AI goes MAD

Deep learning imaging atmospheric Cherenkov telescopes

D. Nieto (d.nieto@ucm.es)

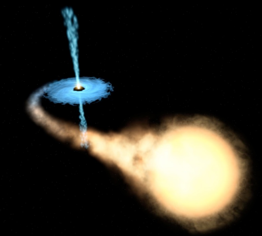
Facultad de Ciencias Físicas

UCM - IPARCOS

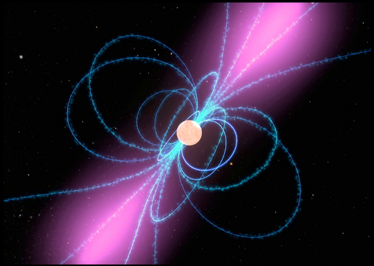
https://apod.nasa.gov/apod/ap200724.html

- Very-high-energy astrophysics in a (very-small) nutshell
- Imaging atmospheric Cherenkov telescopes
- Enhancing IACTs with deep learning

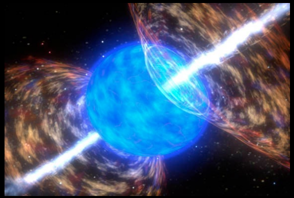
Conventional non-thermal emitters & CR accelerators



Gamma-ray Binaries

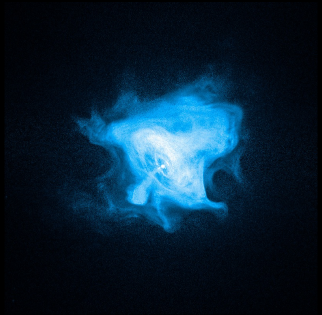


Pulsars

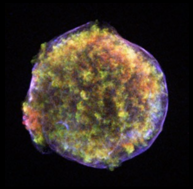


Gamma-ray Bursts

Compact-object mergers



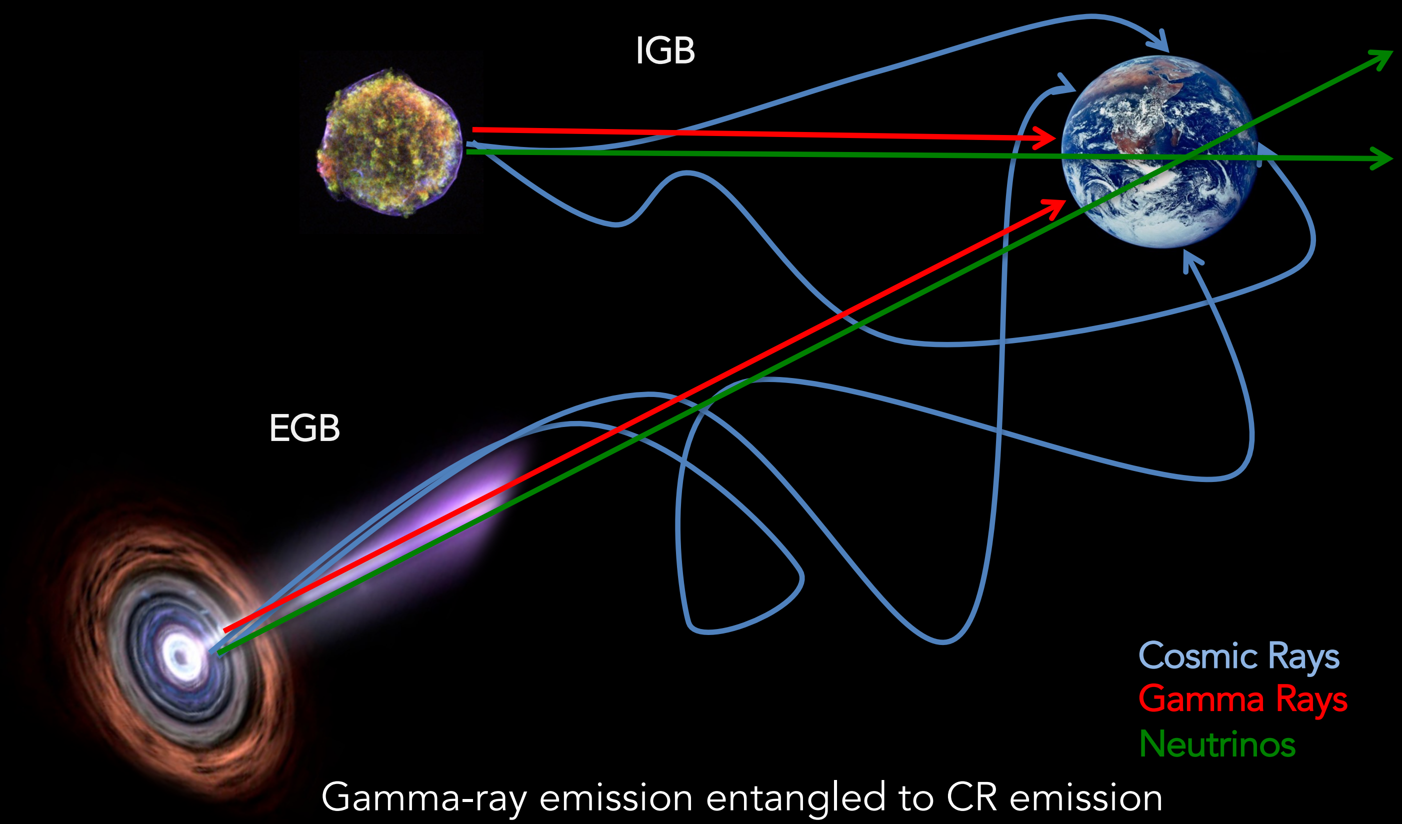
Pulsar Wind Nebulae



Supernova Remnants

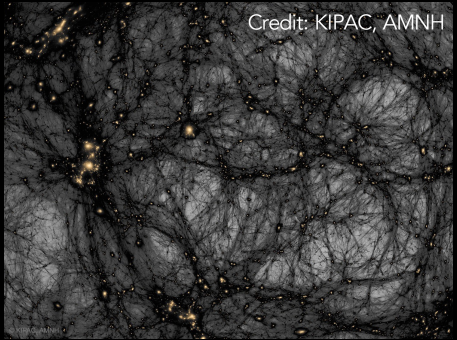
Starburst Galaxies

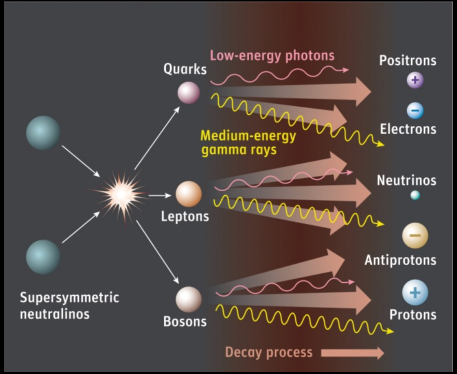
Active Galactic Nuclei



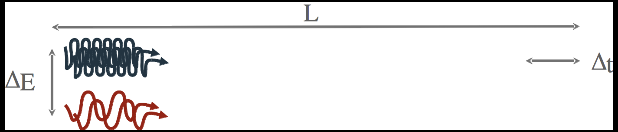
Gamma-ray emission entangled to CR emission
Could carry signatures of hadronic/leptonic production

Dark matter searches

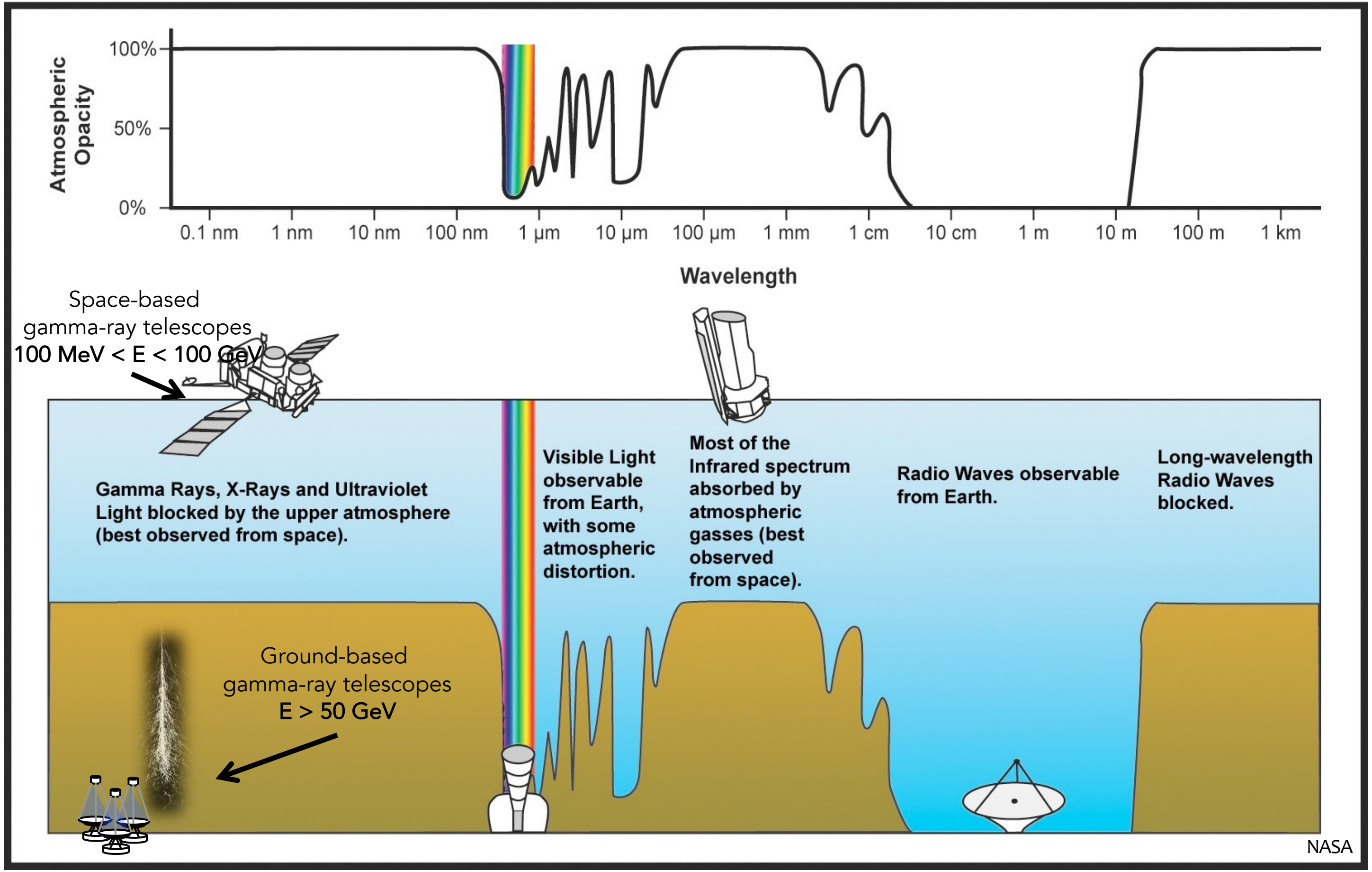




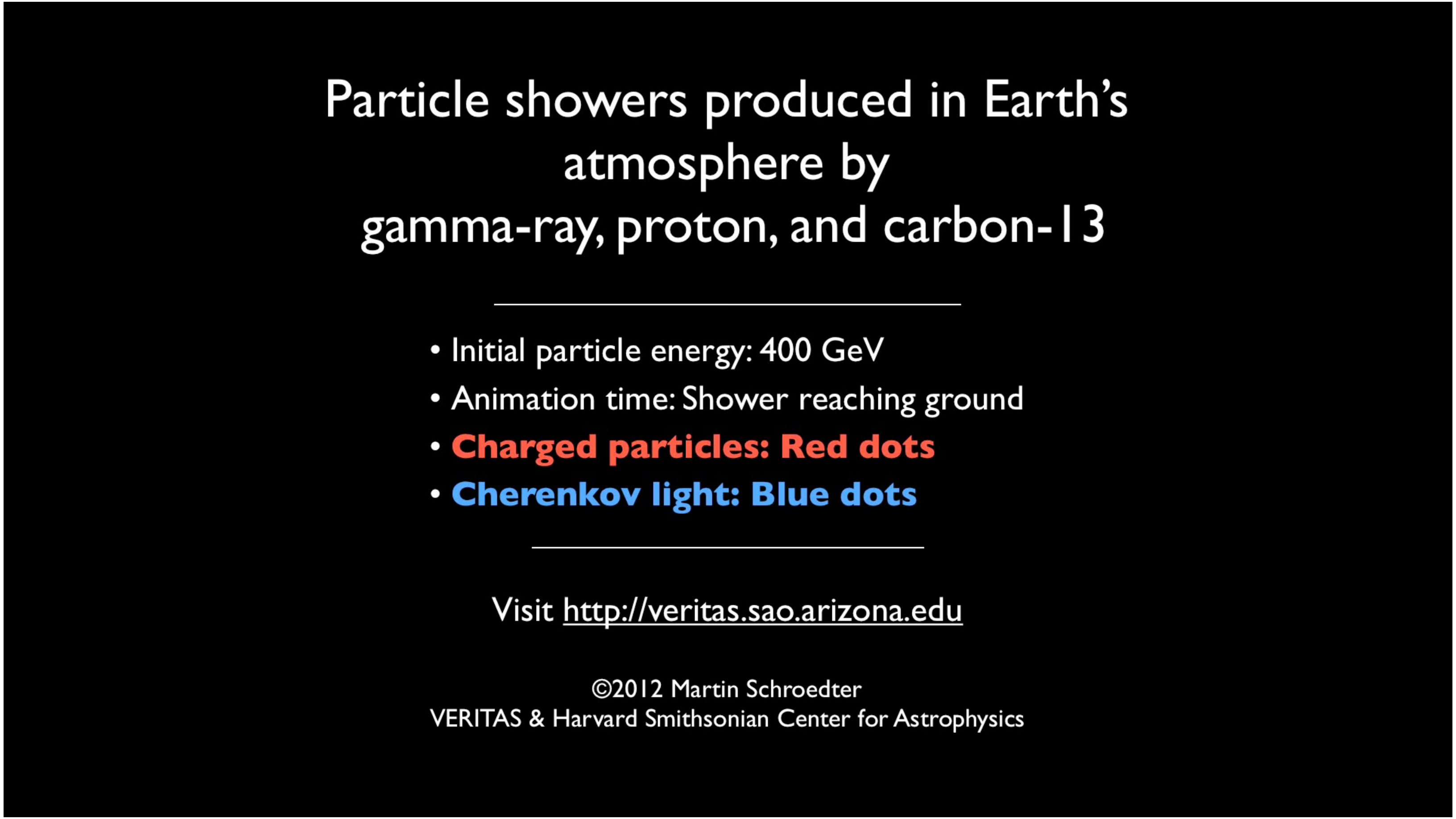
Lorentz invariance



Gamma-ray detectors



Extended atmospheric showers

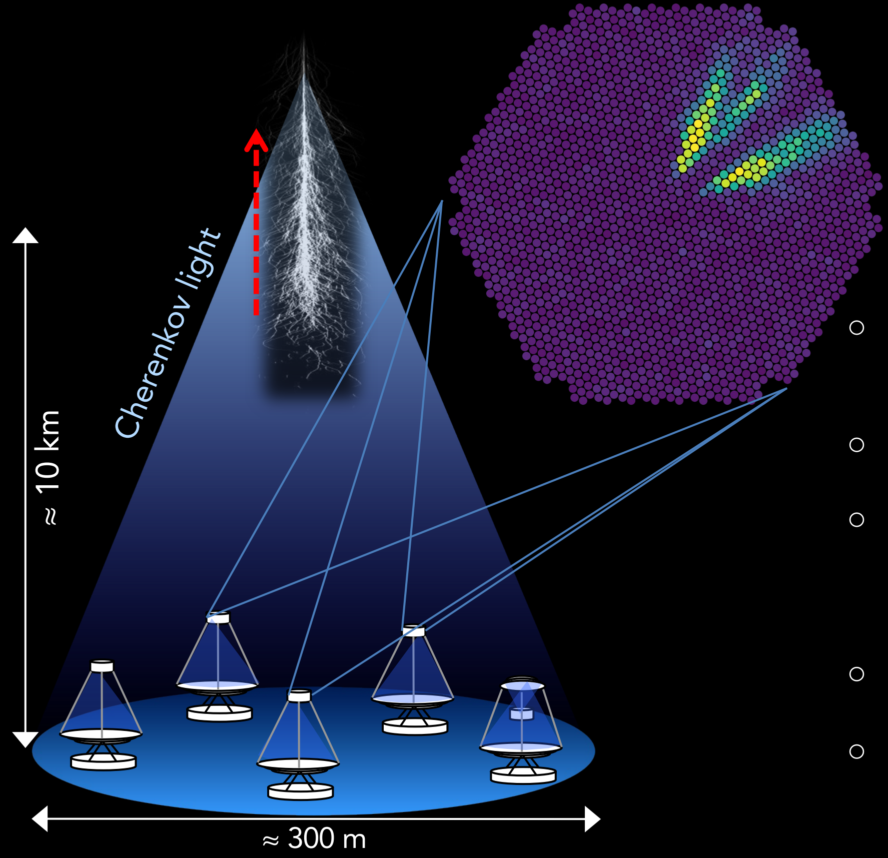


https://www.youtube.com/watch?v=j-BBzWlOai0

Imaging atmospheric Cherenkov technique

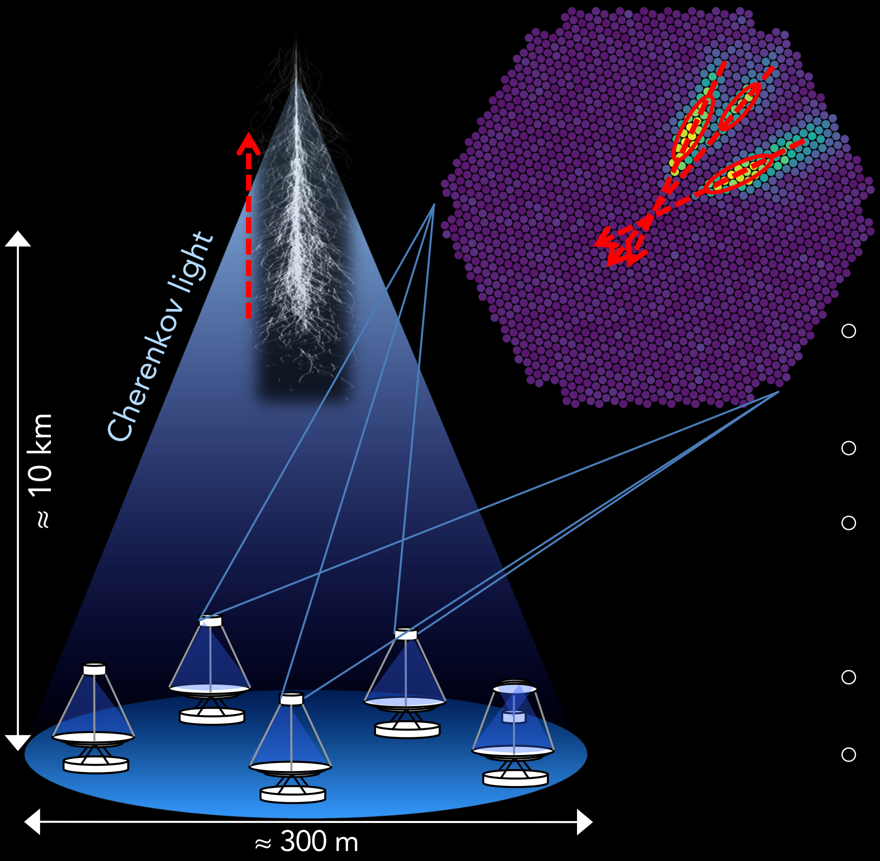
https://apod.nasa.gov/apod/ap200724.html

Imaging atmospheric Cherenkov technique



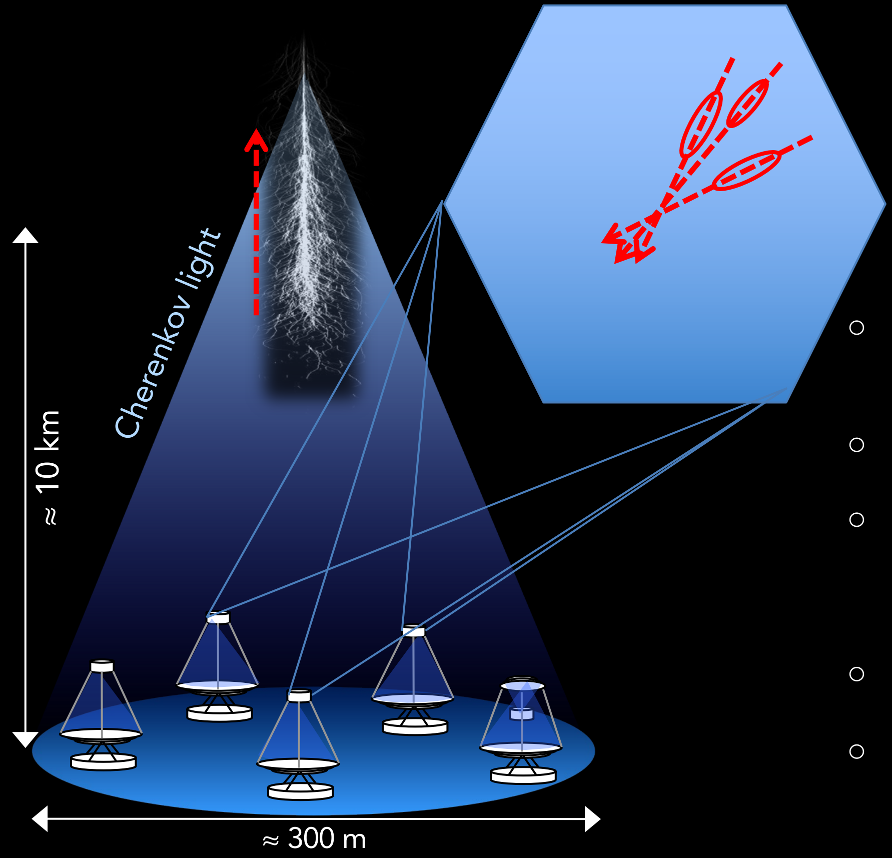
- Detection of extended air showers using the atmosphere as a calorimeter
- Huge γ–ray collection area (~10^5 m^2)
- Large background from charged CR
  - Partly irreducible (e^-/e^+ , single-EM, with current methods)
- Energy window: tens GeV - tens TeV
- Event reconstruction from image:
  - Type of primary event
  - Primary energy estimation
  - Primary arrival direction

