Determination of Spherical Coordinates of Sampled Cosmic Ray Flux Distribution using Principal Components Analysis and Deep Encoder-Decoder Network

Tomasz Hachaj[1] , Marcin Piekarczyk [1] , Łukasz Bibrzycki[2] and Jarosław Wąs[1]

[1]*Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering,*
AGH University of Krakow, Kraków, Poland
[2]*Faculty of Physics and Applied Computer Science,*
AGH University of Krakow, Kraków, Poland
*Corresponding author: Tomasz Hachaj thachaj@agh.edu.pl

Abstract In this paper we propose a novel algorithm based on the use of Principal Components Analysis for the determination of spherical coordinates of sampled cosmic ray flux distribution. We have also applied a deep neural network with encoder-decoder (E-D) architecture in order to filter-off variance noises introduced by sampling. We conducted a series of experiments testing the effectiveness of our estimations. The training set consisted of 92 250 images and validation set of 37 800 images. We have calculated mean absolute error (MAE) between real values and estimations. When E-D is applied, the number of cases (estimations) where MAE < 10 increases from 48% to 79% for θ and from 62% to 65% for ϕ, MAE < 5 increases from 24% to 45% for θ and from 47% to 52% for ϕ, MAE < 1 increases from 6% to 9% for θ and from 12% to 16% for ϕ, where θ is the zenith angle, and ϕ is the azimuthal angle. This is a significant change and it demonstrates the high utility of the E-D network use and shows the accuracy of the PCA-based algorithm. We also publish the source code used in our research in order to make it reproducible.

Keywords: cosmic-ray shower, spherical coordinates, detector grid, Principal Component Analysis, Encoder-Decoder network.

1. Introduction

The ultra-high-energy cosmic radiation reaching the Earth's atmosphere is extensively studied because of its still unknown sources and mechanisms of acceleration as well as because of the implications for the dynamics of the atmosphere, life on Earth, interferences with electronic systems and even possible correlations with seismic phenomena [25], to name just a few [5]. Practical exploration of these phenomena is based on observations obtained from specialized detectors capable of detecting secondary jets produced in the atmosphere and reaching the Earth's surface. These jets can arise due to atmospheric collisions of either single primary high-energy particles or cosmic ray ensembles (CRE), i.e., groups of cosmic rays generated in outer space.

Such observations are made primarily by large-scale infrastructure detectors in projects such as Pierre Auger Observatory in Argentina [48], IceCube in Antarctica [1, 2] or Baikal-GVD at Lake Baikal in Russia [6, 44]. Due to their fixed location, such installations have a limited detection area. To expand the possibilities in this regard,

observational structures involving small-scale detectors distributed around the world have been proposed. Projects such as CRAYFIS [32], DECO [51], CREDO [9, 24] incorporate widely available mobile devices like smartphones and webcams into the citizen science paradigm. Projects allied within the CREDO consortium aggregate observations obtained from a variety of simple and low-cost detectors that can be densely distributed over a local area [29]. The novel image processing algorithms make it possible to detect potential cosmic ray events using even off-the-shelves CMOS cameras [21].

Recently, advanced AI methods have been widely used to analyze such data. The potential of such techniques is used both for low-level recognition of detector signals [8, 19, 37, 53] and globally to detect features and correlations for surface or distributed detector networks [13, 28, 31, 45, 49]. The latest scientific and implementation research allows for real-time detection of potentially anomalous particle tracks and similar particle tracks detection in big data image datasets acquired by CMOS sensors [22, 38]. We need simulations to understand spatial distribution of showers [16, 34]. In this paper, we propose an AI-based method to disentangle the directional information from sparse lateral distributions.

AI methods used in detection of ultra-high-energy particles are basically determined by the types of measurement sensors that are used to detect cosmic rays. Stationary observatories such as Pierre Auger, IceCube or Baikal-GVD use a well-defined spatial arrangement of homogeneous detectors. For this reason, this is a fundamentally different research problem than the one considered in our work, which is a feasibility study aimed at proving that the jet geometry can be reconstructed with non-homogeneous but very flexible cosmic rays detectors set-up. This set-up may consist of various types of detectors, both industrial made and amateur constructions, with almost arbitrary localizations which fit into the general notion of Citizen Science paradigm. The paper proves that it is possible and our method is directly useful in the design of small-scale complex cosmic-ray exposure (CRE) secondary flux detection systems which can be a part of distributed cosmic ray observatories like CREDO.

To the best of our knowledge, the results presented in this work are pioneering in the design of small-scale complex CRE secondary flux detection systems, and it is difficult to point out research results with which to contrast our proposed method.

2. Material and methods

2.1. Muon lateral distribution

In order to generate the simulated cosmic shower, we used the approach previously described in the paper [20]. The equation describing the muon distribution is shown in equation (1) (muon is an elementary particle similar to the electron but with a much greater mass). According to this equation the distribution is singular at $r = 0$ for typical

values of the age parameter s, thus needs to be truncated for distances smaller than $r_{\min}$ [12, 17].

$$\rho_\mu(r) = \begin{cases} \rho_\mu(r_{\min}), & r < r_{\min}\,, \\ \frac{N_\mu}{2\pi r_0^2}\frac{\Gamma(4.5-s)}{\Gamma(s)\Gamma(4.5-2s)}\left(\frac{r}{r_0}\right)^{s-2}\left(1+\frac{r}{r_0}\right)^{s-4.5}, & r \geq r_{\min}\,, \end{cases} \qquad (1)$$

where s is the age parameter [52], N_μ is the shower size parameter, and r_0 describes the characteristic size of the shower.

It should also be taken into account that the cosmic shower can hit the ground at different angles, which can be defined using the spherical coordinates $\theta \in [0, \pi/2]$ and $\phi \in [0, \pi]$ (r is constant), θ is the zenith angle, and ϕ is the azimuthal angle [12]. There is no need to consider a larger range of angles, since the distributions in them at the ground are periodic (see visualizations in [20]). Example distributions depending on the angles of (θ, ϕ) can be seen later in this paper (Figures 6, 7); a detailed analysis can be found in [20]. The paper [20] also presents the relationship between spherical coordinates θ, ϕ and the distribution of projection of cosmic ray shower registered on ground.

