Deep learning at ECAP

Dmitry Malyshev
Gisela Anton, Thomas Eberl, Stefan Funk,
Thorsten Glüsenkamp, Thilo Michel

Erlangen Center for Astroparticle physics

ErUM-Data collaboration meeting
Sept 30 – Oct 1, Karlsruhe

Topics

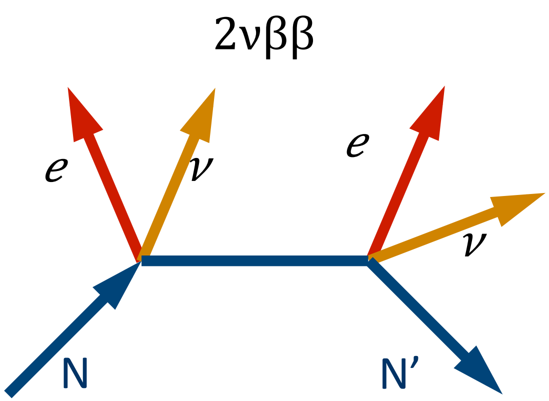
- Deep learning at ECAP
  - Regular informal seminars
    - Presentations by different groups (both neutrino and gamma rays)
    - Exchange of ideas and discussion of technical problems
  - Master level lectures and practical exercises on machine learning (including deep learning)
- Neutrinos
  - EXO-200
  - IceCube
  - KM3NeT
- Gamma rays
  - Fermi
  - H.E.S.S.

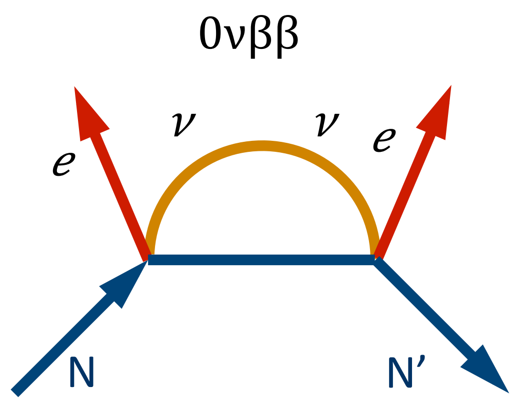
Deep learning at ECAP

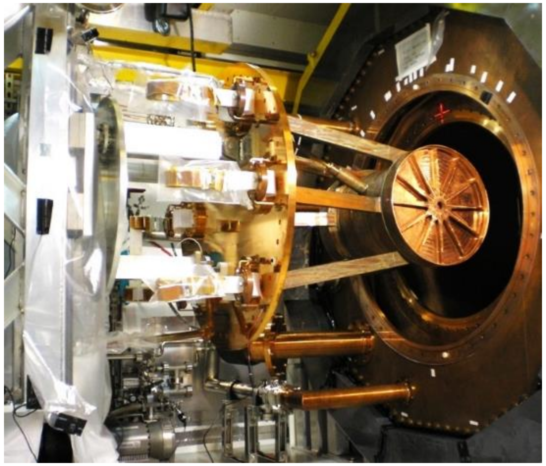
- Current research topics at ECAP
  - Classification
    - Background vs signal
    - Different types of astro sources
  - Regression
    - Reconstruction of energy, direction of astro particls
  - Monte Carlo (with GANs)
- Can DL algorithms do as well as or better than the likelihood methods?
  - Yes, they can and with much less “tuning” by hand.
- New questions (shift of paradigm?):
  - Before:
    - What are important features of the measurement?
    - How does one use them to measure physical quantities?
  - Now:
    - What is the optimal architecture of the networks to measure the physical quantities?

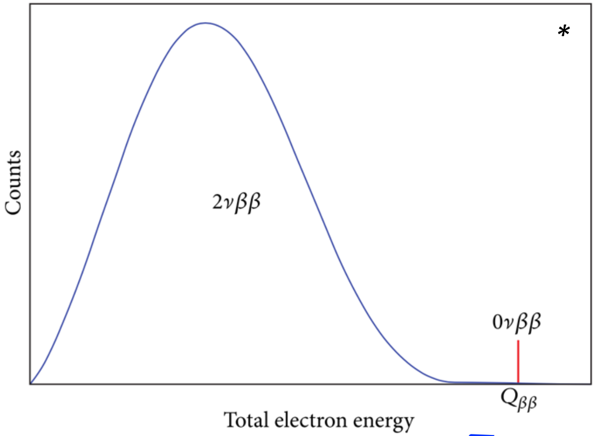
EXO-200 experiment

- Neutrinoless double beta decay
- **Tobias Ziegler**, Thilo Michel, Gisela Anton









