Searches for Mass-Asymmetric Compact Binary Coalescence Events using Neural Networks in the LIGO/Virgo O3 Period

M. Andrés-Carcasona, M. Martínez et al.,

https://dcc.ligo.org/LIGO-P2200184

COST ACTION CA17137
A NETWORK FOR GRAVITATIONAL WAVES, GEOPHYSICS AND MACHINE LEARNING

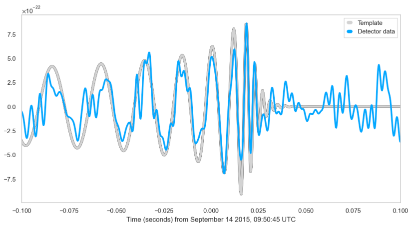
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

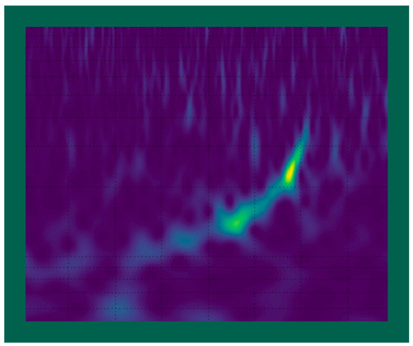
EGO, September 28th 2022

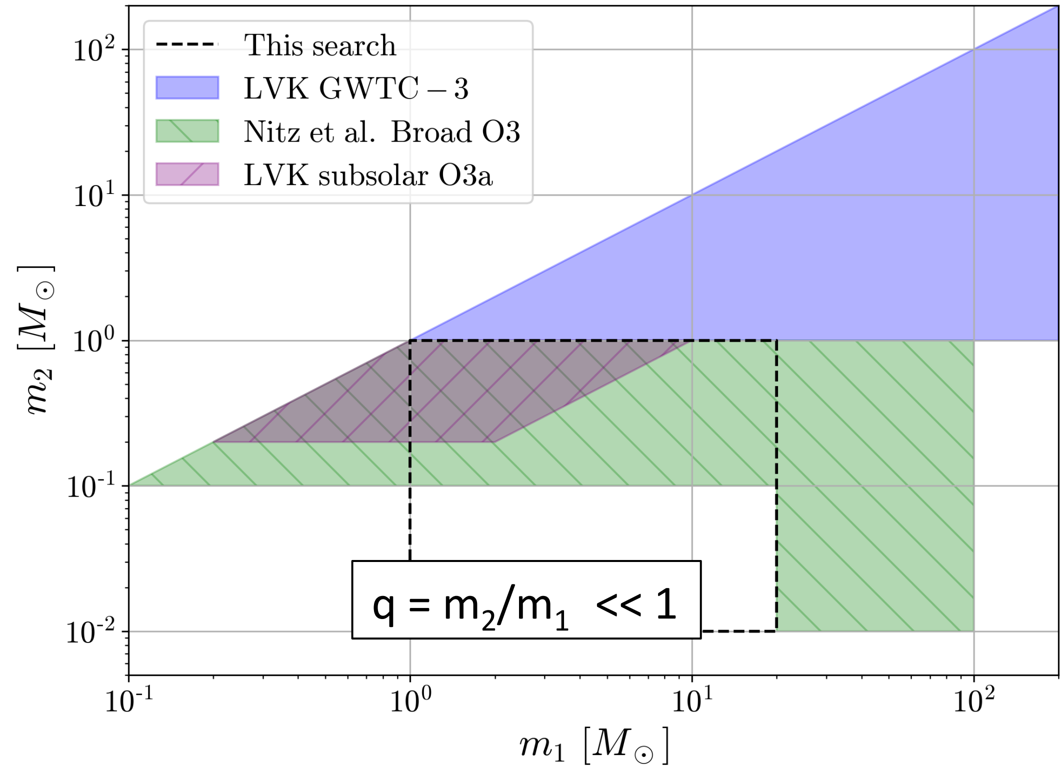
Outline

- Introduction/Motivation
- Data Preparation
- CNN architecture
- Training & Validation
- Results & Interpretations
- Final notes

Introduction/motivation







Inclination	Polarization
$\psi \in [0, \pi/2]$	$\theta_{JN} \in [0, \pi]$
Right Ascension	**Declination**
$\alpha \in [0, 2\pi]$	$\delta \in [-\pi/2, \pi/2]$

Masses $m_1 \in [1, 20]\ M_\odot$, $m_2 \in [0.01, 1]\ M_\odot$ → $q \in [1, 2000]$

$D_L \in [1, 100]\ Mpc$

We extend previous studies using CNNs and spectrograms towards very asymmetric mass configurations motivated by the search for sub solar mass objects (pBHs) in binary systems

Results are based on O3 LVK data set

Data Preparation

We apply the standard Quality requirements (following the work for the GWTC – 3 paper)

