Implementation of new Convolutional Neural Network methods for Imaging Atmospheric Cherenkov Telescope analysis and application to the Large-Sized Telescope of CTA

Supervisor: Dott. Simone Bianco

Co-Supervisor: Dott. Rubén López Coto

Co-Supervisor: Drss. María Isabel Bernardos Martín

Master Dissertation of:
Luca Romanato
Matricola 807691

Academic Year 2019-2020

Curiosity is the essence of human existence

– Gene Cernan

Dedicated to my family.

Abstract

Gamma rays are electromagnetic radiation with millions of times higher energy than visible light. These energies can only be reached by the most violent non-thermal events in the Universe. Unlike cosmic rays; charged particles coming from all directions of the universe at relativistic velocities, gamma rays have a null electric charge which makes them immune to magnetic fields and their deflection effects, leading them to directly point their source. Our planet is constantly struck by all of these rays which interact with atmosphere molecules, creating a cascade of ultra-relativistic particles and generating a blue glow: the Cherenkov light. Imaging Atmospheric Cherenkov Telescopes (IACTs), such as the CTA project's telescopes, are used to detect and observe this phenomenon on Earth's surface.

To optimize the detection of gamma rays (γ rays), one of the telescope's priorities is the ability to distinguish images produced by γ rays from those produced by other cosmic rays (hadrons), being the signal to background ratio lower than $1/1000$. Nowadays, machine learning algorithms, like Random Forest, are used to perform the γ/hadron separation and reconstruction of energy and arrival direction of γ rays. Unfortunately, the large amount of data, in the order of Tera/Petabytes, and the not sufficient computation capabilities of hardware, leads to the necessity of parameterizing the telescope recorded images prior of being analyzed by algorithms. This implies a consequent and inevitable loss of the pixel-wise information. Luckily, thanks to the latest developments in hardware devices, such as Graphic Processing Units (GPUs) and to the implementation of the latest deep learning techniques, such as Convolutional Neural Networks (CNNs), we are nowadays able to process big amounts of data at the pixel-wise level without any loss of information. CNNs architectures are inspired by the organization of the brain visual cortex, imitating its behavior, where a single neuron responds to stimuli only in a restricted region of the entire visual field. Its coverage is entirely due to the overlaps created by the full collection of neurons. Thanks to this feature, CNNs allow to work with raw images and native information by means of special layers called Convolutional and Pooling layers. All of this happens without requiring a strong preprocessing phase such as imaging parametrization, making CNNs the ideal candidate for this analysis process.

In this thesis I propose two new reconstruction algorithms, based on a full use of CNNs trained and evaluated in GPUs, allowing to speed up the decision and regression operations while dealing with a greater amount of information. This strategy will lead to an improvement in the performance of a particular telescope type of the CTA project: the Large Size Telescopes (LSTs).

Sommario

I raggi gamma sono radiazioni elettromagnetiche con un'energia milioni di volte più elevata della luce. Queste energie possono essere raggiunte solamente da processi non termici presenti negli eventi più violenti dell'Universo. Diversamente dalla maggior parte dei raggi cosmici, elettricamente carichi, i raggi gamma sono composti da particelle con carica nulla, detti fotoni, che rendono la loro traiettoria immune agli effetti di attrazione e repulsione dei campi magnetici presenti nel Cosmo. Questa caratteristica è anche il motivo principale del loro studio in quanto è possibile risalire alla loro sorgente. Il nostro pianeta è costantemente colpito da tutti questi raggi che, interagendo con la nostra atmosfera, creano una cascata di particelle ultra relativistiche dando vita ad un bagliore blu, chiamato luce Cherenkov. Per catturare questo fenomeno sul suolo terrestre gli scienziati si servono dei Imaging Atmospheric Cherenkov Telescopes (IACTs), come i telescopi del progetto Cherenkov Telescope Array (CTA).

Per studiare al meglio le proprietà delle sorgenti gamma una delle priorità dei telescopi è la capacità di distinguere eventi prodotti da un raggio gamma rispetto a eventi prodotti dalle restanti tipologie di raggi cosmici, identificati come adroni. Questa abilità è cruciale in quanto, sfortunatamente per l'astronomia gamma, il ratio di osservazioni segnale/rumore è minore di 1/1000. Oggigiorno, per la separazione gamma/adrone, la ricostruzione dell'energia e della direzione d'arrivo dell'evento vengono utilizzati algoritmi di Machine Learning, come Random Forest. Questa scelta è stata dettata dalla grande mole di dati, nell'ordine dei Tera/Petabyte, e le prestazioni hardware non ottime per una computazione accettabile, che hanno richiesto la parametrizzazione delle immagini registrate dal telescopio prima di essere analizzate dagli algoritmi. Purtroppo, come per ogni parametrizzazione, si ha una conseguente ed inevitabile perdita di informazioni contenute nei pixel. Tuttavia, con gli ultimi sviluppi della componentistica hardware e dell'intelligenza artificiale, sono state realizzate nuove unità di calcolo più veloci, come le Graphics Processing Units (GPUs), ed implementati nuovi algoritmi più efficaci nel risolvere problemi complessi, come le Convolutional Neural Networks (CNNs). Quest'ultime hanno un'architettura ispirata alla corteccia visiva del cervello ed un comportamento che tende ad imitarla, dove ogni singolo neurone risponde agli stimoli di un'area ristretta della visuale totale che, anche attraverso delle sovrapposizioni, viene mappata dall'intero insieme di neuroni. Grazie a questa caratteristica le CNNs permettono di lavorare su immagini grezze tramite particolari strati di neuroni detti convoluzionali e di pooling, manipolando direttamente l'informazione nativa senza richiedere una fase di preprocessing importante, come la parametrizzazione. Questa loro peculiarità le rende il candidato ideale per questa tipologia di analisi.

In questa tesi propongo due nuove architetture, basata sul completo utilizzo delle CNN combinato all'utilizzo di GPU. Ciò permetterà di velocizzare le operazioni di decisione e regressione trattando, contemporaneamente, una maggiore quantità di informazioni.

Questa strategia porterà un miglioramento nelle performance di una particolare tipologia di telescopio del progetto CTA, i Large Size Telescopes(LSTs).

Acknowledgements

Luca Romanato
Università degli Studi di Milano-Bicocca
March 7, 2021

Contents

List of Figures

List of Tables

Acronyms

Adam ADAptive Moment estimation.

AGN Active Galactic Nuclei.

AI Artificial Intelligence.

ANN Artificial Neural Networks.

AUC Area Under Curve.

Az Azimuth.

BH Black Hole.

BN Batch Normalization.

CNN Convolutional Neural Network.

CO Carbon monoxide.

CR Cosmic Ray.

CTA Cherenkov Telescope Array.

DL Deep Learning.

DL1DH DL1 Data Handler.

EAS Extended Air Shower.

EM Electromagnetic.

FC Fully Connected.

FoV Field of View.

GPU Graphics Processing Unit.

GRB Gamma-Ray Burst.

H.E.S.S. High Energy Stereoscopic System.

HAWC High-Altitude Water Cherenkov.

HEGRA High Energy Gamma Ray Astronomy.

IACT Imaging Atmospheric Cherenkov Telescope.

INFN Istituto Nazionale di Fisica Nucleare.

ISM Interstellar Medium.

LAT Large Area Telescope.

LR Learning Rate.

LST Large-Sized Telescope.

LUT Look Up Table.

MAGIC Major Atmospheric Gamma-ray Imaging Cherenkov.

MC Monte Carlo.

ML Machine Learning.

MST Medium-Sized Telescopes.

NASA National Aeronautics and Space Administration.

NS Neutron Star.

NSB Night Sky Background.

PMT Photomultiplier Tubes.

PWN Pulsar Wind Nebula.

ReLU Rectified Linear Unit.

RF Random Forest.

SGD Stochastic Gradient Descent.

SN Supernova.

SNR Supernova Remnant.

SST Small-Sized Telescopes.

VERITAS Very Energetic Radiation Imaging Telescope Array System.

VGG Visual Geometry Group.

VHE Very-High Energy.

Zd Zenith Distance.

Chapter 1

Introduction

Gamma-ray astronomy is a young branch of Science whose purpose is to study very high-energy (VHE) photons and astronomical objects under extreme conditions. In 1989 the detection of TeV γ rays from the Crab nebula by the Whipple telescope opened the door of VHE γ-ray astronomy, leading to a subsequent evolution of Imaging Atmospheric Cherenkov Telescopes (IACTs). Starting in 1990, IACTs became more and more sophisticated, from HEGRA[1] to H.E.S.S.[2] and new discoveries have been made, such as the existence of cosmic particles accelerators, possible origin of γ rays, and a more deep view of the most energetic Universe with respect to the one known before. These advances led to the planning of the Cherenkov Telescope Array (CTA)[3] project, consisting in the creation of ~ 100 telescopes, located in both hemispheres, to cover the full sky.

IACTs are peculiar telescopes: they point directly to the γ-ray source but, instead of directly detecting the γ-ray photons, they observe the Cherenkov light generated in the atmosphere, where this electromagnetic radiation is emitted by the interaction between γ rays and air nuclei. IACTs are properly designed to capture these events, known as Extended Air Showers (EASs), and to record images of these showers. Due to the enormous quantity of data, in the order of Tera/Petabyte, and the insufficient hardware performance, the IACT standard analysis requires the cleaning and parametrization of each single image. Image cleaning is a lossy procedure, but necessary to remove the light of the Night Sky Background (NSB), with which the information of most of the pixels is lost. Typically, pixels with significant information after image cleaning are those containing photoelectrons produced by the atmospheric shower. These de-noised images are then parameterized calculating image parameters. After these pre-processing steps, the image parameters are used as input to train a Random Forest model, a machine learning algorithm, with the purpose of performing the event reconstruction. This process consists in a set of operations with the intent of estimating the probability of being a γ-ray event and then reconstruct the energy and the sky position of the source.

Fortunately, in the recent years, new hardware architectures and much faster computing units, such as Graphics Processing Units (GPUs), have been developed allowing to overtake the computational problems and to fully employ Convolutional Neural Network (CNN), such as VGG16 (Simonyan & Zisserman, 2015) and ResNet (He et al., 2016), as neural network architecture for IACT analysis.

---

[1] https://web.archive.org/web/20050209034041/http://www-hegra.desy.de/hegra/
[2] https://www.mpi-hd.mpg.de/hfm/HESS/pages/about/telescopes/
[3] https://www.cta-observatory.org/

The aim of the work presented in this thesis is to implement new methods for Convolutional Neural Network analysis and propose an optimal neural network configuration for each of the reconstruction steps, i.e. signal/background separation, energy reconstruction and arrival direction reconstruction, leading to better telescope performance.
Two solutions are suggested:

- A modified VGG16 architecture, trained on raw images.

- An hybrid ad-hoc architecture, built by the concatenation of a Deep Feed Forward network, trained on the image parameters, and a modified VGG16, trained on raw images.

Since the file format of the given dataset is not suitable for training a neural network, two software libraries are employed and their code has been extended in this thesis. The DL1 Data Handler (DL1DH)[4] package is used for the necessary preprocessing steps, such as preprocess the dataset and the input images; while the CTLearn[5] package is used for the neural networks training and prediction. My contribution in extending both of the software libraries allows to train and implement both of the proposed architectures in this work.

This thesis is the outcome of my work experience in the Group of Multimessenger Astrophysics in Padova, at the Istituto Nazionale di Fisica Nucleare (INFN), and it is divided into five further chapters:
Chapter 2 contains the overall concepts on astrophysics and physical phenomena, such as γ rays and generation of the Cherenkov light. It is organized to give the reader a basic knowledge of γ rays, showing their galactic and extra-galactic origins as well as the evolution of EAS.
Chapter 3 describes Imaging Atmospheric Cherenkov telescopes and the γ-ray data analysis. This chapter presents the IACT techniques and data analysis; in addition it illustrates the CTA project that, thanks to telescopes as the LSTs, will revolutionize γ-ray observation.
Chapter 4 presents an introduction to Convolutional Neural Networks, listing the difficulties faced and the solutions proposed, before describing the neural network architectures used and the libraries for their implementation.
In **Chapter 5** I report the dataset and the strategies used, the obtained results and the performed analysis, in addition to the comparison between the CNN approach and the standard machine learning one.
At last, **Chapter 6** contains the conclusions and discussion of the results of the thesis with a list of possible future works.

---

[4]https://github.com/cta-observatory/dl1-data-handler
[5]https://github.com/ctlearn-project/ctlearn

Chapter 2

Glimpse to VHE gamma-ray astronomy

Since ancient times, the night sky has been studied by humanity which, thanks to technological advances, has succeeded to observe it in various ways more and more precisely. In the beginning the Universe was observable only in the visible light band until, in 1912, the physicist Victor Hess, with an aerostatic balloon, was able to notice what is invisible to our eye. Hess observed that as the altitude increases the density of ionized particles increases as well, concluding that the reason must be an extraterrestrial radiation source. Later, given their origin, these radiations were called cosmic rays by Robert Millikan. (Millikan & Cameron, 1926).

CR studies reveal that their energy range spans from 10^8 eV to above 10^{20} eV while their structure is composed of charged particles, such as protons (87%) and Helium nuclei (12%). This characteristic leads to random deflections of their trajectory, due to magnetic fields encountered in their travel, obliterating any possibility of tracing back their origin once on the Earth's surface. As a consequence, to observe their source, we need to study CR acceleration products in non-thermal processes, such as neutrinos and photons - particles with a neutral charge, unaffected by magnetic fields.

CR acceleration products offer different points of view from which the Universe can be observed. Photons are the fundamental pillar of the Multi-Wavelength Astronomy (Middleton et al., 2017), where a cosmic object is observed in all the possible emission wavelengths, such as Infrared and γ rays, while for the Multi-Messenger Astronomy (Hinton & Ruiz-Velasco, 2020) , not only photons are observed but also gravitational waves (Abbott et al., 2016) and neutrinos emission, in order to reveal different information about their sources.
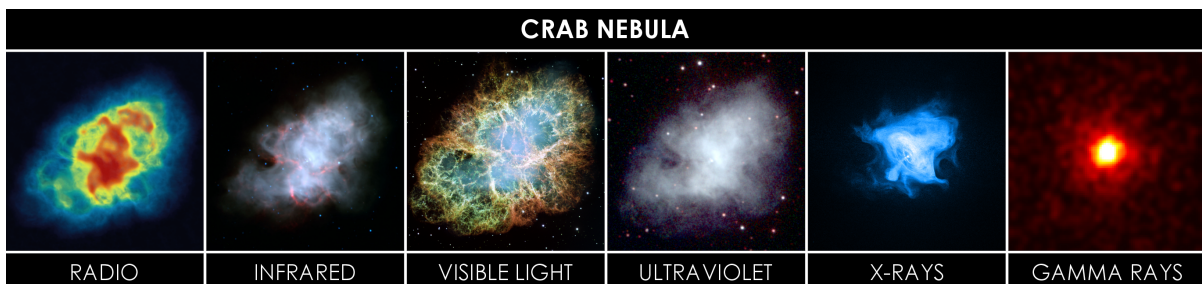
Figure 2.1: Crab Nebula in multiple wavelengths. Credits to NASA et al., 2005

2.1 Types of Atmospheric cascades

The Earth's atmosphere acts as a barrier between us and the Electromagnetic (EM) radiations, letting only few frequencies of the EM spectrum reaching the ground. The most powerful part of this spectrum, where we find the γ-ray radiation, has the peculiarity of interacting with atmospheric air nuclei, generating cascades of relativistic particles, as shown in Fig. 2.2.

According to the role played in EASs, particles can be classified as:

- **Primary particles**: they interact with the high atmosphere and originate the EASs. A primary particle can be a γ ray or a cosmic ray, mainly hadrons.

- **Secondary particles**: they are the product of the interaction between primary particles and atmospheric nuclei. They move faster than the speed of light in the atmosphere and consequently produce the so-called Cherenkov light. There are many types of secondary particles, like positrons or electrons, but also muons and kaons.
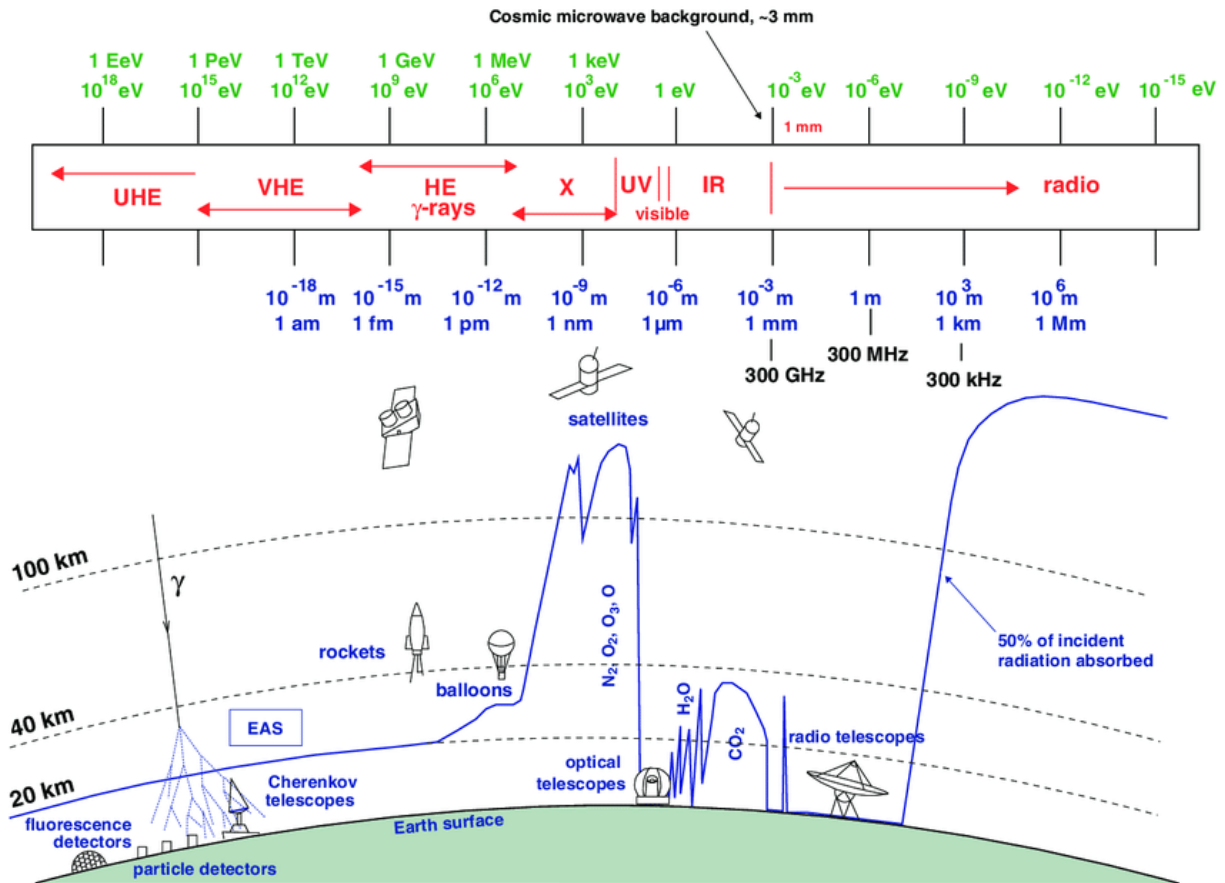
Figure 2.2: The EM spectrum (above) as seen from the altitude where photons are fully absorbed in the atmosphere (below). From (Olaizola, 2000), (Wagner, 2007), (Longair, 2011), (López-Coto, 2017)

Different types of primary particles generate different types of EAS:

- **Electromagnetic air shower**: generated by γ rays or leptons as primary particles. In this case, if a particle with an energy E, where $E \gtrsim 20$ MeV, interacts with air nuclei, the process called pair-production takes place, generating an electron-positron pair. These particles, interacting with the air nuclei and magnetic field, emit γ rays via bremsstrahlung. By the continuous succession of pair-production and bremsstrahlung processes, a cascade of particles is produced until the electrons and positrons reach the so-called critical energy, in air $E_c = 86$ MeV. Below this threshold, energy losses are dominated by ionization.

- **Hadronic air shower**: generated by hadrons as primary particles. In this case, the interaction with air nuclei produces pions and kaons, both undergoing into more collisions or decays, generating more sub-showers. The shower continues to grow up until the energy per nucleon is up to ~ 1 GeV, which is the minimum energy needed for pion production (Fernández Barral, 2017). As illustrated in Figure 2.3, pions decay into photons and muons, generating secondary EM and muon-initiated showers. Hadronic showers constitute the main background signal that we want to filter out, being the γ/hadron ratio $< 1/1000$.

