Friedrich-Alexander-Universität
Naturwissenschaftliche Fakultät

Bachelor Thesis
in Physics

Investigation of Multitask Learning with Graphformer Models for Event Reconstruction at the Southern Wide-field Gamma-ray Observatory (SWGO)

Leoni Kim Decker

Supervisor: Prof. Dr. Christopher van Eldik

Erlangen Centre for Astroparticle Physics

Submission date: 28-10-25

Abstract

The Southern Wide-field Gamma-ray Observatory (SWGO) is a planned next-generation ground-based experiment designed to detect high-energy γ-rays from astrophysical sources. Accurate reconstruction of Extensive Air Showers (EAS) parameters such as energy, arrival direction, core position, and particle type are crucial for identifying γ-ray events among the dominant cosmic-ray background.

This thesis investigates multitask learning as a method to improve event reconstruction performance at SWGO using a Graphformer model, which combines Graph Neural Network (GNN) EdgeConvolutions with Transformer self-attention layers. The primary goal of this work is to evaluate whether jointly training a single model on multiple related tasks enhance reconstruction quality compared to traditional singletask training and classical template-likelihood methods.

Using simulated data, several model configurations were evaluated, including singletask and multitask setups with two, three and four tasks. Their performance was assessed across all relevant metrics. The three-task multitask model, trained exclusively on gamma events to predict energy, core position, and shower axis, achieved the best overall performance. It outperformed the template-likelihood baseline in both energy and core reconstruction while performing on par with the single-task Graphformer. For angular resolution, the model surpassed the template method up to 10^5 GeV and showed improved results compared to the singletask configuration at higher energies. The four-task mixed-particle model, which additionally included gamma/hadron classification, performed slightly worse overall, primarily due to reduced gamma statistics and computational constraints during training.

Overall, this work demonstrates that multitask learning can effectively leverage shared correlation between reconstruction targets to enhance reconstruction accuracy. These findings provide valuable insights for the development of future machine-learning-based analysis pipelines for the SWGO experiment.

Zusammenfassung

Das Southern Wide-field Gamma-ray Observatory (SWGO) ist ein geplantes bodengebundenes Experiment der nächsten Generation, das der Detektion hochenergetischer Gammastrahlung aus astrophysikalischen Quellen dient. Eine präzise Rekonstruktion der Parameter ausgedehnter Luftschauer (Extensive Air Showers (EAS)), wie Energie, Einfallsrichtung, Kernposition und Teilchentyp sind entscheidend, um Gamma Ereignisse zuverlässig vom dominanten kosmischen Strahlungshintergrund zu unterscheiden.

Diese Arbeit untersucht Multitask-Learning als Methode zur Verbesserung der Ereignisrekonstruktion am SWGO mithilfe eines Graphformer-Modells, das Graph Neural Network (GNN)-EdgeConvolutions mit Transformer-Selbstaufmerksamkeitsmechanismen kombiniert. Das Hauptziel besteht darin, zu evaluieren, ob das gleichzeitige Training eines einzigen Modells auf mehreren miteinander verwandten Aufgaben die Rekonstruktionsqualität im Vergleich zu herkömmlichem Singletask-Training und der klassischen Template-Likelihood-Methode verbessert.

Anhand von Simulationsdaten wurden verschiedene Modellkonfigurationen getestet, darunter Single- und Multitask-Ansätze mit zwei, drei und vier Aufgaben. Ihre Leistung wurde über alle relevanten Metriken hinweg bewertet. Das Drei-Task-Multitask-Modell, das ausschließlich auf Gamma Ereignissen trainiert wurde, um Energie, Kernposition und Schauerachse vorherzusagen, erzielte die insgesamt besten Ergebnisse. Es übertraf die Template-Likelihood-Referenz sowohl in der Energie- als auch in der Kernrekonstruktion und erreichte eine vergleichbare Leistung zum Single-Task-Graphformer. Bei der Winkelauflösung übertraf das Modell die Template-Methode bis zu Energien von 10^5 GeV und zeigte auch gegenüber der Singletask-Konfiguration verbesserte Ergebnisse bei höheren Energien. Das Vier-Task-Modell mit gemischten Teilchen, das zusätzlich eine Gamma/Hadron-Klassifikation beinhaltete, schnitt insgesamt etwas schlechter ab, was hauptsächlich auf die geringere Gamma Statistik und begrenzte Rechenressourcen während des Trainings zurückzuführen ist.

Insgesamt zeigt diese Arbeit, dass Multitask-Learning gemeinsame Korrelationen zwischen Rekonstruktionszielen effektiv nutzen kann, um die Rekonstruktionsgenauigkeit zu verbessern. Diese Ergebnisse liefern wertvolle Erkenntnisse für die Entwicklung zukünftiger, auf maschinellem Lernen basierender Auswertungsstrategien für das SWGO-Experiment.

Contents

1 Introduction

High-energy γ-ray astronomy is a field in astroparticle physics that probes the most energetic and non-thermal processes in the universe. γ-rays, the most energetic form of electromagnetic radiation, provide direct insight into the acceleration mechanisms of cosmic particles and the environments in which they occur. Unlike charged cosmic rays, which are deflected by interstellar magnetic fields, γ-rays travel in straight lines and therefore preserve directional information about their sources. This property allows researchers to identify and study astrophysical accelerators such as supernova remnants, pulsar wind nebulae, active galactic nuclei and gamma-ray bursts.

Because high-energy γ-rays are absorbed in the atmosphere, they cannot be detected directly at ground level. Instead, we measure the cascades of secondary particles they produce when interacting with air molecules, known as Extensive Air Showers (EAS). Ground-based detector arrays record the secondary particles or the Cherenkov light they emit, enabling the reconstruction of the energy, arrival direction and primary particle type (Bose et al., 2022).

In recent years, major ground-based experiments such as HAWC (HAWC Collaboration, 2023) and LHAASO (LHAASO Collaboration, 2021), as well as Imaging Atmospheric Cherenkov Telescopes (IACTs) like H.E.S.S. (Hess Collaboration, 2006) and MAGIC (MAGIC Collaboration, 2008) have led to remarkable discoveries in the TeV to PeV range. The upcoming CTAO (CTAO Collaboration, 2023) is expected to further advance these achievements. The next step in this field will be provided by the Southern Wide-field Gamma-ray Observatory (SWGO) (SWGO Collaboration, 2025), a next-generation water-Cherenkov detector array currently under development. SWGO extends monitoring to the southern sky, complementing CTAO. Its wide field of view, high-altitude location and nearly 100% duty cycle make it ideal for surveys and studying transient high-energy sources.

Traditionally, air-shower events have been reconstructed using template-based likelihood methods (Parsons and Hinton, 2014; Joshi et al., 2019), which compare measured detector signals to precomputed distributions derived from Monte Carlo simulations. While effective, these approaches rely heavily on simplified assumptions about shower development and detector response. In recent years, however, Machine Learning (Jung, 2022), and in particular deep learning (Glombitza et al., 2025), has emerged as a powerful alternative for reconstruction. Neural networks can learn complex, nonlinear relationships from raw detector data without the limitations of manual feature engineering.

Among deep learning approaches, Graph Neural Networks (GNNs) (Hamilton, 2020) have emerged as especially powerful architectures for event reconstruction, promising significant uplifts in reconstruction accuracy and sensitivity.

In this thesis, we investigate the application of the Graphformer (Leitl, 2025) architecture that combines aspects of Graph Neural Networks with Transformers (Vaswani et al., 2023) for event reconstruction at SWGO. The primary objectives are reconstructing the key shower parameters, such as the primary particle's energy, arrival direction and core position. Additionally, we perform gamma/hadron separation, distinguishing

electromagnetic from hadronic showers. We focus on exploring multitask learning to further enhance reconstruction performance.

We compare the performance of these models to traditional template-likelihood methods using simulated air-shower data. The analysis includes both single-task and multitask learning scenarios, evaluating performance across a wide energy range. The findings of this work aim to demonstrate the potential of modern Machine Learning approaches to advance the reconstruction capabilities of next-generation gamma-ray observatories such as SWGO.

2 γ-rays and Extensive Air Showers

2.1 High-energy γ-rays

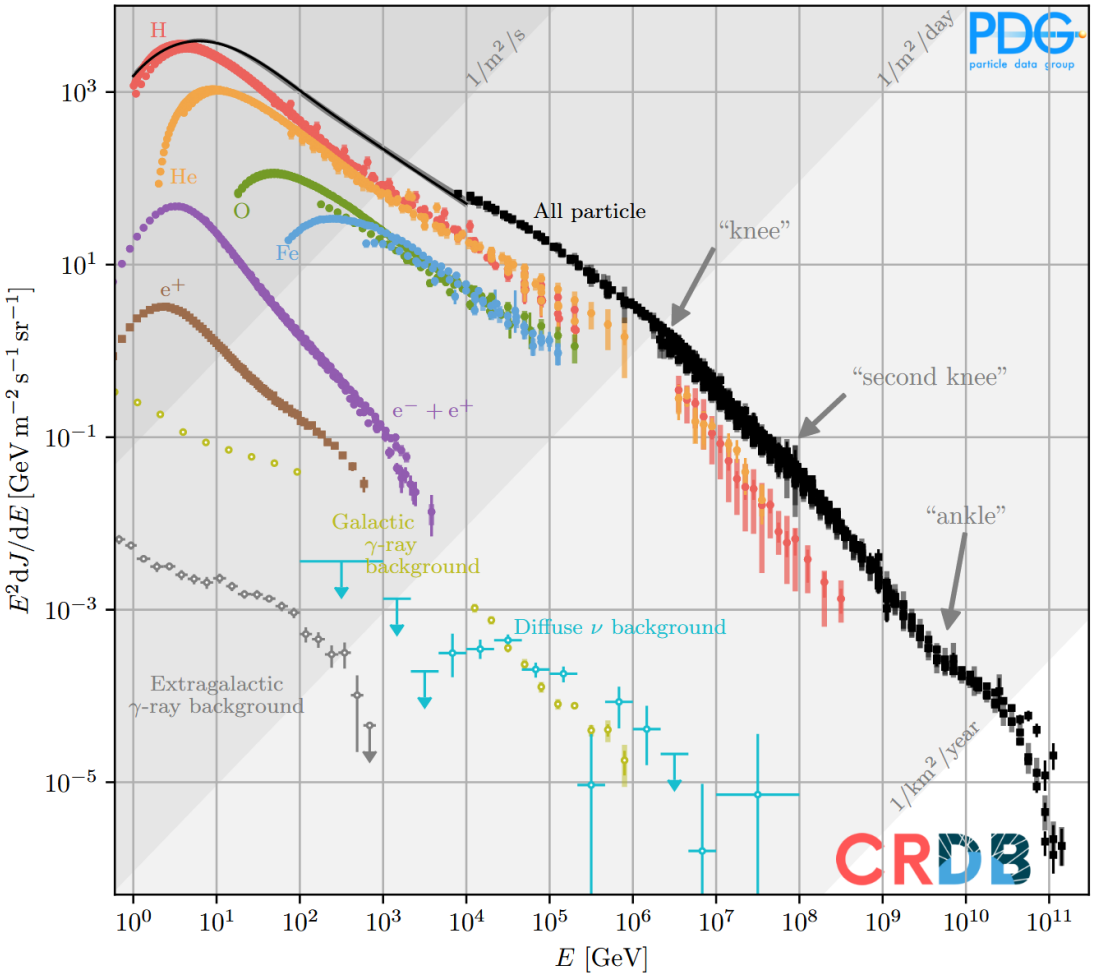


Figure 1: Spectrum of cosmic rays (CRs). The flux is multiplied by E^2 to highlight changes in the spectral index γ. The spectrum steepens at the "knee" ($\sim 3 \cdot 10^{15}$ eV), indicating a transition where Galactic accelerators reach their maximum energy or cosmic rays begin to escape more efficiently from the Galaxy. At the "ankle" ($\sim 4 \cdot 10^{18}$ eV), the spectrum flattens out, which signals a transition to extragalactic origin. Lastly, the spectrum sharply declines beyond the "cut-off" ($\sim 5.5 \cdot 10^{19}$ eV), reflecting energy losses from interactions with the cosmic microwave background (GZK effect). Adapted from Alvarez-Muñiz et al., 2024.