Imaging atmospheric Cherenkov technique



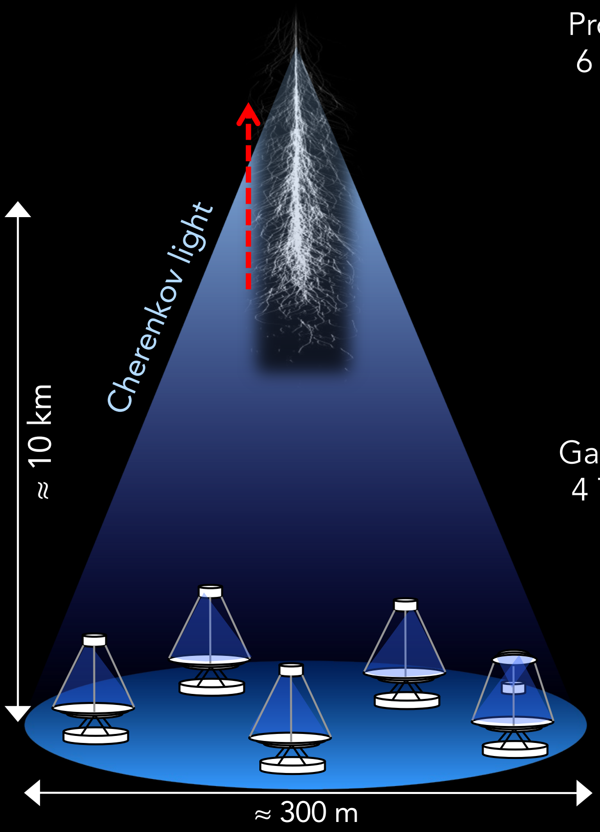
- Detection of extended air showers using the atmosphere as a calorimeter
- Huge γ–ray collection area (~10^5 m^2)
- Large background from charged CR
  - Partly irreducible (e^-/e^+, single-EM, with current methods)
- Energy window: tens GeV - tens TeV
- Event reconstruction from image:
  - Type of primary event
  - Primary energy estimation
  - Primary arrival direction

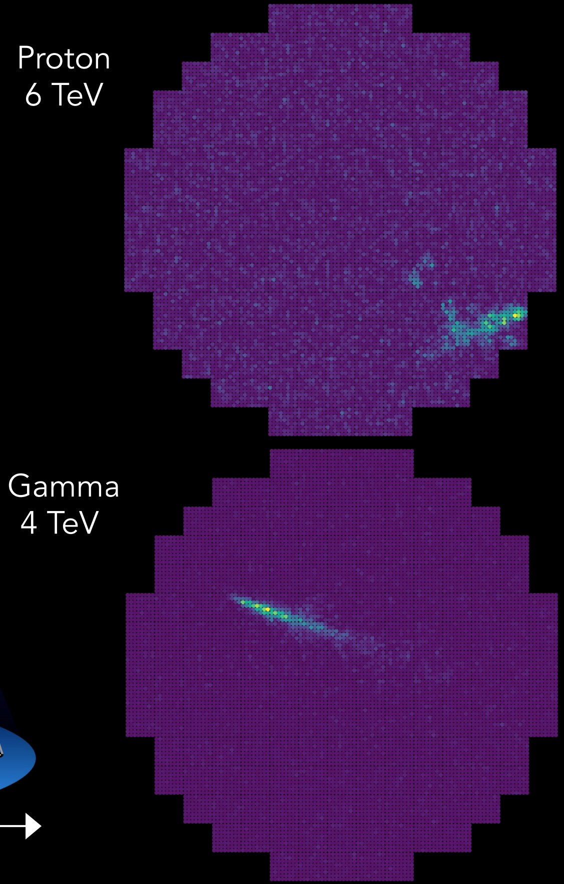
Imaging atmospheric Cherenkov technique



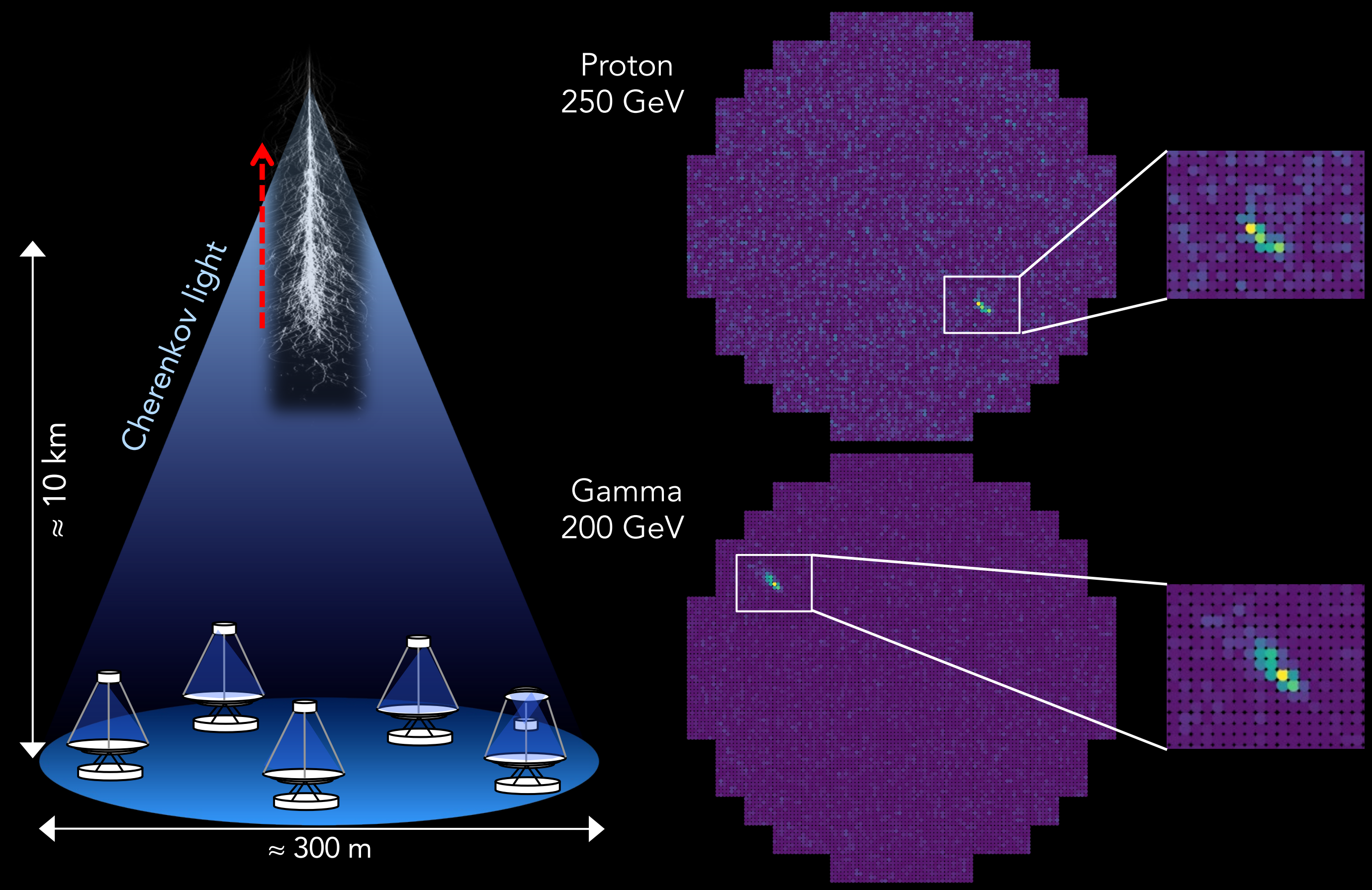
- Detection of extended air showers using the atmosphere as a calorimeter
- Huge γ–ray collection area (~10^5 m^2)
- Large background from charged CR
  - Partly irreducible (e^-/e^+ , single-EM, with current methods)
- Energy window: tens GeV - tens TeV
- Event reconstruction from image:
  - Type of primary event
  - Primary energy estimation
  - Primary arrival direction

Imaging atmospheric Cherenkov technique

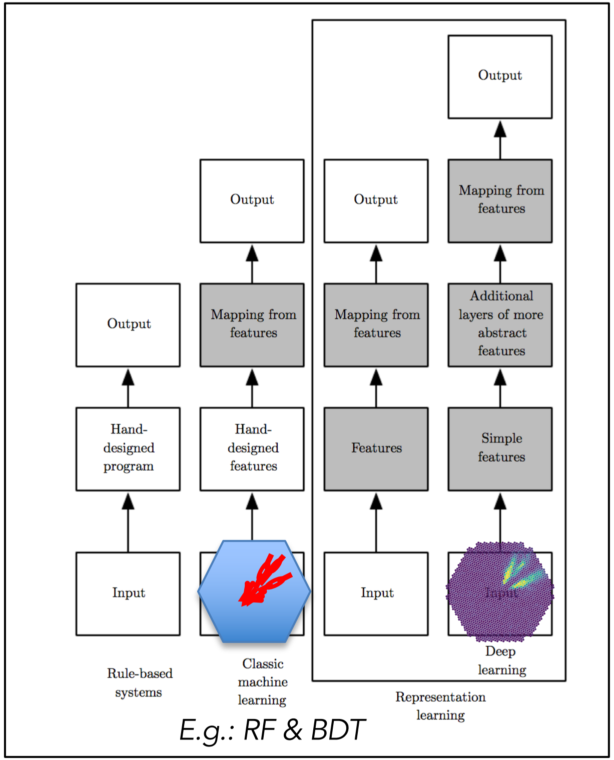


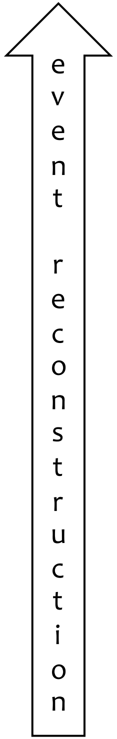


Imaging atmospheric Cherenkov technique



Output: event type, energy, incoming direction



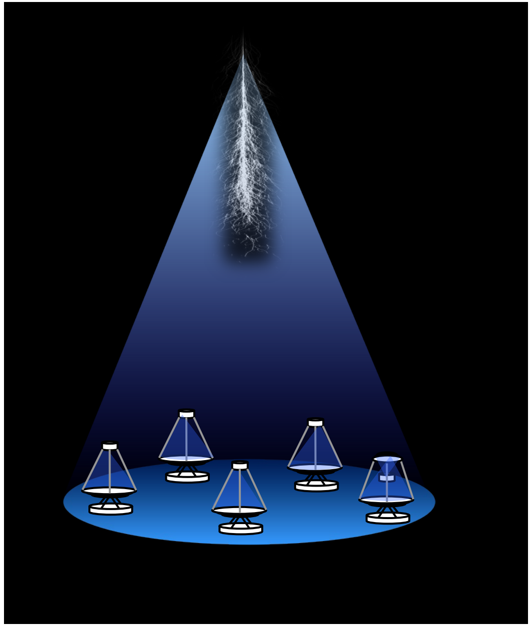


Input: observed events

Problem:
supervised learning requires labelled data

Solution:
to simulate your data!

Problem:
how well does your simulation represent the real world?



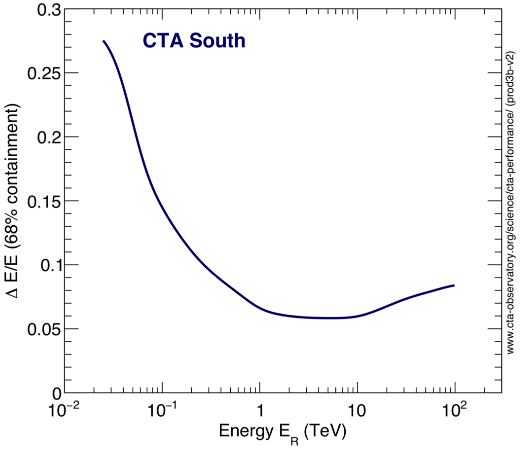
- Stereoscopy:
  - Stereoscopic view of the extended air showers
  - Compact “videos” rather than single snapshots
  - Events effectively recorded in 4D!

CREDIT: DESY/Milde Science Communication

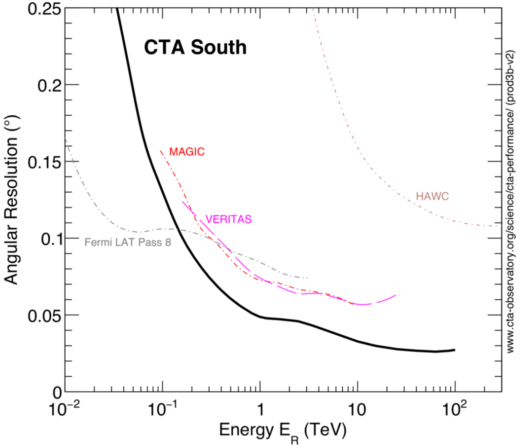
- Heterogeneity of instruments:

- Final metrics are far from trivial and entangled

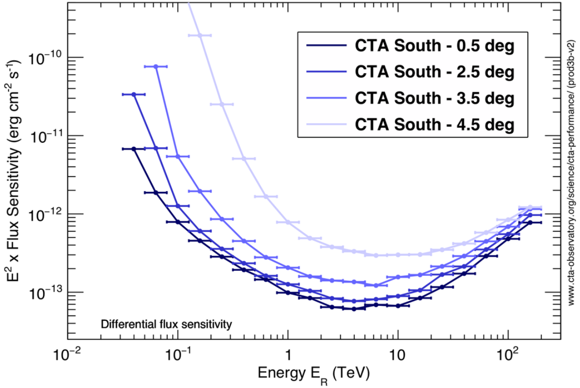
Energy resolution

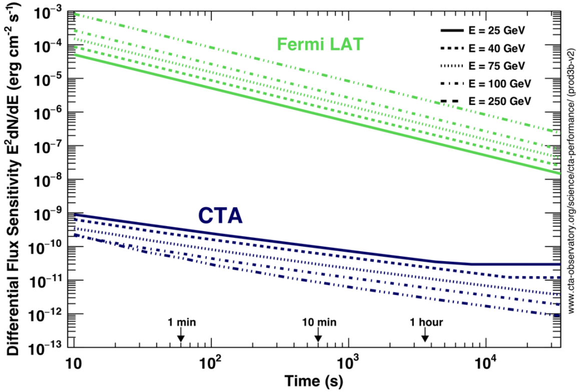


Angular resolution



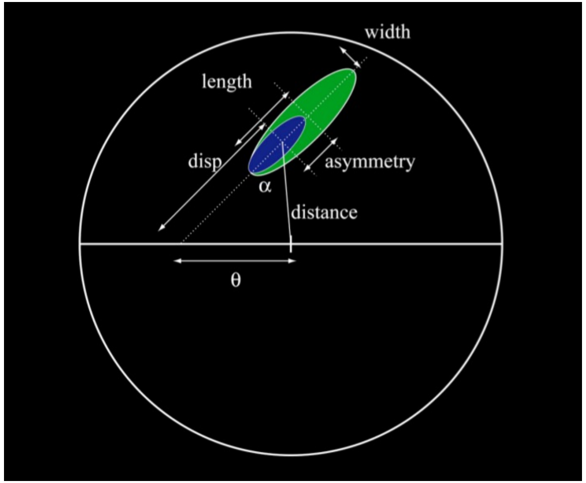
Flux sensitivity

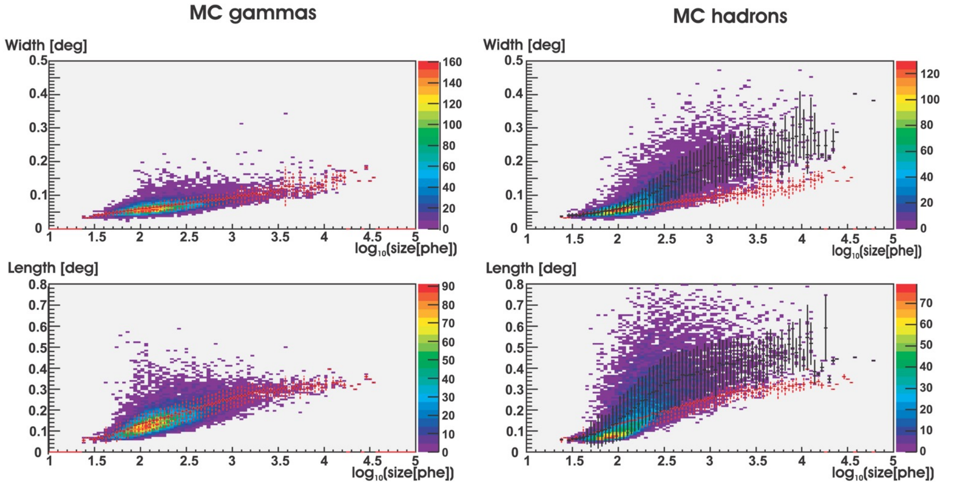




Credit: www.cta-observatory.org

- Based on image parametrization (Hillas parameters)





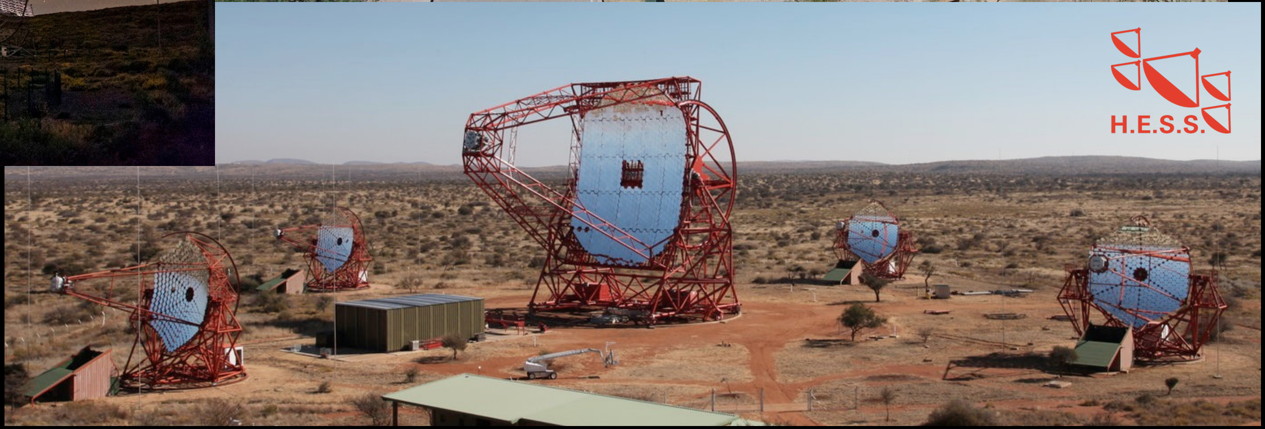
- Event type: box cuts
- Event energy: parametrization
- Event direction: parametrization

$$E = E(size, distance, h_{max})$$

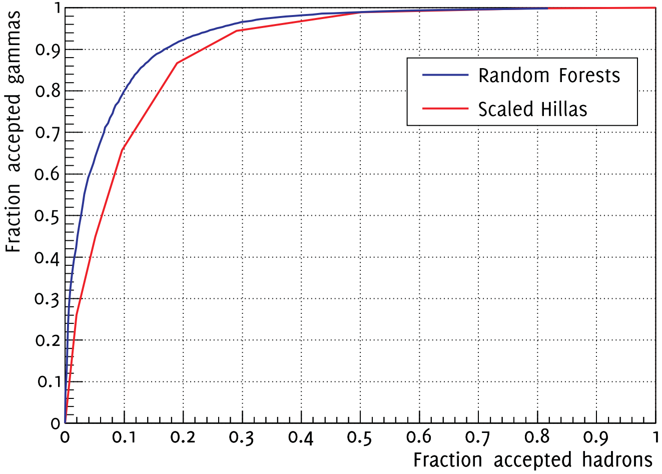
$$DISP = A(SIZE) + B(SIZE) \cdot \frac{WIDTH}{LENGTH + \eta(SIZE) \cdot LEAKAGE2}$$

Albert et al., NIM-A 588:424-432 (2008), JPCS 718(5):052003

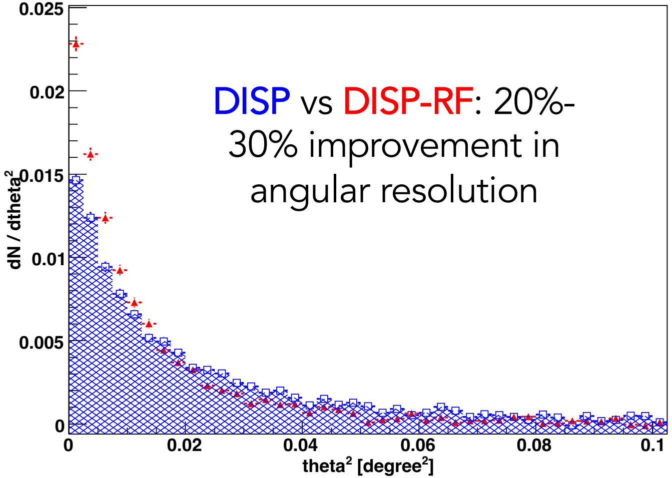
Machine learning & current-generation IACT



- ML method: Random Forest (RF)
- Applied to: background rejection, arrival direction