Comparing EASs, we can observe that hadronic initiated showers are wider and more asymmetrical than the EM air showers. That is due to the major number of sub-showers and the higher transversal momentum of kaons and pions with respect to electrons and positrons. Even timing is different, EM air showers develop in less than 3 nanoseconds while Hadronic air showers take more than 10 nanoseconds to develop.
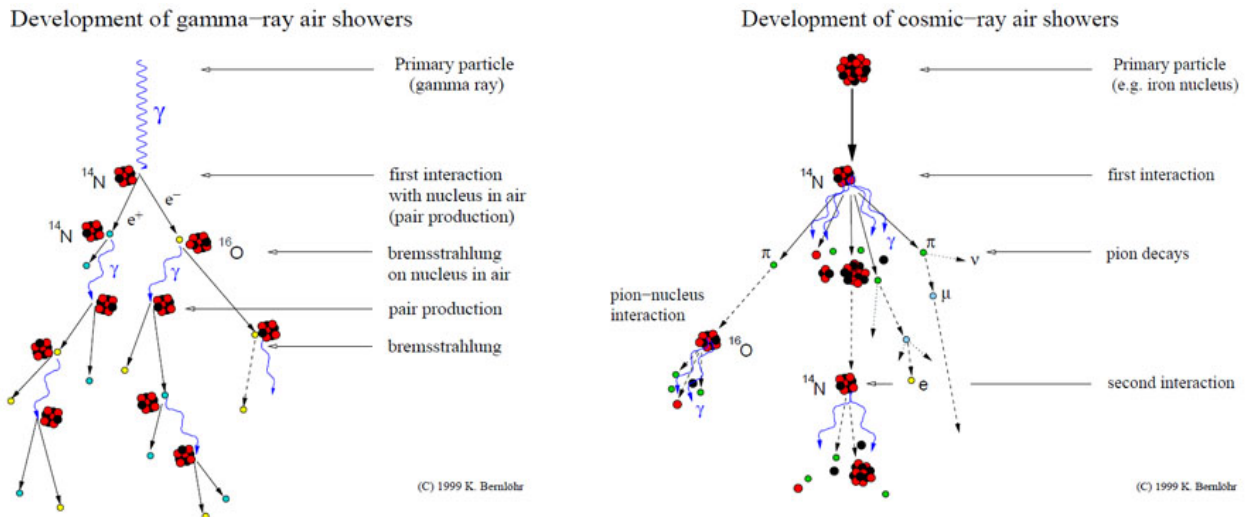
Figure 2.3: Two sketches of EASs development. Credit: K. Bernlöhr.

2.2 Cherenkov radiation

In 1934 a Soviet physicist, Pavel Cherenkov, demonstrated that every time a particle moves through a medium, such as air or water, with a speed v faster than the speed of light in that material, EM radiation is generated. This radiation, called Cherenkov light, propagates around the charged particle track in the form of a cone of angle θ calculated as:

$$\cos\theta = \frac{c}{vn(\lambda)}$$

where c is the speed of light in the vacuum and $n(\lambda)$ is the refraction index of the material, which depends on the wavelength of the Cherenkov light.

The cone shape represents how faster the particle is travelling in the material respect to the light's speed. If speeds are very similar the cone will be very wide, otherwise very narrow. In this particular scenario, where air in the atmosphere is the dielectric material, θ have slightly different values due to the non-uniform density of air's refraction index at different altitudes, being 1° its average.

Hadronic and γ-ray initiated EASs produce a cascade of particles which travel faster than the speed of light in the atmosphere, generating a circular pool of light, product of the superposition of every single interaction: the Cherenkov light pool. Since primary particles hardly fall perpendicular to the ground, the shape of the light pool is often an ellipsoid, especially for γ-ray initiated showers. For EM showers, the amount of Cherenkov light is proportional to the energy of the primary particle, allowing IACTs to reconstruct this energy, based on the recorded image.
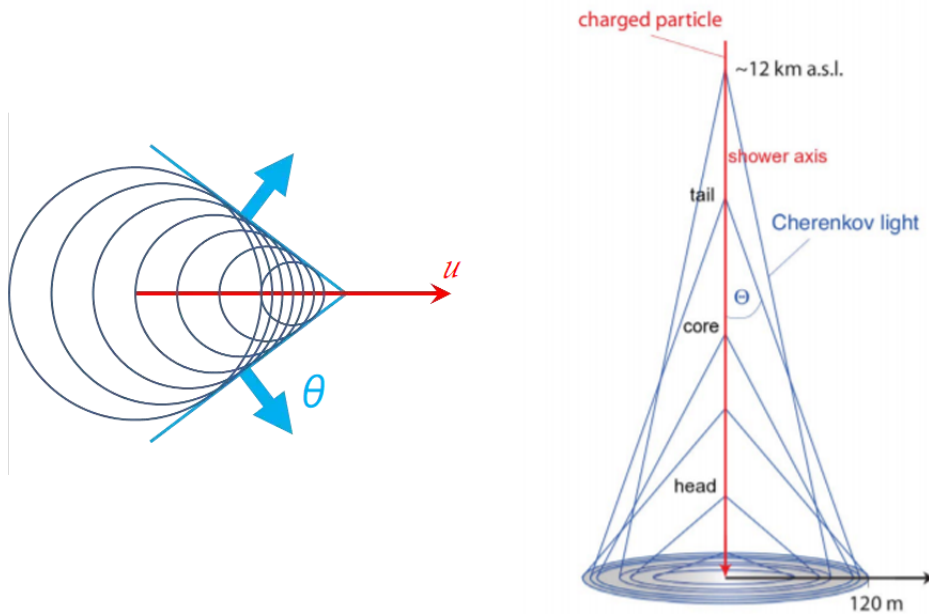
Figure 2.4: Atmospheric Cherenkov angle variation with altitude. (Schultz, 2013)

2.3   Gamma-ray astronomy

CR acceleration and their interaction with the surrounding environment can produce high energy photons, called γ rays, which allows scientists to trace back their origin. For this reason, γ-ray astronomy has gained more and more importance in recent times. The study of γ rays allow scientists to observe the non-thermal Universe, exploring the most violent phenomena within and gain a deeper knowledge to answer unsolved questions, such as the annihilation of dark matter and the nature of particle acceleration itself. γ-ray astronomy deals with the most powerful part of the electromagnetic spectrum, with energy above $\sim$1 MeV, and divided into several energy domains, as shown in Table 2.1.

Name	Abbreviation	Energy range
low energy	LE	1 MeV - 30 MeV
high energy	HE	30 MeV - 50 GeV
very-high energy	VHE	50 GeV - 100 TeV
ultra-high energy	UHE	100 TeV - 100 PeV
extremely-high energy	EHE	>100 Pev

Table 2.1: γ-ray astrophysics energy domains (López-Coto, 2017)

2.4   Gamma-ray sources

Gamma-ray sources are related to the most powerful events in the Universe and, at the time of writing this thesis, there are more than 220 established VHE γ-ray sources[1], whose origins may be galactic, such as:

- **Pulsars**:

    Pulsars are highly magnetized rotating Neutron Stars (NSs) where particles are accelerated in narrow beams in very specific regions along the magnetic field. Here, γ rays are produced due to the extreme magnetic fields that photons, electrons and positrons suffer within the pulsar magnetosphere. The best studied pulsar is the Crab pulsar, first detected by MAGIC (Aliu et al., 2008) and recently observed with LST-1.[2]

- **SNRs**:

    Supernova Remnants (SNRs) are the leftovers of Supernova explosions. It is supposed that γ rays are produced by the interaction between the Supernova (SN) ejecta and the Interstellar Medium (ISM), where particles get accelerated. An example is the SNR G24.7+0.6 where the γ-ray emission can be interpreted as the result of proton-proton interaction between the supernova and the CO-rich surroundings (V. Acciari et al., 2018).

---

[1]http://tevcat.uchicago.edu/
[2]https://www.cta-observatory.org/lst1-detects-vhe-emission-from-crab-pulsar/

- **PWN**:

  Pulsar Wind Nebula (PWN, plural PWNe) is a kind of nebula that lies inside SNR remains, directly fed by the NS at the center. The interaction between the ISM and the pulsar relativistic $e^{\pm}$ wind causes VHE emission (Aleksić et al., 2014).

- **Gamma-ray binaries**:

  γ-ray binaries consist of a massive star and a compact object, such as a black hole (BH) or a NS, which orbit around the common center of mass. Particles are accelerated according to two possible scenarios: in the microquasar one, as SS 433, they get accelerated by a jet; while, in the pulsar wind scenario, particles get accelerated by the interaction between the pulsar wind and the companion wind, as LS 5039.

Alternatively, their origins could be extragalactic, as:

- **AGN**:

  Active Galactic Nuclei (AGN) are actively growing super massive black holes in the center of some galaxies which present two relativistic jets perpendicular to the accretion disk. Here, charged particles get accelerated and produce γ-ray emission. Depending on the position of its relativistic jets towards the Earth, this phenomenon take the name of Blazar, or radiogalaxy, and present some peculiar characteristics, such as an high observed luminosity with a very rapid variation. Two of the most studied and brightest sources in the extragalactic TeV sky are the blazars Markarian 421 (Mrk 421) and Markarian 501 (Mrk 501) (de León et al., 2019).

- **Starburst galaxies**:

  A galaxy is called Starburst Galaxy when the star formation rate is exceedingly high. Consequently, the high rate of SNe leads to a higher CR density allowing shocks strong enough to accelerate particles and emit VHE γ rays. An example is the famous Messier 82, also known as The Cigar Galaxy (Karlsson, 2009).

- **GRBs**:

  Gamma-ray bursts (GRBs) are the most energetic and difficult to observe γ-ray outbursts known. It is believed that their origin take place during a massive star collapse, such as a hypernova, or a merging of two compact objects. Depending on the origin, the GRB created can be classified as short GRB, if its duration is less than 2 seconds, or long GRB, if it persists more time. The first GRB ever reported by an IACT was GRB 190114C (V. Acciari et al., 2019) On 14th January 2019 by the MAGIC telescopes.
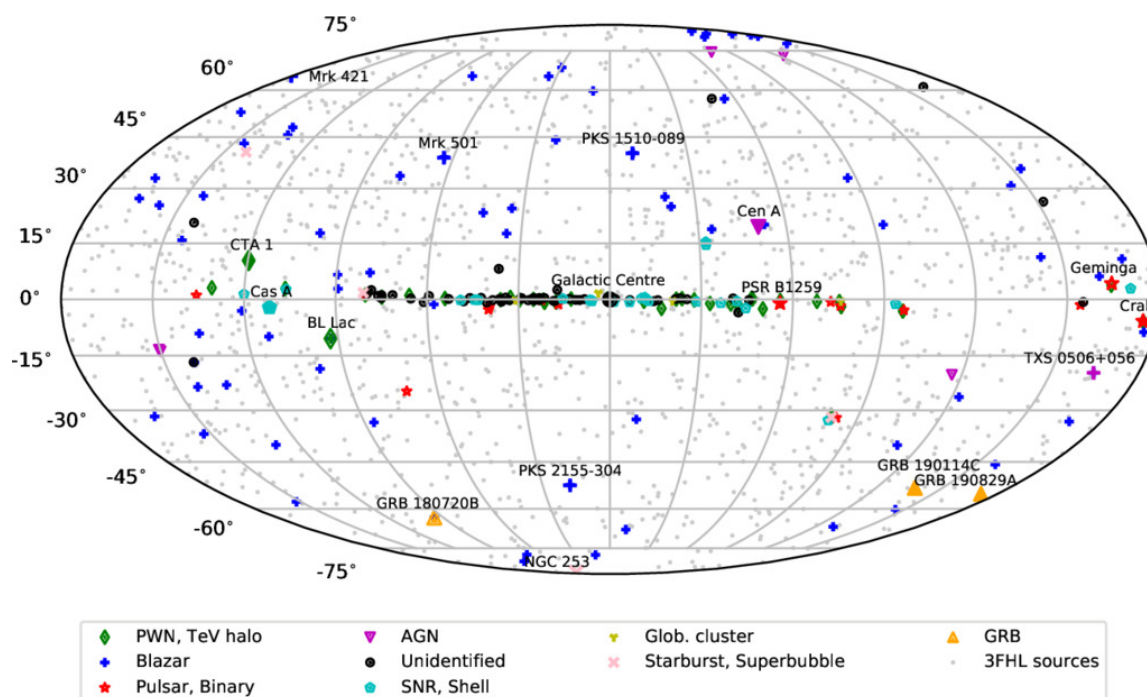
Figure 2.5: Map of the VHE γ-ray sky in mid-2019. Credits: J. Hinton.

Chapter 3

Imaging Atmospheric Cherenkov Telescopes

To detect the broadband γ-ray spectrum, scientists have developed several detection techniques that perform indirect γ-rays observations at ground level, as carried out in ground-based experiments, or direct observation above Earth's atmosphere with space-born satellites.

Thanks to on-board γ-rays detectors, such as the Large Area Telescope (LAT) on the NASA's *Fermi*, satellites detect photons before their interaction with the atmosphere's air nuclei. They perform long observations, examining a large portion of sky, thanks to their great field of view (FoV), and succeed in catching energies from MeV to a few hundred GeV. Over this threshold, satellites are not large enough to allow their detection. Consequently, scientists built different kind of instruments, detectors and ground-based telescopes, with the purpose of increasing the collection area observing EASs. As shown in Figure 3.1, they can also be interconnected with each other, reaching a multi-km^2 collection area.

There are two main techniques:

- **Imaging Atmospheric Cherenkov Telescopes (IACTs)**: They are designed specifically for reconstructing primary particle information using the atmosphere itself as a calorimeter. Because of this, their sensitivity to the light is high, forcing their functioning only during the nighttime. IACT standard operations and best performances are archived under the so-called dark conditions (i.e. the moon is absent from the sky), about 1600 h/yr. The LST-1 and MAGIC are amongst this type of instruments.

- **Extended Air Shower Arrays**: They usually use water tanks equipped with photomultipliers for the detection of the Cherenkov light produced by the secondary particles when they cross the water. They have a worse angular and a worse energy resolution compared to IACTs but they have an almost 100% duty cycle and have larger FoVs. The HAWC observatory is a perfect example of this kind of instrument.
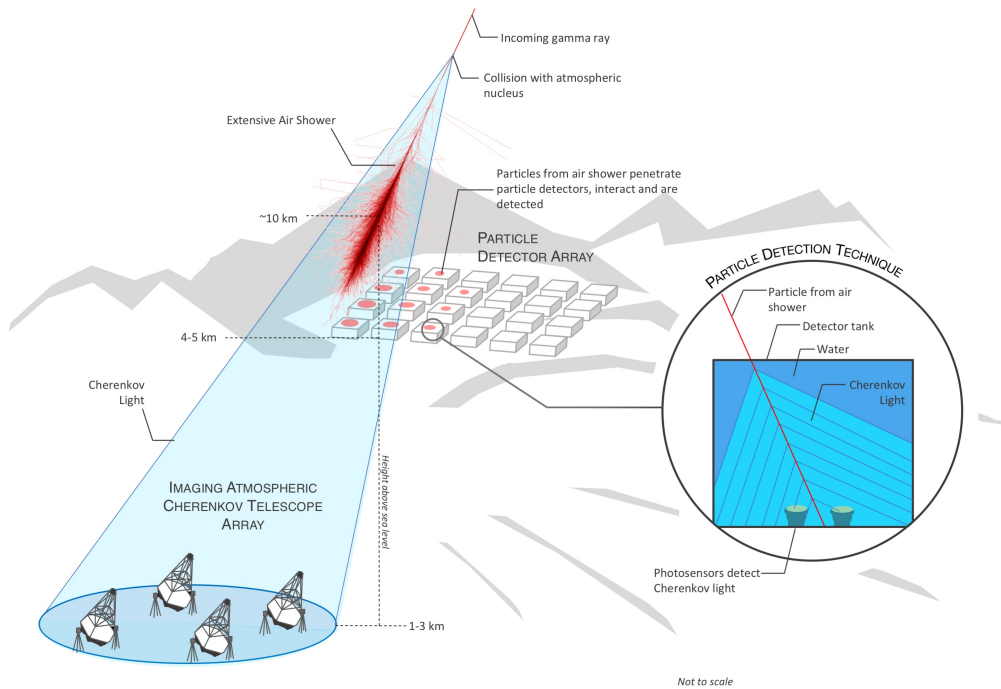
Figure 3.1: Illustration of the complementary detection techniques of high-energy γ rays from the ground. Credit: R. White, F.Schmidt, J.Knapp

3.1   The Imaging Atmospheric Cherenkov Technique

The main purpose of IACTs is to produce images of EAS, combining the spatial and the temporal information of the recorded Cherenkov light. To obtain this result, the telescope requires:

- Large reflectors, to collect the maximum number of photons from the cascade.

- A camera composed of very fast response pixels, usually Photomultiplier Tubes (PMTs).

Each pixel is a photo-detector that converts the collected light into an electrical signal, which is pre-amplified and sent to the readout board and to the triggering system. When several PMTs are above a certain threshold, this system is activated and the camera starts to record images; each image corresponds to a single shower. Once the digital image of the event is captured, the IACT data analysis can be performed starting with the classification of the primary particle of the shower, an example is illustrated in Figure 3.2.
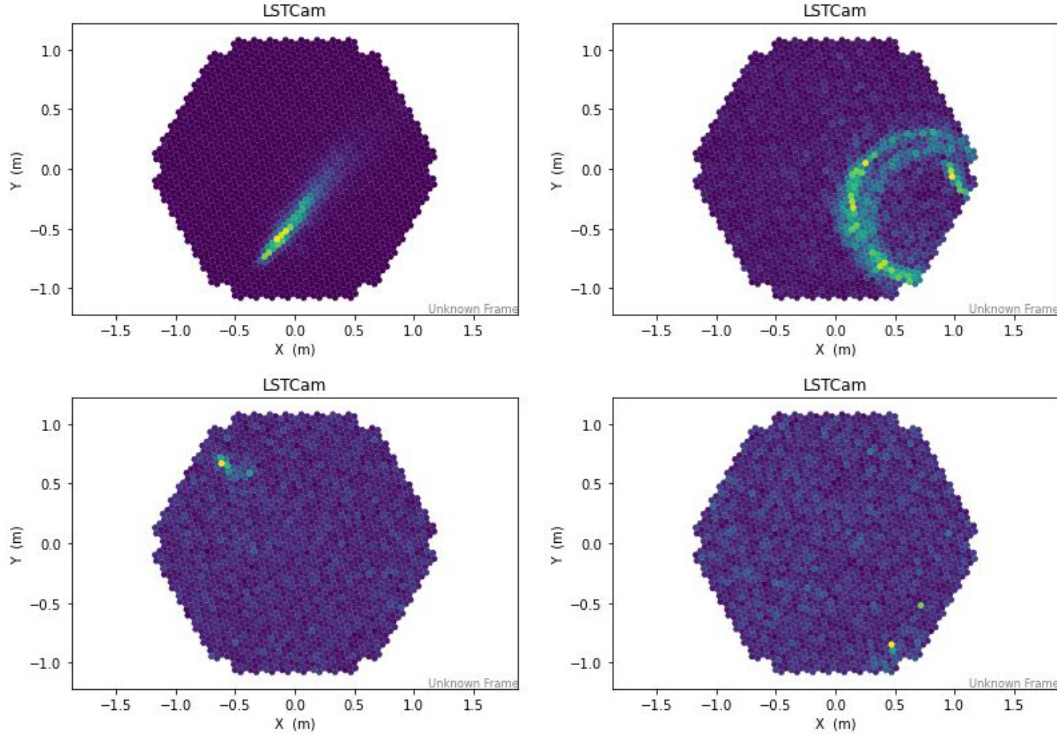
Figure 3.2: Example of γ-ray initiated shower images on the left column and hadron initiated shower images on the right column. As can be seen in these four pictures, it is usually difficult to classify images correctly (second row), while, very few times, it is less tricky (first row).

Since the effective observation time with the Cherenkov telescope arrays is limited to clear sky conditions, atmospheric agents, such as clouds, and the light pollution from cities, represent serious obstacles. To avoid them, at least for the majority, building sites are located at 2000-3000 m.a.s.l.[1], in areas with a low NSB level. Usually, each γ-ray observatory is composed by multiple IACTs working in the so-called Stereo Mode (Caraveo, 2020), in order to improve the event reconstruction process.

3.1.1 Pointing modes

IACTs perform their observations using two different strategies: the ON/OFF and the Wobble pointing modes.

