Analyze and Optimize pipeline for Cherenkov Telescope SST-1M

Hugo Varenne

Master of Science HES-SO in Engineering, Major in Data Science

Under the supervision of Upegui Posada Andres, In collaboration with HEPIA and UNIGE

Introduction

Gamma rays are produced when cosmic rays interact with matter, radiation, or magnetic fields in extreme astrophysical environments, such as supernovae or black holes. When these gamma rays enter the Earth's atmosphere, they cannot be detected directly and instead generate extensive air showers, emitting faint Cherenkov light as secondary particles travel faster than the speed of light in air. This Cherenkov light is captured by Imaging Atmospheric Cherenkov Telescopes (IACTs), which record its spatial and temporal patterns as images and waveforms. These images are then analyzed using machine learning techniques to reconstruct the primary particle type, energy, and arrival direction.

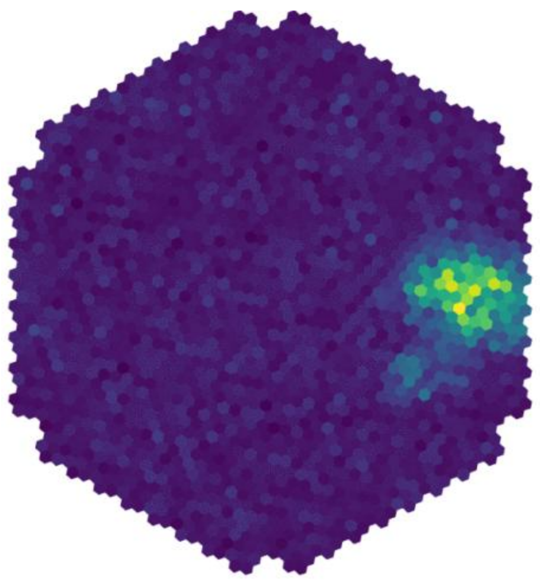


Figure 1.
IACT image

Figure 2.
SST-1M telescope

Material and methods

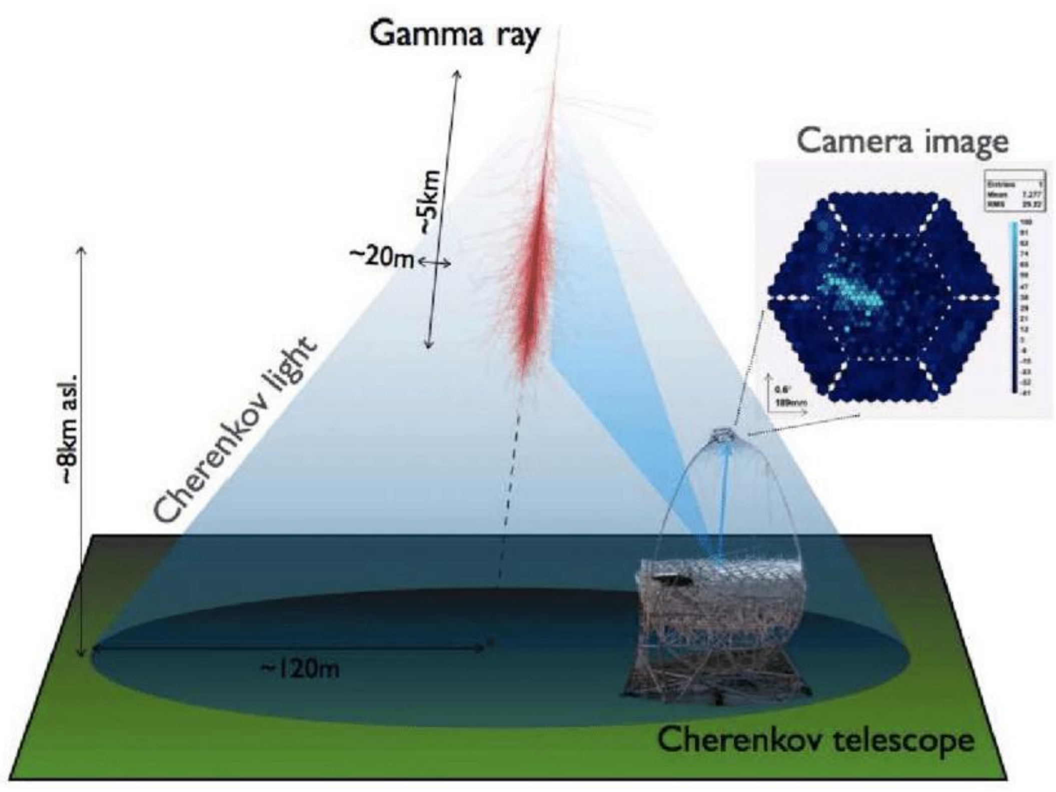
- **CTLearn** is a Python library that applies deep learning to analyze images from Cherenkov telescopes. By helping discriminate gamma-ray events from other particles, it also determines both the energy level and the source of the recorded Cherenkov light

- **MLOps** (Machine Learning Operations) streamlines the data pipeline by automating the transition of collected data through stages such as preprocessing, model training, experiment tracking, and deployment. This ensures that machine learning models are easier to develop, reproduce, and deploy, especially with large datasets on HPC infrastructures.

- An **HPC (High-Performance Computing) cluster** consists of multiple CPUs and GPUs, forming a system capable of running large and computationally intensive tasks in parallel. This architecture enables efficient training of deep learning models on large datasets.

- **Data:** Simulated gamma-ray events → stored in HDF5 → includes calibrated waveforms, image features, reconstructed events & ground truth → used to train models for particle type, energy & direction.