High-energy γ-rays, typically defined as photons with energies above $\sim 100\,\mathrm{MeV}$, are key messengers of non-thermal astrophysical sources where particles are accelerated to relativistic energies, such as:

- **Galactic:** Supernova remnants (SNRs), Pulsar Wind Nebulae (PWNs), binary

systems and the Galactic center are established γ-ray emitters. They are candidate sites for the acceleration of cosmic rays up to the "knee" ($\sim 10^{15}\,\mathrm{eV}$) (cf. Figure 1).

- **Extragalactic:** Active galactic nucleus (AGN), starburst galaxies and gamma-ray bursts (GRBs) are luminous in the GeV to TeV range and probe cosmic-ray acceleration and radiation fields in distant environments.

In both cases, γ-rays originate from either leptonic or hadronic processes. In leptonic scenarios, relativistic electrons up-scatter low-energy photons via inverse Compton scattering or emit bremsstrahlung radiation in ambient gas. In hadronic scenarios, accelerated protons and nuclei collide with gas or photons, producing neutral pions (π^0) that rapidly decay into two γ-rays. The resulting gamma-ray spectra and spatial distributions carry information about the underlying particle acceleration and environment.

Because γ-rays are photons, they travel in straight lines through space, unlike charged cosmic rays, which are deflected by magnetic fields. This means that, in principle, a detected γ-ray points directly back to its source. This property makes gamma-ray astronomy a powerful tool for identifying cosmic accelerators of high-energy particles.

2.2 Extensive Air Showers (EAS)

When such a high-energy γ-ray enters the Earth's atmosphere, it interacts with the molecules in the atmosphere and produces a cascade of secondary particles known as Extensive Air Showers (EASs).

This primary γ-ray initiates a purely electromagnetic cascade. The photon converts into an electron-positron pair, which in turn radiates bremsstrahlung photons that produce further pairs. This multiplication continues until the particle energies fall below the critical energy $E_c \approx 85\,\mathrm{MeV}$, where ionization dominates.

The development of air showers can be approximated by the Heitler model (Matthews, 2005). A primary photon of energy E_0 splits into two particles every radiation length X_0, doubling the number of particles until their energy falls below the critical energy $E_c \approx 85\,\mathrm{MeV}$. The maximum number of particles is

$$N_{\max} \approx \frac{E_0}{E_c},$$

reached after

$$n_{\max} \approx \frac{\ln(E_0/E_c)}{\ln 2}$$

generations, at a depth

$$X_{\max} \approx X_0\, n_{\max}.$$

If the incoming particle is a hadron, the first interaction typically produces mostly pions. Neutral pions (π^0) decay rapidly into γ-rays, feeding electromagnetic subshowers, while charged pions ($\pi^{\pm}$) either decay into muons and neutrinos or re-interact, sustaining the hadronic component. The result is a multi-component shower consisting of electromagnetic, hadronic and muonic parts.

An illustration of electromagnetic and hadronic shower development is shown in Figure 2.

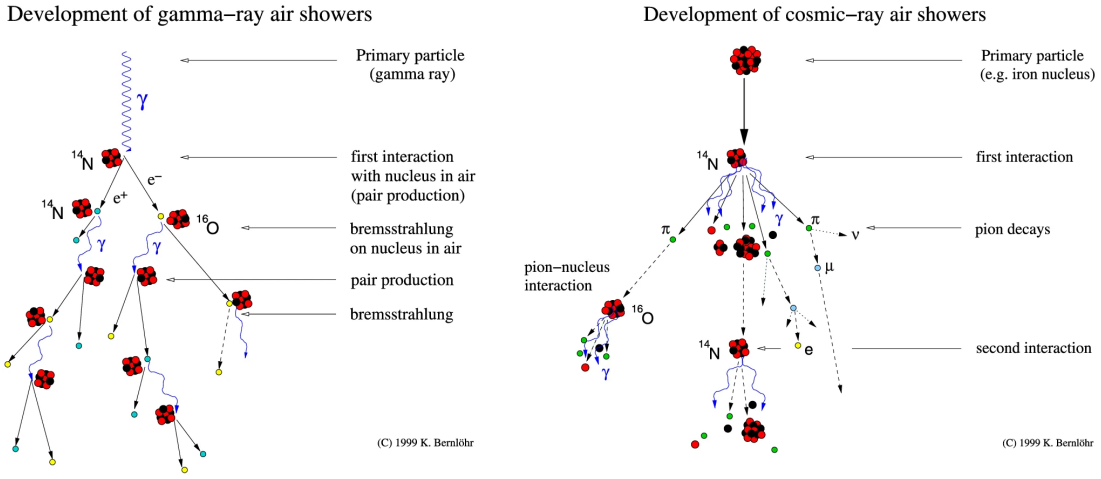


Figure 2: Schematic representation of an extensive air shower, showing the development of secondary particles in the atmosphere (Bose et al., 2022).

γ-ray-induced showers generally contain far fewer muons than hadronic showers, which is a key signature used to distinguish gamma events from cosmic-ray backgrounds. In practice, experiments exploit differences in shower morphology (core size, lateral spread, muon content) to separate gamma-ray events from the much more numerous cosmic-ray events.

2.3 Detection of γ-rays

High-energy γ-rays cannot be directly observed at ground level because they interact with the Earth's atmosphere and initiate EAS, as discussed above. Detection is therefore either direct, using instruments above the atmosphere, or indirect, by observing the air-shower products at ground level.

Direct measurements are usually performed by satellite detectors (e.g. Fermi-LAT (Thompson and Wilson-Hodge, 2023)). These instruments provide high precision but are limited to γ-rays below a few hundred GeV due to their restricted collection areas and exposure times. Higher energies require indirect ground-based detection.

Two main techniques are used on the ground:

- **Imaging Atmospheric Cherenkov Telescopes (IACTs):** These detect Cherenkov light emitted by relativistic charged particles in EAS. Arrays such as H.E.S.S. (Hess Collaboration, 2006), MAGIC (MAGIC Collaboration, 2008), VERITAS (VERITAS Collaboration, 2006), and the upcoming CTAO (CTAO Collaboration, 2023) achieve excellent angular resolution and sensitivity in the $\sim 100\,\mathrm{GeV} - 100\,\mathrm{TeV}$ range. Their limitations include narrow fields of view and operation restricted to clear, moonless nights.
- **Wide-field particle detector arrays:** These measure secondary particles that

reach the ground, often using water-Cherenkov detectors to record the Cherenkov light from particle passage. Examples include HAWC (HAWC Collaboration, 2023), LHAASO (LHAASO Collaboration, 2021), and the planned Southern Wide-field Gamma-ray Observatory (SWGO) (SWGO Collaboration, 2025). Such instruments operate with nearly 100% duty cycle, monitor large fractions of the sky, and are particularly effective for surveys and transient phenomena.

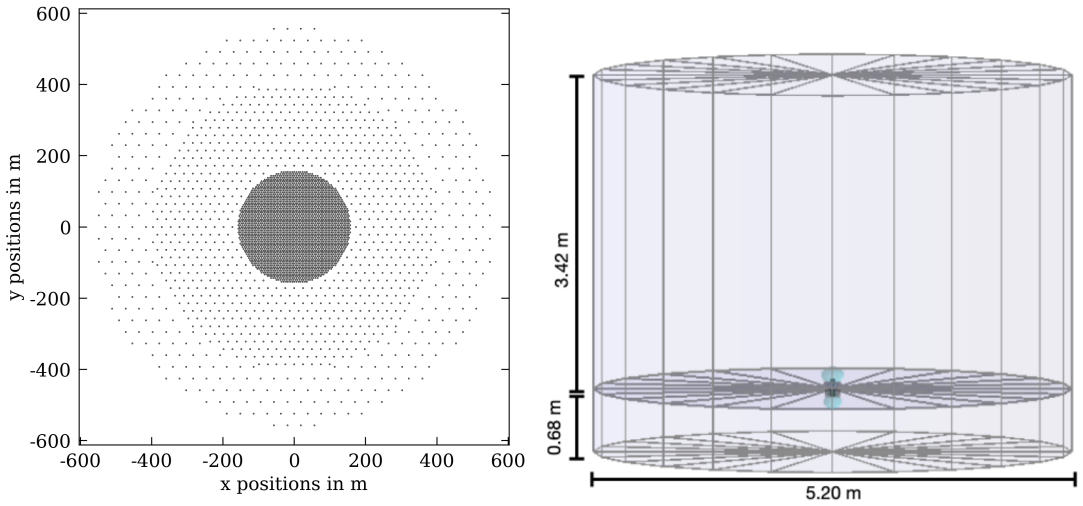
3 Southern Wide-field Gamma-ray Observatory (SWGO)

This section focuses on a specific configuration and its simulation. A detailed description of the science case and design options is found in SWGO Collaboration, 2025.

The Southern Wide-field Gamma-ray Observatory (SWGO) is a next-generation ground-based experiment currently in the design phase. It will be located at Pampa La Bola, at the Atacama Astronomical Park in northern Chile, at an altitude of 4,770 meters. Its primary goal is to provide continuous monitoring of the southern sky in the 100s of GeV up to PeV energy range, with emphasis on the Galactic Center region and transient sources, specifically gamma-ray bursts.

3.1 Layout and design

The SWGO baseline design, shown in Figure 3, is an array of cylindrical Water-Cherenkov Detectors (WCDs), each instrumented with two Photomultiplier Tubes (PMTs). The detector tanks have a radius of 5.2 m, a height of 4.1 m and two volumes inside filled with purified water. When relativistic charged particles from air showers pass through the tank, they produce Cherenkov light, which is detected by the PMTs. Each WCD is equipped with a dual-PMT unit positioned between the chambers comprising a downward-facing 8-10-inch-PMT and a 10-inch-PMT looking upwards.



(a) Detector layout **(b)** Tank design

Figure 3: Detector design of SWGO used in this work. (a) Detector layout featuring three zones with fill factors of 70% (central zone 156 m), 4% (intermediate zone 400 m) and 1.7% (outer zone 560 m). (b) Water-Cherenkov tank design with two PMTs.

The SWGO array is planned in concentric zones of different densities to cover a wide energy range. Zone 1, with a radius of 156 m, 2587 tanks and a fill factor of 70%, enables high sensitivity in the 100 GeV to multi-TeV energy range. This central zone has the highest density because air showers are most concentrated near their cores. Surrounding

this zone there are two outer zones with lower detector density, which cost-effectively extend the instrumented area to approximately $1\,\mathrm{km}^2$. These outer regions enhance sensitivity to the most energetic and spatially extended air showers. Zone 2 has a radius of 400 m with 792 tanks and a fill factor of 4%, while Zone 3 extends to a radius of 560 m with 384 tanks and a fill factor of 1.7%. In total, the configuration contains 3763 WCDs.

