Machine Learning at Telescope Array

Machine Learning for Analysis of High-Energy Cosmic Particles, 01.25

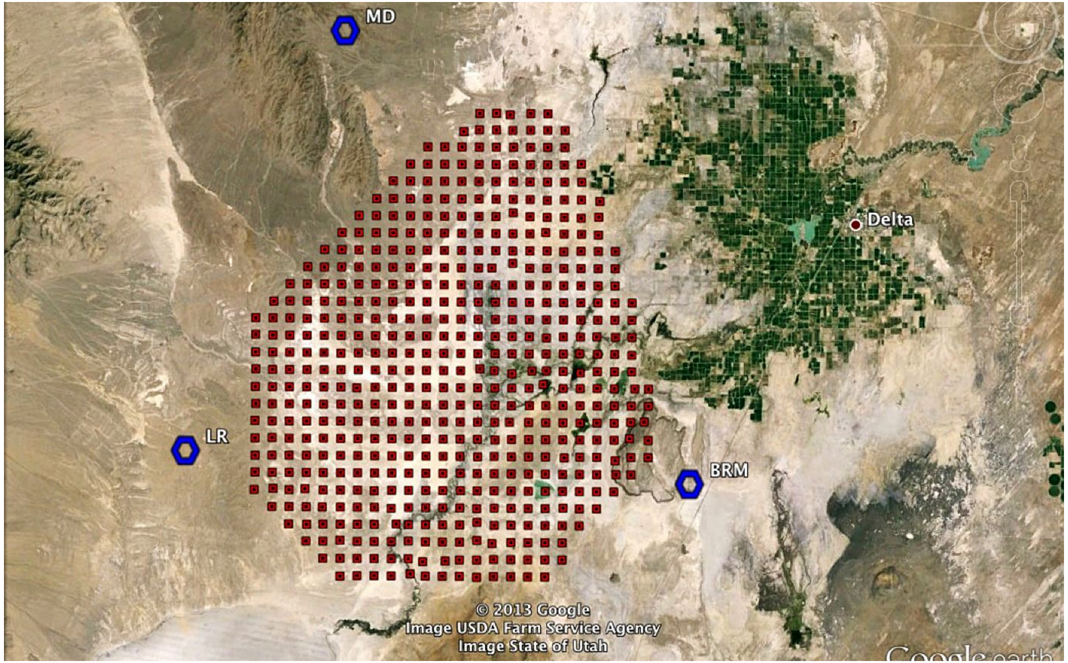
ivan.kharuk@phystech.edu

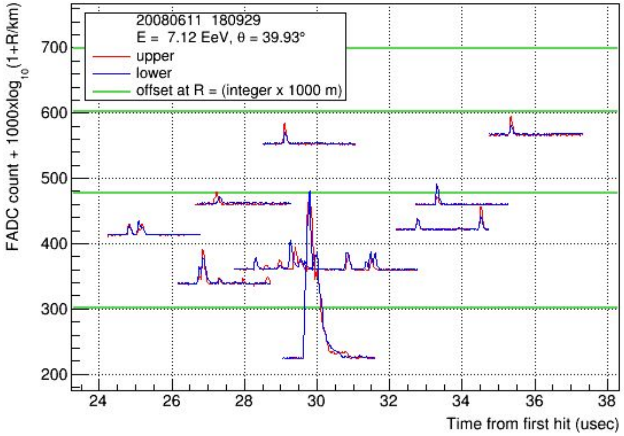
Outline

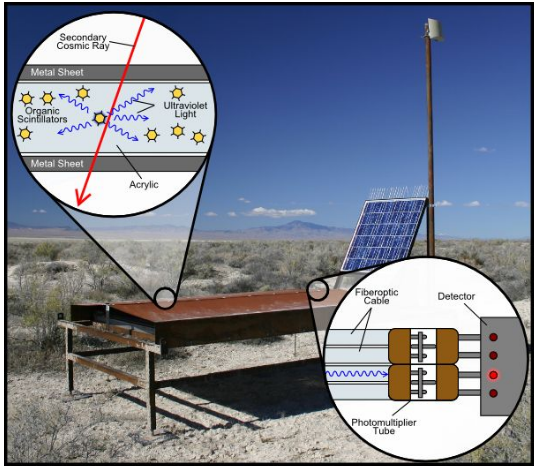
- Composition-spectrum and arrival direction studies
- Towards reliable neural networks
- Autoencoder

Telescope Array

Largest cosmic ray observatory in Northern Hemisphere
(700 km^2, 507 Surface Detector stations + 3 Fluorescent detectors, TA Low Energy Extension)







1. EAS composition, energy and arrival direction reconstruction

Getting most out of the data

- Reconstructed parameters (domain knowledge)

+

- Detector bundle (geometrical features)

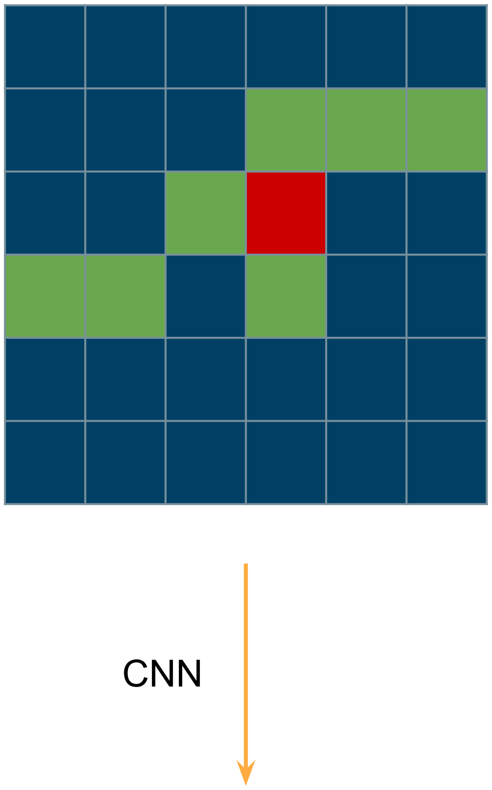
+

- Detector sequence (temporal features)

