Neural Network-Based Shower Core Reconstruction in the HiSCORE Experiment using Autoencoder-Derived Essential Features

Yulia Dubenskaya[1], Stanislav Polyakov[1], Alexander Kryukov[1], Andrey Demichev[1], Pavel Volchugov[1], Elizaveta Gres[2], Dmitry Zhurov[3], Evgeny Postnikov[1], Alexander Razumov[1]

[1] *Skobeltsyn Institute of Nuclear Physics, Moscow State University*
[2] *Research Institute of Applied Physics, Irkutsk State University*
[3] *Irkutsk State University*

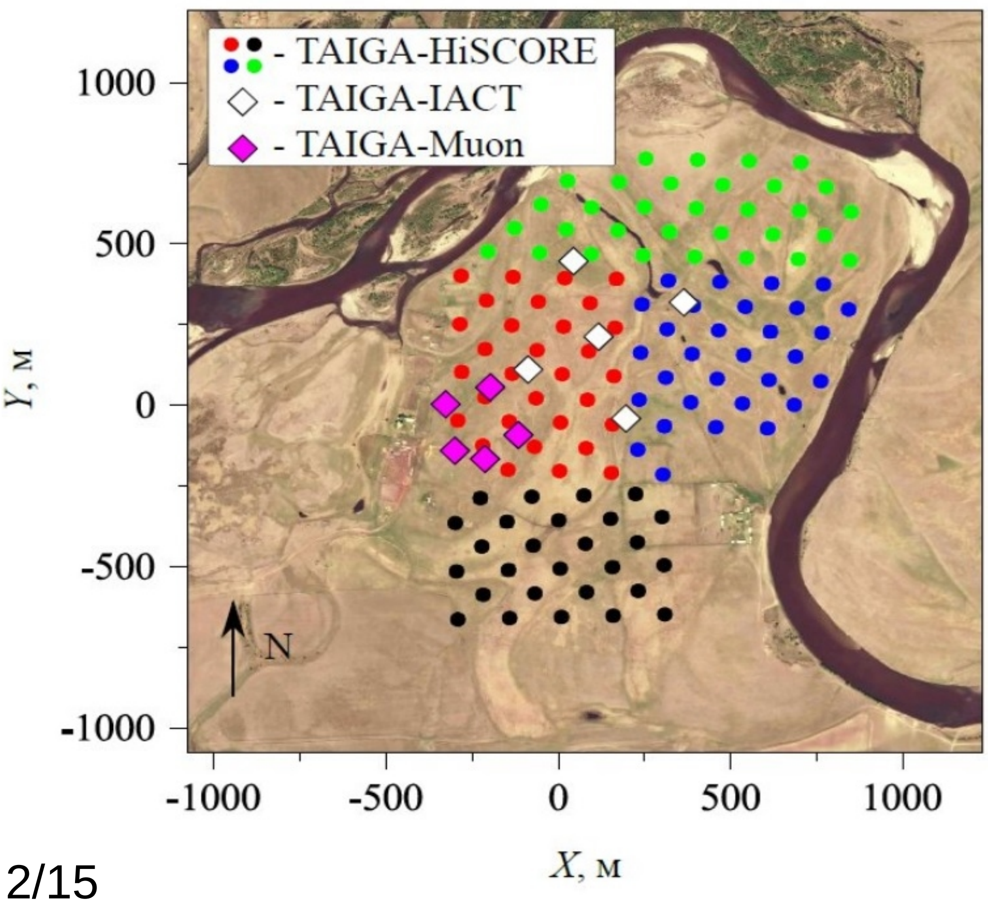
The research was supported by the Russian Science Foundation (grant no. 24-11-00136, https://rscf.ru/project/24-11-00136/), using equipment provided under the Lomonosov Moscow State University Development Program and data from the unique TAIGA scientific facility

07.2026

The TAIGA project

TAIGA stands for the Tunka Advanced Instrument for cosmic ray physics and Gamma-ray Astronomy

It is a large-scale, hybrid astroparticle detector complex located in the Tunka Valley in Siberia, Russia



TAIGA combines multiple detector types:

- Wide-Angle Cherenkov Arrays (HiSCORE and Tunka-133)
- Imaging Air Cherenkov Telescopes (IACTs)
- Scintillation and Muon Detectors (Tunka-Grande and TAIGA-muon)