3.2 Event simulations and template methods

Since SWGO is still in the development phase, no real data are available yet. Instead, the experiment relies on detailed Monte Carlo (MC) simulations to study expected performance and to develop reconstruction methods. Air showers are generated using the software package CORSIKA (COsmic Ray SImulations for KAscade) (Heck et al., 1998). This tool simulates both hadronic and electromagnetic cascades from primary particles of different types, energy and arrival direction. A total of 10^9 events are generated over an energy range from 30 GeV to 1 PeV, assuming a spectral index of $\Gamma = -2$ and a maximum zenith angle of 65°. Each simulated shower is reused for five different core positions within the detector array.

The detector response is simulated using HAWCSim, a framework based on GEANT4 (Agostinelli et al., 2003), which was originally developed for the HAWC collaboration. This includes photon propagation in the water-Cherenkov tanks, Photomultiplier Tube (PMT) response, electronics digitization, and the injection of realistic noise, resulting in simulated signals that closely resemble those expected from the real instrument.

These detailed detector simulations form the basis for traditional reconstruction techniques. A widely used classical approach for reconstructing air-shower parameters is the template-likelihood reconstruction method (Parsons and Hinton, 2014; Joshi et al., 2019). This technique uses the MC simulations of both air showers and detector responses as a reference library and compares these with measured detector signals. Each simulated event defines a template that characterizes the expected lateral or image amplitude distribution as a function of key shower parameters such as the primary energy, impact point, and shower-maximum depth (X_{max}).

During reconstruction, an observed event is fitted to this library of templates using a maximum-likelihood approach that identifies the set of shower parameters that best reproduces the measured detector response. This likelihood formulation naturally incorporates detector fluctuations and noise, providing statistically optimal parameter estimation without relying on analytic approximations. For air-shower arrays such as SWGO, the template-likelihood method typically fits the Lateral Distribution Function (LDF) (Joshi et al., 2019) of photoelectron charges to reconstruct the shower core position and primary energy.

The principles and implementation of angular reconstruction for extensive air showers in the SWGO experiment are described in Robino (2025). The shower front arriving at ground level is slightly curved, and its curvature and time thickness depend on the distance from the shower core. Correcting for this curvature allows the front to be approximated as a plane for easier directional fitting. The `GaussPlaneFit` algorithm,

adapted from HAWC, performs a weighted regression of signal arrival times using inputs such as the estimated core position, a curvature model, and selected PMT signals. Each hit time is corrected for curvature and weighted according to its timing uncertainty to reduce the influence of noisy measurements. The fit iteratively refines the zenith and azimuth angles, rejecting outliers at each step, ultimately yielding the reconstructed arrival direction of the primary particle.

4 Introduction to Machine Learning and Deep Learning

4.1 Machine Learning

Machine Learning (ML) is a subfield of artificial intelligence focused on developing algorithms that improve their performance through experience rather than explicit programming (Jung, 2022). ML systems learn patterns and relationships directly from data, enabling them to perform tasks such as classification and regression.

A typical supervised learning setup consists of a dataset composed of input-output pairs $(\vec{x}, y)$, where the feature vector $\vec{x}$ represents measurable quantities (such as photomultiplier charges and arrival times) and y denotes the target label (e.g. particle type, direction, or energy). The goal of training is to find a function that best maps inputs to outputs by optimizing a set of parameters that minimize prediction error on unseen data.

The most widely used ML models for high-dimensional data are neural networks (NNs, Figure 4), which consist of interconnected layers of nodes (neurons). Each node performs a weighted sum of its inputs and applies a nonlinear activation function. These layers are organized into an input layer, one or more hidden layers, and an output layer. Networks with multiple hidden layers are referred to as deep neural networks (DNNs), forming the foundation of deep learning. The weights and biases of the network are adjusted during training through backpropagation to minimize a defined loss function, allowing the model to learn complex, nonlinear relationships between inputs and outputs.

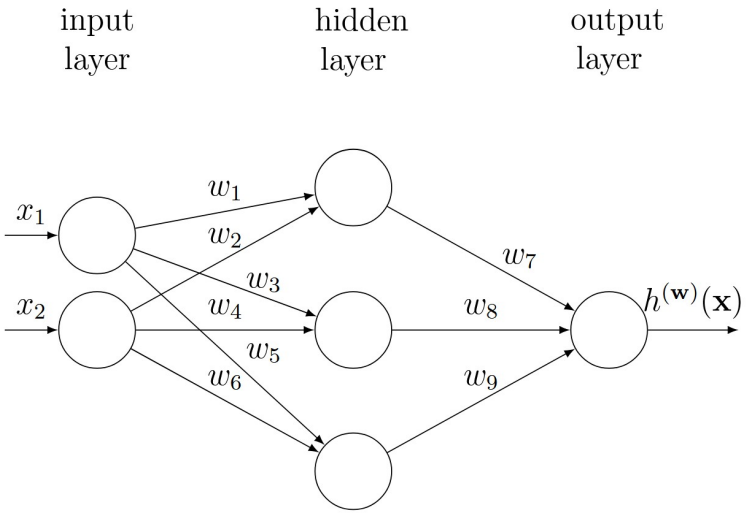


Figure 4: Simple neural network with input feature vector $\vec{x} = (x_1, x_2)^{\mathrm{T}}$, one hidden layer and one output node. Additionally, weights w_i are displayed, which represent the adaptive network parameters. Adapted from Jung (2022).

Machine learning tasks can generally be divided into regression and classification. In regression, the goal is to predict numerical values for given input, such as estimating the energy of a gamma-ray event. In classification, the task is to assign inputs to one of

several discrete categories, such as determining whether a detected air shower originates from a γ-ray or a cosmic ray proton. Both tasks rely on comparing model predictions to known outcomes during training in order to improve performance. To quantify this agreement, machine learning models use a loss function $\mathcal{L}(\hat{y}, y)$, which measures how well the model prediction $\hat{y}$ matches the true target y. The loss function effectively translates model performance into a numerical value that can be optimized. Typical choices depend on the task (Q. Wang et al., 2022; Elharrouss et al., 2025):

Regression tasks often use the **Mean Squared Error (MSE)**

$$\mathrm{MSE} = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 \,, \tag{1}$$

or the **Mean Absolute Error (MAE)**

$$\mathrm{MAE} = \frac{1}{N} \sum_{i=1}^{N} |\hat{y}_i - y_i| \,. \tag{2}$$

MSE squares the errors, which heavily penalizes large errors (outliers), while MAE averages the absolute errors, treating all errors equally.

Classification tasks often use **Binary Cross Entropy (BCE)**

$$\mathrm{BCE} = -\frac{1}{N} \sum_{i} \left[y_i \log \hat{p}_i + (1 - y_i) \log(1 - \hat{p}_i) \right] , \tag{3}$$

where $\hat{p}_i$ is the predicted probability for the correct class.

The learning process involves minimizing the chosen loss function with respect to the model parameters θ. This is achieved through gradient-based optimization, where the gradient of the loss with respect to each parameter indicates how the parameter should be adjusted to reduce error. A simple and widely used algorithm is Stochastic Gradient Descent (SGD), which updates parameters iteratively based on small random batches of data. More advanced optimizers, such as Adam or AdamW, adapt the learning rate and momentum dynamically to accelerate convergence and improve stability during training. Through repeated updates, the model learns to represent the underlying data distribution and generalize to unseen examples. This is then evaluated with a validation and test data set to ensure the model learns the underlying dependencies rather than overfitting statistical noise.

4.2 Deep Learning

Deep learning is a subfield of machine learning that focuses on neural networks with many layers capable of learning complex, hierarchical representations directly from data. Unlike traditional algorithms that rely on manually engineered features, deep learning models automatically extract the relevant information needed for prediction or classification through successive transformations of the input data. This makes them particularly powerful for analyzing large, high-dimensional, and unstructured datasets, as encountered in astrophysics, computer vision, and natural language processing.

Convolutional Neural Network (CNN)

One of the pioneering deep learning architectures is the Convolutional Neural Network (CNN). CNNs were designed to process data with a grid-like structure, such as images, where local spatial relationships are important. A landmark achievement in this area was AlexNet, introduced by Krizhevsky et al. (2012), which achieved a breakthrough in image recognition by dramatically reducing classification errors on the ImageNet dataset.

CNNs operate by applying a set of convolutional filters that detect local patterns such as edges, corners, or textures. Deeper layers combine these low-level features into increasingly abstract concepts, such as shapes or object parts. This hierarchical feature extraction allows CNNs to efficiently learn spatial hierarchies without requiring labels. The use of pooling layers further reduces the spatial resolution while preserving the most relevant information, enabling translation invariance and reducing computational cost. Although CNNs were first developed for image analysis, they have also been successfully adapted to time-series data, speech recognition, and even physics problems with regular detector layouts. However, CNNs rely on the assumption of a fixed, grid-like structure, which limits their ability to model data defined on irregular geometries or relational networks.

Graph Neural Networks (GNNs)

To overcome this limitation, Graph Neural Networks (GNNs) (Hamilton, 2020) were developed to handle data represented as graphs, where relationships between elements (nodes) are irregular and defined by connectivity rather than position. In this framework, each node represents an entity (for example, a detector station), and edges represent relationships between them, such as spatial proximity.

The key idea of GNNs is message passing. During this process, each node updates its representation by aggregating information from its neighboring nodes and edges. This allows the network to learn both local and global patterns within the graph, similar to how convolutional layers aggregate information in CNNs. Repeated message-passing steps expand the receptive field of each node, enabling it to incorporate information from increasingly distant parts of the graph. This makes GNNs especially suitable for physics applications where the geometry is irregular, such as particle detectors.

Transformers

A more recent architecture that has revolutionized deep learning is the Transformer, introduced by Vaswani et al. (2023). Unlike CNNs and GNNs, Transformers do not rely on local connectivity or fixed graph structures. Instead, they use a mechanism called self-attention, which allows each element in the input sequence to directly attend to all other elements and weigh their relevance dynamically.

5 Deep Learning-based Reconstruction for SWGO

This section follows the framework described in Glombitza et al. (2025). We employ simulated SWGO data and a Graph Neural Network (GNN) architecture for event reconstruction. Each air shower event is represented as a graph, where nodes correspond to triggered Photomultiplier Tubes (PMTs) and edges capture spatial relationships between detector stations. The network predicts key event parameters like energy, core position and arrival direction, as well as a gamma/hadron classification score.

5.1 Data preparation and preprocessing

To reconstruct air shower properties and perform γ/hadron separation, we use the spatial and temporal footprint of secondary particles detected at ground level. Rather than using full waveform traces, which are computationally expensive, we rely on the integrated PMT charge as a compact representation of the signal. This preserves the relevant physical information while reducing data dimensionality.

Each PMT contributes four input features:

$$(x, y, t, q)\,,$$

where (x, y) are the tank coordinates on the array, t is the arrival time of the first detected Cherenkov photon, and q is the integrated charge. Together, these form a point cloud describing the shower footprint in space and time.