Removing data segments around identified GWTC-3 events

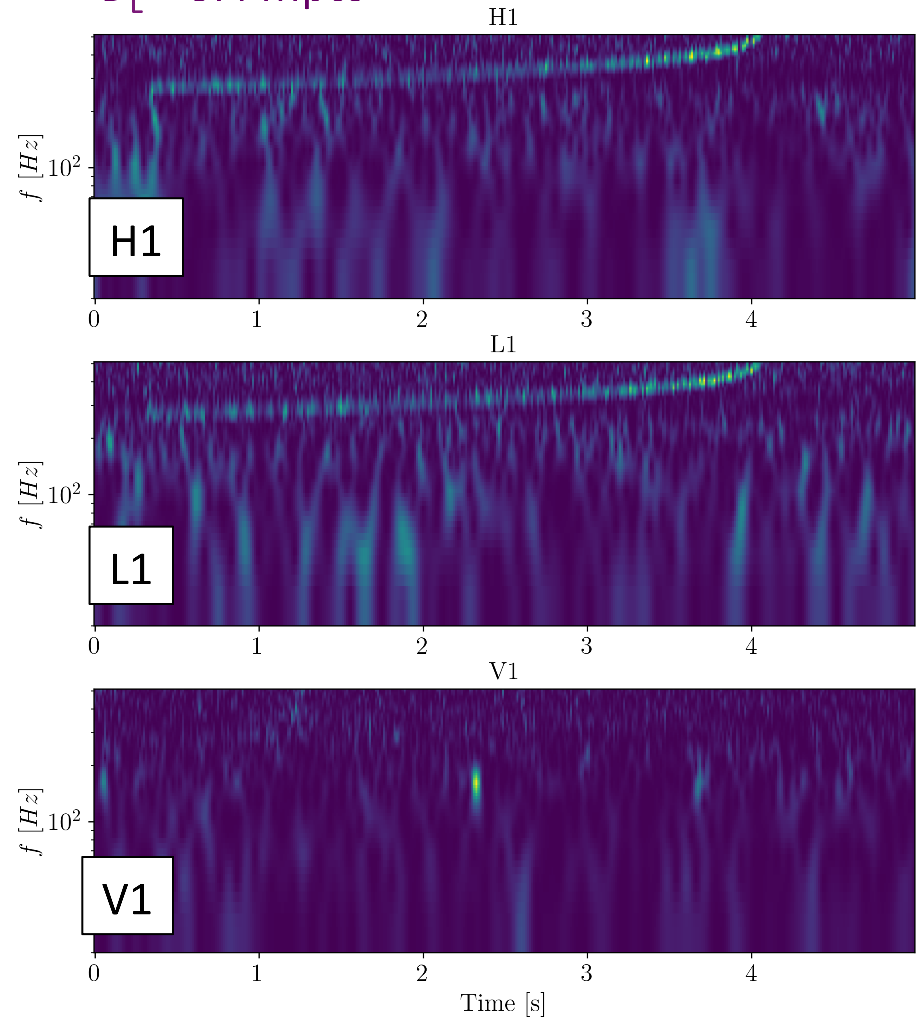
We consider the LIGO/Virgo sample for which the three interferometer are online in physics mode **(this reduces the O3 sample to 155 days)**

2D images (spectrograms) of frequency vs time are constructed in the data (5 s duration each)
→ Data whitened and Q-transformed
→ 45 Hz lower limit to control signal duration

- Background only (50%)
- Background + **injected signals** (50%)

A total of 143.000 images
- 80% for training
- 10% for validation
- 10% for CNN testing

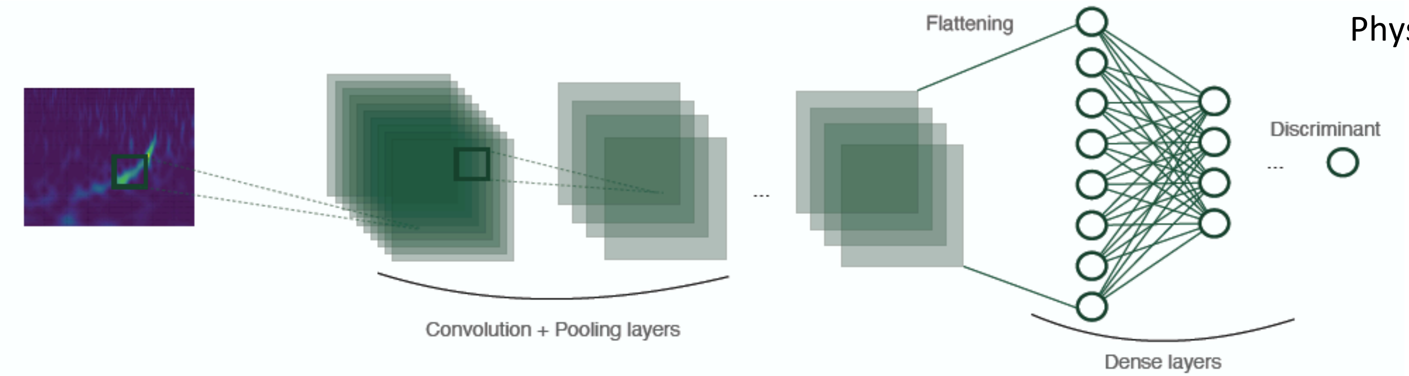
Example of injection with m_1 = 2.6 Msun, m_2 = 0.35 Msun, D_L = 3.4 Mpcs



