Application of Normalizing Flows for Detecting Rare Events in Gamma-Ray Astronomy

A. Kryukov(1), A. Razumov(1), A. Demichev(1), Yu. Dubenskaya(1), S. Polyakov(2), P. Volchugov(1), E. Gres(1,3); D. Zhurov(1,3), E. Postnikov(1)

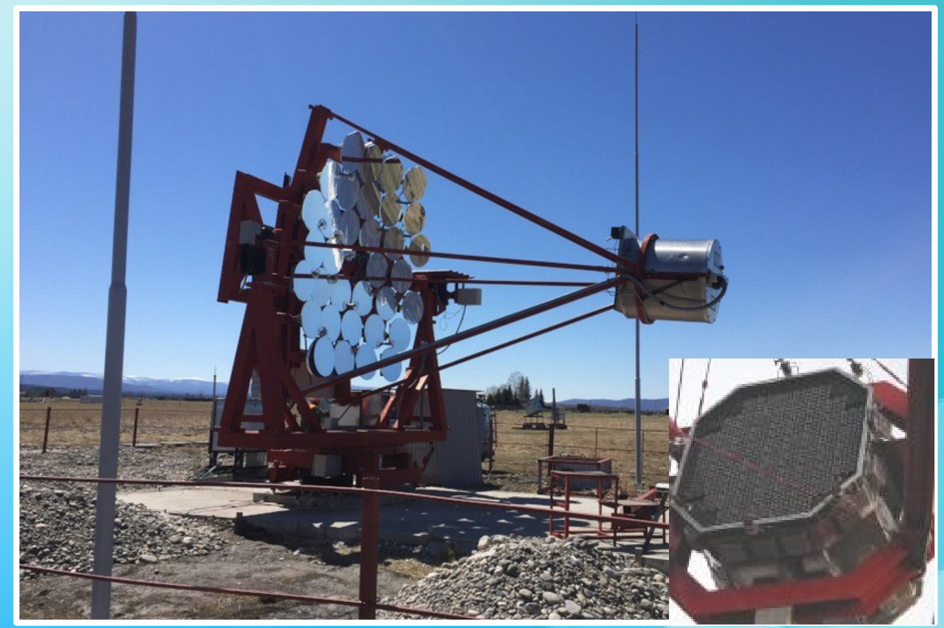
(1) SINP MSU, (2) IIAP NAS RA, (3) IPA IGU

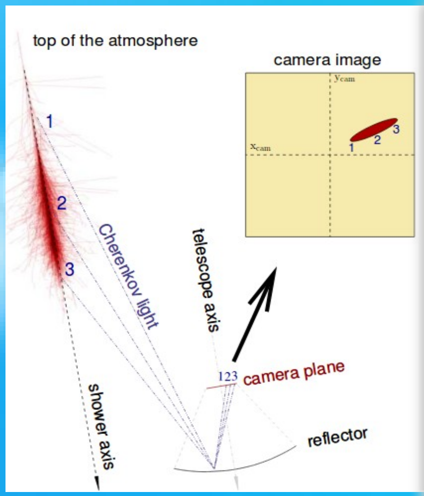
This study was supported by the Russian Science Foundation, grant no. 24-11-00136.

