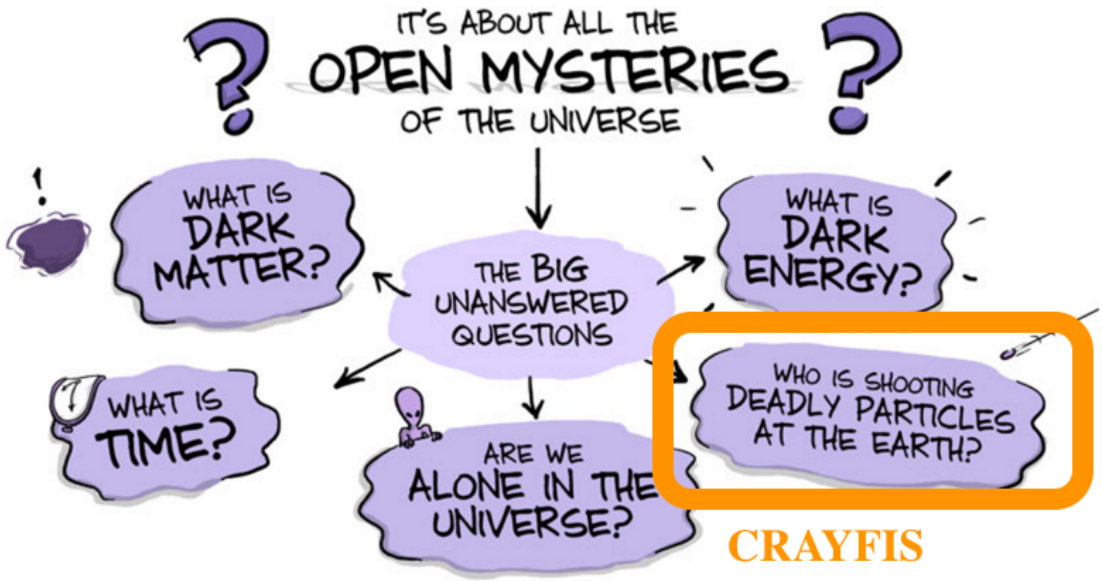
Modeling detector digitization and read-out with adversarial networks

ACAT, Seattle, 2017-08-21

Maxim Borisyak[1], Chase Shimmin[3], Andy Nelson[2], Andrey Ustyuzhanin[1]

[1] Yandex, NRU Higher School of Economics
[2] University of California Irvine
[3] Yale University



Illustrations from the book "We have no idea" by D. Whiteson, J. Cham.

Cosmic RAYs Found In Smartphones Experiment

CRAYFIS experiment proposes usage of private mobile phones for observing Ultra-High Energy Cosmic Rays (UHECR):

- high energies: $> 10^{18}$ eV;
- distributed world-wide observatory;
- mobile phone's camera as cosmic rays detector;
- cluster of mobile phones as intensive air shower detector.

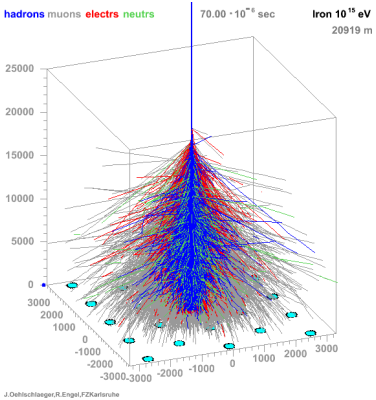
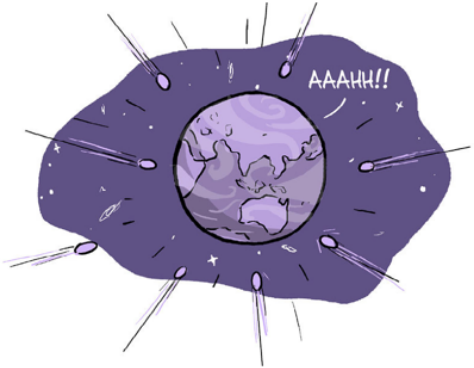


Illustration of an intensive air shower produced by iron ion, 1 PeV, CORSIKA simulation, by J. Oehlschläger and R. Engel.

CRAYFIS in a nutshell



Illustrations from the book "We have no idea" by D. Whiteson, J. Cham.

Challenges

Physics:

- low signal event rate is expected:
  - background cosmic rays ($\approx 10^{13}$eV): $>$ 1000 per second per km^2;
  - UHECR ($\approx 10^{18}$eV): less than once per year per km^2;
- an intensive air shower from UHECR occurs in less than microseconds;

Data processing:

- Getting realistic muon track images:
  - how muons interact with smartphone cameras (no ground-truth)?
- Tracking muons using smart phones:
  - shortage of computational power and storage space (mobile phones);
  - high frame rate processing is required (~ 10 Hz);
  - limited throughput for selected images (end-user Wi-Fi $<$ 1 Mbit$/$s);

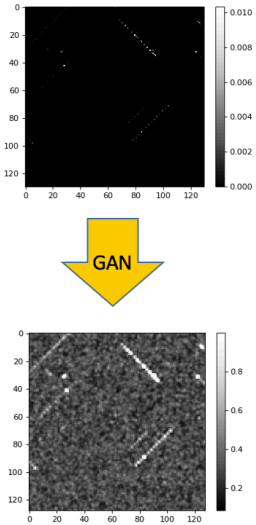
Getting realistic images of muons (Parti-GAN)

The problem

A simulation with simplified geometry and without readout process is relatively simple (GEANT).