Normalization

To stabilize and accelerate training, all features are normalized. Arrival times are standardized using z-score normalization:

$$t' = \frac{t - \mu_t}{\sigma_t}\,,$$

where μ_t and σ_t denote the mean and standard deviation of the arrival time distribution across all events. The charge values are log-scaled to reduce their large dynamic range:

$$q' = \frac{\log_{10}(1 + q)}{\sigma_q}\,,$$

with σ_q representing the standard deviation of the logarithmically scaled charge distribution. Spatial coordinates (x, y) are also z-score normalized. This ensures that all inputs contribute comparably during learning and prevents domination by any single feature scale.

Graph construction

From the point cloud representation, a spatial graph is constructed to encode local correlations between detector stations. This process is illustrated in Figure 5.

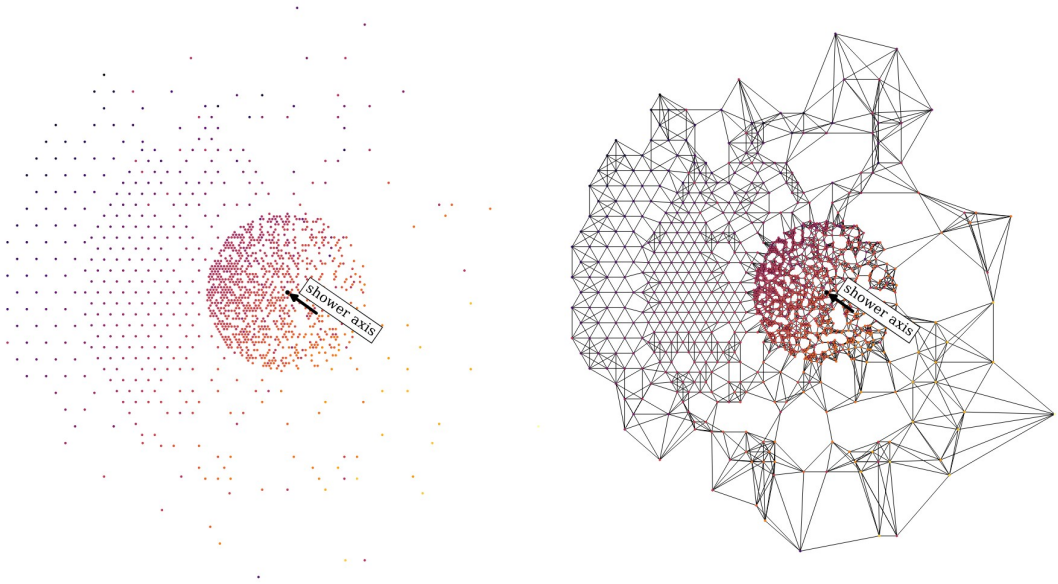


Figure 5: Construction of a spatial graph from an Extensive Air Shower event. The left panel shows the point cloud of detector tanks for a simulated γ-ray event, where lighter colors indicate earlier photon arrivals and darker colors indicate later ones. The right panel illustrates the resulting graph after applying kNN clustering, which connects each detector to its nearest neighbors to encode local spatial correlations. Adapted from Schneider et al., 2025.

A k-nearest neighbor (kNN) algorithm connects each node (station) to its six nearest neighbors in Euclidean space, which is consistent with the approximately hexagonal layout of the SWGO array. Self-connections are included, yielding a directed graph with N nodes and $7 \times N$ edges. Each node stores the feature vector (x, y, t, q), while the edges define the local geometric context used for message passing within the GNN. This structure enables the network to exploit both spatial and temporal correlations in the air-shower footprint.

5.2 Network architecture: The Graphformer Model

The current reconstruction framework in Figure 6, developed for the Southern Wide-field Gamma-ray Observatory (SWGO), is based on a hybrid Graph Neural Network (GNN)-Transformer architecture, referred to as the Graphformer model (Leitl, 2025). This design combines the geometric learning capabilities of GNNs with the global context modeling power of Transformer networks, enabling efficient processing of the spatially distributed and irregular detector data produced by extensive air showers.

After the input layer, the model employs three Edge Convolution (EdgeConv, Y. Wang et al., 2019) layers to extract local spatial and temporal correlations from the event graphs. Each node in the graph represents a triggered detector station, while the edges encode spatial relationships determined by a k-nearest-neighbor (kNN) graph construction. Within each EdgeConv layer, node features are iteratively updated based

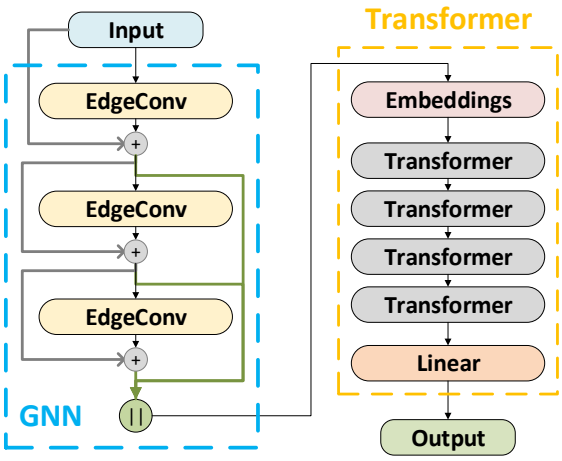


Figure 6: Illustration of the general network structure. This network is used for the γ/hadron separation and energy reconstruction in wide-field Water-Cherenkov-based Gamma-Ray Observatory. Figure is taken from Leitl (2025).

on their own attributes and those of their connected neighbors, allowing the network to learn localized patterns of signal charge and timing that characterize the structure of air showers. Stacking multiple EdgeConv layers enables the model to gradually expand its receptive field, incorporating higher-level contextual information across the array.

The features learned by the GNN are then passed to a Transformer module for global feature aggregation. Before entering the Transformer, node features are projected into a higher-dimensional embedding space, allowing the network to represent complex dependencies between detectors more effectively. The Transformer consists of four layers, each comprising Layer Normalization and Self-Attention blocks. The self-attention mechanism allows every node to dynamically weigh the importance of all other nodes in the event, capturing non-local correlations that cannot be modeled by message passing alone. This global reasoning is essential for tasks such as estimating the shower axis and discriminating between gamma and hadron events, where subtle global patterns can be decisive.

Finally, the output of the Transformer is fed into a fully connected linear layer, which maps the aggregated representation to the desired physical quantities, namely, the primary particle's energy, core position, arrival direction and γ/hadron classification score.

5.3 Training

Training was conducted on a single NVIDIA A100 GPU, with all models containing approximately one million trainable parameters.

Each dataset was divided into training, validation, and test subsets to ensure statistically independent evaluation and to prevent data leakage between optimization and performance assessment. The training set was used for gradient-based parameter updates,

while the validation set monitored generalization and triggered early stopping when no improvement was observed. Another test set, unseen during training, was reserved for the final evaluation of all reconstruction metrics, described in subsection 5.4.

All datasets were filtered to ensure the inclusion of shower cores within 580 m of the array center, zenith angles below 55°, and at least 64 Photomultiplier Tube (PMT) hits per event. For evaluation, slightly stricter cuts were applied, requiring a core distance $\leq 560\,\mathrm{m}$ and zenith angle $\leq 52°$.

Model optimization was performed using the AdamW optimizer, which combines the adaptive gradient updates of Adam with decoupled weight decay regularization to enhance generalization. A learning rate of $2.5 \cdot 10^{-4}$ and a weight decay of $1 \cdot 10^{-2}$ were used. Each model was trained for up to 80 epochs, with early stopping triggered if the validation loss did not improve for 10 consecutive epochs. The Mean Absolute Error (MAE, Equation 2) loss was employed for regression tasks such as energy, core position, and shower-axis reconstruction, while the Binary Cross-Entropy (BCE, Equation 3) loss was used for the gamma/hadron classification task.

5.4 Evaluation metrics

Each reconstruction task is evaluated using metrics specifically chosen to reflect its physical relevance. For energy reconstruction, we compute the logarithmic relative bias, defined as $\log_{10}(E_{\mathrm{reco}}) - \log_{10}(E_{\mathrm{true}})$, which quantifies systematic deviations between reconstructed and true energies. The energy resolution is expressed as the 68% containment width of the distribution $\sigma(\log_{10}(E_{\mathrm{reco}}/E_{\mathrm{true}}))$, representing the spread of reconstructed energies around the true value.

For core position reconstruction, we use the 68% containment radius of the distance distribution between the reconstructed and true shower core positions, providing a measure of spatial accuracy on the ground.

The shower axis (direction) is evaluated through the median angular error and its 68% containment radius. The angular error is defined as $\theta = \arccos(\hat{y} \cdot y^{\mathrm{mc}})$, where $\hat{y}$ is the reconstructed direction from the Graphformer model and y^{mc} is the Monte Carlo truth.

Finally, gamma/hadron separation performance is quantified using the background rejection, the fraction of correctly identified background (hadron) events relative to all background events and the Area Under the Receiver Operating Characteristic (AUROC) curve. The quantity $1-$ rejection is commonly referred to as the background contamination. High background rejection and AUROC values close to one indicate excellent discrimination between gamma and hadron events.

5.5 Baselines: Graphformer singletasking vs. template method

This section compares the performance of the Graph Neural Network (GNN) singletask Graphformer models (depicted in subsection 5.2) to the classical template-likelihood reconstruction (described in subsection 3.2). Two types of simulated datasets are used for evaluation:

(i) a **gamma-only** dataset containing 7,980 simulation files with a combined total of $\approx 5.2 \cdot 10^5$ events, and

(ii) a **mixed dataset** consisting of 6,000 gamma and 1,000 proton files ("6000γ/1000p") with a combined total of $\approx 6.3 \cdot 10^5$ events.

Figures 7 to 13 illustrate the comparison between the singletask Graphformer model and the template baseline for both datasets.

All datasets were filtered to ensure the inclusion of shower cores within 580 m of the array center, zenith angles below 55° and at least 64 Photomultiplier Tube (PMT) hits per event. For evaluation, stricter cuts were applied: core distance ≤ 560 m and zenith angle $\leq 52°$.

The singletask GNNs were trained separately for energy, core position, and shower-axis prediction using the Mean Absolute Error (MAE), defined in Equation 2, as the loss function. For particle classification, the Binary Cross Entropy (BCE) loss, given in Equation 3, was employed. Each training configuration was executed three times to account for stochastic variations in initialization and optimization, and the model achieving the lowest validation loss was selected for evaluation. All loss plots display the evolution of the training and validation loss up to the epoch corresponding to the minimum validation loss. Subsequent epochs are omitted for clarity.

All metrics, described in subsection 5.4, are evaluated in the energy range 100 GeV to 1 PeV in logarithmically spaced energy bins and averaged over zenith-angle ranges to reveal dependencies on energy and geometry. Error bars indicate statistical uncertainties estimated via bootstrap resampling.

For each dataset, 4,000 held-out test files were used exclusively for evaluation. The γ/hadron separation task used 2,000 gamma and 2,000 hadron test files.

5.5.1 Gamma reconstruction

At first, the energy, core and shower axis reconstruction for the gamma-only dataset using the singletask Graphformer model are evaluated by comparing them to the template-likelihood method (see Figure 7, Figure 8 and Figure 9).