Predictions

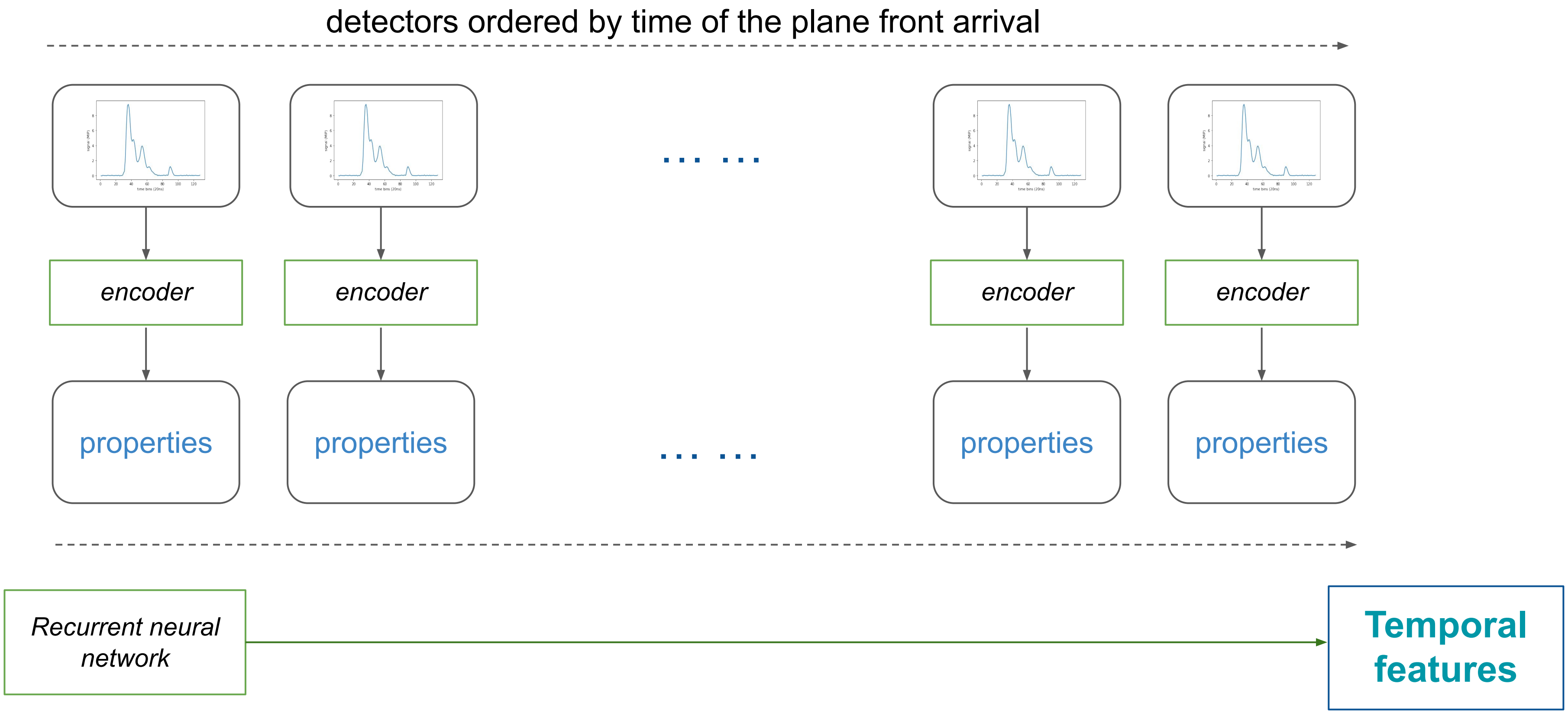
Detectors bundle

- x, y, and z coordinates of the detector
- Detector's total signal
- Time of the plane front arrival
- Difference in time between the start of the recorded signal and the wavefront arrival
- Masks:
  - Was triggered?
  - Was saturated?

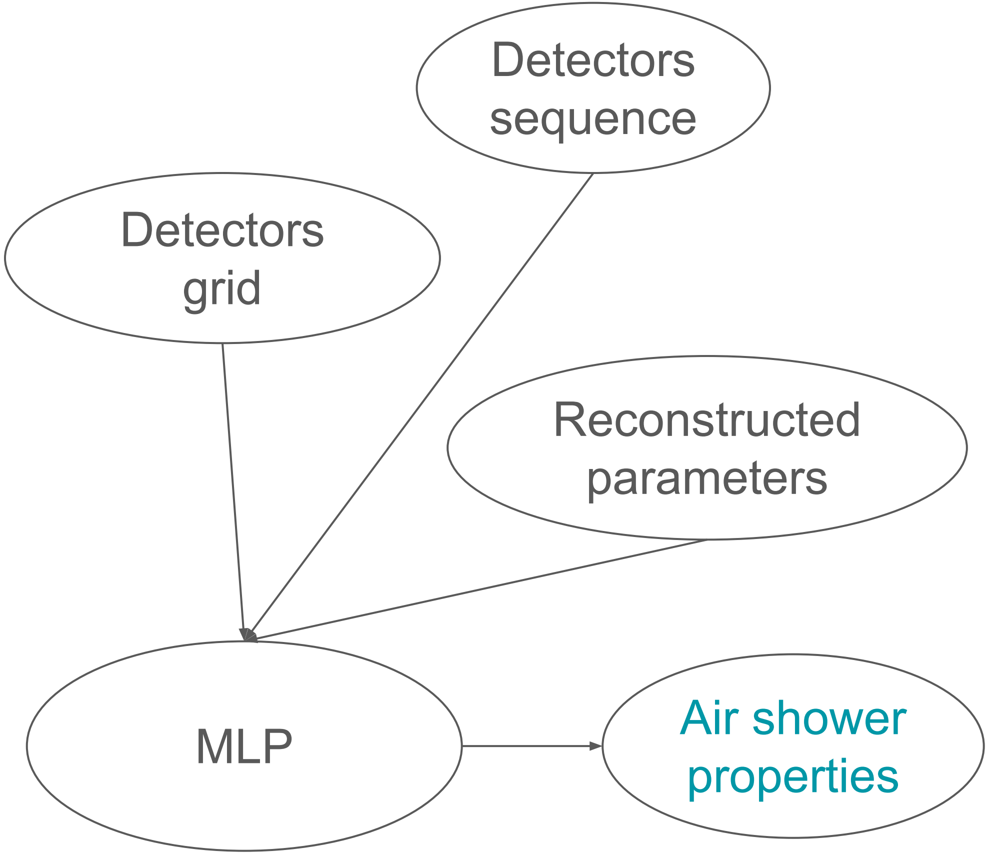


Spatial detectors features

Detectors sequence



Full NN



Technical remarks:

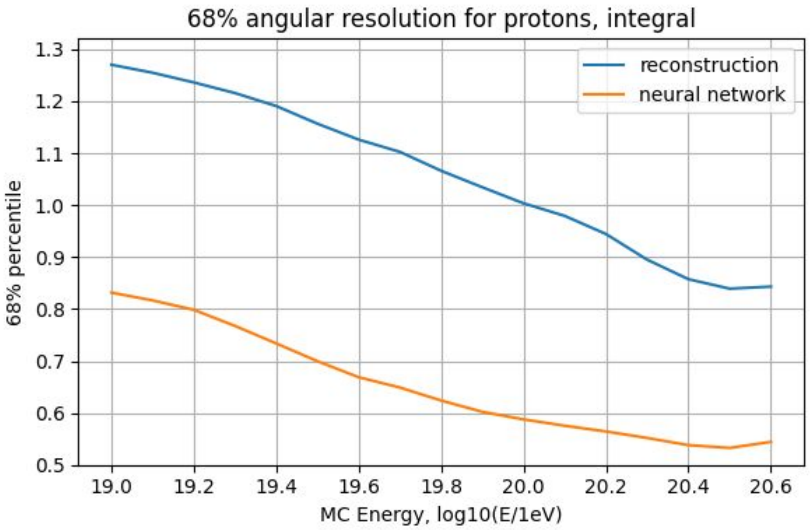
- Cosine similarity for direction reconstruction.
- E^{-1} differential EAS spectrum for training.

- Using Layer Normalization yields better metrics
- Better to train only for one target prediction.
- Small networks ($\sim 10^4$ parameters) perform best.
- Transformers yield similar metrics.