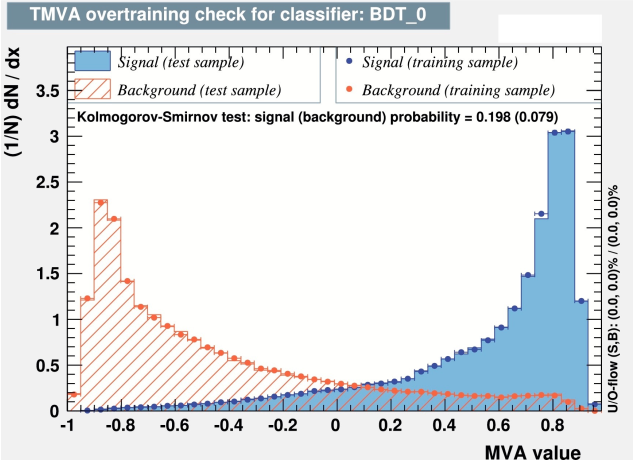


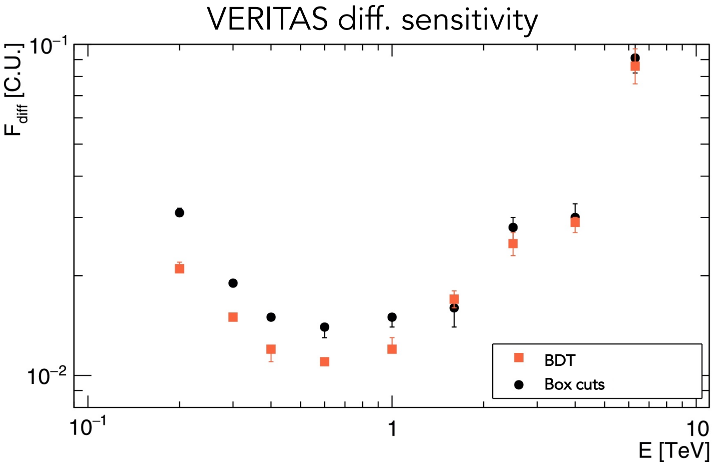
Albert et al., NIM-A 588:424-432 (2008)



Aleksic et al., A&A 524 A77 (2010)

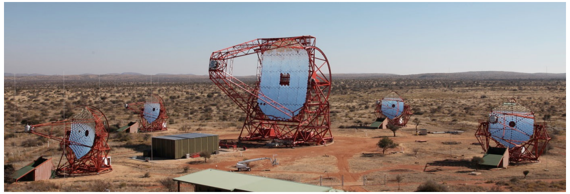
- ML method: Boosted Decision Trees (BDT)
- Applied to: background rejection

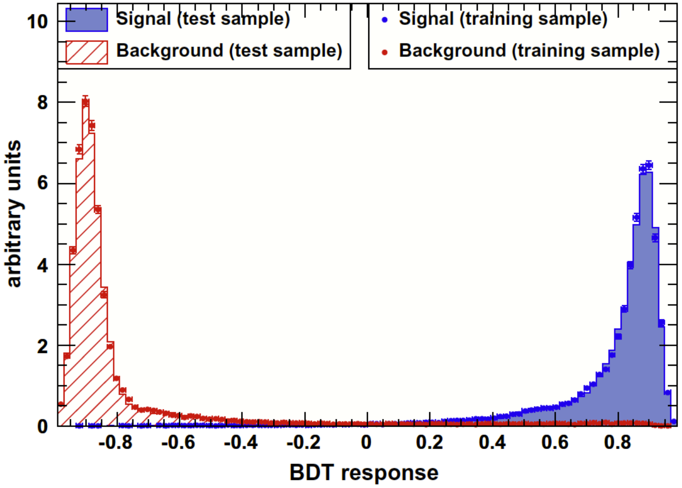


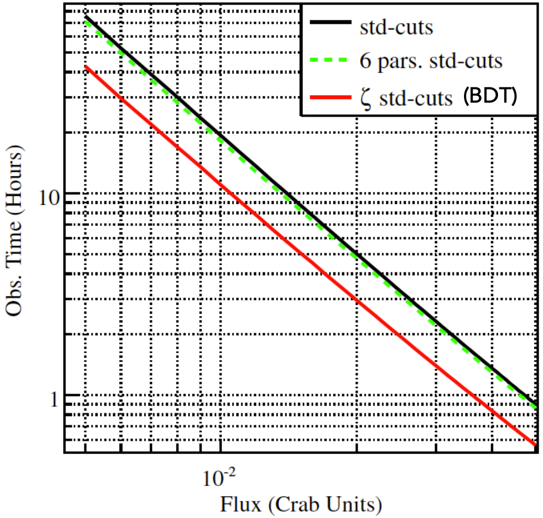


Krause et al., APP V89 P1-9 (2017)

- ML method: Boosted Decision Trees (BDT)
- Applied to: background rejection



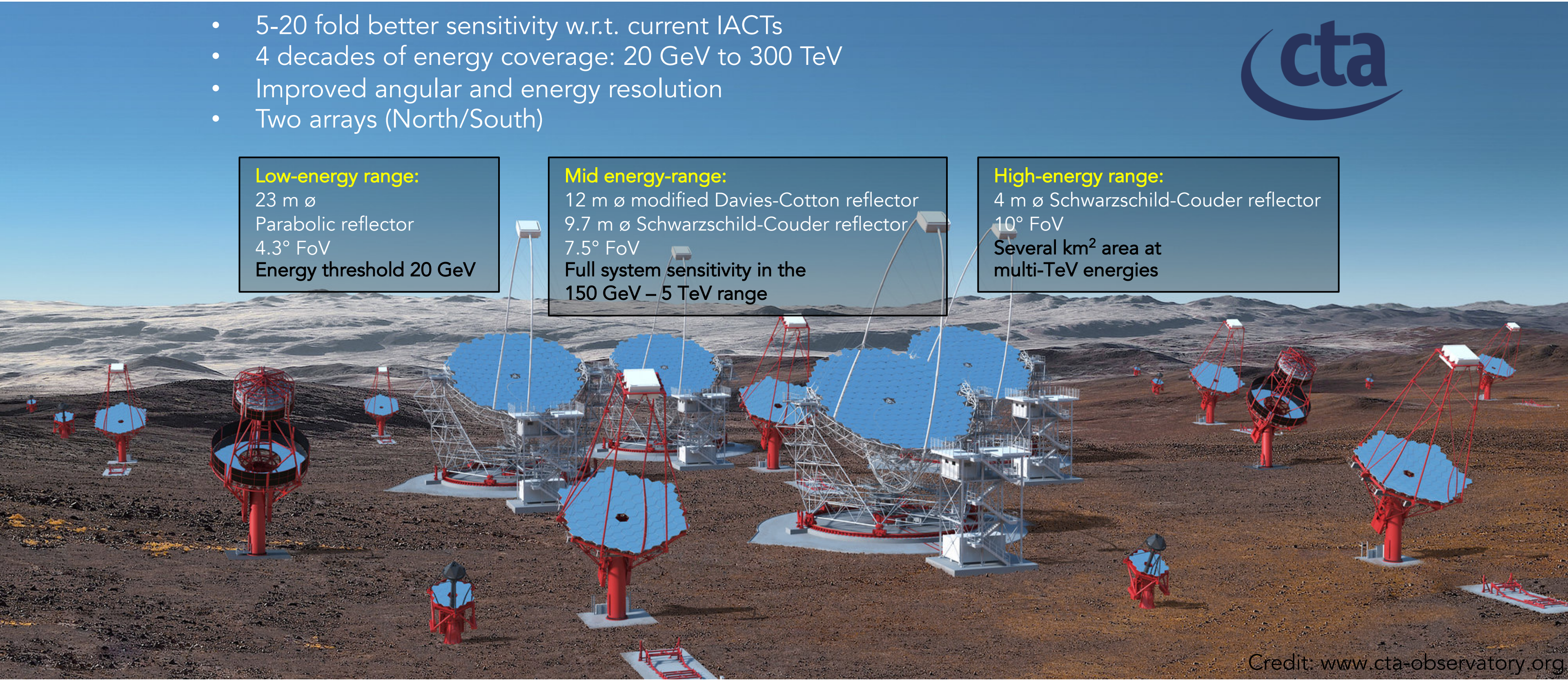




Becherini et al., APP V34-12 P858-870 (2011)

Ohm et al., APP V31-5 P383-391 (2009)

(Results for H.E.S.S. I only)



- 5-20 fold better sensitivity w.r.t. current IACTs
- 4 decades of energy coverage: 20 GeV to 300 TeV
- Improved angular and energy resolution
- Two arrays (North/South)

Low-energy range:
23 m ø
Parabolic reflector
4.3° FoV
Energy threshold 20 GeV

Mid energy-range:
12 m ø modified Davies-Cotton reflector
9.7 m ø Schwarzschild-Couder reflector
7.5° FoV
Full system sensitivity in the 150 GeV – 5 TeV range

High-energy range:
4 m ø Schwarzschild-Couder reflector
10° FoV
Several km^2 area at multi-TeV energies

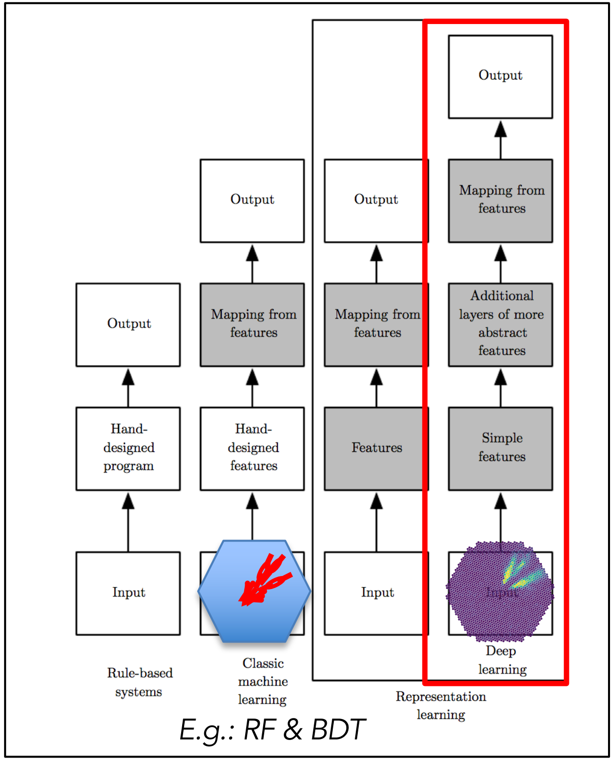
Credit: www.cta-observatory.org

www.cta-observatory.org

Science with CTA, arXiv:1709.07997

Enhancing IACT performance with deep learning?

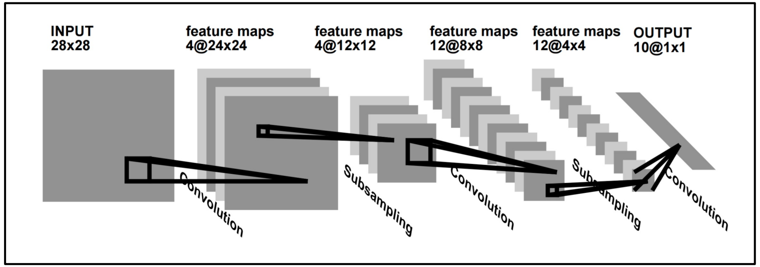
Output: event type, energy, incoming direction



Input: observed events

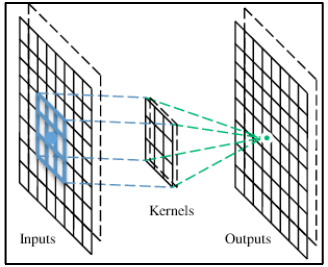
event reconstruction

Convolutional Neural Network (CNN)



LeCunn et al.

Convolution



Guo et al.

- DL capable of ***extracting*** and mapping image features automatically with unprecedented classification accuracy. Hyper-active CS research field constantly improving
- Many HEP/Astro experiments already exploring/utilizing the technique (LIGO, LHC, MicroBooNe, NOVa, etc…)

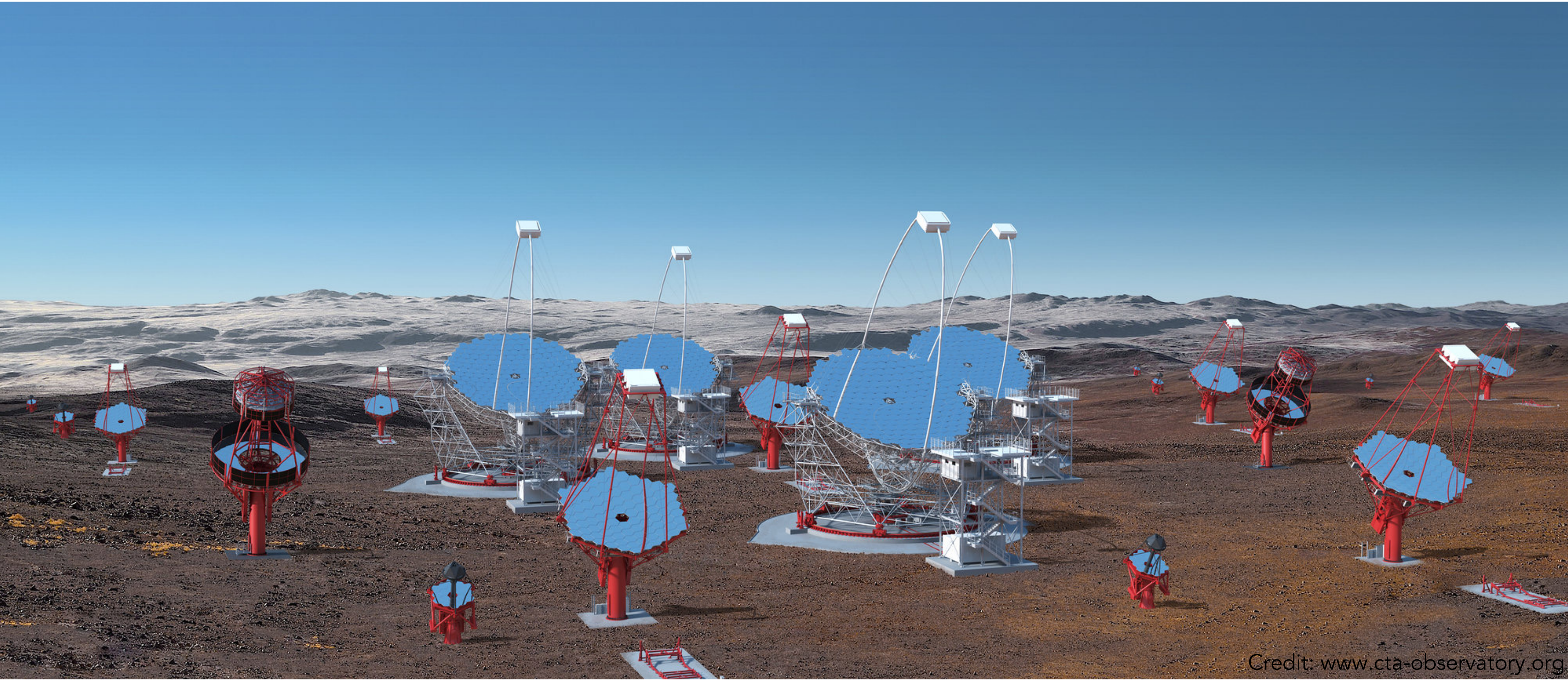
Method:

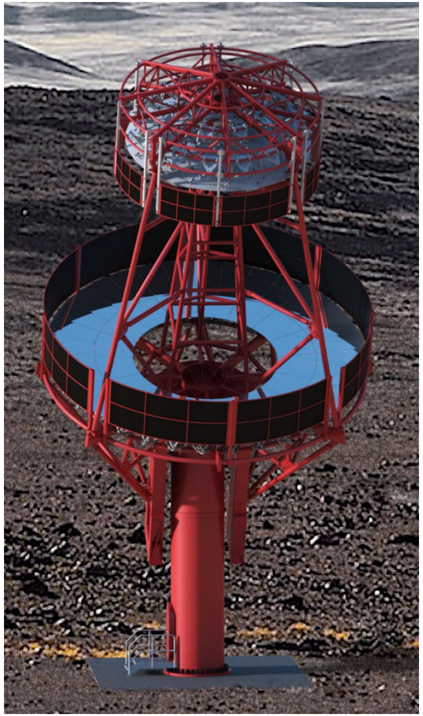
- Use deep learning to reconstruct CTA events from non-parameterized images
  - Performance enhancement -> better sensitivity

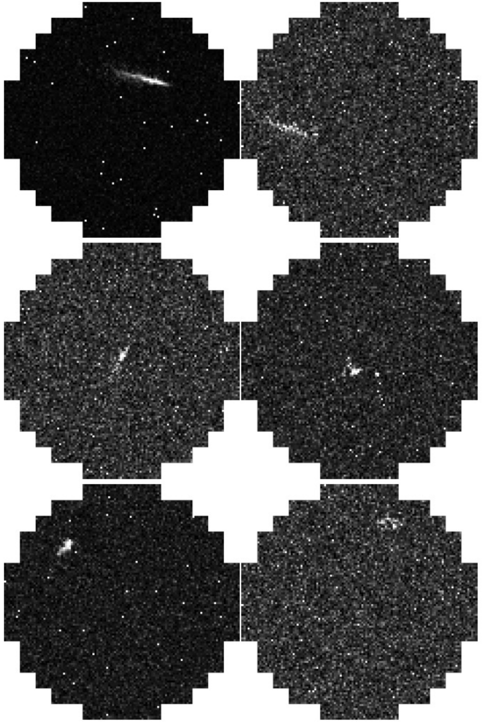
But there are risk…

- MC reliability (e.g. network selecting some features from your MC not present in real data)

Proof of concept: gamma/hadron classification in SC-MST







- Single telescope
- Square pixels
- Only signal charge (no timing)
- Single task: classification

- Three energy bins:

Bin	E_{min} [TeV]	E_{max} [TeV]	N_{gamma}	N_{proton}
Total			4160578	6518742
Low Energy	0.1	0.31	727316	499909
Medium Energy	0.31	1	657397	245912
High Energy	1	10	642034	147012

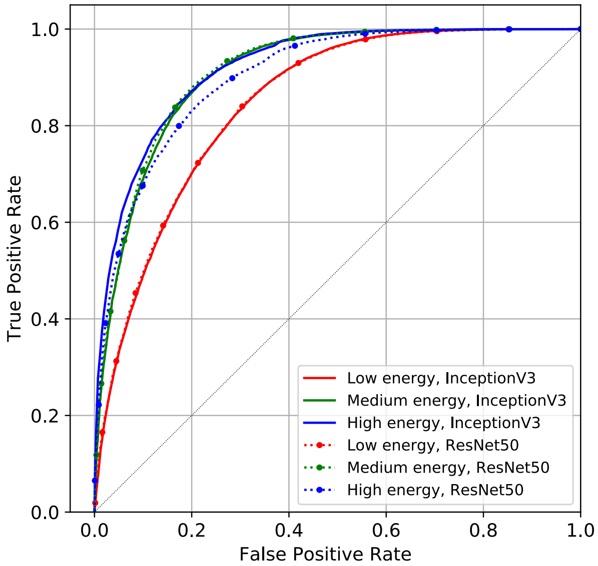
- Sanity cuts prior to BDT training:

Cut
$0 \leq \sqrt{MCxoff^2 + MCyoff^2} \leq 3$
$-2 < MSCW < 2$
$-2 < MSCL < 5$
$EChi2S \geq 0$
$ERecS > 0$
$0 < EmissionHeight < 50$
$dES \geq 0$

Nieto et al., PoS(ICRC2017)809

- Classification happened!