Energy reconstruction

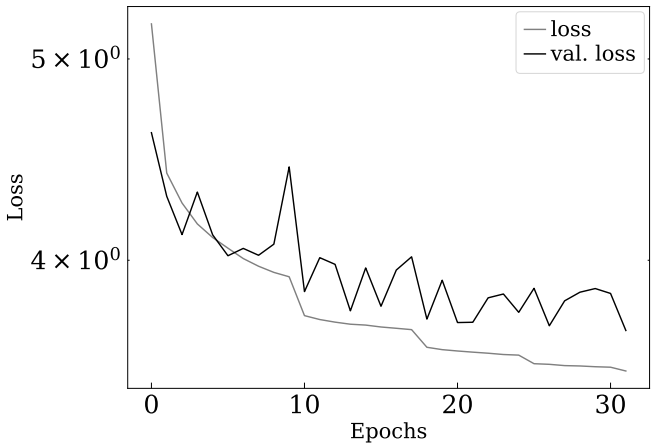
The energy reconstruction is quantified by the energy bias and energy resolution, shown in Figure 7.

Figure 7a shows the evolution of the training and validation loss during the optimization process. The final model corresponds to epoch 31, which achieved the lowest validation loss of 3.7. The noticeable gap between training and validation losses indicates limited generalization to unseen data, suggesting that the network exhibits a degree of overfitting.

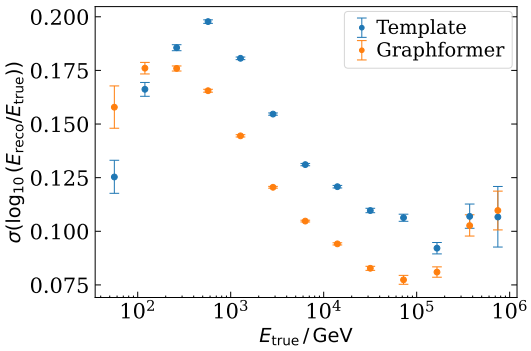
The Graphformer model (GNN), trained using the MAE loss, achieves improved energy resolution across most energy bins relative to the template method (Figure 7b). Smaller

resolution values indicate higher precision. The reduced performance in the lowest and highest energy bins is primarily due to limited statistics, which challenges the model's generalization.

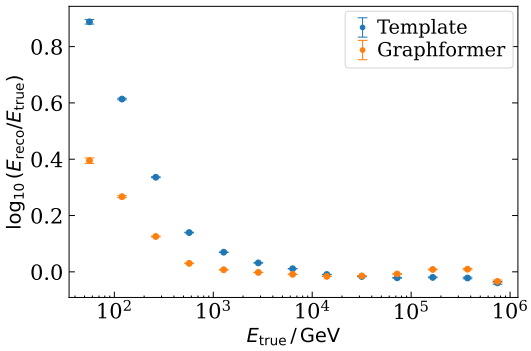
The energy bias (Figure 7c) ideally approaches zero, indicating unbiased reconstruction across the full energy range. At low primary energies, only a few secondary particles reach the detector, resulting in incomplete sampling of the air-shower footprint. In combination with the applied selection cut of at least 64 triggered PMTs per event, this leads to strongly reduced event statistics at low energies. Consequently, the limited data and weaker signals cause larger reconstruction uncertainties and a higher energy bias in this regime. The Graphformer model achieves lower bias than the template method up to $\approx 10^4\,\mathrm{GeV}$, beyond which the template approach performs slightly better. These results demonstrate the Graphformer's ability to learn complex correlations between signal charge, timing, and the spatial configuration of triggered detector tanks, leading to more accurate energy estimation in the core energy range.



(a) Training (gray) and validation (black) loss evolution for energy reconstruction. The final model corresponds to epoch 31, which achieved the lowest validation loss of 3.7.



(b) Energy resolution as a function of true energy.



(c) Energy bias as a function of true energy.

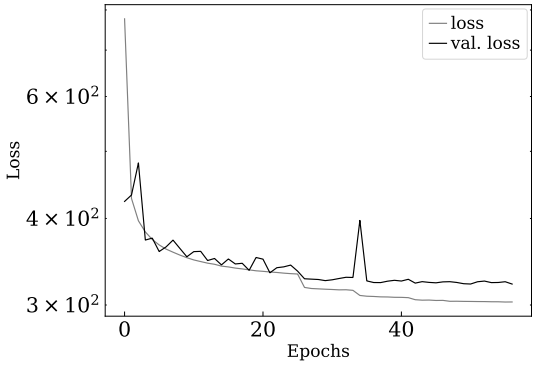
Figure 7: Performance of the single-task Graphformer model (orange), trained on gamma simulation files only, compared to the template-likelihood baseline (blue) for energy reconstruction. (a) Loss evolution during training. (b) Energy resolution. (c) Relative energy bias.

Core reconstruction

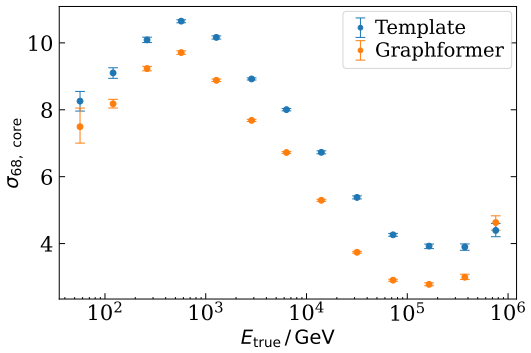
The performance of the core position reconstruction is presented in Figure 8.

The training and validation loss evolution of the singletask trained Graphformer is shown in Figure 8a. The final model corresponds to epoch 56, which achieved the lowest validation loss of $3.2 \cdot 10^2$.

The Graphformer model achieves superior core position reconstruction across all energy bins compared to the template method (Figure 8b), indicating that the learned graph structure effectively captures spatial correlations between neighboring tanks, which are crucial for accurate core localization.



(a) Training (gray) and validation (black) loss evolution for core reconstruction. The final model corresponds to epoch 56, which achieved the lowest validation loss of $3.2 \cdot 10^2$.



(b) Core position resolution (68 % containment radius) as a function of true energy.

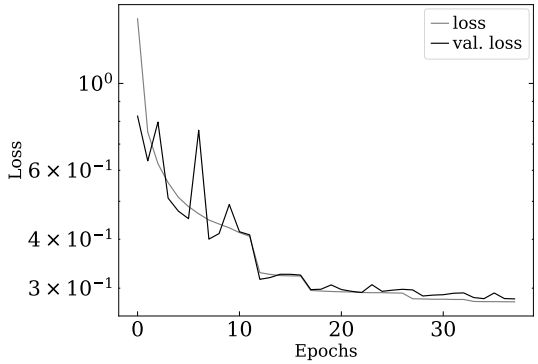
Figure 8: Performance of the single-task Graphformer model (orange), trained on gamma simulation files only, compared to the template baseline (blue) for core position resolution reconstruction. (a) Loss evolution during training. (b) Core position resolution.

Shower axis reconstruction

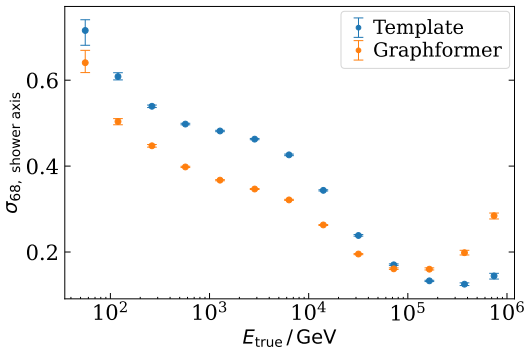
Finally, the shower axis reconstruction is performed in Figure 9.

Figure 9a shows the evolution of the training and validation loss during the optimization process. The final model corresponds to epoch 37, which achieved the lowest validation loss of 0.28.

The GNN outperforms the template method in angular reconstruction (see Figure 9b) at low to mid energies. At high energies, the template method achieves slightly better angular accuracy, which is important for source localization and point-source sensitivity.



(a) Training (gray) and validation (black) loss evolution for shower axis reconstruction. The final model corresponds to epoch 37, which achieved the lowest validation loss of 0.28.



(b) Shower axis resolution (68% containment) as a function of true energy.

Figure 9: Performance of single-task Graphformer (orange), trained on gamma simulation files only, compared to the template-likelihood baseline (blue) for shower axis reconstruction. (a) Loss evolution during training. (b) Shower axis resolution.

5.5.2 Gamma/hadron reconstruction and separation

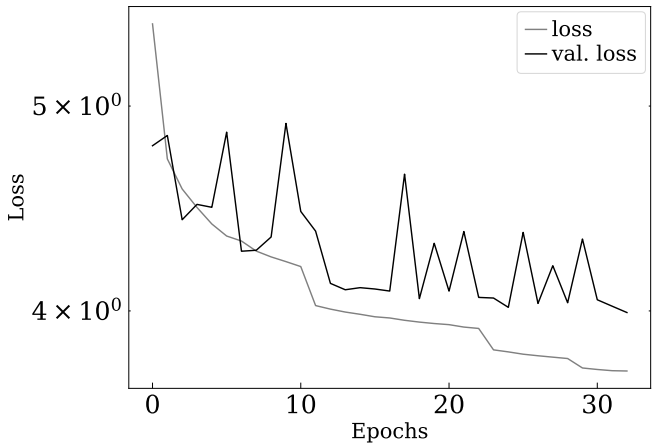
In this part, the mixed 6000γ/1000p dataset is individually trained four times with the tasks of energy, core position, shower axis reconstruction and γ/hadron separation being evaluated. Datasets of equal size were not used, as the focus is on γ-ray reconstruction while also assessing if combining these tasks in multitasking yields an improvement. Evaluations are performed separately for gamma and proton events and for gamma/hadron separation, both datasets are evaluated.

Energy reconstruction

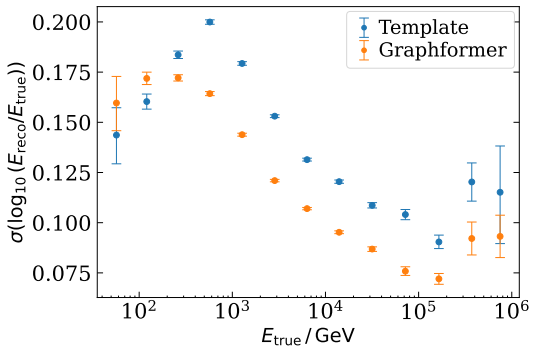
Energy reconstruction for the mixed dataset is presented in Figure 10, with the training loss shown in Figure 10a. The final model corresponds to epoch 32, which achieved the lowest validation loss of 4.0.

The Graphformer outperforms the template across the full energy range in both gamma and proton events (Figure 10b, Figure 10d). As expected, proton reconstruction indicates larger resolution values due to the smaller sample size.

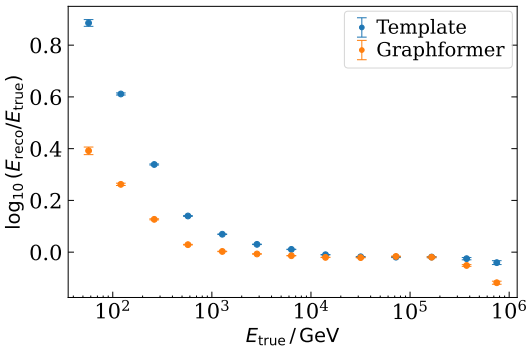
The energy bias for gamma events is comparable to the template method and slightly better at very low and high energies (Figure 10c), while for proton events it is larger than the template (Figure 10e).



