Analysing arrival directions of ultra-high-energy cosmic rays with convolutional neural networks

Oleg Kalashev[1,2], Maxim Pshirkov[1,3,4] and Mikhail Zotov[5]

[1]Institute for Nuclear Research of the Russian Academy of Sciences, Moscow, 117312, Russia
[2]Moscow Institute for Physics and Technology, Dolgoprudny, Moscow Region, 141701 Russia
[3]Sternberg Astronomical Institute, Lomonosov Moscow State University, Moscow, 119992, Russia
[4]Lebedev Physical Institute, Pushchino Radio Astronomy Observatory, 142290, Russia
[5]Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Moscow, 119991, Russia

E-mail: kalashev@inr.ac.ru, pshirkov@sai.msu.ru, zotov@eas.sinp.msu.ru

Abstract.
The problem of identification of ultra-high-energy cosmic ray (UHECR) sources is greatly complicated by the fact that even the highest energy cosmic rays may be deflected by tens of degrees in the galactic magnetic fields. We show that arrival directions of UHECRs from several nearest active galaxies form specific patterns in the sky, which can be effectively recognized by convolutional neural networks. We use one of the recently developed convnet implementations for images defined on the sphere to train the classifier that is able to detect patterns that can be present in the experimental data. We calculate the minimal detectable from-source event fractions for several realistic source candidates and discuss the method limitations.

1. Introduction

The origin of ultra-high-energy cosmic rays (UHECRs, energies $\gtrsim 50$ EeV) remains one of the biggest mysteries of modern high-energy astrophysics. It is natural to expect that UHECR sources somehow trace the non-uniform local matter distribution since interaction with the photon background limits the distance of UHECR propagation to ~ 100. Thus, the distribution of UHECR arrival directions should be anisotropic. An investigation of this anisotropy and its properties is one of the main avenues in the UHECR research as it can strongly constrain characteristics of their sources.

In the present paper, we continue our studies of methods for probing a large-scale anisotropy arising from a presence of a nearby source, which we began in [1, 2] (Papers 1 and 2 further in the text). As a representative of a broader class of models, we used a particular model for cosmic rays by Kachelrieß, Kalashev, Ostapchenko and Semikoz (KKOS) [3]. The KKOS model assumes that ultra-high-energy cosmic rays are accelerated by (possibly a subclass of) active galactic nuclei (AGN). The energy spectra of nuclei are supposed to follow a power-law with a rigidity-dependent cut-off after the acceleration phase. An important prediction of the model is the existence of a nearby AGN (within ~ 20 Mpc) acting as a source of UHECRs. An accelerator of this kind would necessarily result in deviations from an isotropic distribution of arrival directions of UHECR.

In Paper 1, we investigated imprints that this deviation would leave on the angular power spectrum (APS) of the UHECR flux depending on the contribution from the strongest source. We assumed that the source of UHECRs is one of five nearby AGN (Cen A, M82, M87, Fornax A, and NGC 253). It was found that an observation of at least 200–300 events with energies above 57 EeV would be necessary to reject the hypothesis of isotropy with a high level of statistical significance providing that one of the sources forms around 10–15% of the total flux.

However, the APS calculation mostly destroys information about the characteristic shape and size of a region with events coming from a source. This motivated us to employ full information about arrival directions of UHECRs by using methods of machine learning, namely, a convolutional neural network (CNN) classifier. This approach was presented in Paper 2. Since UHECR arrival directions are distributed across the celestial sphere, the CNNs were trained on a HEALPix grid [4]. This approach considerably increased sensitivity, as now the number of events needed for establishing a large-scale anisotropy produced by UHECRs arriving from a nearby source was ~ 4 times smaller than in the case of the traditional APS analysis. For a fixed sample size, the CNN strongly decreased the fraction of UHECRs arriving from the source necessary for a robust detection of an anisotropy. We stress that our results are conservative in the sense that the KKOS model implies a heavy mass composition of UHECRs, which results in significantly more fuzzy patterns of arrival directions than in the case of a light composition.