Signal injection using pyCBC with IMRPhenomD waveforms and no spin

CNN architecture

Phys. Rev. D 103, 062004 (2021)



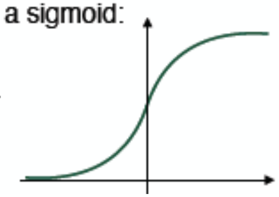
Based on ResNet50

Layer name	Output size	Layer structure
conv1	112×112	7×7, 64, stride 2
conv2_x	56×56	3×3 max pool, stride 2
		[1×1, 64; 3×3, 64; 1×1, 256] ×3
conv3_x	28×28	[1×1, 128; 3×3, 128; 1×1, 512] ×4
conv4_x	14×14	[1×1, 256; 3×3, 256; 1×1, 1024] ×6
conv5_x	7×7	[1×1, 512; 3×3, 512; 1×1, 2048] ×3
	1×1	Global average pool, 1-d fc, sigmoid

Hyper parameters	
Learning rate	0.01
Batch size	32
Number of epochs	10
Optimizer	Adam
Loss function	Binary-cross entropy

The activation function of the last layer is chosen to be a sigmoid:

$$f(x) = \frac{1}{1 + e^{-x}}$$

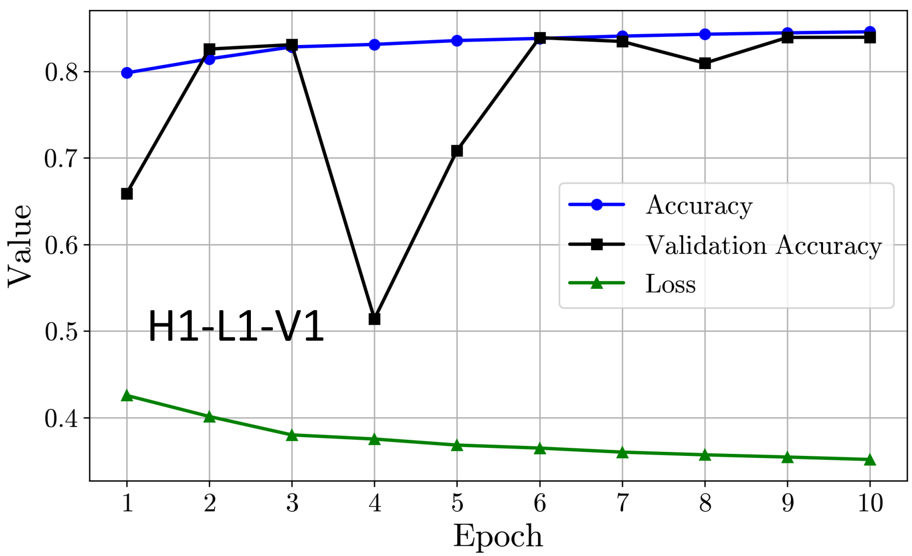


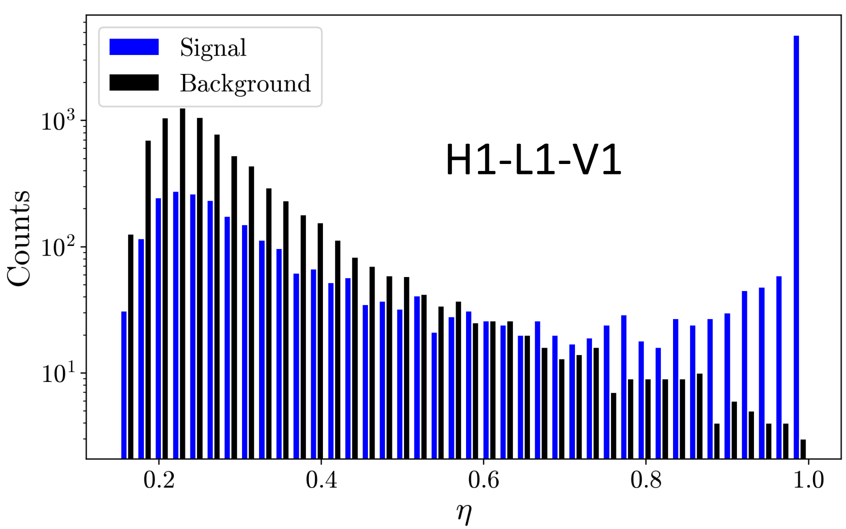
We initially trained 7 different CNNs using as input data from single interferometers, pairs of interferometers and all interferometers together