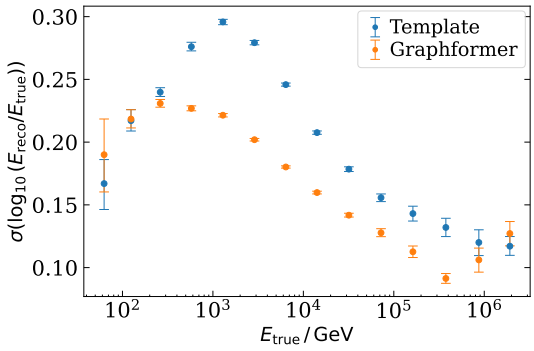
(a) Training (gray) and validation (black) loss evolution for energy reconstruction. The final model corresponds to epoch 32, which achieved the lowest validation loss of 4.0.



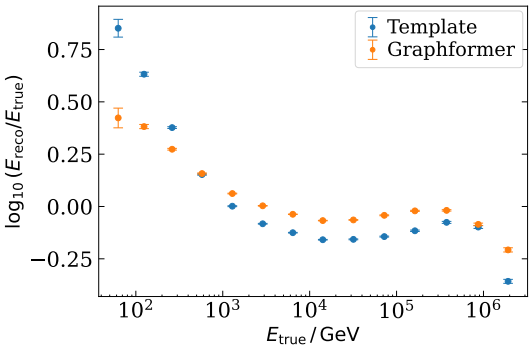
(b) Energy resolution for gamma events as a function of true energy.



(c) Energy bias for gamma events as a function of true energy.



(d) Energy resolution for proton events as a function of true energy.

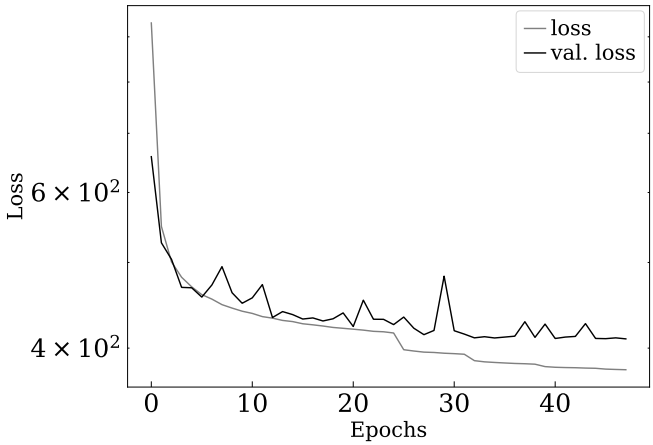


(e) Energy bias for proton events as a function of true energy.

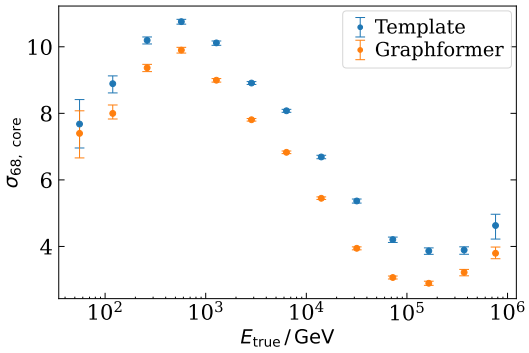
Figure 10: Energy reconstruction performance of the singletask Graphformer (orange), trained on the mixed 6000γ/1000p dataset, and compared to the template-likelihood (blue) method.

Core reconstruction

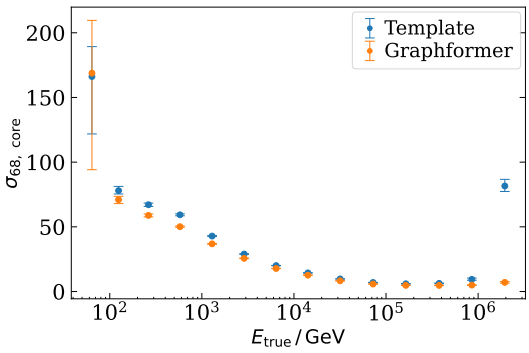
The core reconstruction for both gamma and proton events is illustrated in Figure 11, with the corresponding training loss evolution displayed in Figure 11a. The final model corresponds to epoch 47, which achieved the lowest validation loss of $4.1 \cdot 10^2$. As illustrated in Figure 11b and Figure 11c, the Graphformer consistently outperforms the template-based reconstruction across the entire energy range. The core resolution obtained for γ-ray events is comparable to that of the gamma-only dataset, indicating stable performance. However, the model performs worse for hadronic showers due to the smaller sample size.



(a) Training (gray) and validation (black) loss evolution for core reconstruction. The final model corresponds to epoch 47, which achieved the lowest validation loss of $4.1 \cdot 10^2$.



(b) Core resolution for gamma events.

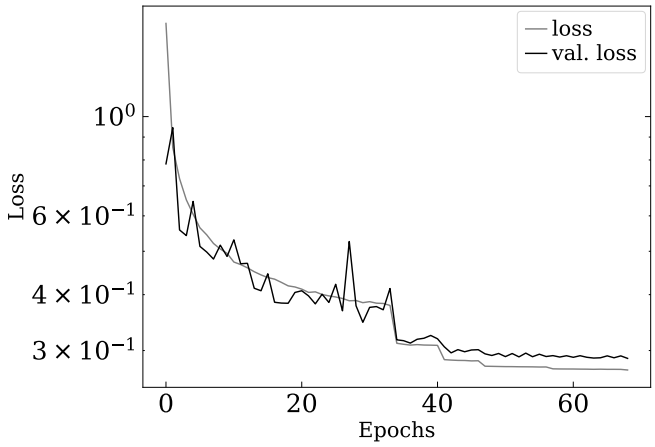


(c) Core resolution for proton events.

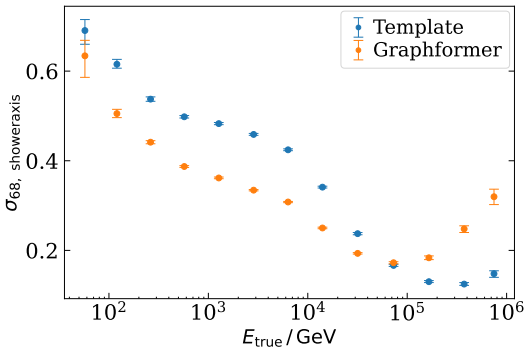
Figure 11: Core reconstruction performance of the singletask Graphformer (orange), trained on the mixed 6000γ/1000p dataset, and compared to the template-likelihood (blue) method.

Shower axis reconstruction

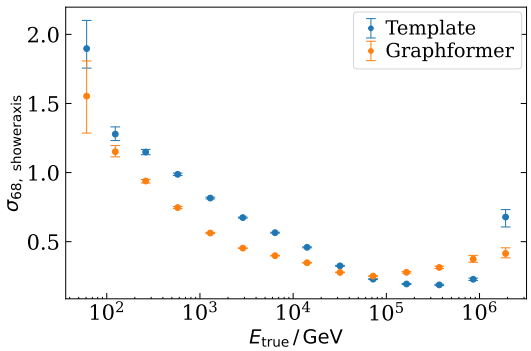
The angular reconstruction results are presented in Figure 12, with the corresponding training loss evolution displayed in Figure 12a. The final model corresponds to epoch 68, which achieved the lowest validation loss of 0.29. Among all reconstruction tasks, the shower-axis prediction shows the most pronounced learning dynamics during training. The Graphformer outperforms the template-based method for gamma-ray events (Figure 12b) at low to intermediate energies (up to approximately $6 \cdot 10^4$ GeV), but exhibits larger angular errors at higher energies. A similar trend is observed for hadronic events (Figure 12c), where the overall angular errors are slightly larger than for γ-ray-induced showers.



(a) Training (gray) and validation (black) loss evolution for shower axis reconstruction. The final model corresponds to epoch 68, which achieved the lowest validation loss of 0.29.



(b) Shower axis resolution for gamma events.

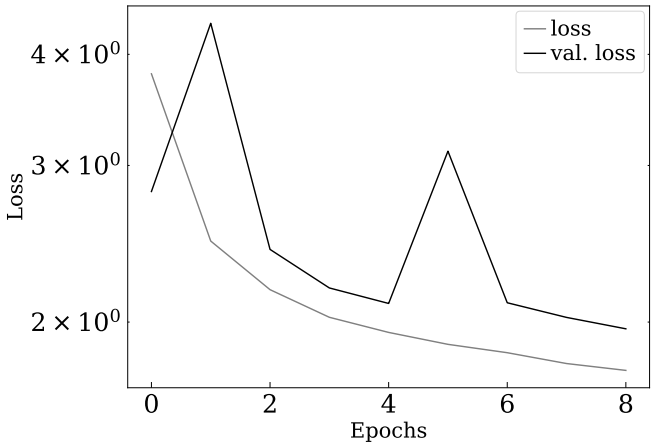


(c) Shower axis resolution for proton events.

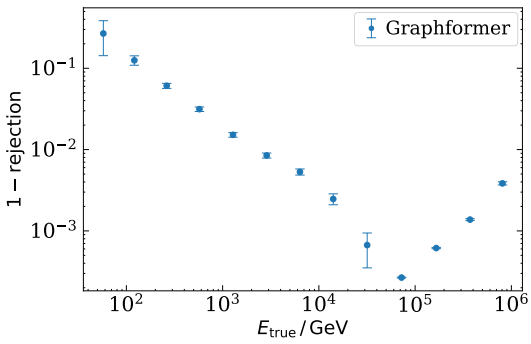
Figure 12: Shower axis reconstruction performance of the singletask Graphformer (orange), trained on the mixed 6000γ/1000p dataset, and compared to the template-likelihood (blue) method.

Gamma/hadron separation

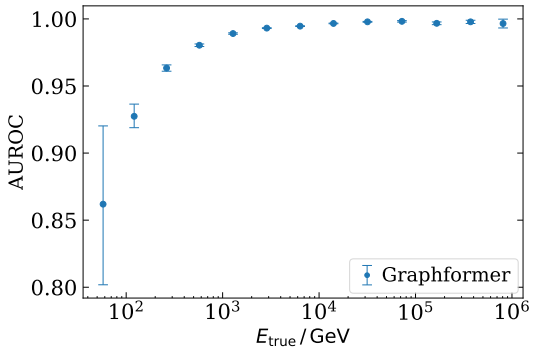
Finally, γ/hadron separation is evaluated in Figure 13. The loss evolution (Figure 13a) exhibits the final model corresponding to epoch 8, which achieved the lowest validation loss of 2.0. Performance metrics include background contamination (Figure 13b) and Area Under the Receiver Operating Characteristic curve (AUROC, Figure 13c), with no classical template available for comparison. The model achieves a background rejection of about 70% at low energies and then exponentially increases to around 99.3% at $2 \cdot 10^4$ GeV, above this energy no background is left. The AUROC curve exhibits values of close to one for energies above 10^3 GeV.



(a) Training (gray) and validation (black) loss evolution for γ/hadron separation. The final model corresponds to epoch 8, which achieved the lowest validation loss of 2.0.



(b) Background contamination as a function of true energy.



(c) AUROC curve for γ/hadron separation. Values close to 1 indicate excellent classification.

Figure 13: γ/hadron separation performance of the singletask Graphformer trained on the mixed 6000γ/1000p dataset.

Summary and comparison

In summary, the gamma event reconstruction results show that, despite the loss operating on different numerical scales across tasks, the Graphformer model consistently outperforms the template-based method in most cases, with the exception of the high-energy regime, where the template achieves better directional resolution. As expected, hadron event reconstruction indicates generally larger resolution values due to the smaller sample size.