Finding a pair of angles (θ, ϕ) on the basis of statistical analysis of the distribution of cosmic shower particles on the earth's surface makes it possible to determine the direction from which the cosmic shower came. The proposition and evaluation of the algorithm for determination of spherical coordinates of sampled cosmic ray flux distribution using data acquired by grid of detectors in the presence of background noise is the main objective of this paper.

2.2. Calculation of angles (θ, ϕ) based on analysis of cosmic ray particle distribution using Principal Components Analysis

Determining the (θ, ϕ) pair of angles allows us to determine from which direction the cosmic rays came.

An Img matrix of size $n \times m$ is given, which represents a discrete measurement grid of cosmic rays. The value in each grid field is equal to the number of particles that have been recorded in that grid field. Suppose there is a recorded cosmic shower inside the grid with a distribution consistent with (1) but with unknown values of (θ, ϕ). The Algorithm 1 allows us to estimate these unknown angles using PCA [27]. PCA is a popular and proven technique, which in different variants allows analyzing the distribution of data depending on angular parameters for example in geodesic [18, 33, 41, 42] or climatological data [40]. Paper [47] reports successful application of PCA-based analysis of cosmic-ray data for extraction of Hale Cycle. In [43] authors perform fully empirical atmospheric correction of cosmic ray data using PCA. Because of those facts, PCA seems to be promising for analysis of another natural phenomena like analysis of spherical coordinates of cosmic ray flux distribution.

Data: Input: Img – input image with dimensions $n \times m$
Result: (θ, ϕ) – a pair of spherical coordinates.
```
// Resample image to range [0,255] making values discrete
```
Img = Integer(255 · (Img - min(Img)) / max(Img));
```
// Create empty list of points
```
Points $\leftarrow \emptyset$;
```
// Iterate through all grid points and add as many times a point with the
    given coordinates as the number of times a particle has been registered
    in it
```
for $x = 0; x < n; x++$ **do**
    for $y = 0; y < m; y++$ **do**
        for $c = 0; c < \text{Img}[x, y]; c++$ **do**
            Points.Append$((x, y))$;
        end
    end
end
```
// mean1, mean2 - mean points value; v1, v2 - PCA components; exv1, exv2 -
    explained variance
```
[mean, $v0$, $v1$, exv0, exv1] = PCA(Points);
```
// modify the first vector in order to have first coordinate positive
```
if $v1[0] < 0$ **then**
    $v1 = -1 * v1$;
end
$\theta = \arccos(\text{exv1}/\text{exv0})$;
$\phi = \pi - \arctan2(v1[0], v1[1])$;
return (θ, ϕ);

Algorithm 1: Estimating the angles (θ, ϕ) of the distribution (1).

In practice, however, we will not have such a dense measurement grid to be able to measure the distribution of particles at discrete points in contact. Suppose we have a square grid in which the distance between the particle detectors horizontally and vertically is constant at d (see Figure 1. Let us denote the sampled Img grid as Img_d.

Analysis of variance based on a set of relatively distant samples might be biased. To increase the spatial density of the samples, we can use a convolution with a Gaussian kernel with a size proportional to the sampling d [26]. In our case, we proposed a filter size equal to $4 \cdot d + 1$ with $\sigma = 0$, which has a large enough diameter to cancel out the sampling "holes" (2):

$$\text{Img}_{d,\text{Gauss}} = \text{Img} \otimes \text{GaussianKernel}_{(4 \cdot d + 1, 4 \cdot d + 1)}, \qquad (2)$$

where $\otimes$ denotes convolution. For the purpose of performing variance analysis with the Algorithm 1, it is not necessary to use a high-resolution grid. If the distribution is inside
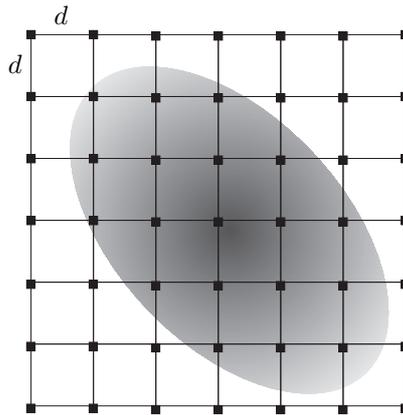
Fig. 1. An illustrative drawing showing a 7×7 grid of detectors, spaced by d in both directions. The detectors are black squares that are equally spaced along the x and y axes. A hypothetical histogram of the density of the particle distribution is shown in the background.

the sampled area, which preserves its spatial contour, it may be possible to resampling the original grid to a given lower resolution. This can be done using, for example, the following proposed Algorithm 2. It allows us to create fewer samples while preserving the sum of the detected particles.

2.3. Enhancing sampled distribution image by deep convolutional encoder-decoder

The method proposed in the previous section for determining the parameters of (θ, ϕ) consists of three steps: registration of radiation samples, Gaussian filtering (2), resampling with the Algorithm 2, and estimation of angles using the Algorithm 1. We can try to improve the reconstruction of the original particle distribution by using an encoder-decoder (E-D) neural network [14, 15, 23, 46]. The role of this network will be to reconstruct the original pre-sampled signal, but after applying Gaussian filtering (2) and rescaling with Algorithm 2. For this purpose, we used the following network:

- Encoder:
  - Convolution layer with 16 3×3 filters followed by ReLU activation and max pooling 2×2;
  - Convolution layer with 8 3×3 filters followed by ReLU activation and max pooling 2×2;
  - Convolution layer with 8 3×3 filters followed by ReLU activation and max pooling 2×2;
- Decoder:
  - Convolution layer with 8 3×3 filters followed by ReLU activation and up-sampling 2×2;
  - Convolution layer with 8 3×3 filters followed by ReLU activation and up-sampling 2×2;
  - Convolution layer with 16 3×3 filters followed by ReLU activation and up-sampling 2×2;
  - Convolution layer with 1 3×3 filter followed by sigmoid activation.