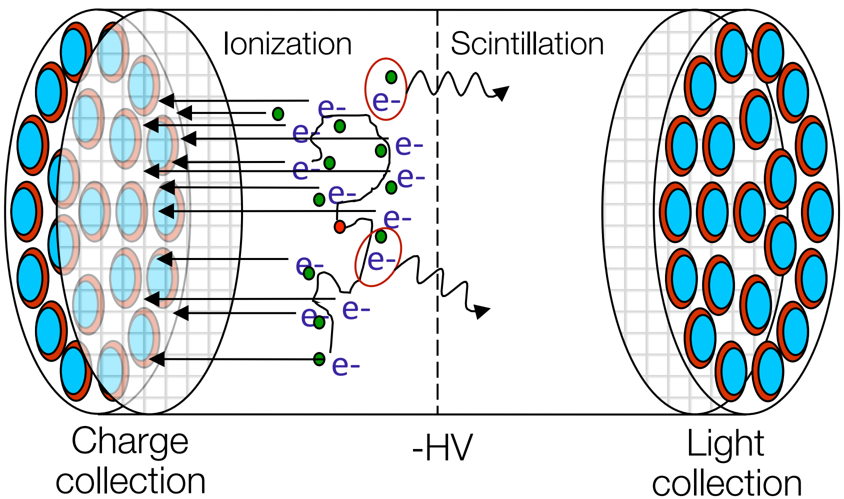
2νββ
- Standard model process

0νββ
- Majorano neutrino
- Lepton number violation

Need a very precise energy measurement

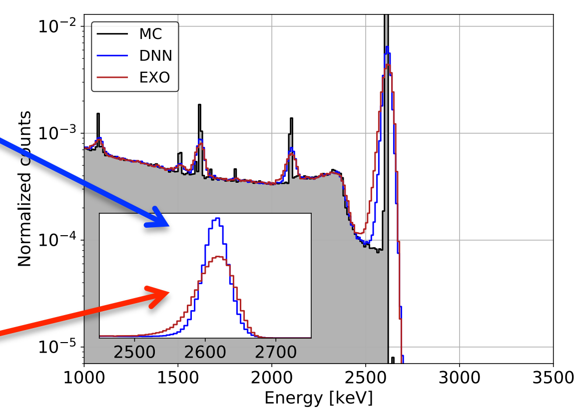
EXO-200: Energy measurement

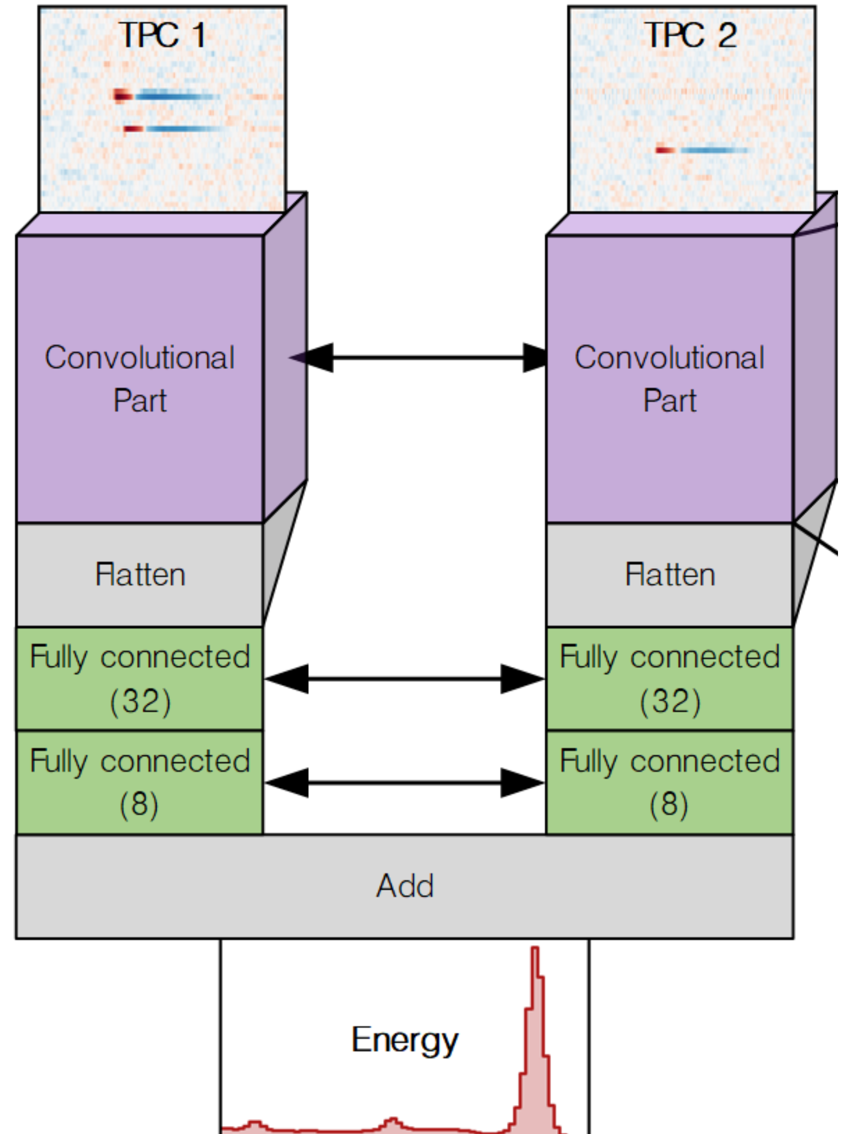
- Energy resolution with deep learning is better than standard methods
  - At the moment the comparison is made for ionization only
  - Future: add scintillation



Deep learning

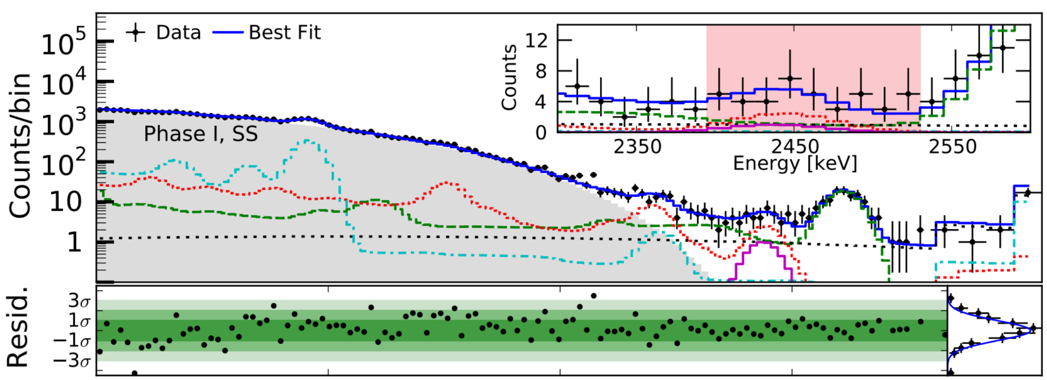
Standard analysis



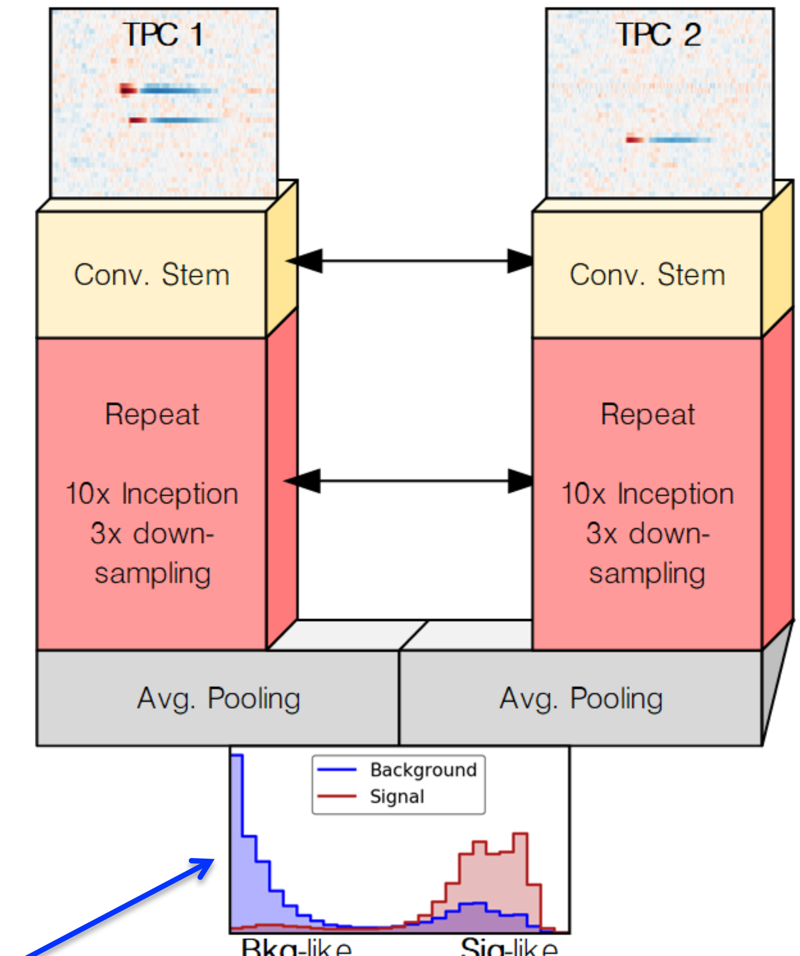


EXO-200: background rejection

- One of the most sensitive searches for $0\nu\beta\beta$ with the full EXO-200 dataset for ^{136}Xe and the first search directly using a DNN discriminator (arXiv: 1906.02723, submitted to PRL).
- Architecture inspired by Inception proposed by Google (kernels of different size at the same level)
- DNN discriminator outperforms BDT discriminator (Phys.Rev.Lett. 120 (2018))