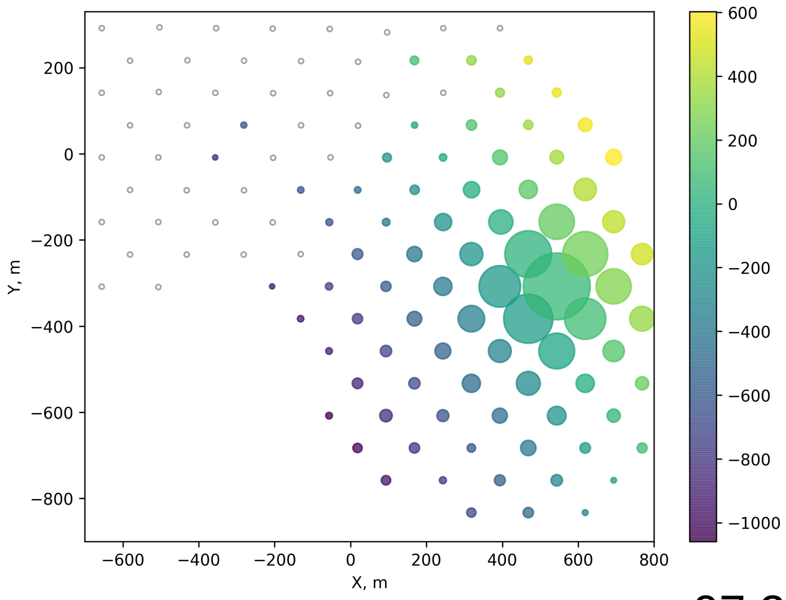
The HiSCORE array

HiSCORE is a large-scale (1 km²) array of wide-angle non-imaging Cherenkov detectors, spaced ~100 m apart, designed to detect and study extensive air showers (EAS) from primary cosmic rays and gamma rays

HiSCORE enables the reconstruction of key EAS parameters: core position, arrival direction, and energy of the primary particle

Each HiSCORE detector records:

- Light amplitudes (Cherenkov photon density)
- Signal arrival times



For the neural network approach, the HiSCORE array data are represented as quasi-images, where each station corresponds to a pixel with amplitude and timing information