- **ON/OFF**: In this mode two different types of observations are performed. To capture the signal, the source is positioned in the center of the camera and tracked, this is called ON observation; while the OFF observation refers to record events, pointing at sky regions where no γ-ray sources are expected. For a better estimation of the background, OFF observations should be taken in the same conditions of Zd/Az and NSB level encountered during ON observations.

- **Wobble**: In the Wobble observation mode (Fomin et al., 1994) the source is located with a certain offset from the center of the camera. This mode saves observation

---

[1]Meters above sea level

time, not requiring OFF observations, and guarantees a properly background estimation using the Wobble positions, as shown in Figure 3.3. To obtain a less biased result, due to the particular location of the source in the camera, the source position in the camera is rotated around the center by an angle of 180° or 90° every $\sim$ 20 minutes.
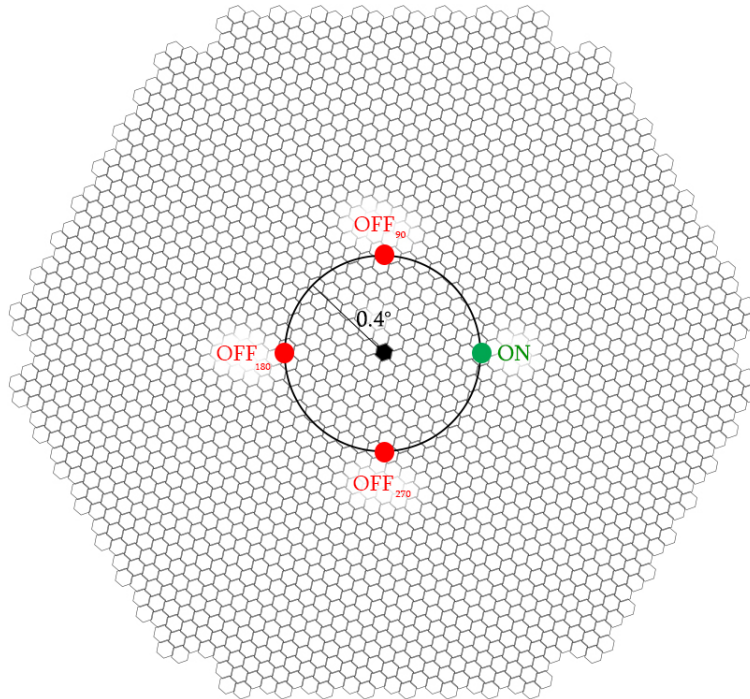
Figure 3.3: Wobble mode with 3 OFFs. A single OFF point, also called Wobble position, is a region in the camera located at a distance equal to the offset from its center. The offset depends on the source, the usual one is around 0.4deg

3.1.2   Data analysis

In γ-ray studies, recording useful images as well as analyzing them properly plays a key role. In this process, called data analysis, meaningful information are extracted from the recorded images and then used as input data for the event-reconstruction algorithm. Since the reconstruction steps are performed only on γ-ray events is crucial that this algorithm is able to distinguish efficiently the signal from the background, looking at the differences between the showers, as the more complex form of those initiated by hadrons (Fig. 3.2).

The currently IACT standard analysis pipeline is divided in 3 processing steps:

- **Low-Level data processing**: Once an image is recorded, an algorithm calibrates it, subtracting the pedestal[2] and integrating the signal. This allows to extrapolate the total number of photons and their averaged arrival times; these values are named charge and peak times respectively. Images are then cleaned from any other remaining source of noise, such as stars in the FoV, selecting only the extended patterns of photoelectrons, see Figure 3.4, which are then parameterized to extrapolate

---

[2]the measured quantity of NSB and readout noise recorded before the observation start

the image parameters, like the Hillas parameters (Hillas, 1985), see Figure 3.6. For a more detailed view see sections 3.1.2.1 and 3.1.2.2.
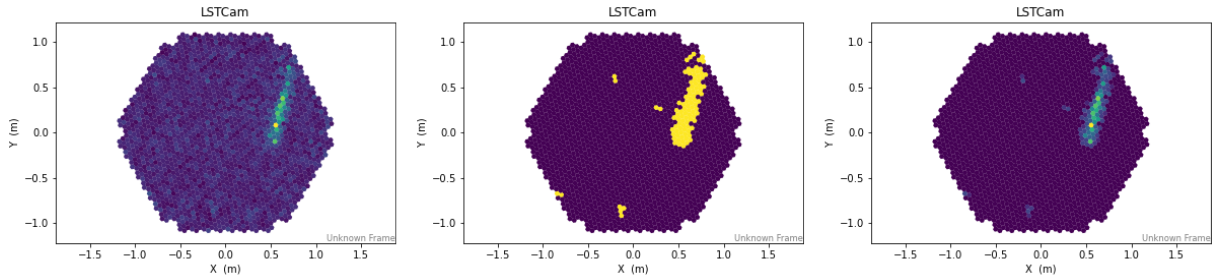
Figure 3.4: Illustration of images before (left) and after (right) the cleaning procedure. At the center is shown the cleaning mask: the preserved pixels are the yellow ones, each group is also called island. The cleaning procedure was performed with the tailcuts_clean(6,3) method.

- **Intermediate data processing**: In this step the event reconstruction process is performed employing algorithms such as Random Forest (RF) (Albert et al., 2008), or Look Up Tables (LUTs). Depending on the required task, RF is trained on the image parameters calculated of specific dataset:

    - For the **classification task**, i.e. γ/hadron separation, the training data is a combination of MC simulated diffuse γ-ray and real hadron events.

    - For the **regression tasks**, i.e. energy and arrival direction reconstruction, a data collection of MC simulated diffuse γ-ray events is used.

- **High-Level data processing**: In this step, graphs and plots, such as the skymaps and θ^2 plot, are displayed. They allow to understand and properly estimate the results leading to the physics analysis.

Since the LST-1 telescope is a prototype its analysis tools are still under development and only MC simulations are employed. Nowadays, the standard LST-1 data analysis pipeline is fully performed by the cta-lstchain package[3], which allows the analysis in mono mode up to high level. This package is based on ctapipe: the low-level data processing pipeline software of CTA. However, in this thesis, a combination of the ctapipe package, DL1 Data Handler and CTLearn is used, defining a new data analysis pipeline. A deep explanation of this last two packages can be found in Chapter 4 section 4.3.1 and 4.3.2. All of these packages are written in Python and are under active development.

3.1.2.1 Image Cleaning

Calibrated images contain the charge and arrival time values for each PMT, but most of them capture only noise, useless for the data analysis. Thus, the aim of the cleaning algorithm is to discard all the noisy pixels and maintaining only the ones that contain useful information about the shower. This process is performed defining two scalars: ζ and κ, with $\kappa < \zeta$, and maintaining a pixel only if:

---

[3]https://github.com/cta-observatory/cta-lstchain

- It detects a number of photons $> \zeta$, or

- It detects a number of photons $> \kappa$ and is adjacent to a pixel that detected more than ζ photons.

In this work the *tailcuts_clean* algorithm has been used with three different cleaning intensities, storing image parameters calculated from *tailcuts_clean(6,3)*, the softer one, which is the standard cleaning and the one employed in the hybrid architecture, *tailcuts_clean(8,4)* and *tailcuts_clean(10,5)* which are harder and experimental cleaning, as illustrated in Fig. 3.5.
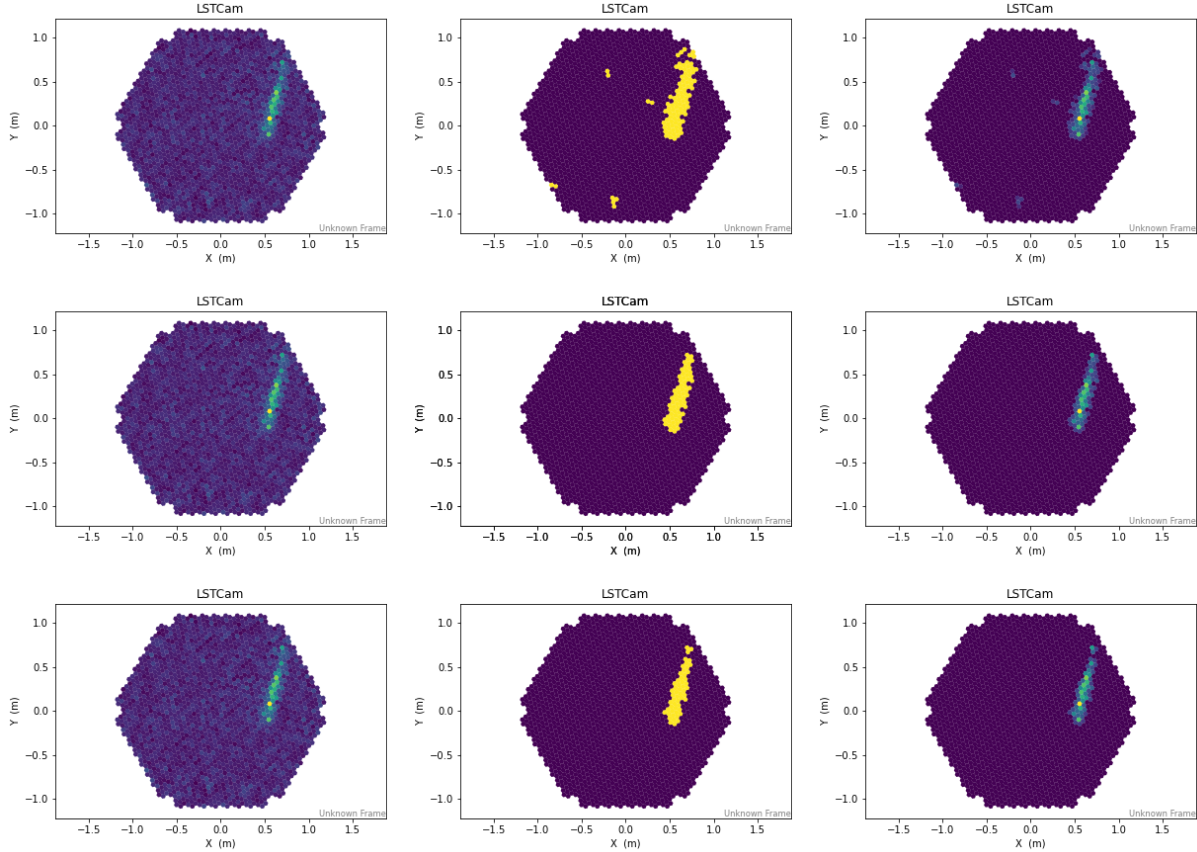
Figure 3.5: Comparison between the three cleaning algorithms employed performed on the same event. The first row refers to the *tailcuts_clean(6,3)*, the second to the *tailcuts_clean(8,4)* and the last to the *tailcuts_clean(10,5)*

3.1.2.2   Image Parametrization

Once the core and boundary pixels are selected, revealing large patterns called islands, the resulting images are parametrized, as illustrated in Figure 3.6. The obtained parameters are a transposed representation of the images in the values domain and their calculation is very useful since they relieve the computational load of the event reconstruction algorithm. The image parameters are divided into several groups to represent every useful detail of the image:

Image Quality parameters: They are used to evaluate the containment of the shower or the quantity of noise in the image.

- **LeakageN**: fraction of the size of the source contained in the N outermost rings of the camera.

- **Number of islands**: number of non-connected groups of pixels that survived the image cleaning.

Source-Dependent parameters: They are still related to the physical properties of the shower, but they depend on the expected position of the source.

- **Dist**: Angular distance from the source expected position in the camera to the CoG of the shower image.

- **Alpha**: Angle between the ellipse major axis and the line connecting the expected source position to the CoG of the shower image.

Time parameters: They are still related to the shower physical properties focusing on the time information.

- **Time gradient**: slope of the linear fit to the arrival time projection along the major axis of the ellipse.

- **Time intercept**: intercept on the ordinate axis of the linear fit to the arrival time projection along the major axis of the ellipse.

Hillas parameters: Represent the shower physical properties and are derived by fitting an ellipse to the maintained pixels:

- **Intensity**: also called size, is the total charge (in phe) contained in the image.

- **Width**: semi-minor axis of the ellipse. It measures the lateral development of the shower.

- **Length**: semi-major axis of the ellipse. It measures the longitudinal development of the shower.

- **CoG**: centre of gravity of the image. It is computed as the mean of the X and Y weighted mean signal along the camera coordinates.

- **Conc(N)**: Fraction of the image concentrated in the N brightest pixels. It measures how compact the shower is and tends to be larger in gamma rays.

- ϕ: angle between the line connecting the CoG with the centre of the camera and the x-axis.

- ψ: angle between major axis of the ellipse and the x-axis.

- **r**: angular distance between the CoG and the centre of the camera.

- **Skewness**: 3rd order Hillas parameter.

- **Kurtosis**: 4th order Hillas parameter.
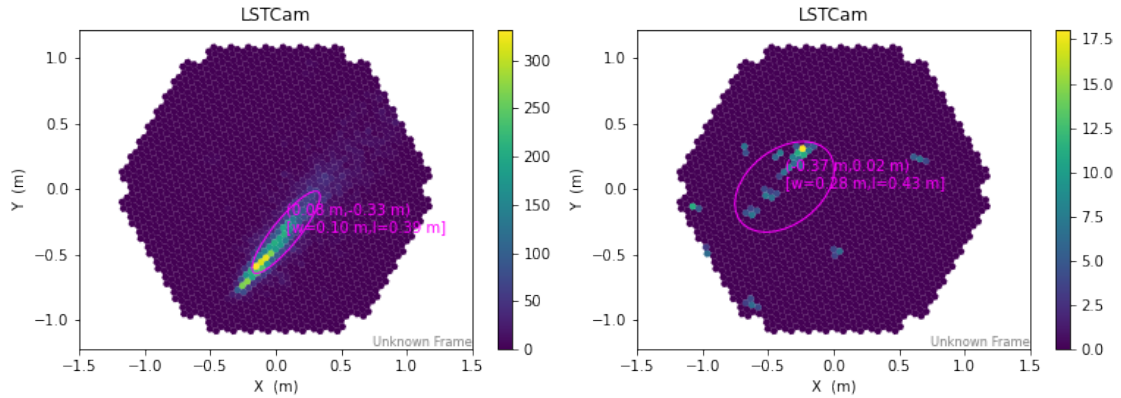
Figure 3.6: Illustration of the Hillas parameters calculated on a single island (left) versus on multiple islands (right).

3.2 IACT arrays

A common characteristic of IACT sites is to consist of multiple telescopes, in order to gain better performance by improving the detection of low γ-ray fluxes and capture different images of the same event. This allows to properly reconstruct the direction of the incident γ ray and have a better background suppression.

Currently, several IACT arrays are present and functioning, observing the sky from different areas of the globe. The most famous are:

- **H.E.S.S.**: The High Energy Stereoscopic System (H.E.S.S.) array is located near the Gamsberg Nature Reserve, Namibia, and went into full operation on July 26, 2012. It is composed of five telescopes: four of them are equipped with a mirror of just under 12 meters in diameter, while the largest one is built in the centre of the array and equipped with a mirror of 28 meters. In 2016, the H.E.S.S. collaboration reported deep γ-ray observations showing the presence of PeV protons originating in the galactic core of the Milky Way (HESS Collaboration, 2016).

- **VERITAS**: The Very Energetic Radiation Imaging Telescope Array System (VERITAS)[4] array is a γ-ray observatory consisting of four 12-meters spherical reflectors, located in southern Arizona, in the United States. It became fully operational in September 2007, discovering in 2009 the first starburst galaxy to emit gamma-ray energies, the Cigar Galaxy (V. A. Acciari et al., 2009).

- **MAGIC**: The Major Atmospheric Gamma-ray Imaging Cherenkov (MAGIC)[5] array is composed of two IACTs: MAGIC-I (2004) and MAGIC-II (2009), of 17-meters diameter parabolic reflector each. It is located in the Roque de Los Muchachos Observatory, on the island of La Palma, and received a major update in the years 2011-2012; the aim was to reduce the differences between the two telescopes to improve stereo performance. The lower energy threshold reached, 25 GeV, and their big size make the MAGIC array ideal for observing high redshift[6] γ-ray sources. MAGIC telescopes have given prolific scientific results such as the detection of the

---

[4]https://veritas.sao.arizona.edu/
[5]https://magic.mpp.mpg.de/
[6]the cosmological measure unit that indicates how far away objects are by measuring their wavelength.

galaxy PKS 1441+25, one of the two most distant galaxies detected at high energies; and the detection of γ-rays from a gigantic explosion occurred in the galaxy QSO B0218+357 thanks to gravitational lensing[7].
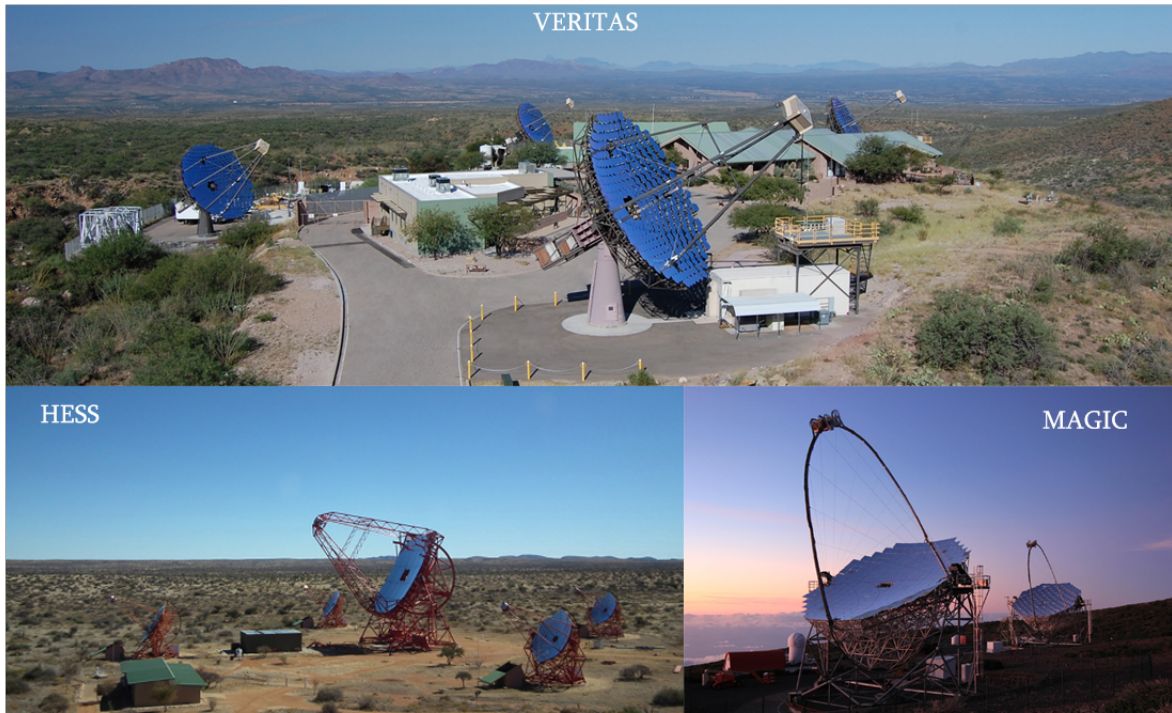
Figure 3.7: Picture of the aforementioned IACT arrays. Credits to: MAGIC, HESS, VERITAS Consortiums.

3.2.1 The Cherenkov Telescope Array

The Cherenkov Telescope Array (CTA) project is considered the future of ground-based observatories, taking the VHE γ-ray astronomy to the next level. CTA will be one of the largest astronomical infrastructures on Earth, composed of ~ 100 telescopes, and operating them as an open observatory (Mazin, 2019). These telescopes will be built on two sites, as illustrated in Figure 3.8, providing a full sky coverage:

- **The north site**, placed on the island of La Palma (Canaries), will focus on the low- and mid-energy ranges from 20 GeV to 20 TeV, looking for extra-galactic γ-ray sources. For this reason, the Northern hemisphere site will host 4 LSTs, to improve the sensitivity of CTA at the lowest energy, and 15 MSTs, to cover CTA's core energy range. Here, on 10 October 2018, the prototype of the LST, called LST-1, was inaugurated, becoming the first CTA telescope ever built on site.

- **The south site**, placed in the Atacama Desert (Chile), will span the entire energy range of CTA, covering γ-ray energies from 20 GeV to 300 TeV, looking for galactic γ-ray sources. The plan is to host 4 LSTs, 25 MSTs and 70 SSTs to cover CTA's highest energy range. To date, February 2021, none of them have yet been built on this site.

