PROCEEDINGS OF SCIENCE

ONLINE ICRC 2021 THE ASTROPARTICLE PHYSICS CONFERENCE Berlin | Germany 37th International Cosmic Ray Conference 12–23 July 2021

Bayesian Deep Learning for Shower Parameter Reconstruction in Water Cherenkov Detectors

C. R. Bom,[a,*] **L. O. Dias,**[a] **R. Conceição,**[b,c] **B. Tomé,**[b,c] **U. Barres de Almeida,**[a] **A. Moraes,**[a] **M. Pimenta,**[b,c] **R. Shellard**[a] **and M. P. Albuquerque**[a]

[a] *Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro, Brazil*

[b] *LIP - Laboratório de Instrumentação e Física Experimental de Partículas, Lisbon, Portugal*

[c] *IST - Instituto Superior Técnico, Lisboa, Portugal*

E-mail: debom@cbpf.br

Deep Learning methods are among the state-of-art of several computer vision tasks, intelligent control systems, fast and reliable signal processing and inference in big data regimes. It is also a promising tool for scientific analysis, such as gamma/hadron discrimination. We present an approach based on Deep Learning for the regression of shower parameters, namely the its core position and ground energy, using water Cherenkov detectors. We design our method using simulations. We evaluate the limits of such estimation near the borders of the arrays, including when the center is outside the detector's range. We used Bayesian Neural Networks and derived and quantified systematic errors arising from Deep Learning models and in an EfficientNet model design. The method could be easily adapted to estimate other parameters.

37th International Cosmic Ray Conference (ICRC 2021)
July 12th – 23rd, 2021
Online – Berlin, Germany

*Presenter

PoS(ICRC2021)739

1. Introduction

The cosmic rays and gamma rays entering the Earth's atmosphere can cause Extensive air showers (EAS) generating Cherenkov light. Many EAS are connected to high energy extragalactic or interstellar medium sources that present a window to study the high energy Universe. Imaging Air Cherenkov Telescopes (IACTs) collect Cherenkov light produced and can reconstruct the energy and the direction of the primary gamma-ray. These detections have a typical energy resolution of around 20% for TeV gammas and zenith angles of about 20°. The resolution of MAGIC-II was measured to be 15% in Ground-based gamma-ray observatories sample the particles (mainly electrons and photons). From their time and position distributions, it can be determined, with reasonable accuracy, the shower core position and the direction of the primary gamma-ray. The energy measured by a detector array with electromagnetic calorimetric capabilities, like Water Cherenkov Detectors (WCD), is by definition highly correlated with the ground energy (S_{em}) [1]. In this contribution, we aim to investigate WCD shower parameters recovery from a novel perspective: using Deep and Bayesian Deep Learning techniques.

Deep learning algorithms [2] are a class of Neural Network methods that have changed scientists perspectives on how to model data. Neural Networks and DL models can approximate a wide variety of functions in compact $\mathbb{R}_n$[2–4]. This is achieved by a training phase that produces data-driven models. Therefore, Deep learning has been successfully applied to several regression problems [5–7]. To construe a deep model for regression instead of the standard classification problem, one may fit the model for a function similar to log-likelihood or a simply Absolute deviation. In this contribution, we adopt a Bayesian Deep Learning perspective based on the Concrete Dropout procedure [henfeforth CD; 8]. This method allows us to approximate a two-dimensional probability of both the predicted parameters, shower's core coordinates and ground energy, for a given deep model. In the following sections, we present a short overview of the architecture and the uncertainty estimation in the Bayesian Deep Learning framework.

The main focus of this contribution is to explore the newly proposed method to achieve an accurate energy reconstruction of EAS between several hundreds of GeV and a few TeV, which is the most challenging energy region.

