table 1, we summarize the results of training our network for 50 epochs for each range a) from scratch, and b) for curriculum learning, i.e., sequentially transferring the weights from easier to harder datasets. In these first experiments, we predict the *complete* injections; i.e., the label vector is 1 whenever an injection is present (*naïve labels*). In Table 1, we show the average loss and accuracy on the test sets. Curriculum learning yields only marginal improvements.

Next, we evaluate the detection task as described in section 2, i.e, we train and predict (for 50 epochs, using our *fuzzy zones* for both) on the FWHM intervals of the signal envelopes. As for the previous experiments, we performed minimal fine-tuning on our hyper-parameters. In Table 1, we report the loss and accuracy (for comparison), as well as the detection ratio $\text{TP}/(\text{TP} + \text{FN})$ and false alarm ratio $\text{FP}/(\text{TP} + \text{FP})$. The results show that ConvWave reliably recovers coalescences over a wide range of SNRs, while triggering few false alarms. For an example prediction, see Figure 2.

As a final test, we apply the model that we learned on the GW151226 data using FWHM labels (distances 1000–1700 Mpc) to the full 4096 s around the GW150914 event. Again, we smooth the prediction with a rolling average (window size 200). We indeed successfully recover the event at the correct location without a single false alarm, proving that our method is also able to generalize beyond the characteristics of the particular waveforms and noise it was trained on.

4 Conclusion

ConvWave is a fully convolutional neural net architecture with dilated kernels to detect gravitational wave signals from compact binary coalescences in strain measurements from any number of detectors.

[4] There is no notion of true negatives in a continuous prediction.

[5] We remark that this differs from the SNR definition used in the context of matched filtering (Allen, 2005).

[6] Previous detections were estimated at 410, 440, 880, 540 and 40 Mpc (Abbott et al., 2016a,b, 2017a,b,c).

We evaluate CONVWAVE on real measurements from the LIGO Hanford and Livingston detectors and simulated waveforms with mass and distance parameter ranges that include the estimates for all actual events detected so far. Further, we use fuzzy zones as a trick to improve convergence during training, and introduce an evaluation procedure for coalescence detection based on the FWHM of the signal envelope. Following our encouraging preliminary results, we plan to investigate how CONVWAVE can be developed into a complementary trigger generator for the LIGO search pipeline, and properly compare its performance to matched filtering.

References

J. Aasi, B. P. Abbott, R. Abbott, et al. Advanced LIGO. *Classical and Quantum Gravity*, 32(7): 074001, 2015. DOI: 10.1088/0264-9381/32/7/074001.

B. P. Abbott, R. Abbott, T. D. Abbott, et al. Observation of Gravitational Waves from a Binary Black Hole Merger. *Physical Review Letters*, 116:061102, 2016a. DOI: 10.1103/PhysRevLett.116.061102.

B. P. Abbott, R. Abbott, T. D. Abbott, et al. GW151226: Observation of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence. *Physical Review Letters*, 116:241103, 2016b. DOI: 10.1103/PhysRevLett.116.241103.

B. P. Abbott, R. Abbott, T. D. Abbott, et al. Characterization of transient noise in advanced LIGO relevant to gravitational wave signal GW150914. *Classical and Quantum Gravity*, 33(13):134001, 2016c. DOI: 10.1088/0264-9381/33/13/134001.

B. P. Abbott, R. Abbott, T. D. Abbott, et al. GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2. *Physical Review Letters*, 118:221101, 2017a. DOI: 10.1103/PhysRevLett.118.221101.

B. P. Abbott, R. Abbott, T. D. Abbott, et al. GW170814: A Three-Detector Observation of Gravitational Waves from a Binary Black Hole Coalescence. *Physical Review Letters*, 119:141101, 2017b. DOI: 10.1103/PhysRevLett.119.141101.

B. P. Abbott, R. Abbott, T. D. Abbott, et al. GW170817: Observation of gravitational waves from a binary neutron star inspiral. *Physical Review Letters*, 119(16), 2017c. DOI: 10.1103/physrevlett.119.161101.

T. S. Adams, D. Meacher, J. Clark, et al. Gravitational-Wave Detection using Multivariate Analysis. *Physical Review D*, 88(6), 2013. DOI: 10.1103/PhysRevD.88.062006.

B. Allen. A χ^2 time-frequency discriminator for gravitational wave detection. *Physical Review D*, 71 (6), 2005. DOI: 10.1103/physrevd.71.062001.

B. Allen, W. G. Anderson, P. R. Brady, et al. FINDCHIRP: An algorithm for detection of gravitational waves from inspiraling compact binaries. *Physical Review D*, 85(12), 2012. DOI: 10.1103/physrevd.85.122006.