However, we did not take into account the intrinsic uncertainties in the energy of primary cosmic rays arriving to Earth in Paper 2. This made samples used for training and testing the CNNs more accurate than in any experiment. Less energetic particles would be deflected by larger angles generally, populating the periphery of the region filled with UHECRs arriving from a particular nearby source. It is natural to expect that taking this uncertainty into account would diminish our results making them less relevant for experimental data. This motivated us to make the next step in the development of a deep-learning-based approach to anisotropy studies. It is demonstrated below that considering energy of UHECRs as an additional observable allows one to mitigate the effect of this uncertainty.

2. Method

Our analysis employs mock maps of UHECR arrival directions that we generated using the KKOS scenario as the starting point. In our scenario, all of the considered AGN share the same injection spectrum and composition. However, these properties evolve during propagation of nuclei from the source to the Milky Way, and we took this into consideration by using the TransportCR code [5] and considering only UHECRs with energies above 56 EeV. The high energy threshold allowed us to neglect deflections in extragalactic magnetic fields [6] and to assume that UHECRs arrive to the Milky Way within $\pm 1^\circ$ from the source position. Propagation from the Galactic boundary to the Earth was simulated using the CRPropa 3 code [7]. Employing the HEALPix grid with $N_{\rm side} = 512$, which provides the angular resolution of $7'$, allowed us to map accurately arrival directions of UHECRs of different rigidities at the Galactic boundary to arrival directions observed at Earth using the method of backpropagation.

Mock maps were generated in the following way. Given a source located at a given distance from the Galaxy, we took the propagated spectrum at the Galactic boundary calculated with TransportCR and sampled it $N_{\rm src}$ times, each time extracting some nuclei with an energy and charge (E, Z). The mapping obtained with CRPropa was used then to find an observed arrival direction of a cosmic ray for each (E, Z). For a sample of size N, the remaining events were generated following the isotropic distribution. The procedure was repeated thousands times for all sources. Fig. 1 provides an example of a sample of size $N = 100$ for Cen A assuming the Jansson–Farrar model of the GMF [8].

In Paper 1, we analysed these mock maps by calculating the angular power spectrum of UHECR arrival directions following a method suggested by the IceCube and the Pierre Auger

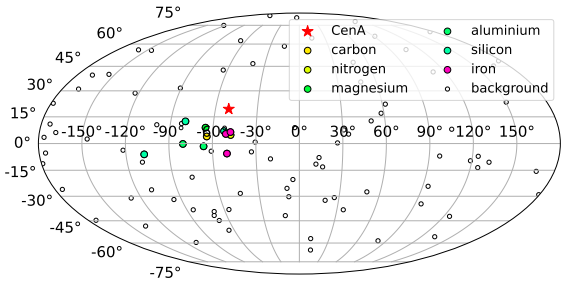

Figure 1. An example of a map of arrival directions of 100 UHECRs with 11 of them arriving from Cen A (shown in colour) and others distributed isotropically. The CNN developed in Paper 2 allows testing H_0 with samples like this one. The map is in the Mollweide projection with galactic coordinates.

Observatory collaborations [9, 10]. The APS was calculated using maps of the relative intensity of the UHECR flux obtained from these mock maps. After that, we calculated an estimator that quantified difference between an APS calculated for a flux containing an UHECR source and an APS of the isotropic flux. The null hypothesis H_0 was that arrival directions of a mixed sample of UHECRs follow an isotropic distribution. We assumed the value of the error of the second type $\beta = 0.05$ and searched for a minimal fraction of events arriving from a particular source in the total flux such that the error of the first type $\alpha \lesssim 0.01$.

