Insitute for Nuclear Research of Russian Academy of Sciences

Machine learning techniques applications for the ENDA experiment data analysis

on behalf of ENDA collaboration

3rd ISCRA, 8-10 June 2021

Oleg Shchegolev shchegolev@inr.ac.ru

Layout

1. Approaches for data analysis of ENDA
2. Motivation for machine learning (ML)
3. Monte-Carlo simulations
4. Convolutional neural networks (CNN)
5. Comparison of target metrics on simulation data
6. Conclusions

Approaches for data analysis of ENDA: 1st way

1. Maximum likelyhood estimation of shower direction based on times of electronic component recording
2. Maximum likelyhood estimation of core position, shower age and size using NKG function
3. Recover primary energy using shower size and fit of simulation data based on some model

Approaches for data analysis of ENDA: 2st way

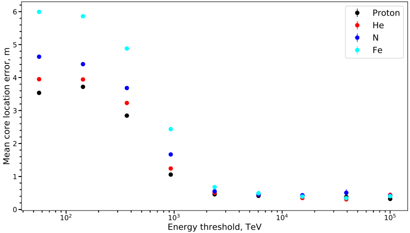
1. Maximum likelyhood estimation of shower direction based ontimes of electronic component recording

2. Maximum likelyhood estimation of core position, shower age and size using NKG function

3. Recover primary energy and atomic number using multivariate analysis on (shower size, shower age, zenith angle, number of neutrons)

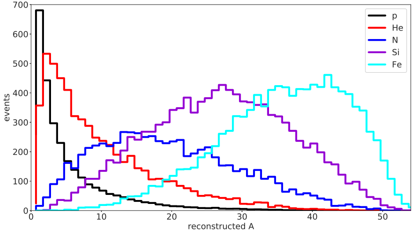
   The last step could be done machine learning techniques as linear regression model (improved with polynomial features), random forest, boosted decision trees or fully connected neural networks.

Approaches for data analysis of ENDA: 2st way



Mass composition

Core location



Approaches for data analysis of ENDA: 3rd way, full deep learning approach

Using deep convolutional neural network to directly reconstruct all the parameters: core location, zenith angle, primary energy and mass without any intermediate steps from measurements to target

Motivation for machine learning (ML)

1. Some experiments have published ML approaches to their data analysis about 20 years ago but only last years it is a mainstream and usually gives state of the art precision.
2. ML techniques are now well developed and studied and could be used almost "from the box"
3. Using ML techniques we can avoid some unreasonable things like NKG function or linear fit

Monte-Carlo simulations

1. Full Monte-Carlo simulation was performed for data analysis techniques validation and improvement
2. Events were simulated in energy range of 30 TeV - 100 PeV with differential slope of -2.7
3. Extensive air showers (EAS) were simulated using CORSIKA7.74 and detector response was simulated using GEANT4.10.6
4. Primary protons, helium, nitrogen, silicon and iron were simulated
5. Over 400K simulated events were "recorded" by array

Some examples of ML/DL usage in EAS experiments

COSMIC RAY SPECTRUM AND COMPOSITION FROM

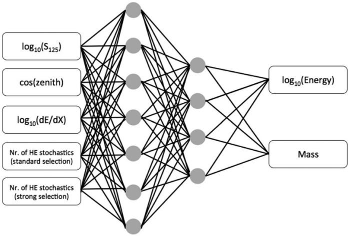


FIG. 10. The neural network architecture of the best performing neural network. This network maps five input variables onto two output variables using two hidden layers with, respectively, seven and four neurons using a tanh activation function. It is therefore called a 5-7-4-2 network.