Data: Input: Img – input image with dimensions $n \times m$; scale – resampling factor
        (scale > 1)
Result: Img_{res} – resampled image.
```
// calculate the size of the resampled image
```
xSize = Integer(n / scale);
ySize = Integer(m / scale);
```
// initialize the resulting matrix with zeros
```
Img_{res} = zeros(xSize, ySize);
```
// iterate through all grid points
```
for $x = 0; x < \text{xSize}; x++$ **do**
    for $y = 0; y < \text{ySize}; y++$ **do**
        sum = 0;
        for $a = 0; a < \text{scale}; a++$ **do**
            for $b = 0; b < \text{scale}; b++$ **do**
                sum = sum + Img$[(\text{scale} * x) + a, (\text{scale} * y) + b]$;
            end
        end
        $\text{Img}_{\text{res}}[x, y]$ = sum;
    end
end
return Img_{res};

Algorithm 2: Resampling with preservation of the sum of detected particles.

In training, we will use the binary cross entropy loss function. The training data consisted of pairs:

- as an input Img filtered by (2), resampled by Algorithm 2 and scaled to discrete (integer) range [0,255]
- as an output Img_d filtered by (2), resampled by Algorithm 2 and scaled to discrete (integer) range [0,255]

The resulting image generated by the encoder-decoder described in this section is then subjected to the Algorithm 1 to estimate the (θ, ϕ) angle pair. In Figure 2 we present a diagram that explains the proposed method.

3. Results

In this section, we will describe the validation tests of our PCA-based Algorithm 1 and how the use of E-D network affects the resulting angle pair estimates (θ, ϕ).

To test the performance of our method, we specified the following particle flux parameters (1): $N_\mu = 10^6$ (corresponds to a primary particle energy of more than $10^{16} \, eV$ [39]), $r_0 = 100$ (this value is compatible with the Molière radius in the Earth's atmosphere at
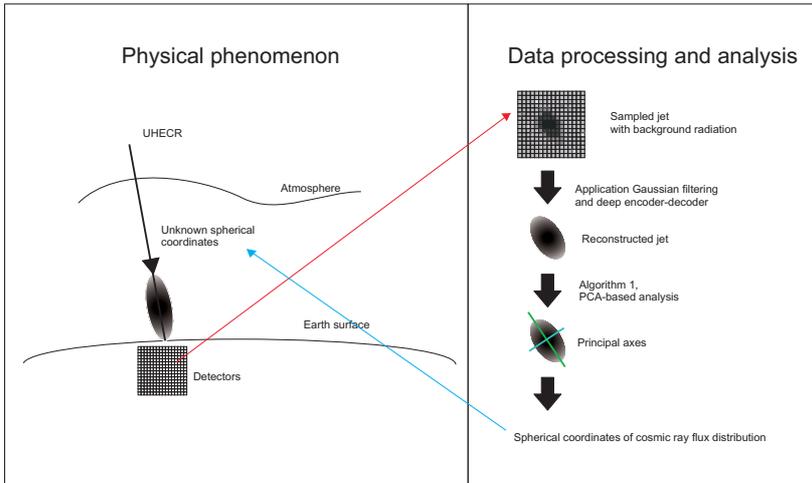
Fig. 2. Diagram that explains the proposed method. Ultra-high energy cosmic ray (UHECR) with unknown spherical coordinates generates a jet that is observed on Earth surface by the detectors. Data registered by detectors is sampled and mixed with background radiation (noise). After applying Gaussian filtering, deep encoder-decoder network reconstructs the original jet and Algorithm 1 is used to calculate spherical coordinates of the cosmic ray flux distribution.

the ground level [3, 4]), $s = 1.3$ [20]. We assumed that the particle flux is recorded over an area of 800×800 cm. The area is divided by a 1x1 cm grid. This means that the initial Img image has a resolution of 800×800 ($n = m = 800$). Let us assume that in the 800×800 cm area at intervals of $d = 25$ cm there are detectors equally spaced horizontally and vertically, each with an area of 1×1 cm. This means that we sample Img with $32 \times 32 = 1024$ samples (detectors) thus obtaining Img_d. We assume that each detector is capable of recording all the radiation particles that hit it during a single event. We have assumed the value of background radiation according to [35, 36], where the muon flux density at the earth's surface is $1 \frac{\text{muon}}{\text{cm}^2 \cdot \text{min}}$. Thus, we can assume that over a period of 1 second, the background radiation density at 1 cm^2 of the earth's surface averages $\rho = \frac{1}{60} \frac{\text{muons}}{\text{s} \cdot \text{cm}^2} = 0.01(6) \frac{\text{muons}}{\text{s} \cdot \text{cm}^2}$.

We conducted a series of experiments testing the effectiveness of the Algorithm 1.

1. Angle estimation based on Img with resolution 800×800;

2. Angle estimation based on Img filtered by (2) and resampled with Algorithm 2 to a resolution of 80×80 pixels.

3. Angle estimation based on Img_d filtered by (2) and resampled with the Algorithm 2 to a resolution of 80×80 pixels.

4. Angle estimation based on Img_d filtered by (2) and resampled with Algorithm 2 to a resolution of 80×80 pixels and reconstructed by E-D network.

Encoder–decoder training
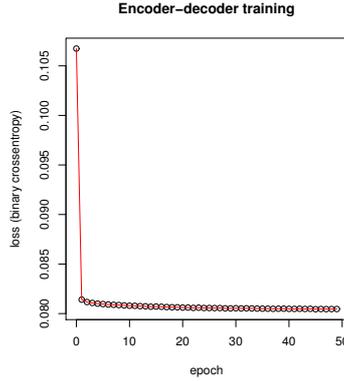
Fig. 3. A plot of loss function value during training.

