Image Data Augmentation for the TAIGA-IACT Experiment with Conditional Generative Adversarial Networks

Dubenskaya Yu.[1], Kryukov A.[1], Demichev A.[1], Gres E.[2], Polyakov S.[1], Postnikov E.[1], Volchugov P.[1], Vlaskina A[1], Zhurov D.[2]

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

The work was supported by RSF, grant no.24-11-00136

06.2024

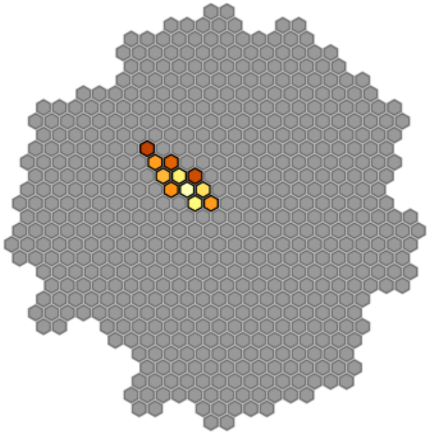
Air showers detection with IACT

Charged cosmic rays and high energy gamma rays interact with the atmosphere

The result is extensive air showers (EAS) of secondary particles emitting Cherenkov light

Imagine Atmospheric Cherenkov Telescopes (IACT) detect the light

The figure on the right shows the reflecting telescope of the TAIGA-IACT project



Once detected by the telescope's camera the light results in an "image" of the corresponding air shower event

The two types of events are gamma and hadron events

Artificial images generation

For each IACT to operate correctly a large amount of experimental data, including simulated data, is required

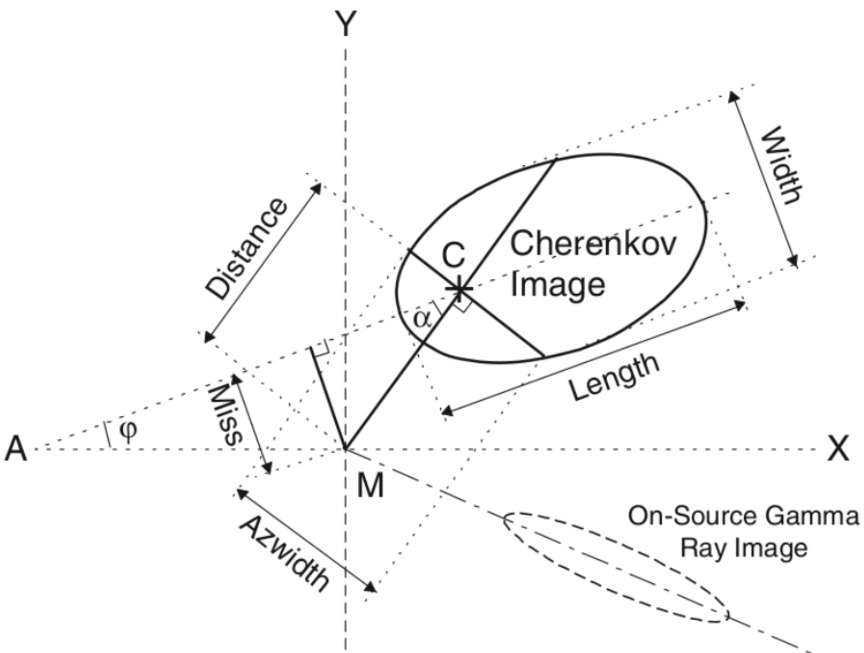
Moreover, solutions based on neural networks are increasingly being used to classify incoming events. Training these networks also requires simulated data

Traditionally, event images are modeled by Monte Carlo simulation (by performing direct simulation of extensive air showers), thereby producing reasonably accurate but resource-intensive and time-consuming results

Machine learning techniques such as generative adversarial networks (GANs) significantly reduce the time to generate images

In this work we focus on generation of artificial <u>gamma</u> images using conditional GANs (cGANs)