Eventually, the problem of source identification could be reformulated as pattern recognition task, which can be addressed with convolutional neural networks [11]. Such neural networks use local feature maps at different scales to extract valuable information and perform a classification task. In the simplest case, it answers a question whether a map belongs to an "isotropic" or "with source" class. In cosmic ray physics, one needs to study an "image" on a sphere. Some implementations of spherical CNNs were proposed recently [12, 13, 14]. We used the code developed in [14] with only minor fixes and additions. The code implements the convolution and down-sampling operations on the HEALPix grid using Keras library [15]. The CNN developed in Paper 2 takes one feature map in the HEALPix grid with $N_{\rm side} = 32$ as an input.[1]

In the present work, we extended the initial method by increasing the number of feature maps analyzed by the CNN. We split the energy range of interest to bins of size Δ_b in $\lg E$ scale and calculate expected event density maps for each energy bin before sampling individual events. The from-source event density map calculation procedure is slightly different from the one used in Paper 2. As before, we take the propagated spectrum calculated with TransportCR for a source located at a given distance from the Galaxy and sample it $N_{\rm ini} = 100{,}000$ times, each time extracting some nuclei with an energy E and charge Z. An observed arrival direction of a cosmic ray is then found for each (E, Z) pair using the previously calculated mapping obtained with CRPropa code by backpropagation. Due to the uncertainty in energy determination, each sampled (E, Z) pair contributes to several neighbour energy bins with a weight given by the normal distribution integral taken with the limits defined by the bin boundaries:

$$w_i = \int_{\lg E_i}^{\lg E_i + \Delta_b} N(E, \Sigma_{\lg E}),$$

where $\Sigma_{\lg E}$ is proportional to the relative energy determination uncertainty:

$$\Sigma_{\lg E} \simeq \frac{1}{\ln 10} \frac{\Delta E}{E}.$$

[1] $N_{\rm side} = 32$ corresponds to the sphere divided into 12,288 cells with the angular size of 1.83°. This size is similar to the angular resolution of modern UHECR experiments.

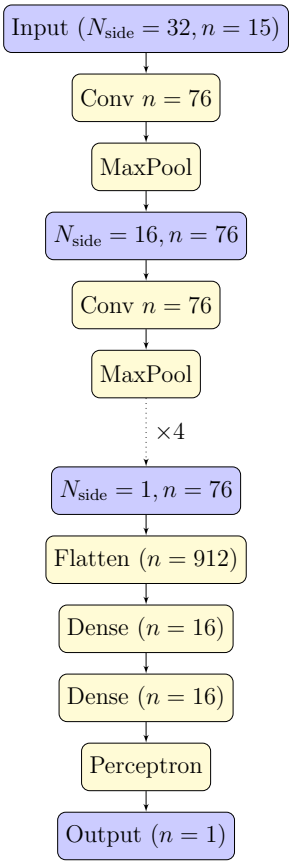

Figure 2. Architecture of the CNN developed in this work. Blue boxes are used to show feature vectors and maps. Yellow boxes show operations.

Maps generated this way are used then to sample $N_{\rm src}$ from-source events, characterized by HEALPix angular coordinates and an energy bin. The remaining $N - N_{\rm src}$ events are generated following the isotropic distribution assuming the same model energy spectrum, with the sampling procedure accounting for the energy determination error in the same way. The event sampling process is repeated multiple times in order to generate a sufficiently large number of maps for each source. Note that although we impose the lower cut $E > 56$ EeV on the observed event energies as in Paper 2, this time we self-consistently account for the influence of the lower energy part of the model on the analysis because of the non-vanishing energy determination uncertainty. This is one of the major advantages of the new approach. We assume the value of 20% for $\Delta E/E$, which follows an estimate for the POEMMA mission [16].

The binary classifier neural network is trained to discriminate the maps generated assuming zero and alternative hypotheses. Because of the energy binning introduced, the architecture (see Fig. 2) has been modified to take as an input a collection of maps instead of single map. We also had to repeat the model meta-parameter optimization procedure. As before, we tried to adjust the HEALPix grid size along with number of filters and convolutions. Besides this, we found it useful to add two extra dense layers after the sequence of convolutions. The log energy bin size Δ_b was also one of the free parameters being optimized. Among the three values probed, 1/5, 1/10 and 1/20, the smallest one was found to be optimal, which results in 15 energy bins in the range $56 < E/\text{EeV} < 315$ EeV.