---

[7]https://magic.mpp.mpg.de/outsiders/results/magic-highlights/
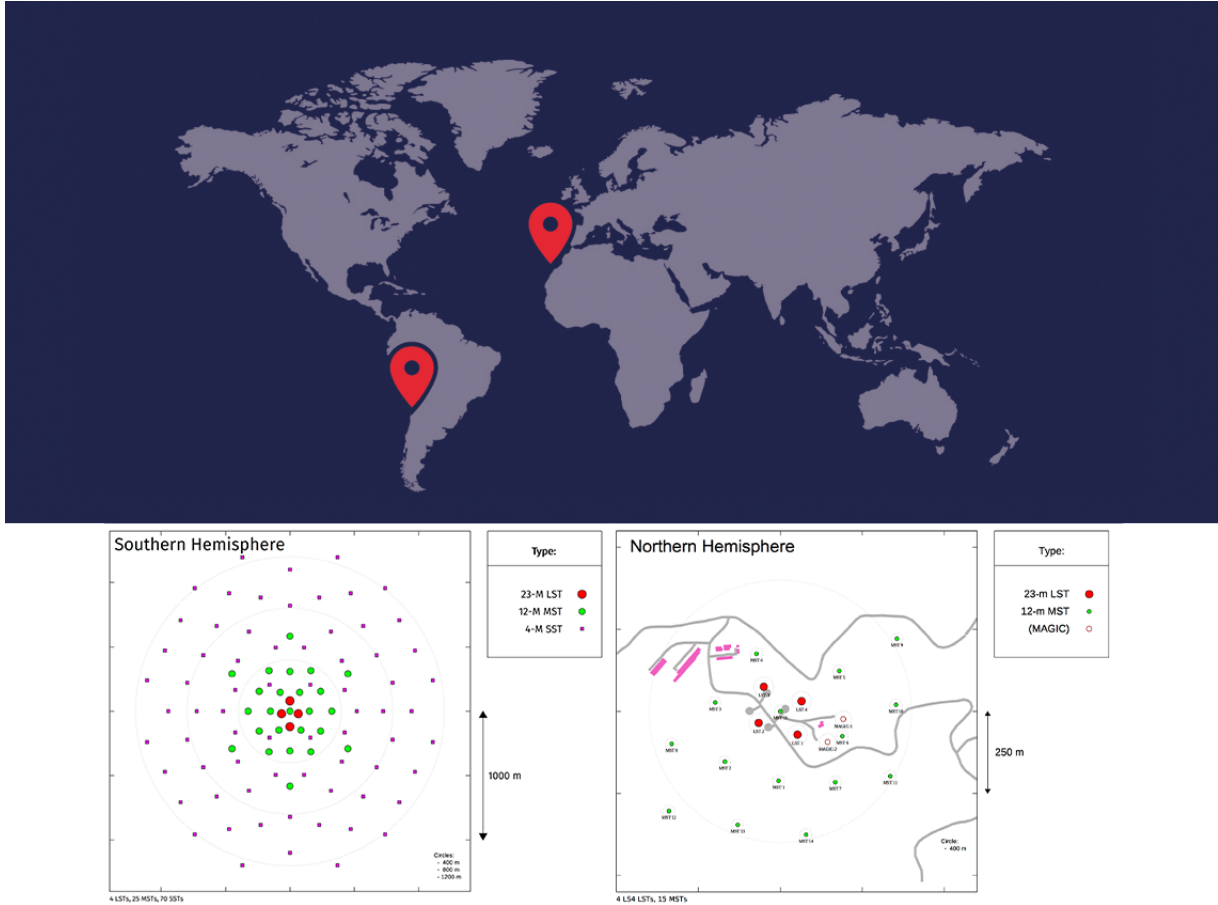
Figure 3.8: CTA's sites locations (top) and array layout in the Northern (right) and Southern (left) hemispheres (bottom), respectively. Credits to: CTA Consortium.

Besides the large collection area that this project will reach, the real innovation is the graded approach used in telescope arrays of CTA, where different typologies of Cherenkov telescopes are employed and focused on different parts of the energy range. In details:

- **Small-Sized Telescopes (SST)** : SSTs[8], with the smallest mirror area (8 m^2) of the project, will cover energies between a few TeV and 300 TeV, improving the chances to detect the highest energy γ rays. They will outnumber all the other telescopes in order to cover the largest collection area possible, increasing the chances to capture VHE γ rays, which are very uncommon. In addition, VHE showers produce a large amount of Cherenkov light on a large zone, therefore, a larger collection area is required.

- **Medium-Sized Telescopes (MST)** : MSTs[9] will cover the core energy range of CTA, from about 150 GeV to 5 TeV, with a mirror area of 88 m^2 and a 16-meters focal length. An MST can be equipped with a FlashCam or a NectarCam, which have slightly different features, such as the number of pixels: 1764 and 1855

---

[8]https://www.cta-observatory.org/project/technology/sst/
[9]https://www.cta-observatory.org/project/technology/mst/

respectively, and the width of the field of view: 7.5deg for FlashCams and 7.7deg for NectarCams.

- **Large-Sized Telescope (LST)**: LSTs[10] will be arranged at the centre of both the sites arrays in order to cover the low energy sensitivity of CTA, between 20 and 150 GeV, with a mirror area of 370 m^2. Unlike other CTA's telescopes, LST is an alt-azimuth telescope, which stands 45 meters tall and weighs around 100 tons. Its structure is made of reinforced carbon fibre tubes and the camera is held by two carbon-fibre arches, allowing it to re-point itself of 180 degree within 20 seconds to capture brief low-energy γ-ray signals.
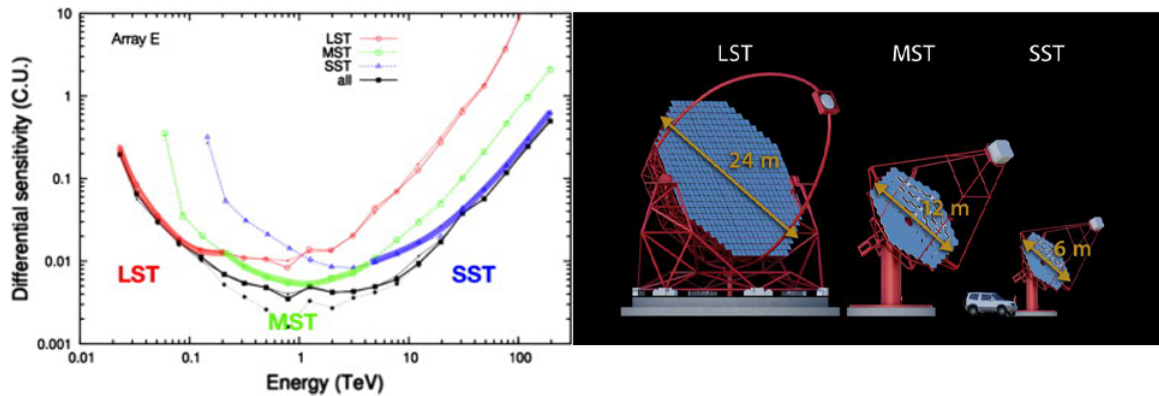
Figure 3.9: Illustration of the energy range of CTA (left) covered by its telescopes (right). The graded approach also allows many observation modes, depending on the part of the spectrum that has been decided to emphasize. Credits to: CTA Consortium.

---

[10]https://www.cta-observatory.org/project/technology/lst/

Chapter 4

Deep Learning and Convolutional Neural Networks

Machine Learning (ML) is the subfield of Artificial Intelligence (AI) that provides systems with the ability to automatically learn and improve from experience, without being explicitly programmed. ML algorithms focus on gaining experience by exploring data, identifying patterns on their own and leveraging them to learn. This new approach has revolutionized the data analysis process, enabling to perform tasks that would be too hard for a human, as the event reconstruction in the IACT data analysis. The new developments and improvements of Deep Learning (DL) have established it as one of the most promising branches of ML, with outstanding results in applications of various domains, such as medical imaging, self-driving cars and fraud detection. This great success lies in the DL's solving strategies for representation learning problems, where useful information is extracted directly from the raw data and their representation used to perform classification and regression tasks.

Given these great successes and the current development of the analysis tools for the LST prototype, it was decided to employ and test a DL strategy: the Convolutional Neural Networks (CNNs). As it was described in section 3.1.2, the standard analysis in the LST-1, and in IACTs in general is based on the parametrization of the recorded images of the showers. These parameters are then fed to a ML algorithm (in the LST-1 particular case, a RF) that performs the regression and classification tasks needed to reconstruct the original parameters of the event.
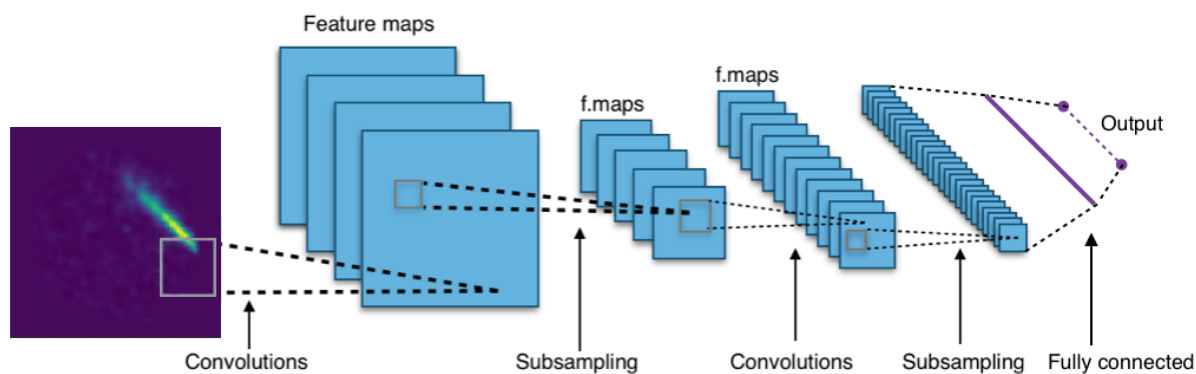
Figure 4.1: Illustration of how Convolutional Neural Networks work, from the input image to the final prediction.

4.1   Convolutional Neural Networks

Convolutional Neural Networks (Lecun et al., 1998) are Artificial Neural Networks (ANN) specially designed for image analysis. They are able to successfully capture the spatial and temporal dependencies in the images through the application of relevant filters. Generally, the input data is a 3-ranks tensor of shape $h \times w \times d$, where $h \times w$ is the number of pixels and d is the depth, corresponding to the image channels. Since image analysis is computationally intensive, the primary role of CNNs is to reduce images to a form that is easier to process, without losing features. A simple scheme is illustrated in Fig. 4.1. Their architecture is basically composed of:

- **Convolutional layers**: The role of a convolutional layer is to perform the convolution operation, which consists in extracting features from images. To achieve this process a square filter is applied to the image, which moves to the right until it parses the complete image's width. Then, it hops down to the beginning (left) of the image and repeats the process until the entire image is covered. At the end of the process, a feature map is filled.
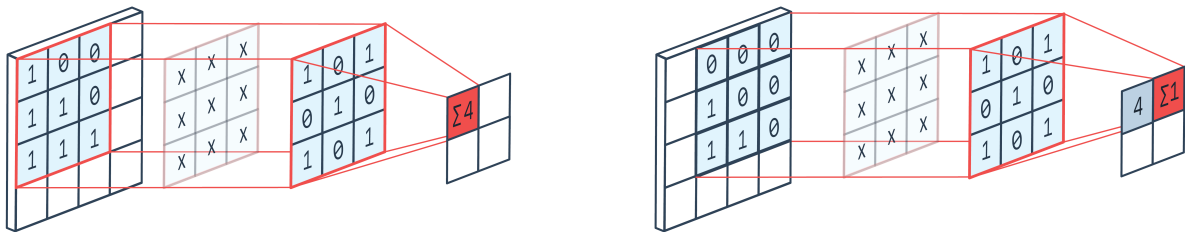
Figure 4.2: Illustration of the convolution operation. The applied filter moves with a certain stride value, in this case equal to 1, filling the feature map with the total sum of the pixels selected by the filter.

- **Pooling layers**: The main role of a pooling layer is to decrease the network dimensionality; downsampling each feature map independently by reducing the height and width but keeping the depth intact. This results in less computing power required to process the data, preventing also the problem of overfitting, explained in section 4.1.3. In this work different types of pooling layers were employed, such as the *Max pooling* and the *Global Average pooling* layers.
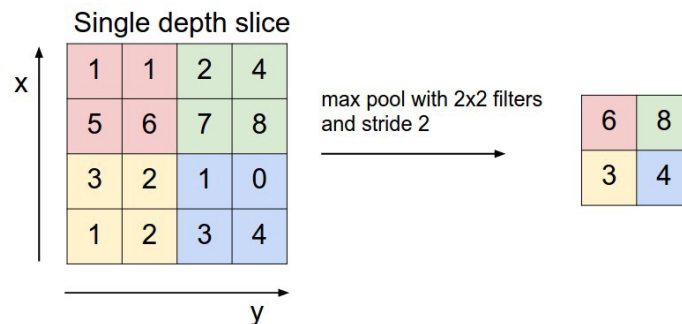
Figure 4.3: Illustration of how a max pooling layer reduce the dimenstionality of the feature map by considering only the highest value inside the applied filter.

- **Activation functions**: The role of the activation functions is to determinate if a neuron should activate or not, based on its input. There are several functions, such as the Rectified Linear Unit (ReLU) and the Sigmoid, which are applied to the convolutional layers to modify the learning behaviour of the network. If no activation function was present, the entire network would be equivalent to a single large linear system, losing the ability to represent all the information extracted into the output categories, such as "bike-ness" or "truck-ness". In our case it will be gammaness, energy and direction. In this work, the use of the ReLU activation function for hidden layers was preferred to the others since it is less likely to saturate, leading to a tougher training process.

Sigmoid

$\sigma(x) = \frac{1}{1+e^{-x}}$

tanh

$\tanh(x)$

ReLU

$\max(0, x)$

Leaky ReLU

$\max(0.1x, x)$

Maxout

$\max(w_1^T x + b_1, w_2^T x + b_2)$

ELU

$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$
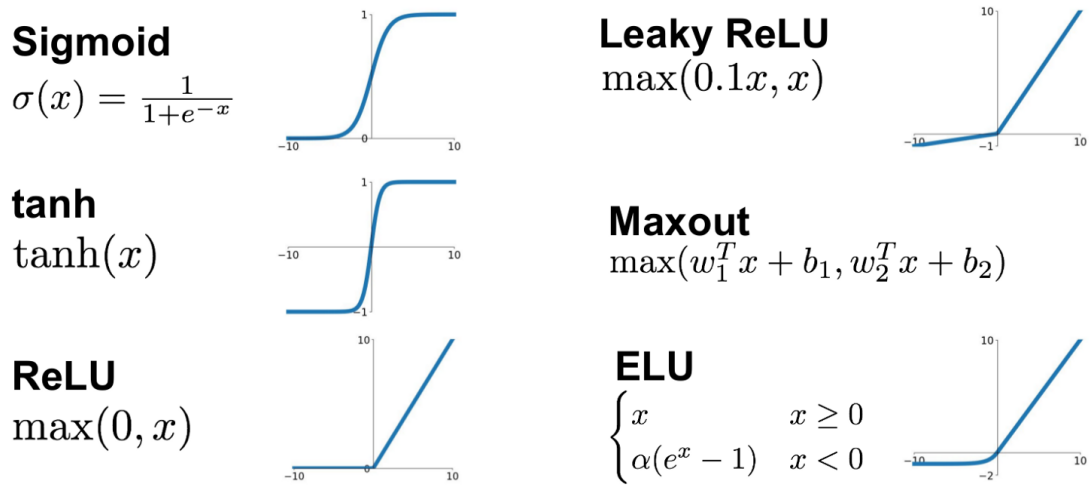
Figure 4.4: An illustration of the most common activation functions with their corresponding graph.

One fundamental aspect of CNNs is the depth of the network. Multiple convolutional layers can be stacked, taking the name of Convolutional Block, increasing the depth and allowing to capture more specific features. The first layers capture the low-level features such as edges, colour and gradient orientation while the high-level features, such as objects, are captured by the last ones. The implementations of VGGs and Deep Residual Networks highlighted by showing great results in image recognition.

4.1.1 Training process

Before accomplishing the required task, an ANN demands a training phase, where the algorithm sees the whole training dataset a precise number of times, called epochs. In detail, this training set comes from splitting the entire dataset into three parts, labelled as: training set ($\sim 70\%$), validation set ($\sim 10\%$) and test set ($\sim 20\%$). During the training step, the weights[1] are adjusted via the back-propagation algorithm, searching for the best combination to minimize the desired loss function L(s,ŝ)[2]. Meanwhile, the validation set is used to monitor the improvement of the network by estimating the performance on the unseen data. Interpreting its behaviour allows understanding if the

---

[1] the parameters contained in the hidden layers

[2] the measure on how far the predicted score values s are from the true values $\hat{s}$

network is specializing too much on the training set or if it is not learning at all. Lastly, the model is benchmarked on the test set, performing its predictions for the task.

4.1.2 Optimization

The training process of ANN can be optimized with different strategies, like the use of an optimizer or the dataset normalization, obtaining better results and more robust networks on unknown data.

Optimizers are algorithms that, based on their hyperparameters values, reduce the loss function by adjusting the weights of the network. The most popular optimization algorithm is the Stochastic Gradient Descent (SGD), which evaluates the gradient on the current batch of data and updates the weights in the opposite direction. The Learning Rate (LR) is the most influential hyperparameter, which must be carefully tuned for a good optimization process, as shown in Figure 4.5. In this work, an extension of SGD was used: the Adam[3] optimizer; since it computes individual adaptive LRs and combine the advantages of AdaGrad (Duchi et al., 2011) and RMSProp (Tieleman & Hinton, 2012) (Kingma & Ba, 2017).

Figure 4.5: The impact of LR on the training process.

A best practice in DL image applications is normalizing the input data so the network can learn the optimal weights for each node more quickly. Different strategies are allowed, such as normalizing the whole dataset and centring the images, which is a very expensive procedure, or use the Batch Normalization (BN) (Ioffe & Szegedy, 2015). This last strategy considers only the current batch, the number of samples seen by the network, and addresses the internal covariate shift by normalizing the layers input, i.e. subtracting the mean and dividing by the variance which are estimated during the training phase. This strategy increases the stability of the network training, allowing the possibility of employing much higher learning rates. In this work, the use of BN has been preferred since the pixel values of the input images are not bound between 0 and 255, like common RGB images, but contains the information of the charge which is necessary for the event-reconstruction process. In IACT data analysis input images are usually composed of 2 layers, namely charge and arrival time information, with pixels values which span from negative to positive quantity.

---

[3]ADAptive Moment estimation

4.1.3 The Overfitting problem

Training a deep neural network that can generalize well to new data is a challenging problem. It is a common reality to fall into *overfitting* by defining a model that learns too well the data patterns of the training dataset, specializing in them. The consequence is a bad performance of the network when applied on unknown data.

Fortunately, overfitting can be prevented and resolved with various existing techniques, such as adding more training data or reducing the network complexity. The most popular option is the employment of dropout layers (Srivastava et al., 2014), which randomly turn off a defined percentage of neurons at each step[4], preventing them from updating their weights. In this work, instead of dropout layers, weight decay and global average pooling layers were employed to deal with the overfitting problem, in addition to paying attention to find the right compromise between batch size, LR and network complexity. This strategy has been necessary as dropout layers are not already available in CTLearn.

4.2 Architectures

A CNN architecture is the composition and succession of distinct layers which, stacked together, define the complexity and structure of the network. In recent years several CNN architectures have been designed, focusing on deeper architectures, like ResNet, or wider ones, like InceptionNet (Szegedy et al., 2015). In this work, two different architectures are used and presented: a modified VGG16 and a hybrid architecture. Both have been defined in the CTLearn package.

4.2.1 Modified VGG16 architecture

This architecture is inspired by the standard architecture of VGG16, as the great model performance in the image analysis: it achieves 92.7% top-5 test accuracy on ImageNet, one of the largest data-sets available during the Large Scale Visual Recognition Challenge 2014 (ILSVRC2014)[5] (Simonyan & Zisserman, 2015). This model was also used in other DL thesis on IACTs: Grespan, 2020, Marinello, 2019. Unlike the VGG16 model, the proposed architecture is designed to be composed of the canonical five convolutional blocks of VGG16, paired with the ReLU activation function (Simonyan & Zisserman, 2015), without the Fully Connected (FC) layers[6] at the end, for a total of $\sim 1.4 \cdot 10^7$ parameters, described in A. A BN layer is added at the beginning for normalizing input and also speed up the training process while at the end of the convolutional block, a global average pooling is appended to reduce the network dimensionality. The implementation of the last layer depends on the task:

- For the **classification task** it is employed a FC layer with two neurons with the consequently use of the softmax_crossentropy activation function.