Hillas parameters and the energy of the primary particle



For each event image we can calculate the so-called Hillas parameters, which form a set of geometric features of the image

The most important Hillas parameters are:

- Image brightness
- Width and length of the ellipse
- Number of triggered pixels and distance
- Angles: alpha, phi

In both image analysis and image generation, the main parameter of interest is the <u>energy of the primary particle</u>

Unlike other image parameters, energy can only be approximately determined from the image (with or without neural network solutions)

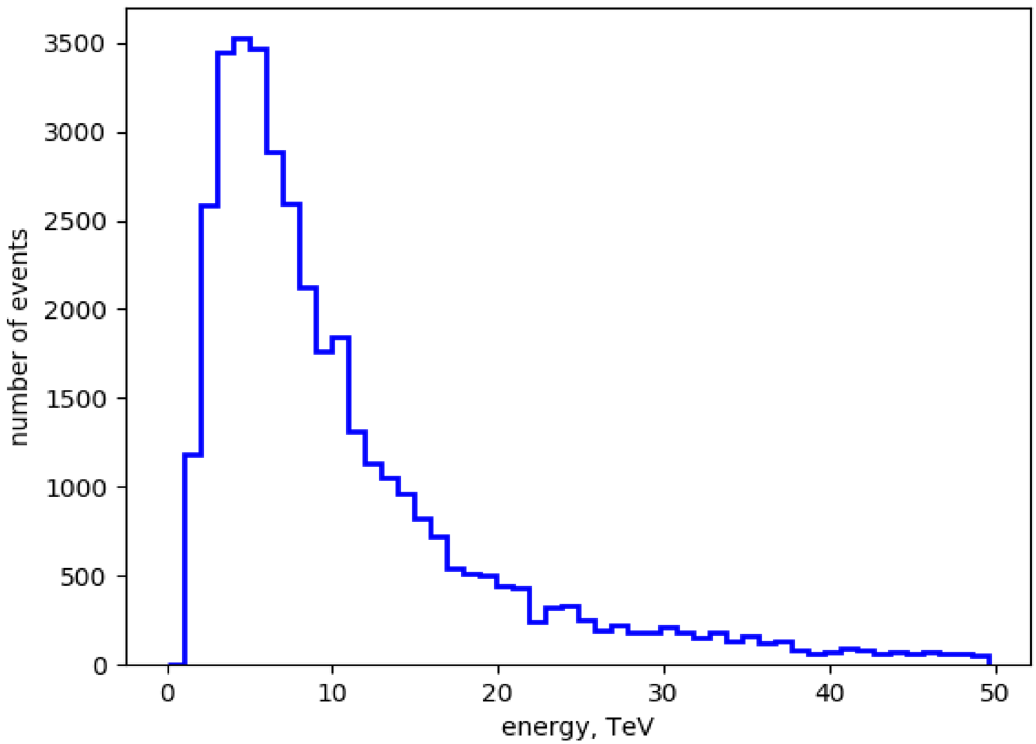
The energy is only known accurately for the Monte Carlo simulated data

Energy distribution

The real-life energy distribution of the IACT data is uneven and asymmetrical

For some scientific applications it is important to reproduce the energy distribution of incoming events

To train other neural networks, the resulting energy distribution must be close to uniform. To do this, one needs to generate more events, which are de facto rare. This is called <u>data augmentation</u>



In this work, we propose a way to quickly obtain data with any of these distributions

Conditional generative adversarial network (cGAN)

cGAN is a modification of a traditional GAN that allows you to divide images into multiple classes according to the value of some property of the image