2. Bayesian Deep Neural Networks

In DL predictions, uncertainty estimation is one of the key aspects to define the model reliability. Nevertheless, intuitive error estimation and uncertainty quantification in deep learning is the subject of active research [9]. For the purposes of this contribution, we consider two main classes of errors in many machine learning applications: *aleatoric* (statistical) and *epistemic*. Aleatoric uncertainties represent stochastic effects mainly related to data quality level. Epistemic uncertainties enclose the unknowns of a given DL model. Due to the nature of data-driven models, contrary to first-principles modeling, the epistemic errors can change if the training set is updated. Thus, the trained model might be refined. It is worth noticing that epistemic errors are systematic, though not every systematic is epistemic. Also, statistical errors are aleatoric, but some aleatoric errors might also be systematic. The total DL error is known as predictive error. One possible approach to derive uncertainty from DL models is representing it in a Bayesian scheme [10]. Let a set of DL weights matrices be $\omega = \{\mathbf{W}_{l=1}^{L}\}$, with L layers, the weight's matrix dimension is K_l. The Probability Density Function (henceforth PDF) for a prediction y and input x is

$$p(y|x, \mathbf{X}, \mathbf{Y}) = \int p(y|x, \omega)\, p(\omega|\mathbf{X}, \mathbf{Y})\ d\omega\,, \tag{1}$$

where $\mathbf{X}$ and $\mathbf{Y}$ are the training set's inputs and truth table, respectively. In the Concrete Dropout method, we define a dropout probability p used to drop some weights and thus derive the model's error (epistemic) from the $p(y|x, \mathbf{X}, \mathbf{Y})$ distribution [8]. The level of p is defined by a parametric function $q_\theta(\omega)$ that replaces the unknown weights probability $p(\omega|\mathbf{X}, \mathbf{Y})$. The Loss function optimizes the parameters θ are, including p in the training process defined as

$$\mathcal{L} = \int q_\theta(\omega)\ \log p(\mathbf{Y}|\mathbf{X}, \omega)\ d\omega - \mathrm{KL}(q_\theta(\omega)||p(\omega))\,. \tag{2}$$

The first term in the loss definition is the cross-entropy. The prior $p(\omega)$ used is weights initialization, and the Kullback-Leibler (KL) divergence can be approximated by a regularization [2]. The PDF of each prediction is then defined by applying those dropout probabilities in multiple forward passes of the trained network.

2.1 EfficientNet Models

EfficientNets [11] are a set of CNN based models. They were particularly scaled to optimize accuracy and computational budget. They are derived from an initial network model similar to a mobile CNN [12], this model is optimized by a multi-objective neural architecture search [13] for accuracy and Floating-point Operations Per Second. This output of this first optimization is a model known as $B0$. The following models are defined from a set of scaling relations regarding network width, depth and resolution:

$$\begin{aligned} \text{depth: } d &= \alpha^\phi \\ \text{width: } w &= \beta^\phi \\ \text{resolution: } r &= \gamma^\phi. \end{aligned} \tag{3}$$

ϕ is an integer named compound coefficient. The coefficients α, β, and γ are defined by a small grid search optimization in the $B0$ model with fixed $\phi = 1$. From this procedure, one may derive the $B1$ to BN neural network architectures using a standard ImageNet dataset for training where $\phi = N$. In this work, we test all models defined by the original EfficientNet paper [11] adapted to our loss and trained with our simulations. We find the best results regarding training stability and performance using the $B0$.

3. Shower Simulations

The shower simulations used as input to our deep learning methods were produced using CORSIKA [14] (version 7.5600) and the detector response using the Geant4 toolkit (with version 4.10.05.p01)[15]. We assumed an observatory with an altitude of 5200 a.s.l. The stations have a base diameter of 4 m. We simulated gamma-ray showers up to 1 TeV with 10 deg zenith angle. For a more detailed description of the simulation parameters used, we refer to [1]. We leave a detailed study of other zenith angles and a broader set of detector configurations for a future contribution, including hadronic showers.

We present the distribution of X_{core}, Y_{core} and ground energy (S_{em}) in figure 1. The dashed lines in the histogram present the limits of our detectors at $160m$ from the center. Thus any results near these lines are close to the border and outside are out of the detector limits. We define the regions of interest in the figure 2. The central region is the region that $X_{core} < 120m$ and $Y_{core} < 120m$. The region outside the detector limits is where at least one of the core coordinates are in the region $> 160m$.