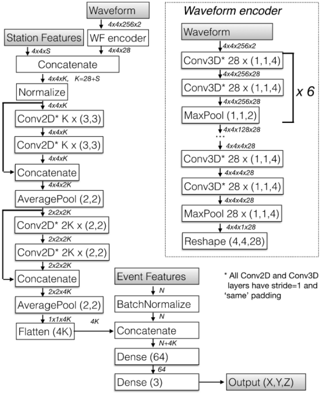


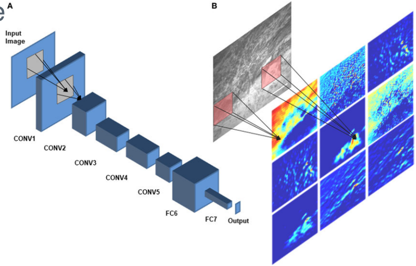
Figure 2. Neural network architecture.

Figure 1: IceTop FCNN

Figure 2: TA CNN

Convolutional Neural Networks (CNN)

Convolutional neutral networks are essential approach for image processing imitating work of eyes and brain.



CNN show robustness to image rotations, transitions and scaling

In EAS ground surface experiments we deal with "images" of showers

ENDA EAS data could be represented as a three layer image: array response for electromagnetic component, number of recorded neutrons for each detector, detectors relative delays (times).

Convolutional Neural Networks (CNN)

Similar approach was presented in *Erdmann M., Glombitza J., Walz D. A deep learning-based reconstruction of cosmic ray-induced air showers //Astroparticle Physics. – 2018. – . 97. – . 46-53.*

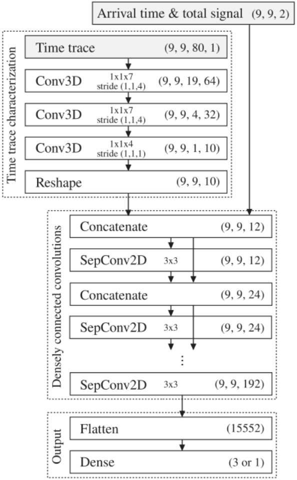


Fig. 4. Network architecture: A series of 3D convolutions characterizes the time trace of each detector by 10 features which are then stacked with the arrival time and the total signal. The main part consists of five separable 2D convolution layers whose in- and output are concatenated for the next layer. This creates a shortcut of the original input and the output of each convolution layer to all subsequent layers. A final dense layer outputs the shower property.

CNN for ENDA-64: current version

1. Input consisting of 64 detector responses for electrons and neutrons (64x2)
2. 3 Convolutional layers (2x8), (8x16), (16x32)
3. MaxPooling layer
4. 3 Fully connected layers (32x120), (120x80), (80x2)
5. Output consisting of x, y (core position)

Core location error

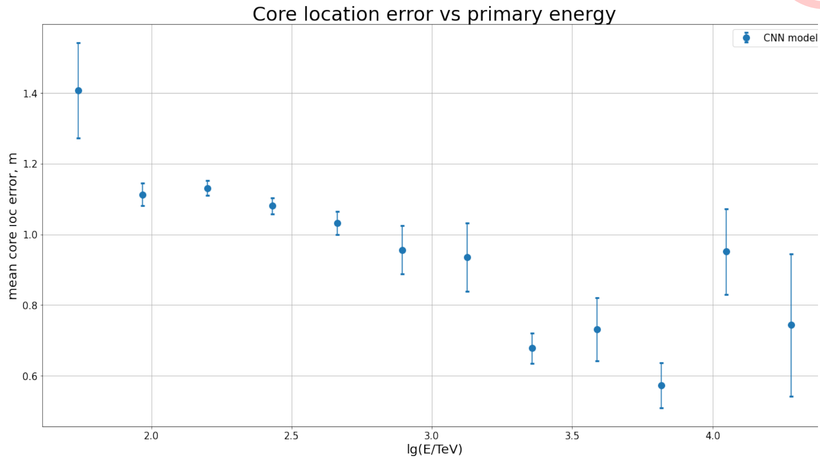


Figure 3: core location error

Conclusions

1. Using of machine learning algorithms allows to improve reconstruction of EAS parameters

2. Simple convolutional neural network was trained for core location and showed better result than NKG function fitting way

3. Work on training more complex CNN for energy and mass reconstruction is continued

Thank you for attention!