But there is no CMOS sensors details, precise enough for reliable simulation of particle-sensor interaction:

- various type of sensors;
- impossible to tune for every phone;
- muons are difficult to find & to prove.



Let's see if GANs can solve it

- Dataset:
  - simulated by GEANT images of energy deposition;
  - real images from radioactive source. no labels;
  - simple toy model adjusted to real data;
  - approximate ratio of signal/background: 0.001.
- Classical GAN - doesn't deal with images as input and converges poorly;
- Cycle-GAN - should help finding 1:1 mapping, but does not solve all problems alone;
- Energy-Based - should help convergence, but still doesn't work.

What if we add physics-based insights into the training of the generator?

Overwhelming amount of noise

Trick 1: importance sampling (real batch)

- increase number of events with higher signal probability in the batch reduce variance of the gradient;
- events are reweighted to keep original signal/noise ratio:

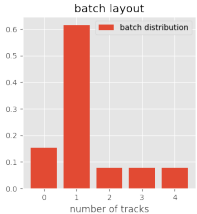
$$\mathcal{L}_{\text{real}} = \frac{1}{n} \sum_i w_i l_{\text{real}}(y_i)$$

where:

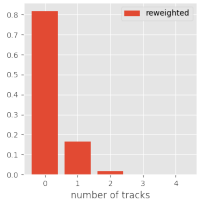
- $w_i \sim \frac{1}{p_i}$ - weight to compensate for change in sample distribution;
- p_i - sampling probability for i-th sample;
- we can use image brightness of the image as a proxy for p_i.

Structure of a batch

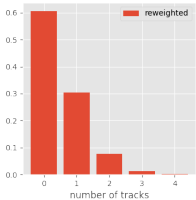
$8\times$ $32\times$ $4\times$ $4\times$ $4\times$



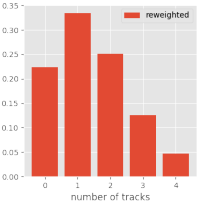
original



$\lambda = 0.2$



$\lambda = 0.5$



$\lambda = 1.5$

Helping GAN to learn the signal

- simulation does not account for signal event rate (just provides examples of interaction);
- signal event rate λ in real observations is not known exactly;

Trick 2: introduce λ as a GAN parameter to optimize;

- generate m_k samples with k events up to a large k;
- apply reweighting to redistribute number of events to match $\mathrm{Poisson}(\lambda)$:

$$\mathcal{L}_{\mathrm{pseudo}} = \sum_k w^k \sum_{j=1}^{m_k} l_{\mathrm{pseudo}}(G(x_j^k))$$

where: $w^k = \frac{1}{m_k} \frac{\lambda^k e^{-\lambda}}{k!}$ - redistribution term.

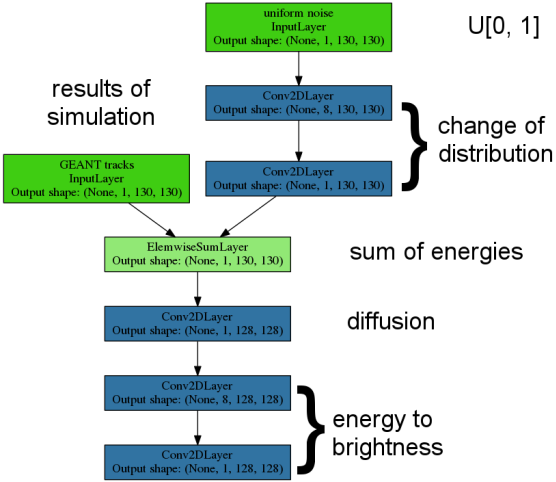
Technical details. Overview

- Energy-Based GAN for [fast] convergence;
- Cycle-GAN to ensure bijection mapping between GEANT and generated samples;
- Batch reweighting to ensure convergence:
  - importance sampling for real images (decrease variance of gradient estimations);
  - physics process parameter λ for generated samples.

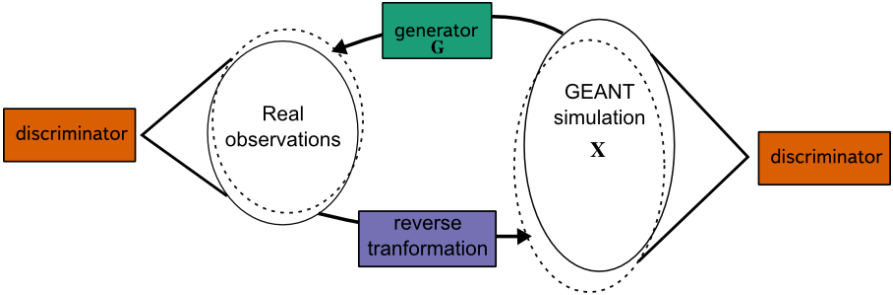
Constructing generator. Assumptions

- Observed samples are result of two independent 'processes':
  - various kinds of noise;
  - particle-sensor interaction;
- Track and noise energies can be added to each other:
  - simulation results can be added to various noise;
- The simulated process is local:
  - restricted perception field of generator (3×3);
- CMOS pixel brightness is of functional dependency on pixel energy deposit (adjacent pixels).