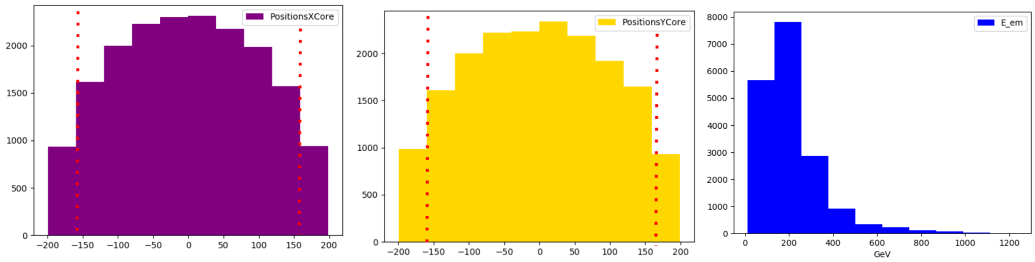

Figure 1: Shower's parameter distribution: X_{core} coordinates (center), the Y_{core} coordinates and ground energy S_{em}(right).

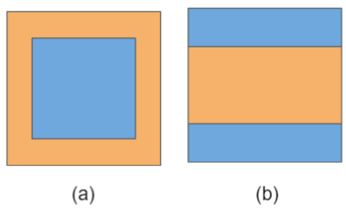

Figure 2: The detector region of interest of shower center: (a) The central region, where $X_{core} < 120m$ and $Y_{core} < 120$ (b) define the outside the border the region between where at least one of the core center coordinates ,X_{core} or Y_{core} is higher than $160m$.

4. Results

We split randomly the data into two different sets: Training and Validation/Testing. The first consist of 90% of the simulated showers. They were used to train the Neural Network. A set of 10% was used for validation and test purposes, they were not used to optimize the loss function, but they were a consistency check of the training generalization. This set is also used to evaluate the model performance. The training performance is presented in fig. 3. The model optimized for the loss function, and the comparison between the validation set and training set did not present a strong overfitting tendency, particularly for epoch 40 and beyond.

We present the summary of the best results within the Neural Networks tested in 4. We assess our model performance using the bias, i.e., $\delta X, Y_{core} = X, Y_{core} - X, Y_{pred}$ in the core position recovered. We observe that in most of the ranges considered, the bias is lower than $\pm 5m$, and it is lower than $\pm 2m$ in some regions. Not surprisingly, the scenario changes near the borders;

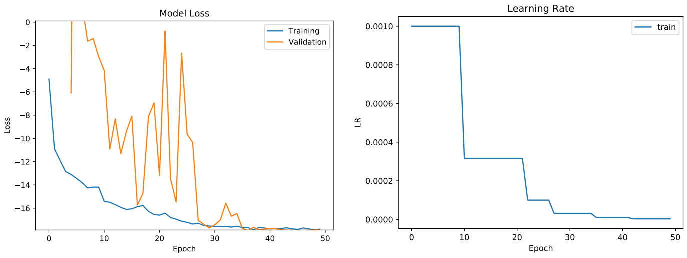

Figure 3: Training performance. Left: the loss per epoch. Right: Learning rate per epoch.