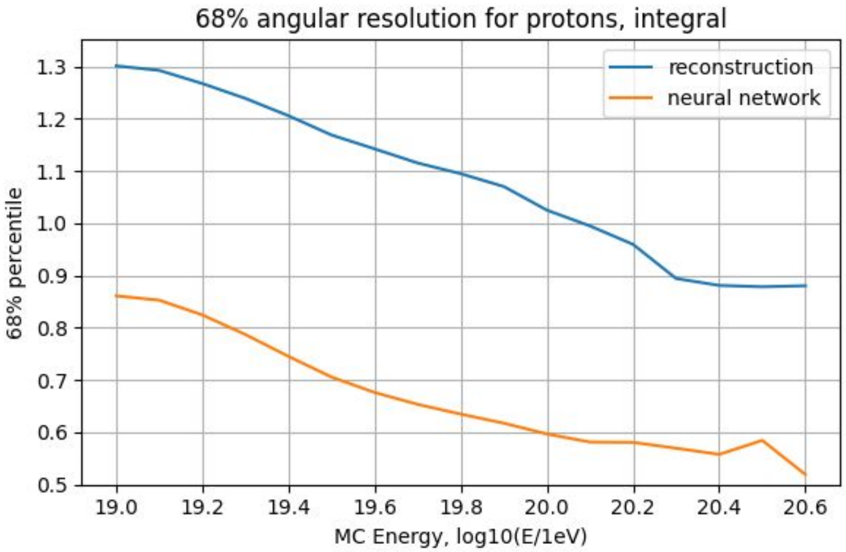
Angular resolution

- ~0.4° angular resolution improvement
- Weak model dependence

"Native" QGSJET-II-4



EPOS-LHC

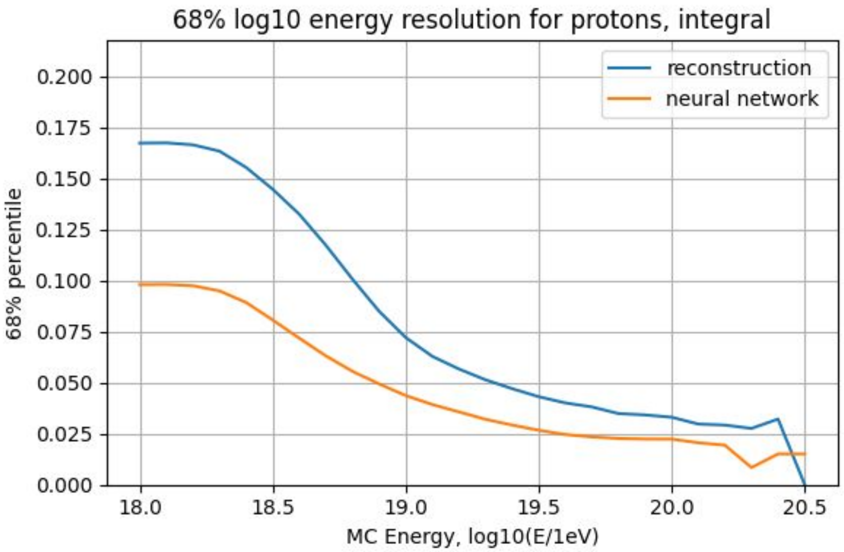


HiRes spectrum

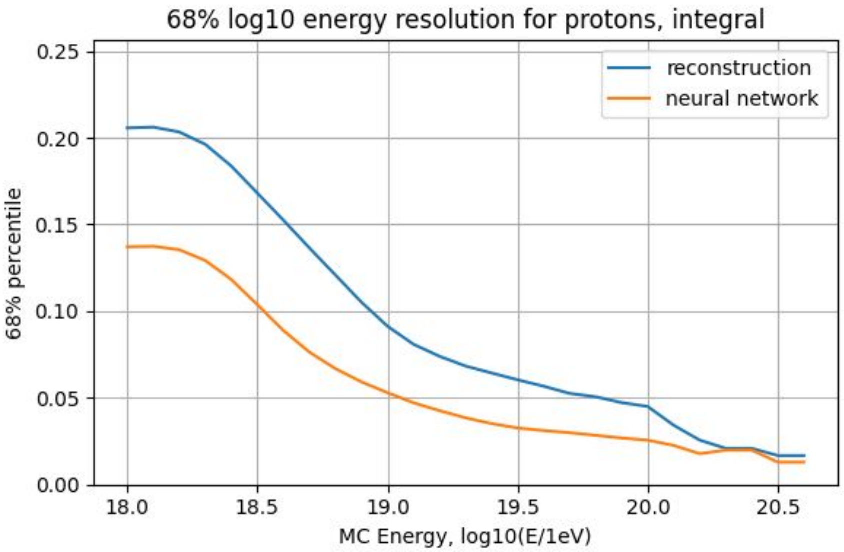
Energy resolution

- Better energy reconstruction with NN
- Moderate model dependence

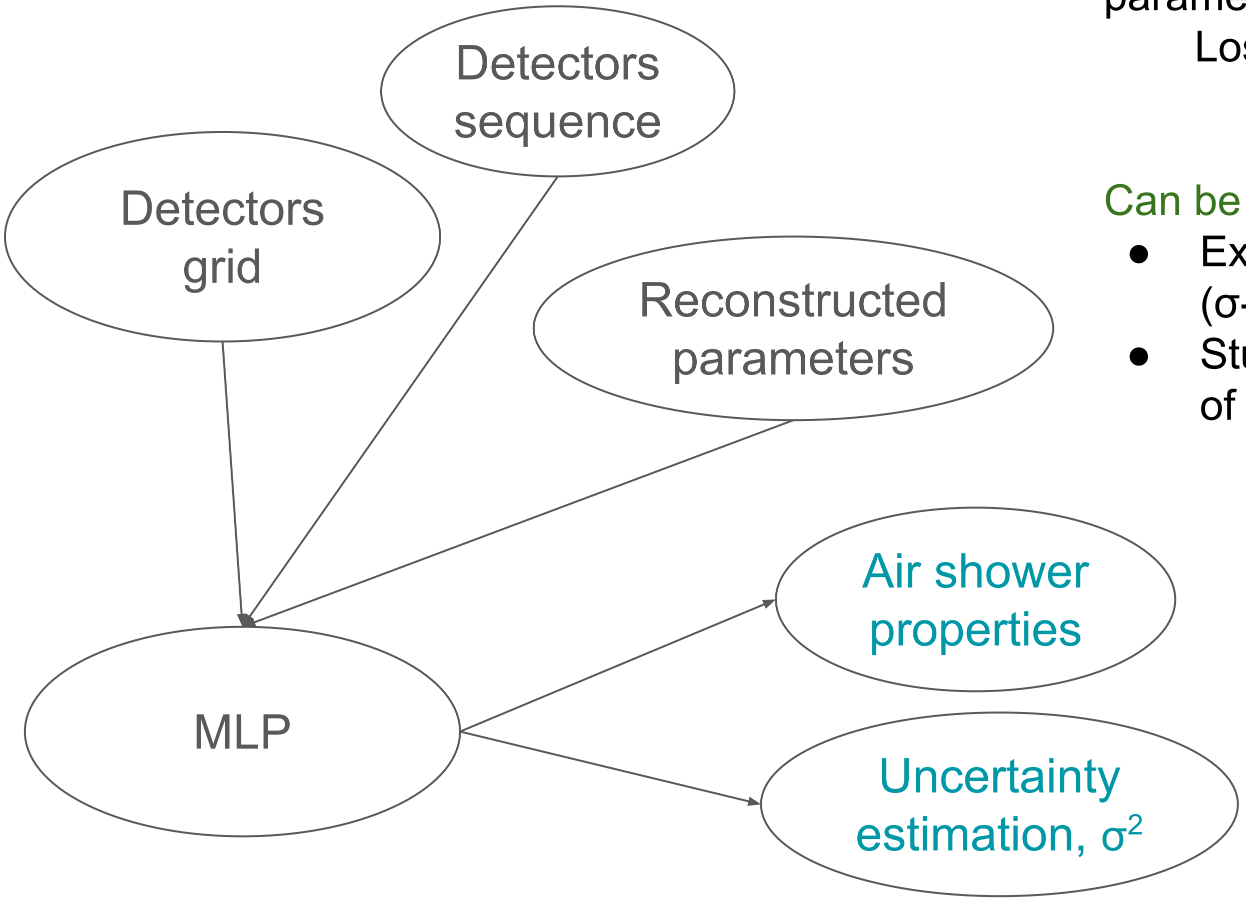
“Native” QGSJET-II-4



EPOS-LHC



Uncertainty estimation



Estimate error of reconstructed parameters:

$$\text{Loss} = L_{reco} + \ln\sigma^2 + L_{reco} / \sigma^2$$

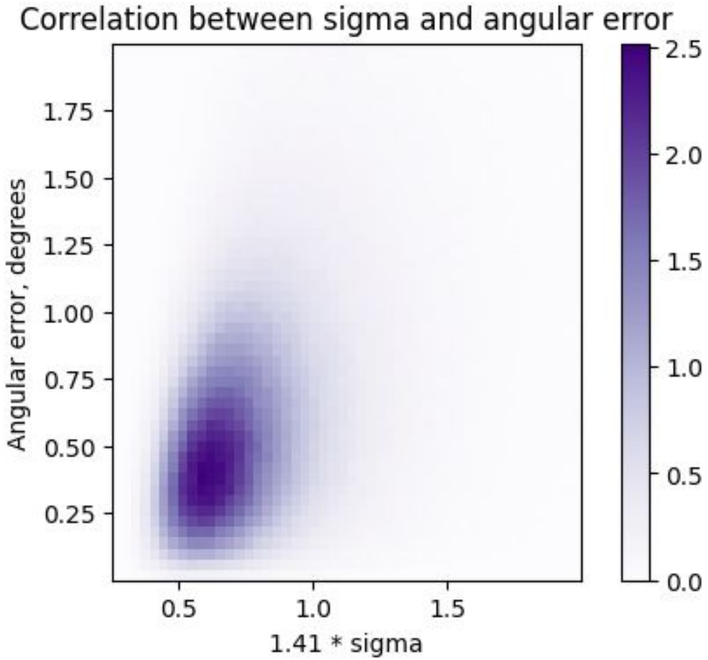
Can be used to:

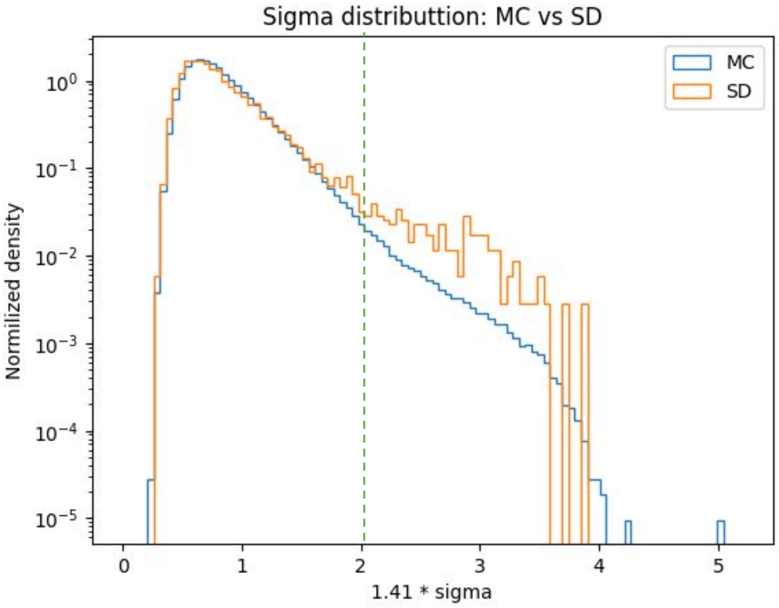
- Exclude events of bad quality (σ-based cuts)
- Study if neural network is unsure of prediction on new data

Uncertainty estimation, direction

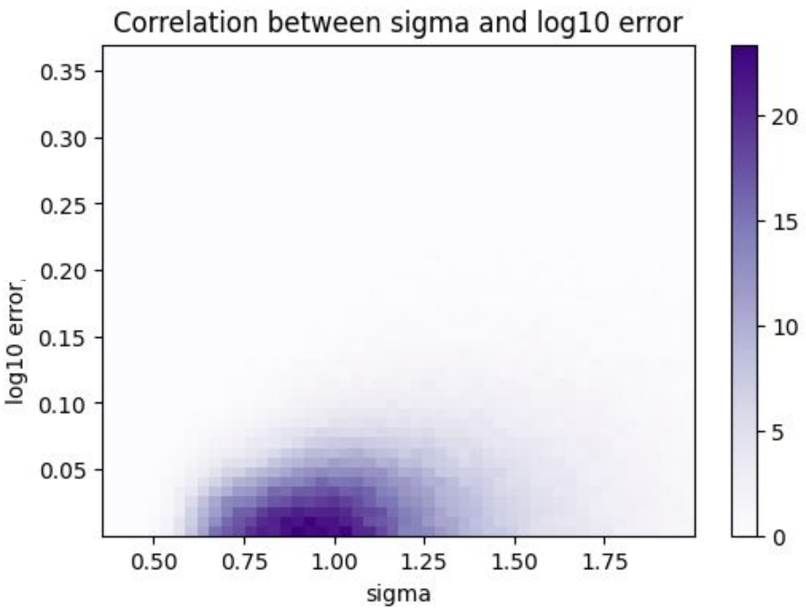
NN correctly identifies events of bad quality

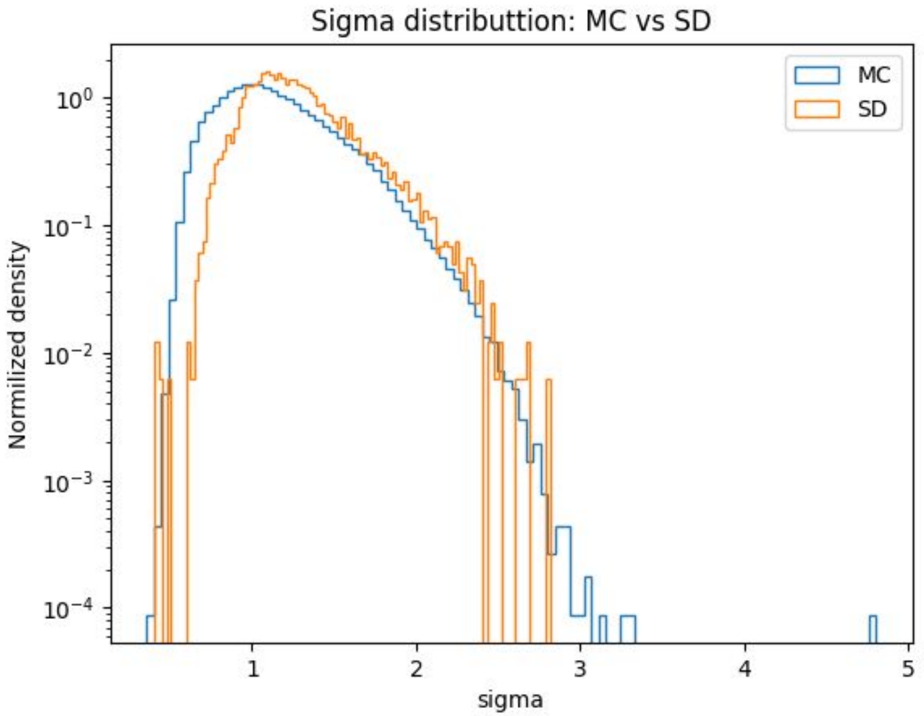
1.41 x sigma cut	0.5	1.0	1.5	2.0	3.0
exposition	0.08	0.77	0.96	0.99	0.999





Uncertainty estimation, energy





2. Towards reliable neural networks

Domain adaptation

Domain adaptation - techniques used to address challenge of training a model on one data distribution and applying it to a related but different data distribution:

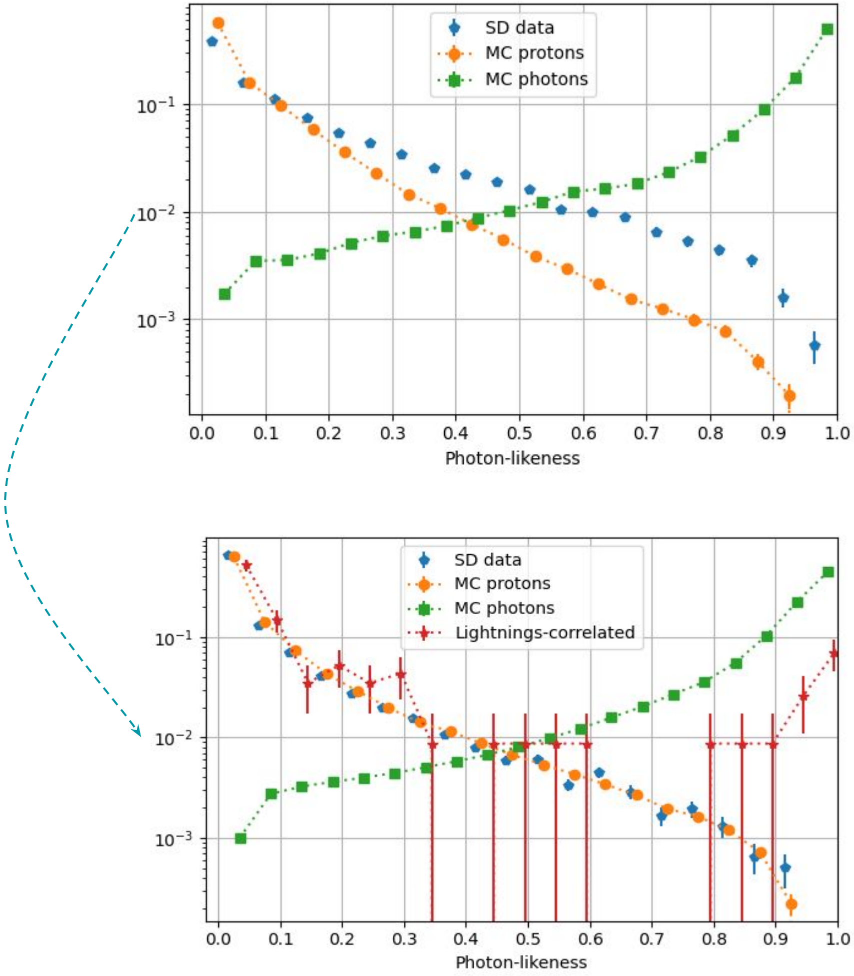
- Train on various source domains (different hadronic models)