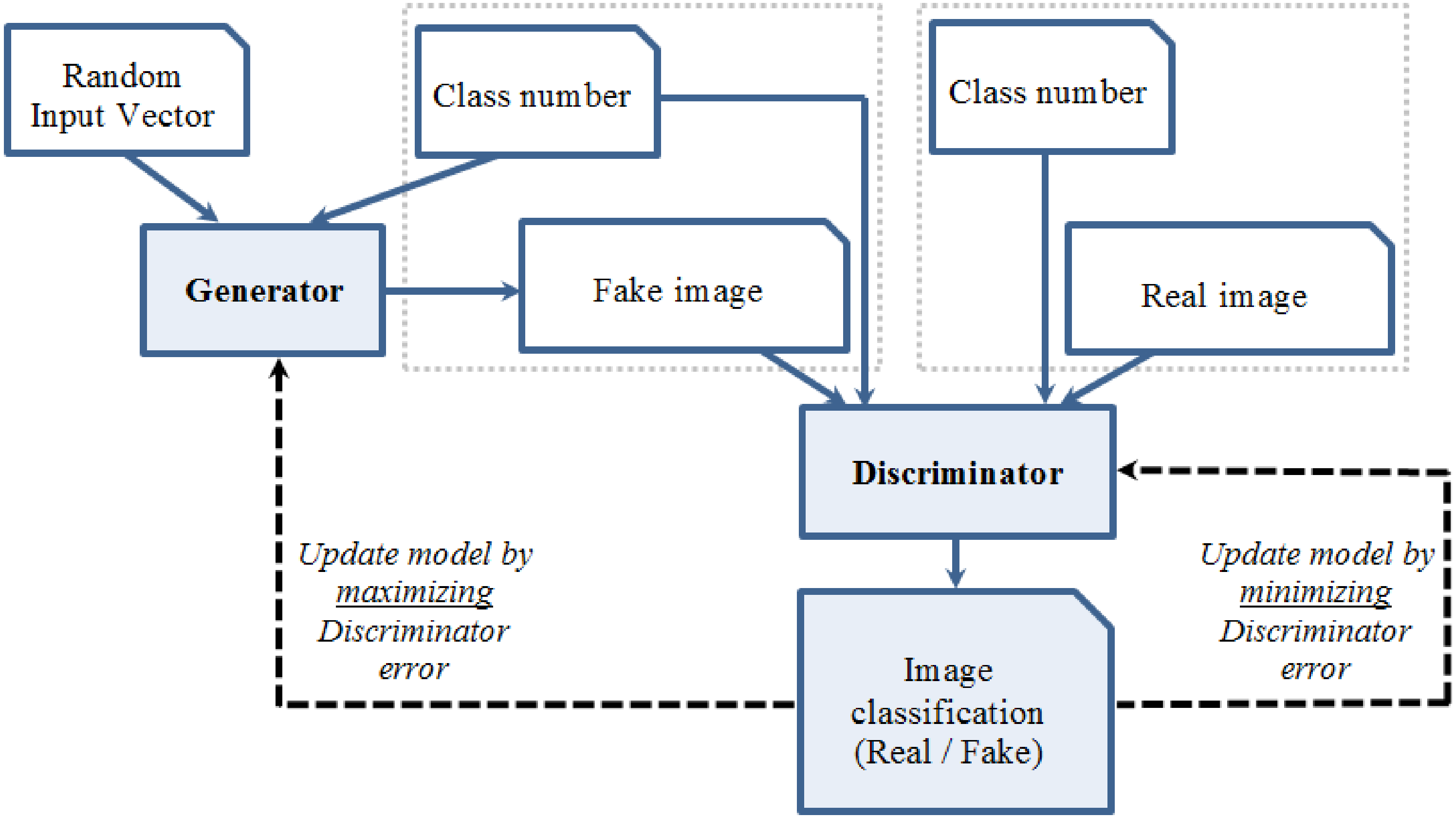


Image generation algorithm

A cGAN is trained to simulate the IACT images

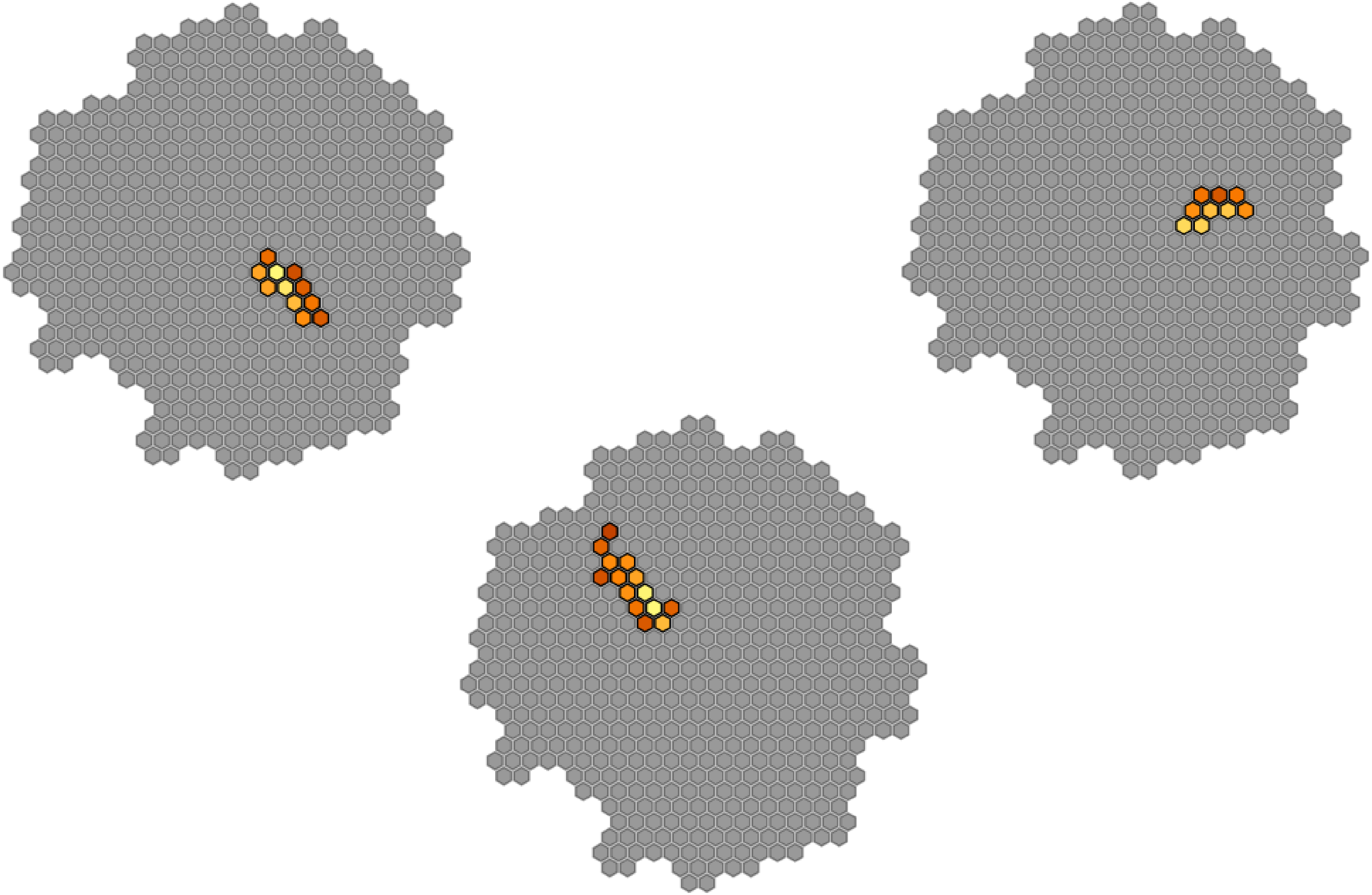
The Monte Carlo simulated data with known energy is used for training this cGAN

We divided our training sample of gamma images by energy, so that the images with the similar energy fall into the same class. Thus we got 10 classes by energy. We division was done so that each class has the same number of images

Similarly, we divided the images of each energy class into 10 classes by distance. As a result, we received 100 classes to train the cGAN

