Summary: Summer fellowship in convolutional neural networks for gravitational waves detection

Abstract

This document includes a summary of Teodor Parella Dilmé's achievments along the 2020 IFAE summer internship in convolutional neural nets (CNN) for gravitational waves (GW) detection at Virgo department.

Several 2 channel ResNet-50 CNN for asymmetric binaries with $M_2 = 2$ - $5\ M_\odot$, and a q factor ranging from $q = 0.1$ - 0.3 have been developed, and their performance has been eventually represented throughout their corresponding ROC curves. A further study of the developed CNN plus another high mass CNN created by Alexis Menendez, has been carried out: CNNs performance has been studied as a function of the chirp mass, distance and effective distance of the sources, giving a broad vision of how the CNNs performance extend far way beyond the training range. Finally, a study regarding how the images signal to noise ratio (SNR) within the training range influence the CNN effectiveness has been executed. A study of the relationship between the distributions of euclidean distance and effective distance has been carried out as well.

I. CNN for asymetric binaries

A model for a 2-channel ResNet-50 CNN [1] has been developed for binary systems with a low mass ranging from $M_2 = 2$ - $5\ M_\odot$, and a q factor ranging from $q = 0.1$ - 0.3. The input specrograms belonged to L1 and H1 spectrometers, and offered a reslution of 400x100 bins.

For the training, theoretical spectrograms have been created by adding a theoretical GW signals over noise backgrounds from O2. This procedure has been carried out with the PYCBC and KERAS libraries, and the available data from O2 in the Virgo department.

Several CNN have been developed for the model, varying the distance or effective distance while generating the spectrograms at the training phase. Three remarkable CNN correspond to effective distances (D_{eff}, see section II) from 10 Mpc to 300 Mpc, 600 Mpc and 1000 Mpc (for at least one interferometer). The corresponding ROC curves are shown in fig(1).

The fact is that, while increasing the limit effective distance, spectrograms' signals weaken, making the CNN struggle to detect the least intense signals. That is, the CNN performance decays (within the range) with an increasing effective distance threshold. An in-depth study of the signal to noise ratio (SNR) and CNN scope is carried on the following sections of this document.

II. A effective distance study

Effective distance is a key magnitude, as will be seen further on. This magnitude takes into account the geometric factors of the binary system with respect to the interferometers: it represents an equivalent euclidean distance for an optimally-orientated source.

A stochastic approach for uniform random parameters has enabled to obtain the count rate as a function of the distance and D_{eff}.

Fixing a distance d, the normalized probability for a given effective distance is shown in fig(2).

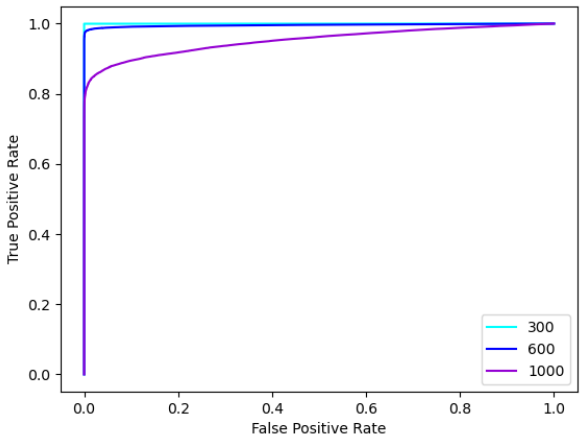

Figure 1: *ROC curves for the three CNN with 10-300 Mpc, 10-600 Mpc and 10-1000 Mpc effective distance threshold.*

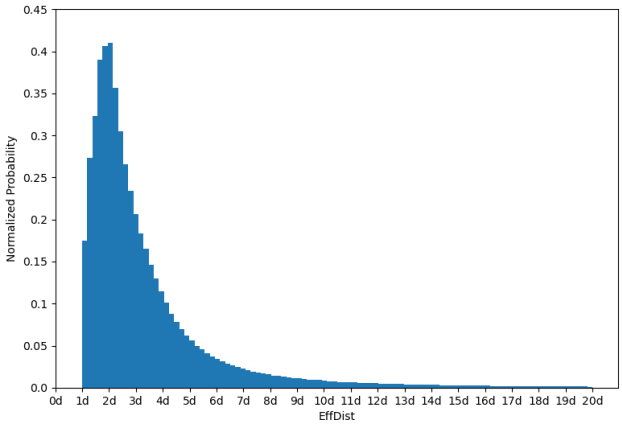

Figure 2: *Stochastic normalized probability as a function of the effective distance for a fixed distance d.*

