PROCEEDINGS OF SCIENCE

Search for optimal deep neural network architecture for gamma ray search at KASCADE

M. Kuznetsov,[a] **N. Petrov,**[a,b,c,*] **I. Plokhikh,**[a,b,d] **V. Sotnikov**[a,e] **and M. Tsobenko**[f]

[a] *Institute for Nuclear Research of the Russian Academy of Sciences, 117312, Moscow, Russia*

[b] *Novosibirsk State University, 630090, Novosibirsk, Russia*

[c] *Budker Institute of Nuclear Physics, 630090, Novosibirsk, Russia*

[d] *Institute of Thermophysics SB RAS, 630090, Novosibirsk, Russia*

[e] *JetBrains Limited, 2409 Nicosia, Cyprus*

[f] *Higher School of Economics University – St. Petersburg 194100, St. Petersburg, Russia*

E-mail: n.petrov@g.nsu.ru

We present the first steps of a search for high-energy (> 1 PeV) gamma rays in archival data of the KASCADE experiment. With the data collected from 1996 to 2013 the KASCADE statistics is comparable with that of modern observatories. The data is provided by the KASCADE Cosmic ray Data Center (KCDC) and publicly available. We employ methods of machine learning to distinguish between air showers produced by hadronic and gamma-ray primaries. For that we design primary particle type classifiers and train them with the KASCADE Monte-Carlo simulations. We compare results of several deep learning methods: a graph neural network, a self-attention network and a compact convolutional transformer. The level of hadronic background suppression with respect to gamma-ray signal in the best of these methods exceeds that of the original KASCADE method by more than an order of magnitude.

**** 27th European Cosmic Ray Symposium - ECRS ****
**** 25-29 July 2022 ****
**** Nijmegen, the Netherlands ****

[*]Speaker

PoS(ECRS)091

1. Introduction

Since the middle of the last century, one of the key issues in the physics of high energy cosmic rays was the question of presence or absence of gamma rays with energies above 100 TeV. This is very important test for theoretical models of acceleration and diffusion of galactic cosmic rays. Until recently, only the upper limits for the flux of gamma rays of these energies were set by several experiments [1–3]. The reason for this was the lack of sufficient sensitivity of the arrays and the methods of analysis used. In 2019 gamma rays with energies higher than 100 TeV were registered for a first time from the Crab Nebula by Tibet-ASγ [4] and HAWC [5]. It was followed by the observation of diffuse gamma rays with energies higher than 100 TeV by the Tibet-ASγ [6] and up to 1.4 PeV by LHAASO [7]. These observatories, that are situated at an altitude of more than 4 km above sea level and supplied with underground muon detectors, have extremely high sensitivity to distinguish between hadronic and gamma-ray air showers. On the other hand, it is possible to increase the respective sensitivity using modern methods of data analysis — methods based on the latest achievements in machine learning. Applying these techniques to cosmic ray data has already given good results in experiments such as IceCube [8], Pierre Auger [9], and Telescope Array [10]. Therefore, it is interesting to apply these methods to the archival data of similar experiments, which were processed by standard methods during their operation. One of these experiments with extensively documented and publicly available data and Monte-Carlo simulations is the KASCADE experiment, which collected and analyzed data from 1996 to 2013.

The KASCADE data are extensive air showers recorded by a ground-based array of scintillators and underground muon detectors, with energies ranging from $\sim$ 300 TeV to $\sim$ 100 PeV. In this study we use the data and Monte-Carlo provided by KCDC service [11]. The total exposure of the KASCADE experiment, accumulated over 15 years of measurements is approximately half of the exposure of LHAASO used to discover gamma rays with energies up to 1.4 PeV from galactic sources [7]. Thus, we can argue that the KASCADE data surely contains the events induced by the primary gamma rays. The final goal of our studies is to separate these events from the hadronic background. This implies that the survival rate of hadronic background with respect to gamma-ray signal for our machine learning techniques should be comparable to that of modern experiments: 10^{-5} — 10^{-6}. In this study we use several machine learning architectures: a graph neural network, a self-attention network and a compact convolutional transformer. We are reporting the recent status of these methods development and calibration with KASCADE Monte-Carlo simulations.