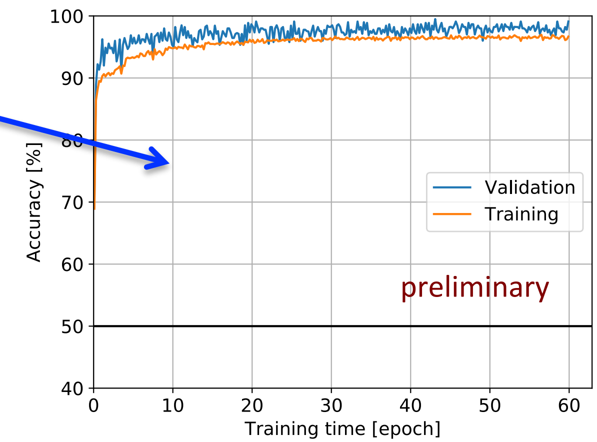
Anton et al, arxiv:1906.02723



Main background – gamma rays

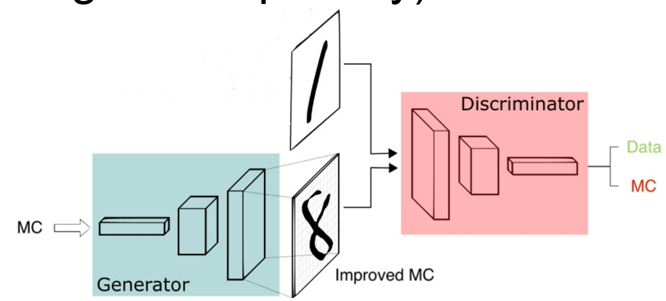
EXO-200: MC simulations with GANs

- **Federico Bontempo**, Thilo Michel, Gisela Anton
- MC simulations can be easily distinguished from real data images by a discriminator NN
  - Can one use Generative Adversarial Networks (GANs) to improve MC generator?

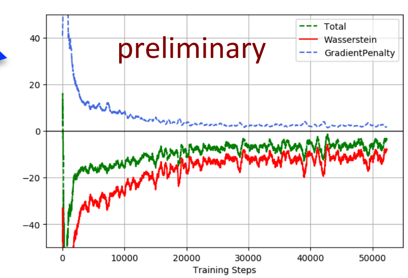


- Use Wasserstein distance in training the generator to have only minimal changes necessary to make MC images realistic.
  - Loss function for discriminator training (with gradient penalty):

$$L = \underbrace{\underset{\tilde{\boldsymbol{x}} \sim \mathbb{P}_g}{\mathbb{E}} \left[D(\tilde{\boldsymbol{x}}) \right] - \underset{\boldsymbol{x} \sim \mathbb{P}_r}{\mathbb{E}} \left[D(\boldsymbol{x}) \right]}_{\text{Original critic loss}} + \underbrace{\lambda \underset{\hat{\boldsymbol{x}} \sim \mathbb{P}_{\hat{\boldsymbol{x}}}}{\mathbb{E}} \left[\left(\| \nabla_{\hat{\boldsymbol{x}}} D(\hat{\boldsymbol{x}}) \|_2 - 1 \right)^2 \right]}_{\text{Our gradient penalty}} .$$



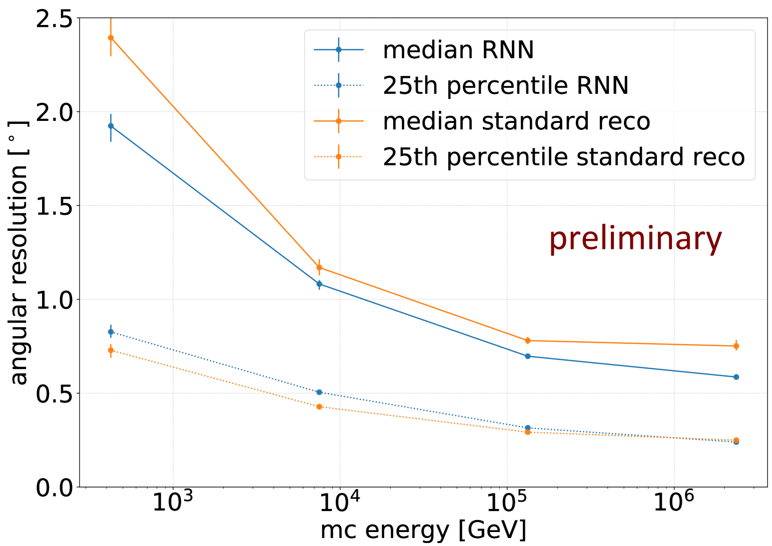
  - Generator + discriminator training is converging

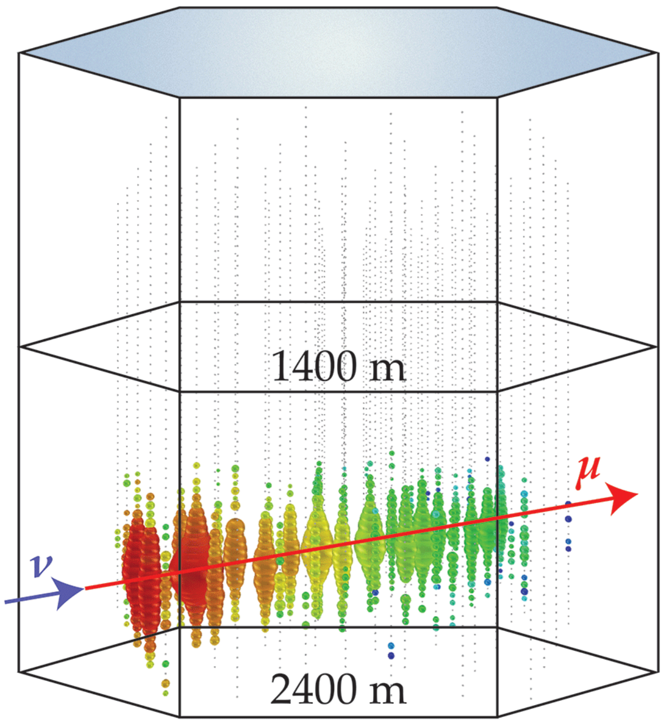


- Work in progress:
  - Comparison with physical parameters of events (x, y, z, E)