Whereas fixing an effective distance D_{eff}, the normalized probability distribution for the distance is given in fig(3)

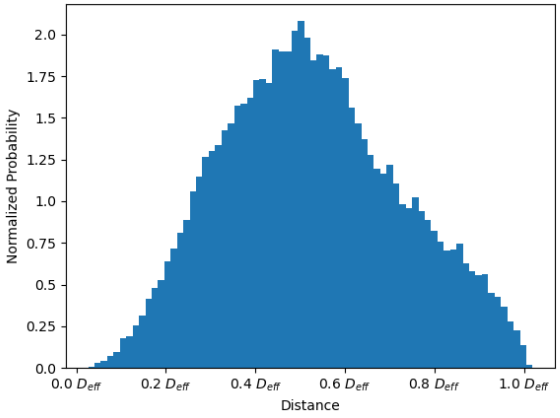

Figure 3: *Stochastic normalized probability as a function of the distance for a fixed effective distance D_{eff}.*

As can be seen in fig(2) and fig(3), an effective distance will always be equal or larger than the corresponding distance.

III. A neural network scope study

While scanning through the whole O2 data with CNNs presented in section I, several GW beyond the training scope were detected. This fact gave urge to develop a study of the CNN effectiveness in front of the binary parameters.

First of all, one should remind that the spectrograms introduced in the neural nets size 100x400 pixels. The neural nets are just able detect the characteristic merging pattern in the time-frequency spectrogram, since they appear blurry and no details apart from this can be apreciated. This detection depends on 3 main conditions:

A) The merging time range, both in the merge position through the 5 seconds that the spectrogram lasts and the merge duration.

B) The merging frequency range.

C) The merging signal intensity.

For the CNN developed, since merging times are located at random through the 5 seconds, the time location should not be an inconvenience.

The merge duration and frequency range are two factors depending mainly on the masses m_1 and m_2 of the binary, which can be englobed in the chirp mass $\mathcal{M}$ as

$$\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}} \quad . \tag{1}$$

That is, facts A and B (shape of the merge) can be easily solved by training the CNN with all the cases wanted to deal with.

The huge problem for the CNN leans on the signal intensity. The signal intensity of a gravitational wave is directly proportional to the strain h of the gravitational wave. An approach to this strain [2] [3] may be given by:

$$h(t) = A(t)cos(2\phi_0 + 2\phi(t)) \tag{2}$$

Where:

$$A(t) = -(\frac{G\mathcal{M}}{c^2 D_{eff}})(\frac{t_0 - t}{5G\mathcal{M}/c^3})^{-1/4} \tag{3}$$

Being $\mathcal{M}$ is the chirp mass, D_{eff} the effective distance, ϕ_0 the termination phase and $\phi(t)$ the phase.

Essentially, the strain amplitude can be rewritten as:

$$A(t) = \frac{\mathcal{M}^{5/4}}{D_{eff}}\alpha(t) \tag{4}$$

Overall, for the signal intensity, all that is important is the given quotient from the previous equation (4), that is, the signal intensity depends on the chirp mass and effective distance.

Considering binaries with $\mathcal{M} = 1-61\ M_\odot$ and $D_{eff} = 10-2000$ Mpc, a 20x20 partition of these intervals has been done, and around 2000 spectrograms for each interval have been created. The 1000 Mpc effective distance CNN weights from section 1 has been tested with this data, and the CNN detection rate for each interval has been saved, always with a threshold of 0.95 for deciding ether if a detection was positive or negative. The results have been plotted in fig(4).