(Note: results for single-telescope images)

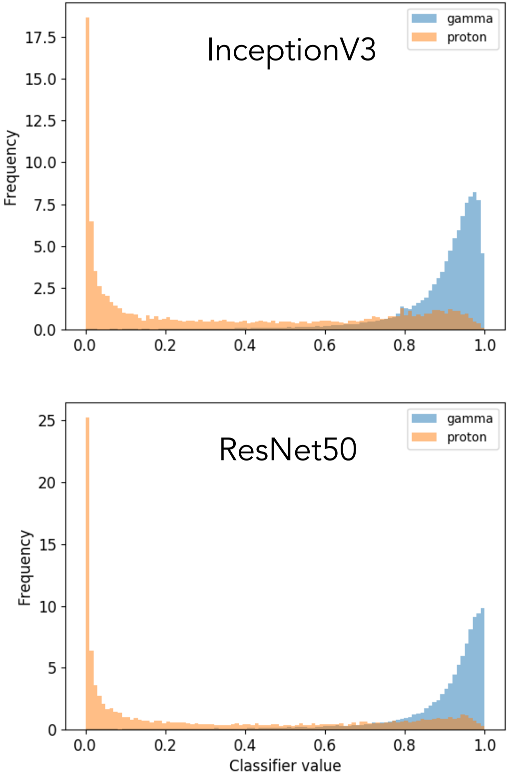


Area Under the Curve

Model/Energy	Low E.	Med. E.	High E.
InceptionV3	84.7%	91.1%	92.0%
ResNet50	84.8%	91.4%	90.2%

100% -> perfect classification
50% -> random classification

Medium energies
(0.3 TeV < E < 1 TeV)

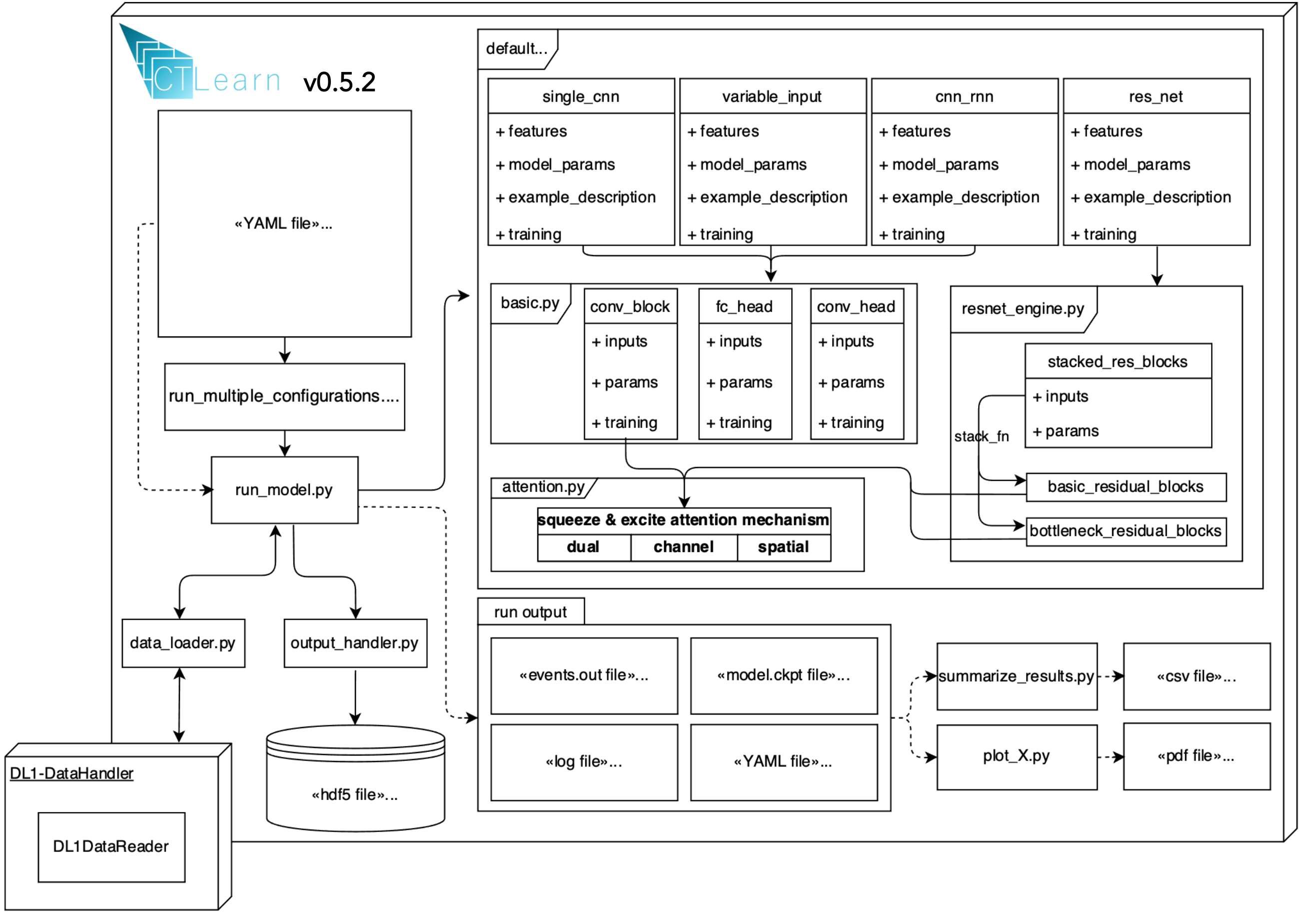


Nieto et al., PoS(ICRC2017)809

- High-level Python package for using deep learning for IACT event reconstruction
- Configuration-file-based workflow and installation with conda drive reproducible training and prediction
- Supports any TensorFlow model that obeys a generic signature
- Open source on GitHub:

  https://github.com/ctlearn-project/ctlearn
  https://pos.sissa.it/358/752
  DOI 10.5281/zenodo.3345947
  (Latest release: **CTLearn v0.5.2**, 02/02/22)

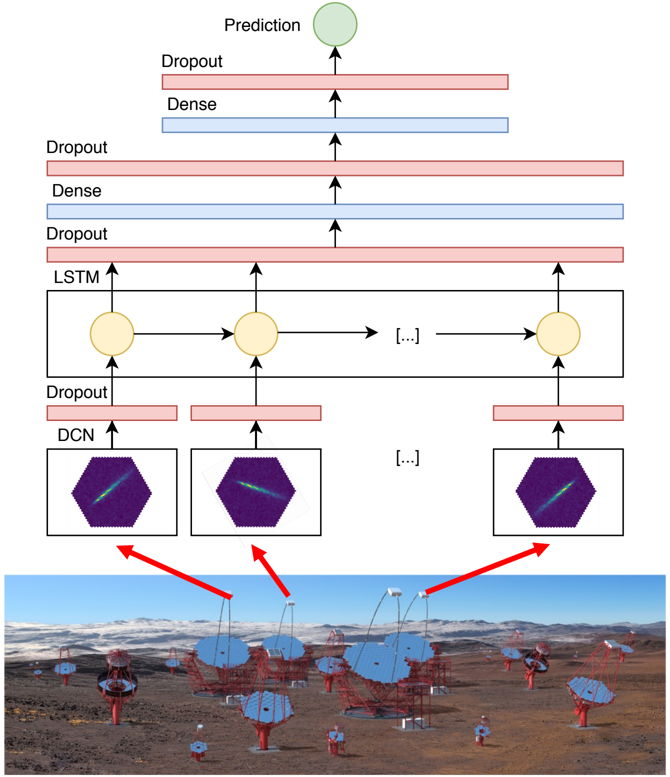
<u>Core developers</u>
Tjark Miener, DN (**IPARCOS-UCM**)
Ari Brill, Qi Feng (Columbia)
Bryan Kim (UCLA, now at Stanford)
(See contributors here)



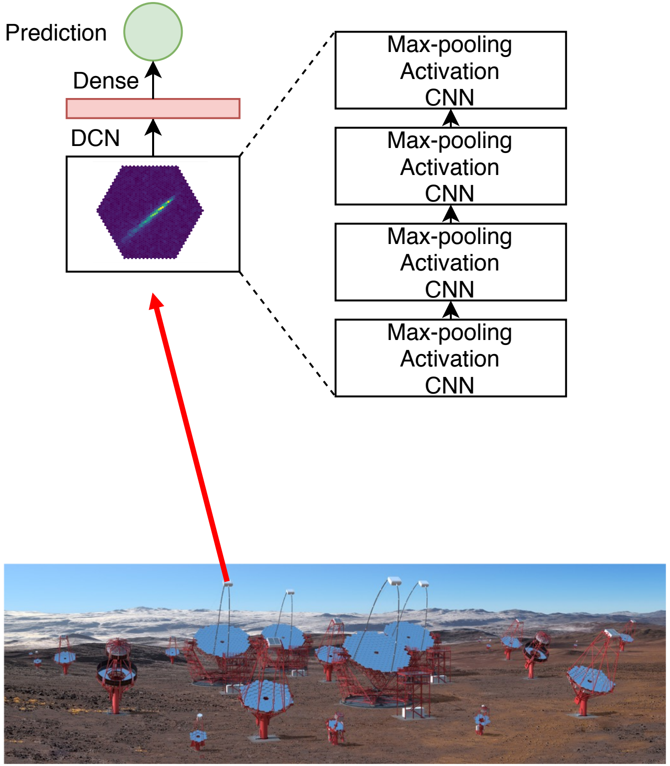
CTLearn v0.5.2
«YAML file»...
run_multiple_configurations....
run_model.py
data_loader.py
output_handler.py
DL1-DataHandler
DL1DataReader
«hdf5 file»...
default...
single_cnn
+ features
+ model_params
+ example_description
+ training
variable_input
+ features
+ model_params
+ example_description
+ training
cnn_rnn
+ features
+ model_params
+ example_description
+ training
res_net
+ features
+ model_params
+ example_description
+ training
basic.py
conv_block
+ inputs
+ params
+ training
fc_head
+ inputs
+ params
+ training
conv_head
+ inputs
+ params
+ training
resnet_engine.py
stacked_res_blocks
+ inputs
+ params
stack_fn
basic_residual_blocks
bottleneck_residual_blocks
attention.py
squeeze & excite attention mechanism
dual
channel
spatial
run output
«events.out file»...
«model.ckpt file»...
«log file»...
«YAML file»...
summarize_results.py
«csv file»...
plot_X.py
«pdf file»...

CTLearn: gamma/hadron classification

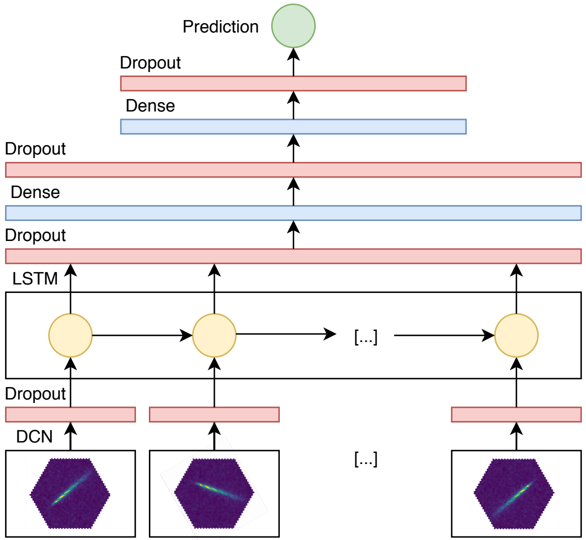
CNN-RNN model

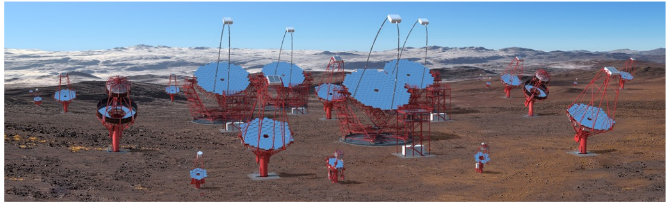


Single-tel model

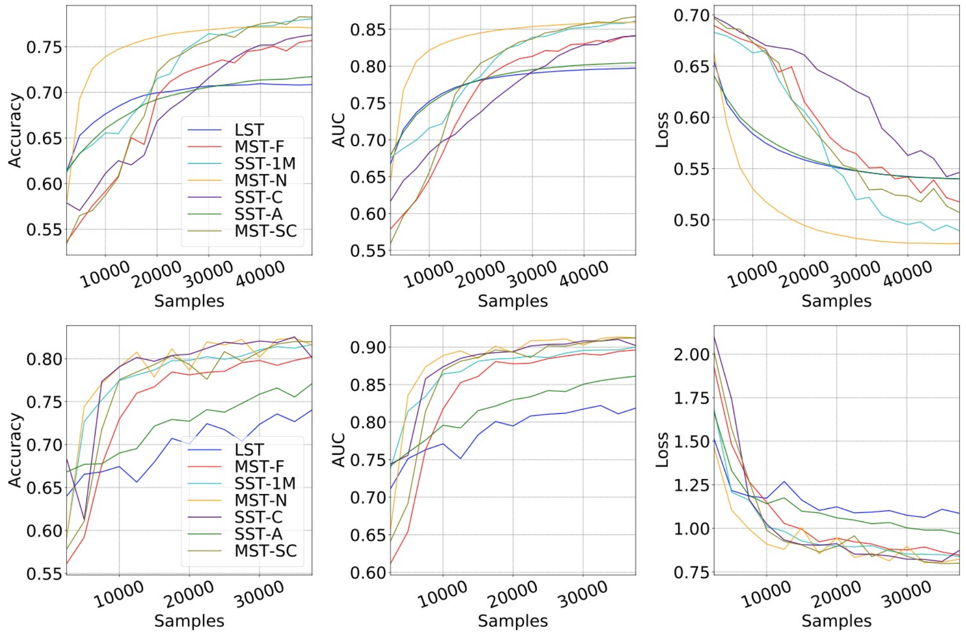


CNN-RNN model



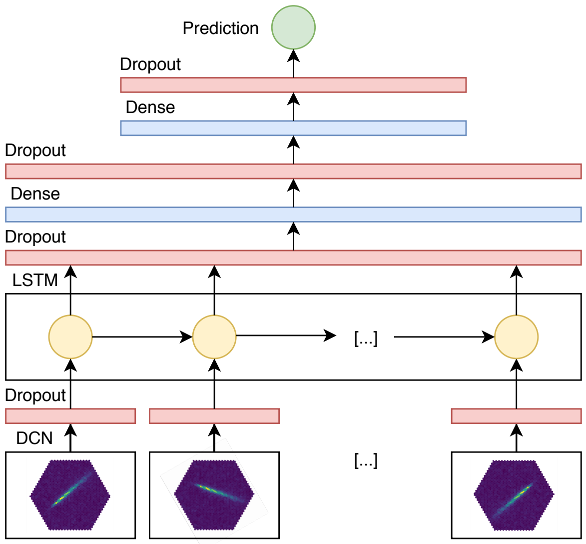


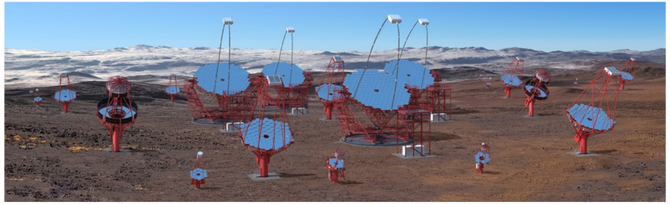
Gamma/hadron classification



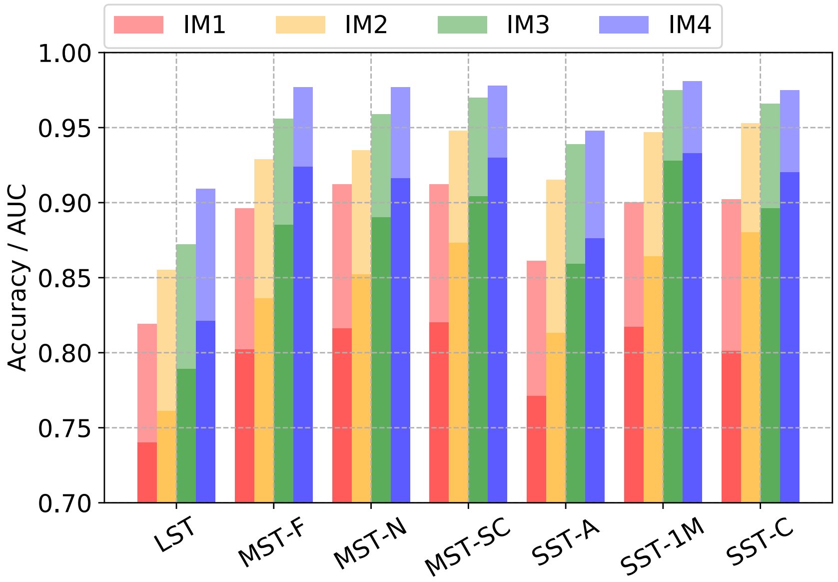
D. Nieto et al. PoS(ICRC2019)752

CNN-RNN model





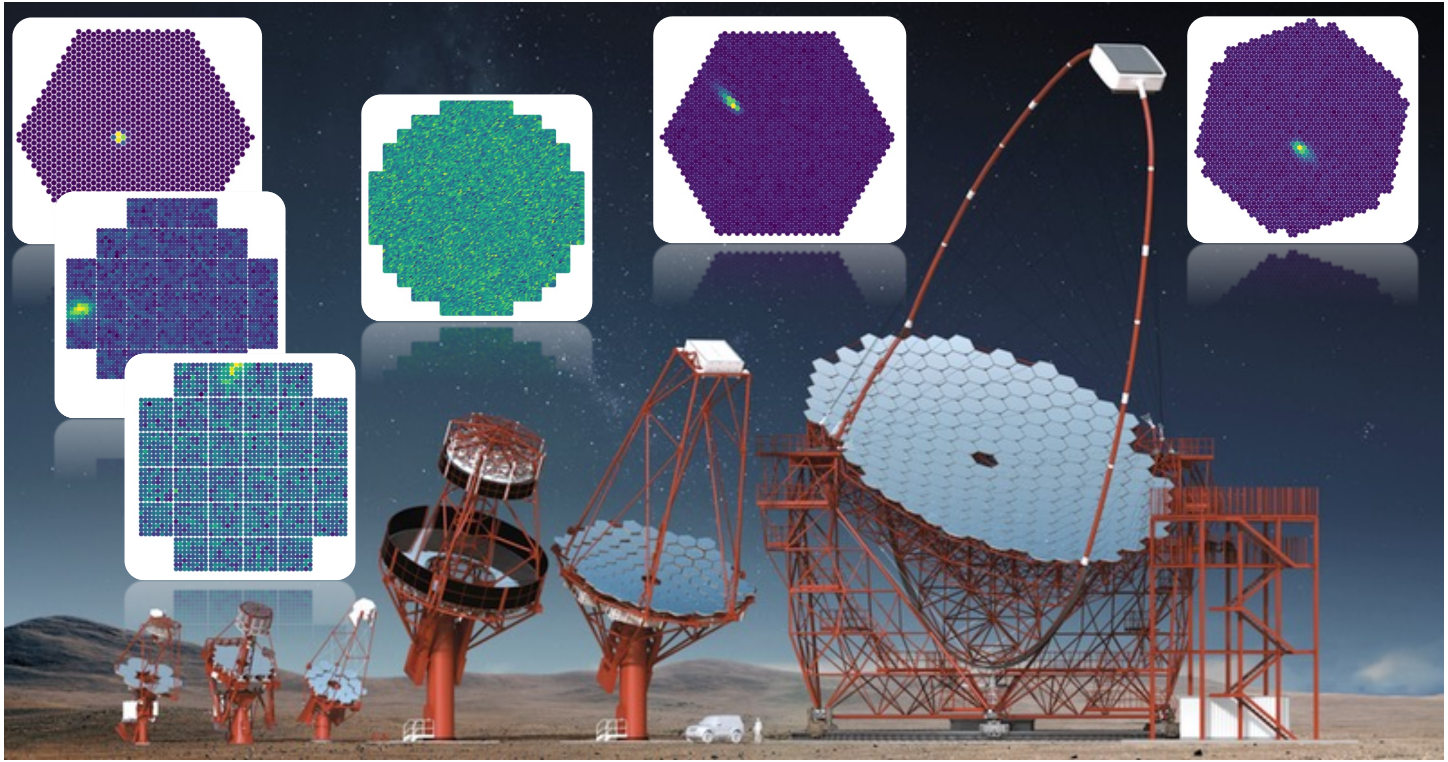
Gamma/hadron classification



D. Nieto et al. PoS(ICRC2019)752

- Heterogeneity of instruments:

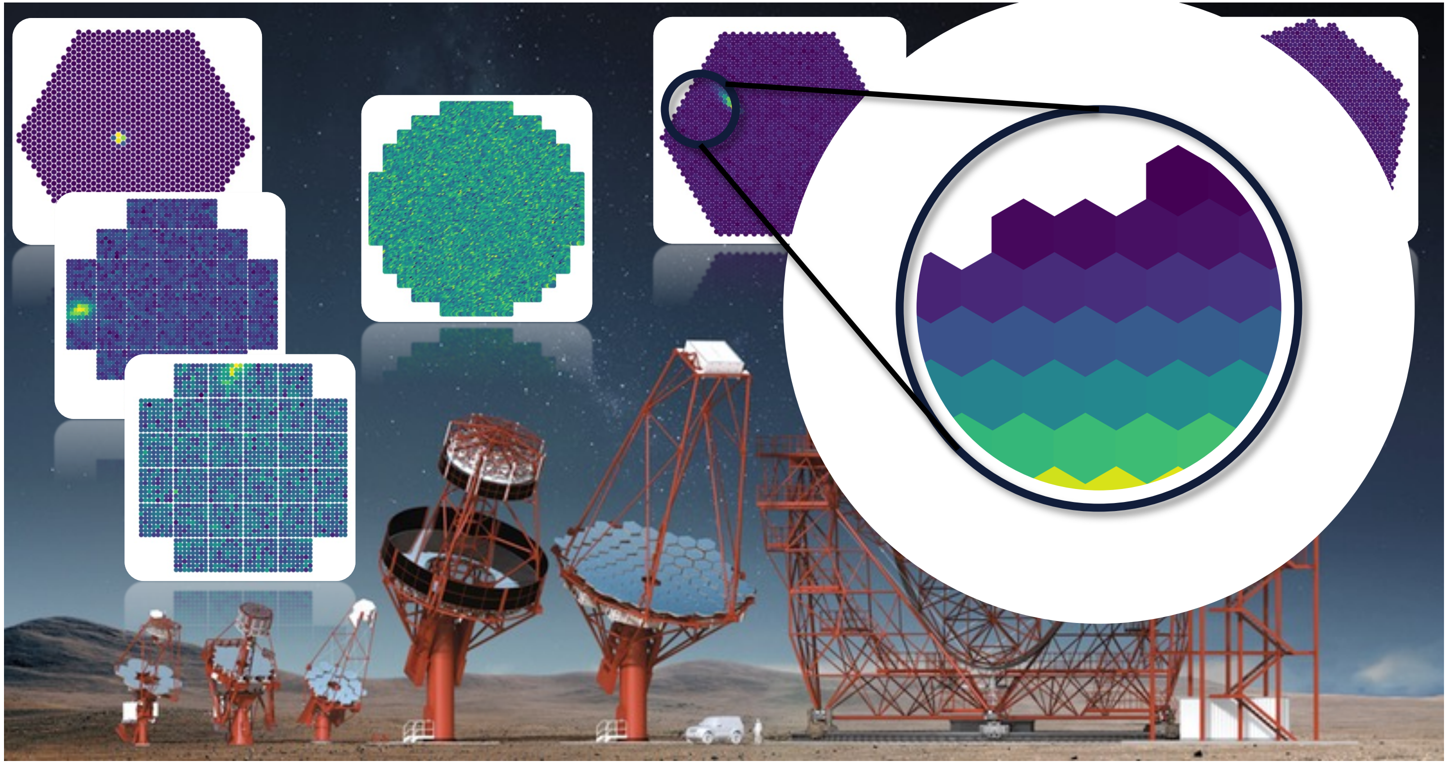
Camera images courtesy of T. Vuillaume



- Heterogeneity of instruments:

Hexagonal pixels

Camera images courtesy of T. Vuillaume

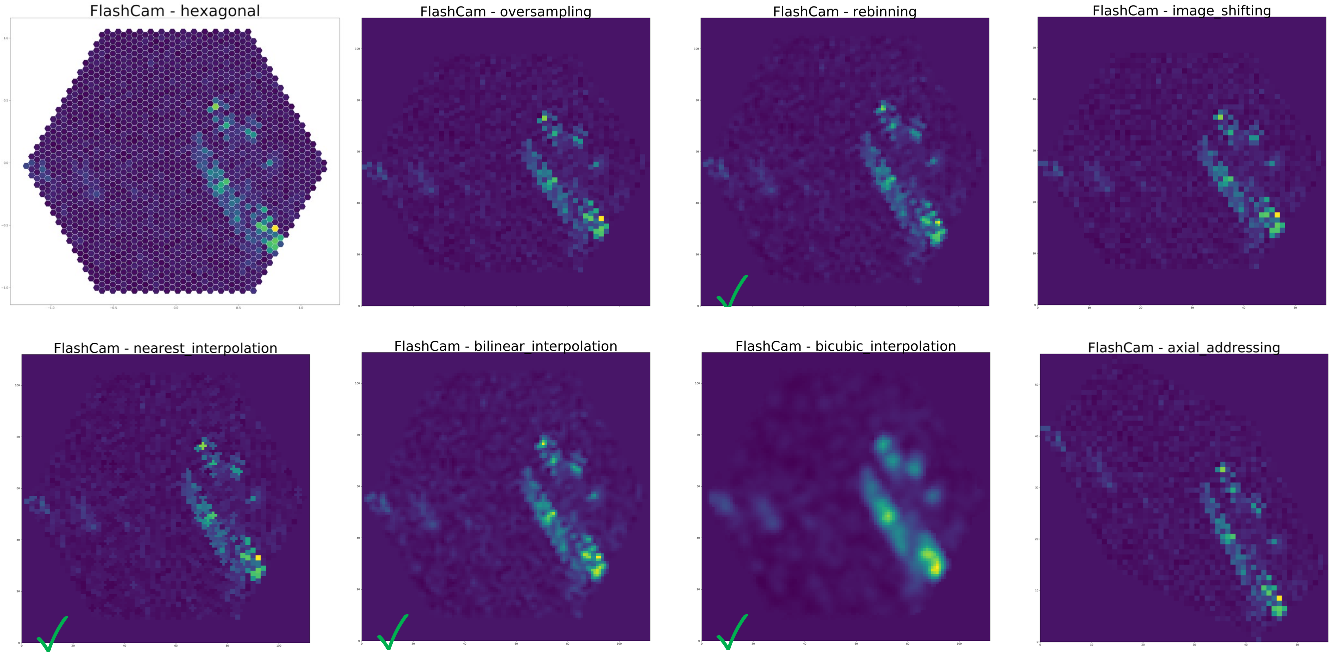


Tackling the hexagonal-pixel challenge

A. Brill, B. Kim, Q. Feng
D. Nieto, T. Miener,
et al.

https://github.com/ctlearn-project/

- Image mapping (preprocessing)

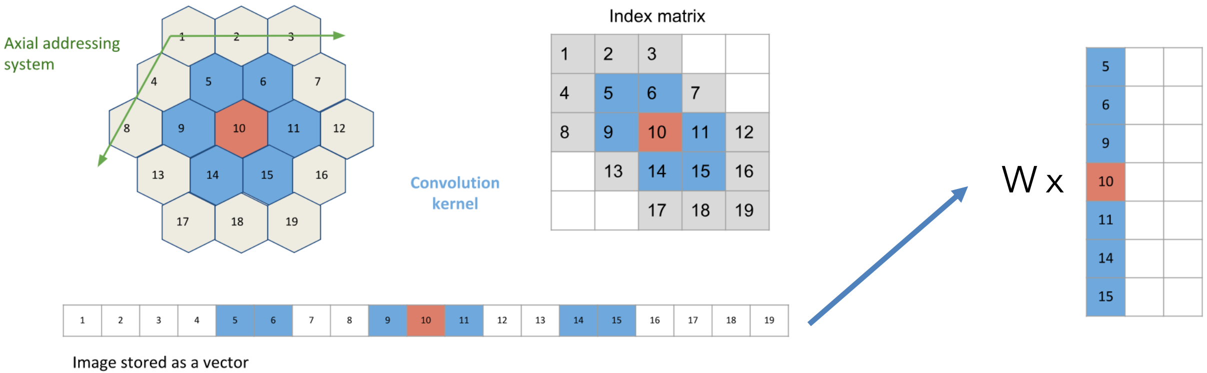


✓ Angles and distances preserved

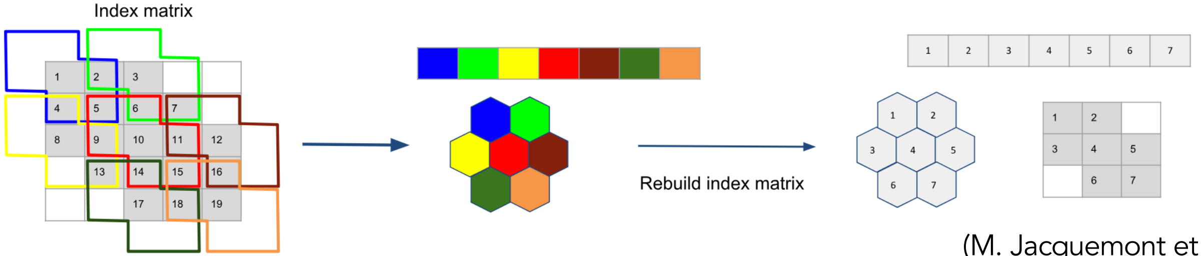
T. Vuillaume,
M. Jaquemont, et al.

https://github.com/IndexedConv

- Hexagonal convolution
  - Convolution

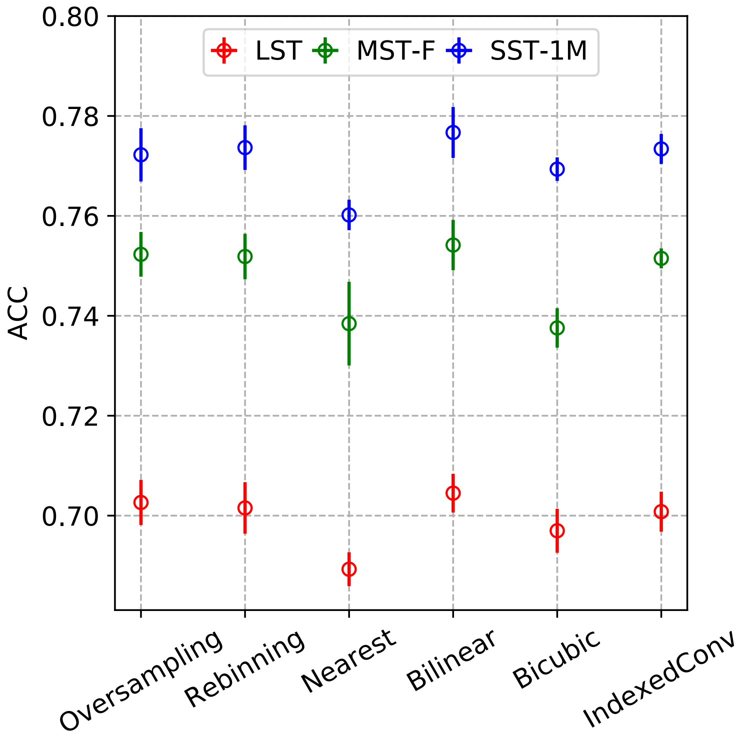


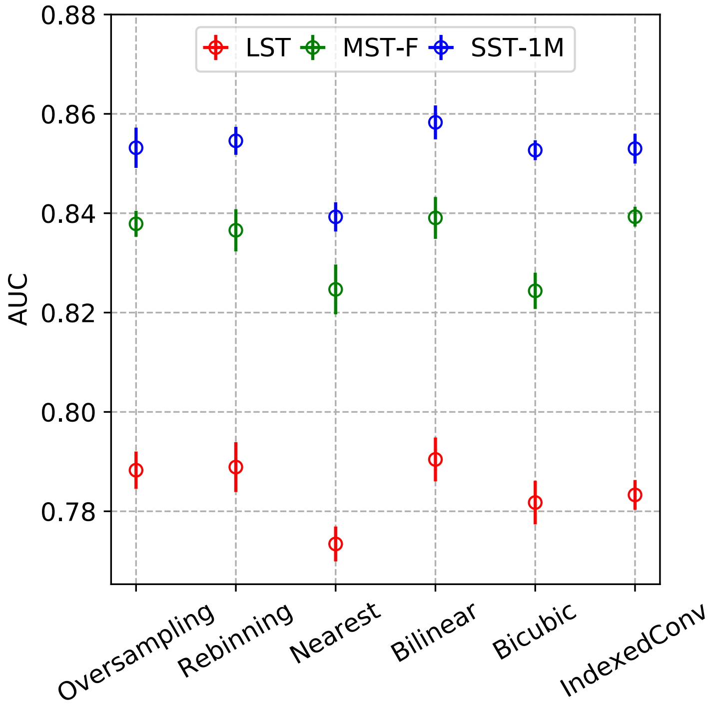
  - Pooling



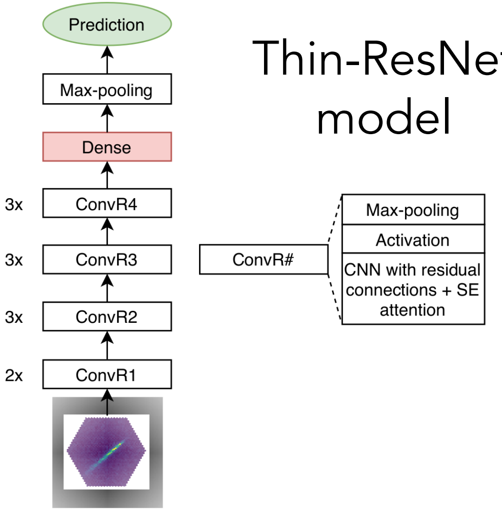
(M. Jacquemont et al. 2019)

- Comparison of methods for classification task

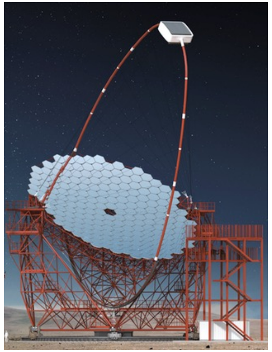




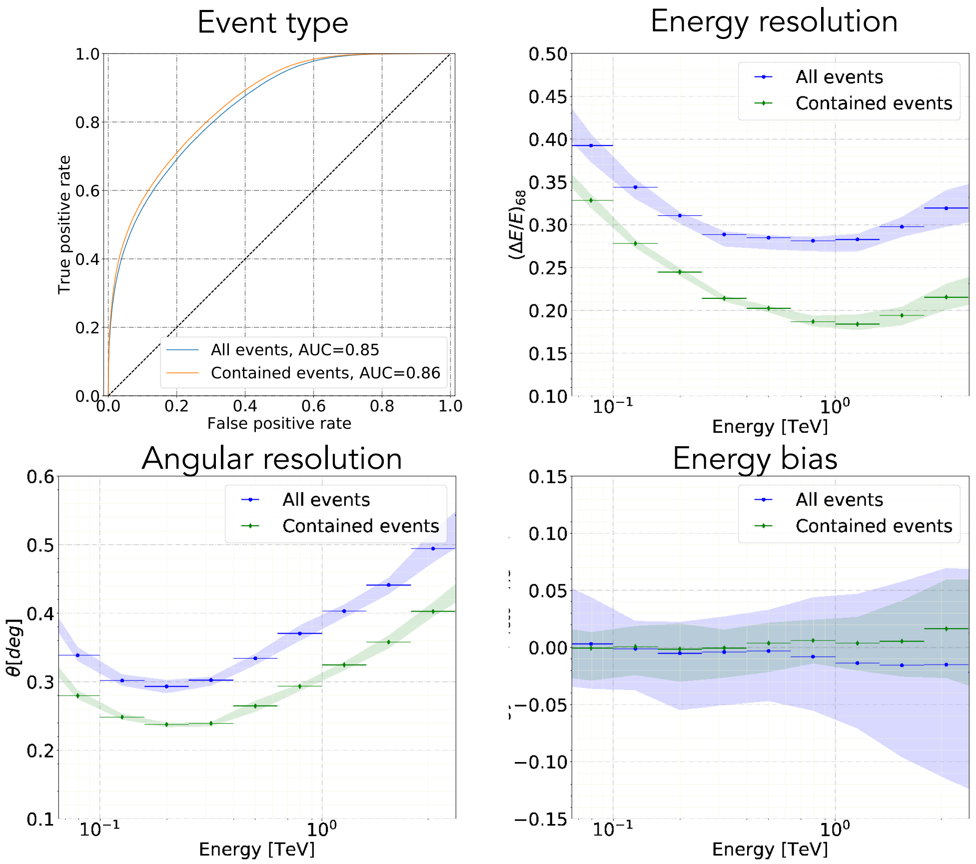
D. Nieto et al. PoS(ICRC2019)753



Thin-ResNet model



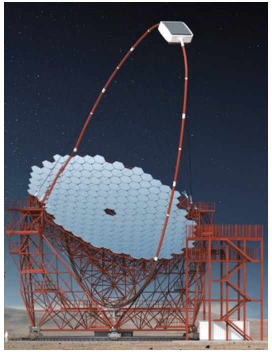
Full-event reconstruction for single-telescope data achieved!

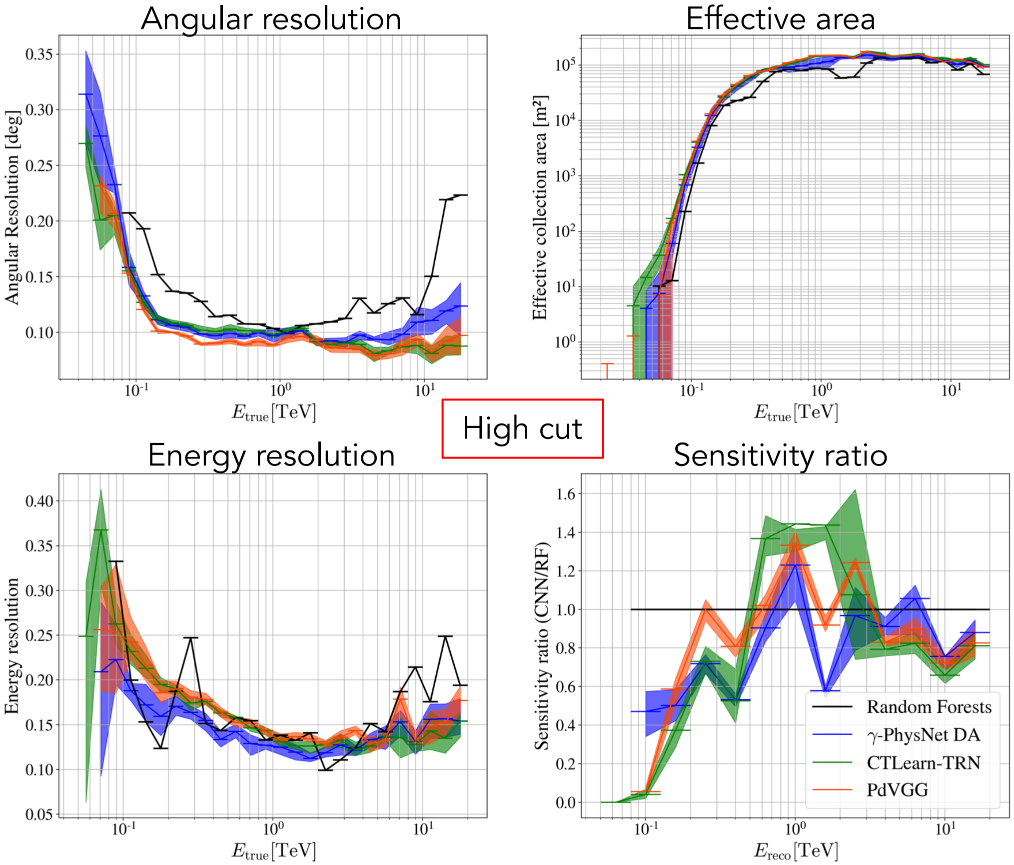


D. Nieto et al. ADASS XXX 2020

CTLearn: crosschecking results

- Crosschecking three different implementations
- Same datasets, same cuts
- Different models
- Comparison against standard analysis (RF)

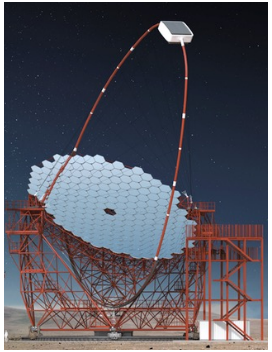


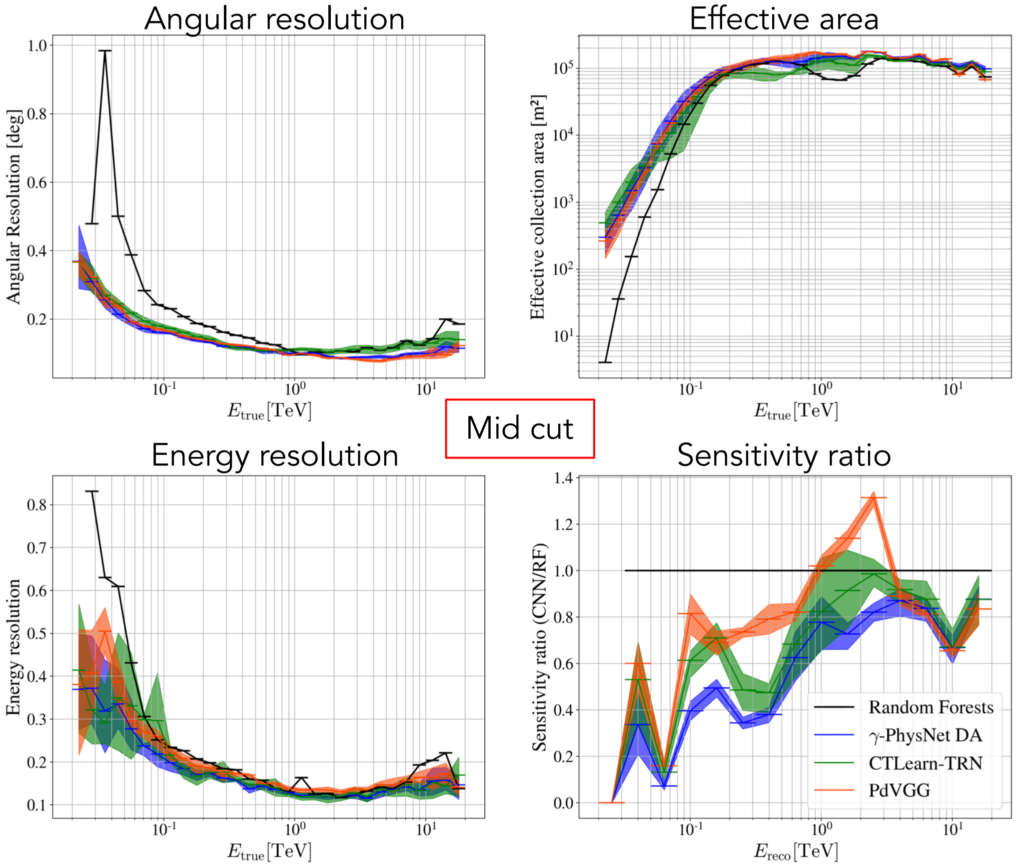


[P. Grespan et al. PoS(ICRC2021) 771](#)

CTLearn: crosschecking results

- Crosschecking three different implementations
- Same datasets, same cuts
- Different models
- Comparison against standard analysis (RF)

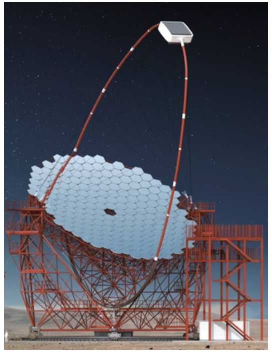


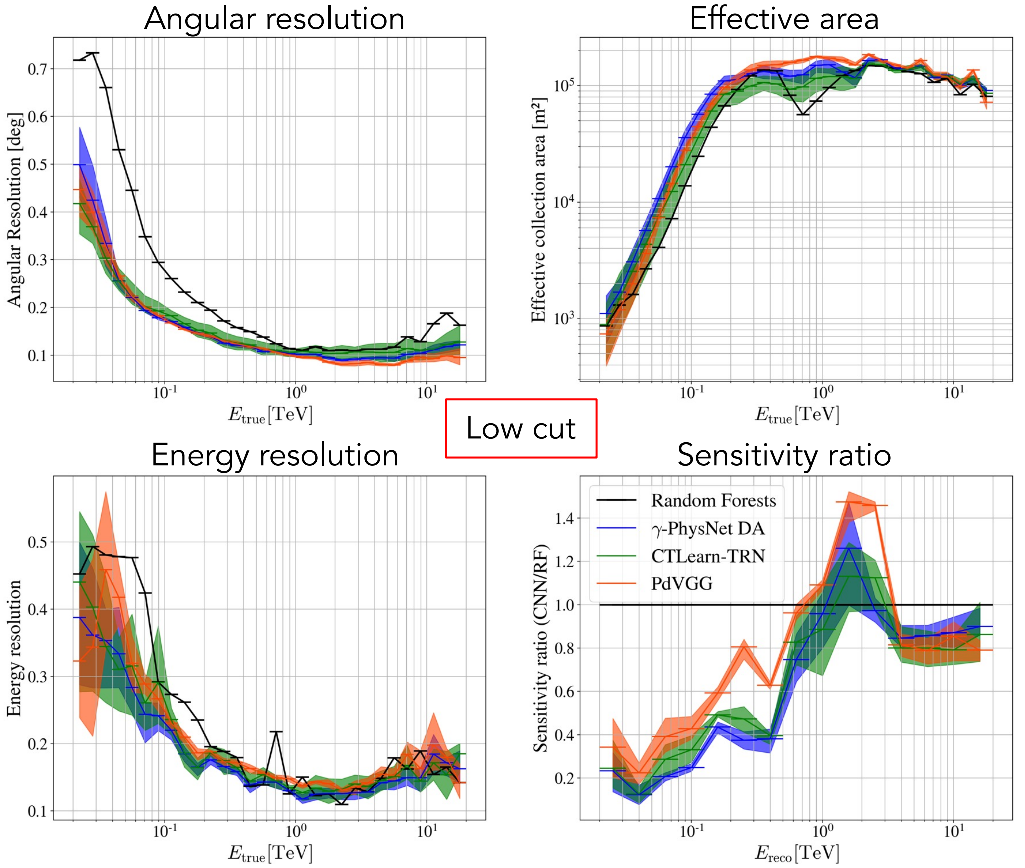


P. Grespan et al. PoS(ICRC2021) 771

CTLearn: crosschecking results

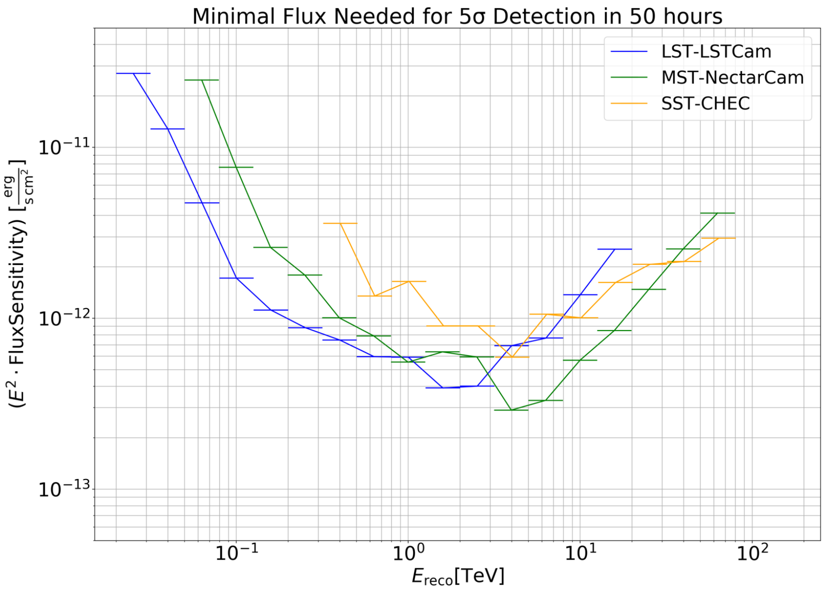
- Crosschecking three different implementations
- Same datasets, same cuts
- Different models
- Comparison against standard analysis (RF)