The last two cases in the list above (3 and 4) are real-life scenarios. Cases 1 and 2 are used to check the validity of the Algorithm 1 assumptions, since in practice in these cases we have a very densely sampled distribution, which is not very realistic.

In order to train the E-D network, we generated distributions (1) in which the angle of $\theta \in \{0, 2, 4, ..., 80\}$, $\phi \in \{0, 2, 4, ..., 178\}$ (a total of 3690 distributions). We then discretized the distributions to an 800×800 grid by adding offsets along the x and y axes of $\{0, 3, 6, 9, 12\}$ cm. Thus, the final training set consisted of 92 250 discrete images. We filtered the 800×800 image (2) and resampled with the algorithm 2 to resolution 80×80. These images were the input to the E-D network. In the output we used the same data Img_d where $d = 25$. We filtered Img_d by (2) and resampled with Algorithm 2 to resolution 80×80. In this case, the validation set was not needed because the validation was done as part of the validation of the entire Algorithm 1 (see discussion below). We used the optimization algorithm Adam [30] with learning rate = 0.001. The training lasted 50 epochs. A plot of loss function is presented in Figure 3.

We implemented our approach in Python 3.8 using Keras/Tensorflow 2.8, scipy 1.8 and opencv-python 4.5 libraries. Significant speed-ups in generating distributions (1) were achieved using the numba 0.56 library. The entire experiment including data generation, network training and validation Algorithm 1 on a PC computer with Intel Core i7 3.00 Ghz; 64 GB RAM, Windows 10 OS took more than 3 days to execute. Some of the figures were made in R 3.6 language using dplyr 0.8 library and ggplot2 3.4.

In order to test the performance of our method, we generated a validation set of distributions (1) in which the angles $\theta \in \{1, 3, 5, ..., 83\}$, $\phi \in \{1, 3, 5, ..., 179\}$. In addition, we introduced a random offset of the distribution along the x and y axes in the range of values [0,12] cm which corresponds to half the distance between the positions of the simulated particle detectors. For each test configuration of the (θ, ϕ) pair, we performed 10 independent simulations (1) and background radiation with random offset values.

There were 37 800 images in total. We then applied the Algorithm 1 to each test case 1, 2, 3 and 4. The results for each pair (θ, ϕ) were averaged and the estimated angles (θ, ϕ) are shown in Figures 4 and 5. All the results are shown in degrees.

The source codes we have written can be downloaded from [10]. That online repository contains all source codes that are required to replicate the study including data generation script, network training, evaluation and plotting the results. All calculations included in this work were made using source codes from that repository.

The examples of results form the presented analysis made with the Algorithm 1 are shown in Figs. 6 and 7.

4. Discussion

The training of the E-D network, which changes in the loss function in successive iterations can be seen in Figure 3 was stable. After 50 epochs, loss had a binary cross entropy value of 0.080, which has remained virtually unchanged since epoch 40.

As can be seen in the Figures 6 and 7, the Algorithm 1 using PCA to detect the directions along which the largest variance is found works as expected. The axes are found with relatively small error, so that the estimation of the angle θ, which is calculated directly from the first axis of the PCA, is precise. In the case of the angle ϕ, for the calculation of which the variance ratio along the PCA axis is used, images that have more noise such as Img+sampled+Gauss and Img+sampled+Gauss+resampled have a less precise estimate of the ϕ angle than the img+Gauss+resampled and Img+sampled+Gauss+resampled+E-D example. Figures 6 and 7 also show that the proposed E-D is effective in cleaning Img+sampled+Gauss+resampled from measurements that cause disturbances in the variance estimation, which directly translates into improved estimation of θ.

These conclusions are supported by detailed statistic studies, the results of which we present in Figures 4 and 5. These figures show mean absolute error (MAE) estimates of (θ, ϕ). In the case of the full Img image, the θ angle estimate is very accurate and decreases due to resampling. The largest error in determining the angle θ is in the area of small value of this angle (less than 10 to 20 degrees) because then the change in the statistical distribution of radiation particles is not large enough to be accurately estimated by PCA. A similar phenomenon also occurs in the case of the angle ϕ, when the largest estimation error is in the interval [0, 10] and [170, 180]. This is due to trigonometric periodicity. Both of the phenomena discussed above are expected and easily explained, which proves the stability of our proposed method. As can be seen in Figure 5, the use of the E-D network significantly improves the accuracy of the angle θ estimate, which is calculated based on the variance along the PCA axis sometimes by as much as 20 degrees on average, compared to the estimate without the E-D architecture. The larger the actual θ angle is, the more the particle distribution is "stretched" in space

(a) Estimation of θ on Img (800×800).

(b) Estimation of ϕ on Img (800×800).

(c) Estimation of θ on Img convolved by Gaussian and resampled (80×80).

(d) Estimation of ϕ on Img convolved by Gaussian and resampled (80×80).

Fig. 4. MAE of estimations of (θ, ϕ) from Algorithm 1. Each measurement is averaged over 10 trials.

(a) Estimation of θ on Img_d convolved by Gaussian and resampled (80×80).

(b) Estimation of ϕ on Img_d convolved by Gaussian and resampled (80×80).

(c) Estimation of θ on Img_d convolved by Gaussian, resampled (80×80) and processed by E-D.

(d) Estimation of ϕ on Img_d convolved by Gaussian, resampled (80×80) and processed by E-D.