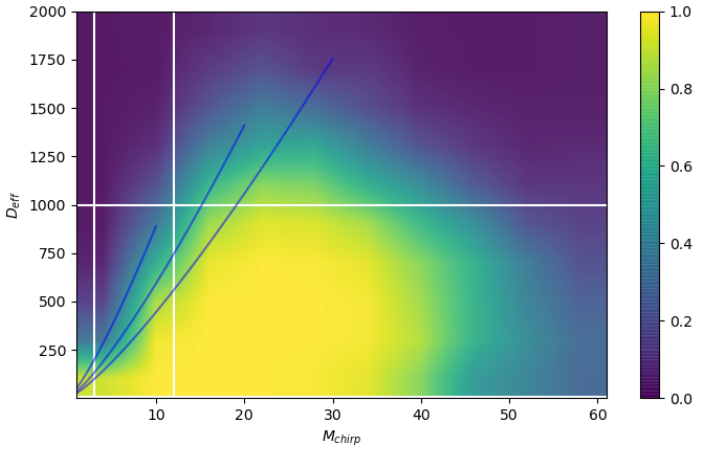

Figure 4: *Detection rate as a function of the effective distance and chirp mass for the CNN with D_{eff} = 10-1000 Mpc. The square within the white lines represent the training range, whereas the blue curves represent equal values of the quotient in equation (4).*

It can be seen that the scope goes far way beyond the training range. Within the training range, the CNN is able to detect the most intense spectrograms, or a huge fraction of them (corresponding to the yellow and green zone inside the white square), whereas when the signal intensity diminishes, the CNN is unable to detect

the merges, or only detects a few percentage (violet and dark blue zone inside the white square). This plot has complete sense with the corresponding ROC curve in fig(1), as a considerable part of GW within the training range aren't detected.

The blue curves represent equal values of the quotient in equation (4), which are same intensity merges (from left to right: 0.02, 0.03 and 0.04). While increasing the chirp mass, merges shape and duration change: they become more vertical and last less time, as can be appreciated in fig(5).

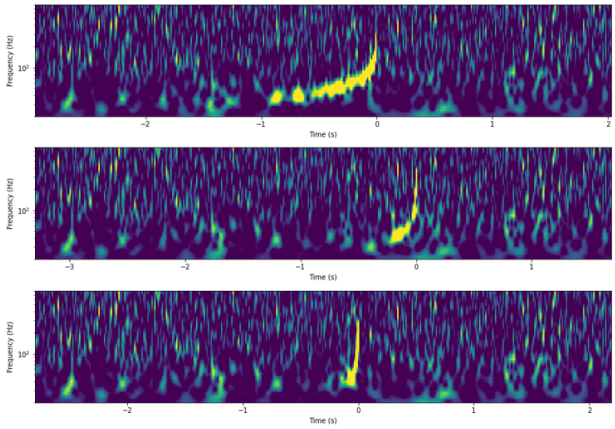

Figure 5: *Shape of three different gravitational waves with chirp masses of 10, 20 and 30 $M_\odot$ from top to bottom. The shape becomes more vertical with an increasing chirp mass.*

It is not strange that the CNN looses performance at high chirp masses. The kernels, or other instruments from the neural net, may be no longer useful to detect this kind of shapes.

A study of a high mass 2-channel CNN for distances ranging from 100-1000 Mpc and masses ranging from 25-100 $M_\odot$, developed by Alexis Menendez, has been carried out with this method. The same analysis of the detection rate for a 0.95 threshold is ploted in fig(6).

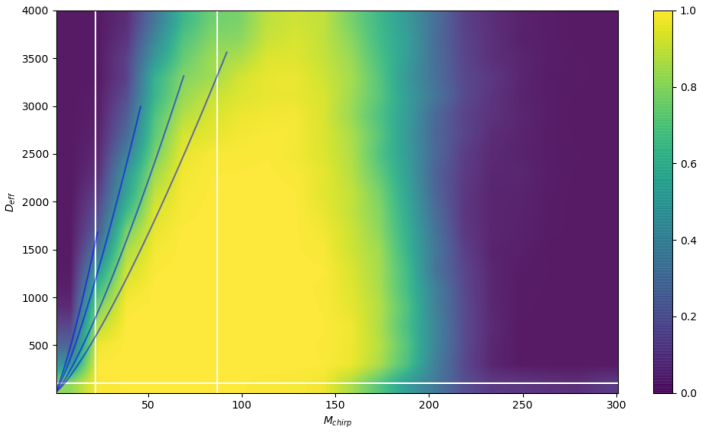

Figure 6: *Detection rate as a function of D_{eff} and chirp mass for the high mass CNN with d = 100-1000 Mpc. The rectangle within the white lines represent the training range, whereas the blue curves represent equal values of the quotient in eq.(4).*