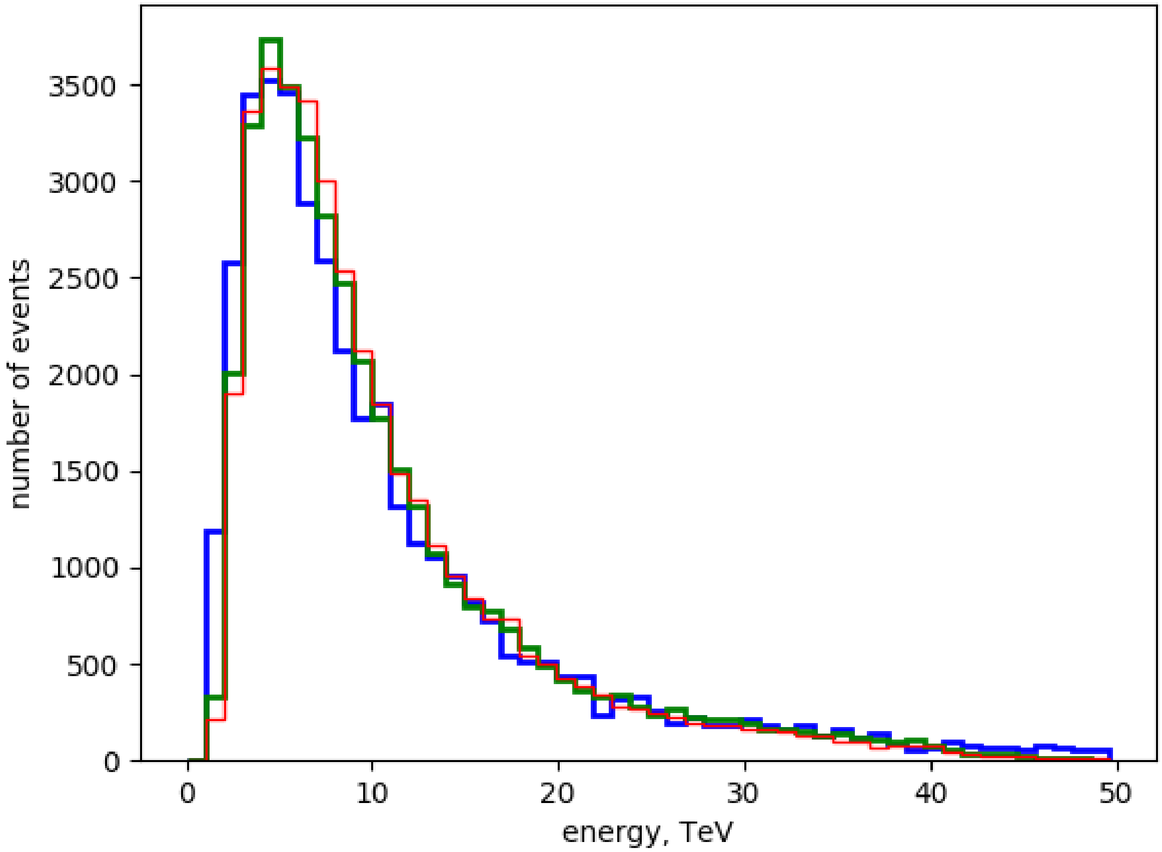
To determine the energy of the resulting dataset we used a separate neural network previously trained on Monte Carlo data

Image generation examples



Reproducing the incoming energy distribution

For each class we generate the same number of images. The energy distribution summed over all classes is close to the original distribution of the training set



Monte Carlo energy values (for the training set)

Energy estimation by neural network (for the training set)

Energy estimation by neural network (for generated images)

Data augmentation

To get the energy distribution for each class we need to calculate the number of images to generate. The idea is that the wider the class boundaries are, the more events we need to generate

First, for each energy class we calculate the coefficient (i=0..9):

