Machine Learning aided noise filtration and signal classification for CREDO experiment

Łukasz Bibrzycki [1], Olaf Bar [1], Marcin Piekarczyk [1], Krzysztof Rzecki [2], Michał Niedźwiecki [3], Sławomir Stuglik [4], Piotr Homola [4]
Institute of Computer Science, Pedagogical University of Krakow, Kraków, Poland [1], AGH University of Science and Technology, Kraków, Poland [2],
Department of Computer Science, Cracow University of Technology, Kraków, Poland [3], Institute of Nuclear Physics PAN, Kraków, Poland [4]

What is CREDO?

- The **C**osmic **R**ay **E**xtremely **D**istributed **O**bservatory (CREDO) [1] experiment is the large-scale study of various radiation forms that reach the Earth from space, collectively known as cosmic rays.
- It utilizes the CREDO Detector mobile application [4].
- The project follows the citizen science philosophy and engages various groups of people, mostly non-scientists.
- Installing the CREDO Detector App turns their smarthones into mobile observatories.
- CREDO user base:

	number		number
Users	$1.4 \cdot 10^4$	Candidate detections	$1.0 \cdot 10^7$
Devices	$1.5 \cdot 10^4$	Operation time (days)	$3.9 \cdot 10^5$ ($\sim$1050 years)

Introduction and motivation

- The wealth of data collected by the CREDO infrastructure greatly surpasses the capabilities of manual analysis. So, efficient means of rejecting the non-cosmic-ray noise and identification of signals attributable to extensive air showers are necessary.
- To address these problems we use:
  - ▷ Convolutional Neural Network-based trigger for artefact rejection and
  - ▷ Statistical Classifiers for signal morphological classification.
- Both approaches are based on supervised learning, so we need to provide a representative subset of the CREDO dataset for training and validation.

Detections - Signals (spots, tracks, worms)

- Therefore over 2300 images were chosen and manually labeled by 5 annotators. The images were split into spots, tracks, worms (collectively named signals) and artefacts classes.
- Types of observed signals are shown in Fig. 1:
  - ▷ spots (1st row),
  - ▷ tracks (2nd row) and
  - ▷ worms (3rd row).

Fig. 1: Examples of signal types observed by CREDO detectors.

Introduction and motivation (continued)

Detections - Artefacts

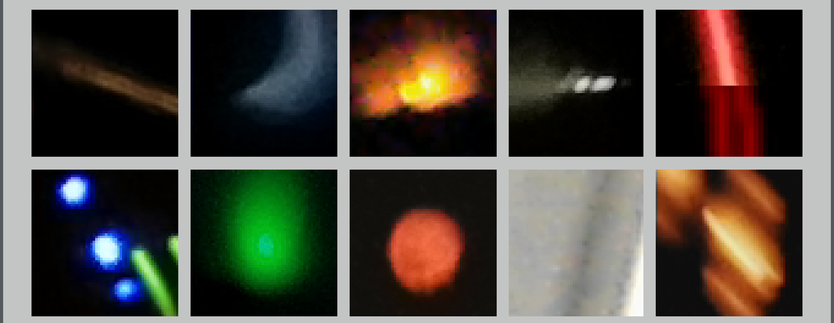


Fig. 2: An example of various bad detections - artefacts. They are usually created by incorrectly covering the camera's CMOS array.

Artefact rejection - CNN based trigger

- Artefact rejection is preceded by three preprocessing steps:
  1. Grayscale conversion by summing up the color channels.
  2. Noise reduction by applying the adaptive threshold which depends on the average brightness (pixels below the threshold are cut off).
  3. Wavelet transform.
- Preprocessing steps are illustrated in Fig. 3.

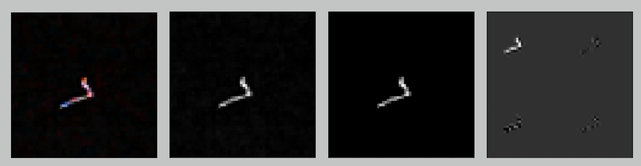


Fig. 3: Example of the worm-type image. Left to right: original image (color), grayscale accumulation, adaptive thresholding and Daubechies wavelet transformation.

- Preprocessed images are fed to the CNN based classifier for training and evaluation. The trained classifier works as a signal trigger.
- Evaluation results in terms of signal and artefact recognition rate:

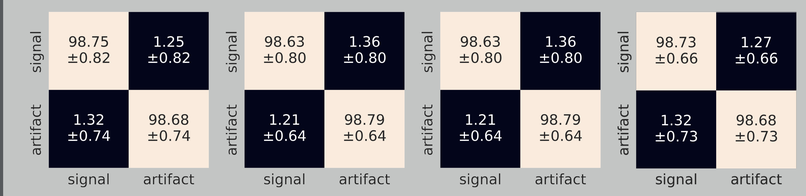
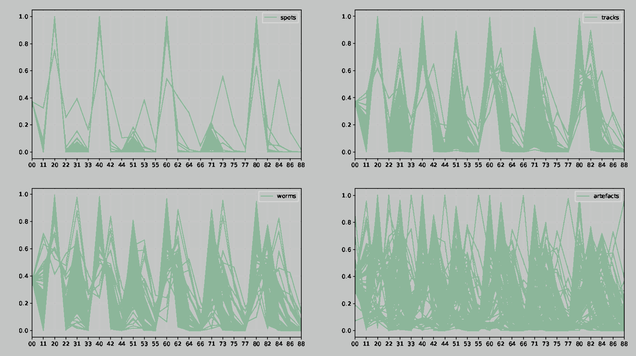


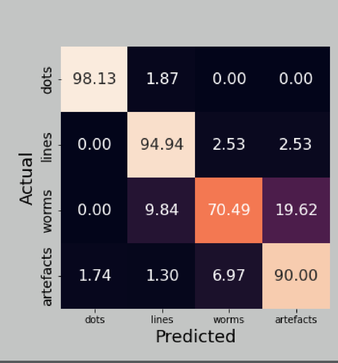
Fig. 4: Confusion matrices for four configurations of input tensors (Raw RGB, D2, D20, D2:D20). The horizontal and vertical dimensions refer to predicted and judged labels, respectively.

Morphological classification of Signals

- Zernike moments of adaptively thresholded input images are used as features for the 4-class statistical classifiers.
- Plots show that the Zernike moments spectra of spots and artefacts are very different from the spectra of tracks and worms.
- On the other hand tracks and worms have similar moments spectra.



- Several classifiers were tested of which the best for shape recognition was: ν-Support Vector Classifier, which e.g.
  - ▷ out of 79 track samples correctly recognised 74 ($\mathbf{94.94\%}$),
  - ▷ out of 107 spot samples correctly recognised 105 ($\mathbf{98.13\%}$).



Recognition rates

- Signal vs. Artefact: 99%, 99%,
- Spots vs. Tracks vs. Worms vs. Artefacts: 98%, 95%, 70%, 90%,

Bibliography

[1] P. Homola et al., Cosmic-Ray Extremely Distributed Observatory, Symmetry 2020,12, **1835**. doi:10.3390/sym12111835.41210.

[2] M. Piekarczyk et al., CNN based classifier as an offline trigger for the CREDO experiment. (in review).

[3] O. Bar et al. Zernike moment based classification of radiationsignals in CMOS sensors. (in preparation).

[4] Ł. Bibrzycki et al, Towards A Global Cosmic Ray Sensor Network: CREDO Detector as the First Open-Source Mobile Application Enabling Detection of Penetrating Radiation, Symmetry 2020, 12, 1802.