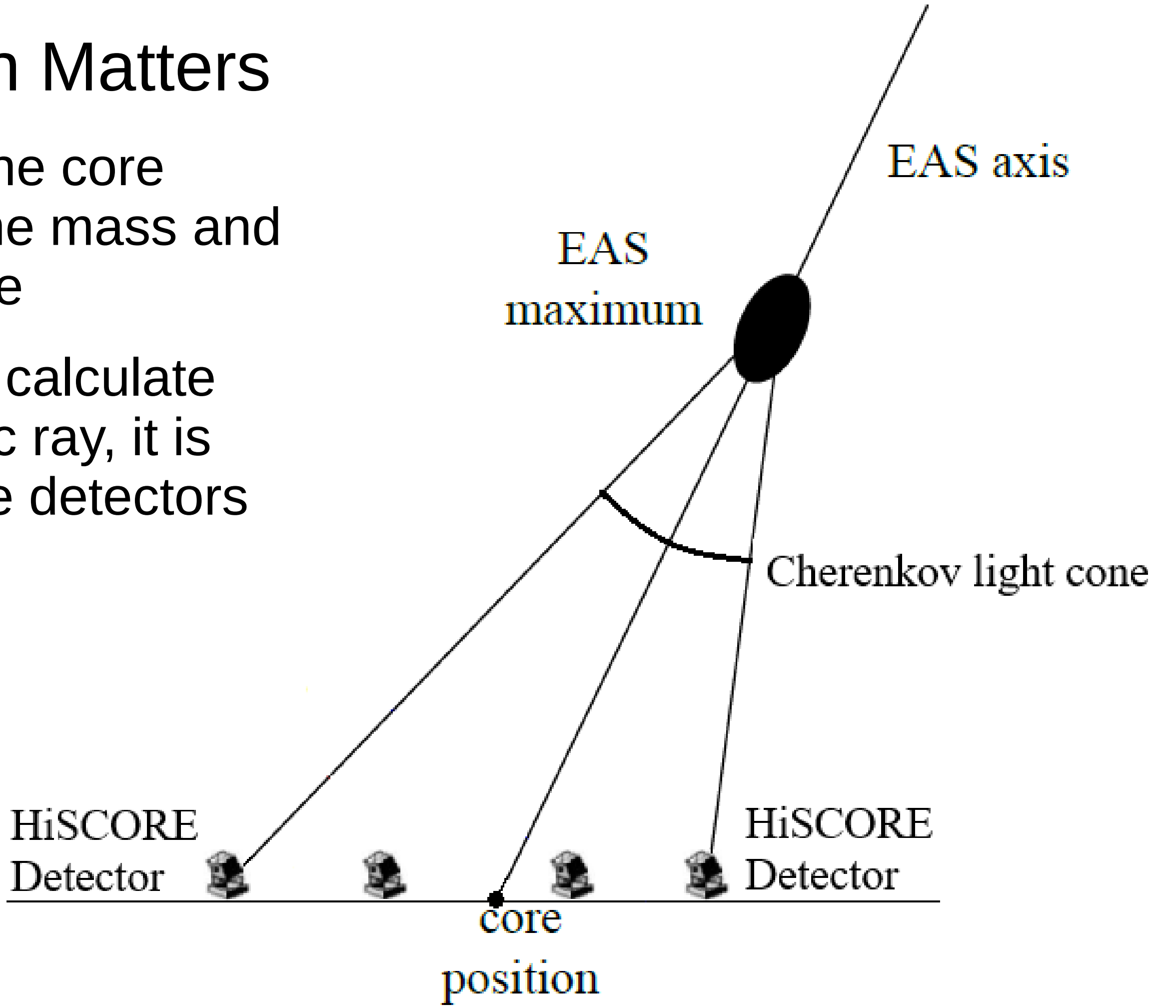
EAS core position

Core position is the exact point on the ground where the central axis of an EAS hits the Earth

Why the EAS Core Position Matters

Primary Particle Identification: The core area carries vital information about the mass and type of the original cosmic-ray particle

Energy Calibration: To accurately calculate the total energy of the primary cosmic ray, it is important to know how far the particle detectors are from the shower core



Towards the multimodal analysis

Multimodal analysis – joint processing of data from different detector types

The global goal of our project is to use ML-techniques to prepare and process data from various TAIGA detectors so that they can be easily combined, as well as to further jointly analyze these data

Our main idea is to use neural networks that reconstruct different EAS parameters using essential features (i.e., a compact representation of the data) as input. As essential features we propose to use a latent space of an autoencoder (AE) trained on HiSCORE data

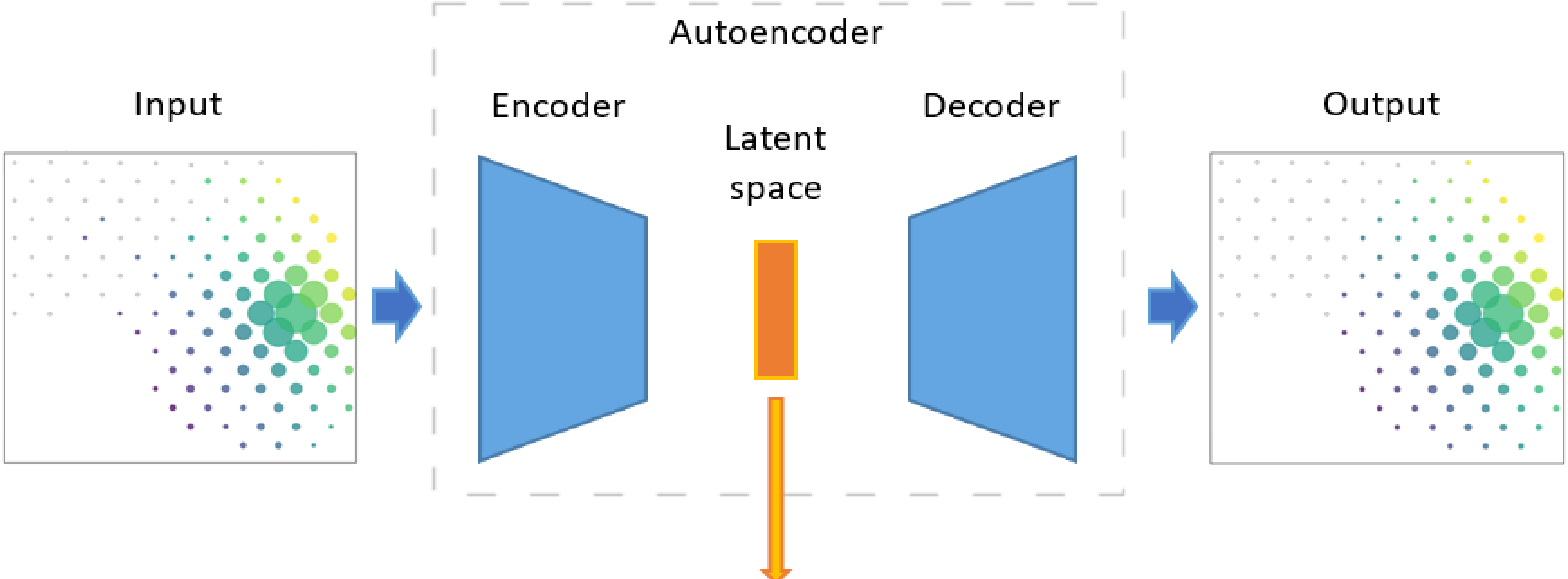
The essential features are not related to the dimensions and format of the data