Again, the scope of the CNN goes far way beyond the training range (in this case can't be represented by a closed rectangle, as the training range has been done in distance instead of effective distance, but it is englobed in the middle-rectangle of the figure), In this case, training the CNN with higher masses, and hence higher chirp masses, allows the detection to reach higher chirp masses (detection rate arround 1 up to $\mathcal{M}$ = 150 $M_\odot$). The fact is that this CNN is expected to detect almost any high mass system so far, since the biggest binary collision observed until today is GW190521, with $\mathcal{M} = 64 \pm 13\ M_\odot$.

Again, after a certain $\mathcal{M}$, the CNN is unable to recognize the merges. It can be seen in the last image of fig(5), that with the 400x100 pixel resolution used, after a chirp mass of 30 $M_\odot$, the merge is almost a vertical line. Beyond this point, the merge timing becomes narrower and narrower, until a point where the merge is so narrow in time that it can not be seen in the spectrograms with this resolution. Hence, after 150 $M_\odot$ chirp masses, the neural net begins to loose accuracy due to image low resolution.

The blue lines in this case represent quotient values of (from left to right) 0.03, 0.04, 0.06 and 0.08. Regarding conditions where low intensity merges take place and the net begins to loose accuracy due to this (around the blue curves), any comments commenting performance at lower intensities can be done at first sight. Despite this, the big mass CNN extends the scope further away from the training limits than the asymmetric CNN does.

Performing the same analysis for the high mass CNN, but instead of effective distance, considering the distance, fig(7) is obtained:

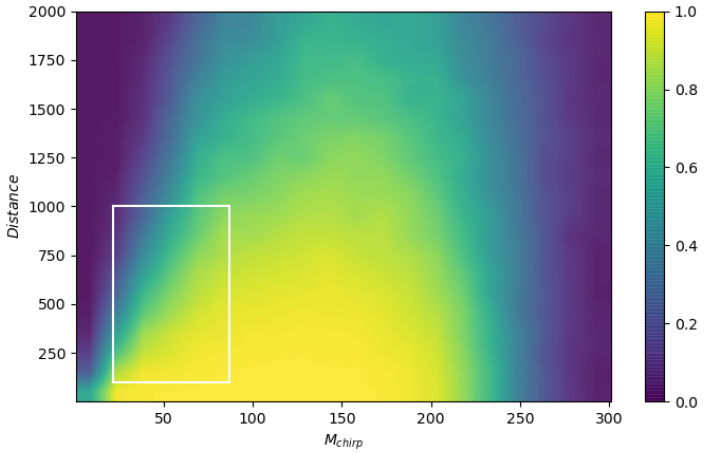

Figure 7: *Detection rate as a function of the distance and chirp mass for the CNN with d = 100-1000 Mpc. The white square represents the training range. (4).*

Which may be more intuitive at first sight.

IV. SNR detection

A study of the CNN performance as function of the signal to noise ratio (SNR) has been developed for the high-mass CNN from Alexis Menendez. A theoretical power spectral data (PSD) generated from the PYCBC library has been used. Theoretical GW signals within the training range have been created, which then have been adjusted to fit a specific value of SNR [4]. 100 partitions have been made from 0 to 99 and have been used as SNR, with 4080 spectrograms for each number. Then, the high-mass CNN weights have been run over the spectrograms in order to get the detection rate for different thresholds. The results of the analysis can be seen in fig(8).

