Understandable Deep Learning Analysis for Very High Energy Astroparticle Physics

Ettore Mariotti[1]

Abstract. This work addresses the machine learning problems related to Imaging Air Cherenkov Telescopes analysis with the intent of leveraging a richer description of the captured events by using Convolutional Neural Networks, avoiding the potential information loss due to the hand-crafted feature engineering used in the standard analysis. An explainable and minimal model pointed toward the limitation of the simulation used for training. The full analysis is then tested both on simulated and real data, showing a significant improvement of the sensitivity of the system for different energy bands.

1 INTRODUCTION

In the universe there is more than we can see, our eyes are only sensitive to a small portion of the light spectra but outside of that lie information about a wild variety of phenomena. In particular the MAGIC telescopes[2] are built to capture the most energetics photons (gamma-rays) that reach our planet and interact with the atmosphere. The system is built exploiting a phenomenon known as Cherenkov radiation, a brief and dim blue light emitted by objects traveling faster than the speed of light in the medium they are in. One of the big problems with this technique is being able to discriminate the signal from the background. Unfortunately gamma rays are not the only particles that produce Cherenkov light, as there is a wide and diverse catalog of events that make the telescopes trigger, outnumbering gamma rays at about 2000:1. In order to discover and characterize a new gamma-ray source the analysis must recover only the events caused by gamma photons and then reconstruct their energy and their direction of impact in the sky.

When an energetic particle interact with the atmosphere, it produces a cone of light due to the Cherenkov effect. The telescopes focus with a big array of mirrors this secondary shower of light toward the camera. The camera is composed of 1024 photodetectors which produce a voltage proportional to the light that reach the sensors (raw signal). This voltage is then processed to estimate the number of photons arrived to the camera (Calibrated signal). After a number of triggers that aim to grab plausible gamma events, a set of handcrafted features is computed (Hillas parameters). The physical underlying process is reproduced with a Montecarlo Simulation (MC) which allows to have a set of expected observations with their ground-truth from which the analysis can learn. Once the analysis chain is completed, in order to check its correctness the telescopes are pointed toward the Crab Nebula, a well-known strong and stable source of gamma-rays, and compares its spectra with the one established in the literature by other experiments.

[1] Universitá di Padova, Italy, email: mr.ektor@gmail.com
[2] http://www.magic.iac.es/

In order to select and reconstruct the energy and the direction of the events, the standard analysis makes use of machine learning from a parametrization of the grabbed events. In particular, Aleksic et al. [1] solves the tasks in this way:

1. *Separation*: A random forest for classification is trained on the Hillas parameters for MC gamma rays (Signal) and real observations taken when the telescopes point to a dark region of the sky (Background)
2. *Energy*: A Nearest Neighbour trained on the MC estimates the energy of the events
3. *Direction*: Some additional features are computed with a geometric rationale, and the final result is then fine-tuned with a Random Forest Regressor.

2 DEEP LEARNING ANALYSIS

The feature engineering is (as of any parametrization) destroying information that could be potentially useful for the purposes of the analysis. The approach proposed in this work is based on Deep Learning (DL). More precisely we use Convolutional Neural Network (CNNs) starting from the Calibrated signal. Anyway since the feature engineering provides an already valuable source of predictive information, Hillas parameters are passed to the network as side-information. The architectures for each task thus are composed of two differentiable stages:

1. a CNN body that process the Calibrated signal
2. a Self-Normalizing Neural Network [6] that processes the hand-crafted features.

These two branches of processing are then merged together at the final layer in order to produce the final estimation. In particular different CNN bodies were found more suit to different tasks:

1. *Separation* uses a VGG-19 [7]
2. *Energy* uses an Inception V3 [8] with squeeze and excitation blocks [2]
3. *Direction* uses a Densenet-121 [4] with squeeze and excitation blocks [2]

Each model was trained with stochastic weight avearging [5] of checkpointed models computed with a cyclical learning rate [3]. In order to being able to interface with the DL frameworks the calibrated data was interpolated from a native hexagonal grid to a regular rectangular grid. Due to the big memory footprint of the training data (>500GB), the events were structured in a custom-designed SQL server.

Proceedings of the 1^{st} Doctoral Consortium at the European Conference on Artificial Intelligence (DC-ECAI 2020), pages 57–58, Santiago de Compostela, Spain, August 29-30, 2020.

3 UNDERSTANDING THE SIMULATION/REALITY GAP

When separation is trained and tested on the MC-Real data (as in the standard analysis) the model reach an accuracy of 99.99%, which is unrealistic because we still expect a small 1% of gamma-like events in the Real data.

In order to better investigate what happened in the separation, we designed a minimal and explainable CNN model (called SimplicioNet) based only on the Calibrated signal. Its explainable nature lies in its simple structure:

1. One layer of four convolutional filters 20x20 pixels with ReLU activation.
2. One maxpool layer that selects the most intense activation.
3. One dense layer without bias that combines the values of the previous activation and squeezes the result with a sigmoid.