Fig. 5. MAE of estimations of (θ, ϕ) from Algorithm 1. Each measurement is averaged over 10 trials.
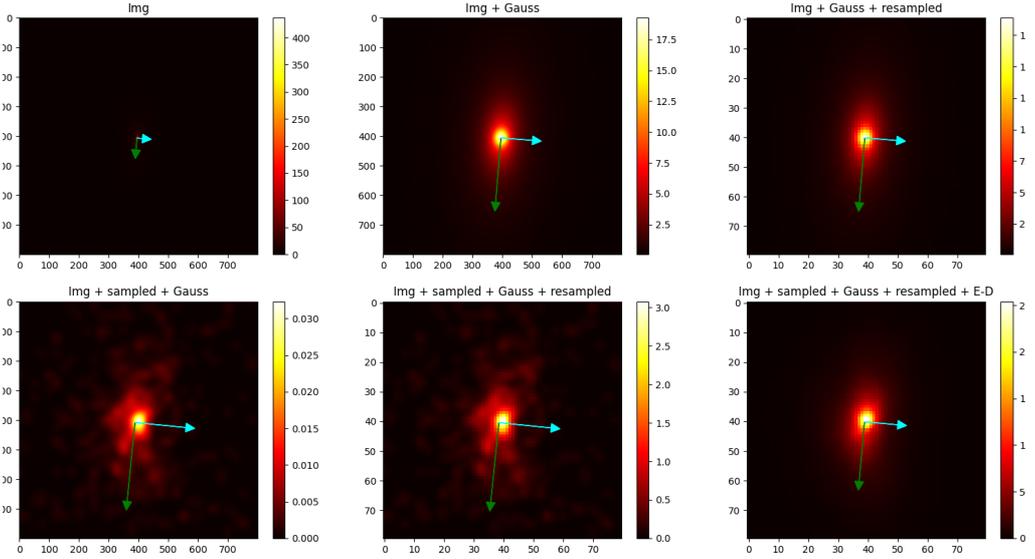
Fig. 6. Examples of results of Algorithm 1 for $(\theta = 62, \phi = 85)$. From left to right, from top to bottom (see also the titles of images): estimation of these angles for Img (800×800) equals $(\theta = 62.64, \phi = 85.15)$, for Img convolved by Gaussian it is $(\theta = 60.96, \phi = 85.38)$, for Img convolved by Gaussian and resampled to 80×80 it is $(\theta = 60.77, \phi = 85.21)$, for sampled Img_d convolved by Gaussian (800×800) it is $(\theta = 50.17, \phi = 84.34)$, for sampled Img_d convolved by Gaussian and resampled to 80×80 it is $(\theta = 49.43, \phi = 84.19)$, for sampled Img_d convolved by Gaussian, resampled to 80×80 and processed by E-D it is $(\theta = 56.55, \phi = 84.84)$.

and more noise appears on Img_d through the sampling process. The E-D network does an excellent job of reducing this unfavorable phenomenon. As in the previously discussed cases, this phenomenon is expected and easily explained, which proves the stability of Algorithm 1. When E-D is applied for Img_d convolved and resampled, the number of cases (estimations) where MAE < 10 increases from 48% to 79% for θ and from 62% to 65% for ϕ, MAE < 5 increases from 24% to 45% for θ and from 47% to 52% for ϕ, MAE < 1 increases from 6% to 9% for θ and from 12% to 16% for ϕ. This is a significant change and demonstrates the high utility of the E-D network used.

5. Conclusion

The proposed algorithm based on the use of PCA for the determination of spherical coordinates of sampled cosmic ray flux distribution proved to be an effective and precise method in the experiment we conducted. The additional use of a deep neural network with an encoder-decoder architecture significantly increases its efficiency in the area of
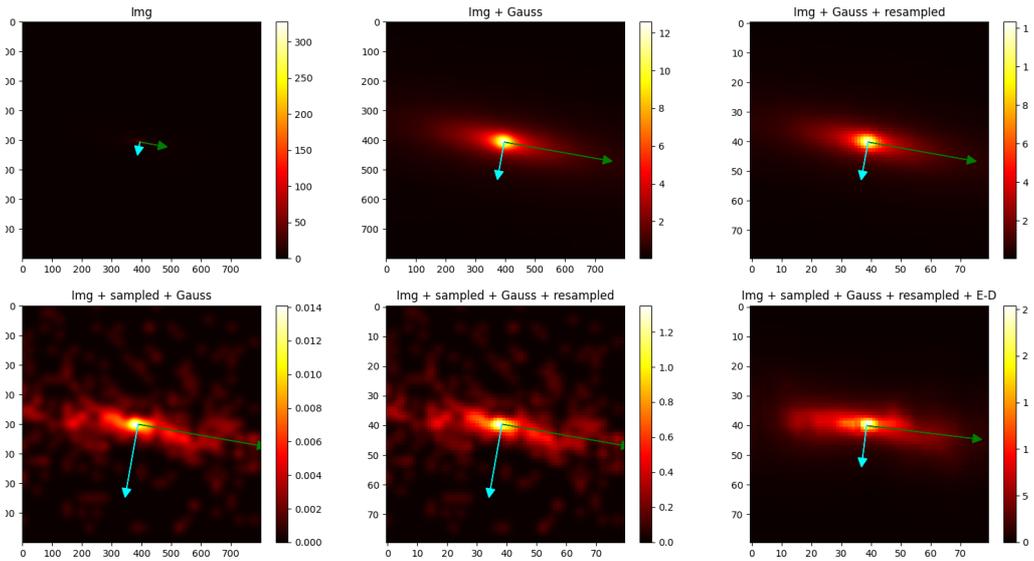
Fig. 7. Examples of results of Algorithm 1 for ($\theta = 76, \phi = 170$). From left to right, from top to bottom (see also the titles of images): estimation of these angles for Img (800×800) equals ($\theta = 75.87, \phi = 169.54$), for Img convolved by Gaussian it is ($\theta = 72.96, \phi = 169.85$), for Img convolved by Gaussian and resampled to 80×80 it is ($\theta = 72.85, \phi = 169.81$), for sampled Img_d convolved by Gaussian (800×800) it is ($\theta = 57.13, \phi = 169.97$), for sampled Img_d convolved by Gaussian and resampled to 80×80 it is ($\theta = 57.04, \phi = 170.01$), for sampled Img_d convolved by Gaussian, resampled to 80×80 and processed by E-D it is ($\theta = 72.00, \phi = 173.06$).

high values of angles (θ, ϕ) making the proposed approach even more effective. Our Algorithm 1, together with the E-D network, is a very important method that will find its application in the research related to physical observations of fundamental astronomical processes. In particular, the introduced scheme can be directly useful in the design of small-scale complex CRE secondary flux detection systems. As we mentioned earlier, to the best of our knowledge, the results presented in this paper are pioneering in the field of small-scale complex CRE secondary flux detection systems, and it is difficult to point out research for the direct comparison. However, based on published research describing the use of a deep encoder-decoder for image denoising and original probabilistic distribution reconstruction [7, 11, 50, 54] we obtained the expected results, that is, the removal of unwanted noise at various frequencies while enhancing the signal with the desired statistical distribution. The effect of this denoising was an increased accuracy in estimating the rotation angles of the particle distribution described by the equation (1).