The text is organised as follows: in section 2 we outline the data and Monte-Carlo sets we use, in section 3 we briefly describe the machine learning methods we develop for the data analysis, in section 4 we show the performance of our methods on KCDC Monte-Carlo simulations, we conclude in section 5.

2. Data

In this research we use the KASCADE preselection data sets and Monte-Carlo simulations provided by the KCDC service [11]. This data includes information about the detected cosmic air showers events. The full archive consists of $\sim$ 300 M air shower events in energy range $\sim$ 1 – 100 PeV, detected by a 16x16 array of scintillating detectors during experiment operation from 1996 to 2013.

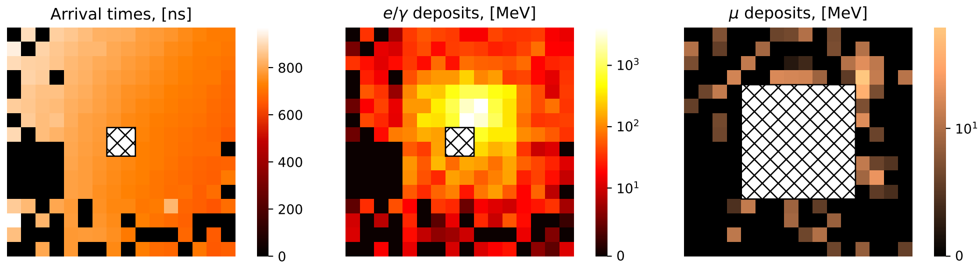


Figure 1: Example of the experimental event in the dataset. The matrices of arrival times, e/γ and μ deposits are shown. Reconstructed features of the event: $\log_{10} E$, [eV] = 15.54, θ = 19.81°. Note, KASCADE does not have detector stations in the central part 2x2 for arrival times and e/γ deposits and 8x8 for μ deposits, due to that we represent these areas as unresponsive.

Each event consists of energy deposits from e/γ and μ detectors and shower arrival times, and a set of reconstructed features. These features are: shower energy (E), shower core coordinates (x and y), zenith (θ) and azimuthal (ϕ) angles of the primary particle, muon ($\log_{10} N_\mu$) and electron ($\log_{10} N_e$) numbers and shower age (s).

The example of the experimental event is depicted in the FIG. 1. Arrival times and deposits in e/γ and μ detectors are represented as images of 16x16 shape. KASCADE does not have detectors in the central part (2x2 for e/γ, 8x8 for μ) and due to this we mark these areas as unresponsive.

For Monte-Carlo we use the CORSIKA [12] simulations provided by KCDC service. These simulations (MC) have the same properties as in real data, reconstructed using the actual detector response. Namely, we simulate events for gamma-ray and proton primaries using QGSJet-II.04 [13], EPOS-LHC [14] and Sybill 2.3c [15] hadronic interaction models. The primary energy spectrum is set to $\sim E^{-2}$. The number of simulated events for gamma rays and protons are $\sim 6 \cdot 10^4$ and $\sim 10^6$ respectively for QGSJet-II.04 hadronic interaction model and $\sim 3 \cdot 10^4$ and $\sim 6 \cdot 10^5$ respectively for others models.

According to the KASCADE photon search study [2] we use the following quality cuts: $x^2 + y^2 < 91$ m, $\log_{10} N_e > 4$. We relaxed the zenith angle cut from $\theta < 20°$ that was used in the original KASCADE study to $\theta < 40°$. We threat the ranges $\theta < 20°$ and $20° < \theta < 40°$ separately to test the reliability of the large zenith angle region.

We also use the real data set, we have divided it into "blind" and "unblind" parts in 80%:20% ratio. The unblind part is used in the process of the analysis. The blind part is set aside and will be used only to produce the physical results, when the research methodology will be tested and fixed. This way allows us to get rid of unconsciously changing the methodology in the process of the research to get a better result, but at the same time allows us to compare different distributions to make sure the methodology is correct. Therefore, in this paper all the figures mentioning experimental data represent its unblind part only.