The aim of this work is to test several network architectures for reconstructing the core position and select the best performing one

The same approach can be applied to IACT data, enabling seamless data fusion across detector systems

Extracting essential features

The AE was trained to compress/restore HiSCORE data



Essential features

used as input for the subsequent neural networks

The AE is convolutional without fully connected layers. The number of trainable encoder parameters is approximately 3 million

In our previous work, we investigated how the accuracy depends on the dimensionality of the latent space AE (N), and now according to these results we choose N=12

The core position is predicted using a subsequent neural network (NN)

Neural network (NN) to determine the core position

Input (*we compare two variants*):

- 12 latent parameters
- 12 latent parameters + 2 angle values (zenith and azimuth angles)

In our previous work, we learned how to determine zenith and azimuth angles with neural networks with high accuracy, so we can use them to improve the reconstruction accuracy of other EAS parameters

Output:

2 coordinates of the core position: X and Y

Architecture:

We compare different NN architectures:

- fully connected networks
- networks with a combination of fully connected and convolutional layers

Activation functions: leakyReLU

Loss function: MSE

Architectures compared

We compared the following fully-connected architectures:

1) Input (**K**) – 128-256-512-256-128 – Output (2)
2) Input (**K**) – 2048-512-128 – Output (2)
3) Input (**K**) – 2048-512 – Output (2)

with the following combined architectures:

4)
Input (**K**)
Dense(2048)
Reshape((8, 8, 32))
Conv2D(64, (2, 2), strides=(2, 2))
Conv2D(128, (2, 2), strides=(2, 2))
Flatten
Dense(**M**)
Output (2)

5)
Input (**K**)
Dense(2048)
Reshape((8, 8, 32))
Conv2D(64, (2, 2), strides=(2, 2))
Conv2D(128, (2, 2), strides=(2, 2))
Flatten
Dense(512)
Dense(128)
Output (2)

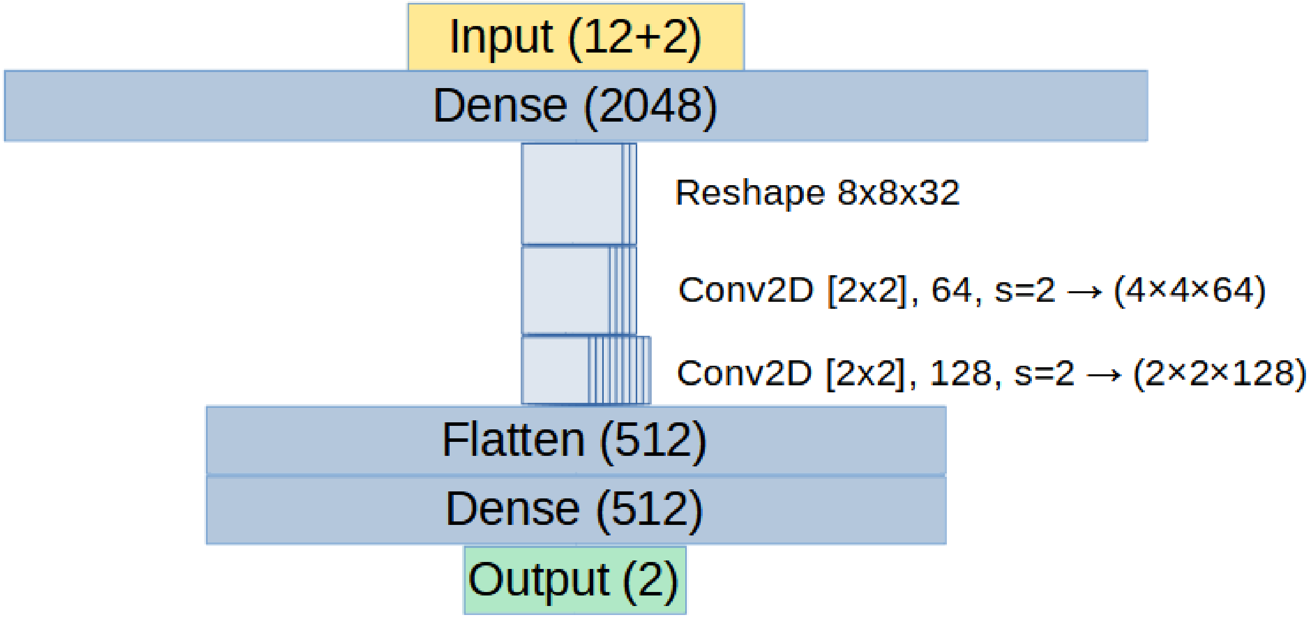
Here:
K = 12 *or* (12+2) input parameters
M = 128, 256, 512 *or* 1024 neurons