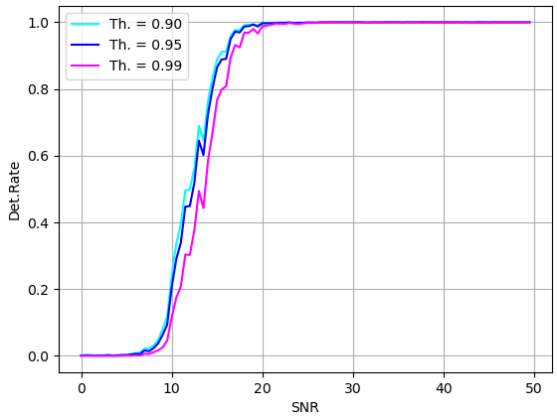

Figure 8: *High mass CNN from Alexis Menendez: detection rate as a function of the SNR for 0.90, 0.95 and 0.99 CNN output thresholds.*

The CNN performance clearly increases with an increasing spectrogram SNR. Low SNR merges are not detected at all whereas after reaching about 8 SNR values, the detection rate begins to increase until SNR values of 20. At this point almost every GW is detected by this CNN. Another identical SNR analysis has been developed for the 1000 Mpc D_{eff} asymmetric binaries from section I. The detection rate as function of the SNR is plotted in fig(9).

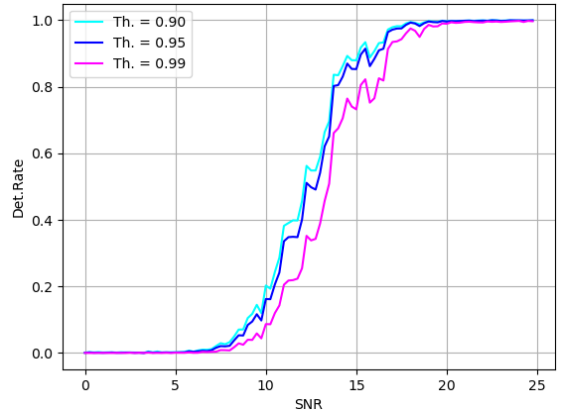

Figure 9: *Asymmetric binaries $D_{eff} = 1000$ Mpc: detection rate as a function of the SNR for 0.90, 0.95 and 0.99 CNN thresholds.*

Despite that fig(8) and fig(9) have been acquired through the same procedure, strangely the second one does not look as smooth as the first one. This SNR analysis procedure does not probably work in low chirp mass binaries CNN as good as it does with high mass ones.

The comparison for both fig(8) and fig(9) for 0.90 and 0.99 thresholds can be seen in fig(10).

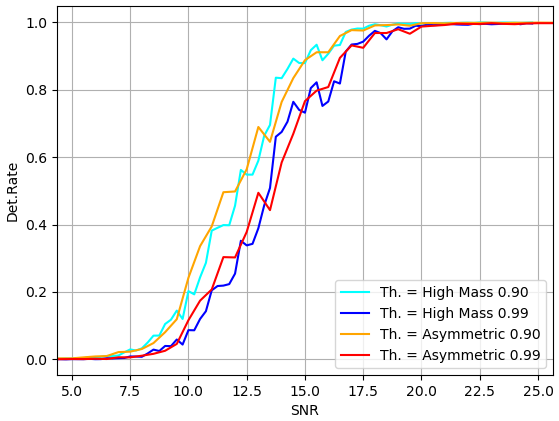

Figure 10: *Comparison of detection rate as a function of the high mass CNN (Cyan and Blue) and asymmetric CNN (Orange and Red) for 0.90 and 0.99 CNN thresholds.*

The detection rate as function of the SNR behaves very similarly in both cases. Due to low resolution of the final curves, small differences between them may not be appreciated. Despite this, both CNN seem to have a similar detection rate dependence on the SNR.

A more accurate study of the detection rate as function of the SNR would involve the calculus of the real PSD from the noise images instead of a theoretical calculation with the PYCBC library.

References

[1] Deep Residual Learning for Image Recognition. Kaiming He et al. (2015)

[2] FINDCHIRP: An algorithm for detection of gravitational waves from inspiraling compact binaries. Bruce Allen et al. (2012)

[3] Implementing a search for aligned-spin neutron star - black hole systems with advanced ground based gravitational wave detectors. Tito Dal Canton et al. (2018)

[4] Convolutional neural networks: A magic bullet for gravitational-wave detection? Timothy D. Gebhard et al. (2019)