- Auto-labeling:
  - Train NN on labeled (for example, MC) data
  - Label target data using NN
  - Fine-train NN on the merged dataset

- Search for common representation space:
  - Force data representations to be the same for source and target distribution

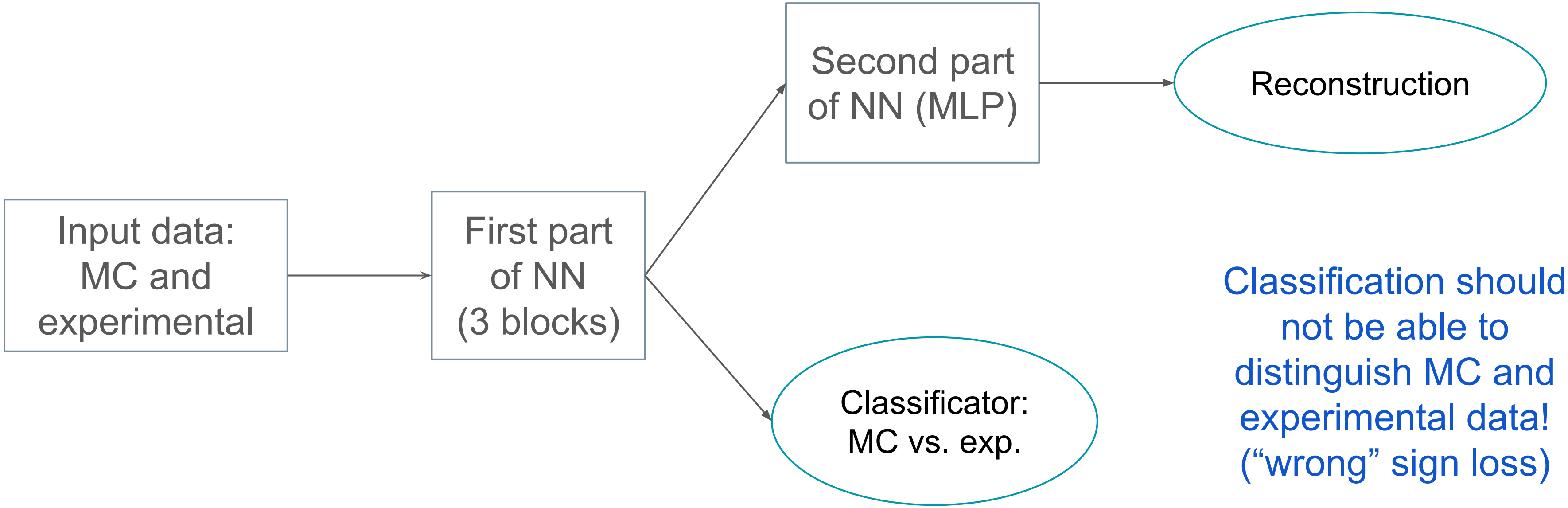
Auto-labeling: photon search

1. Train NN on MC data to distinguish protons and photons
2. Apply NN to experimental data
3. Select events that are classified as protons with high confidence ($\xi < 0.2$)
4. Include these protons to the training data set
5. Fine-train NN on the resulting mix



Enforcing common representation space (project)

One should force NN to disregard differences between various hadronic interaction models and experimental data



Classification should not be able to distinguish MC and experimental data! (“wrong” sign loss)

3. Autoencoder

Autoencoder

Goals:

- Obtain latent (contracted) event representations:
  - Search for anomalies
  - MC - experimental data comparison
  - Additional parameters for NN training

- Model independent event representation

Towards model independence (next step)

Usual AE:

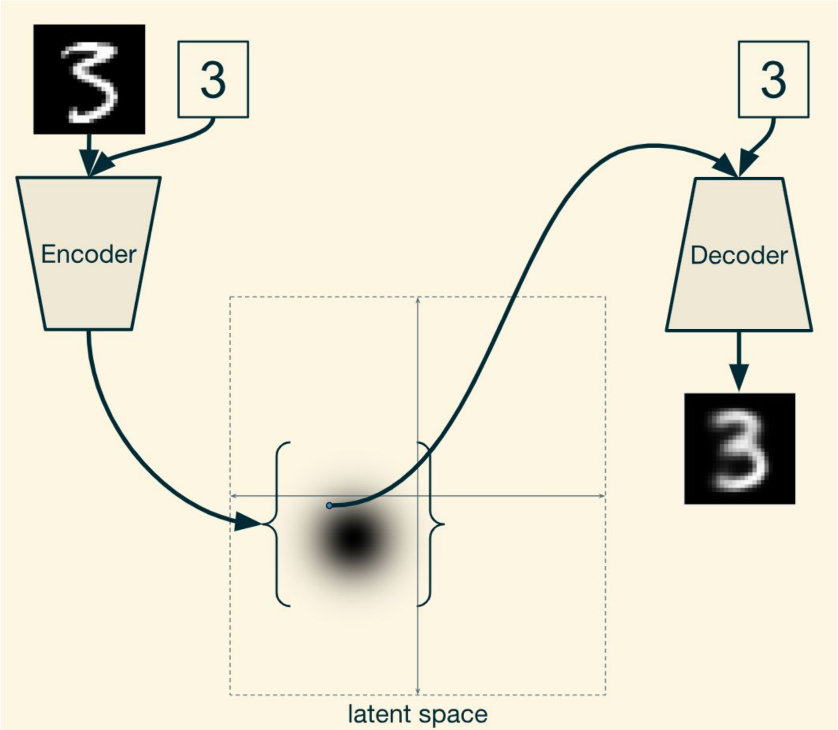
- Input data: images
- Latent space: full information to reproduce data

Conditional AE:

- Input data: image plus label
- Latent space: information to reproduce data (given additional label information)
- Decoder: given point in latent space and label, reconstruct data

Latent space does not store information on the label!