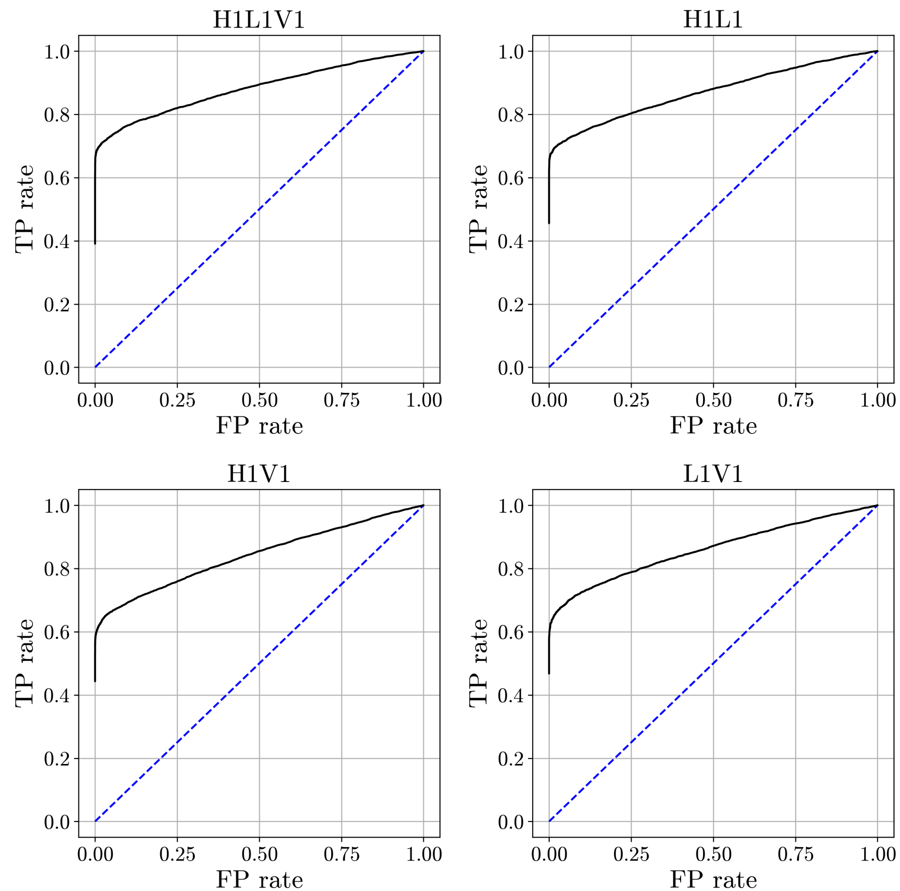
→ The simultaneous use of multiple interferometers help in figuring out correlations across IFTs

→ **We discarded the CNNs using single Interferometer inputs for the final O3 scan**

Training and ROC







There is a healthy evolution of training process
We observe a discriminating power in CNNs

The detection efficiency is limited to ~ 70%
The CNN operating point (η) is determined
by the ROC information **and tolerable fake rate**

FARs and combined CNNs

- We compute the False Alarm Rate (FAR) using the same time slide technique as in the case of matched filtering approach
- → testing O(10^9) images and reaching FAR values of 1/153 years^{-1}

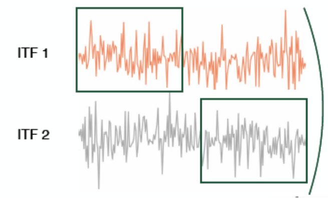
$$\text{FAR}(\eta) = N(\eta > \eta_0) / T$$

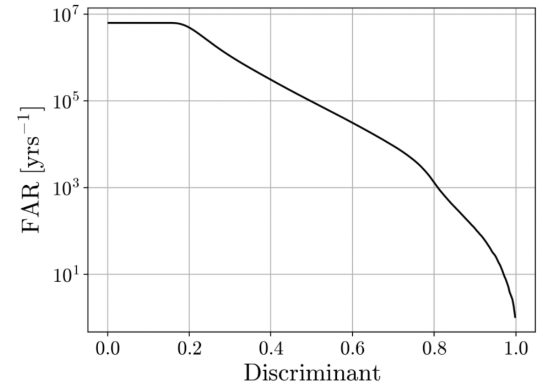
- We aimed for a CNN operating point (η_0) with FAR of 1/year but FAR remains large for CNNs as $\eta_0 \to 1$

- We further improved the sensitivity by combining the CNNs outputs in a single discriminant