6 Multitasking with Graphformer models for SWGO

A key advantage of deep learning is the to perform multitask learning, where a single model is trained to predict multiple related tasks simultaneously (Crawshaw, 2020). In this work, a single GNN Graphformer model is trained to reconstruct the energy, arrival direction, core position, and particle type of each event. This approach can improve overall performance because the tasks share underlying physical correlations. For instance, the spatial footprint that determines the shower core position also provides information about the primary energy and particle type.

In multitask learning, it is essential to properly balance the individual task losses since they typically differ in magnitude. This prevents any single task from dominating the optimization process. The total loss is therefore expressed as a weighted sum of task-specific losses:

$$\mathcal{L}_{\text{total}} = \sum_t w_t \mathcal{L}_t \,,$$

where the weights w_t may be fixed or adaptively learned, for example, using uncertainty-based weighting (Kendall et al., 2018).

6.1 Gamma reconstruction

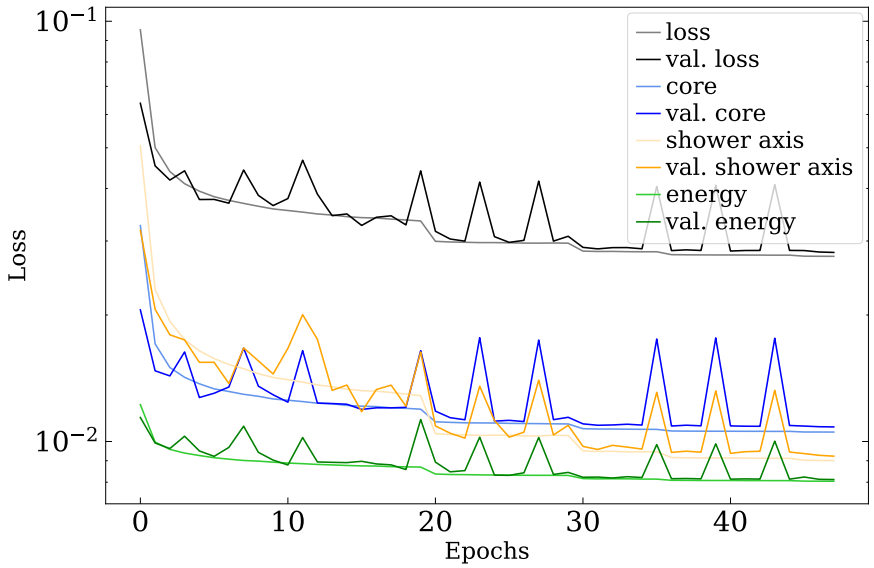
This section presents multitask training results using simulated γ-ray events. A single Graphformer network was configured with a joint output head for energy, core position, and shower-axis direction. A static custom loss weighting was applied to balance learning across tasks:

$$w_{\text{energy}} = 7 \cdot 10^{-3} \,, \quad w_{\text{core}} = 10^{-4} \,, \quad w_{\text{shower axis}} = 10^{-1} \,. \tag{4}$$

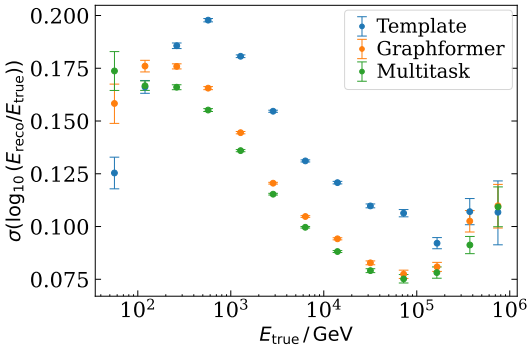
These weights were determined empirically through trial and error, prioritizing the shower-axis reconstruction as the most important task, followed by core and energy.

Figure 14a shows the evolution of the training and validation losses for the three individual tasks and the combined model. The final model corresponds to epoch 47, which achieved the lowest validation loss of $2.8 \cdot 10^{-2}$.

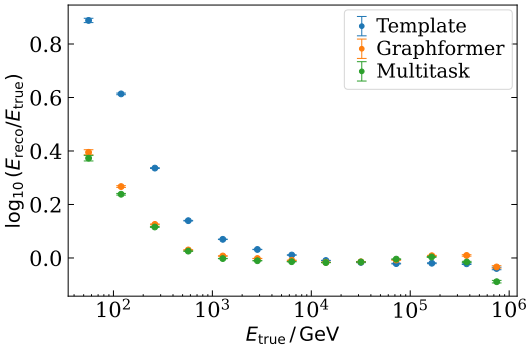
Compared to the single-task Graphformer, the multitask model achieves improved energy resolution (Figure 14b) and reduced energy bias (Figure 14c). The shower-axis resolution (Figure 14e) is comparable to the single-task result at low and mid energies but surpasses it above 10^5 GeV, remaining slightly less precise than the template method at the highest energies. The core resolution (Figure 14d) exceeds the template performance across the entire energy range, although the single-task Graphformer performs marginally better above 10^4 GeV.



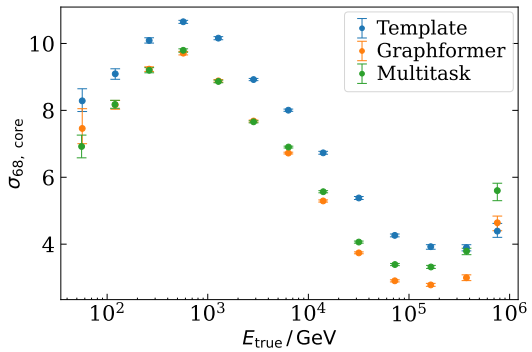
(a) Training and validation loss evolution of energy (green), core (blue), shower axis (yellow) and the combined model (black). The final model corresponds to epoch 47, which achieved the lowest validation loss of $2.8 \cdot 10^{-2}$.



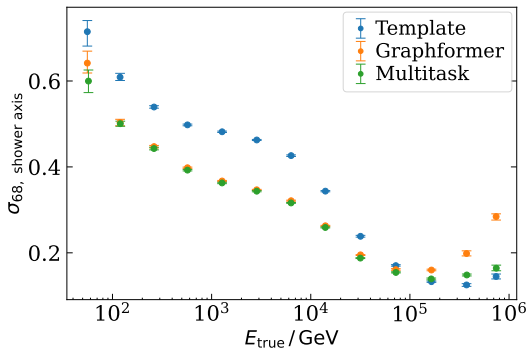
(b) Energy resolution as a function of true energy.



(c) Energy bias as a function of true energy.



(d) Core position resolution (68% containment radius) as a function of true energy.



(e) Shower axis resolution (68% containment) as a function of true energy.

Figure 14: Gamma reconstruction for three-task multitasking (green) against template method (blue) and singletask Graphformer model (orange).

Performance of two-task reconstruction models

To investigate whether the Graphformer model trained on two tasks simultaneously yields improved performance compared to the single-task and three-task configurations, we trained and evaluated models combining pairs of reconstruction targets: core and shower axis, energy and shower axis, and energy and core. All experiments employed the same custom loss weighting defined in Equation 4.

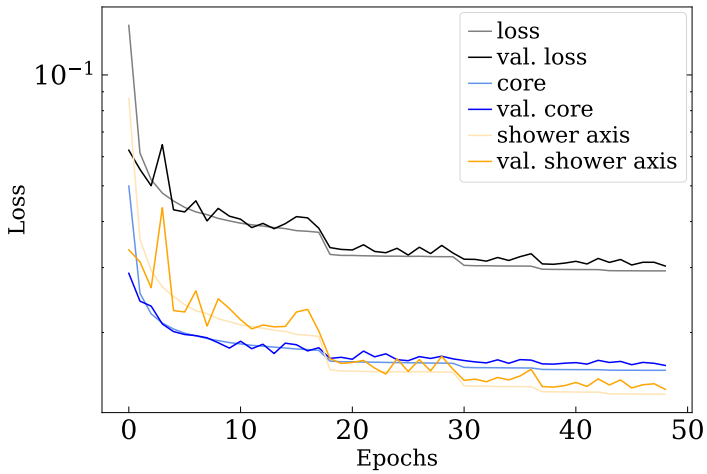
Core and shower axis

The results for the combined core and shower-axis reconstruction are shown in Figure 15. Figure 15a displays the evolution of the training and validation losses for both tasks and the overall combined loss. The final model corresponds to epoch 48, which achieved the lowest validation loss of $3.0 \cdot 10^{-2}$.

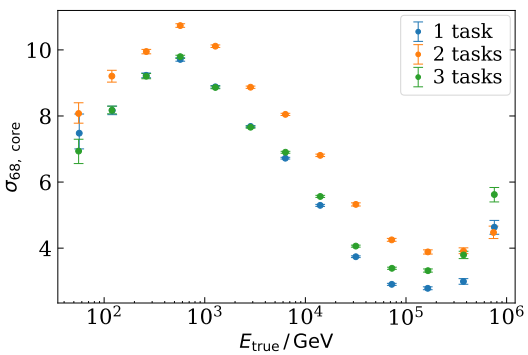
The combination of these two tasks yields no improvement for core reconstruction (Figure 15b) and as mentioned before, the singletask model is still slightly better at high energies.

As shown in Figure 15b, combining core and shower axis reconstruction does not improve the core position performance; the single-task Graphformer remains slightly better at higher energies. The direction resolution in Figure 15c shows a small improvement above 10^4 GeV compared to the three-task setup but performs worse at intermediate and lower energies.

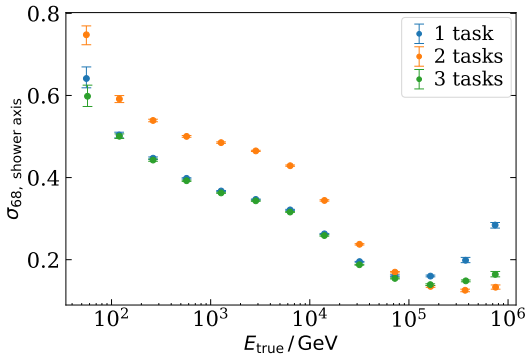
In summary, the combination of core and shower-axis reconstruction provides a modest benefit at high energies for angular resolution but offers no significant improvement for core reconstruction compared to the single- or three-task models.



(a) The evolution of the training and validation losses for core (blue), shower axis (yellow) and the combined model (black). The final model corresponds to epoch 48, which achieved the lowest validation loss of $3.0 \cdot 10^{-2}$.



(b) Core position resolution (68% containment radius) as a function of true energy.



(c) Shower axis resolution (68% containment) as a function of true energy.

Figure 15: Comparison of the reconstruction resolution for the two-task model trained on core and shower-axis prediction (orange) against the three-task multitask model (green) and the singletask Graphformer baseline (blue).