IceCube experiment

- **Gerrit Wrede**, Thorsten Glüsenkamp, Gisela Anton
- Neutrino converts to charged particles, e.g., muons
- Need to measure the direction and energy of the muons
- Data – light deposited in PMTs in the ice as a function of time
- Use Recurrent Neural Network (RNN) with Long short-term memory (LSTM) layers





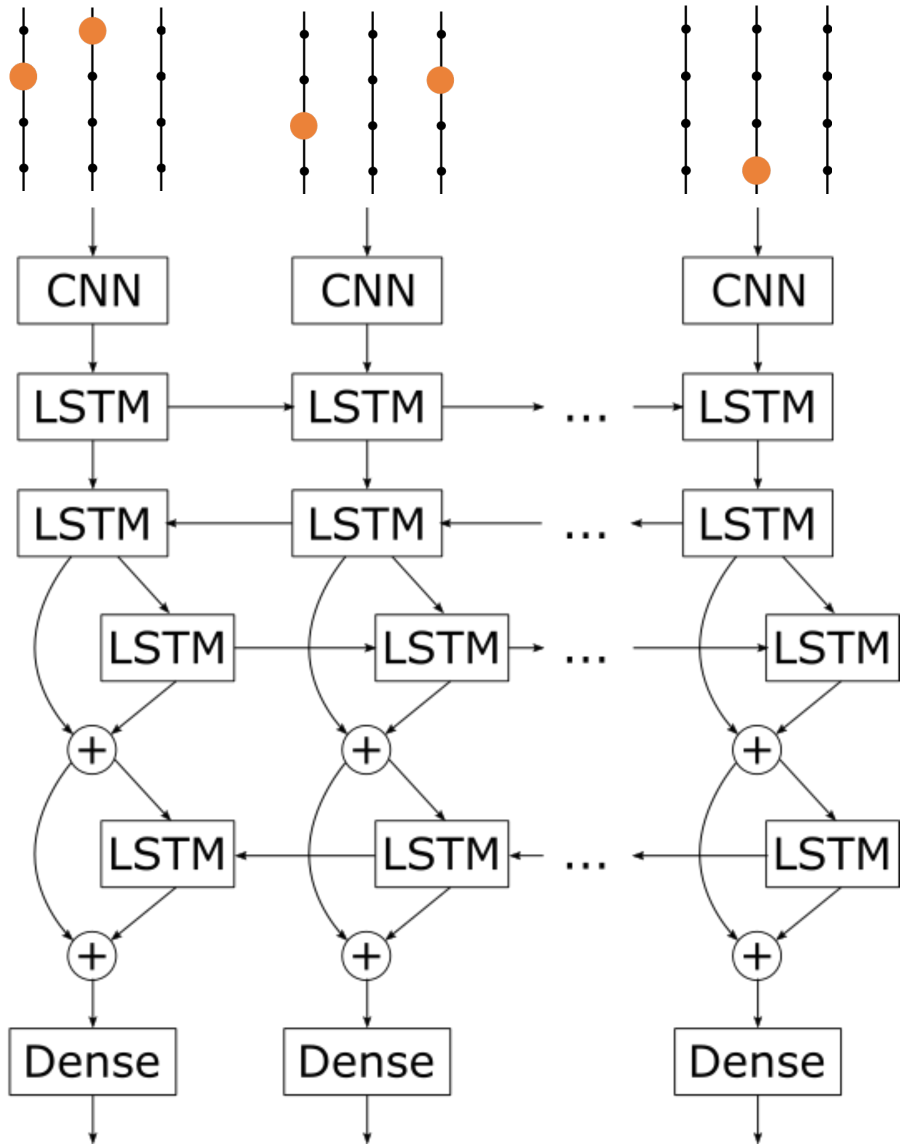
	reconstruction time
RNN	$\approx$ 60 ms
standard reco	$\approx$ 1 s
standard reco + seeds	$\approx$ 1 min

IceCube: Position Reconstruction

- Input: Light detected in PMTs binned in time (30ns bins)
- 3 1D-Convolutional layers
  - along z-axis
- 1 forward LSTM layers
- 1 reverse LSTM layers
- 2 residual LSTM layers
- Predict position of muon in each time step
- Loss function:

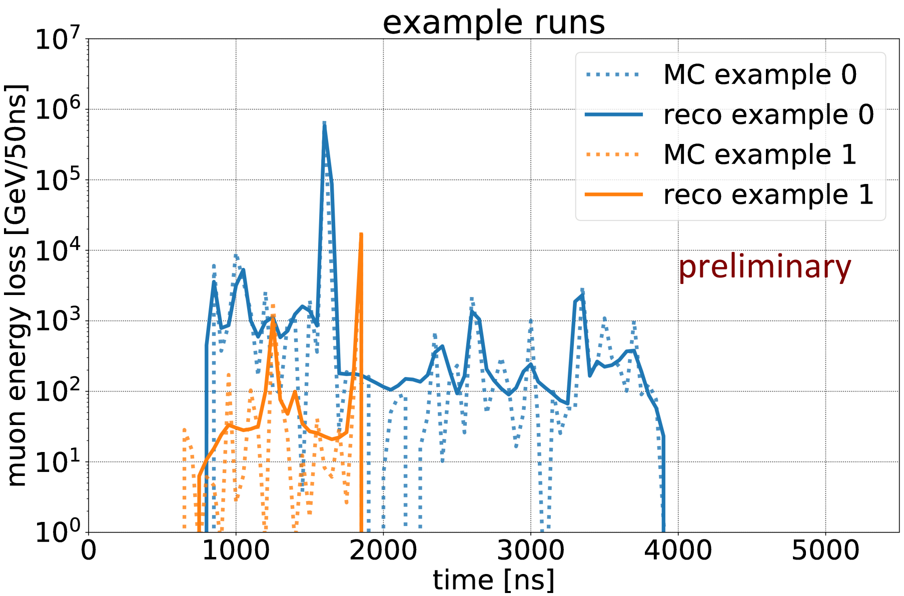
$$L = \sum_i \ln \sigma_i^2 + \frac{(\mu_i - x_i)^2}{\sigma_i^2}$$