$K_e^i=(E_{max}^i-E_{min}^i)/(E_{max}-E_{min})$

E_{max}^i, E_{min}^i – class boundaries by energy

E_{max}, E_{min} – max and min energy values

Next for each energy class we calculate the distance coefficients (j=0..9):

$K_d^{ij}=(D_{max}^{ij}-D_{min}^{ij})/(D_{max}^i-D_{min}^i)$

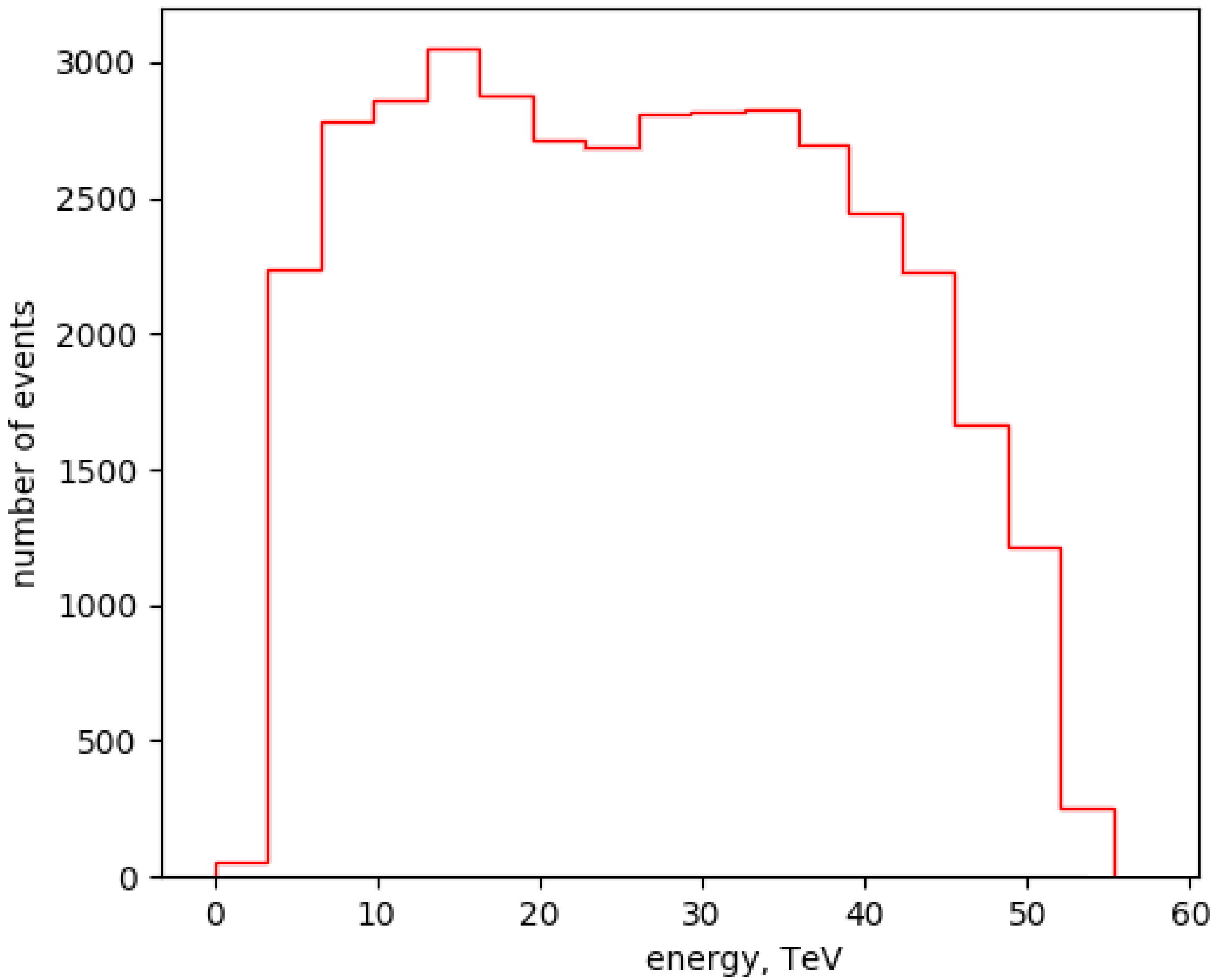
D_{max}^{ij}, D_{min}^{ij} – class boundaries by distance