Several research problems have arisen in the preparation of this study that need to be addressed in future work. These include the effect of the topology of the detector grid

on the efficiency of estimating the angles (θ, ϕ), the estimation of the appropriate grid density on the correctness of the estimate, and the dependence of the estimate on the energy of the particle flux and the offset of the flux center with respect to the detector grid center. These issues will be the subject of future research; nevertheless, it can already be summarized that our proposed method is a very effective approach.

References

[1] M. G. Aartsen, M. Ackermann, J. Adams, J. Aguilar, M. Ahlers, et al. The IceCube Neutrino Observatory: instrumentation and online systems. *Journal of Instrumentation*, 12(03):P03012, 2017. doi:10.1088/1748-0221/12/03/P03012.

[2] M. G. Aartsen, M. Ackermann, J. Adams, J. Aguilar, M. Ahlers, et al. Erratum: The IceCube Neutrino Observatory: instrumentation and online systems. *Journal of Instrumentation*, 19(05):E05001, 2024. doi:10.1088/1748-0221/19/05/E05001.

[3] J. Abraham, P. Abreu, M. Aglietta, C. Aguirre, E. Ahn, et al. Atmospheric effects on extensive air showers observed with the surface detector of the Pierre Auger observatory. *Astroparticle Physics*, 32(2):89–99, 2009. doi:10.1016/j.astropartphys.2009.06.004.

[4] J. Abraham, P. Abreu, M. Aglietta, C. Aguirre, E. Ahn, et al. Erratum to "Atmospheric effects on extensive air showers observed with the surface detector of the Pierre Auger observatory" [Astroparticle Physics 32(2) (2009), 89–99]. *Astroparticle Physics*, 33(1):65–67, 2010. doi:10.1016/j.astropartphys.2009.10.005.

[5] R. Aloisio. Ultra High Energy Cosmic Rays an overview. *Journal of Physics: Conference Series*, 2429(1):012008, 2023. Proc. 12th Cosmic Ray International Seminar (CRIS 2022), 12-16 Sep 2022, Napoli, Italy. doi:10.1088/1742-6596/2429/1/012008.

[6] A. D. Avrorin, A. V. Avrorin, V. M. Aynutdinov, R. Bannash, I. A. Belolaptikov, et al. Baikal-GVD. *EPJ Web of Conferences*, 136:04007, 2017. Proc. 6th Roma International Conference on Astroparticle Physics (RICAP16), 21-24 Jun, Roma, Italy. doi:10.1051/epjconf/201713604007.

[7] K. Bajaj, D. K. Singh, and M. A. Ansari. Autoencoders based deep learner for image denoising. *Procedia Computer Science*, 171:1535–1541, 2020. Proc. 3rd International Conference on Computing and Network Communications (CoCoNet'19), 18-21 Dec, Trivandrum, Kerala, India. doi:10.1016/j.procs.2020.04.164.

[8] O. Bar, Ł. Bibrzycki, M. Niedźwiecki, M. Piekarczyk, K. Rzecki, et al. Zernike moment based classification of cosmic ray candidate hits from CMOS sensors. *Sensors*, 21(22):7718, 2021. doi:10.3390/s21227718.

[9] Ł. Bibrzycki, D. Burakowski, P. Homola, M. Piekarczyk, M. Niedźwiecki, et al. Towards a global cosmic ray sensor network: CREDO detector as the first open-source mobile application enabling detection of penetrating radiation. *Symmetry*, 12(11):1802, 2020. doi:10.3390/sym12111802.

[10] browarsoftware. `cosmic_ray_spherical`. GitHub. https://github.com/browarsoftware/cosmic_ray_spherical.

[11] Y. Choi, S. Park, and S. Kim. Development of point cloud data-denoising technology for earthwork sites using encoder-decoder network. *KSCE Journal of Civil Engineering*, 26(11):4380–4389, 2022. doi:10.1007/s12205-022-0407-8.

[12] M. T. Dova, L. N. Epele, and A. G. Mariazzi. Particle density distributions of inclined air showers. *Nuovo Cimento C*, 24(4–5):745–750, 2001. https://www.sif.it/riviste/sif/ncc/econtents/2001/024/04-05/article/5.

[13] D. Droz, A. Tykhonov, X. Wu, and M. Deliyergiyev. Neural networks for TeV cosmic electrons identification on the DAMPE experiment. In: *Proc. The European Physical Society Conference on High Energy Physics PoS(EPS-HEP2021)*, vol. 398 of *Proceedings of Science*, p. 045. Online – Hamburg, Germany, 26-30 Jul 2021, published in 2022. doi:10.22323/1.398.0045.

[14] V. Dutta, M. Choraś, M. Pawlicki, and R. Kozik. A deep learning ensemble for network anomaly and cyber-attack detection. *Sensors*, 20(16):4583, 2020. doi:10.3390/s20164583.

[15] V. Dutta, M. Pawlicki, R. Kozik, and M. Choraś. Unsupervised network traffic anomaly detection with deep autoencoders. *Logic Journal of the IGPL*, 30(6):912–925, 2022. doi:10.1093/jigpal/jzac002.