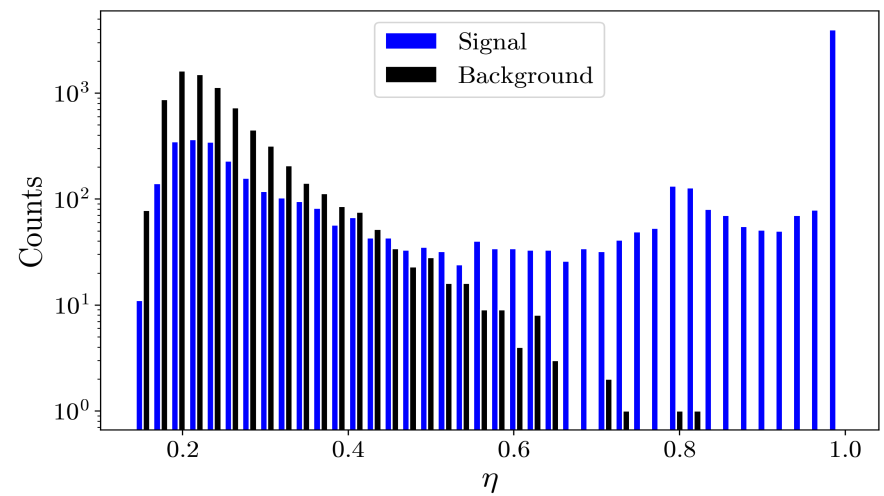
→ a simple average was implemented after considering other more sophisticated combination methods

→ A FAR below 1/year is reached



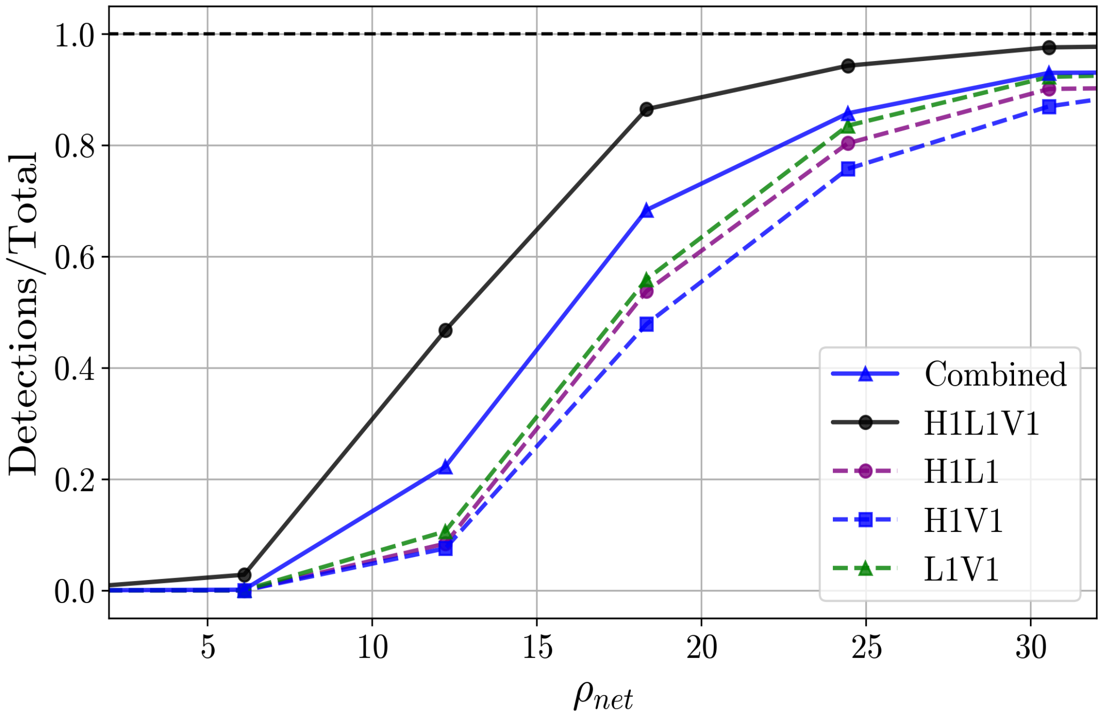


CNN	Threshold	TP rate	FP rate	FAR($\eta_0 = 1$) [yrs^{-1}]
H1 – L1	1.0	0.46	$\lesssim 2 \cdot 10^{-4}$	$\sim 10^2$
L1 – V1	1.0	0.47	$\lesssim 2 \cdot 10^{-4}$	$\sim 10^3$
H1 – V1	1.0	0.44	$\lesssim 2 \cdot 10^{-4}$	$\sim 10^2$
H1 – L1 – V1	1.0	0.58	$\lesssim 2 \cdot 10^{-4}$	$\sim 10^3$
Combined	0.998	0.50	$\lesssim 2 \cdot 10^{-4}$	$\sim 10^{-2}$