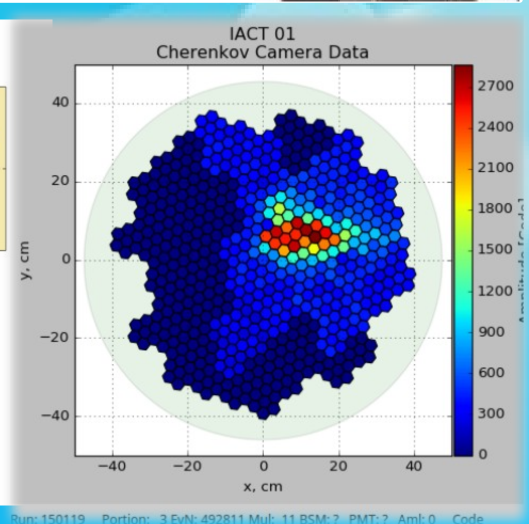
**) Email: kryukov@theory.sinp.msu.ru*

TAIGA experiment

- The TAIGA experimental setup is located in the Tunka Valley (50 km from Lake Baikal) in the Republic of Buryatia. The setup includes several different detectors: TAIGA-IACT, HiSCORE, TUNKA-133, and others.
- We consider the problem of identifying rare gamma events using the example of data from the Imaging Atmospheric Cherenkov Telescope (IACT).
- The TAIGA-IACT setup consists of 4 IACTs, which are reflector telescopes with a 4-meter segmented spherical mirror. At the focus of the telescope is a camera with about 600 photomultipliers for recording Cherenkov radiation emitted by extensive air showers (EAS).