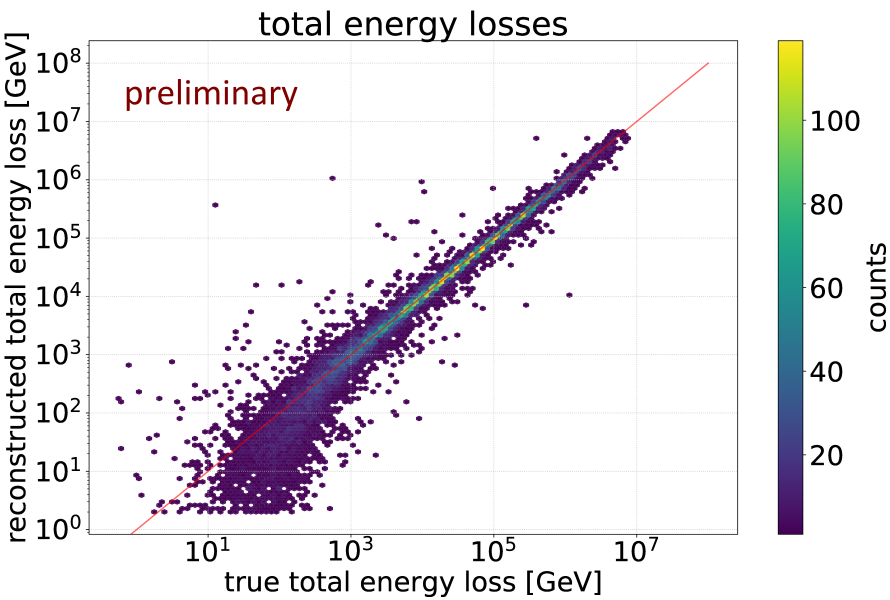
- For the final track reconstruction use the linear fit through the DL position estimates at each time step



IceCube: Energy Reconstruction

- Reconstruct stochastic energy losses of muons
- RNN with similar architecture



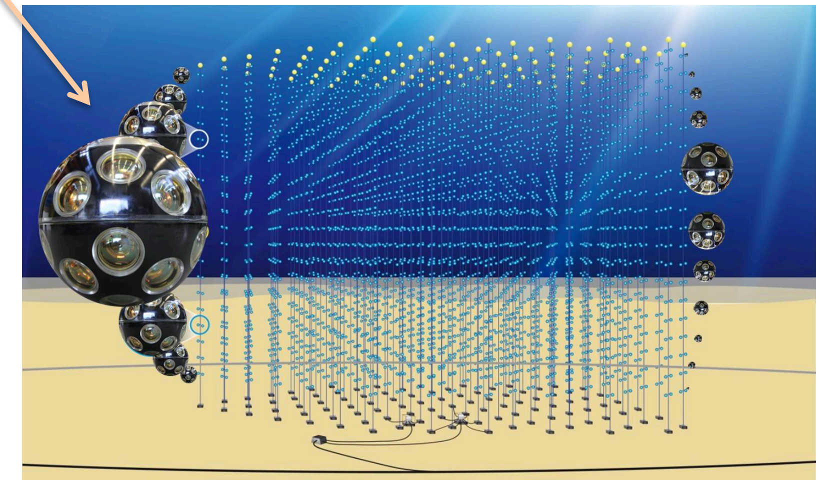


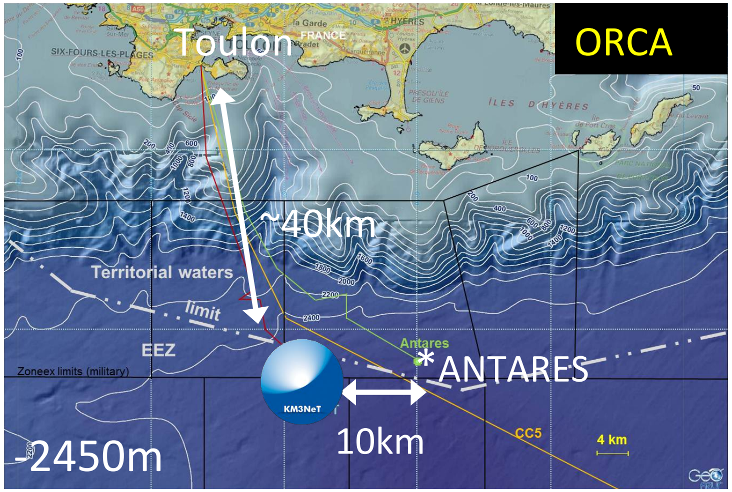
- Next steps:
  - Background rejection with RNNs

KM3NeT

- **ARCA: A**stroparticle **R**esearch with **C**osmics in the **A**byss
  - TeV-PeV neutrino astronomy
- **ORCA: O**scillation **R**esearch with **C**osmics in the **A**byss
  - neutrino mass ordering with few-GeV atmospheric neutrinos

Digital Optical
Module (DOM)





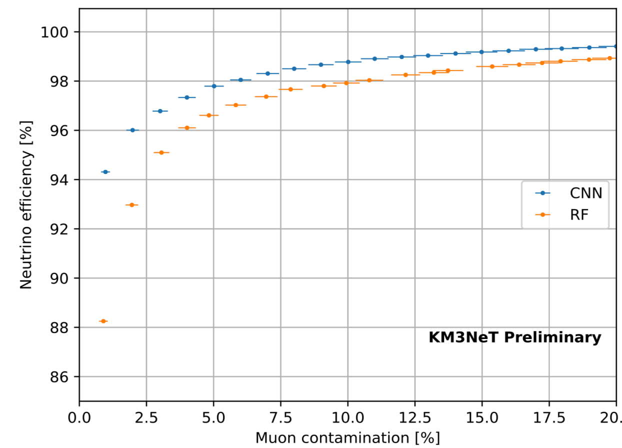
31 PMTs:
19 up, 12 down

KM3NeT Deep Neural Networks

Michael Moser, Steffen Hallmann, Stefan Reck, Thomas Eberl, Gisela Anton

Classification

- Background suppression
- Event topology identification (flavor proxy)