Improved separation of signal vs background using the combined discriminant

Signal Injections



For a single interferometer

$$\rho^2 = 4 \int_{f_{\min}}^{f_{\max}} df \; \frac{|\tilde{h}(f)|^2}{S_n(f)}$$

For multiple interferometers

$$\rho_{\rm net}^2 = \sum_i \rho_i^2$$

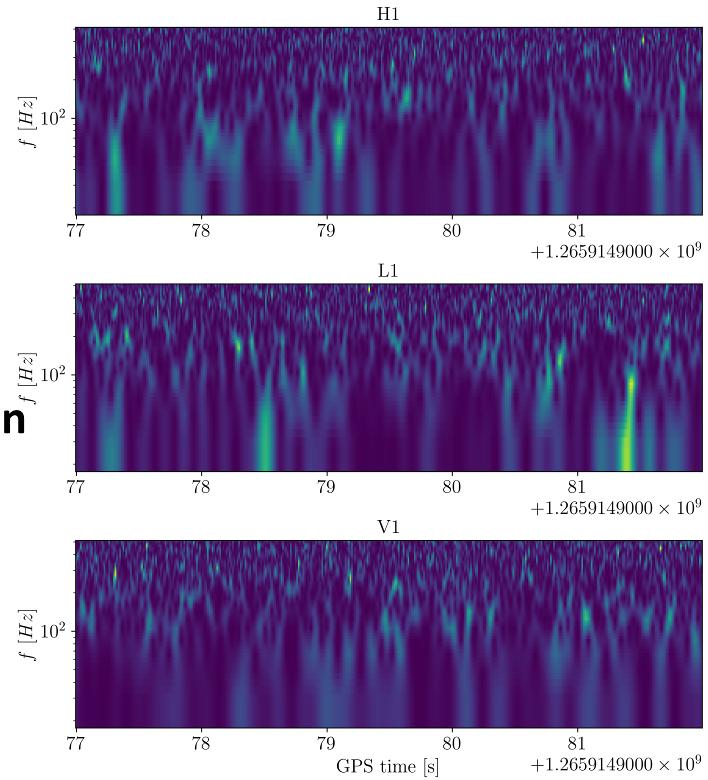
We evaluate the performance of the CNNs and their combination in terms of SNR (ρ) using the injection of signals

→ **Best results are obtained by the H1-L1-V1**
→ The performance of the combination is a compromise between H1-L1-V1 and the rest

CNN	$\rho_{\rm net}(80\%)$	$\rho_{\rm net}(90\%)$	$\rho_{\rm net}(95\%)$
H1 – L1	24.4	30.5	42.0
L1 – V1	23.7	29.0	41.2
H1 – V1	26.8	21.1	25.8
H1 – L1 – V1	17.3	21.1	25.8
Combined	22.4	28.0	40.1

O3 scan

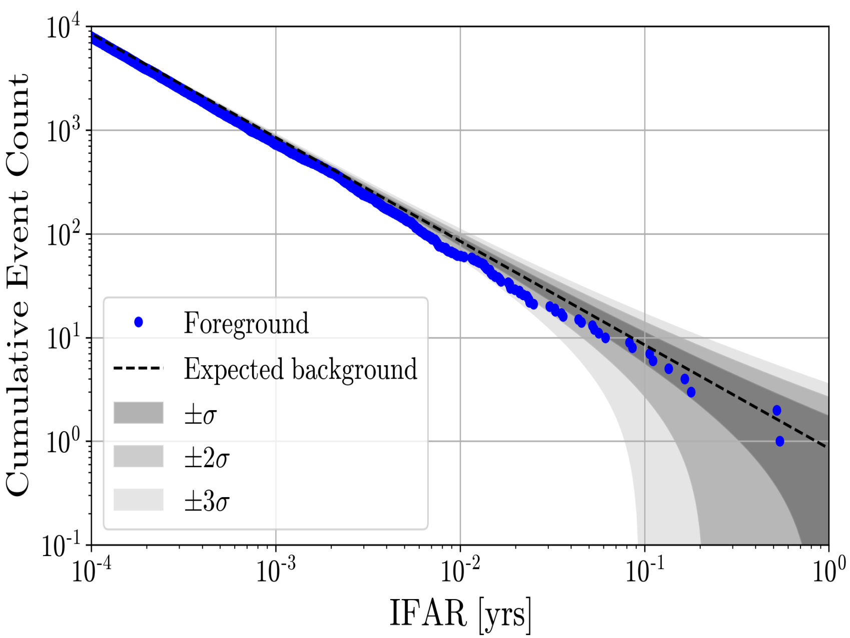
most significant event
with FAR of 1.9 $years^{-1}$
Combined CNN → 0.9635



We performed a full scan over the O3 data using images of 5s duration with a slicing window of 2.5s
→ This implies the testing of more than 8M images

The combined CNN is used and different values between 0 and 1 of the discriminant are explored
→ In each case the (inversed FAR) is computed

No significant deviation from background only prediction
→ No claim for detection of asymmetric events