3. Machine learning methods

We apply machine learning methods to separate KASCADE events produced by primary gamma rays and primary hadrons. We design, train and check out performance of the Self-

	MLP	**CCT**	**GCN**
Input	$e/\gamma + \mu$ deposits, flatten to 1x512	$e/\gamma + \mu$ deposits, as an image 2x16x16	a graph: $e/\gamma + \mu$ deposits as nodes; edges between stations + reconstructed features
Peculiarities	Attention	2D Convolutions, Attention, Pooling	Graph Convolutions
Spatial invariance	Non-invariant	Partially invariant	Non-invariant
Number of parameters	30 183	30 531	29 520

Table 1: The comparison of the neural net architectures used.

attention Multi Layer Perceptron (MLP), Compact Convolutional Transformer (CCT) and Graph Convolutional Network (GCN). These networks are considered as event-by-event classifiers in the research. The brief comparison of the models is shown in a Table 1.

All the neural networks are trained with proton and gamma Monte-Carlo sets. The gamma set is compiled of all hadronic models sets, while for proton we use QGSJet-II.04 set only. We devide each of these sets into train, validation and test subsets. We use train and validation sets for training models. In particular, validation set is used for early stopping to avoid overfitting and hyperparameter tuning of neural nets. All the metrics are evaluated on test subsets of the corresponding Monte-Carlo sets.

3.1 Self-attention Multi Layer Perceptron

The key feature of this model is the self-attention layers [16]. They are based on Vision Transformers [17], that are competitive to state-of-art convolutional networks in image classification tasks. The inputs for the MLP are energy deposits from e/γ and μ detector station represented as flatten 1D array. This classifier is implemented in TensorFlow [18] and has $\sim 30\,000$ trainable parameters.

3.2 Compact Convolutional Transformer

The second model we use is CCT [19]. This network architecture combines the convolutions and the standard vision transformer. Typically, this architecture has smaller number of trainable parameters than other transformers (while achieving similar performance), therefore it can be trained more easily. The input of the CCT are energy deposits from e/γ and μ detector stations represented as 2 channel (one for e/γ and one for μ) 16x16 image. In this research we implemented CCT using PyTorch [20]. The model has $\sim 31\,000$ trainable parameters.

3.3 Graph Convolutional Network

The last model we test is GCN. The key feature of this network is the graph convolutions [21]. This architecture is successfully used for cosmic rays classification in LHAASO [22]. The inputs of the model are the graph whose nodes are e/γ and μ deposits, while edges represent the distances between neighboring detector stations. This classifier is implemented with PyTorch Geometric [23] library. The model has $\sim 30\,000$ trainable parameters.

4. Results

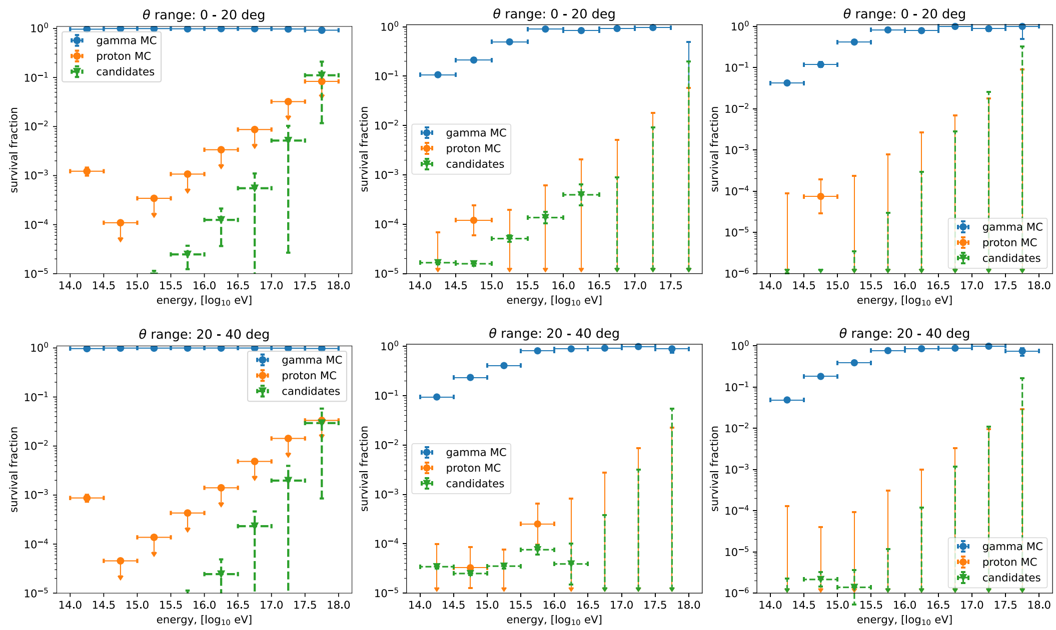
After the implementation of the models we evaluate their performance. We train the models on the mix of gamma-ray and proton Monte-Carlo. The latter is considered as a background for our classification task. We use protons as a background because they are the lightest nuclei and their air showers are the most similar to that of gamma rays, among other nuclei. The characteristic value we use for testing the neural nets performance is a gamma ray candidate survival rate (survival fraction) that is a ratio of a number of events classified as gammas in a given event set to a total number of events in the set. For an accurate classifier the survival rate should be high for gamma MC events and low for proton MC events. We compare the survival rates for gammas and protons for different primary energies and zenith angles. Additionally, we plot the rates for the experimental data to check the credibility of the classifier. For reliable classifiers we expect the data survival rates to be compatible with the proton rates.