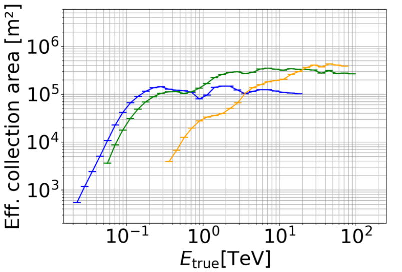


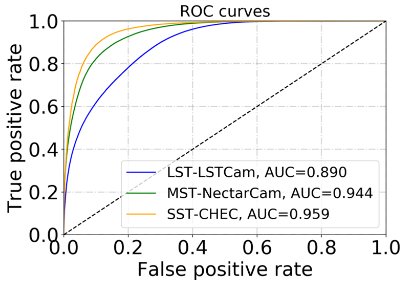


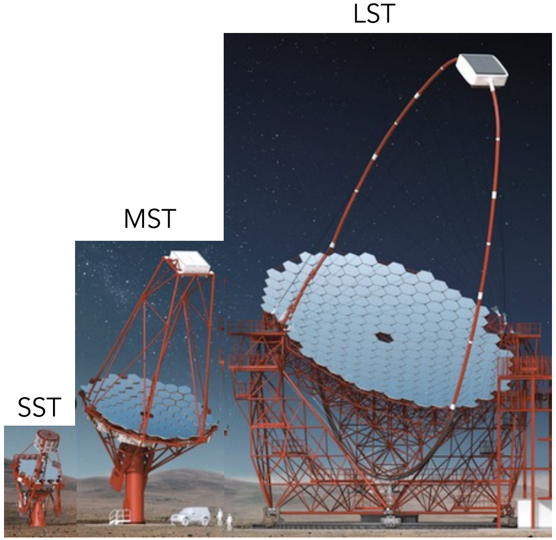
P. Grespan et al. PoS(ICRC2021) 771

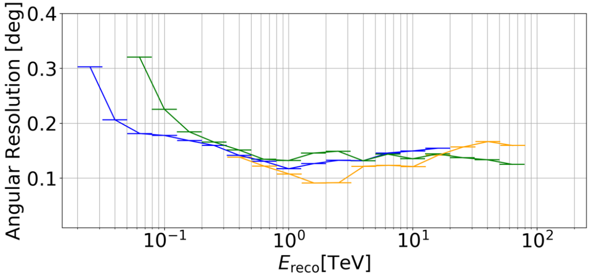
CTLearn: single-telescope full-event reconstruction

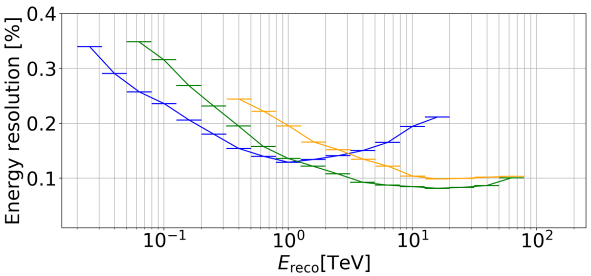






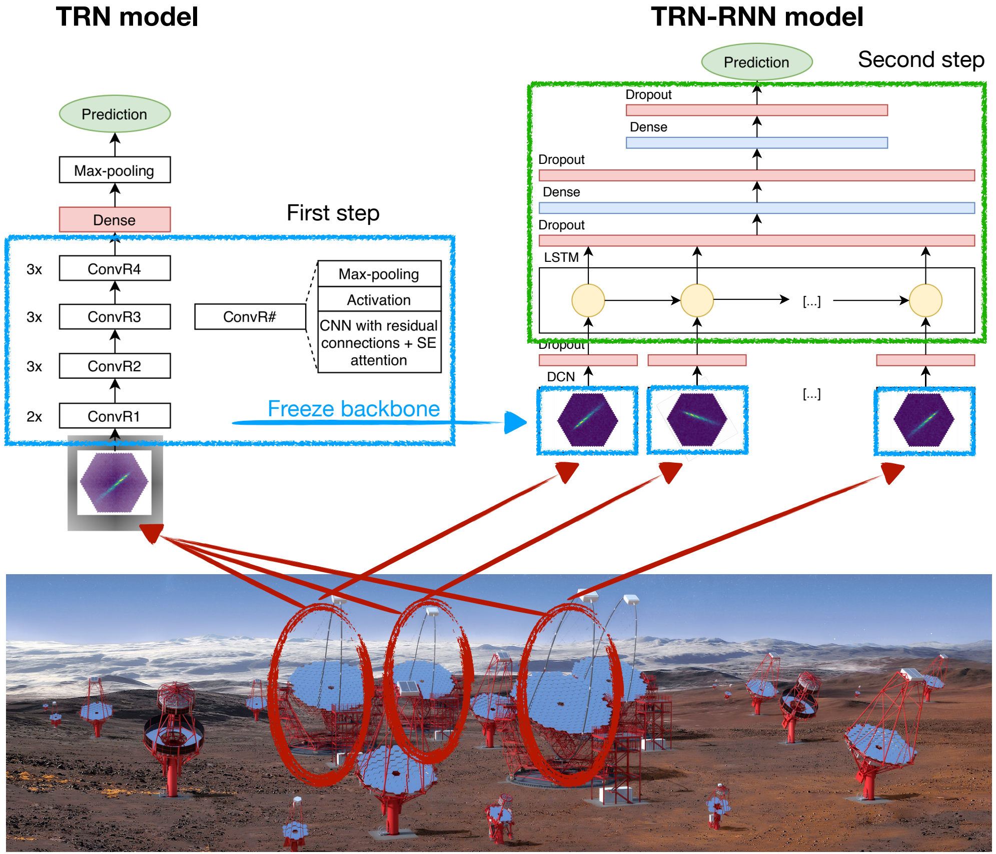






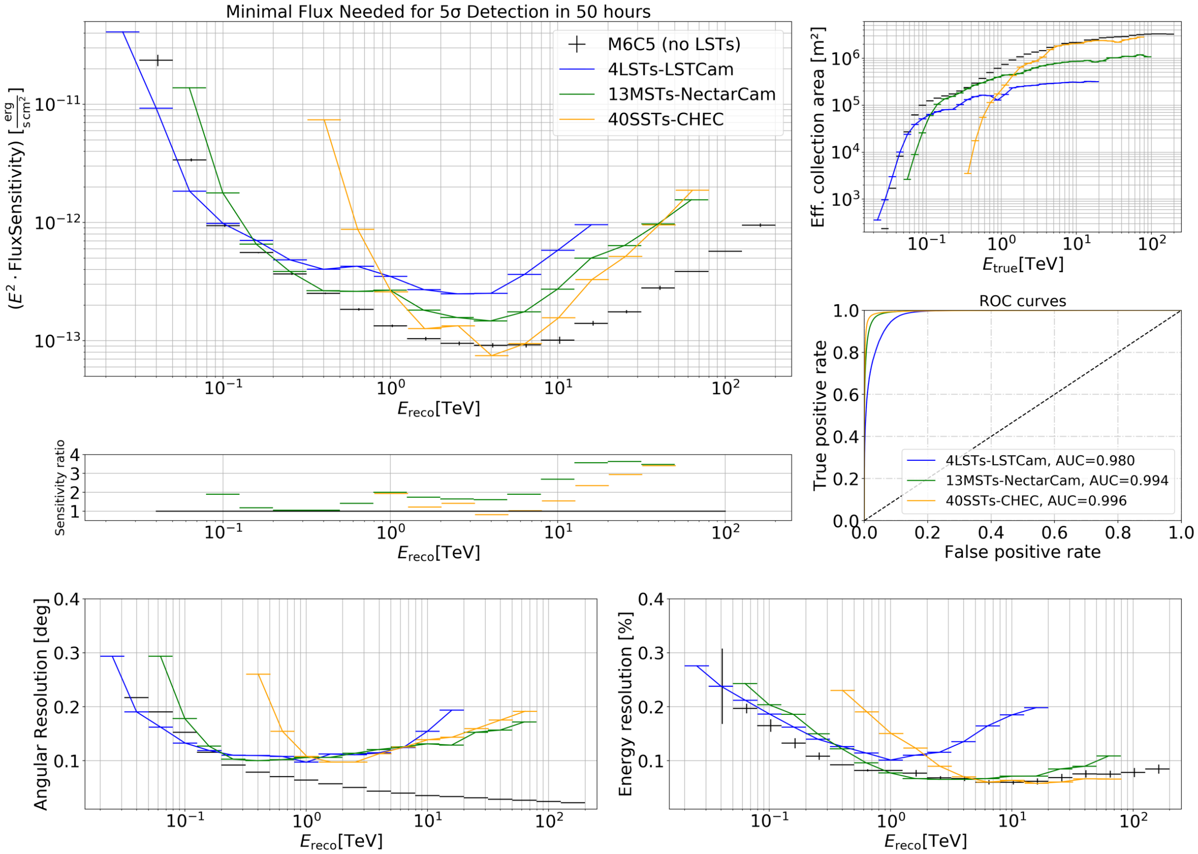
T. Miener et al., PoS(ICRC2021) 730

CTLearn: multiple-telescope full-event reconstruction



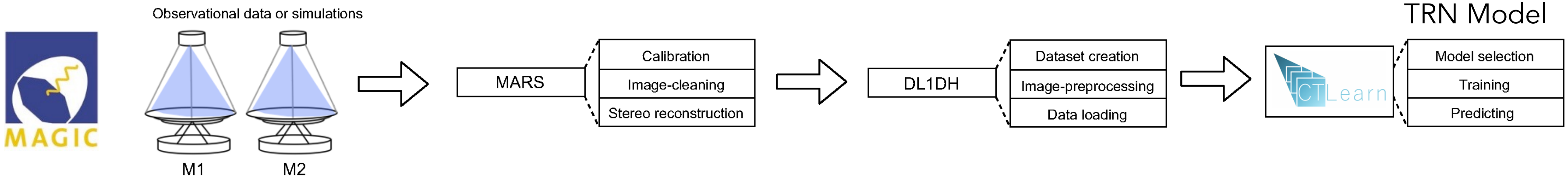
[T. Miener et al., PoS(ICRC2021) 730]

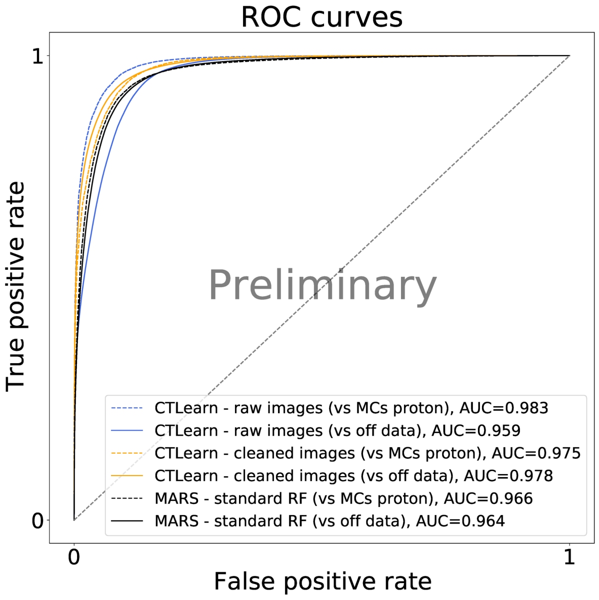
CTLearn: multiple-telescope full-event reconstruction

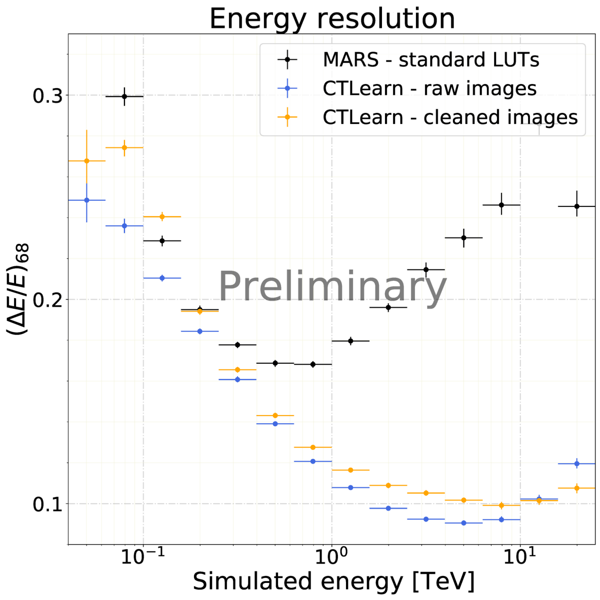


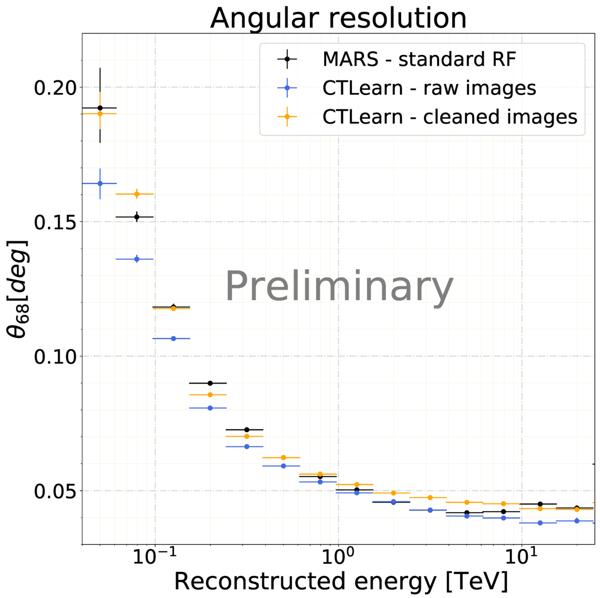
[T. Miener et al., PoS(ICRC2021) 730]

CTLearn: application to real data

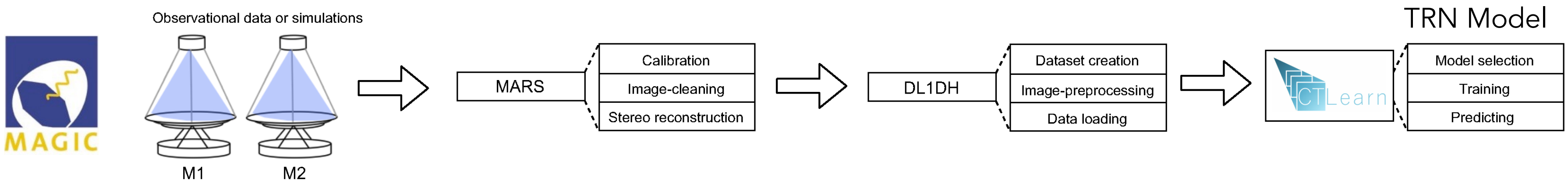


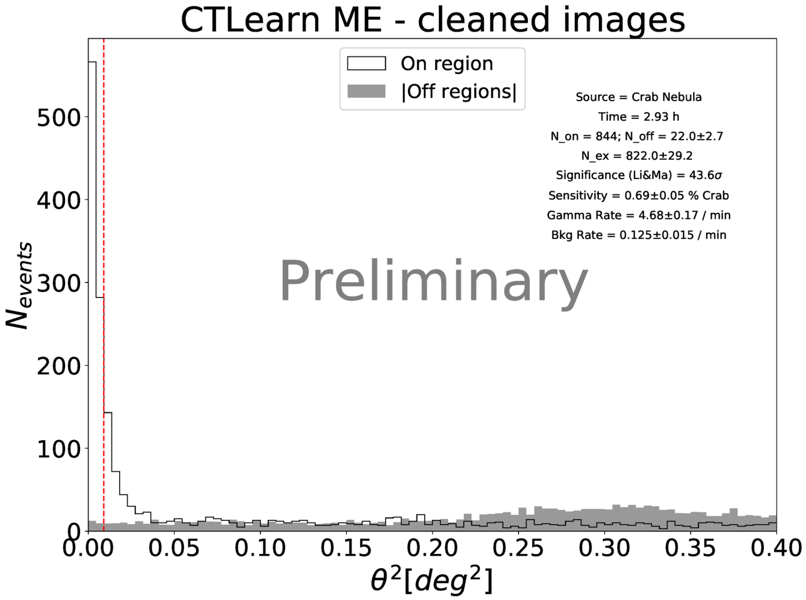






T. Miener et al. 2021 (ADASS XXXI)





Analysis	γ rate [/min]	bkg rate [/min]	Sen. [% Crab]	Sig. (Li&Ma)
MARS – ME	4.54 ± 0.16	0.119 ± 0.015	0.70 ± 0.05	43.0σ
CTLearn – ME (raw)	3.45 ± 0.14	0.133 ± 0.018	0.97 ± 0.08	36.5σ
CTLearn – ME (cleaned)	4.68 ± 0.17	0.125 ± 0.015	0.69 ± 0.05	43.6σ
MARS – LE	16.49 ± 0.35	3.861 ± 0.086	1.09 ± 0.03	61.1σ
CTLearn – LE (raw)	11.70 ± 0.32	3.832 ± 0.114	1.53 ± 0.05	47.5σ
CTLearn – LE (cleaned)	16.24 ± 0.35	3.872 ± 0.086	1.11 ± 0.03	60.4σ

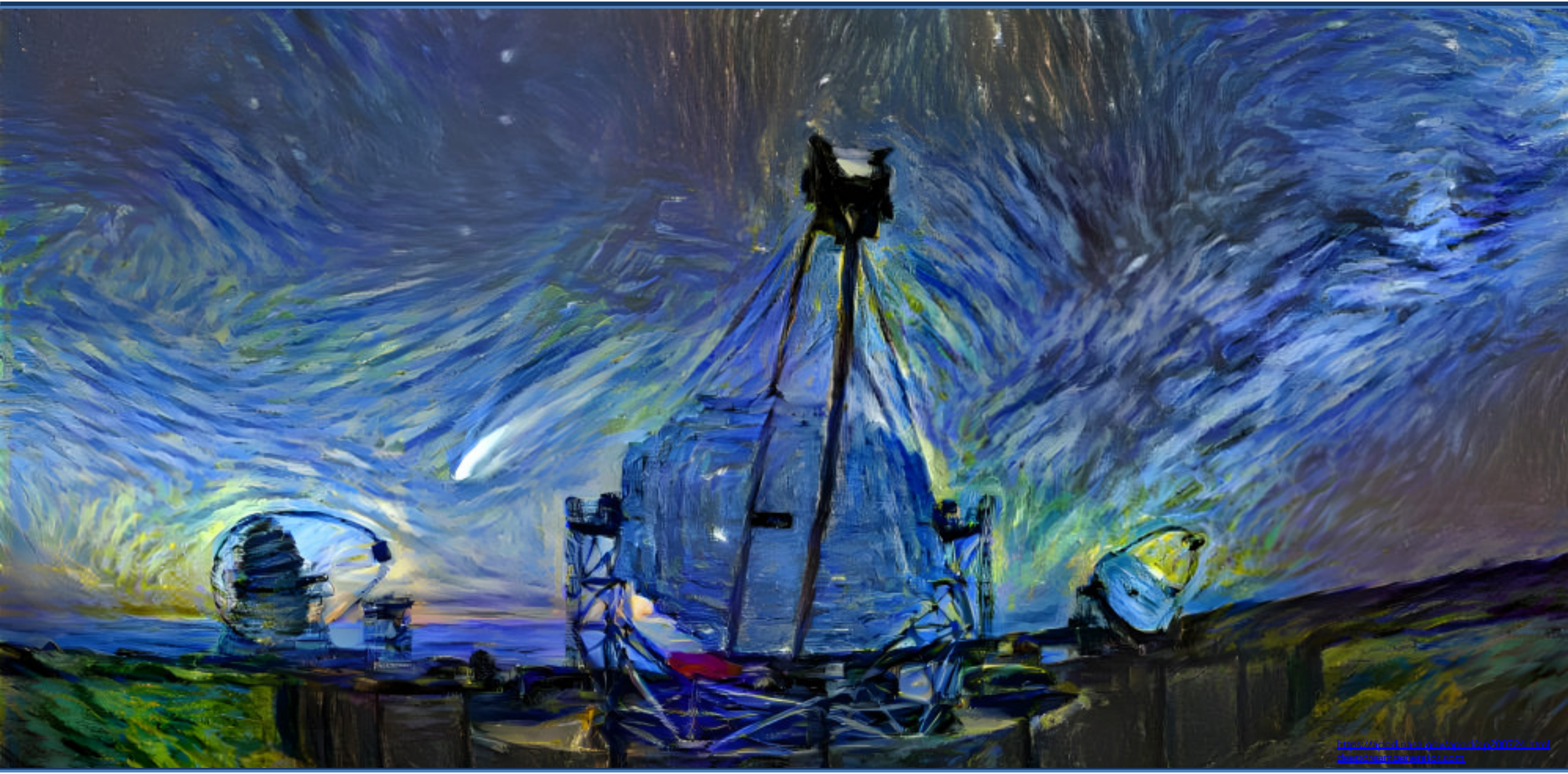
Analysis	N_{on}	N_{off}	N_{ex}
MARS – ME	819	21.0 ± 2.6	798.0 ± 28.7
CTLearn – ME (raw)	629	23.3 ± 3.1	605.7 ± 25.3
CTLearn – ME (cleaned)	844	22.0 ± 2.7	822.0 ± 29.2
MARS – LE	3579	679.0 ± 15.0	2900.0 ± 61.7
CTLearn – LE (raw)	2730	673.7 ± 20.0	2056.3 ± 56.0
CTLearn – LE (cleaned)	3536	680.7 ± 15.1	2855.3 ± 61.3