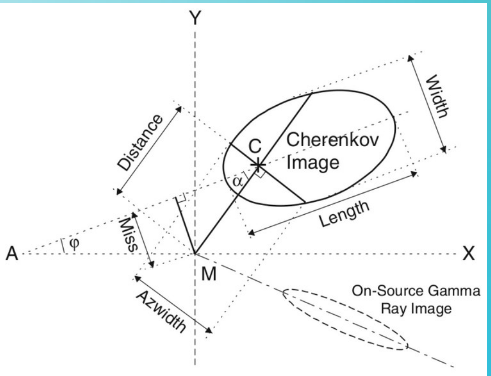


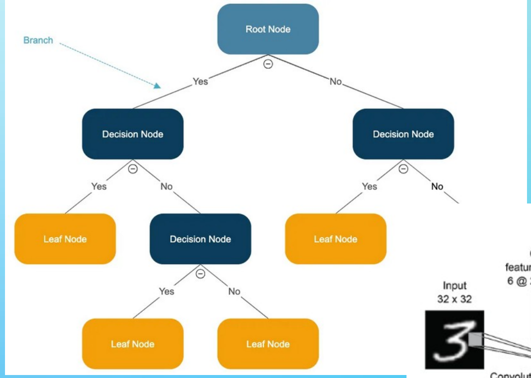


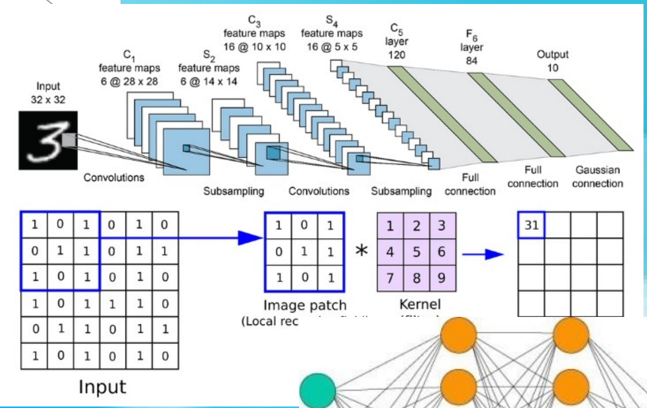


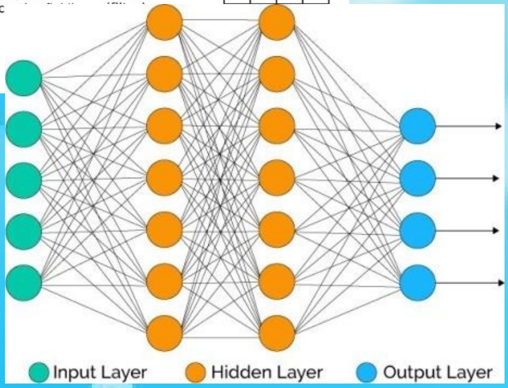
Methods for extracting gamma events

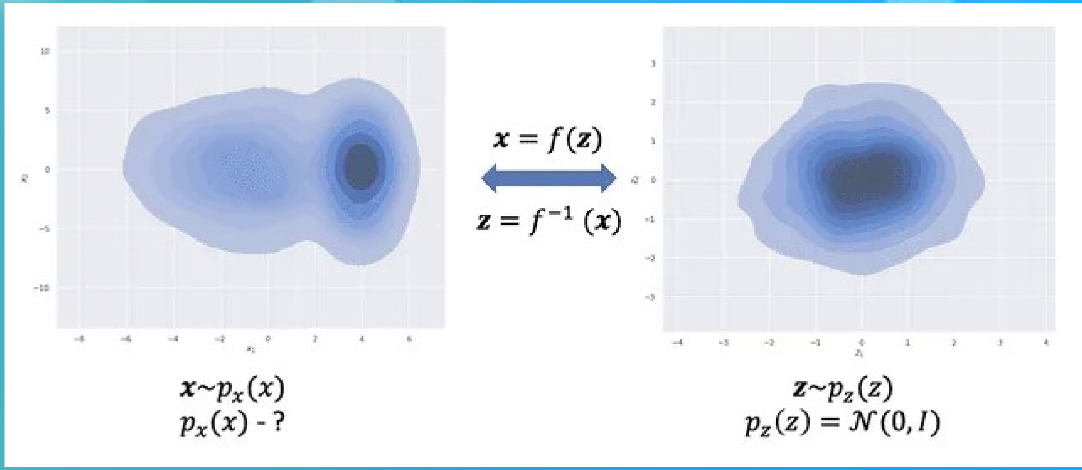
- Traditional method - Hillas parameter cutting
- Machine learning
  - Decision trees
  - Fully connected networks
  - Convolutional networks
  - …
- We propose to use Normalization Flow









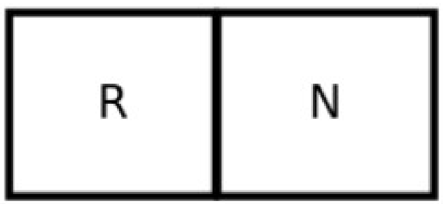


Features of experimental data in gamma-ray astronomy

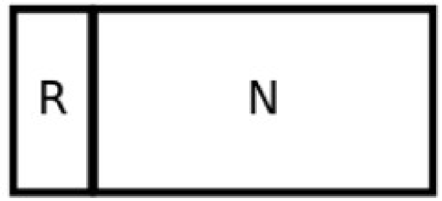
- The main flux of cosmic rays is charged particles:
  - Protons, helium nuclei,
  - Oxygen, nitrogen, …, iron, …
- The fraction of ultra-high energy gamma quanta is less than 0.1%.
- Cherenkov light from EAS, registered by the IACTs, is very weak - from several tens of photoelectrons.
  - High noise level.
- Low resolution: the camera matrix of only several hundred PMTs.
  - TAIGA-IACT ~ 600 PMTs