- For **regression tasks** it is employed a FC layer with one neuron paired with the absolute difference activation function.

---

[4]a step corresponds to a gradient update
[5]http://www.image-net.org/challenges/LSVRC/2014/
[6]layers that connect every neuron in one layer to every neuron in the next layer

4.2.2 Hybrid architecture

This architecture is inspired by the idea of using the pixel information combined with the image parameters used in the standard IACT analysis method, section 3.1.2.2, taking the advantages of both. It is composed by two macro blocks:

- **The convolutional macro block**: This block allows to analyze the raw images without losing any information in the pixel-wise data, which is one of the main advantages of using DL methods. It is inspired by the VGG16 architecture and composed by the first four canonical convolutional block, followed by a global average pooling, which replaces the flatten layer, and a variable number of FC layers. At the beginning a BN layer is added.

- **The image parameter macro block**: This block allows to analyze the image parameters, adding new useful information to the network. It is composed by a variable number of ReLU-activated FC layers, usually five, taking as input the image parameters selected.

These two macro blocks are then concatenated and followed by one last layer as for the previous architecture, for a total of $\sim 7.9 \cdot 10^6$ parameters, described in B.

4.3 Software Libraries

In this work, two software libraries have been used: The input data is read and generated by the DL1DH; The definition of neural network models, the training and testing processes are performed by the CTLearn. In order to obtain the results presented in this work, the author of this thesis has contributed to the extension of the code of both libraries.

4.3.1 DL1 Data Handler

The DL1 Data Handler (Kim et al., 2020) is a python package of utilities for writing, reading, and applying image processing to calibrated data from IACT in a standardized format. It is under active development and can perform two crucial steps:

- Processing the raw dataset, a collection of files containing the telescope simulations, into a proper one, a collection of *hdf5*[7] files, for DL and ML analysis.

- Loading the data and preprocess the images, interpolating them for neural network training. A deeper explanation is illustrated at Section 5.1.1.

The extension of this package, performed specifically for this thesis, consisted on adding new information to the files and adapt the entire pipeline to the new data format. In detail, each single *hdf5* file has been updated by adding the image parameters and the cleaning masks. The new file structure consists of tables named as:

- **Array_Information**: which collect the telescope array information, such as the geographic coordinates of the telescope under consideration with respect to the center of the array.

---

[7]Hierarchical Data Format

- **Events**: where the information of each single simulated event is stored, such as the azimuth angle of the shower and its energy.

- **Telescope_Type_Information**: storing the telescope details, like the pixels of the camera.

- **Images**: Here the information extrapolated from images are stored, such as the arrival time and the charge. Also the cleaning mask and the inverse cleaning mask.

- **Parameters** : used to store the image parameters of each event. There can be multiple parameter tables corresponding to different cleaning algorithms. In that case:

  - **Parameters0** table, which is always present, contains the image parameters calculated with the default cleaning algorithm. In this work it refers to tailcuts_clean(6,3).
  - **Parameters 1 - N** are optional tables and contain image parameters coming from optional cleaning algorithms. In this work they refer to tailcuts_clean(8,4) and tailcuts_clean(10,5), respectively.

Thanks to the new file structure, each single file can store different image parameters tables by running the preprocessing step once on the entire dataset production, saving time and disk memory, as no duplicate data is generated, besides to allow more in-depth analysis and quality checks on results, quickly and easily. Last but not least it speeds up the image filtering process of CTLearn.

4.3.2 CTLearn

CTLearn is a Python package, under active development, that uses the DL technique to analyze data from IACT arrays (Nieto, Brill, Feng, Humensky, et al., 2020). It implements several DL models based on TensorFlow (Abadi et al., 2016) and provides a high-level interface for reproducible training and prediction by automatically recording the training hyperparameters and the model architecture, which are explicitly defined by the user through a YAML[8] configuration file. CTLearn is organized around the *run_model* module, as shown in Fig. 4.6, which parses the configuration, loads the data calling the DL1DH's *reader* module and initializes the model. Depending on the mode set by the user, it can train the network or run predictions only, otherwise, it can perform both operations in sequence.

In order to obtain the results for this work, the author of this thesis has contributed to CTLearn by adapting the handling of the new data format and implementing two new model architectures:

- The **VGG16 architecture** is implemented by extending an already existing code, adapting it to define more convolutional layers in sequence, interspersed by a Max pooling layer.

- The **Hybrid architecture** is implemented from scratch, also using components in common with the previous model.

---

[8]YAML Ain't Markup Language

These new implementations expand the number of possible models to find the best one for IACT data analysis, since different combinations are possible.
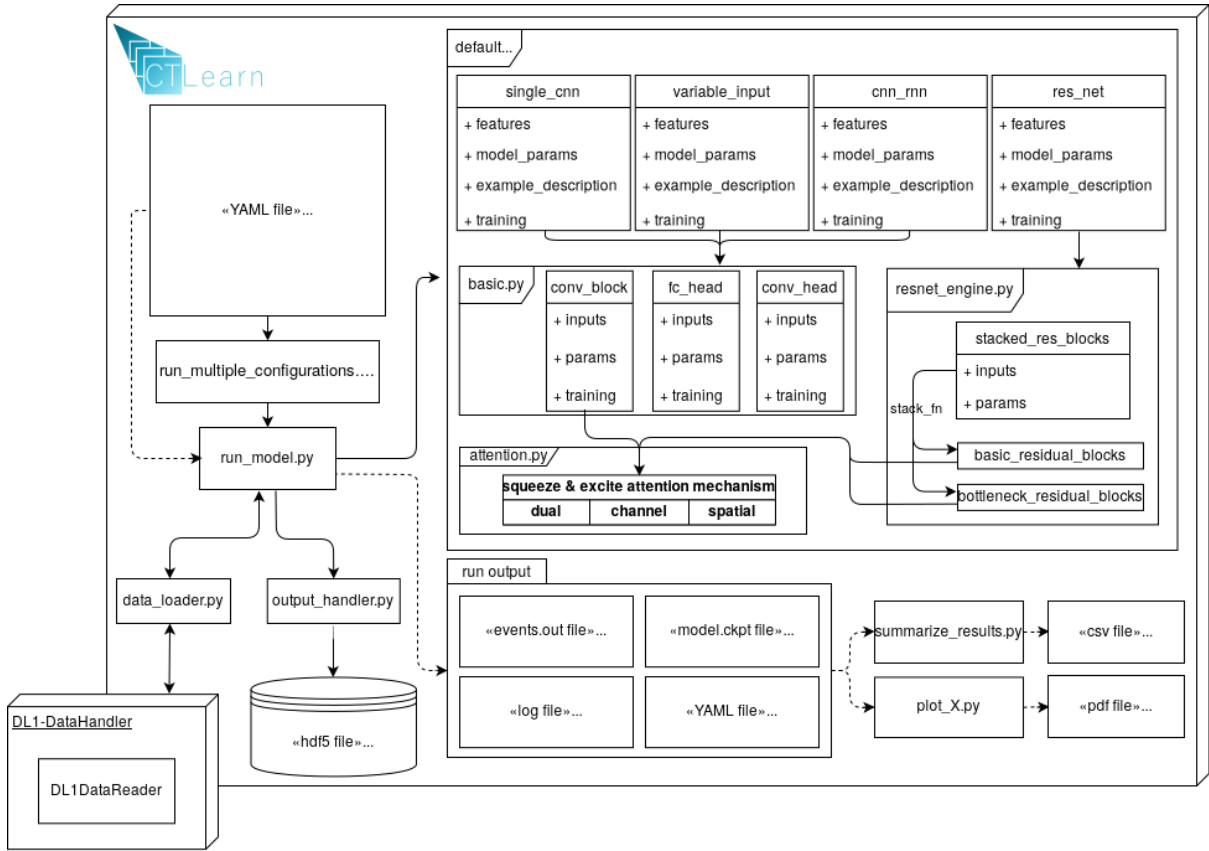
Figure 4.6: CTLearn diagram.

4.4 LST Event Reconstruction with CNNs

Recalling Chapter 2, in typical IACT observations, the number of γ rays is more than three orders of magnitude smaller than that of background events generated by CR-initiated showers. Therefore, IACTs must be able to recognize γ rays among the massive coming swarm of hadrons at certain Zd/Az values, evaluating their energy to study the spectral properties of the sources. Event reconstruction can be summarized in three main tasks:

1. **Signal/background separation**, a binary classification task. Each input event is given a score called gammaness;

2. **Energy reconstruction** of the γ-initiated events, a regression task with one output;

3. **Direction reconstruction**, a linear regression task with two outputs, i.e. The coordinates of the arrival direction of the source of γ rays.

As already mentioned, CNNs accomplish these three tasks by directly analyzing the raw pixel-wise information, avoiding the cleaning and parametrization steps, as shown in Figure 4.7.
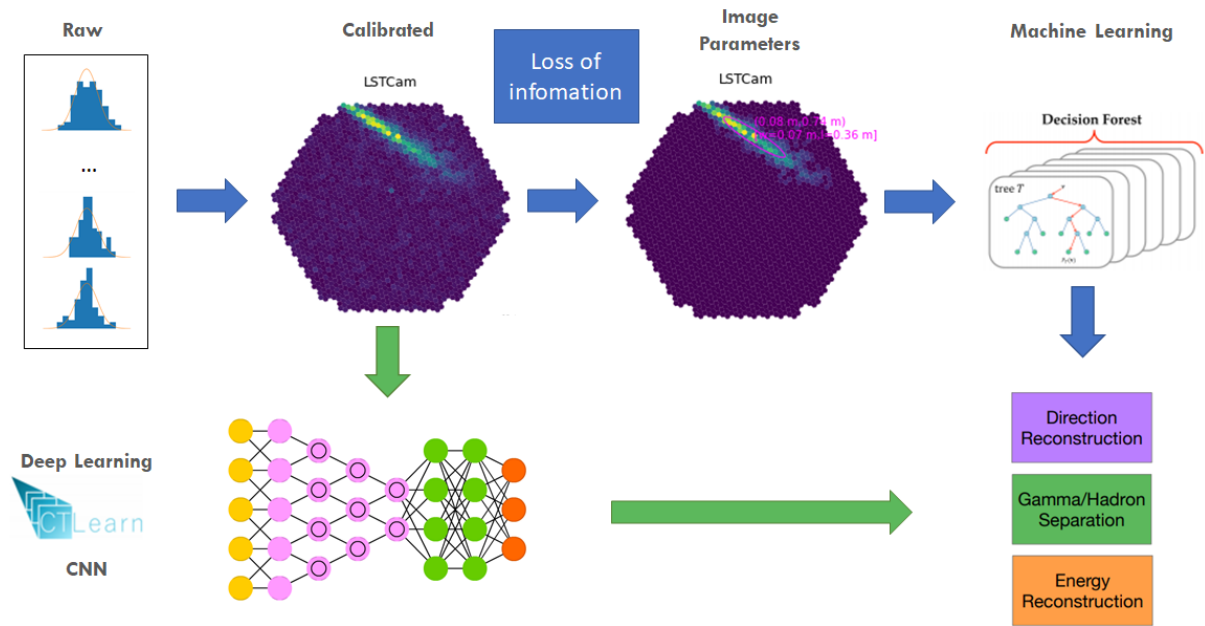
Figure 4.7: LST-1 data analysis pipeline. The standard pipeline is the one identified by the blue arrows while the experimental one, on which this thesis is based, is identified by the green arrows.

Since the various training process steps, such as convolution and backpropagation, are quite simple but very expensive computationally speaking, the use of GPUs is strongly recommended. Thanks to their bandwidth, GPUs can parallelize a great amount of calculations, speeding up the training phase and allowing to work with a massive dataset. This work was carried out using the INFN's cloud computing service Cloud Veneto - Area PD, with a VM that used two Titan XP GPUs connected in parallel. The analysis tools were developed in CTLearn, using Tensorflow 1.15. Details and results of the analysis are presented in Chapter 5.

Chapter 5

Analysis and results

In this chapter, the results obtained during the realization of this thesis are presented, together with the explanation of the analysis methods followed.

In the first section of this chapter, a description of the dataset used and the preprocessing steps executed to perform the experiments are shown. Then, in the following ones, the results obtained from the event reconstruction process for the LST-1 are presented and discussed. As it is illustrated in section 4.4, this process consists in separating the signal from the background and reconstructing the γ-ray energy and arrival direction. The proposed strategies are two: a modified VGG16 and a Hybrid architecture, both trained and tested on different selections of data, according to the cut-levels defined in Table 5.1. The image parameters used in the hybrid architecture training and used to define the cut-levels threshold are obtained with *Tailcuts_ clean(6,3)*. The performance obtained is at first compared between the two networks proposed, then with those obtained on Prod3b in Grespan, 2020 and the ones performed by the standard RF method.

5.1   Dataset

The dataset utilized in this thesis is a collection of simulated shower images, called Prod5 and produced in September 2020. Thanks to the first year of LST-1 observations, the knowledge of the technical conditions of the telescope has increased, allowing to perform these more realistic simulations. Therefore, they are much closer to representing reality than Prod3b, utilized in Grespan, 2020 and Marinello, 2019. The MC production was performed by Y. Ohtani, using the CORSIKA[1] program [Heck et al., 1998] with the IACT/ATMO[2] extension for the atmospheric simulation, while the whole telescope response was simulated with the sim-telarray[3] package. The dataset is composed of six telescopes, 4 LSTs, MAGIC I and MAGIC II, accounting for events that triggered at least one of them. Given the purpose of this work, only the trigger events of LST-1 are considered.

The shower images belong to four classes of particles:

- **pointlike γ rays**: Differentiated by the telescope pointing position with respect to the source: γ pointlike 0.4, $\sim 2.3 \cdot 10^6$ events, are generated with an offset of $0.4°$ while γ pointlike 0.0, $\sim 2.6 \cdot 10^6$ events, are generated with an offset of $0.0°$ (center

---

[1]https://www.iap.kit.edu/corsika/
[2]https://www.mpi-hd.mpg.de/hfm/bernlohr/iact-atmo/
[3]https://www.mpi-hd.mpg.de/hfm/bernlohr/sim_telarray/

of the camera) and with energies spanning from 5 GeV up to 50 TeV. Pointlike γ rays are used during testing time;

- **diffuse γ rays**: $\sim 2.5 \cdot 10^6$ events generated in a view cone of 6° , with energies ranging from 5 GeV up to 50 TeV. They are used for the test part of the CNN production;

- **protons**: $\sim 1.6 \cdot 10^6$ events generated in a view cone of 10° around the telescope pointing position, with energies spanning from 10 GeV up to 100 TeV;

- **electrons**: $\sim 1.4 \cdot 10^6$ events generated in a view cone of 6° around the telescope pointing position, with energies spanning from 5 GeV up to 5 TeV.

All the events are simulated at a zenith angle = 20° and an azimuth = 180°.

As illustrated in Chapter 3, IACTs indirectly observe γ rays, but point directly to their sources since VHE γ-ray events are expected to originate from a certain location in the sky and arrive with unchanged direction. Training only on pointlike events would introduce a bias in the learning process, nullifying any effort to generalize it. To overcome this problem, diffuse γ rays are used for training while pointlike γ rays for testing.

5.1.1 Image Preprocessing

The state-of-the-art python API for CNN analysis require the images to have rectangular pixels, for that reason it is necessary to apply a pre-processing to LST images, where pixels have hexagonal shape.

In this work, the image pre-processing step is performed by the DL1DH *imageMapper* module, which implements different resampling methods, such as rebinning and interpolation, as described in Shilon et al., 2019. Each method has its pros and cons, such as intensity conservation (rebinning) and shape conservation (interpolation). In this work, the bilinear interpolarion was preferred, since it demonstrates to achieve better performance on IACT images (Brill et al., 2019). In detail, the *ImageMapper* performs the bilinear interpolation by mapping the input pixels, in the hexagonal layout, to pixels in the Cartesian output layout through a weighted linear combination. Weights can be stored in a sparse array dubbed a mapping table, with shape (*camera pixels, output x, output y*). After this step, the *ImageMapper* selects, for each output pixel, its three closest input pixels, which will be involved in the interpolation through the Delaunay triangulation method, as shown in 5.1. The normalization over the mapping tables allows to approximately preserve the input image charge after the interpolation (Nieto, Brill, Feng, Jacquemont, et al., 2020).
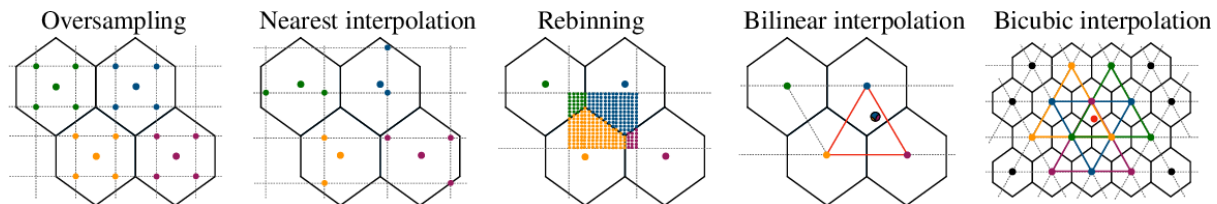
Figure 5.1: Diagrams depicting all the implemented mapping methods in DL1DH (Nieto, Brill, Feng, Jacquemont, et al., 2020). Courtesy of CTLearn Team.
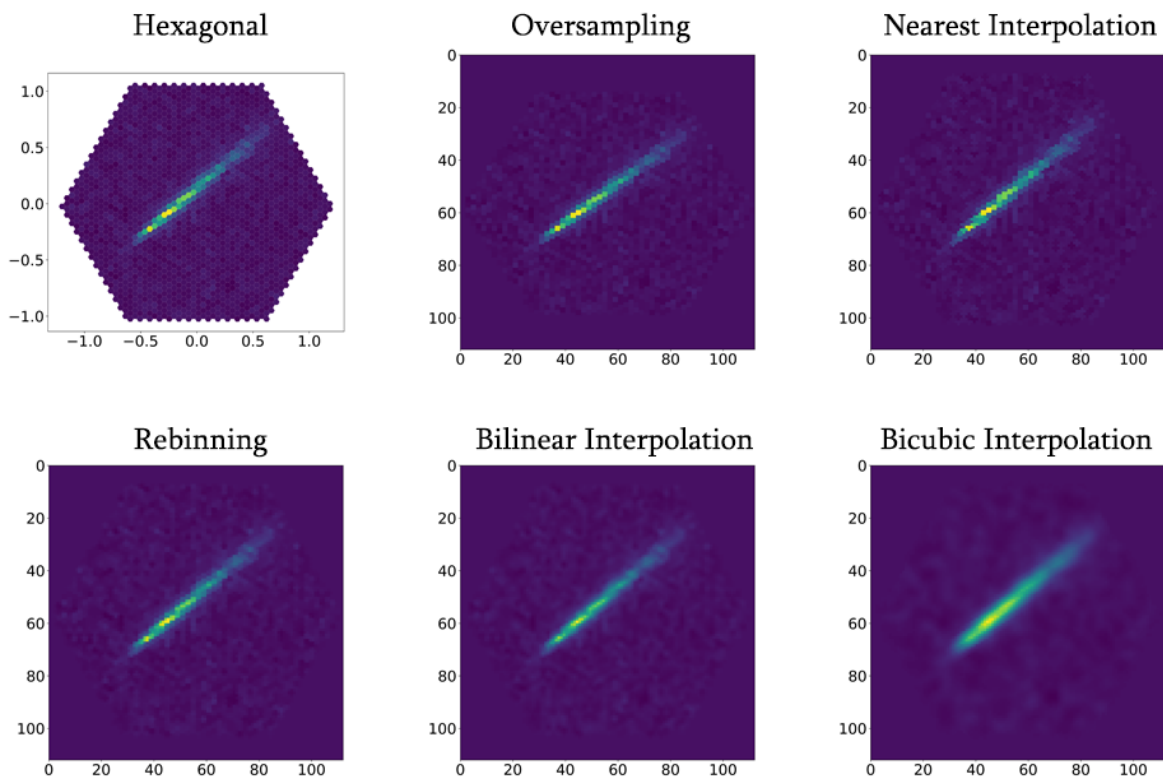
Figure 5.2: Image from a simulated γ-ray event as seen by a camera with pixels arranged in a hexagonal lattice (top left), and the result of applying each of the implemented mapping method in DL1DH(Nieto, Brill, Feng, Jacquemont, et al., 2020). Courtesy of CTLearn Team.