Generator structure



Cycle-GAN architecture (Parti-GAN is the same)

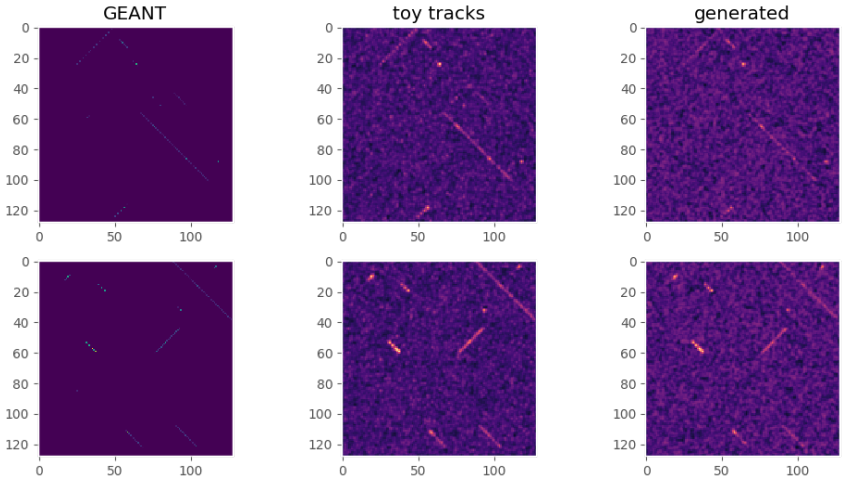


Parti-GAN Loss function

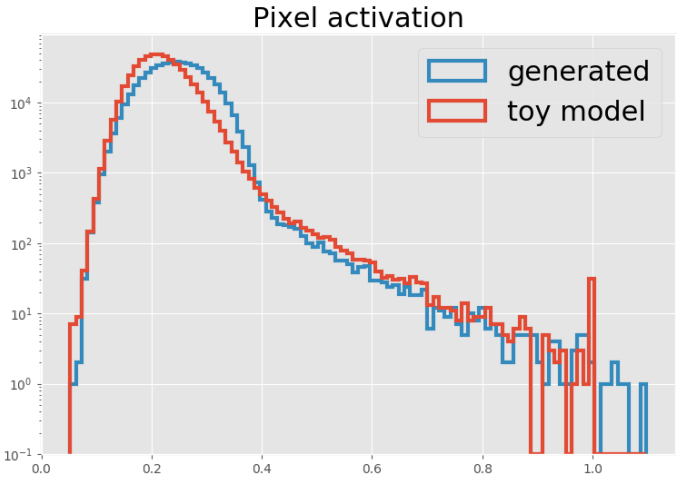
$$L_D^X(X, X') + L_D^Y(Y, Y') + L_C^X(X, X'') + L_C^Y(Y, Y'')$$

- X - bunch of GEANT images, Y - bunch of real images;
- X' - 'real' images generated from X, $G(X)$ in the batch;
- Y' - 'GEANT' images generated from real images Y, $\tilde{G}(Y)$ in the batch;
- $X'' : \tilde{G}(G(X))$, $Y'' : G(\tilde{G}(X))$;
- L_D^X, L_D^Y - EBGAN loss functions;
- L_C^X, L_C^Y - cycle loss functions (MSE);
- X are weighted by physical sampling coefficients;
- Y are weighted by importance sampling coefficients.

Parti-GAN results



Results cross-check: pixel intensity



Summary

Parti-GAN learns actual physical process (λ) + preprocessing algorithm:

- corrections on electron drift;
- thermal noise, readout noise;
- physical/hardware readout systems.

Parti-GAN matches simulation to real observations (”unpaired image translation”)
Parti-GAN is based on CycleGAN with EBGAN loss function + importance sampling

It can generate realistic images of muon tracks for any phone model!

Thank you for attention!

Backup

CRAYFIS collaboration

University California Irvine:

- Daniel Whiteson;
- Homer Strong;
- Jay Karimi;
- Kyle Brodie;
- Rob Porter;
- Eric Albin;
- Emma He;
- Zichao Ziliver Yuan;
- Jeff Swaney;

Yale University:

- Chase Shimmin;

University of California Davis:

- Michael Mulhearn;
- Dustin Burns;
- Brandon Buonacarsi;

New-York University:

- Kyle Cranmer;

Yandex School of Data Analysis:

- Andrey Ustyuzhanin;
- Maxim Borisyak;
- Mikhail Usvyatsov;

National Astronomical Observatory of China:

- Jianrong Deng;

Unaffiliated:

- Danielle Cranmer;
- Jodi Goddard.