NN training results for all the architectures

Input size	Network architecture	Number of NN parameters	Core position (X, Y), meters		
			Mean	Std	MAE
12	Dense 128-256-512-256-128	335,874	(0.44, 0.2)	(28.4, 33)	(21.4, 25.1)
12+2	Dense 128-256-512-256-128	336,130	(0.15, 0.61)	(26.9, 31.1)	(20.1, 23.2)
12	Dense 2048-512-128	1,152,386	(-0.39, 0.26)	(24.7, 29.2)	(18.2, 21.3)
12+2	Dense 2048-512-128	1,156,482	(0.67, 0.65)	(23.8, 27.9)	(17.4, 20.4)
12	Dense 2048-512	1,086,978	(0.13, 0.26)	(21.9, 24.4)	(16.3, 18.1)
12+2	Dense 2048-512	1,091,074	(-0.24, -0.27)	(21.1, 23.7)	(15.7, 17.4)
12	Dense 2048 + Conv2D + Dense 128	143,170	(0.72, -0.03)	(23.8, 27.2)	(17.4, 19.9)
12+2	Dense 2048 + Conv2D + Dense 128	147,266	(0.2, 0.03)	(22.3, 25.9)	(16.2, 18.7)
12	Dense 2048 + Conv2D + Dense 256	209,602	(-0.21, 0.32)	(19.4, 22.1)	(14.1, 16.2)
12+2	Dense 2048 + Conv2D + Dense 256	213,698	(0.6, -0.18)	(18.1, 20.9)	(13.4, 15.4)
12	Dense 2048 + Conv2D + Dense 512	342,466	(0.24, 0.62)	(16.6, 17.8)	(12, 13.1)
12+2	Dense 2048 + Conv2D + Dense 512	346,562	(-0.54, 0.48)	(15.8, 17)	(11.2, 12.3)
12	Dense 2048 + Conv2D + Dense 1024	608,194	(0.88, -1.16)	(15.9, 17.1)	(11.3, 12.4)
12+2	Dense 2048 + Conv2D + Dense 1024	612,290	(-1.28, -0.6)	(15.6, 16.7)	(11.1, 12)
12	Dense 2048 + Conv2D + Dense 512-128	407,874	(0.4, -0.51)	(22.7, 27.2)	(16.7, 19.9)
12+2	Dense 2048 + Conv2D + Dense 512-128	411,970	(-0.03, -0.88)	(21.8, 26.1)	(15.9, 18.9)

Selected architecture of the NN to determine the core position

The network architecture that showed the best results among all the options tested is as follows:

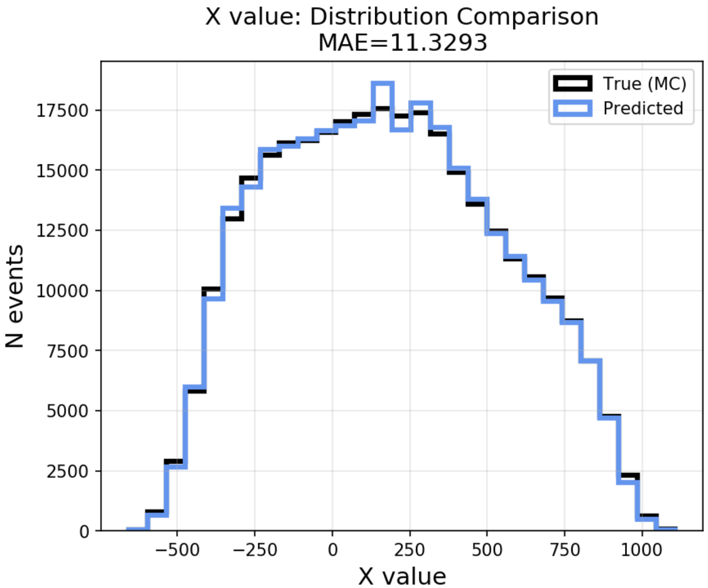


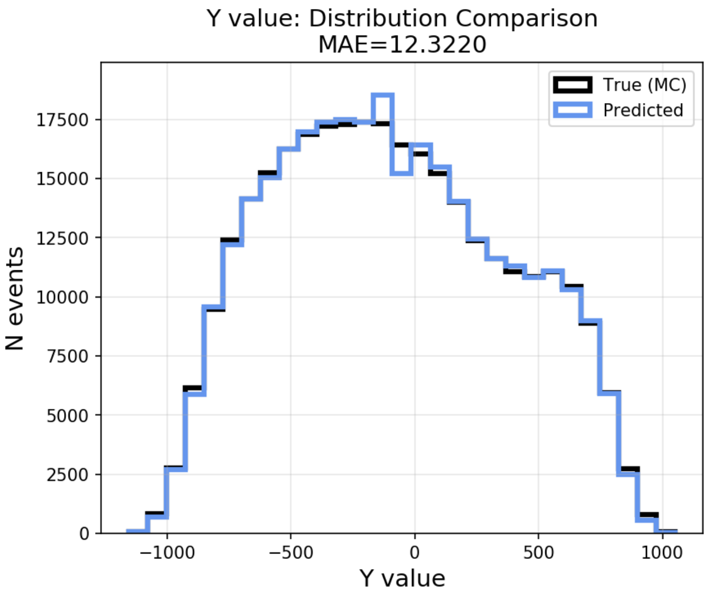
We train this network on 930,000 sets of latent features along with the zenith and azimuth angle values and the (X,Y) core position coordinate values of the corresponding Monte Carlo HiSCORE events

Number of epochs: 60; batch size: 256

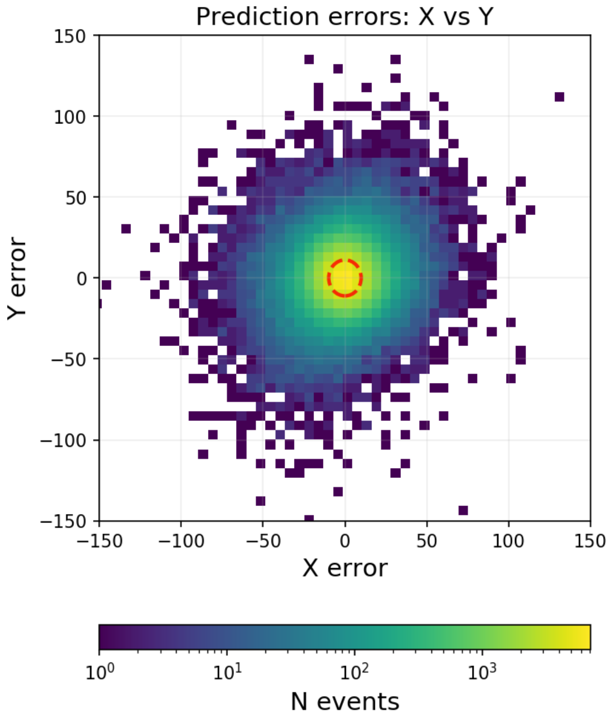
Both the test and validation sets contain 310,000 events