By having a sigmoid as the last non-linear activation we can know that the feature map combined with a positive coefficient is responsible for the gamma class and the one combined with a negative coefficient is responsible for the non-gamma class. By tracing the location of the most prominent feature (maxpool) it is possible to trace the precise activation field of the Calibrated signal that caused the prediction. When inspected, it was clear that the model was making decisions not by looking at the shower itself, but by looking at other places of the image. The model was solving the separation problem by focusing on some feature of the MC simulation that was different from the Real-Data observations. In other words it was discriminating the Simulation from the Real data instead of gamma from non-gamma. The separation problem was then addressed in two different ways:

- by performing a strong preprocessing of the calibrated signal that cleaned the image;
- by using a MC simulation of the background (which unfortunatly is not really representative of the richness of the real events).

4 RESULTS

The whole pipeline was evaluated both on MC and on Real observations.

Regarding the MC set, we observed that the DL proposal showed significant improvements with respect to the standard analysis. Energy resolution is enhanced of 30% for events above 1 TeV while direction reconstruction is enhanced for each energy band ranging from a 5% to 40%. This is especially important since improvements in the direction reconstruction translates proportionally to improvements in the sensitivity of the system. The whole pipeline, thus including also the separation stage, globally enhance the sensitivity, particularly in the lower energies bands (Fig. 1).

The final pipeline was then evaluated on a set of 5h of real observations of the crab nebula. The results showed performances comparable with the one of the standard analysis, thus proving the proof of concept that this richer analysis can in principle work in a real environment. Most importantly, as the simulation will become more and more realistic it is reasonable to expect a much more dramatic improvement in line with the results on the MC set.

5 CONCLUSIONS

This work leveraged the steep advancements of computer vision by adapting DL solutions to the specific field of Imaging Air Cherenkov Telescopes. Results on MC showed superior performances in terms of sensitivity of the instrument. The implementation of such approach would allow to require less observation time in order to claim a discovery, which translates in more discoveries in the same operational time. Moreover the results point toward improvement in certain energy bands that allows the discovery of previously-undetectable sources. SimplicioNet made evident the limits of the simulation, thus investment of energy and time in its refinement will be a valuable investment as it will resolve in higher performances on the real observations.

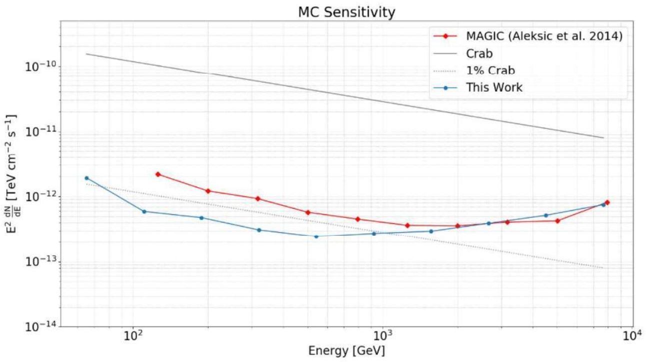

Figure 1. Sensitivity computed on the MC set obtained with the proposed approach (blue line) and compared with the one actually in operation for the telescopes (red line). The lower the better.

ACKNOWLEDGEMENTS

We would like to thank NVIDIA for donating appropriate hardware with which was possible to conduct this investigation. We would also like to thank the software board of the experiment for the valuable feedback obtained at different stages of the project.

This presentation is associated to the NL4XAI project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860621.

REFERENCES

[1] J. Aleksić, S. Ansoldi, L.A. Antonelli, P. Antoranz, A. Babic, P. Bangale, M. Barceló, J.A. Barrio, J. Becerra González, W. Bednarek, and et al., 'The major upgrade of the magic telescopes, part ii: A performance study using observations of the crab nebula', *Astroparticle Physics*, **72**, 76–94, (Jan 2014).

[2] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, and Enhua Wu. Squeeze-and-excitation networks, 2017.

[3] Gao Huang, Yixuan Li, Geoff Pleiss, Zhuang Liu, John E. Hopcroft, and Kilian Q. Weinberger. Snapshot ensembles: Train 1, get m for free, 2017.

[4] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks, 2016.

[5] Pavel Izmailov, Dmitrii Podoprikhin, Timur Garipov, Dmitry Vetrov, and Andrew Gordon Wilson. Averaging weights leads to wider optima and better generalization, 2018.

[6] Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. Self-normalizing neural networks, 2017.

[7] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2014.

[8] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision, 2015.

Proceedings of the 1^{st} Doctoral Consortium at the European Conference on Artificial Intelligence (DC-ECAI 2020), pages 57–58, Santiago de Compostela, Spain, August 29-30, 2020. Copyright held by the owner/author(s).