KM3NeT collab. PoS(ICRC2019)904

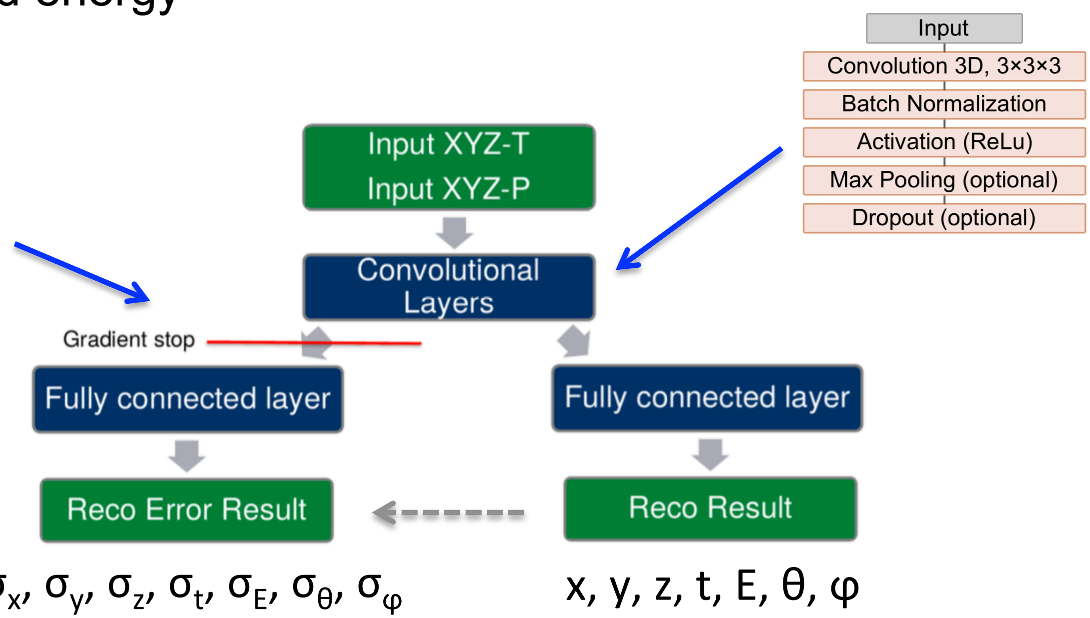
Regression

- Neutrino direction and energy reconstruction
- Uncertainty estimate

Add a neuron per reconstructed regression variable to predict absolute residual of the mean error.

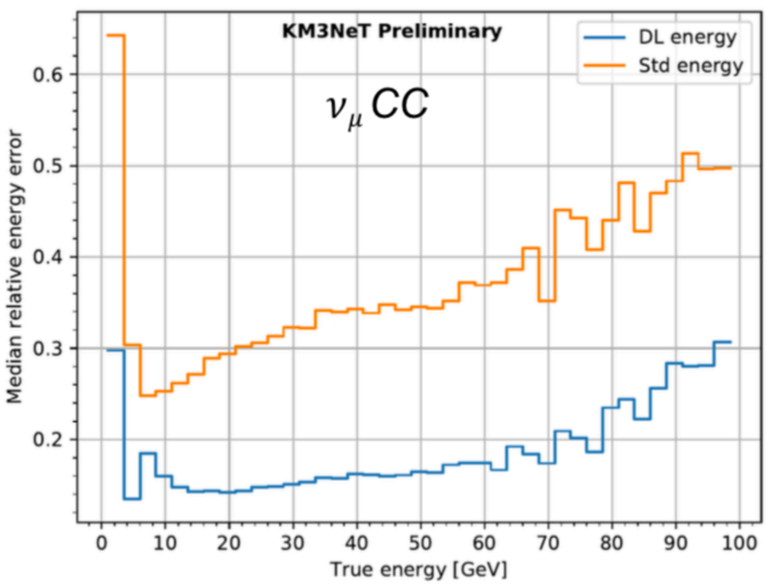
Loss function:

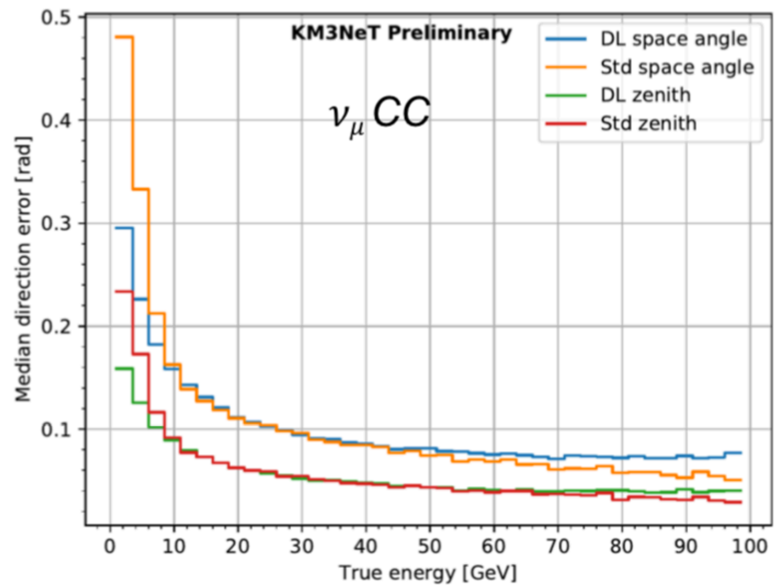
$$L = \frac{1}{n}\sum_{i=1}^{n}\left(\sigma_{\mathrm{reco}} - |y_{\mathrm{true}} - y_{\mathrm{reco}}|\right)^2$$



KM3NeT: Energy and Direction

- Energy and direction for track events
  - Energy error is much smaller for DL than the standard methods
  - Direction reconstruction is comparable for DL and the standard methods



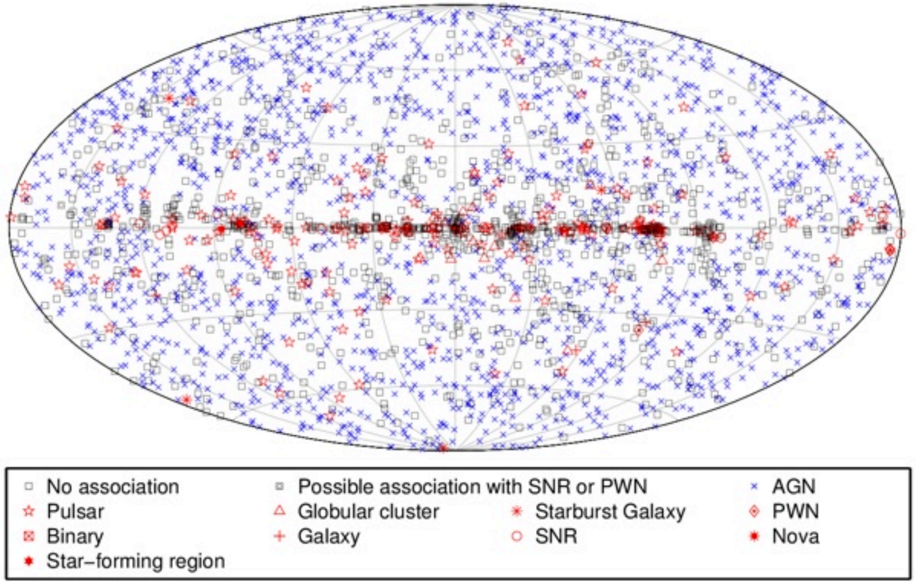


Michael Moser et al. TAUP 2019 poster

- Next steps:
  - Regression and classification with CNNs for ARCA and ANTARES
  - Use of autoencoders to train directly on data
    - master thesis of Stefan Reck