Anomalous Events vs Rare Events

Rear events

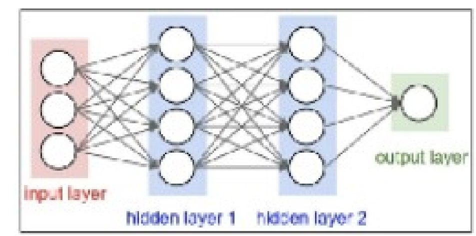


Training set: R~N



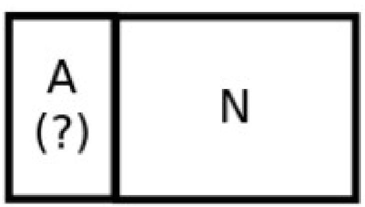
Tesing set: R«N

Two-class learning

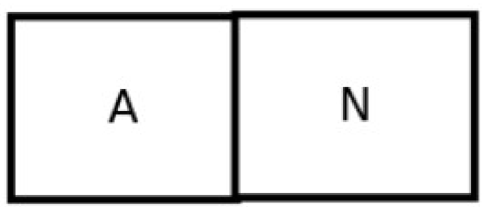


Classifier

Anomalous events

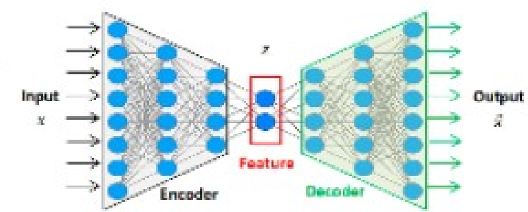


Training set: A«N



Tesing set: A~N

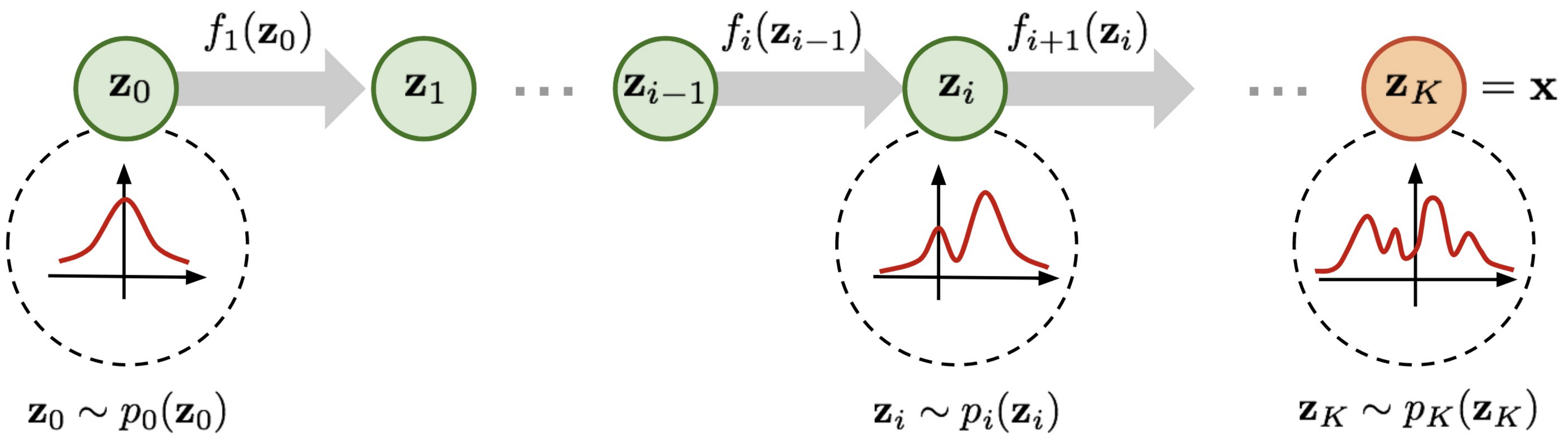
One-class learning



min "MSE"