Results for the core position reconstruction

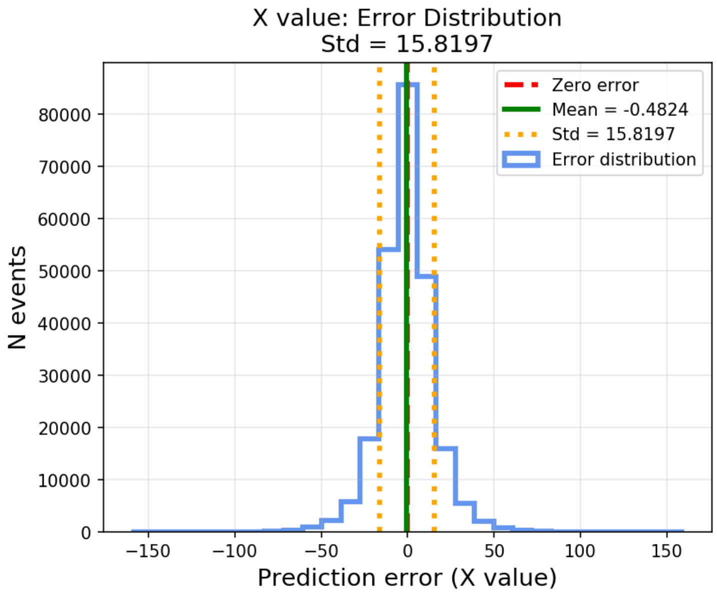


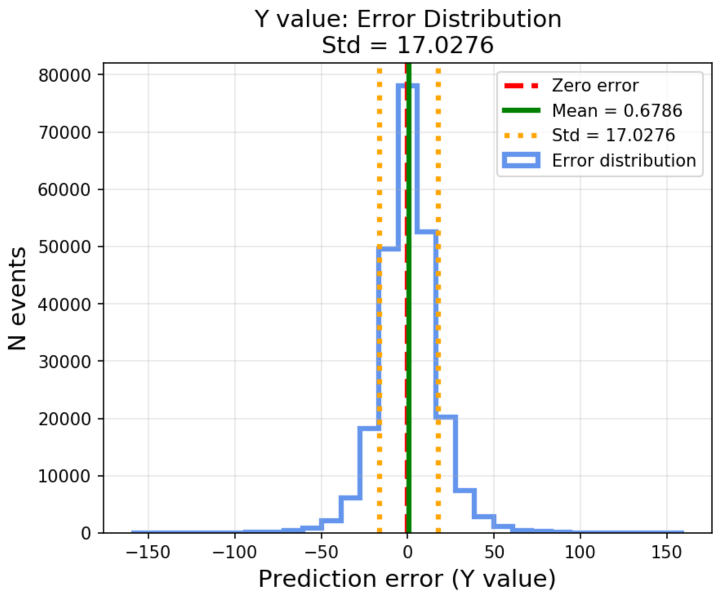


$$X \text{ error} = X_{\text{predicted}} - X_{\text{true}}$$
$$Y \text{ error} = Y_{\text{predicted}} - Y_{\text{true}}$$



The red dashed ellipse shows the median errors: (X = 9.8 m, Y = 11.1 m)





Full test dataset vs reduced test dataset

We investigated the effect of EAS core position on the reconstruction accuracy due to the following reasons:

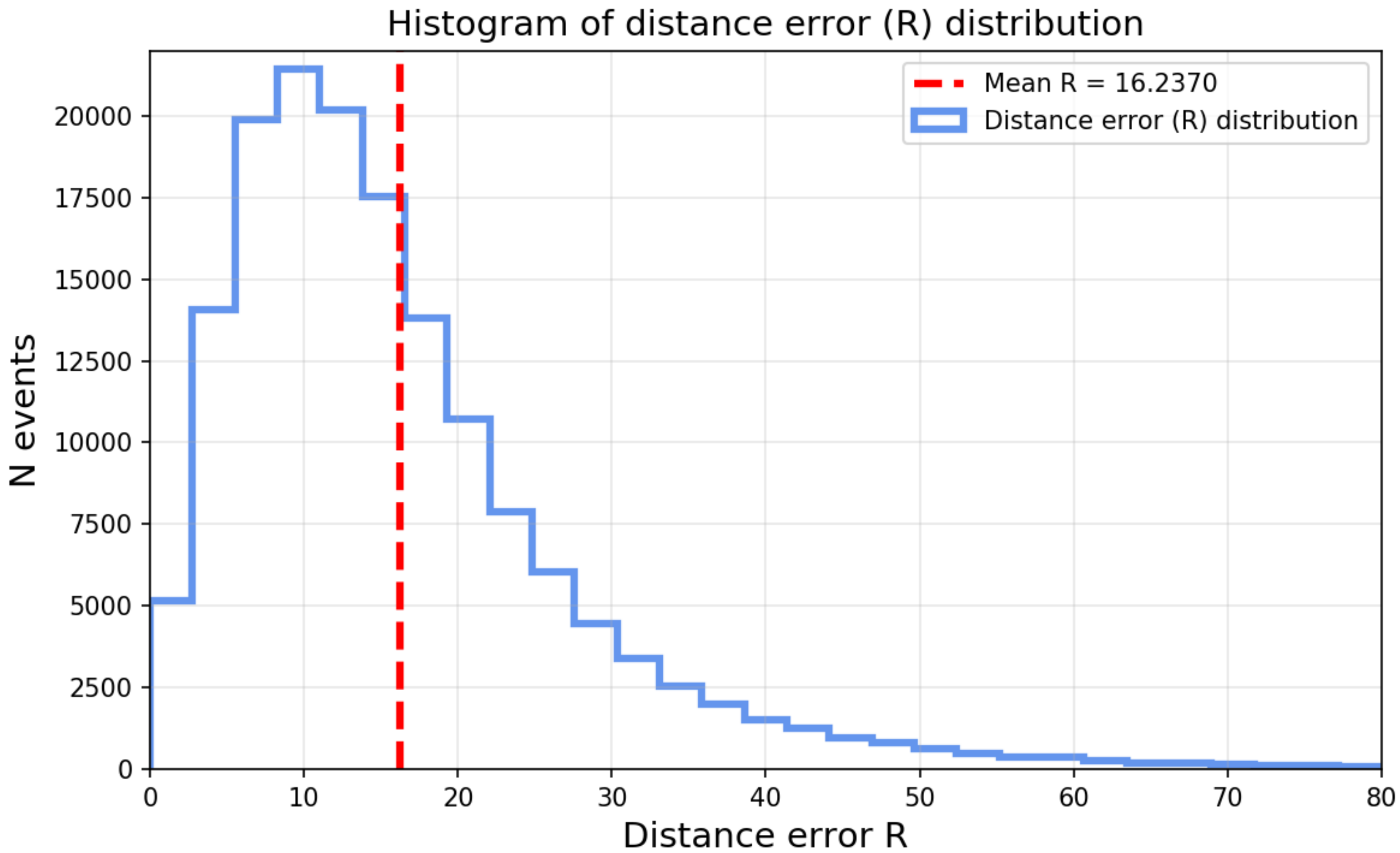
- The parameters of boundary events falling at the edge of the detector array are typically reconstructed with lower accuracy
- The HiSCORE array is not rectangular. This may further reduce the accuracy of boundary events

For testing, we limited the test dataset to only those events falling within the inner region of the detector array (within 500 meters of the average X and Y coordinates). This reduced the test set size by a factor of 2, but increased the accuracy of EAS core position reconstruction

	Number of events	MAE, m		X error, m		Y error, m	
		X	Y	Mean	Std	Mean	Std
The whole test set	100%	11.2	12.3	-0.54	15.8	0.48	17
Reduced set ($X_{mean} \pm 500$, $Y_{mean} \pm 500$)	50%	10	10.6	-0.65	13.8	0.8	14.6

Distance error histogram

R (distance error) = distance between the ground truth core position and the core position predicted by the neural network (in meters)



Comparison of Results with Other Methods

In the conventional method, the accuracy of reconstructing EAS core position ranges from **6 to 15 meters**, depending on the primary particle energy, shower zenith angle, the number of stations activated, and other factors

We also tested the accuracy of EAS core position reconstruction using a neural network directly from HiSCORE quasi-images, without using an autoencoder. The standard deviation of the reconstruction error for this method is **12-13 meters**

The standard deviation of the core position reconstruction error for the proposed method is **15-17 meters** for the whole test set and **13-14 meters** for the reduced test set (without boundary events), which is comparable with the results obtained by other methods

Conclusion

The architecture of the neural network for reconstructing the EAS core position has been selected

We demonstrate that the latent space of a properly trained autoencoder can be used to reconstruct the core position of the EAS, and that adding additional pre-computed angle parameters to the input further improves the reconstruction accuracy

The standard deviation of the core position reconstruction error for the proposed method (with pre-computed angle parameters added to the input) is 15-17 meters for the whole test set and 13-14 meters for the reduced test set (without boundary events), which is comparable with the results obtained by other methods

We expect that these results will be useful in multimodal analysis (joint processing of data from different detector types), enabling seamless integration of latent parameters and improving overall reconstruction accuracy

Thank you for attention!