P. T. Baker, S. Caudill, K. A. Hodge, et al. Multivariate classification with random forests for gravitational wave searches of black hole binary coalescence. *Physical Review D*, 91(6), 2015. DOI: 10.1103/physrevd.91.062004.

A. Bohé, L. Shao, A. Taracchini, et al. Improved effective-one-body model of spinning, nonprecessing binary black holes for the era of gravitational-wave astrophysics with advanced detectors. *Physical Review D*, 95(4), 2017. DOI: 10.1103/physrevd.95.044028.

T. Dal Canton et al. Implementing a search for aligned-spin neutron star-black hole systems with advanced ground based gravitational wave detectors. *Physical Review D*, 90(8):082004, 2014. DOI: 10.1103/PhysRevD.90.082004.

A. Einstein. Zur Allgemeinen Relativitätstheorie. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften*, 1915:778–786, 1915.

A. Einstein. Näherungsweise Integration der Feldgleichungen der Gravitation. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften*, 1916:688–696, 1916.

H. Gabbard, F. Hayes, C. Messenger, and M. Williams. Matching Matched-Filtering with Deep Networks for Gravitational wave Astronomy. *In Preparation*, 2017.

D. George and E. A. Huerta. Deep Neural Networks to Enable Real-time Multimessenger Astrophysics. 2016. arXiv: 1701.00008v2.

D. George and E. A. Huerta. Deep Learning for Real-time Gravitational Wave Detection and Parameter Estimation: Results with Advanced LIGO Data. 2017. arXiv: 1711.03121.

D. George, H. Shen, and E. A. Huerta. Deep Transfer Learning: A new deep learning glitch classification method for advanced LIGO. 2017. arXiv: 1706.07446.

K. A. Hodge. *The Search for GravitationalWaves from the Coalescence of Black Hole Binary Systems in Data from the LIGO and Virgo Detectors*. PhD thesis, California Institute of Technology, 2014. URL: http://resolver.caltech.edu/CaltechTHESIS:06022014-104457554.

R. A. Hulse and J. H. Taylor. Discovery of a pulsar in a binary system. *The Astrophysical Journal*, 195:L51, 1975. DOI: 10.1086/181708.

S. J. Kapadia, T. Dent, and T. Dal Canton. A classifier for gravitational-wave inspiral signals in non-ideal single-detector data. 2017. arXiv: 1709.02421.

D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. 2014. arXiv: 1412.6980.

N. Mukund, S. Abraham, S. Kandhasamy, et al. Transient classification in ligo data using difference boosting neural network. 2016. arXiv: 1609.07259v2.

A. Nitz, I. Harry, D. Brown, et al. ligo-cbc/pycbc: Post-O2 Production Release 2, 2017. DOI: 10.5281/zenodo.888262.

J. Powell, D. Trifirò, E. Cuoco, et al. Classification methods for noise transients in advanced gravitational-wave detectors. *Classical and Quantum Gravity*, 32(21):215012, 2015. DOI: 10.1088/0264-9381/32/21/215012.

J. Powell, A. Torres-Forné, R. Lynch, et al. Classification methods for noise transients in advanced gravitational-wave detectors ii: performance tests on advanced ligo data. 2016. arXiv: 1609.06262v2.

J. H. Taylor, L. A. Fowler, and P. M. McCulloch. Measurements of general relativistic effects in the binary pulsar PSR1913+16. *Nature*, 277(5696):437–440, 1979. DOI: 10.1038/277437a0.

A. Torres-Forné, A. Marquina, J. A. Font, and J. M. Ibáñez. Denoising of gravitational wave signals via dictionary learning algorithms. *Physical Review D*, 94(12), 2016. DOI: 10.1103/physrevd.94.124040.

S. A. Usman et al. The PyCBC search for gravitational waves from compact binary coalescence. *Classical and Quantum Gravity*, 33(21):215004, 2016. DOI: 10.1088/0264-9381/33/21/215004.

M. Vallisneri, J. Kanner, R. Williams, A. Weinstein, and B. Stephens. The LIGO Open Science Center. In *Journal of Physics: Conference Series*, volume 610, pp. 012021, 2015. DOI: 10.1088/1742-6596/610/1/012021.

A. van den Oord, S. Dieleman, H. Zen, et al. WaveNet: A Generative Model for Raw Audio. 2016. arXiv: 1609.03499.

M. Zevin, S. Coughlin, S. Bahaadini, et al. Gravity Spy: integrating advanced LIGO detector characterization, machine learning, and citizen science. *Classical and Quantum Gravity*, 34(6): 064003, 2017. DOI: 10.1088/1361-6382/aa5cea.