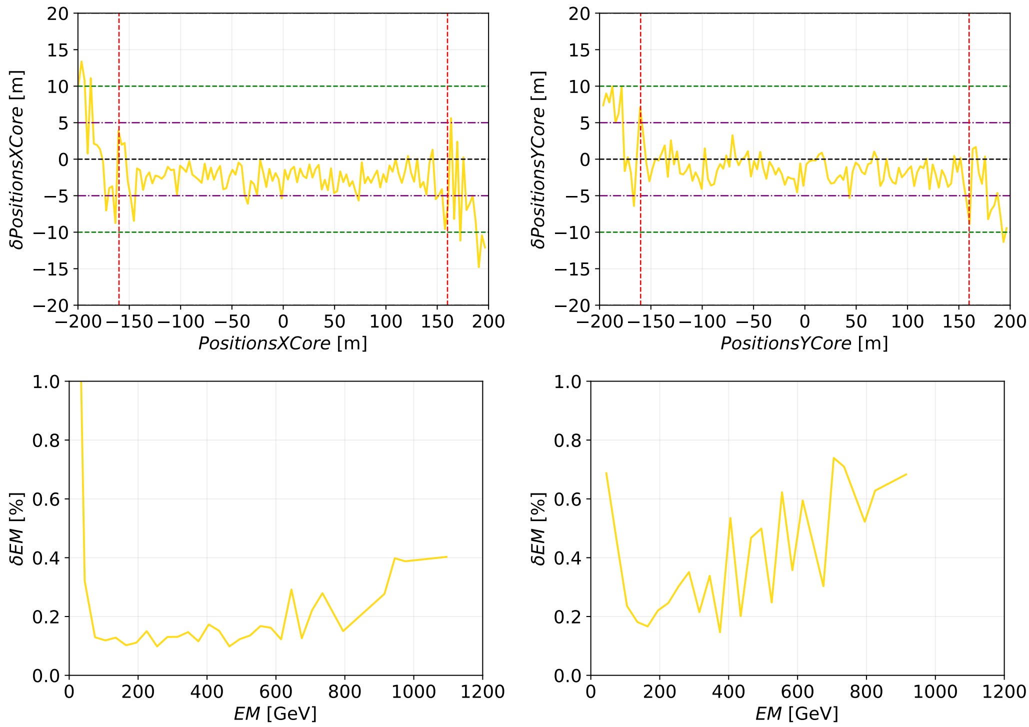

Figure 4: The panels summarizes the results using a Bayesian Deep Neural Network using EfficientNet $B0$ architecture. The upper panels presents the X_{core} (left) and Y_{core}(right) bias. The lower panels present the ground energy bias (%) for the region near the detector center (left) and outside detector range (right).

however, the bias still lower than $5m$ even outside the detectors. Due to the zenith angle, we observe that the two boundaries are not symmetric. The best result for the ground energy deviation, $\delta EM = (S_{em} - S_{pred})/S_{em}$ is presented in the lower panel of fig. 4 for two regions, in the left we for a central region, i.e. the (a) region displayed in 2. The left plot presents the result for the region (b) of the same figure, i. e., at least one coordinate outside the detector. We also present in fig. 5 the results for ground energy for the EfficientNet model without the Bayesian approach. The ground energy is constrained to less than 20% accuracy in the central regions for most of the range. The result with an EfficientNet (not in a Bayesian version) could reach a 10% or less around 200 – 600 GeV. The estimate outside the detector has a higher deviation as expected. However, we still have around 40% – 50% agreement in this region up to 600 GeV in the traditional Neural Network and

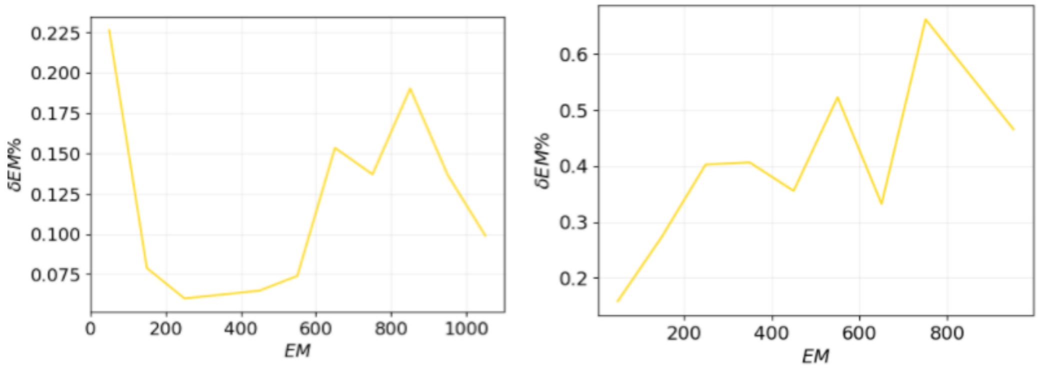

Figure 5: The results for a non bayesian EfficientNet regression for ground energy. Left: ground energy bias (%) for the region near the detector center. Right: ground energy bias (%) outside detector range.