5.1.2 Data selection

The simulated events, due to their spatial distribution and stochastic component, are imaged with different quality levels. This quality is strongly correlated to the total intensity of the image. Showers from higher energy events produce more Cherenkov light, and will show images with higher intensity and more definite shape, while for images with lower intensity the shower may not be fully contained in the camera. Making cuts in the intensity would improve the overall performance of the reconstruction, at the expense of raising the energy lower threshold of the telescope. One of the aims of this work is to investigate the impact of the cut levels shown in Table 5.1, on the network performances.

Cut level	Intensity threshold [phe]	Leakage2 threshold
no-cut	None	None
low-cut	50	0.2
mid-cut	200	0.2
high-cut	1000	0.2

Table 5.1: The four cut levels adopted in the analysis.

The effect of the different cuts on the dataset can be observed on the event energy distributions, shown in Figure 5.3. Since each cut-level is basically a filter on two proprieties of the image, the intensity and the leakage, changing the cleaning algorithm influences the filter itself. Indeed, harder cleaning results in a less noisy image and lead to more separated spectra of each cut-level, as shown in the same figure.
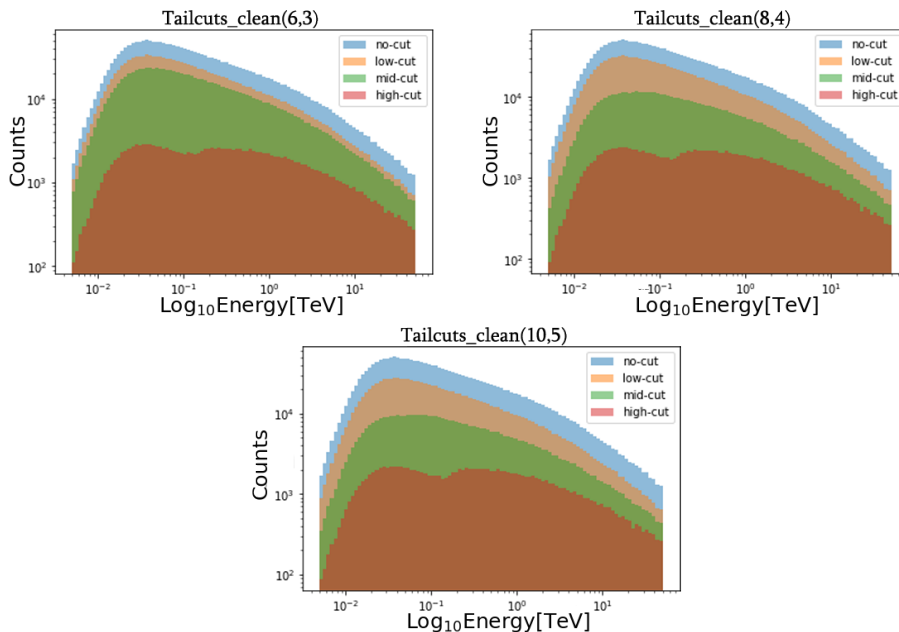
Figure 5.3: Training set energy distributions of diffuse γ rays with the three different cleaning algorithms. Note the impact of harder cleanings on the energy mid range, decrease the number of images belonging to the mid-cut and low-cut levels compared to the standard ones.

For each task, both networks have been trained with the same hyperparameters: a batch size of 128 and the *Adam* optimizer, with a LR value close to $0.1 \cdot 10^{-3}$ or $0.1 \cdot 10^{-4}$, depending on the cut-level. The number of epochs have been decided based on the validation loss behaviour obtained in different training processes. The aim was to find the best one for each cut-level. The value selected represents the number of epochs required to arrive at a minimum, a flat zone where the loss does not decrease anymore, or right before the loss increases, a symptom of overfitting. This strategy is a very basic emulation of the early stopping (Yao et al., 2007), which unfortunately is not implemented yet in CTLearn. The models have been then tested on the pointlike γ rays, as already illustrated in section 5.1, and evaluated by comparing the prediction with the ground truth, i.e. the shower simulation input parameters.

5.2   Gamma/Hadron separation

The first operation to be accomplished, requires a training set composed of diffuse γ rays and protons in equal quantity, $\sim 1.34 \cdot 10^{6}$ events each. The loss function chosen to minimize is the softmax cross-entropy, usually used in multi-classification tasks, since both networks have two neurons in the last layer. This function estimated the gammaness for each event, the histogram results are shown in Figure 5.4. The Area Under Curve (AUC) has been used to evaluate performance, which is a measure, with a value between 0 and 1, that represents how well the model is able to distinguish between classes. The higher the AUC, the better the model predicts, while an AUC value of 0.5 indicates that the model is unable to make predictions, behaving like a random decision maker.

AUC has been performed on a test set of $\sim 2.3 \cdot 10^{5}$ γ pointlike events and $\sim 2.5 \cdot 10^{5}$ proton events, by calculating the true positive rate ($TPR = TP/(TP+FP)$) against the false positive rate ($FPR = FP/(FP+TP)$), where T/F stand for *True/False* and P/N for *Positive/Negative*. Both the proposed architectures perform very close together; the

Cut level	Hybrid	Mod. VGG16	Random Forest	Prod3b
nocut	0.866	0.870	0.752	0.868
lowcut	0.873	0.858	0.794	0.883
midcut	0.919	0.905	0.877	0.928
highcut	0.989	0.984	0.975	0.986

Table 5.2: AUC approximated value for each cut levels adopted in the analysis.

difference is rather nil in the cuts at the extremes where the modified VGG16 performs better than the hybrid architecture for *no-cut*, by ~ 0.04, while for *high-cut* the hybrid one performs better by ~ 0.05 . For the core cuts hybrid architecture has better performance than the VGG16, increasing the *low* and *mid-cut* performance by ~ 0.015 and ~ 0.014, respectively.

Comparing the Prod5's AUCs with those obtained in Prod3b, shown in Table 5.2, it is evident that *low-cut* and *mid-cut* perform worse in Prod5, with a difference of ~ 0.01 and ~ 0.009, respectively, while *high-cut* and *no-cut* perform very closely with the corresponding one, if not better. Looking at 5.2, it is clear that the proposed DL models outperform the RF model among all the cuts; the difference is more perceptible at lower

energies where RF performance decreases of ~ 0.14 for *no-cut* and ~ 0.065 for *low-cut.*
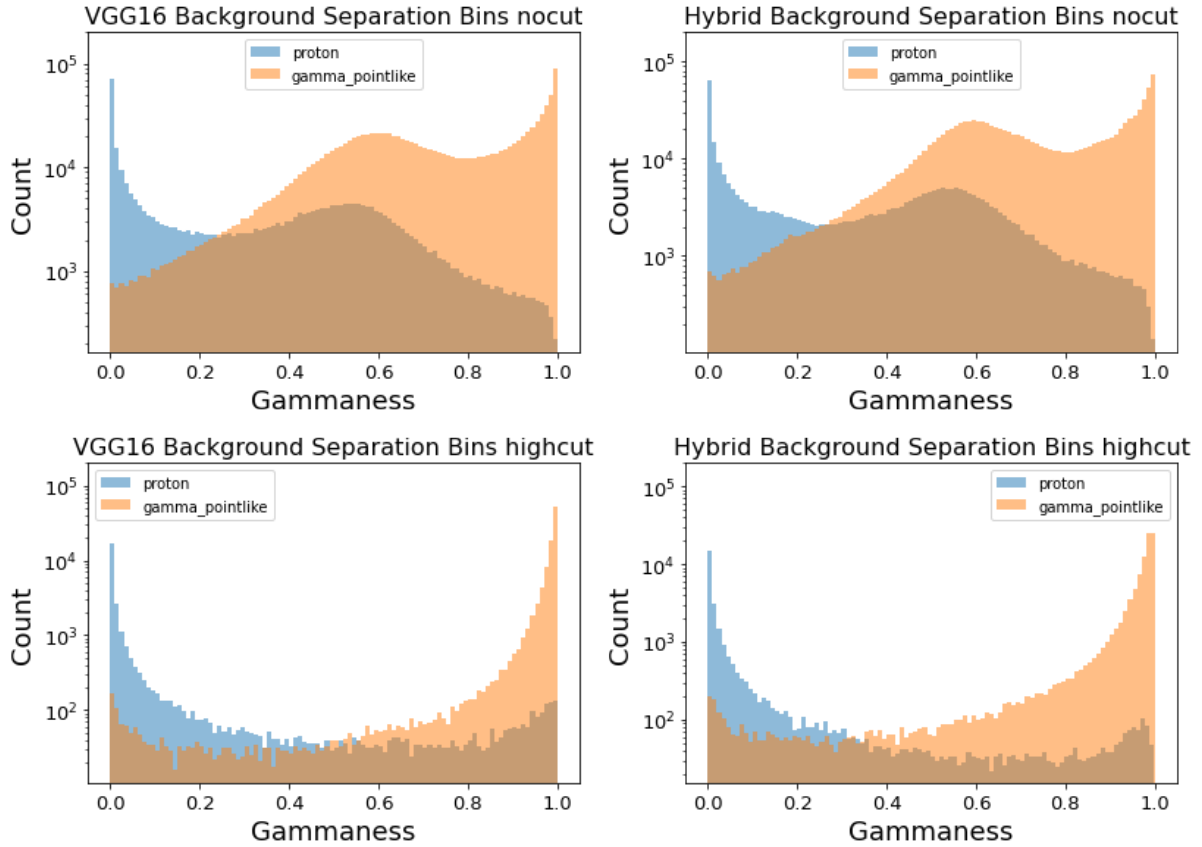
Figure 5.4: Gammaness distribution of pointlike γ rays and protons for the *no-cut* (first row) and *high-cut* (second row). It can be seen that for the highest cut, the networks classify the particles with more confidence, unlike the lower cut where a large component of events is classified with uncertainty, i.e. gammaness = 0.5 .

5.3   Energy reconstruction

For this task, the whole set of $\sim 2.5 \cdot 10^6$ diffuse γ-ray events has been used for the train/validation and the networks were trained with the values of the hyperparameters illustrated at the beginning of this chapter. Given the purpose of this task, the output of the model is the energy of the γ-ray event: $\epsilon =$ E. The loss function chosen was the absolute difference: various internal studies by the CTLearn and GammaLearn[4] teams agreed that it was the best loss for regression activities on these data.
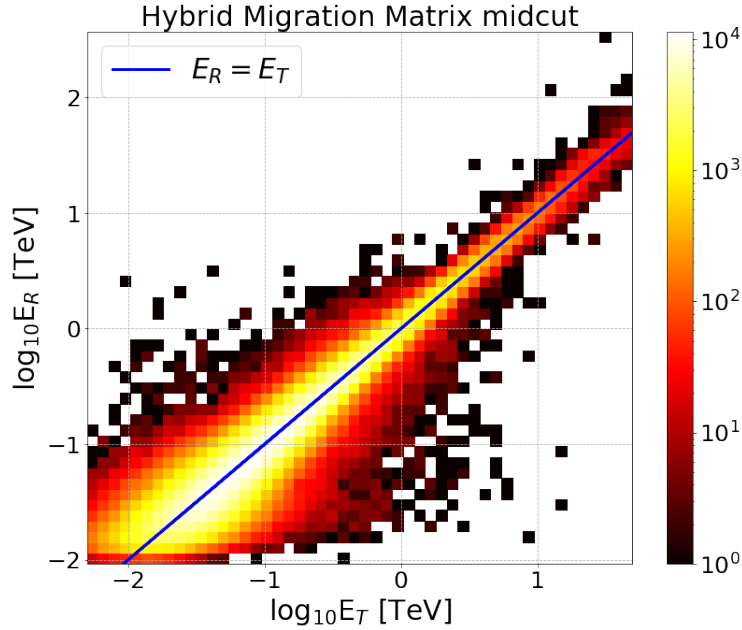
Figure 5.5: Energy migration matrix of *mid-cut* of the Hybrid network. The blue line represents the ideal reconstruction.

To visualize the performances of cut-level the *migration matrix*, which is the 2D-histogram of the reconstructed energy against the MC energy, as shown in Fig 5.5, is plotted; while, for evaluating them, the energy spectrum was split in bins where, for each bin, the relative energy error is histogrammed, namely:

$$\frac{\Delta E}{E} = \frac{E_R - E_T}{E_T} \tag{5.1}$$

where T stands for "true" and R for "reconstructed". The distribution is then bias-corrected, i.e. the median is subtracted, and the energy resolution is defined as the 68th percentile of the histogram $|E_R - E_T|_{corr} / E_T$.

---

[4]https://gitlab.lapp.in2p3.fr/GammaLearn/GammaLearn
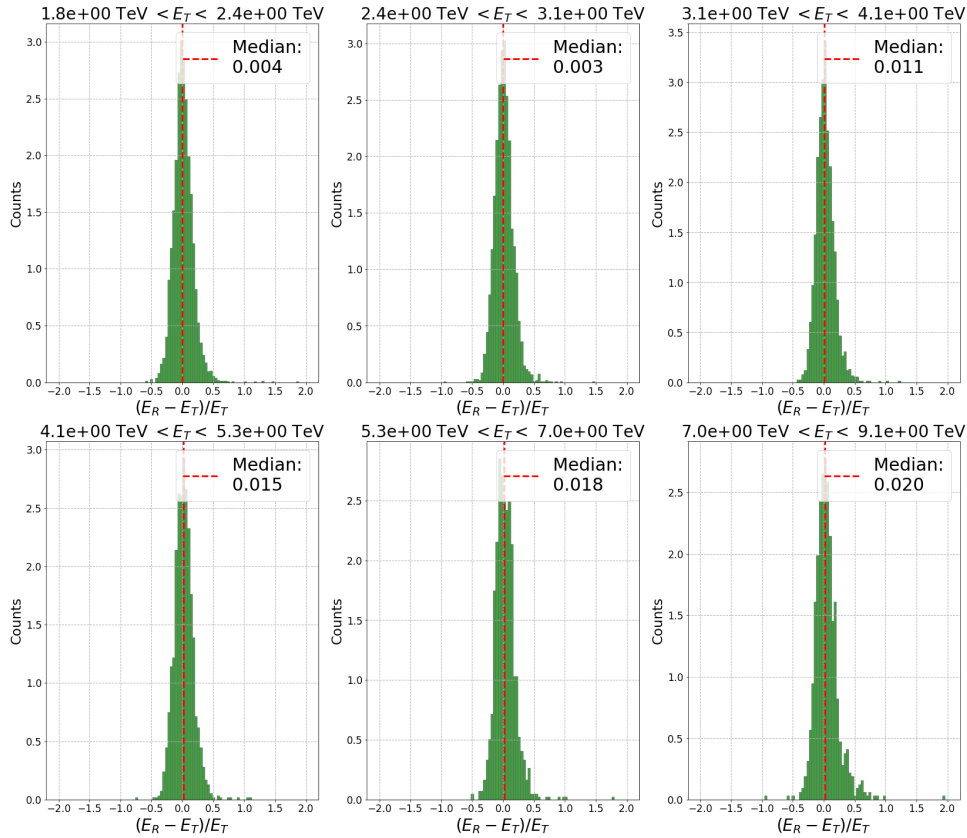
Figure 5.6: Five histograms of the relative energy error, regarding the range of 1.8 TeV - 9.1 TeV, obtained with the *high-cut* model. The median of each distribution represents the relative energy bias.
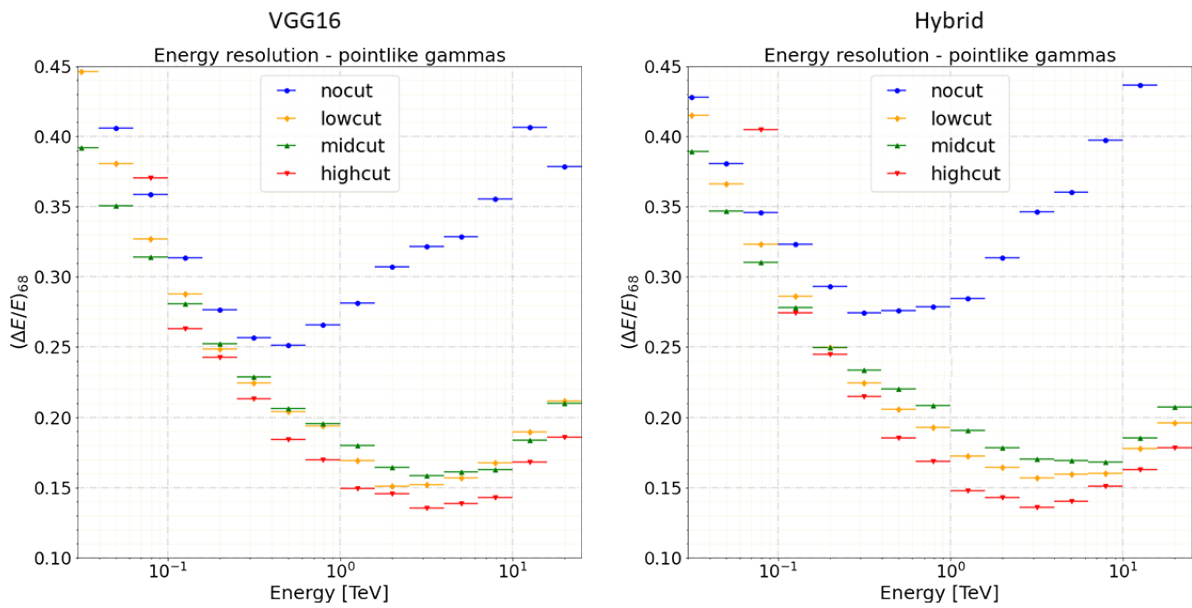
Figure 5.7: Energy resolution curves for the Hybrid architecture (right) and for the modified VGG16 (left).

37

Figure 5.7 shows the energy resolutions for the different cut levels of both architectures. It is expected that *high-cut* has worst performance at the lowest energies (< 100 GeV) since the higher cut in intensity discards the lower energy events. Observing both architectures separately, the *low*, *mid* and *high-cut* show the same trend across all the energy ranges, reaching their best performance at 500 GeV-10 TeV; while the *no-cut* reconstruction ability degrades before 1 TeV probably due to the large number of images that contain a very small fraction of the shower, or do not contain it at all. As expected, the best performance is achieved at 500 GeV-10 TeV since, at lower energies (< 500 GeV) resolution is poorer due to the less luminous cascade produced by low-energy γ rays, resulting in less light collected in the camera; meanwhile, the deterioration in performance for higher energies (> 10 TeV) is clearly due to the lower statistics at higher energy, due to the simulation spectrum. Furthermore, it is thought to be fault of brighter showers. This condition leads to more events with a large impact parameter[5] which trigger more telescopes but produces shower images with few photoelectrons, as for the lowest energies.
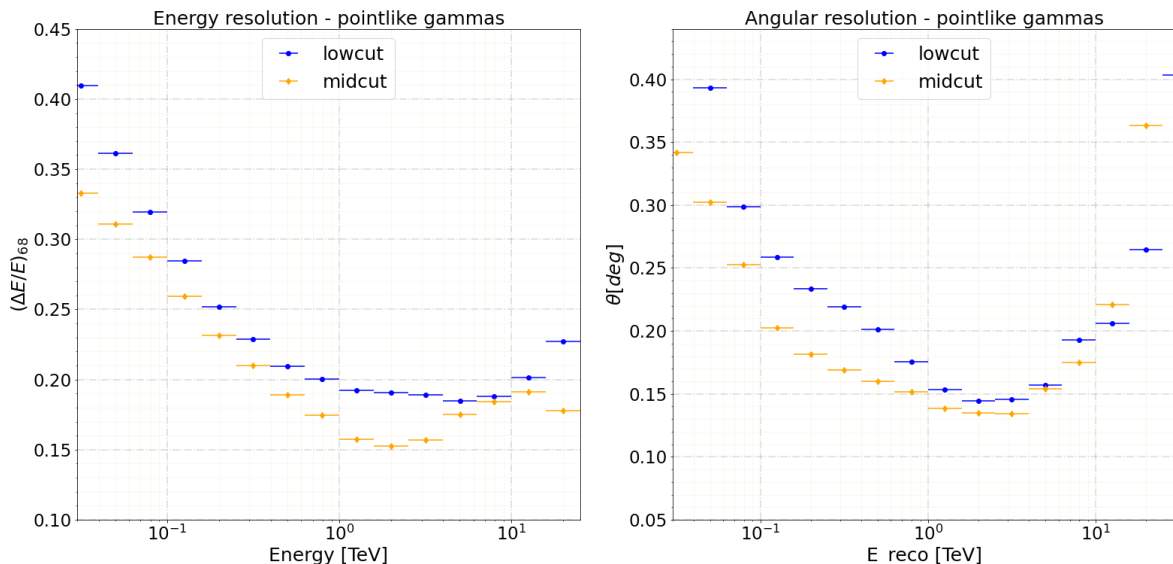
Figure 5.8: Energy resolution angular resolution curves for the Hybrid architecture with the *Tailcuts_ clean(10, 5)*.