The results are shown in Fig. 2. In this figure, "gamma MC" means the survival rate for gamma primaries, "proton MC" — for protons and "candidates" — for the unblind experimental data. At some energies only the upper limits stands for protons, this corresponds to a situation when zero proton events were classified as gammas in the particular energy bin. In this case the upper limits are inferred as 68% binomial confidence intervals for a given total number of events and given number of candidates. It is clear that the real survival fractions for protons are lower for these bins, while the present situation reflects only the fact that we do not have enough proton MC events at the given energies. The much lower survival rates obtained for the (almost totally hadronic) data also supports this interpretation and points to a possible true survival rates of the background events for a given classification method.

Comparing the results of the different classifiers one can see that in the observationally most interesting region (0.3 — 10 PeV) MLP and GCN have a survival rate for the data at the level 10^{-6} — $3 \cdot 10^{-5}$, that is comparable to the modern gamma observatories [6, 7] and more than order of magnitude stronger than the survival rate achieved in the original KASCADE analysis [2]. Also, one should note that the MLP method have an advantage of almost 100% survival rate for a gamma signal at all energies. The second thing we can infer from these plots is that at least for the MLP method the survival rates for the background and the signal do not depend much on the zenith angle (in fact for larger zenith angles the suppression of the background is even stronger). Both of these findings can help us to achieve the sensitivity enough to decipher the gamma signal from the KASCADE data. Although the additional proton MC is needed to verify the performance of the classifiers.

5. Conclusion

We have shown the first steps of the search for ultra-high-energy gamma rays based on archival data of the KASCADE experiment, provided by the KASCADE Cosmic ray Data Center (KCDC). We implemented and tested three different machine learning architectures: MLP, CCT and GCN for the event-by-event classification of the gamma-ray signal on top of the proton background. We also estimated the performance of these models with an unblind part of experimental data and found a reasonable agreement between data and MC in this aspect. The gamma-hadron separation



(a) **(b)** **(c)**

Figure 2: The dependence of the survival rate on the energy of the primary particle for (a) MLP, (b) CCT, (c) GCN. The upper panel shows the zenith angle range from 0 to 20 degrees, the lower panel — from 20 to 40 degrees.

power of our methods was found to be more than order of magnitude stronger than that of the original KASCADE gamma-ray search method. However, we need at least 10 times more hadronic Monte-Carlo events to make more thorough test of our classifiers.

Acknowledgments

The work is supported by the Russian Science Foundation grant 22-22-00883. We appreciate the contribution on the initial stages of this project made by our colleagues: Dmitry Kostunin, Vladimir Lenok and Victoria Tokareva.

References

[1] CASA-MIA collaboration, *Limits on the isotropic diffuse flux of ultrahigh-energy gamma radiation*, *Phys. Rev. Lett.* **79** (1997) 1805 [astro-ph/9705246].

[2] KASCADE Grande collaboration, *KASCADE-Grande Limits on the Isotropic Diffuse Gamma-Ray Flux between 100 TeV and 1 EeV*, *Astrophys. J.* **848** (2017) 1 [1710.02889].