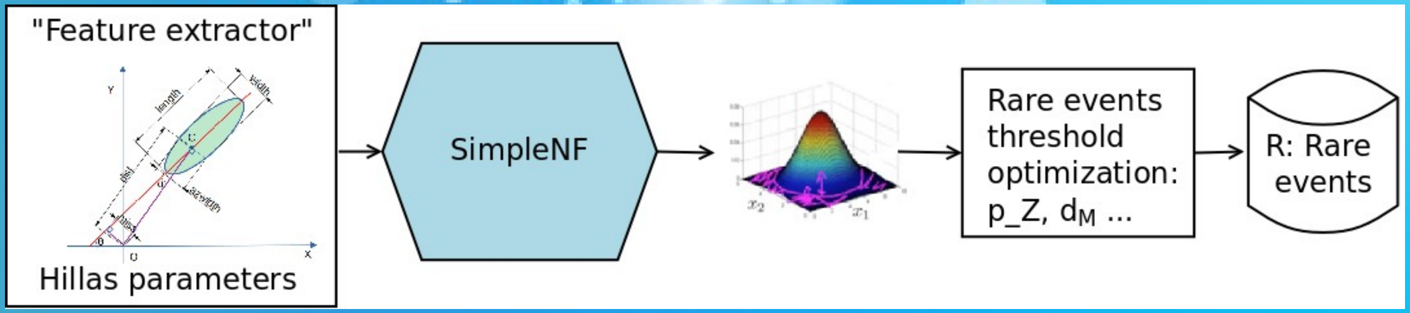
Normalizing Flow

- NF transforms a simple distribution of a random vector into a complex one.
- This is done using a chain of one-to-one transformations.
- The algorithm is based on the theorem on the change of variables under the integral.



Algorithm for identifying anomaly events

1) Selection/extraction of essential features (dimensionality reduction) of data.
2) Transformation of the distribution of non-anomalous data in the training set into a normal one N(0,1).
   - Events in the original distribution will be concentrated around the mean values (zero) of the normal distribution.
   - Since the mapping is optimized for a sample of non-anomalous data, anomalous events will have a low probability.
3) In the original space with a complex distribution (we only have a finite sample), it is difficult to draw a boundary between normal and anomalous data.
4) For a normal distribution, it is easy, for example, the boundary is set at 2σ(95%)



NF transformation chain

$$\mathbf{z}_{i-1} \sim p_{i-1}(\mathbf{z}_{i-1})$$
$$\mathbf{z}_i = f_i(\mathbf{z}_{i-1}), \text{ thus } \mathbf{z}_{i-1} = f_i^{-1}(\mathbf{z}_i) \quad (1)$$

$$\begin{aligned} p_i(\mathbf{z}_i) &= p_{i-1}(f_i^{-1}(\mathbf{z}_i)) \left| \det \frac{df_i^{-1}}{d\mathbf{z}_i} \right| \\ &= p_{i-1}(\mathbf{z}_{i-1}) \left| \det \left(\frac{df_i}{d\mathbf{z}_{i-1}} \right)^{-1} \right| && \Leftarrow \text{the inverse func theorem.} \\ &= p_{i-1}(\mathbf{z}_{i-1}) \left| \det \frac{df_i}{d\mathbf{z}_{i-1}} \right|^{-1} && \Leftarrow \text{Jacobians of invertible func.} \end{aligned}$$

- It is more convenient to work with the sum of terms:

$$\log p_i(\mathbf{z}_i) = \log p_{i-1}(\mathbf{z}_{i-1}) - \log \left| \det \frac{df_i}{d\mathbf{z}_{i-1}} \right| \quad (2)$$

NF transformation chain

It is necessary to construct a chain of transformations such that:

- The Jacobian of the transformation is easily calculated → triangular matrix.
- The inverse functions are easily calculated → affine transformation.
- The derivative of the function is easily calculated → exponent.

$$\begin{cases} \mathbf{y}_{1:d} & = \mathbf{x}_{1:d} \\ \mathbf{y}_{d+1:D} & = \mathbf{x}_{d+1:D} \odot \exp(\sigma(\mathbf{x}_{1:d})) + \mu(\mathbf{x}_{1:d}) \end{cases} \Leftrightarrow \qquad (2)$$

$$\Leftrightarrow \begin{cases} \mathbf{x}_{1:d} & = \mathbf{y}_{1:d} \\ \mathbf{x}_{d+1:D} & = (\mathbf{y}_{d+1:D} - \mu(\mathbf{y}_{1:d})) \odot \exp(-\sigma(\mathbf{y}_{1:d})) \end{cases} \qquad (3)$$