Comparing the proposed architectures in Figure 5.7, the best performance is achieved by the modified VGG16 basically in all the cuts. Also, it is possible to notice that *low-cut* performs slightly better than *mid-cut*, which is more evident in the Hybrid one. We performed some tests to study this unexpected behaviour and found that the current default image cleaning that is applied is too low and therefore leads to inconsistency in the results of images not properly cleaned. This is more accentuated in Prod5 results because of the different strategies for signal integration and the more-up-to-date parameters used for the simulation. We tested this hypothesis by performing the same analysis in data cleaned with higher cleaning levels *Tailcuts_ clean(10, 5)* and the results are shown in Fig. 5.8. As it can be seen, the results follow the expected hierarchy. The LST collaboration is aware of this fact and is working on a more sophisticated cleaning (using arrival time information and noise fluctuations as well) that will allow for a compromise between

---

[5]The distance between the shower core projected onto the ground and the telescope.

reaching the lowest possible energy threshold without accepting too many signal from noise fluctuations and that will be released in the forthcoming cta-lstchain v0.7.0 .

The confrontation between Prod5 and Prod3b, Fig. 5.9, highlights a common behaviour of the performance curves and worse performance in all the cuts, as expected, except for both the *high-cut* models of the Prod5, which have better performance. A substantial difference is the equal and best performances of filtered models in the 1 - 3 TeV of Prod3b, while in Prod5 curves are more divided. Lastly, comparing the two neural networks proposed with the RF, Fig. 5.10, it is evident that DL strategies perform better. In detail, for *low* and *high-cut*, they outperform the standard method at all the energies; while for *mid-cut*, the performance is pretty equal in the range 200 GeV - 5 TeV.
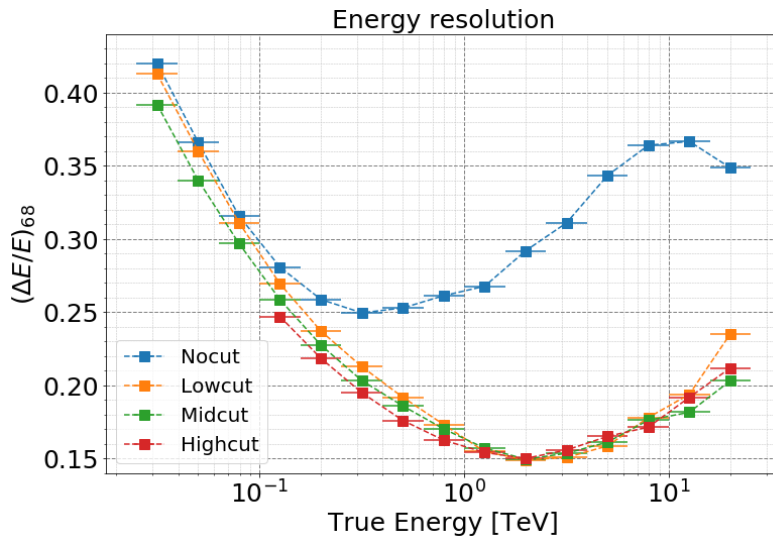
Figure 5.9: Energy resolution of the modified VGG16 on Prod3b. From Grespan, 2020.

Figure 5.10: Comparison of energy resolution between Hybrid architecture (Orange), modified VGG16 (Blue) and Random Forest (Green).

5.4 Arrival Direction reconstruction

In this task the aim of the network is to reconstruct the two coordinates, the *altitude* $\hat{\alpha}$ and the *azimuth* $\hat{\rho}$, which describe the arrival direction of the γ-ray initiated shower. To learn how to perform this regression task, the MC data coordinates are provided. Since LST-1 has a different coordinate-frame, this two values need to be converted into the alt-azimuthal difference between the source position and the pointing direction of the telescope: $\mathbf{z} = (\Delta\hat{\alpha}, \Delta\hat{\rho})$. As for the energy reconstruction task, the loss function chosen was the absolute difference.
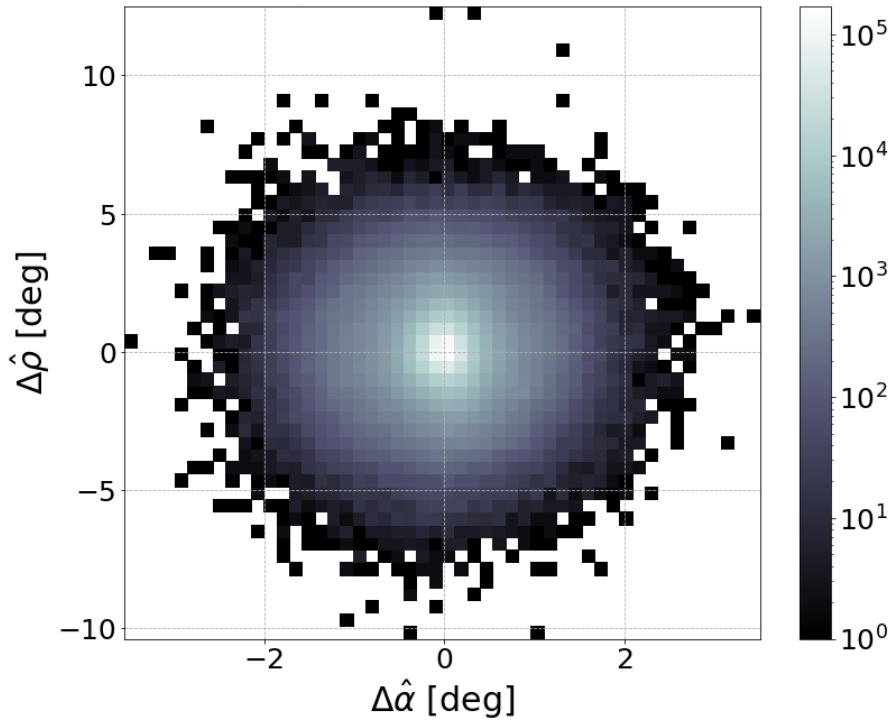
Figure 5.11: 2D distribution of the reconstructed position of events for the Hybrid network, with the *no-cut* filter. The coordinate (0,0) represents the ideal reconstruction.

The 2D distribution of the reconstructed position of events can be seen in Fig. 5.11, While, to evaluate the performances of each cut-level the θ^2 plot is performed for each energy bin. θ^2 plot is the histogram of the squared angular distance (θ) between the estimated shower arrival direction and the expected source position; it displays the quantity:

$$\theta_i^2 \equiv \|\mathbf{z}_i - \hat{\mathbf{z}}_i\|^2 \tag{5.2}$$

where $\hat{\mathbf{z}}$ represents the true coordinates and $\mathbf{z}$ the reconstructed counterpart. From each θ^2 histogram, examples in Figure 5.12, is extracted the angular resolution for that energy bin, defined as the square root of the 68% containment radius, shown in Figure 5.13
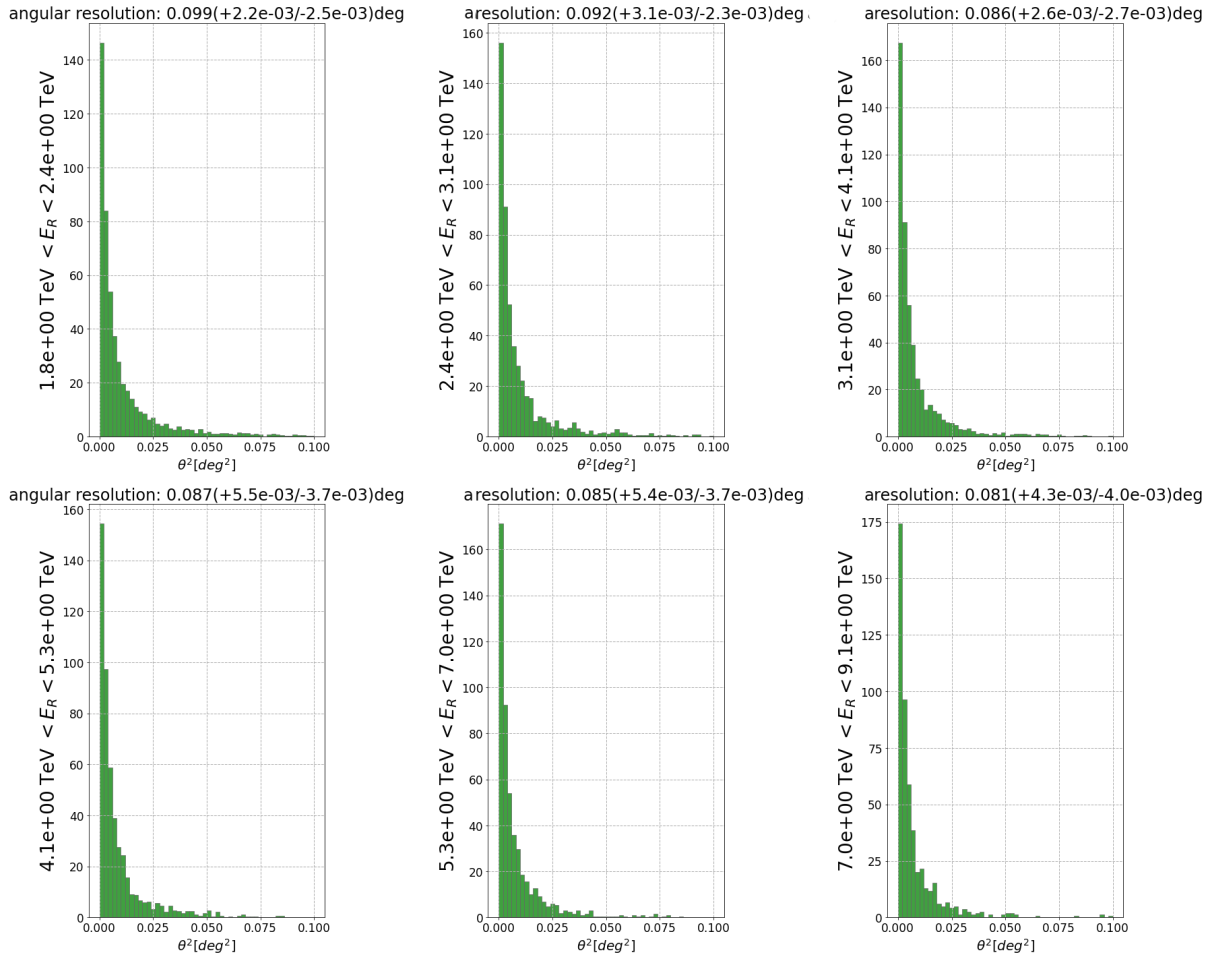
Figure 5.12: θ^2 histograms of test events in the 1.8 TeV - 9.1 TeV range, obtained with the *high-cut* model. The 68th percentile, used to calculate angular resolution, is displayed at the top of each plot.
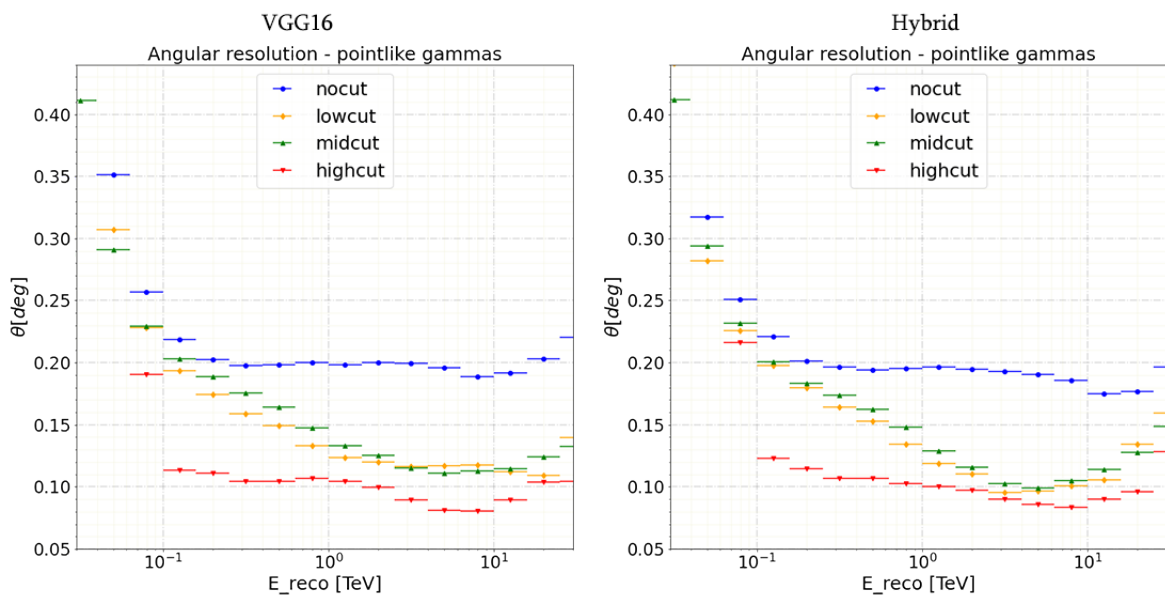
Figure 5.13: Angular resolutions for all the cut-levels (lower is better).

42

Fig. 5.13 highlights that, increasing the energies, all models acquire better angular resolution; only *no-cut* cannot improve as much as the others. At energies > 10 TeV, performance starts worsening because of the lower quality of the images, as the energy reconstruction task. *Low* and *mid-cut* have slightly different performance, very close together, once again due to the cleaning, as it was discussed already in the previous section. We repeated the tests with higher cleaning levels (see Fig. 5.8) and we can see that the expected behaviour is reproduced.

Comparing both architectures, the best performance is achieved by the Hybrid architecture in all the cuts, except for the VGG16 *high-cut* model, which performs better in the 3 - 10 TeV energy range. Comparing the two productions, we can notice a worse performance of Prod5 with respect to Prod3b in the *low* and *mid-cut* and very similar one for the *high*. *No-cut* is the only one which performs better with respect to the Prod3b. Lastly, comparing neural networks and RF performance, Fig. 5.15, for the arrival direction reconstruction, is notable that DL improvement is definitely evident since RF has worse performance in all the energy range, in particular at lower energies.
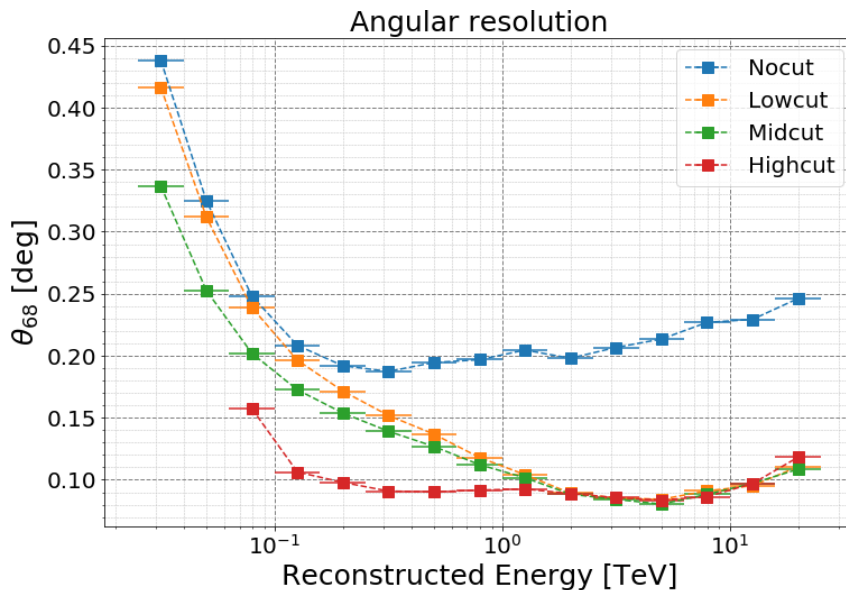
Figure 5.14: Angular resolution of the modified VGG16 on Prod3b. From Grespan, 2020.
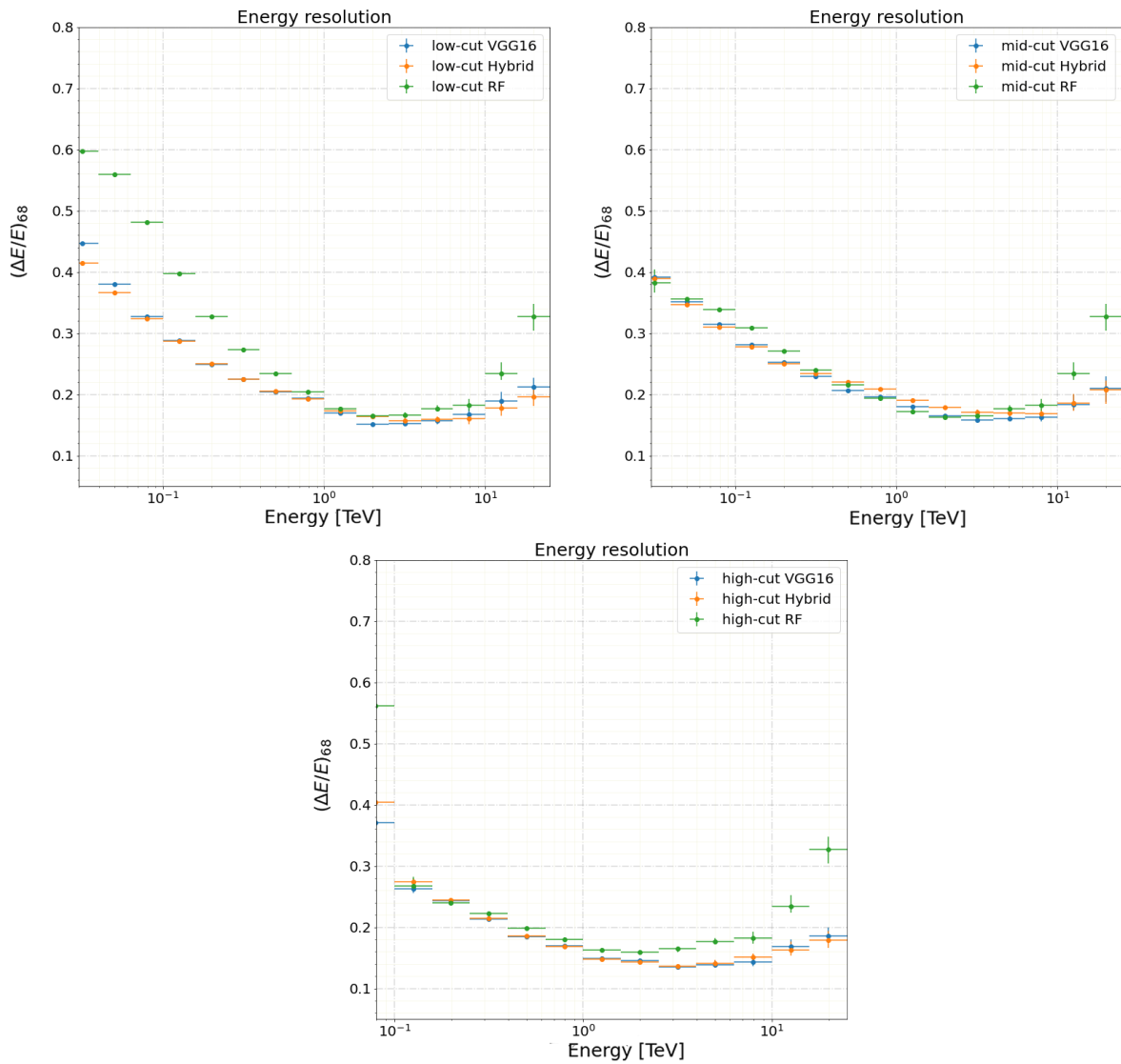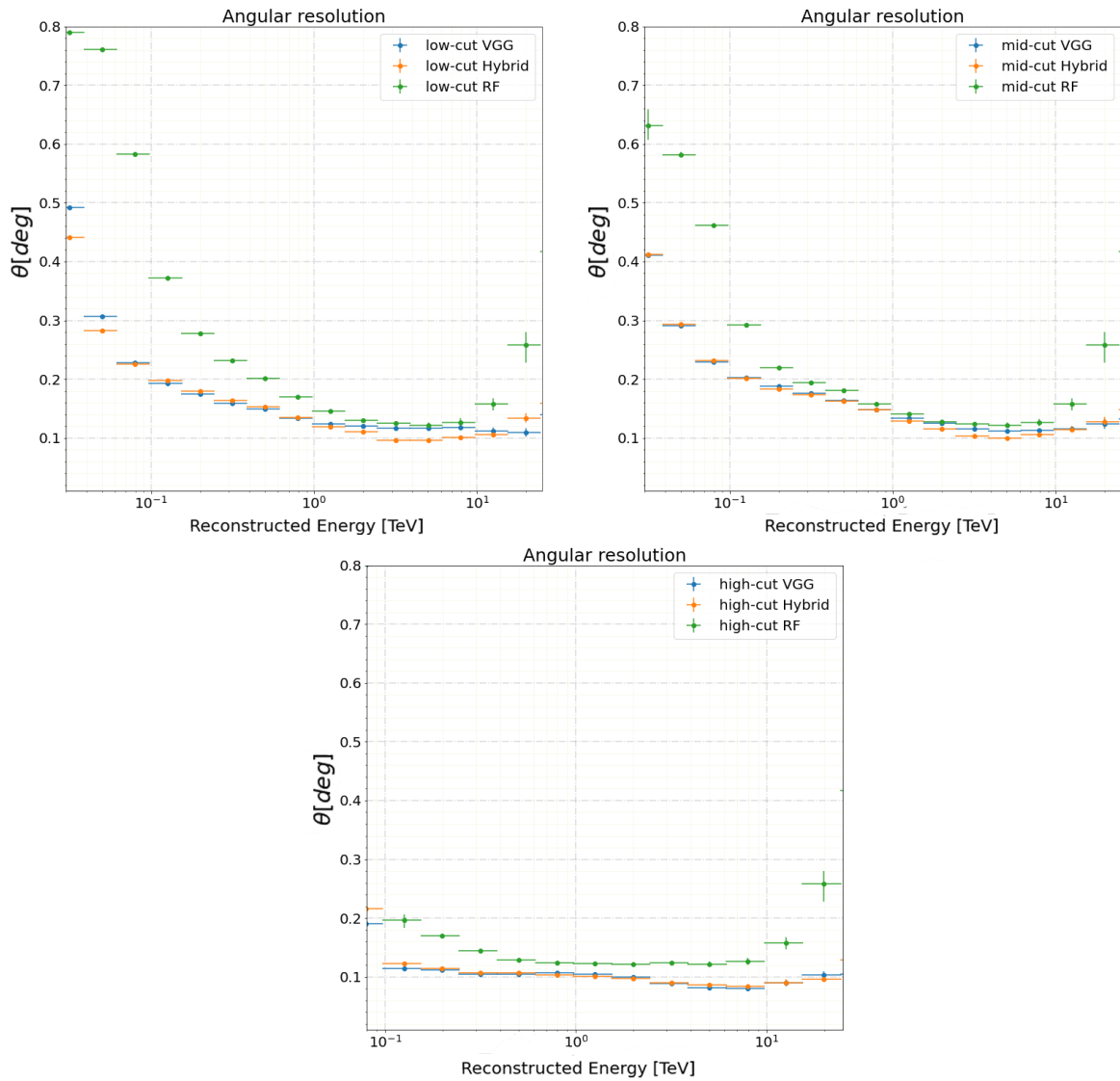
Figure 5.15: Comparison of angular resolution between Hybrid architecture (Orange), modified VGG16 (Blue) and Random Forest (Green).