For astrophysics, label = MC model or experimental data



source: ijdykeman.github.io

Conclusion

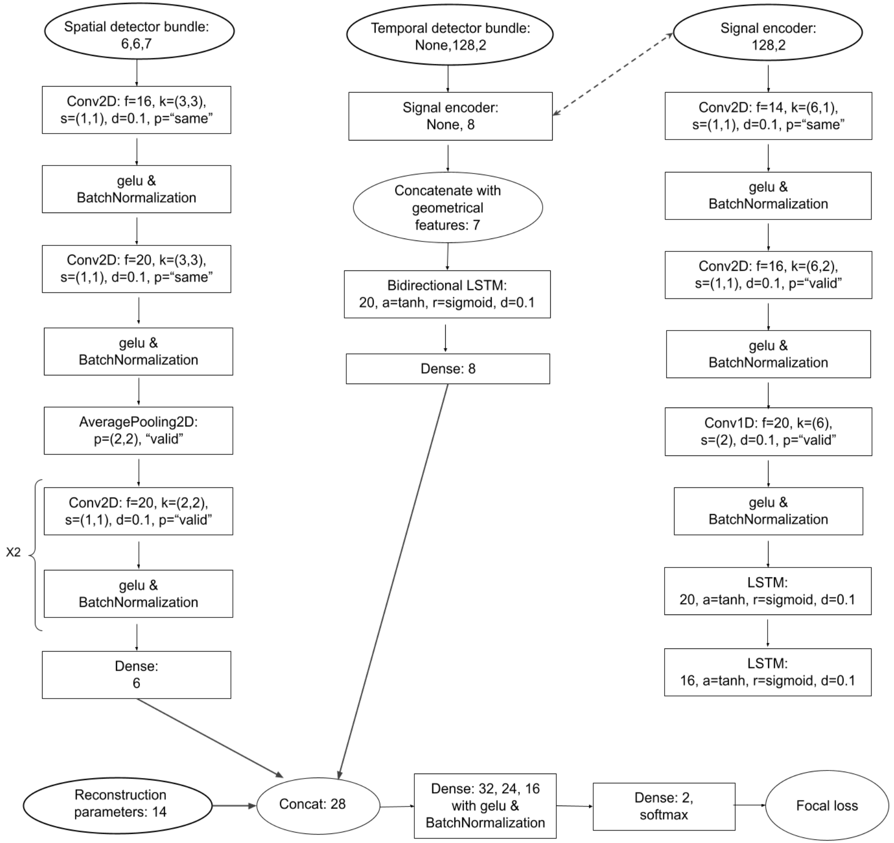
- Main problem: validating NN predictions on experimental data
  - Domain adaptation and CAE as way to solve this
- ML for FD, TALE and others
- Unified framework for data analysis (Anton’s talk)

Backup

Neural network

Neural network's blocks:

- Spatial detectors bundle (geometrical features)
- Temporal detector bundle (overall information)
- Reconstruction parameters (high-level information)



Finding optimal classification threshold

Optimize merit function :

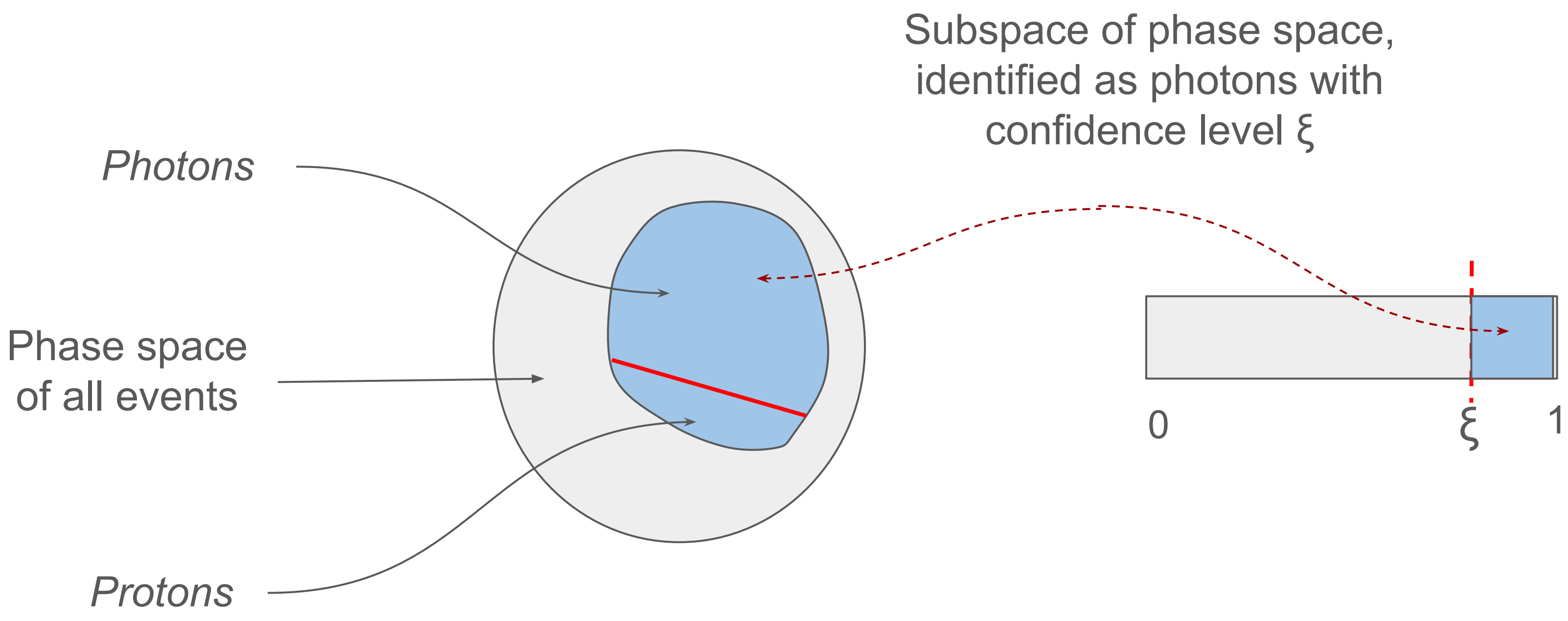
$$L^{95} = \frac{M\left(\sigma^{95}(n_{\text{cand}}(\xi))\right)}{S(\xi)}, \quad S(\xi) = \frac{n_\gamma(\xi)}{n_\gamma^0}$$

L^{95} provides a blind estimation of upper limit on photon flux:

$n_{cand}(\xi)$ - expected number SD events classified as photons
σ^{95} - 95% upper bound on the expected number $n_{cand}(\xi)$
$M(\cdot)$ - expectation value with respect to Poisson distribution

$n_\gamma(\xi)$ - number of photons classified as photons
n_γ^0 - total number of photons

Finding optimal classification threshold



Optimize merit function :

$$L^{95} = \frac{M\left(\sigma^{95}(n_{\text{cand}}(\xi))\right)}{S(\xi)}, \quad S(\xi) = \frac{n_\gamma(\xi)}{n_\gamma^0}$$

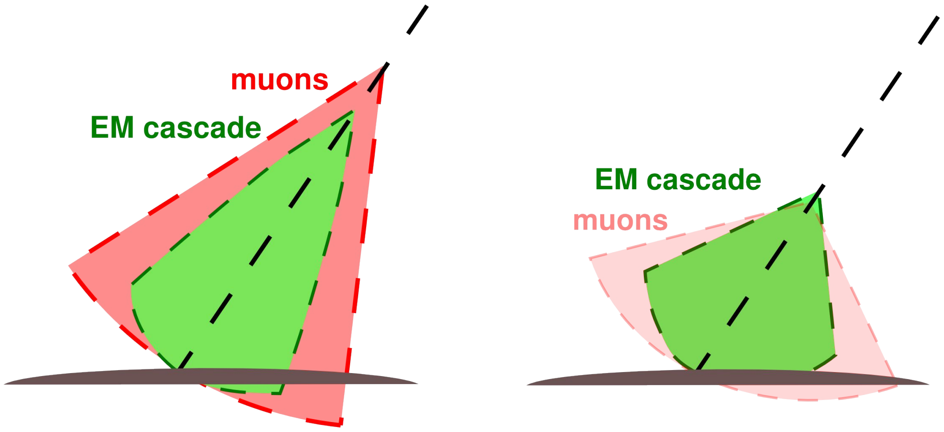
Motivation

Detection of EeV photon-induced air showers would indicate new physics

To search for such events, it is desirable to have as big exposure time as possible

- Surface detectors operate almost at all times.