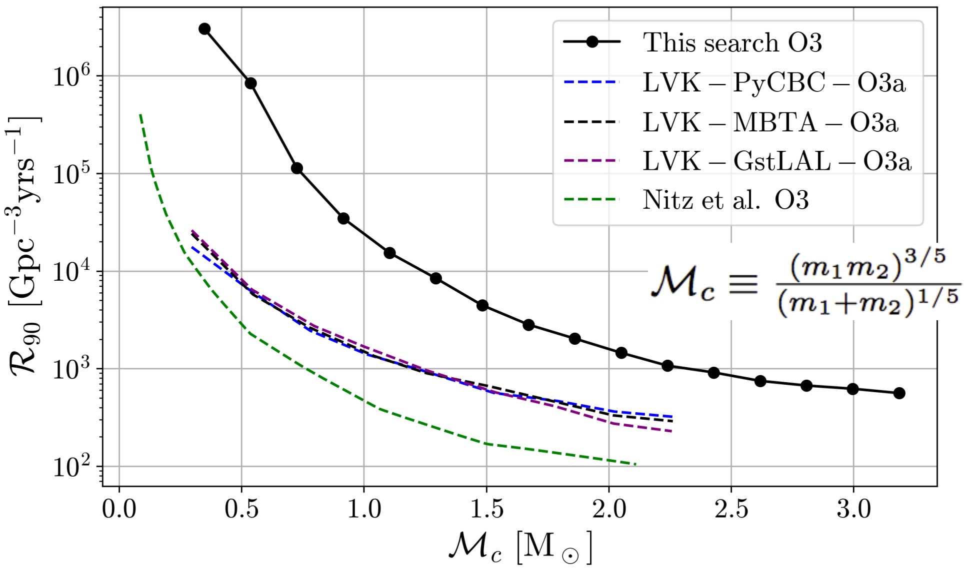
The O3 scan fast → 2000 CPU hours

However, FAR computation required using GPU Minotauro Cluster in the Barcelona's Super computing center (BSC)

arXiv:2109.12197

90% CL on merger rates

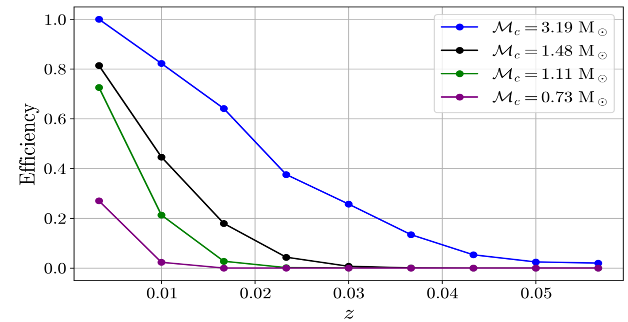
$$\mathcal{R}_{90} = \frac{2.3}{\langle VT \rangle}, \ \langle VT \rangle = T \int dz \frac{1}{1+z} \frac{dV_c}{dz} \varepsilon(z)$$



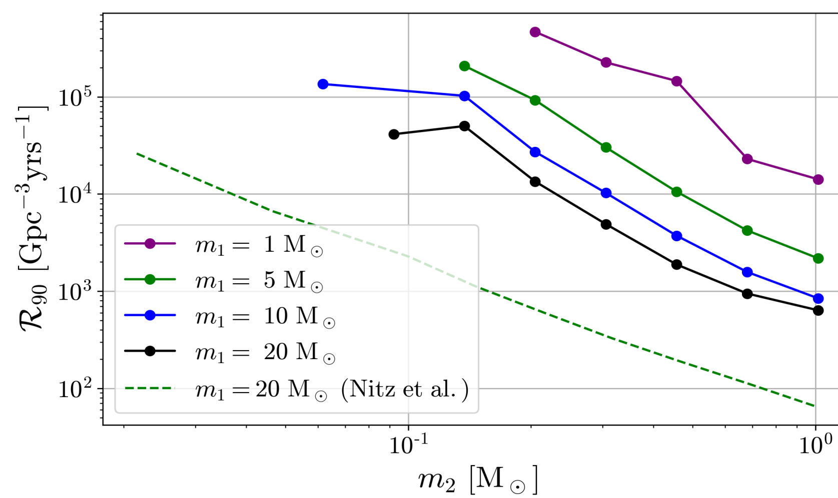
Using loudest event statistics

→ depends on surveyed time-volume

Detection efficiency ε(z) computed using injected signals (z < 0.06)



Other parameters homogeneously distributed in comoving volume



Results are expressed in terms of 90% CL upper limits on CBC merging rates vs mass

→ Not improving existing bounds from matched filtering results (but extended in mass)

→ Partially attributed to the reduced (1/2) in T imposed by using H1-L1-V1 configurations

Final notes

We present results on the search for asymmetric binary events using CNNs and O3 data

https://dcc.ligo.org/LIGO-P2200184

- **No claim for detection is made after full scan and calculating the FAR for each configuration**
- **Results translated into 90% CL upper limits on merging rates versus chirp mass**

Not competitive with existing matched filtering based bounds but extends the mass range

→ **Due to the required H1-L1-V1 configurations to control fake rates but reducing ½ the observation time.**