Chapter 6

Conclusions and outlook

This thesis represents a perfect example of the powerful impact of AI in the research field by developing a new analysis technique for a new generation of telescopes, surpassing the standard and current analysis. These results are very encouraging as the dataset used is a collection of simulations, based on parameters measured on real observations, and therefore more representative of reality.

Particularly, this thesis proposes two different strategies, with two different approaches, for the Large-Sized Telescope Prototype of CTA. Both are based on CNN technology: the modified VGG16, which represents a common architecture for image analysis tasks in several domains and the hybrid architecture is a new model for IACT data analysis. While VGG16 was already applied to the LST analysis, it is the first time that an hybrid architecture is used to perform full data analysis. Networks are trained on different data quality and energy cuts (*no-*, *low-*, *mid-* and *high-cuts*), showing the crucial role of image quality for better performance in the three tasks. With no surprise, the *no-cut* models always perform worst than any other cut, the reason is addressed to the great number of poor quality images, where showers are not well-contained or faint.

To evaluate the results obtained on the cut-levels, three different comparisons were made:

- A comparison between Prod5 MC simulations, which is based on the measured parameters using real data observations of the LST-1, and Prod3 MC simulations, whose events are simulated with ideal parameters;

- A comparison between the two proposed architectures, a modified VGG16 and the Hybrid one;

- A comparison between the DL models and the standard RF model.

On the **background rejection** task the proposed models significantly outperform the RF by increasing overall the AUC: *low-cut* $\sim 9.9\%$, *mid-cut* $\sim 4.8\%$, *high-cut*: $\sim 1.4\%$. DL strategies are so powerful that the *no-cut* performance is close to the one of the RF *mid-cut*. Both the presented architectures share pretty much the same values, even if the Hybrid architecture performs slightly better; while, comparing to Prod3b, only *low-* and *mid-cut* performance worsens.

Regarding **energy reconstruction**, VGG16 generally performs better than the hybrid architecture, also showing similar trends. RF is outperformed by both the DL models in the lowest energies of *low-cut*, in the higher energies for *high-cut* but has pretty much the same trend in the *mid-cut*. Comparing the two productions it is evident the strange

behaviour of *low-* and *mid-cut*, but trends are similar.

Both the networks show the best performance increase, with respect to RF, in the **arrival direction reconstruction** by outperforming it in all the energy range. This is so evident in the lower energies of *low-cut*. Best performance is achieved by the Hybrid architecture, which reach an angular resolution down to 0.081°-0.105° for all the cuts except *no-cut*.

A common behaviour in all the tasks is the performance decreasing, in comparison with Prod3, and the strange trend of *low-* and *mid-cut*. This is addressed to the cleaning algorithm *Tailcuts_ clean(6,3)*, demonstrating that it is too soft, even if it is the standard one, since this new production is noisier than the previous ones. New results were produced using an harder cleaning algorithm: *Tailcuts_ clean(10,5)* and we demonstrated that the expected behaviour of the CNN performance is recovered. At the moment the LST collaboration is implementing more sophisticated cleaning algorithms to maintain a good performance at the lowest energies without the expense of accepting several signals coming from spurious events.

Concerning the two software libraries employed, DL1DH and CTLearn, this thesis push further their functionality, defining a new file format that works with various IACT cameras, not only for LST, and implementing new DL architectures which can be easily re-used in future analyses. The whole implementation of the new code has been finalized to allow the user to perform a large number of combinations by deciding a large number of hyperparameter values, such as the number of layers and neurons, but also, the number and the type of cleaning algorithms.

All the results contained in this thesis confirm that DL methods, in particular the use of CNNs, can be the future methods to be used in the data analysis of IACT. They perform better than the standard parameter-based analysis methods currently used, providing solid results. This thesis is not the first that highlights this aspect, even in the ones of Grespan, 2020 and Marinello, 2019 CNNs outperform RF, enforcing the superiority of DL. There is certainly the possibility of further improvement and testing new architectures, maybe focusing on extreme conditions, as cloudy night or moon sky. However, this thesis and the previous ones have shown very promising results, performing better than the standard analysis. This makes neural networks ready for the big step of using real data for their analysis.

Appendix A

Modified VGG16 architecture

Layer type	Layer
Batch Normalization	-
Conv2D 1	3×3, 64, stride 1
Conv2D 2	3×3, 64, stride 1
Max Pooling	2×2, stride 2
Conv2D 3	3×3, 128, stride 1
Conv2D 4	3×3, 128, stride 1
Max Pooling	2×2, stride 2
Conv2D 5	3×3, 256, stride 1
Conv2D 6	3×3, 256, stride 1
Conv2D 7	3×3, 256, stride 1
Max Pooling	2×2, stride 2
Conv2D 8	3×3, 512, stride 1
Conv2D 9	3×3, 512, stride 1
Conv2D 10	3×3, 512, stride 1
Max Pooling	2×2, stride 2
Conv2D 11	3×3, 512, stride 1
Conv2D 12	3×3, 512, stride 1
Conv2D 13	3×3, 512, stride 1
Max Pooling	2×2, stride 2
Global Average Max Pooling	-
Fully Connected	1-d (2-d) softmax (linear)

Table A.1: Architecture of the VGG16. The last layer represent the two possible scenarios: 1-d for 1 neuron in regression tasks while 2-d for 2 neurons in classification task, where d stay for dimension.

Appendix B

Hybrid architecture

Convolutional	macro block		
Layer type	**Layer**		
Batch Normalization	-		
Conv2D 1	3×3, 64, stride 1		
Conv2D 2	3×3, 64, stride 1		
Max Pooling	2×2, stride 2		
Conv2D 3	3×3, 128, stride 1		
Conv2D 4	3×3, 128, stride 1		
Max Pooling	2×2, stride 2		
Conv2D 5	3×3, 256, stride 1		
Conv2D 6	3×3, 256, stride 1		
Conv2D 7	3×3, 256, stride 1		
Max Pooling	2×2, stride 2		
Conv2D 8	3×3, 512, stride 1	**Image Params**	**macro block**
Conv2D 9	3×3, 512, stride 1	**Layer type**	**Layer**
Conv2D 10	3×3, 512, stride 1	Fully Connected 4	64
Max Pooling	2×2, stride 2	Fully Connected 5	128
Global Average Max Pooling	-	Fully Connected 6	256
Fully Connected 1	64	Fully Connected 7	512
Fully Connected 2	64	Fully Connected 8	256
Fully Connected 3	12	Fully Connected 9	12
Concatenating output layers			
Fully Connected 10	2-d (1-d) softmax (linear)		

Table B.1: Hybrid architecture. Output layers of the two macro blocks are merged with *tf.keras.layers.concatenate()* to create a unique input layer for the last layer. The last layer represent the two possible scenarios: 1-d for 1 neuron in regression tasks while 2-d for 2 neurons in classification task, where d stay for dimension.

Bibliography

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., . . . Zheng, X. (2016). Tensorflow: A system for large-scale machine learning, In *12th usenix symposium on operating systems design and implementation (osdi 16)*. https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf

Abbott, B. P., Abbott, R., Abbott, T. D., Abernathy, M. R., Acernese, F., Ackley, K., Adams, C., Adams, T., Addesso, P., Adhikari, R. X., Adya, V. B., Affeldt, C., Agathos, M., ..., & Zendri, J.-P. (2016). Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.*, *116*, 061102. https://doi.org/10.1103/PhysRevLett.116.061102

Acciari, V. A., Aliu, E., Arlen, T., Aune, T., Bautista, M., Beilicke, M., Benbow, W., Boltuch, D., Bradbury, S. M., Buckley, J. H., Bugaev, V., Byrum, K., Cannon, A., Celik, O., Cesarini, A., Chow, Y. C., Ciupik, L., Cogan, P., Colin, P., . . . Collaboration, T. V. (2009). A connection between star formation activity and cosmic rays in the starburst galaxy m82. *Nature*, *462*(7274), 770–772. https://doi.org/10.1038/nature08557

Acciari, V., Ansoldi, S., Antonelli, L. A., Arbet Engels, A., Arcaro, C., Baack, D., Babić, A., Banerjee, B., Bangale, P., & et al. (2018). Discovery of tev γ-ray emission from the neighbourhood of the supernova remnant g24.7+0.6 by magic. *Monthly Notices of the Royal Astronomical Society*, *483*(4), 4578–4585. https://doi.org/10.1093/mnras/sty3387

Acciari, V., Ansoldi, S., Antonelli, L. A., Engels, A., Baack, D., Babic, A., Banerjee, B., Barres de Almeida, U., Barrio, J. A., Becerra-Gonzalez, J., Bednarek, W., Bellizzi, L., Bernardini, E., Berti, A., Besenrieder, J., Bhattacharyya, W., Bigongiari, C., Biland, A., Blanch Bigas, O., & Nava, L. (2019). Teraelectronvolt emission from the γ-ray burst grb 190114c. *Nature*, *575*, 455–458. https://doi.org/10.1038/s41586-019-1750-x

Albert, J., Aliu, E., Anderhub, H., Antoranz, P., Armada, A., Asensio, M., Baixeras, C., Barrio, J., Bartko, H., Bastieri, D., Becker, J., Bednarek, W., K., & et al. (2008). Implementation of the random forest method for the imaging atmospheric cherenkov telescope magic. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, *588*(3), 424–432. https://doi.org/https://doi.org/10.1016/j.nima.2007.11.068

Aleksić, J., Ansoldi, S., Antonelli, L. A., Antoranz, P., Babic, A., Bangale, P., Barrio, J. A., Becerra González, J., Bednarek, W., Bernardini, E., & et al. (2014). Discovery of γ-ray emission from the pulsar wind nebula 3c 58 by magic. *Astronomy & Astrophysics*, *567*, L8. https://doi.org/10.1051/0004-6361/201424261

Aliu, E., Anderhub, H., Antonelli, L. A., Antoranz, P., Backes, M., Baixeras, C., Barrio, J. A., Bartko, H., Bastieri, D., & et al. (2008). Observation of pulsed -rays above 25 gev from the crab pulsar with magic. *Science*, *322*(5905), 1221–1224. https://doi.org/10.1126/science.1164718

Brill, A., Feng, Q., Humensky, T. B., Kim, B., Nieto, D., & Miener, T. (2019). Investigating a deep learning method to analyze images from multiple gamma-ray telescopes, In *2019 new york scientific data summit (nysds)*. https://doi.org/10.1109/NYSDS.2019.8909697

Caraveo, P. (2020). The golden age of high-energy gamma-ray astronomy: The cherenkov telescope array in the multimessenger era. *La Rivista del Nuovo Cimento*, *43*. https://doi.org/10.1007/s40766-020-00006-3

Celli, S., Morlino, G., Gabici, S., & Aharonian, F. A. (2019). Exploring particle escape in supernova remnants through gamma rays. *Monthly Notices of the Royal Astronomical Society*, *490*(3), 4317–4333. https://doi.org/10.1093/mnras/stz2897

Chernyakova, M., & Malyshev, D. (2020). Gamma-ray binaries.

de León, S. C., Alonso, A. C., Rosa-González, D., & Longinotti, A. L. (2019). Spectral analysis of the blazars markarian 421 and markarian 501 with the hawc gamma-ray observatory.

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, *12*, 2121–2159.

Fernández Barral, A. (2017). *Extreme particle acceleration in microquasars jets and pulsar wind nebulae with the MAGIC telescopes* (Doctoral dissertation). Barcelona, Autonoma U.

Fomin, V., Stepanian, A., Lamb, R., Lewis, D., Punch, M., & Weekes, T. (1994). New methods of atmospheric cherenkov imaging for gamma-ray astronomy. i. the false source method. *Astroparticle Physics*, *2*(2), 137–150. https://doi.org/https://doi.org/10.1016/0927-6505(94)90036-1

Grespan, P. (2020). *Convolutional neural network data analysis development for the large sized telescope of cta and broadband study of the blazar 1es 1959+650* (Master's thesis). Università di Padova.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition, In *2016 ieee conference on computer vision and pattern recognition (cvpr)*. https://doi.org/10.1109/CVPR.2016.90

HESS Collaboration. (2016). Acceleration of petaelectronvolt protons in the galactic centre. *Nature*, *531*(7595), 476–479. https://doi.org/10.1038/nature17147

Hillas, A. M. (1985). Cerenkov Light Images of EAS Produced by Primary Gamma Rays and by Nuclei, In *19th international cosmic ray conference (icrc19), volume 3.*

Hinton, J., & Ruiz-Velasco, E. (2020). Multi-messenger astronomy with very-high-energy gamma-ray observations. *Journal of Physics: Conference Series*, *1468*, 012096. https://doi.org/10.1088/1742-6596/1468/1/012096

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift (F. Bach & D. Blei, Eds.). In F. Bach & D. Blei (Eds.), *Proceedings of the 32nd international conference on machine learning*, Lille, France, PMLR. http://proceedings.mlr.press/v37/ioffe15.html

Karlsson, N. (2009). Discovery of vhe gamma-ray emission from the starburst galaxy m82.

Kim, B., Brill, A., Miener, T., Nieto, D., & Feng, Q. (2020). *DL1-Data-Handler: DL1 HDF5 writer, reader, and processor for IACT data* (Version v0.8.2). Zenodo. https://doi.org/10.5281/zenodo.3979698

Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization.

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791

Li, J., Torres, D. F., Liu, R.-Y., Kerr, M., de Oña Wilhelmi, E., & Su, Y. (2020). Gamma-ray heartbeat powered by the microquasar ss 433. *Nature Astronomy*, *4*(12), 1177–1184. https://doi.org/10.1038/s41550-020-1164-6

Longair, M. S. (2011). *High energy astrophysics* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511778346

López-Coto, R. (2017). *Very-high-energy gamma-ray observations of pulsar wind nebulae and cataclysmic variable stars with magic and development of trigger systems for iacts.* https://doi.org/10.1007/978-3-319-44751-3

Marinello, N. (2019). *Convolutional neural network single-telescope reconstruction for the large size telescope of cta* (Master's thesis). Università di Padova.

Mazin, D. (2019). The Cherenkov Telescope Array, In *Proceedings of 36th international cosmic ray conference — pos(icrc2019)*. https://doi.org/10.22323/1.358.0741

Middleton, M., Casella, P., Gandhi, P., Bozzo, E., Anderson, G., Degenaar, N., Donnarumma, I., Israel, G., Knigge, C., Lohfink, A., & et al. (2017). Paving the way to simultaneous multi-wavelength astronomy. *New Astronomy Reviews*, *79*. https://doi.org/10.1016/j.newar.2017.07.002

Millikan, R. A., & Cameron, G. H. (1926). High frequency rays of cosmic origin iii. measurements in snow-fed lakes at high altitudes. *Phys. Rev.*, *28*, 851–868. https://doi.org/10.1103/PhysRev.28.851

NASA, ESA, NRAO/AUI, Bietenholz, M., Uson, J., Cornwell, T., JPL-Caltech, Gehrz, R., Hester, J., Loll, A., Swift, Hoversten, E., PSU, NASA/CXC/SAO/F.Seward, & Buehler, N. L. (2005). Picture of crab nebula in multiple wavelengths.

Nieto, D., Brill, A., Feng, Q., Humensky, T. B., Kim, B., Miener, T., Mukherjee, R., & Sevilla, J. (2020). CTLearn: Deep Learning for Gamma-ray Astronomy. *PoS*, *ICRC2019* arXiv 1912.09877, 752. https://doi.org/10.22323/1.358.0752

Nieto, D., Brill, A., Feng, Q., Jacquemont, M., Kim, B., Miener, T., & Vuillaume, T. (2020). Studying deep convolutional neural networks with hexagonal lattices for imaging atmospheric Cherenkov telescope event reconstruction. https://doi.org/10.22323/1.358.0753

Olaizola, A. M. (2000). *Búsqueda de fuentes cósmicas de radiación gamma de muy alta energía con el detector airobicc* (Doctoral dissertation). Universidad Complutense de Madrid.

Schultz, C. (2013). *Development of New Composite Mirrors for Imaging Cherenkov Telescopes and Observations of the Two Blazar Objects 1ES 0806+524 and 1ES 1011+496 with MAGIC* (Doctoral dissertation). Padua U.

Shilon, I., Kraus, M., Bächele, M., Egberts, K., Fischer, T., Holch, T., Lohse, T., Schwanke, U., Steppa, C., & Funk, S. (2019). Application of deep learning methods to analysis of imaging atmospheric cherenkov telescopes data. *Astroparticle Physics*, *105*, 44–53. https://doi.org/https://doi.org/10.1016/j.astropartphys.2018.10.003

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research, 15*(56), 1929–1958. http://jmlr.org/papers/v15/srivastava14a.html

Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions, In *2015 ieee conference on computer vision and pattern recognition (cvpr)*. https://doi.org/10.1109/CVPR.2015.7298594

Tieleman, T., & Hinton. (2012). *G. Lecture 6.5 - RMSProp, COURSERA: Neural Networks for Machine Learning* (tech. rep.).

Wagner, R. (2007). Measurement of very high energy gamma-ray emission from four blazars using the MAGIC telescope and a comparative blazar study. *Publications of the Astronomical Society of the Pacific, 119*(860), 1201–1203. https://doi.org/10.1086/522380

Yao, Y., Rosasco, L., & Caponnetto, A. (2007). On early stopping in gradient descent learning. *Constructive Approximation, 26*(2), 289–315. https://doi.org/10.1007/s00365-006-0663-2