Figure 3. Gamma ray interaction with the atmosphere captured by a Cherenkov telescope



Methodological Contributions and Implemented Tools

- **MLOps Tools:** Modular tasks for training, testing, report generation, and model comparison enable independent execution and fault tolerance. Configuration files manage experiment parameters for reproducibility. Reports and visualizations provide metrics and performance insights. The setup supports automation, monitoring, and scalable, production-ready model management.

- **Custom Model:** A template has been integrated into CTLearn to build custom models, separating backbone (high-level features) and head (task-specific predictions). This structure allows fine-tuning, reusability, and integration of pretrained weights while ensuring compatibility with the library.

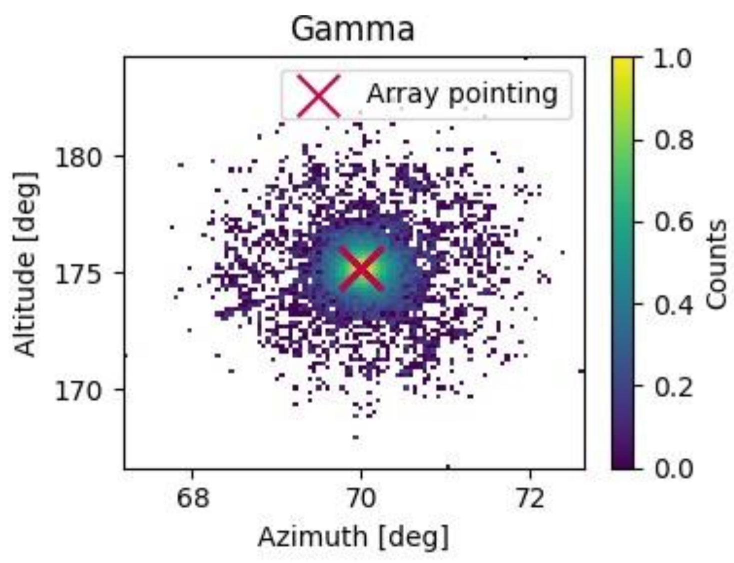


Figure 4.
Altitude-Azimuth distribution

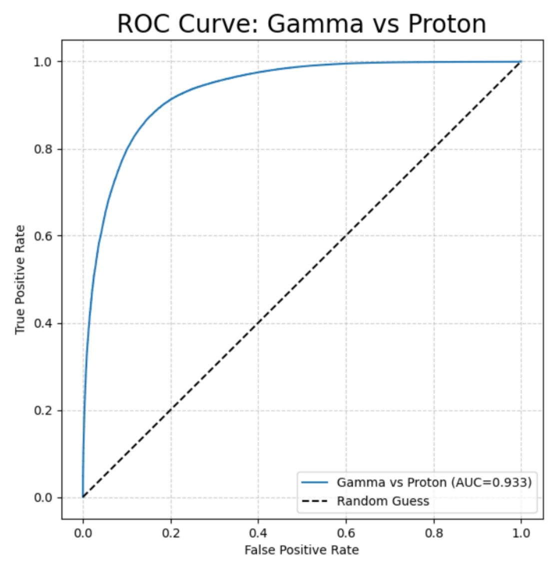


Figure 5.
ROC Curve

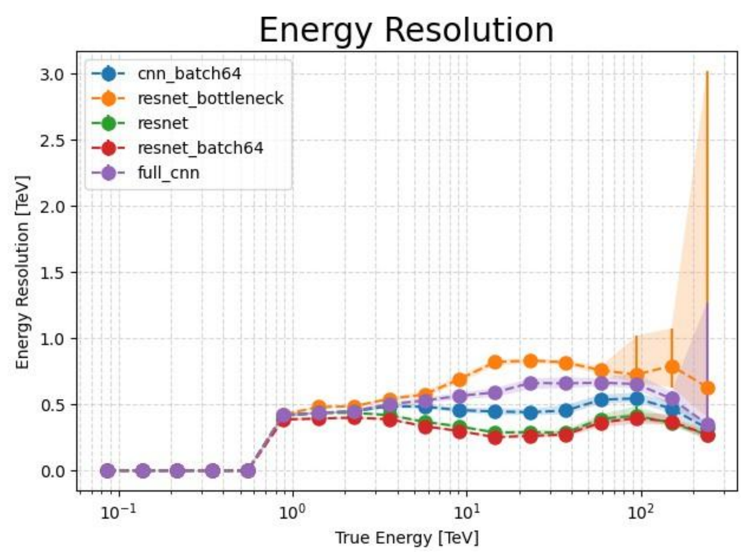


Figure 6. Comparison of model for the energy resolution

Conclusion

Modern MLOps practices greatly enhance deep learning workflows for gamma-ray reconstruction with Cherenkov Telescopes. By integrating automation, configuration-driven experiments, standardized reporting, and scalable HPC workflows into the CTLearn framework, the thesis bridges astrophysics expertise with machine learning engineering. Beyond model performance, the results highlight that reproducibility, scalability, and maintainability are critical for complex deep learning architectures in scientific environments.

Overall, this project serves as a proof of concept for MLOps' role in enabling sustainable model development within the CTLearn ecosystem.

Acknowledgments

I would like to thank all those involved in the CTLearn library, especially Andres Upegui Posada, Matthieu Heller, Laurent Gantel, and Jakub Kvapil, who contributed to this project with their guidance, feedback, and support. Their contributions had a positive impact on this thesis and its results.

References

1. SST-1M consortium, https://sst-1m.space/
2. CTLearn, https://arxiv.org/abs/1912.09877, Research paper about the Python library
3. Ultra-High-Energy Gamma-Ray Astronomy, **https://doi.org/10.1146/annurev-nucl-112822-025357**