D_{max}^i, D_{min}^i – max and min distance values

Then we calculate the number of images for each class by multiplying the total number of the generated images by the corresponding coefficients K_e^i and K_d^{ij}

Data augmentation results

The energy distribution summed over all classes is not uniform, but is much closer to a uniform distribution than the original one



Energy estimation by neural network (for generated images)

Conclusion

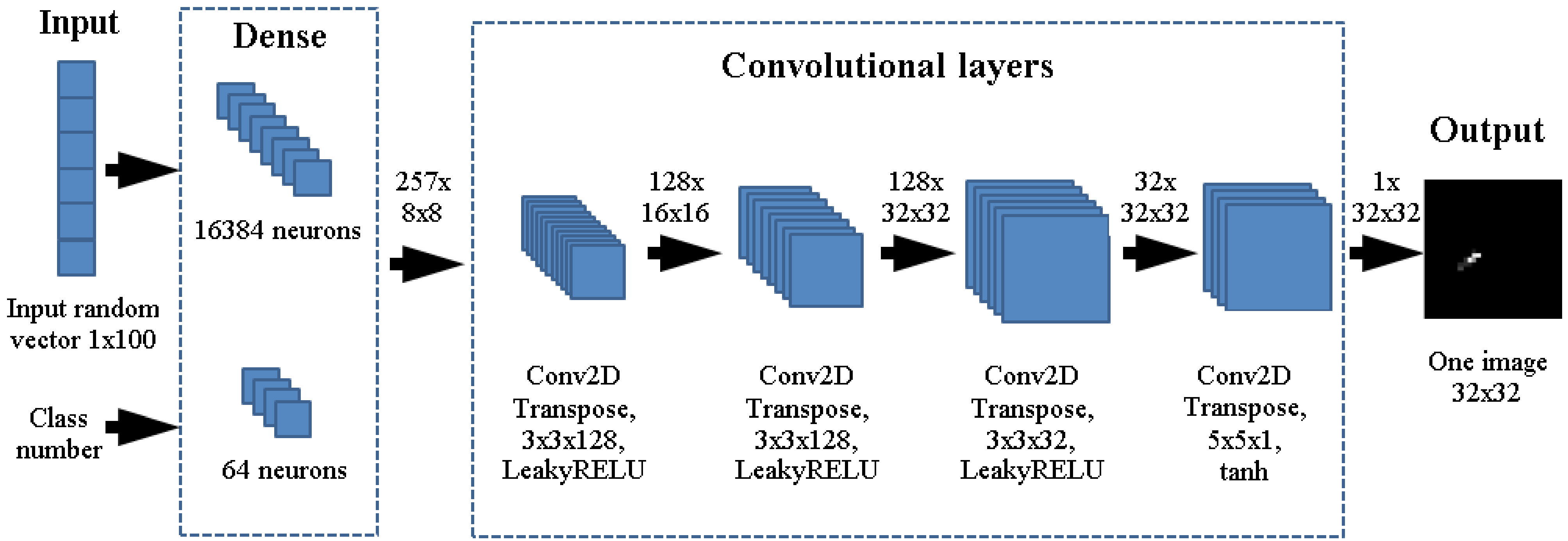
A conditional generative adversarial network simulate images for the TAIGA-IACT experiment with a very good degree of accuracy

Using the proposed algorithm, we can quickly generate images with either the original incoming energy distribution or perform data augmentation to obtain a nearly uniform energy distribution

The rate of images generation using cGAN is more than 1000 times higher than the rate of generation by the traditional method

Thank you for attention!

Generator architecture



Discriminator architecture