The implementation [14] for the convolution operations on a sphere which we use, provides an

Table 1. Percentage of UHECRs arriving from the candidate sources in samples of sizes $N = 50, 100, \ldots, 500$ such that the error of the first type $\alpha \lesssim 0.01$ for the null hypothesis of isotropy H_0 providing the second kind error $\beta = 0.05$, obtained with the new binned convolutional neural network classifier (BCNN) and with the previous CNN model. One source at a time was considered.

Source	Method	50	100	200	300	400	500
NGC 253	BCNN	12	8	5	4	3.25	3.2
	CNN	12	7	4.5	3.67	3	2.6
Cen A	BCNN	16	10	7	5.67	4.75	4.2
	CNN	16	11	7	5.67	5	4.4
M 82	BCNN	20	14	8.5	6.67	5	4.4
	CNN	20	12	7	6	4.75	4.2
M 87	BCNN	22	15	11.5	7	6.25	5.4
	CNN	22	14	9	8	6.25	5.2
Fornax A	BCNN	16	10	6	5	4.5	3.8
	CNN	16	9	6	5	4.5	3.8

analogue for the 3×3 convolutions on a flat image. The resulting CNN takes $N_{bins} = 15$ feature maps in the HEALPix grid with $N_{\rm side} = 32$ as an input; $N_f = 76$ feature maps are built at the first step using the convolution operation with $3 \times 3 \times N_f \times N_{bins}$ free parameters and max-pooling the image to $N_{\rm side} = 16$. The sequence of convolutions and max-pooling operations is repeated until $N_{\rm side} = 1$ with the persistent number of feature maps, which means that each intermediate convolution operation has $3 \times 3 \times N_f^2$ trainable weights. The rectified linear activation function is used for all intermediate layers. Finally, 76 feature maps with $N_{\rm side} = 1$ are flattened and processed by a multilayer-layer perceptron with two inner layers of size 16 followed by a single neuron output with sigmoid activation.

To avoid overfitting, we used an early-stop technique. Each batch of training sets was generated randomly which also helped to prevent ovefitting. For this reason we did not use any other regularization techniques, such as dropout and the L2 regularization. The validation set consisted of 10,000 samples generated using different random seeds. The weights were optimized using the Adadelta adaptive learning rate method [17]. The trained model was evaluated on the test set consisting of 50,000 map samples. The output layer of the classifier had sigmoid activation function. The output value, a number between 0 and 1 was used as the test statistic.

The minimal fractions η of from-source UHECRs needed to reject the null hypothesis of an isotropic flux with the same demands on α and β as above, are presented in Table 1. It is clearly seen that results obtained with energy uncertainties taken into account are essentially the same as those obtained in Paper 2 despite of the more fuzzy patterns of UHECR arrival directions. This became possible due to the binned classifier that employs energy as an additional observable.

3. Conclusions

We demonstrated in Paper 2 how one can strongly improve the efficiency of an analysis of arrival directions of UHECRs by introducing a deep convolutional neural network trained on a HEALPix grid. The basic idea was to train a classifier which discriminates samples generated assuming

null (isotropy) and alternative (anisotropy) hypotheses and to use the classifier output as a test statistic. An application of models that involve pattern recognition such as the suggested CNN gave a qualitative enhancement in terms of sensitivity to deviations from an isotropic distribution of arrival directions. It was shown in particular that the method allows decreasing the minimal number of events necessary to reject the null hypothesis by ~ 4 times. This reduces technical demands and the required total exposure of an UHECR experiment drastically.