NF transformation chain

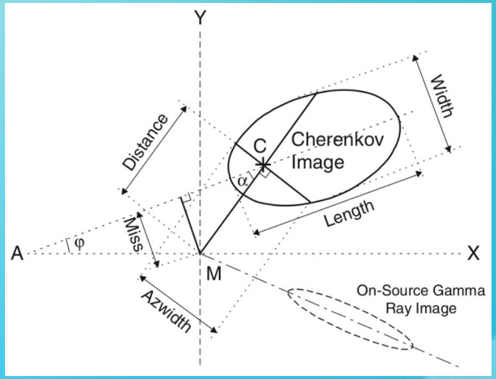
- Jacobian matrix is triangle
- f^{-1} does not require computing the inverse of σ or μ
- not involve computing the Jacobian of σ or μ, ⇒ can be arbitrarily complex;
- both σ and μ can be modeled by deep neural networks
- To make sure all the inputs have a chance to be altered, the model reverses the ordering in each layer so that different components are left unchanged.

$$\mathbf{J} = \begin{bmatrix} \mathbb{I}_d & \mathbf{0}_{d\times(D-d)} \\ \frac{\partial \mathbf{y}_{d+1:D}}{\partial \mathbf{x}_{1:d}} & \mathrm{diag}(\exp(\sigma(\mathbf{x}_{1:d}))) \end{bmatrix}$$

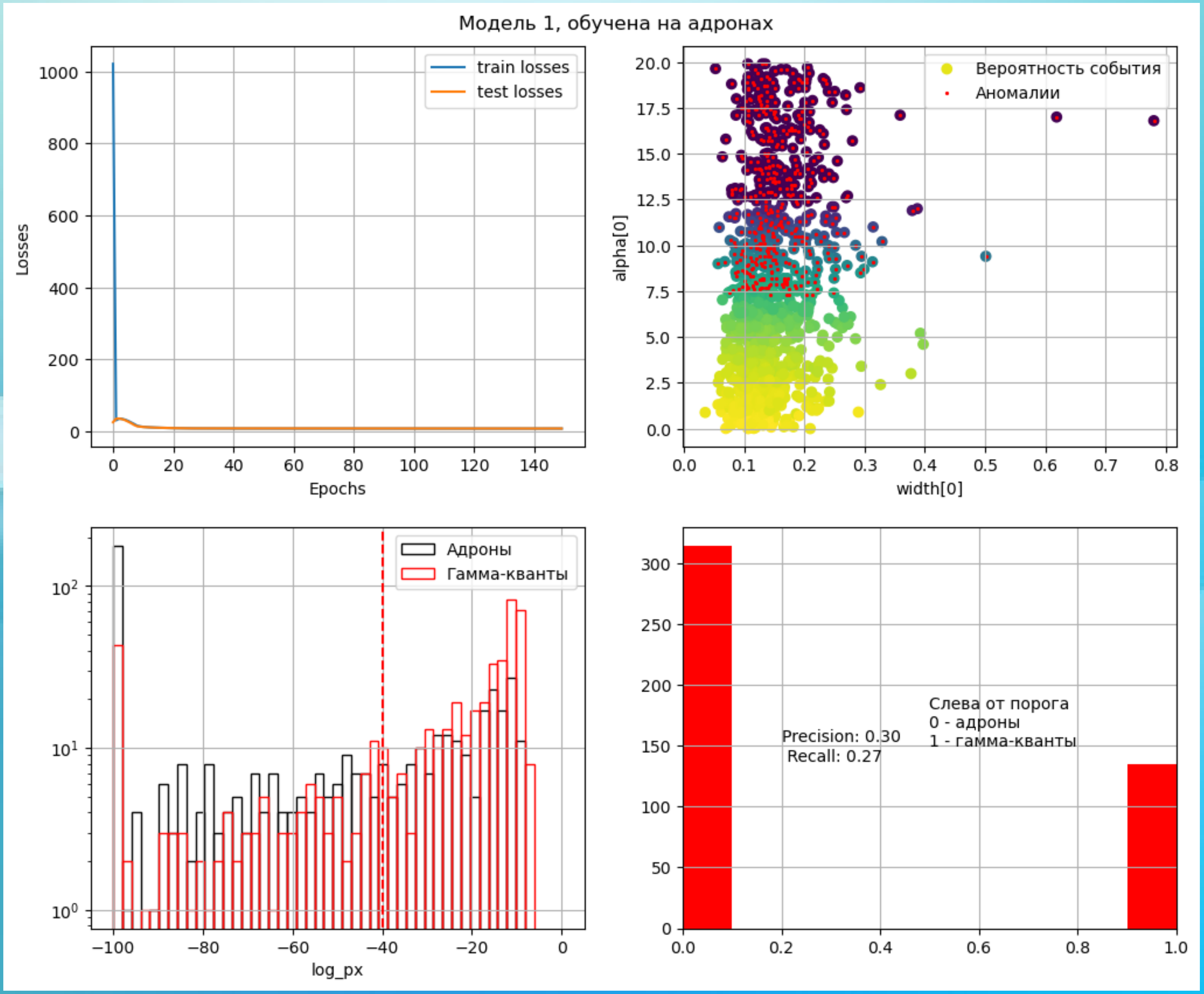
$$\det(\mathbf{J}) = \prod_{j=1}^{D-d} \exp(\sigma(\mathbf{x}_{1:d}))_j = \exp(\sum_{j=1}^{D-d} \sigma(\mathbf{x}_{1:d})_j)$$