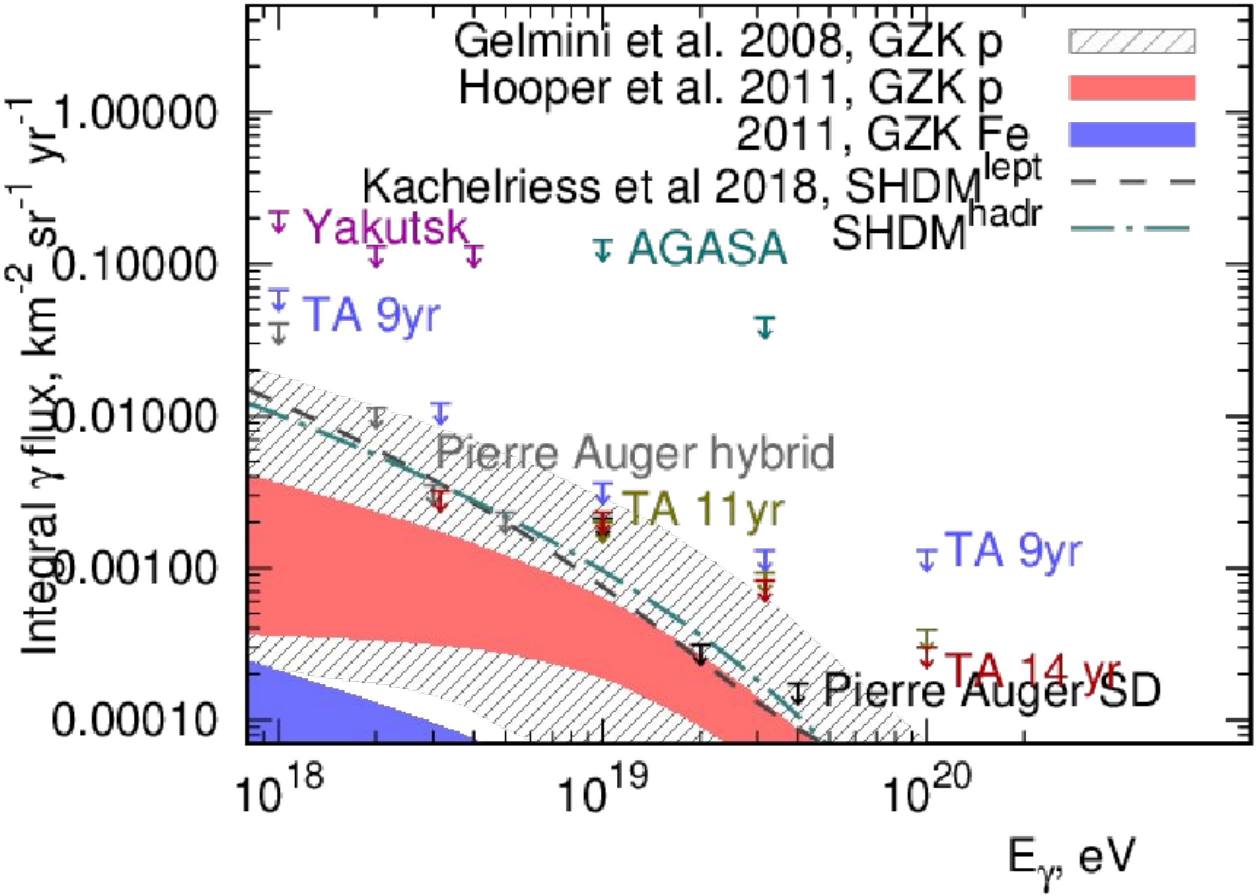
- To get the best separation of photon- and hadron-induced air showers, employ Neural Networks.



hadrons

photons

Results

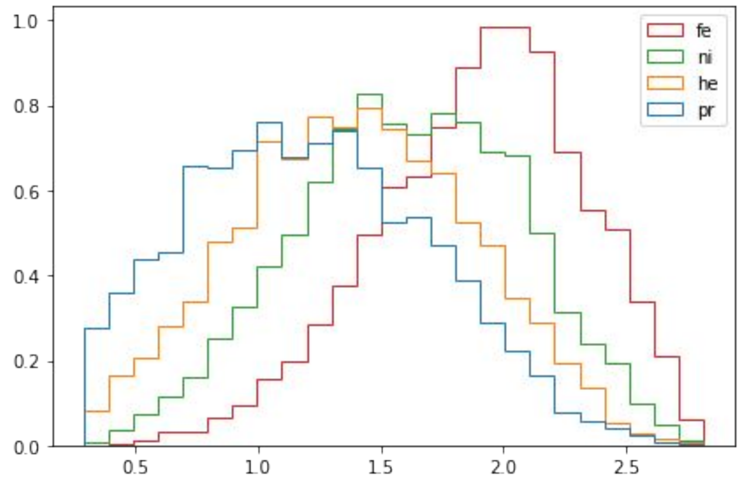


Gelmini et al. 2008, GZK p
Hooper et al. 2011, GZK p
2011, GZK Fe
Kachelriess et al 2018, SHDM^{lept}
SHDM^{hadr}
Yakutsk
AGASA
TA 9yr
Pierre Auger hybrid
TA 11yr
TA 9yr
TA 14 yr
Pierre Auger SD
1.00000
0.10000
0.01000
0.00100
0.00010
Integral γ flux, km^{-2} sr^{-1} yr^{-1}
10^{18}
10^{19}
10^{20}
E_γ, eV

4. Mass composition analysis

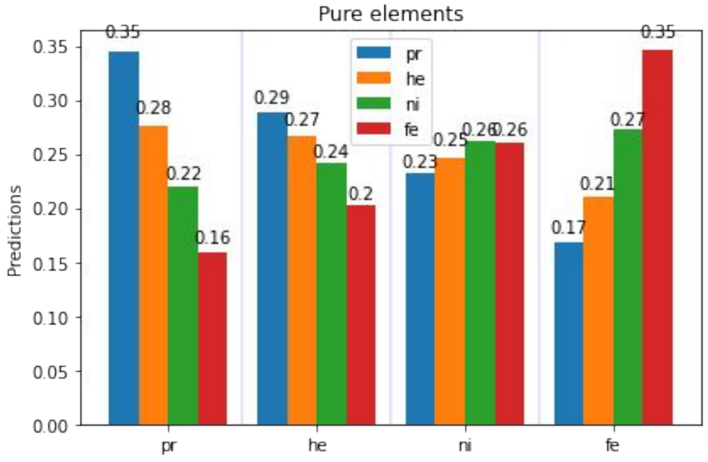
Mass composition analysis

Evolution of air showers is stochastic.
Data may be similar for different primaries



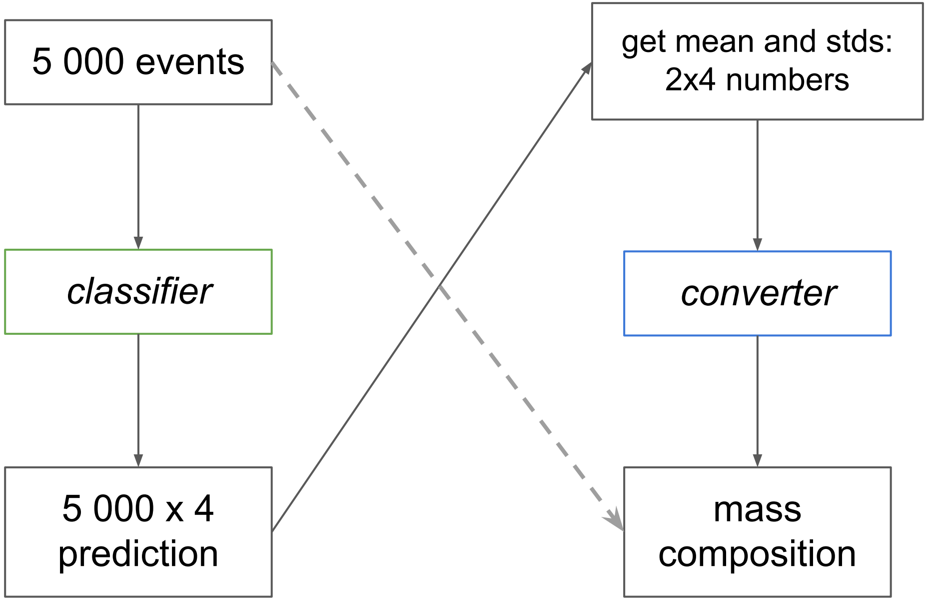
~39% success in 4-class model

Can we do better on ensembles of events?