[3] Y.A. Fomin, N.N. Kalmykov, I.S. Karpikov, G.V. Kulikov, M.Y. Kuznetsov, G.I. Rubtsov et al., *Constraints on the flux of* $\sim (10^{16} - 10^{17.5})$ *eV cosmic photons from the EAS-MSU muon data*, *Phys. Rev. D* **95** (2017) 123011 [1702.08024].

[4] M. Amenomori et al., *First Detection of Photons with Energy Beyond 100 TeV from an Astrophysical Source*, *Phys. Rev. Lett.* **123** (2019) 051101 [1906.05521].

[5] HAWC collaboration, *Measurement of the Crab Nebula at the Highest Energies with HAWC*, *Astrophys. J.* **881** (2019) 134 [1905.12518].

[6] Tibet ASgamma collaboration, *First Detection of sub-PeV Diffuse Gamma Rays from the Galactic Disk: Evidence for Ubiquitous Galactic Cosmic Rays beyond PeV Energies*, *Phys. Rev. Lett.* **126** (2021) 141101 [2104.05181].

[7] Z. Cao et al., *Ultrahigh-energy photons up to 1.4 petaelectronvolts from 12 γ-ray Galactic sources*, *Nature* **594** (2021) 33.

[8] IceCube collaboration, *Graph Neural Networks for low-energy event classification & reconstruction in IceCube*, *JINST* **17** (2022) P11003 [2209.03042].

[9] Pierre Auger collaboration, *Deep-learning based reconstruction of the shower maximum X_{max} using the water-Cherenkov detectors of the Pierre Auger Observatory*, *JINST* **16** (2021) P07019 [2101.02946].

[10] Telescope Array collaboration, *Mass composition of ultrahigh-energy cosmic rays with the Telescope Array Surface Detector data*, *Phys. Rev. D* **99** (2019) 022002 [1808.03680].

[11] A. Haungs et al., *The KASCADE Cosmic-ray Data Centre KCDC: Granting Open Access to Astroparticle Physics Research Data*, *Eur. Phys. J. C* **78** (2018) 741 [1806.05493].

[12] 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) .

[13] S. Ostapchenko, *QGSJET-II: towards reliable description of very high energy hadronic interactions*, *Nuclear Physics B-Proceedings Supplements* **151** (2006) 143.

[14] T. Pierog, I. Karpenko, J.M. Katzy, E. Yatsenko and K. Werner, *EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider*, *Phys. Rev. C* **92** (2015) 034906 [1306.0121].

[15] F. Riehn, R. Engel, A. Fedynitch, T.K. Gaisser and T. Stanev, *Charm production in SIBYLL*, *EPJ Web Conf.* **99** (2015) 12001 [1502.06353].

[16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez et al., *Attention Is All You Need*, *arXiv e-prints* (2017) arXiv:1706.03762 [1706.03762].

[17] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner et al., *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*, *arXiv e-prints* (2020) arXiv:2010.11929 [2010.11929].

[18] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*, *arXiv e-prints* (2016) arXiv:1603.04467 [1603.04467].

PoS(ECRS)091

[19] A. Hassani, S. Walton, N. Shah, A. Abuduweili, J. Li and H. Shi, *Escaping the Big Data Paradigm with Compact Transformers*, *arXiv e-prints* (2021) arXiv:2104.05704 [2104.05704].

[20] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan et al., *PyTorch: An Imperative Style, High-Performance Deep Learning Library*, *arXiv e-prints* (2019) arXiv:1912.01703 [1912.01703].

[21] T.N. Kipf and M. Welling, *Semi-Supervised Classification with Graph Convolutional Networks*, *arXiv e-prints* (2016) arXiv:1609.02907 [1609.02907].

[22] LHAASO collaboration, *Classifying cosmic-ray proton and light groups in LHAASO-KM2A experiment with graph neural network*, *Chin. Phys. C* **44** (2020) 065002 [1910.07160].

[23] M. Fey and J.E. Lenssen, *Fast Graph Representation Learning with PyTorch Geometric*, *arXiv e-prints* (2019) arXiv:1903.02428 [1903.02428].