Summary of all performed analyses of the same Crab Nebula sample

T. Miener et al. 2021 (ADASS XXXI)

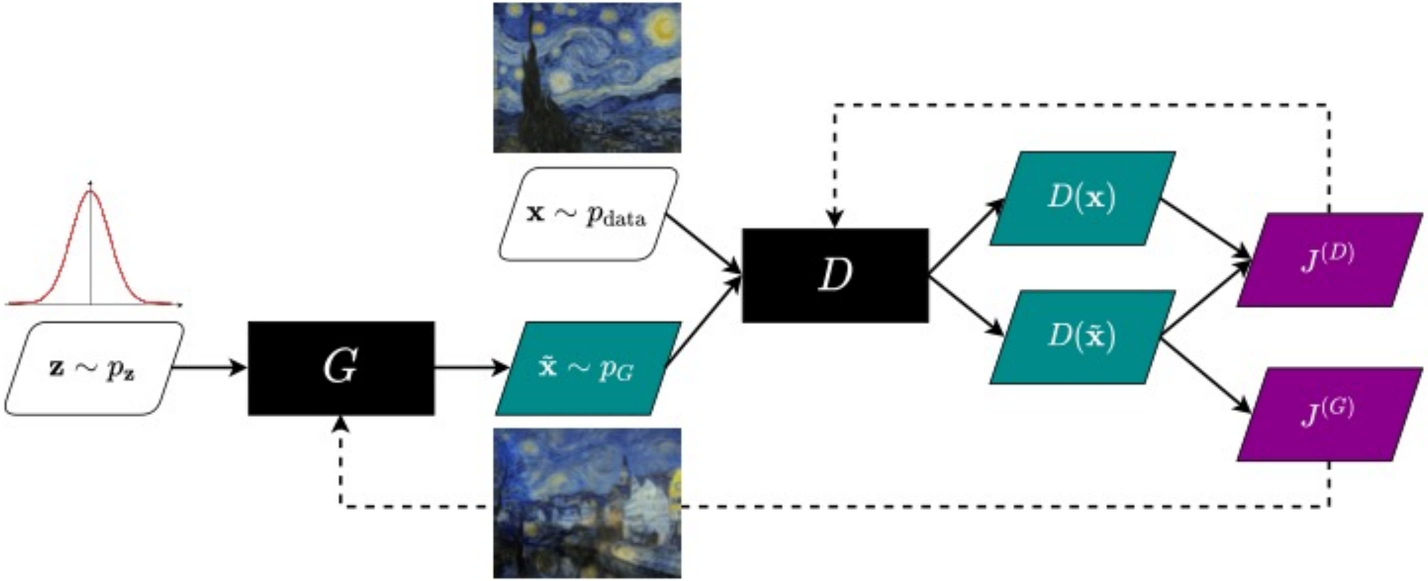
Dreaming of IACTs

Dreaming of IACTs



https://apod.nasa.gov/apod/ap200724.html
deepdreamgenerator.com

- Generative adversarial networks (GANs)

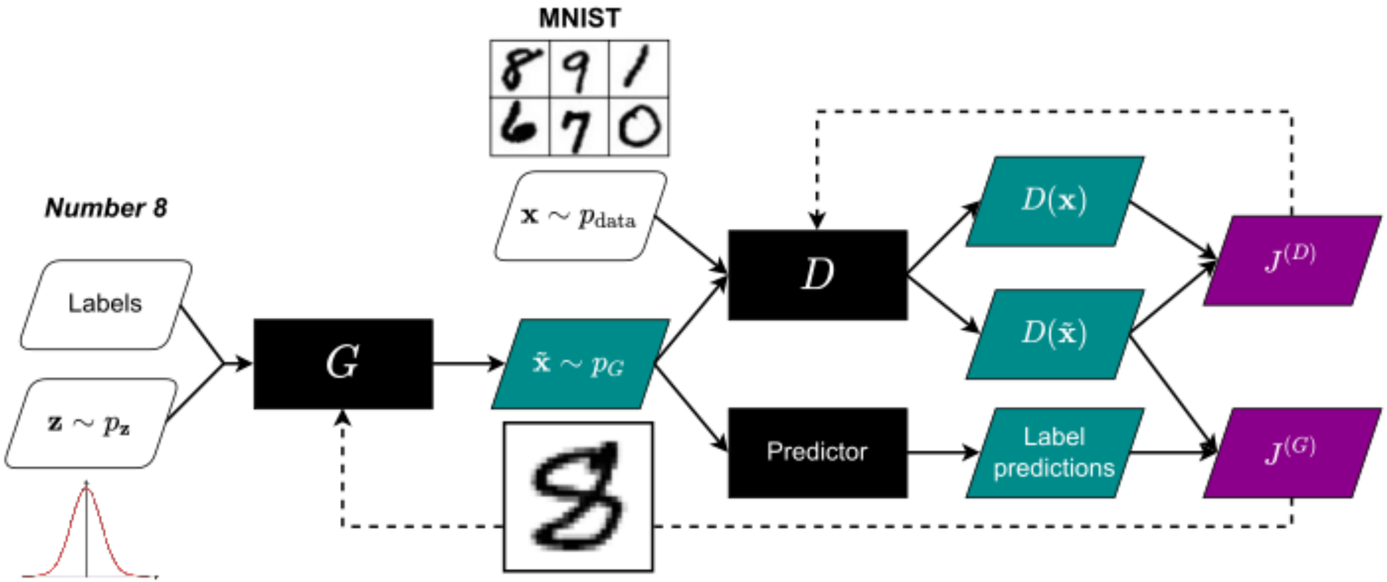


$$J^{(D)} = \mathbb{E}_{\mathbf{z}\sim p_{\mathbf{z}}}[D(G(\mathbf{z}))] - \mathbb{E}_{\mathbf{x}\sim p_{\text{datos}}}[D(\mathbf{x})]$$

$$J^{(G)} = -J^{(D)} = \mathbb{E}_{\mathbf{x}\sim p_{\text{datos}}}[D(\mathbf{x})] - \mathbb{E}_{\mathbf{z}\sim p_{\mathbf{z}}}[D(G(\mathbf{z}))] \rightarrow -\mathbb{E}_{\mathbf{z}\sim p_{\mathbf{z}}}[D(G(\mathbf{z}))]$$

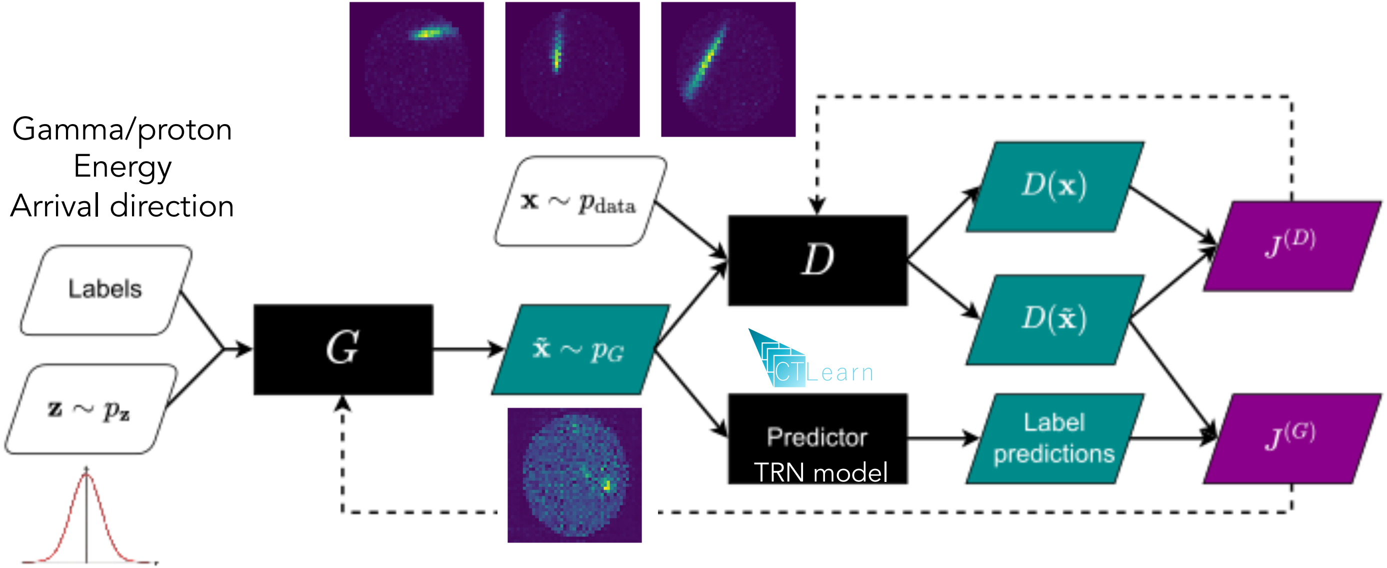
S. García-Heredia et al.

- Auxiliary conditional generative adversarial networks (AC-GANs)



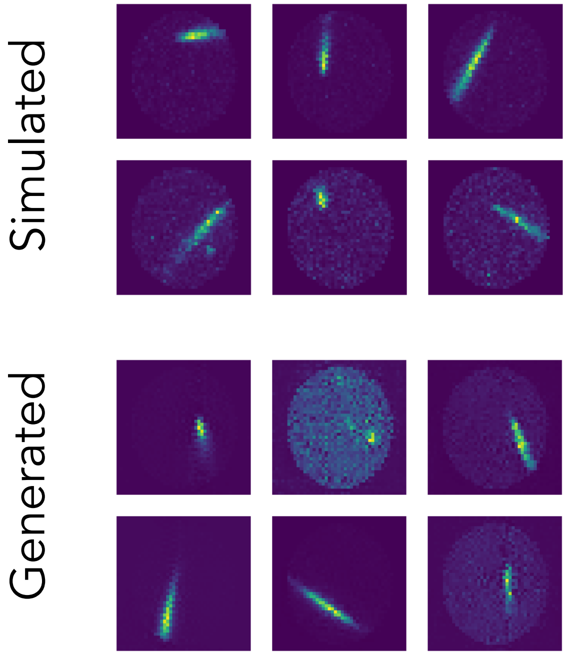
S. García-Heredia et al.

- Auxiliary conditional generative adversarial networks (AC-GANs)

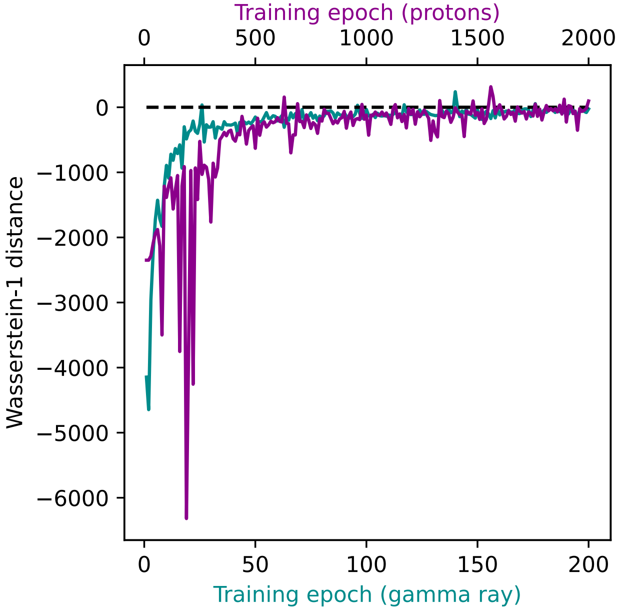


S. García-Heredia et al.

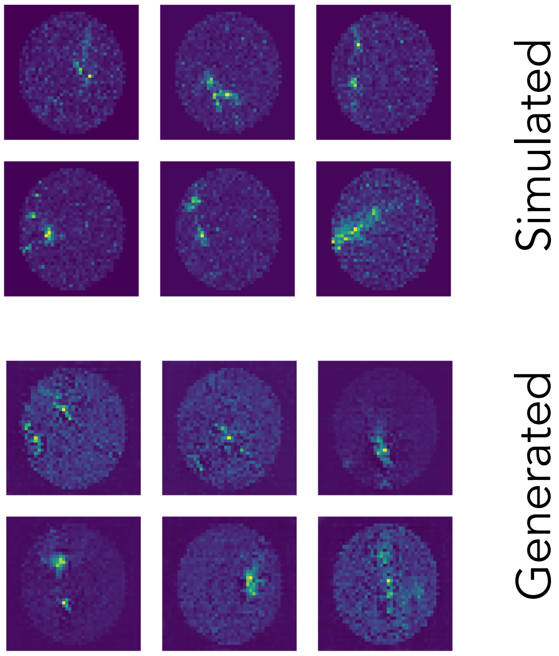
GAMMA RAYS



WASSERSTEIN-1 DISTANCE



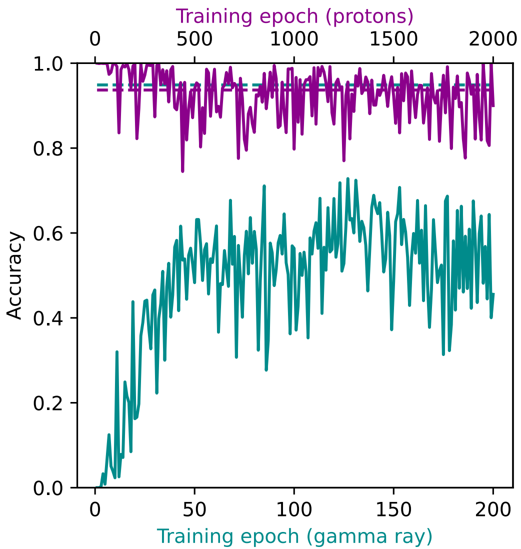
PROTONS



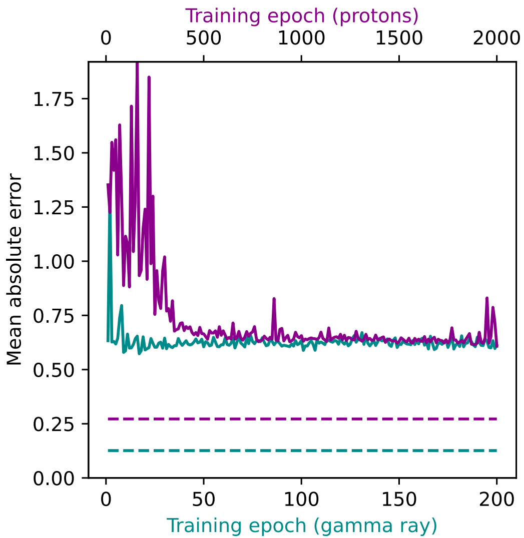
S. García-Heredia et al.

Generating IACT events with GANs

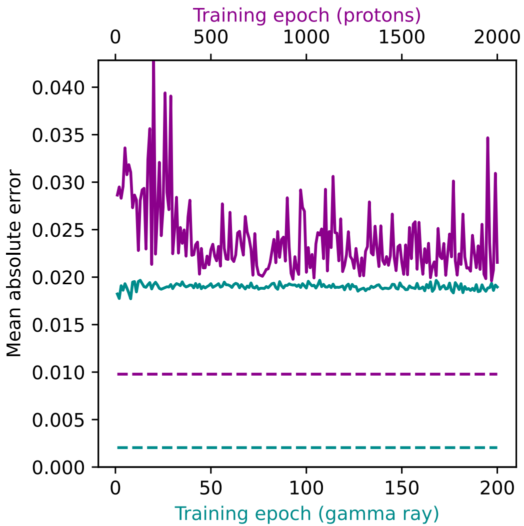
PARTICLE TYPE



ENERGY



ARRIVAL DIRECTION



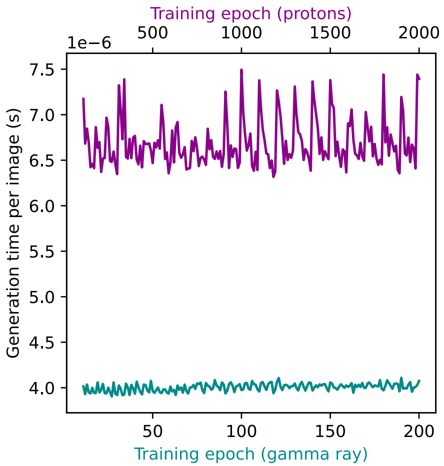
(-- Simulated and — Generated)

S. García-Heredia et al.

- Generation time

GANS

Using a GPU Nvidia 3090:



SIMULATIONS

- $\sim 1\ \mathrm{s/event}$
- Each event consists of one image for each detector
- Depends on what is being simulated and the computational capacity.

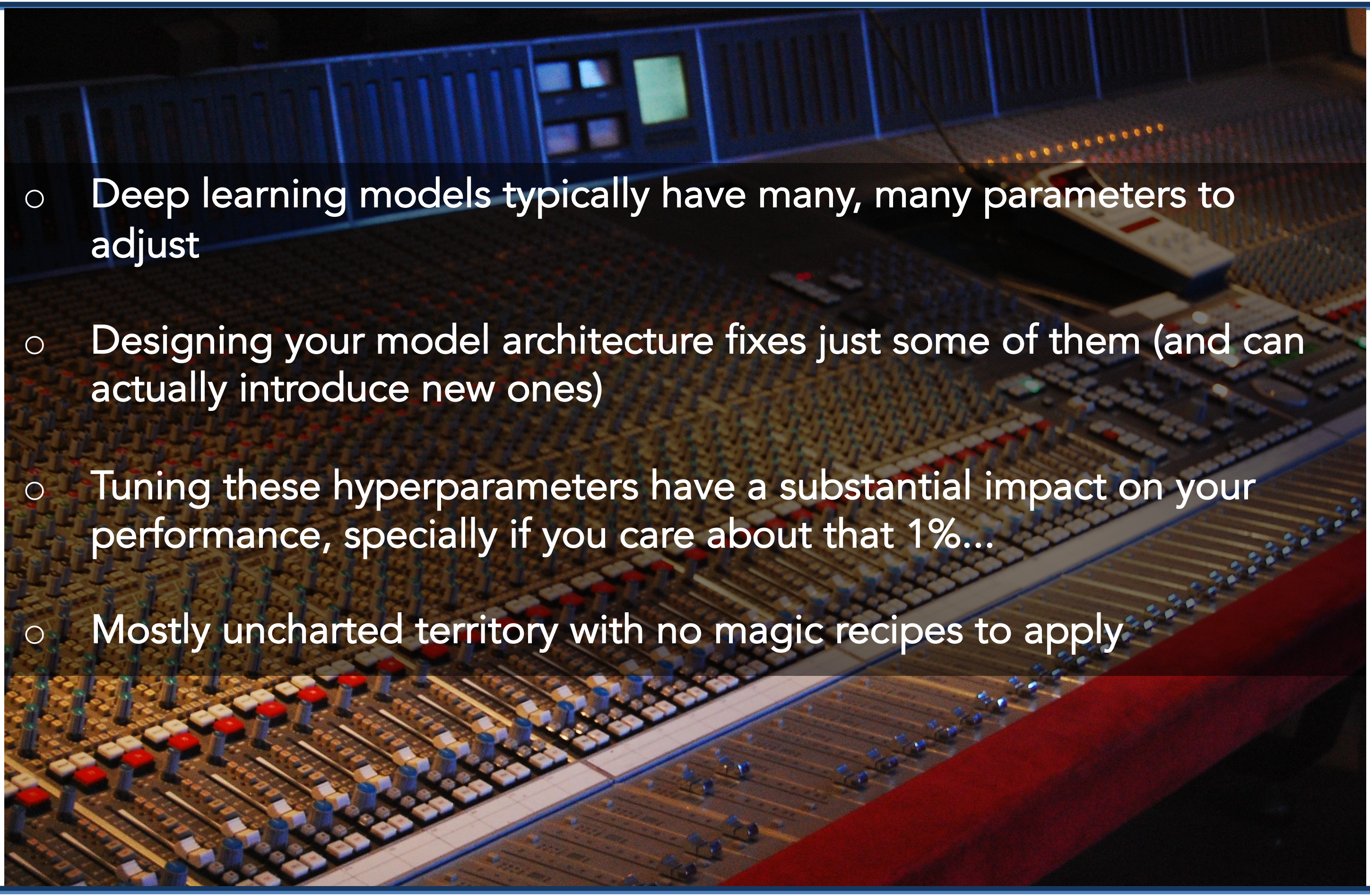
S. García-Heredia et al.

Next step -> find the best performing model for event reconstruction

The curse of dimensionality haunts us here too!