[16] J. Glombitza, M. Erdmann, M. Vieweg, and M. Dohmen. Deep learning based air shower reconstruction at the Pierre Auger Observatory. In: *Proc. DPG Spring meeting 2019*, vol. Aachen 2019 issue of *Verhandlungen der Deutschen Physikalischen Gesellschaft*. Aachen, Germany, 25-29 Mar 2019. https://inis.iaea.org/records/dwvnb-hhf07.

[17] K. Greisen. Cosmic ray showers. *Annual Review of Nuclear Science*, 10(1):63–108, 1960. doi:10.1146/annurev.ns.10.120160.000431.

[18] Y. Guo and B. W.-K. Ling. Spherical coordinate-based kernel principal component analysis. *Signal, Image and Video Processing*, 15(3):511–518, 2021. doi:10.1007/s11760-020-01771-8.

[19] T. Hachaj, Ł. Bibrzycki, and M. Piekarczyk. Recognition of cosmic ray images obtained from CMOS sensors used in mobile phones by approximation of uncertain class assignment with deep convolutional neural network. *Sensors*, 21(6):1963, 2021. doi:10.3390/s21061963.

[20] T. Hachaj, Ł. Bibrzycki, and M. Piekarczyk. Fast training data generation for machine learning analysis of cosmic ray showers. *IEEE Access*, 11:7410–7419, 2023. doi:10.1109/ACCESS.2023.3237800.

[21] T. Hachaj and M. Piekarczyk. The practice of detecting potential cosmic rays using CMOS cameras: Hardware and algorithms. *Sensors*, 23(10):4858, 2023. doi:10.3390/s23104858.

[22] T. Hachaj, M. Piekarczyk, and J. Wąs. Searching of potentially anomalous signals in cosmic-ray particle tracks images using rough k-means clustering combined with eigendecomposition-derived embedding. In: *Proc. International Joint Conference on Rough Sets (IJCRS 2023)*, vol. 14481 of *Lecture Notes in Computer Science*, pp. 431–445. Springer, Kraków, Poland, 5-8 Oct 2023. doi:10.1007/978-3-031-50959-9_30.

[23] P. Hasiec, A. Świtoński, H. Josiński, and K. Wojciechowski. Anomaly detection of motion capture data based on the autoencoder approach. In: *Proc. International Conference on Computational Science (ICCS 2023 2023)*, vol. 14074 of *Lecture Notes in Computer Science*, pp. 611–622. Springer, 3-5 Jul 2023. doi:10.1007/978-3-031-36021-3_59.

[24] P. Homola, D. Beznosko, G. Bhatta, Ł. Bibrzycki, M. Borczyńska, et al. Cosmic-ray extremely distributed observatory. *Symmetry*, 12(11):1835, 2020. doi:10.3390/sym12111835.

[25] P. Homola, V. Marchenko, A. Napolitano, R. Damian, R. Guzik, et al. Observation of large scale precursor correlations between cosmic rays and earthquakes with a periodicity similar to the solar cycle. *Journal of Atmospheric and Solar-Terrestrial Physics*, 247:106068, 2023. doi:10.1016/j.jastp.2023.106068.

[26] J. Jaworek-Korjakowska and R. Tadeusiewicz. Determination of border irregularity in dermoscopic color images of pigmented skin lesions. In: *Proc. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 6459–6462. IEEE, Chicago, IL, USA, 26-30 Aug 2014. doi:10.1109/EMBC.2014.6945107.

[27] H. Josiński, D. Kostrzewa, A. Michalczuk, A. Świtoński, and K. Wojciechowski. Feature extraction and HMM-based classification of gait video sequences for the purpose of human identification. In:

A. Nawrat and Z. Kuś, eds., *Vision Based Systemsfor UAV Applications*, vol. 481 of *Studies in Computational Intelligence*, pp. 233–245. Springer, 2013. doi:10.1007/978-3-319-00369-6_15.

[28] O. Kalashev, I. Kharuk, G. Rubtsov, on behalf of the Baikal-GVD Collaboration, et al. Machine learning based background rejection for Baikal-GVD neutrino telescope. *Journal of Physics: Conference Series*, 2438:012099, 2023. Proc. 20th International Workshop on Advanced Computing and Analysis Techniques in Physics Research, 29 Oct – 3 Dec 2021, Virtual and Daejeon, South Korea. doi:10.1088/1742-6596/2438/1/012099.

[29] M. Karbowiak, M. Orzechowski, T. Wibig, Ł. Bibrzycki, P. Kovacs, et al. Small shower array for education purposes-the CREDO-Maze Project. In: *Proc. 37th International Cosmic Ray Conference PoS(ICRC2021)*, vol. 395 of *Proceedings of Science*, p. 219. Online – Berlin, Germany, 21-23 Jul 2021, published in 2022. doi:10.22323/1.395.0219.

[30] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In: *Proc. 3rd Int. Conf. Learning Representations, ICLR 2015*. San Diego, CA, 7-9 May 2015. Accessible in arXiv. doi:10.48550/arXiv.1412.6980.

[31] P. Koundal, R. Abbasi, M. Ackermann, J. Adams, IceCube Collaboration, et al. Composition analysis of cosmic-rays at IceCube Observatory using graph neural networks. In: *Proc. 27th European Cosmic Ray Symposium – PoS(ECRS)*, vol. 423 of *Proceedings of Science*, p. 085. Nijmegen, The Netherlands, 25-29 Jul, 2022, published in 2023. doi:10.22323/1.423.0085.

[32] R. Kumar. Tracking cosmic rays by crayfis (cosmic rays found in smartphones) global detector. In: *Proc. The 34th International Cosmic Ray Conference PoS(ICRC2015)*, vol. 236 of *Proceedings of Science*, p. 1234. The Hague, The Netherlands, 30 Jul – 6 Aug 2015, published in 2016. doi:10.22323/1.236.1234.