40% – 60% for a Bayesian.

5. Discussion and Concluding remarks

5.1 Summary

In this proceeding, we present a new methodology using Bayesian Neural Networks to constrain the core center coordinates and its ground energy in Water Cherenkov Detectors. We describe the method and present some preliminary results focusing on the region where traditional methods face their major challenges, i. e., near detector borders and with the shower's core outside the detector's range.

5.2 Discussion

The presented technique can constrain the core position and its energy at some level near and outside the detector range. The Deep Learning methods such as presented here require a detailed exploration to fine-tune the architecture and hyper parameters, which can significantly improve the results. However, such an approach needs detailed and realistic simulations and should be carefully designed for a specific scientific experiment. It is worth noticing that the method has a better performance around 200GeV which is the region where more simulations are available. This suggests that the method's accuracy might be limited to the number of simulations in the range of energy. A change in array geometry, for instance, would require a new study in order to find the best solution. Nevertheless, the pursuance of such tailored made application is still a scientific case of interest as we can investigate the limits imposed by the data of one can analyse and constrain EAS near the borders and even outside the detector's range.

References

[1] Conceição, R., González, B. S., Guillén, A., Pimenta, M. and Tomé, B., *Muon identification in a compact single-layered water cherenkov detector and gamma/hadron discrimination using machine learning techniques*, Eur. Phys. J. C **81** (2021) 542.

[2] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, MIT Press (2016).

[3] B.C. Csáji, *Approximation with artificial neural networks*, *Faculty of Sciences, Etvs Lornd University, Hungary* **24** (2001) 48.

[4] B. Hanin, *Universal function approximation by deep neural nets with bounded width and relu activations*, *arXiv preprint arXiv:1708.02691* (2017) .

[5] C. Bom, J. Poh, B. Nord, M. Blanco-Valentin and L. Dias, *Deep learning in wide-field surveys: Fast analysis of strong lenses in ground-based cosmic experiments*, *arXiv preprint arXiv:1911.06341* (2019) .

[6] S. Lathuilière, P. Mesejo, X. Alameda-Pineda and R. Horaud, *Deepgum: Learning deep robust regression with a gaussian-uniform mixture model*, in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 202–217, 2018.

[7] V. Belagiannis, C. Rupprecht, G. Carneiro and N. Navab, *Robust optimization for deep regression*, in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2830–2838, 2015.

[8] Y. Gal and Z. Ghahramani, *Dropout as a bayesian approximation: Representing model uncertainty in deep learning*, in *international conference on machine learning*, pp. 1050–1059, 2016.

[9] J. Caldeira and B. Nord, *Deeply uncertain: Comparing methods of uncertainty quantification in deep learning algorithms*, *arXiv preprint arXiv:2004.10710* (2020) .

[10] J. Lee, Y. Bahri, R. Novak, S.S. Schoenholz, J. Pennington and J. Sohl-Dickstein, *Deep neural networks as gaussian processes*, *arXiv preprint arXiv:1711.00165* (2017) .

[11] M. Tan and Q. Le, *Efficientnet: Rethinking model scaling for convolutional neural networks*, in *International Conference on Machine Learning*, pp. 6105–6114, 2019.

[12] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand et al., *Mobilenets: Efficient convolutional neural networks for mobile vision applications*, *arXiv preprint arXiv:1704.04861* (2017) .

[13] M. Tan, B. Chen, R. Pang, V. Vasudevan, M. Sandler, A. Howard et al., *Mnasnet: Platform-aware neural architecture search for mobile*, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2820–2828, 2019.

[14] D. Heck, J. Knapp, J. Capdevielle, G. Schatz, T. Thouw et al., *Corsika: A monte carlo code to simulate extensive air showers*, *Report fzka* **6019** (1998) .

[15] J. Allison, K. Amako, J. Apostolakis, P. Arce, M. Asai, T. Aso et al., *Recent developments in geant4*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **835** (2016) 186.

PoS(ICRC2021)739