→ **Future improvements → implement a nested NN to reduce glitches and fake rates**

Results already blessed by LIGO and Virgo and will be submitted for publication once the LVK subsolar mass paper is finally submitted

(we are under embargo)

Searches for Mass-Asymmetric Compact Binary Coalescence Events using Neural Networks in the LIGO/Virgo Third Observation Period

M. Andrés-Carcasona,[1] A. Menéndez-Vázquez,[1] M. Martínez,[1,2] and Ll. M. Mir[1]
[1] *Institut de Física d'Altes Energies (IFAE), Barcelona Institute of Science and Technology, E-08193 Barcelona, Spain*
[2] *Catalan Institution for Research and Advanced Studies (ICREA), E-08010 Barcelona, Spain*
(Dated: September 27, 2022)

We present results on the search for the coalescence of compact binary mergers with very asymmetric mass configurations using convolutional neural networks and the LIGO/Virgo data for the O3 observation period. Two-dimensional images in time and frequency are used as input. Masses in the range between 0.01 $M_\odot$ and 20 $M_\odot$ are considered. We explored neural networks trained with input information from a single interferometer, pairs of interferometers, or all three interferometers together, indicating that the use of the maximum information available leads to improved performance. A scan over the O3 data set using the convolutional neural networks for detection results into no significant excess from an only-noise hypothesis. The results are translated into 90% confidence level upper limits on the merger rates as a function of the mass parameters of the binary system.

PACS numbers: 95.85.Sz, 04.80.Nn, 95.55.Ym, 04.30-w, 04.30.Tv

I. INTRODUCTION

Since the discovery of Gravitational Waves (GW) in 2015 [1], generated by a compact binary coalescence (CBC) of black holes (BH), the LIGO and Virgo experiments have improved their sensitivity and observed an increasing number of GW signals, including also events attributed to the coalescence of neutron stars (NS), as well as the coalescence of BH-NS binary systems. The latest catalogue of events, from O1, O2 and O3 observation runs, collects a total of 90 events, dominated by BH-BH candidates [2–4]. The data indicate that the masses in the binary systems range between 1.17 $M_\odot$(GW191219_163120) and 105 $M_\odot$(GW190426_190642), with a mass ratio $q \equiv m_1/m_2$, where m_1 denotes the heaviest of the two objects, in the range between 1.1 (GW170817) and 26.5

Since there is no well-established astrophysical explanation for the origin of SSM BHs, their discovery would point to the presence of new physics. The presence of SSM BHs are predicted by different models, including primordial black holes (PBHs) from the the collapse of overdensities in the early universe [15–18]; gravitational collapse of dark matter halos [19–22]; the accumulation of dark matter by neutron stars leading to SSM BHs [23]; or SSM boson stars [24–26]. As illustrated in Figure 1, this study complements the phase space in mass considered by previous searches for SSM events using O3 data and matched-filtering based selections [4, 27, 28]. Previous results using other observational periods are included in Refs. [29–32].

II. DATA PREPARATION

Notes on other combinations

TRAINING RESULTS - Combination of outputs

To enhance the detection of events we have explored the combination of the outputs of the CNNs trained with the information coming from different ITFs.

Method 1 from Physics of the Dark Universe 35 (2022) 100932

$$Loss = \sum_{SNR=5}^{SNR=40} \sigma^2_{SNR} \longrightarrow \sigma^2_{SNR} = \langle D_{SNR} - \langle D_{SNR} \rangle^2 \rangle$$

After running it for our case:

$$D = \beta_1 D_{L1H1V1} + \beta_2 D_{L1H1} + \beta_3 D_{L1V1} + (1 - \beta_1 - \beta_2 - \beta_3) D_{H1V1}$$

$$\beta_1 = 0.03 \quad \beta_2 = 0.30 \quad \beta_3 = 0.33$$

Method 2 from Physics of the Dark Universe 35 (2022) 100932

$$Loss = \sum_{SNR=SNR_0+1}^{SNR=40} \frac{N(D_{SNR} < D_0)}{N_{SNR}} \qquad \frac{N(D_{SNR_0} > D_0)}{N_{SNR_0}} = 0.5 \rightarrow D_0$$

SNR_0: SNR at which to accept 50% of the events

After running it for our case: Optimizer didn't converge, the function is not smooth enough

Minimization of the binary cross-entropy

$$Loss = -\frac{1}{N} \sum_{i=1}^{N} y_i \log(D_i) + (1 - y_i) \log(1 - D_i)$$

After running it for our case:

$$D = \beta_1 D_{L1H1V1} + \beta_2 D_{L1H1} + \beta_3 D_{L1V1} + (1 - \beta_1 - \beta_2 - \beta_3) D_{H1V1}$$

$$\beta_1 = 0.45 \quad \beta_2 = 0.50 \quad \beta_3 = 0.0$$

Arithmetic mean

Gives a good FAR distribution

Example