Fermi LAT: point source classification

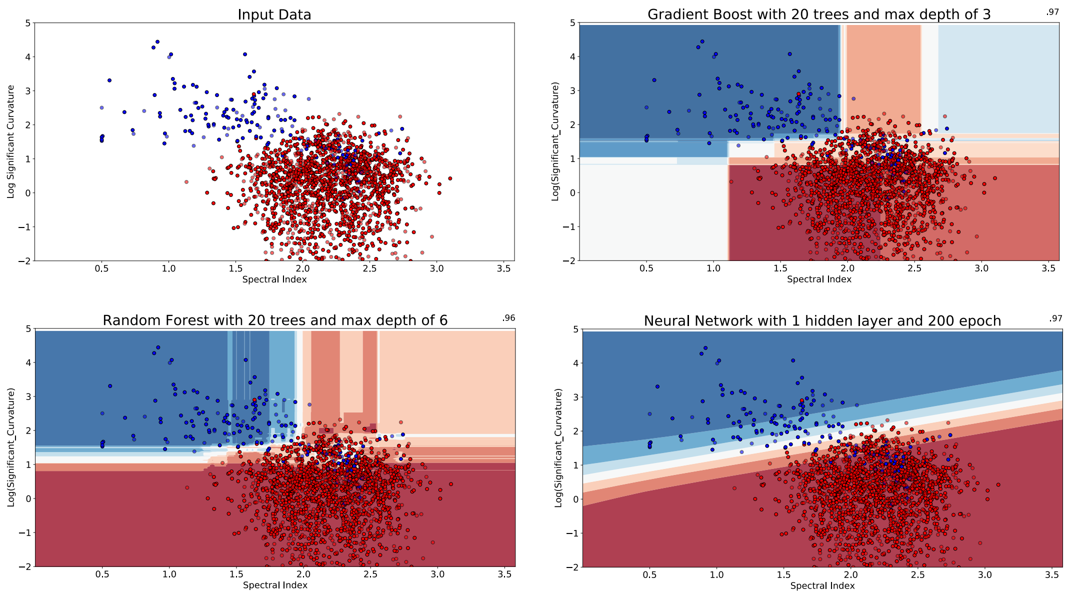
- Aakash Bhat, Dmitry Malyshev
- There are unidentified sources in the Fermi LAT catalogs
- Propose classification of unidentified sources into active galactic nuclei (AGNs) and pulsars based on their properties:
  - Features: position, spectral index, cutoff, variability
- Use ML algorithms:
  - Random forests
  - Boosted decision trees
  - Neural networks
  - Logistic regression
  - Support vector machines
  - …
- Future directions:
  - Classification based on raw photon data



Acero et al. ApJS 218 (2015)

Fermi LAT: PS classification

- Accuracy of classification as a function of the complexity, e.g., depth of the trees in RF, number of layers in NN
- Examples of domains for a pair of parameters
- Result of the classification for unidentified sources in 3FGL catalog (4 years of data)
- Goals:
  - Compare with the new 4FGL catalog (8 years of data)
  - Make predictions for the unidentified sources in 4FGL
  - Extend classification for subclasses
- Future direction:
  - PS reconstruction with CNNs
  - PS classification using light curve with RNN

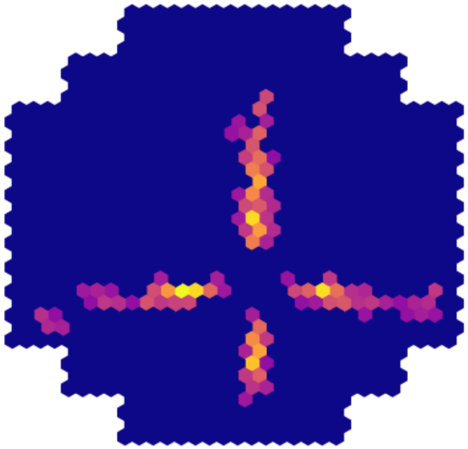


Algorithm	Accuracy	Expected new PSRs
RF	98.3	148/1056
BDT	96.9	184/1056
NN	98.1	213/1056

H.E.S.S. experiment

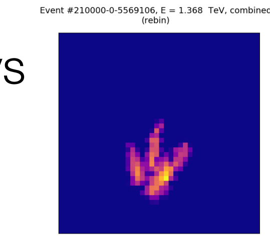
- Christina Hillig, Matthias Buchele, Stefan Funk
- IACT – detection of high energy gamma rays (E > 50 GeV)
- Neural Networks
  - Event classification
    - Gamma rays vs nuclei
  - Energy reconstruction
- Input: images of Cherenkov light from atmospheric showers
  - Used the data from the four small telescopes in this analysis