[33] R. Nirwan and N. Bertschinger. Rotation invariant householder parameterization for Bayesian PCA. In: K. Chaudhuri and R. Salakhutdinov, eds., *Proc. 36th International Conference on Machine Learning*, vol. 97 of *Proceedings of Machine Learning Research*, pp. 4820–4828. PMLR, Long Beach, CA, USA, 9-15 Jun 2019. https://proceedings.mlr.press/v97/nirwan19a.html.

[34] R. D. Parsons and S. Ohm. Background rejection in atmospheric Cherenkov telescopes using recurrent convolutional neural networks. *The European Physical Journal C*, 80(5):363, 2020. doi:10.1140/epjc/s10052-020-7953-3.

[35] Particle Data Group, R. L. Workman, V. D. Burkert, V. Crede, E. Klempt, et al. Review of particle physics. *Progress of Theoretical and Experimental Physics*, 2022(8):083C01, 2022. doi:10.1093/ptep/ptac097.

[36] Particle Data Group, P. A. Zyla, R. M. Barnett, J. Beringer, O. Dahl, et al. Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2020(8):083C01, 2020. doi:10.1093/ptep/ptaa104.

[37] M. Piekarczyk, O. Bar, Ł. Bibrzycki, M. Niedźwiecki, K. Rzecki, et al. CNN-based classifier as an offline trigger for the CREDO experiment. *Sensors*, 21(14):4804, 2021. doi:10.3390/s21144804.

[38] M. Piekarczyk and T. Hachaj. On the search for potentially anomalous traces of cosmic ray particles in images acquired by CMOS detectors for a continuous stream of emerging observational data. *Sensors*, 24(6):1835, 2024. doi:10.3390/s24061835.

[39] J. S. Pryga, W. Stanek, K. W. Woźniak, P. Homola, K. Almeida Cheminant, et al. Analysis of the capability of detection of extensive air showers by simple scintillator detectors. *Universe*, 8(8):425, 2022. doi:10.3390/universe8080425.

[40] D. B. Reusch, R. B. Alley, and B. C. Hewitson. Relative performance of self-organizing maps and principal component analysis in pattern extraction from synthetic climatological data. *Polar Geography*, 29(3):188–212, 2005. doi:10.1080/789610199.

[41] K. Sargsyan, J. Wright, and C. Lim. GeoPCA: a new tool for multivariate analysis of dihedral angles based on principal component geodesics. *Nucleic Acids Research*, 40(3):e25–e25, 2012. doi:10.1093/nar/gkr1069.

[42] K. Sargsyan, J. Wright, and C. Lim. GeoPCA: a new tool for multivariate analysis of dihedral angles based on principal component geodesics. *Nucleic Acids Research*, 43(21):10571–10572, 2015. (This is a correction to [41]). doi:10.1093/nar/gkv1000.

[43] M. Savić, A. Dragić, D. Maletić, N. Veselinović, R. Banjanac, et al. A novel method for atmospheric correction of cosmic-ray data based on principal component analysis. *Astroparticle Physics*, 109:1–11, 2019. doi:10.1016/j.astropartphys.2019.01.006.

[44] J. Stasielak, P. Malecki, D. Naumov, V. Allakhverdian, on behalf of the Baikal-GVD Collaboration, et al. High-energy neutrino astronomy—Baikal-GVD neutrino telescope in Lake Baikal. *Symmetry*, 13(3):377, 2021. doi:10.3390/sym13030377.

[45] Z. Szadkowski and K. Pytel. Trigger based on a fuzzy logic for a detection of very inclined cosmic rays in the surface detector of the Pierre Auger Observatory. In: *Proc. 2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 1–5. IEEE, Auckland, New Zealand, 20-23 May 2019. doi:10.1109/I2MTC.2019.8827075.

[46] R. Tadeusiewicz, R. Chaki, and N. Chaki. *Exploring neural networks with C#*. CRC Press, Boca Raton, 2015. doi:10.1201/b17332.

[47] J. Takalo. Extracting hale cycle related components from cosmic-ray data using principal component analysis. *Solar Physics*, 297(9):113, 2022. doi:10.1007/s11207-022-02048-8.

[48] The Pierre Auger Collaboration. The Pierre Auger Cosmic Ray Observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 798:172–213, 2015. doi:10.1016/j.nima.2015.06.058.

[49] The Pierre Auger Collaboration, A. Aab, P. Abreu, M. Aglietta, J. Albury, et al. Extraction of the muon signals recorded with the surface detector of the Pierre Auger Observatory using recurrent neural networks. *Journal of Instrumentation*, 16(07):P07016, 2021. doi:10.1088/1748-0221/16/07/P07016.

[50] C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, et al. Deep learning on image denoising: An overview. *Neural Networks*, 131:251–275, 2020. doi:10.1016/j.neunet.2020.07.025.

[51] J. Vandenbroucke, S. BenZvi, S. Bravo, K. Jensen, P. Karn, et al. Measurement of cosmic-ray muons with the distributed electronic cosmic-ray observatory, a network of smartphones. *Journal of Instrumentation*, 11(04):P04019, 2016. doi:10.1088/1748-0221/11/04/p04019.

[52] T. Wibig. Small shower CORSIKA simulations. *Chinese Physics C*, 45(8):085001, 2021. doi:10.1088/1674-1137/ac0099.

[53] M. Yu, T. B. Anderson, Y. Chen, S. Coutu, T. LaBree, et al. Machine learning applications on event reconstruction and identification for ISS-CREAM. In: *Proc. 37th International Cosmic Ray Conference PoS(ICRC2021)*, vol. 395, p. 061. Proceedings of Science, Online – Berlin, Germany, 21-23 Jul 2021, published in 2022. doi:10.22323/1.395.0061.

[54] D. Zheng, S. H. Tan, X. Zhang, Z. Shi, K. Ma, et al. An unsupervised deep learning approach for real-world image denoising. In: *Proc. 8th International Conference on Learning Representations (ICLR)*. Virtual, 26 Apr – 1 May 2020. Published in OpenReview. https://openreview.net/forum?id=tIjRAiFmU3y.