However, the model considered was a simplified one since it did not take into account uncertainties in the energy of detected UHECRs, which influence the spectrum and thus the shape of patterns formed by nuclei arriving from a source. As a result, our estimates could be too optimistic. We addressed the problem in the present work by introducing a binned classifier by modifying the neural network developed in Paper 2. It was demonstrated that even under conditions of a 20% uncertainty in the energy resolution (which is higher than in contemporary ground experiments), the quality of the new classifier did not decrease. More than this, an introduction of energy as a new observable opens a way for an improvement of the efficacy of discriminating samples generated assuming isotropy and anisotropy hypotheses. This will be studied in more details elsewhere.

The source code and supplemental materials for this work, including trained classifier models, can be downloaded from the project web page [18].

Acknowledgments

The development of the classification method and the architecture of the corresponding deep convolutional neural network is supported by the Russian Science Foundation grant 17-72-20291. The authors acknowledge support of the Interdisciplinary Scientific and Educational School of Lomonosov Moscow State University "Fundamental and Applied Space Research."

References

[1] Kalashev O, Pshirkov M and Zotov M 2019 *JCAP* **09** 034 (*Preprint* 1810.02284)
[2] Kalashev O, Pshirkov M and Zotov M 2020 *JCAP* **11** 005 (*Preprint* 1912.00625)
[3] Kachelrieß M, Kalashev O, Ostapchenko S and Semikoz D V 2017 *Phys. Rev.* **D96** 083006 (*Preprint* 1704.06893)
[4] Górski K M, Hivon E, Banday A J, Wandelt B D, Hansen F K, Reinecke M and Bartelmann M 2005 *Astrophys. J.* **622** 759–771 (*Preprint* astro-ph/0409513)
[5] Kalashev O E and Kido E 2015 *J. Exp. Theor. Phys.* **120** 790–797 (*Preprint* 1406.0735)
[6] Pshirkov M S, Tinyakov P G and Urban F R 2016 *Phys. Rev. Lett.* **116** 191302 (*Preprint* 1504.06546)
[7] Alves Batista R, Dundovic A, Erdmann M, Kampert K H, Kuempel D, Müller G, Sigl G, van Vliet A, Walz D and Winchen T 2016 *JCAP* **1605** 038 (*Preprint* 1603.07142)
[8] Jansson R and Farrar G R 2012 *Astrophys. J.* **757** 14 (*Preprint* 1204.3662)
[9] Hülss J and Wiebusch C 2007 Search for signatures of extra-terrestrial neutrinos with a multipole analysis of the AMANDA-II sky-map *Proceedings, 30th ICRC: Merida, Mexico, July 3–11 2007* (*Preprint* 0711.0353)
[10] Aab A *et al.* (Pierre Auger) 2017 *JCAP* **1706** 026 (*Preprint* 1611.06812)
[11] LeCun Y, Boser B, Denker J S, Henderson D, Howard R E, Hubbard W and Jackel L D 1989 *Neural Computation* **1** 541–551 ISSN 0899-7667
[12] Cohen T S, Geiger M, Koehler J and Welling M 2018 Spherical CNNs *International Conference on Learning Representations* p 542 (*Preprint* 1801.10130)
[13] Perraudin N, Defferrard M, Kacprzak T and Sgier R 2019 *Astron. Comput.* **27** 130–146 (*Preprint* 1810.12186)
[14] Krachmalnicoff N and Tomasi M 2019 *Astron. Astrophys.* **628** A129 (*Preprint* 1902.04083)
[15] Chollet F *et al.* 2015 Keras https://keras.io
[16] Olinto A V *et al.* (POEMMA) 2021 *JCAP* **06** 007 (*Preprint* 2012.07945)
[17] Zeiler M D 2012 *CoRR* (*Preprint* 1212.5701) URL https://arxiv.org/abs/1212.5701
[18] Kalashev O, Pshirkov M and Zotov M 2020 Supplemental materials and source code for this work https://github.com/okolo/ml_cr_aniso