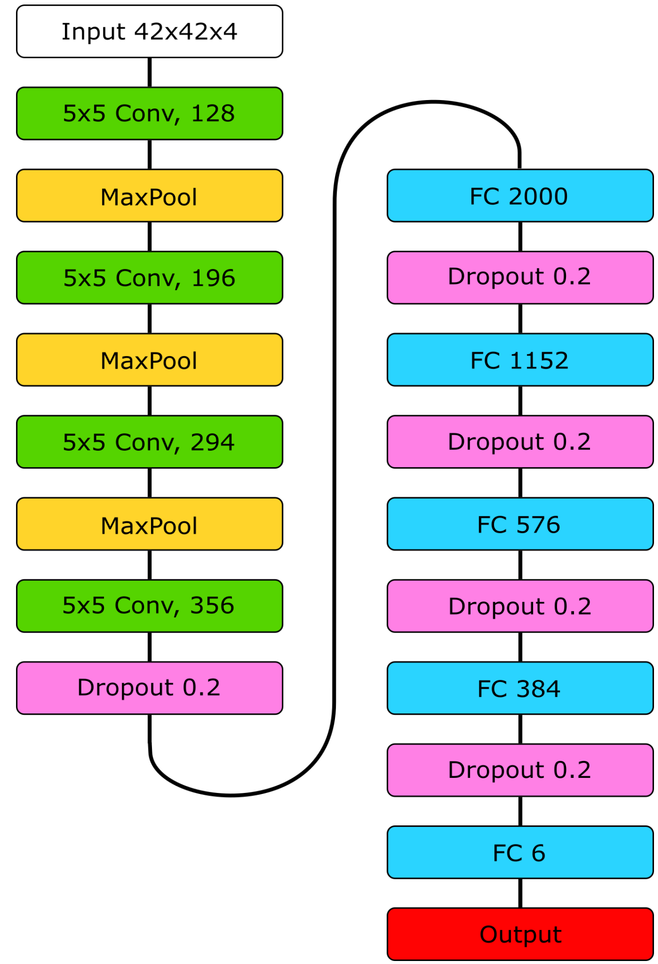
CT1-4 Combined Image

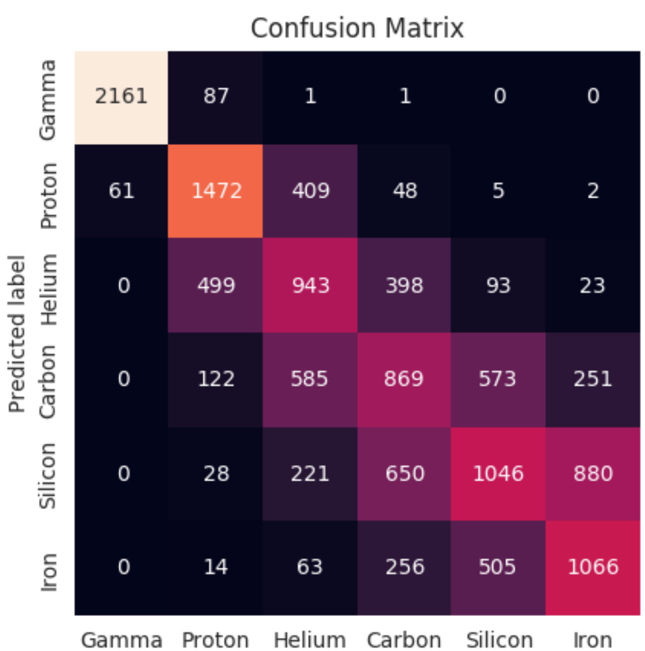


H.E.S.S.: Event Classification

- A reasonable separation (~ 1/200) of gamma rays from nuclei was achieved
- One can also use CNNs to separate light nuclei (e.g., Hydrogen or Helium) from heavy ones (e.g., Iron)



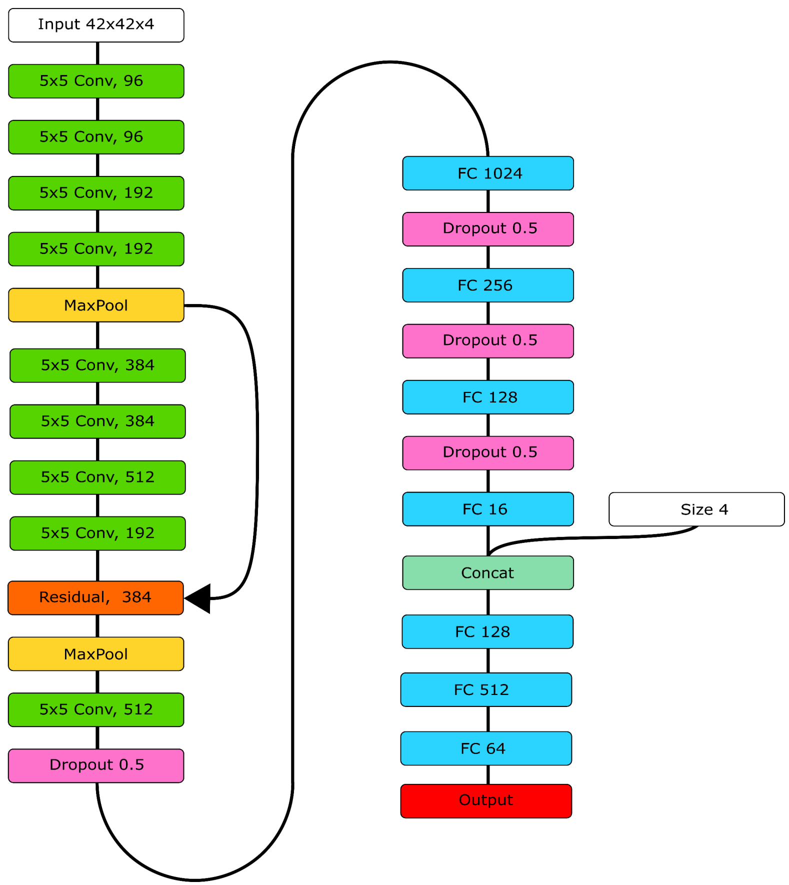


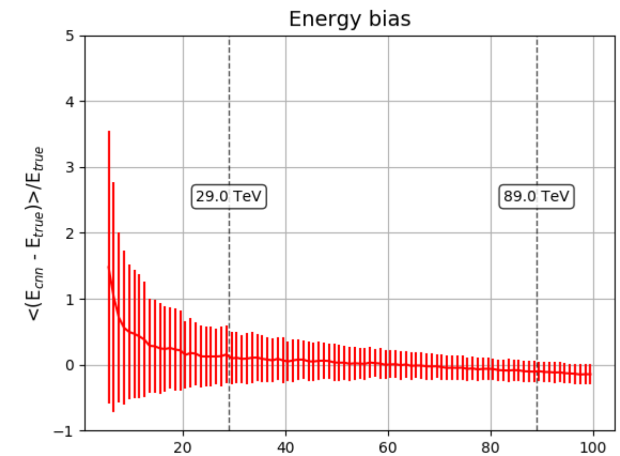


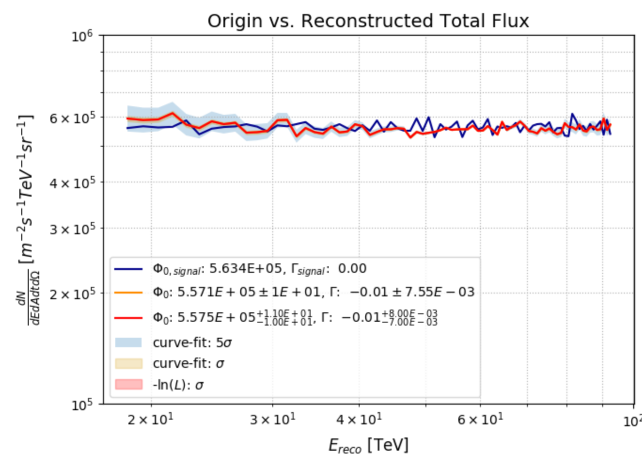
Christina Hillig, Master thesis

H.E.S.S.: Energy Reconstruction

- Use a similar network to reconstruct the energy
  - Results are reasonable between ~ 30 – 90 TeV







Christina Hillig, Master thesis

Conclusions

- DL methods have been used in ECAP in analysis of EXO-200, IceCube, KM3NeT, Fermi-LAT, H.E.S.S. data for
  - Background rejection
  - Energy reconstruction
  - Direction reconstruction
- The DL methods work similar to or better than the “standard methods” such as likelihood analysis or other machine learning technics, e.g., BDT
- Advantages:
  - Little additional input
    - in many cases the input is the raw data such as shower images, electrical current, PMT responses
    - the DL algorithms learn to determine the most important features “by themselves” – no need to guess or construct the features
    - this typically leads to much shorter overall development time, although the algorithm training time can be several days
  - Data analysis time is often shorter than for standard methods
- Challenges:
  - No standard procedure – how does one choses a particular network, can another network be better?
  - Sometimes interpretation of results in terms of, e.g., statistical uncertainties or confidence is not straightforward