NF using TAIGA-IACT as an example

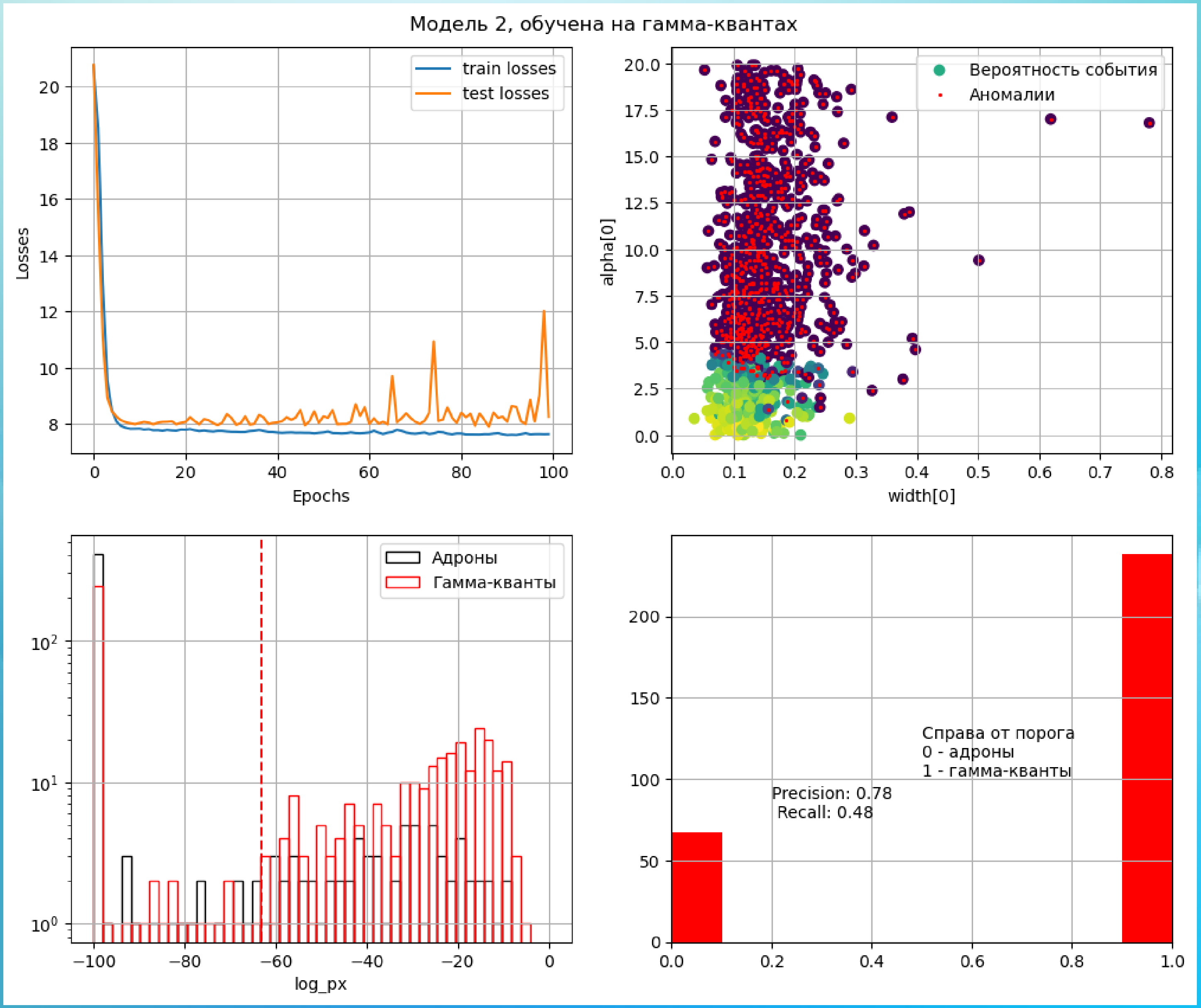
- Hillas parameters included in NF training
  - Size > 100
  - 0.02 < Width < 0.8
  - Alpha < 20
  - Dist < 3
  - Thus, the feature vector has a dimension of 4.
- Training and validation set
  - Train_size = 3000
  - Test_size = 500
- 8 normalization blocks were used
  - Each block μ and σ are neural networks, which contain 2 hidden fully connected layers of 512 neurons