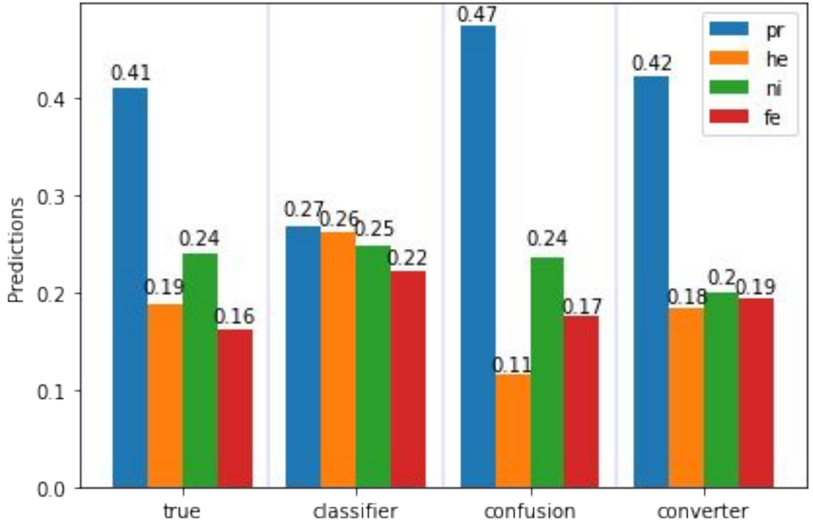
Making use of statistics

We are interested in obtaining mass composition of an ensemble of events!



Converter is the second neural network, which improves *classifier* predictions for ensembles of events

Making use of statistics

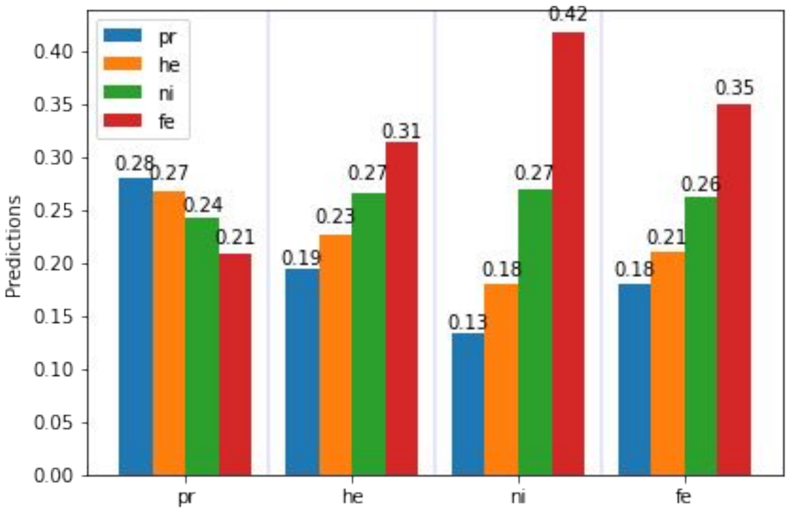


	proton	helium	nitrogen	iron
classifier	0.1	0.14	0.12	0.09
converter	0.03	0.07	0.06	0.02

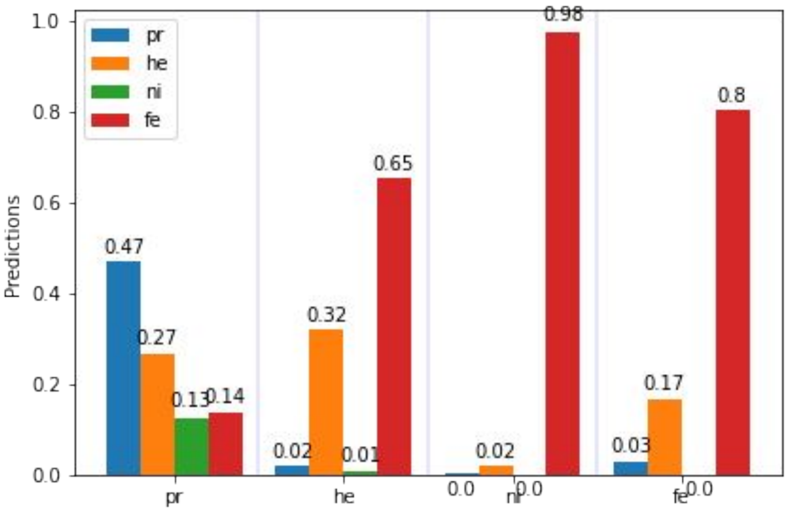
Error: mean absolute error (averaging over events) on 2000 ensembles

Model dependence

Neural network, trained on QGSJET II-03, observing events generated with QGSJET II-04:



Classifier predictions



Converter predictions

High systematic error: up to 100%

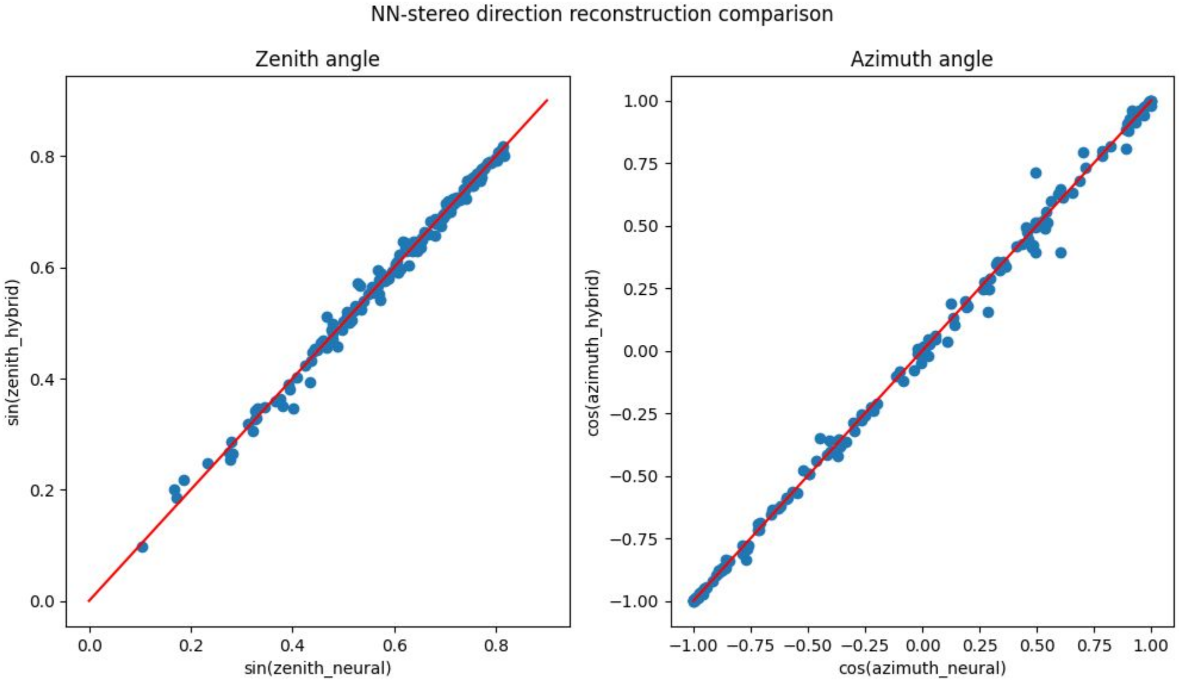
Comparison with stereo data

File from “Stereo from Stroman” page, 167 common events

NN - stereo:
<Δn> : 1.15°
<ΔΘ> : 0.05°
<Δφ> : 0.02°

NN - SD:
<Δn> : 0.73°
<ΔΘ> : 0.12°
<Δφ> : 0.13°

SD - stereo:
<Δn> : 1.20°



Comparison with hybrid data

File from “User: Whanlon” page, 9.5 years; 149 common events

NN - hybrid:
<Δn> : 0.95°
<ΔΘ> : 0.33°
<Δφ> : -0.16°

NN - SD:
<Δn> : 0.93°
<ΔΘ> : 0.24°
<Δφ> : -0.10°

SD - hybrid:
<Δn> : 1.0°