- Hyperparameter space for deep learning architecture design
  - Number of CNN layers
  - Kernel size
  - Activation function
  - Dropout rate
  - Number of FC layers
  - Batch size
  - Learning rate
  - Optimizer
  - …

- Optimization strategies
  - Grid searches
  - Random searches
  - Bayesian optimization
  - Evolutionary algorithms
  - Reinforcement learning
  - …



- Deep learning models typically have many, many parameters to adjust
- Designing your model architecture fixes just some of them (and can actually introduce new ones)
- Tuning these hyperparameters have a substantial impact on your performance, specially if you care about that 1%...
- Mostly uncharted territory with no magic recipes to apply

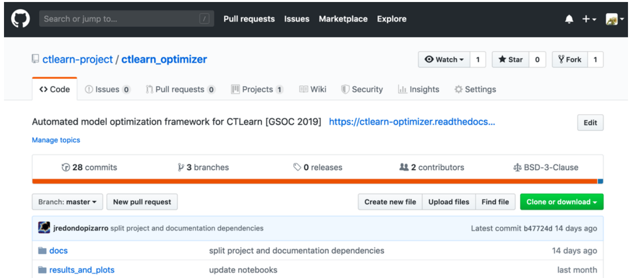
- Framework for hyperparameter optimization of CTLearn models
  (Although can be adapted to any config-file based DCN framework)

- Based on Tune: a scalable hyperparameter tuning library

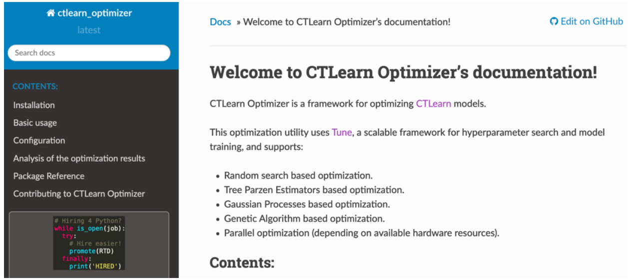
- Supported optimization strategies:
  - Random search
  - Tree Parzen Estimators
  - Gaussian Processes
  - Genetic Algorithms
  - Parallel optimization (depending on available hardware)

Tree Parzen Estimators, Gaussian Processes, Genetic Algorithms: Bayesian optimization

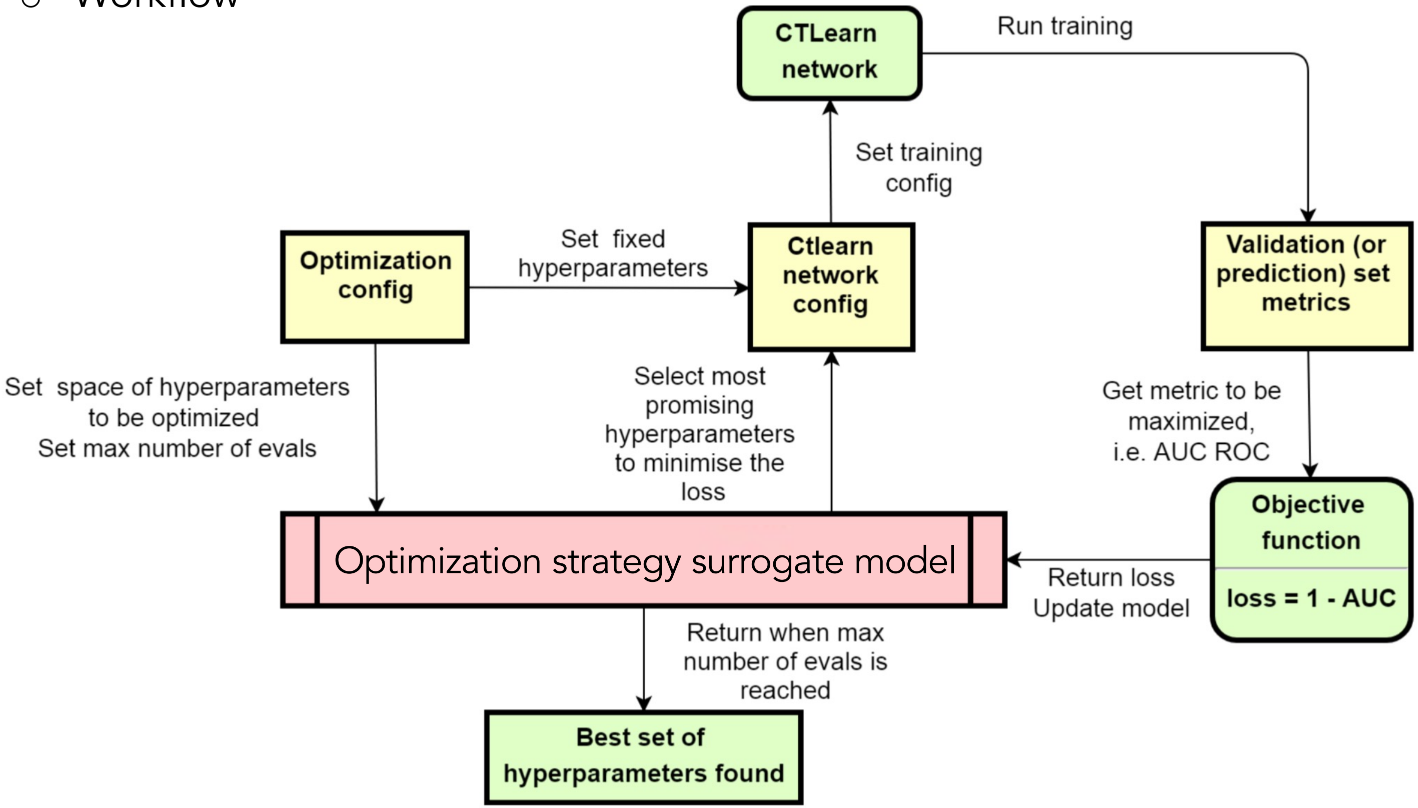
github.com/ctlearn-project/ctlearn_optimizer



ctlearn-optimizer.readthedocs.io

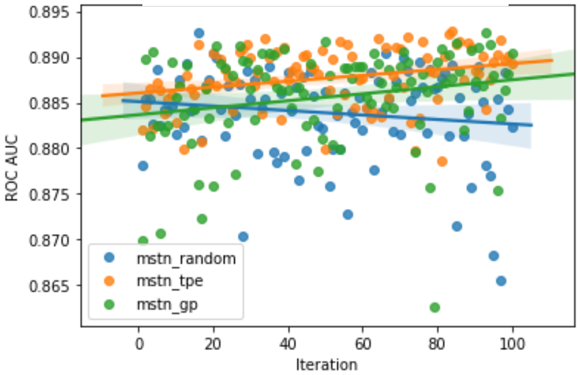


- Workflow

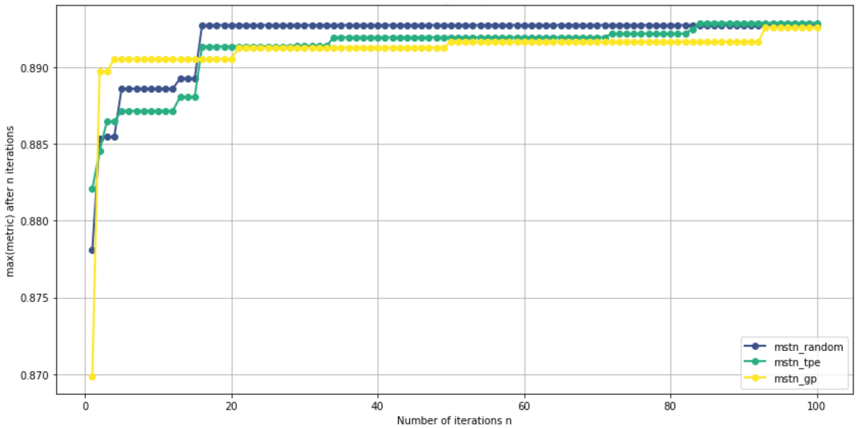

CTLearn network
Run training
Set training config
Optimization config
Set fixed hyperparameters
Ctlearn network config
Validation (or prediction) set metrics
Set space of hyperparameters to be optimized
Set max number of evals
Select most promising hyperparameters to minimise the loss
Get metric to be maximized, i.e. AUC ROC
Objective function
loss = 1 - AUC
Optimization strategy surrogate model
Return loss
Update model
Return when max number of evals is reached
Best set of hyperparameters found

- Visualization

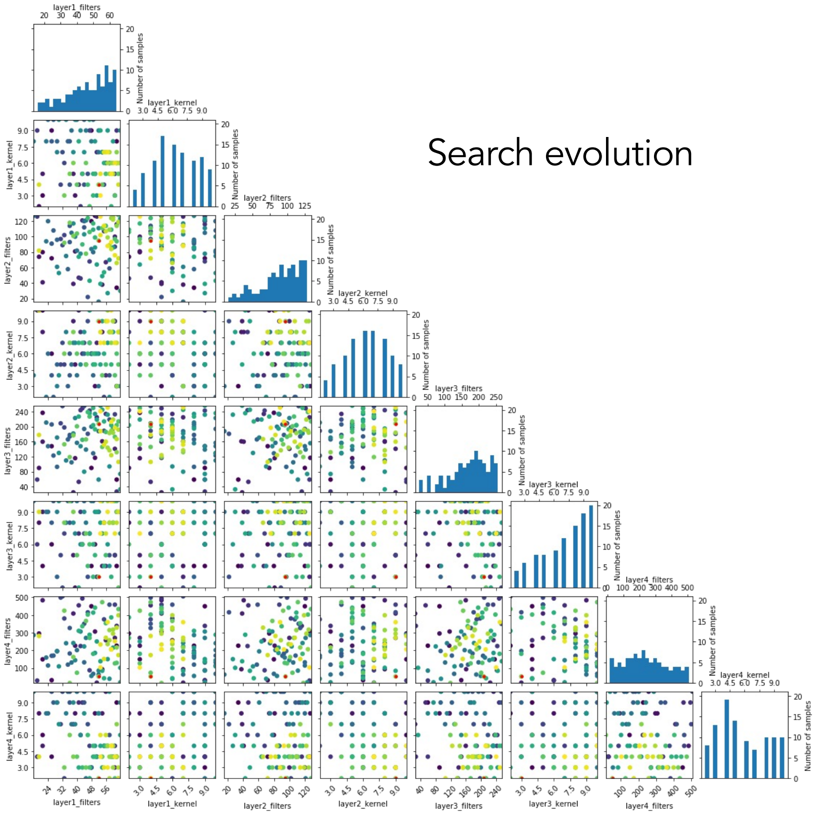
Evolution of the metric



Convergence of the metric

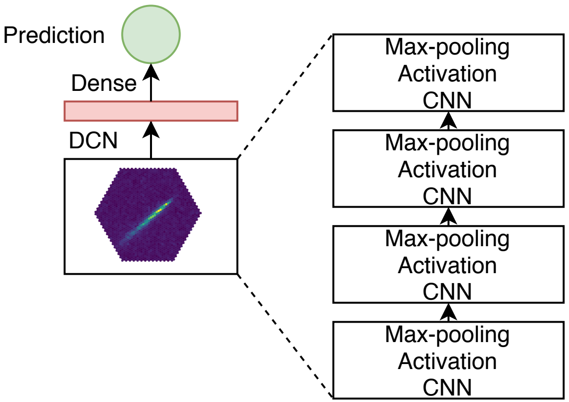


Search evolution



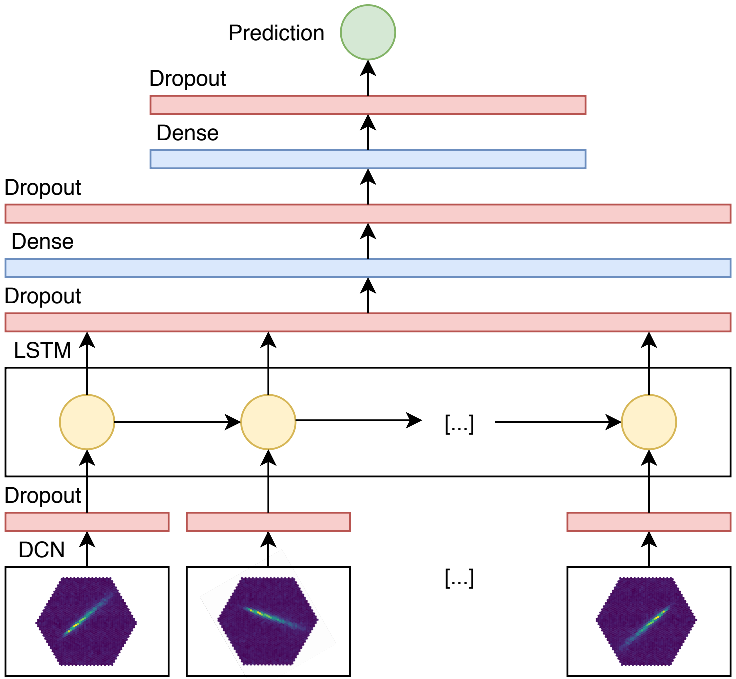
CTLearn Optimizer: some results

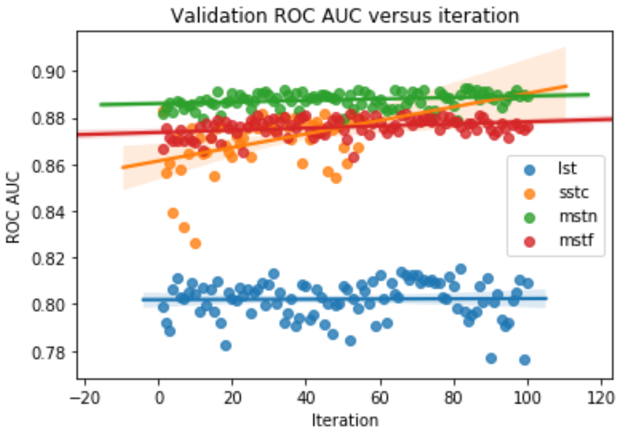
Single-tel model

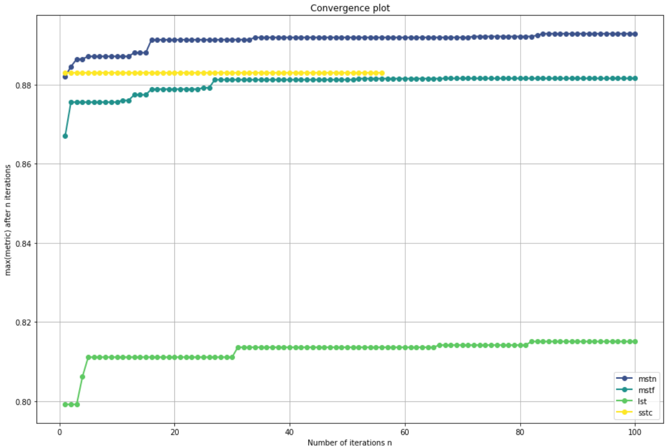


Number of kernels
Size of kernels
Number of CNN layers

CNN-RNN model





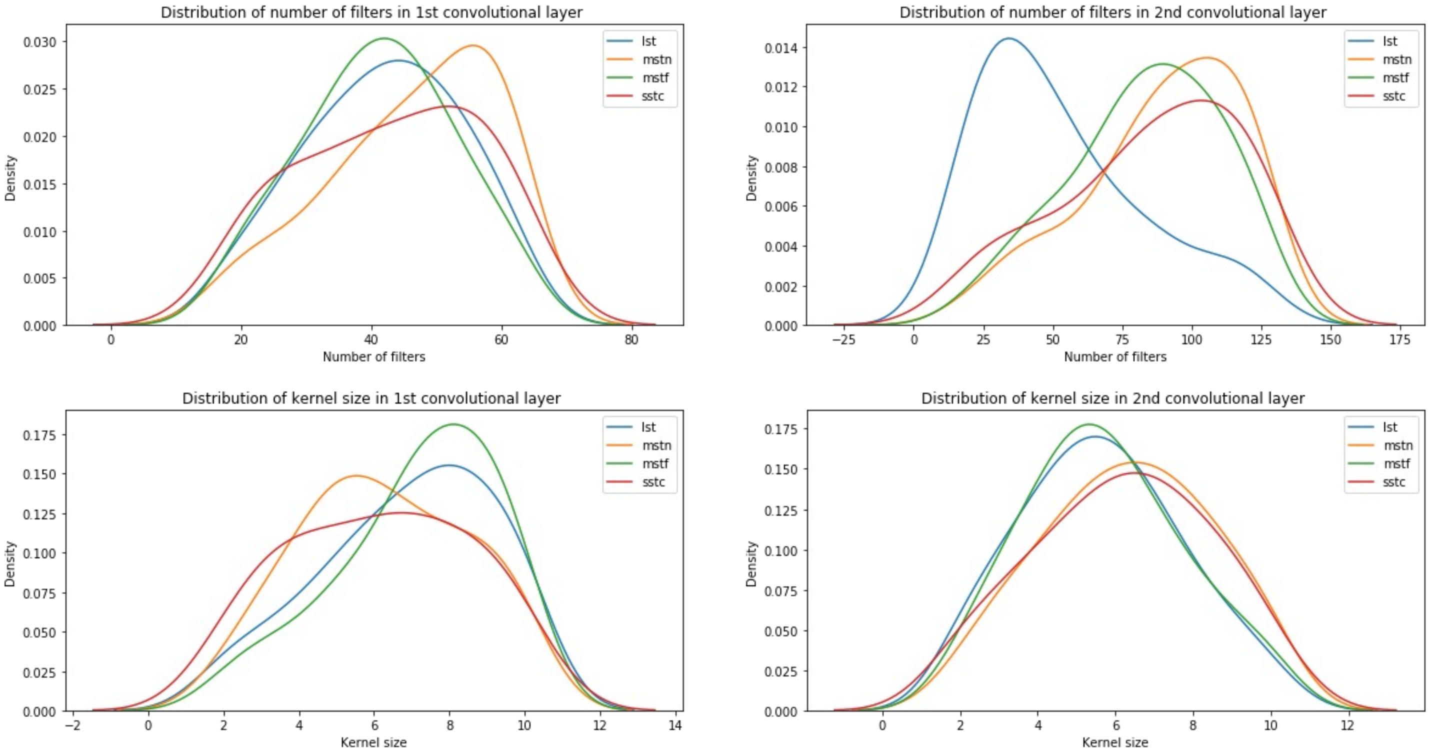


Hyperparameters	Telescope Type	Validation Accuracy	Validation AUC	Training Time
Base	LST	70.38%	0.7887	0h 41m 22s
Optimized	LST	72.45%	0.8150	0h 39m 14s
Base	SSTC	73.90%	0.8118	0h 42m 4s
Optimized	SSTC	79.87%	0.8830	1h 16m 4s
Base	MSTN	78.04%	0.8659	0h 58m 10s
Optimized	MSTN	80.11%	0.8929	0h 52m 48s
Base	MSTF	74.60%	0.8360	0h 55m 0s
Optimized	MSTF	79.01%	0.8816	0h 48m 37s

Telescope Type	Metric	Improvement
LST	Validation Accuracy	2.07%
LST	Validation AUC	2.63%
SSTC	Validation Accuracy	5.97%
SSTC	Validation AUC	7.12%
MSTN	Validation Accuracy	2.07%
MSTN	Validation AUC	2.70%
MSTF	Validation Accuracy	4.41%
MSTF	Validation AUC	4.56%

CTLearn Optimizer: some results

- Single_tel & TPE search



Optimized hyperparameters seem to be telescope-type dependent

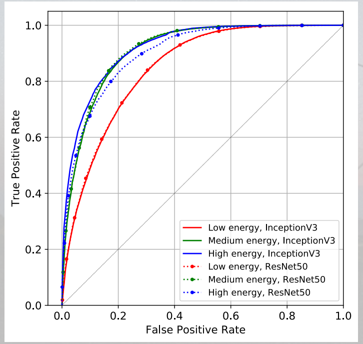
- Single_tel & TPE search: transfer to CNN-RNN

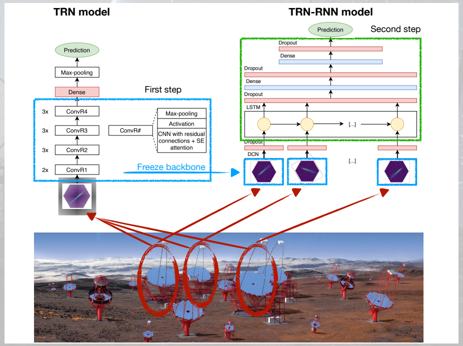
Hyperparameters	Telescope Type	Validation Accuracy	Validation AUC	Training Time
Base	LST	73.43%	0.8285	0h 41m 22s
Optimized	LST	74.96%	0.8422	0h 46m 53s
Base	SSTC	80.64%	0.9072	1h 51m 5s
Optimized	SSTC	83.49%	0.9217	3h 31m 43s
Base	MSTN	83.10%	0.9169	2h 15m 52s
Optimized	MSTN	84.20%	0.9313	6h 43m 14s

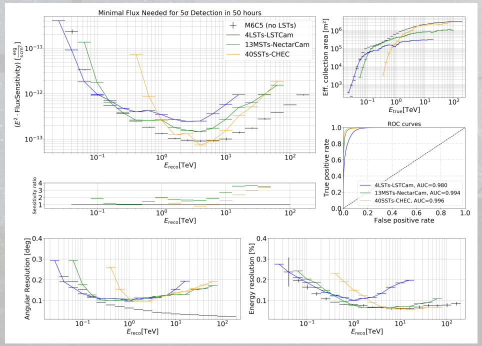
Telescope Type	Metric	Improvement
LST	Validation Accuracy	1.53%
LST	Validation AUC	1.37%
SSTC	Validation Accuracy	2.85%
SSTC	Validation AUC	1.45%
MSTN	Validation Accuracy	1.10%
MSTN	Validation AUC	1.44%

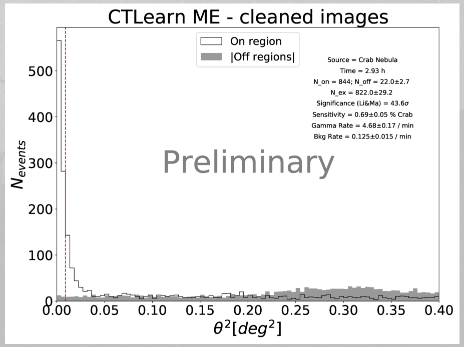
Conclusions & Outlook

- Current-generation IACTs have enhanced their performances through ML
- Next-gen (even current-gen!) IACT may profit from latest developments in ML
- Ongoing efforts to exploit deep learning as an event reconstruction method for IACTs
  - Full-event reconstruction over simulated IACT events demonstrated
  - Application to real observations works!
  - Working on optimizing architectures & multi-task learning
  - Using AC-GANs as pseudosimulators
  - Tackling the real-data problem









Acknowledgment

The research here presented has been partially supported by the former Spanish Ministry of Economy, Industry, and Competitiveness / ERDF grants FPA2015-73913-JIN and FPA2017-82729-C6-3-R, the Spanish Ministry of Science and Innovation grant PID2019-104114RB-C32, NSF awards PHY-1229205, 1229792, and 1607491, and the European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures funded by the European Union's Horizon 2020 research and innovation program under Grant Agreement no. 824064. The authors acknowledge support from Google LLC through the Google Summer of Code program and NVIDIA Corporation with the donation of a Titan X Pascal GPU used for part of this research.