Hadrons as normal events



Gammas as normal events



Conclusions

- It was demonstrated on small neural network models that normalizing flows have the potential to isolate rare gamma events.
- The most promising application of NF is the case when protons events rather than gamma events are considered anomalous.
- Preliminary analysis has shown that simple NF models do not provide reliable extraction of gamma events. Possible reasons are:
  - small size of the training set;
  - it is necessary to use advanced neural network NF models with a large number of transformation blocks, layers and neurons in layers;
  - a poor choice of the input parameter set.
- We have plans to continue research this method on more complex models and much larger training samples.
  - In particular, a variant of two-class classification with NF, when a normal event has a positive weight in the loss function, and anomalous events have a negative weight.
- We propose to study the possibility of replacing Monte Carlo events with experimental data far from gamma sources as non-anomalous events.

Литература

- hhttps://towardsdatascience.com/unsupervised-learning-for-anomaly-detection-44c55a96b8c1/ttps://towardsdatascience.com/unsupervised-learning-for-anomaly-detection-44c55a96b8c1/
- Yu, Jiawei, et al. "Fastflow: Unsupervised anomaly detection and localization via 2d normalizing flows."arXiv preprint arXiv:2111.07677 (2021).
- Chiu, Li-Ling, and Shang-Hong Lai. "Self-supervised normalizing flows for image anomaly detection and localization."In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 2927-2936. 2023. https://openaccess.thecvf.com/content/CVPR2023W/VAND/papers/Chiu_Self-Supervised_Normalizing_Flows_for_Image_Anomaly_Detection_and_Localization_CVPRW_2023_paper.pdf
- I Kirichenko, Polina, Pavel Izmailov, and Andrew G. Wilson. "Why normalizing flows fail to detect out-of-distribution data."Advances in neural information processing systems 33 (2020): 20578-20589. https://proceedings.neurips.cc/paper/2020/file/ecb9fe2fbb99c31f567e9823e884dbec-Paper.pdf
- Dohi, Kota, Takashi Endo, Harsh Purohit, Ryo Tanabe, and Yohei Kawaguchi. "Flow-based self-supervised density estimation for anomalous sound detection."In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 336-340. IEEE, 2021. https://arxiv.org/pdf/2103.08801

Thank you!

Questions?