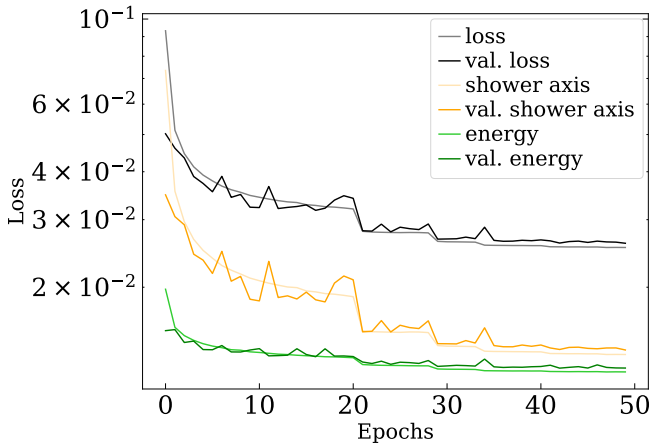
Energy and shower axis

The results for the combined energy and shower-axis reconstruction are presented in Figure 16. Figure 16a shows the evolution of the training and validation losses for both tasks and the joint model. The final model corresponds to epoch 49, which achieved the lowest validation loss of $2.6 \cdot 10^{-2}$.

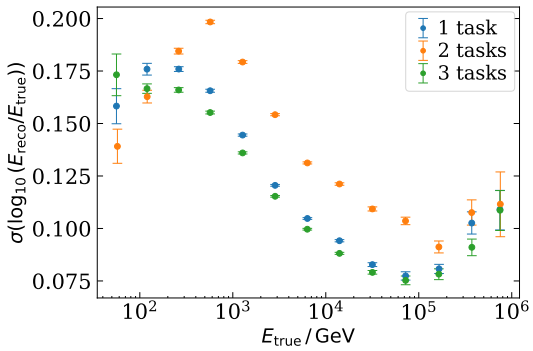
As illustrated in Figure 16b, the energy resolution is significantly worse than in both the single-task and the three-task models, particularly at intermediate energies. Similarly, the energy bias (Figure 16c) deteriorates for low-energy events, indicating that the combined learning objective hampers the network's ability to accurately estimate the shower energy.

The shower-axis resolution (Figure 16d) yields the same improvement visible only at the highest energies compared to that of the core and shower axis combination in Figure 15c.

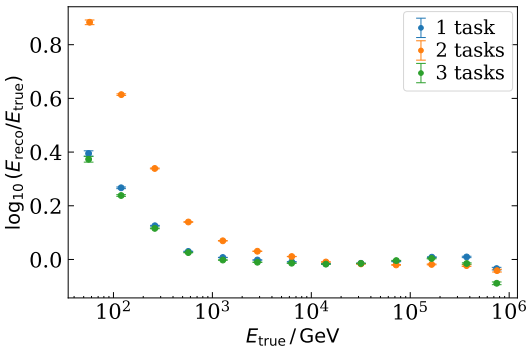
In conclusion, combining energy and shower-axis reconstruction does not yield meaningful gains. The inclusion of the energy task negatively impacts overall energy accuracy while offering only marginal improvements in angular resolution at high energies.



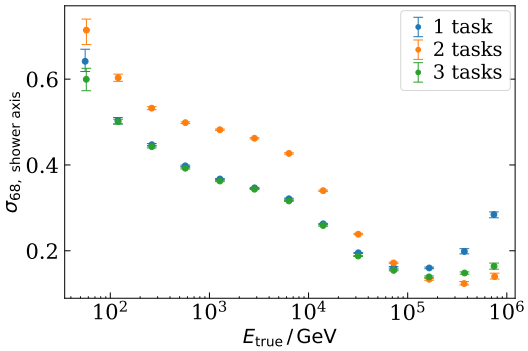
(a) The evolution of the training and validation losses for energy (green), shower axis (yellow) and the combined model (black).



(b) Energy resolution as a function of true energy.



(c) Energy bias as a function of true energy.



(d) Shower axis resolution (68% containment) as a function of true energy.

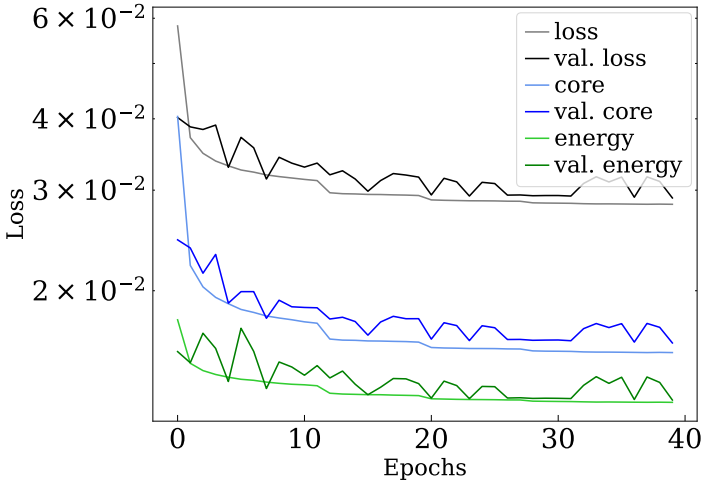
Figure 16: Comparison of the reconstruction resolution for the two-task model trained on energy and shower-axis prediction (orange) against the three-task multitask model (green) and the singletask Graphformer baseline (blue).

Energy and core

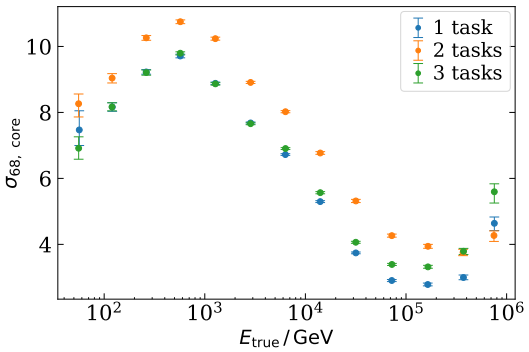
The results for the combined energy and core reconstruction are shown in Figure 17. Figure 17a displays the training and validation loss evolution for both tasks and the overall model. The final model corresponds to epoch 39, which achieved the lowest validation loss of $2.9 \cdot 10^{-2}$.

As seen in Figure 17b and Figure 17c, both core resolution and energy resolution are noticeably worse across the entire energy range compared to the single-task and three-task Graphformer models. The energy bias (Figure 17d) also exhibits larger deviations, particularly at low energies, indicating that the joint optimization interferes with stable energy calibration.

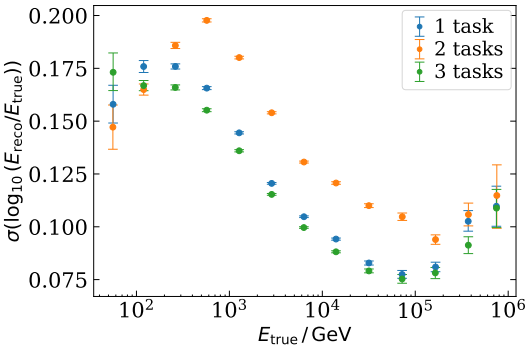
Overall, this task combination fails to provide any improvement in either energy or core reconstruction, suggesting that these tasks do not share sufficiently complementary information to benefit from joint learning.



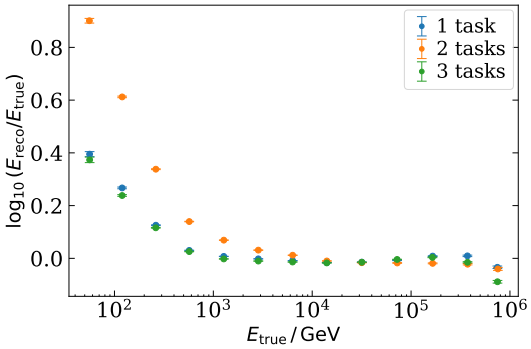
(a) The evolution of the training and validation losses for energy (green), core (blue) and the combined model (black).



(b) Core position resolution (68% containment radius) as a function of true energy.



(c) Energy resolution as a function of true energy.



(d) Energy bias as a function of true energy.

Figure 17: Comparison of the reconstruction resolution for the two-task model trained on core and energy prediction (orange) against the three-task multitask model (green) and the singletask Graphformer baseline (blue).

Summary and comparison

These findings indicate that single-task training remains robust for specialized objectives, while multitask training with all three tasks combined (energy, core and direction) achieves the most balanced and generally improved performance. The three-task Graphformer benefits from shared feature representations between the related tasks, while avoiding the instability observed in partial task combinations.

6.2 Gamma/hadron reconstruction and separation

In this final experiment, we train a multitask Graphformer model simultaneously on all four tasks: energy, core position, shower axis and particle classification, using a mixed dataset containing both γ-ray and hadron-initiated events. The objective is to evaluate whether the shared representations learned across all tasks can improve reconstruction and classification performance in a mixed-particle scenario.

Custom task weights were assigned to balance the loss contributions, as summarized below:

$$w_{\text{energy}} = 0.01\,, \quad w_{\text{core}} = 10^{-4}\,, \quad w_{\text{shower axis}} = 0.15\,, \quad w_{\text{primary}} = 0.02\,. \tag{5}$$

We again evaluate gamma events and hadron events separately.

The model was trained for 44 epochs, achieving its minimum validation loss of $4.2 \cdot 10^{-2}$ (Figure 18). Performance was then evaluated separately for gamma and hadron events to isolate potential differences in reconstruction behavior between the two primary particle types.

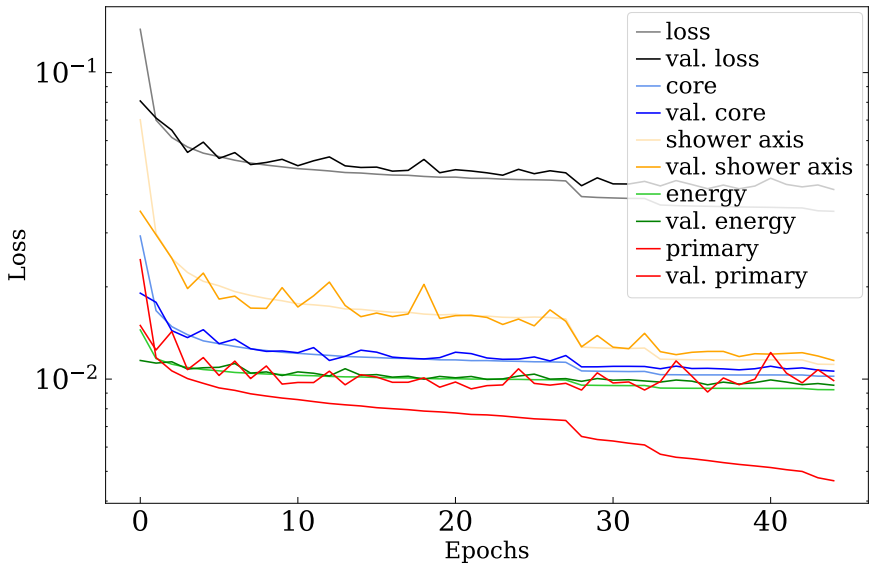
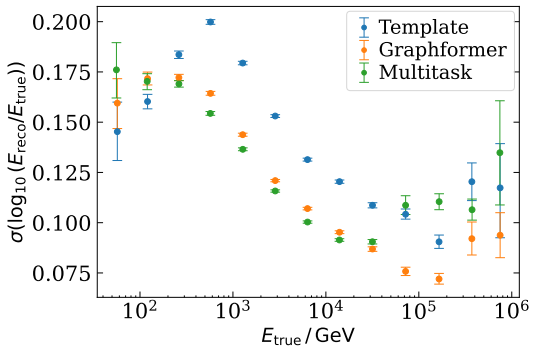


Figure 18: The evolution of the training and validation losses for primary (red), energy (green), shower axis (yellow), core (blue), and the combined model (black).

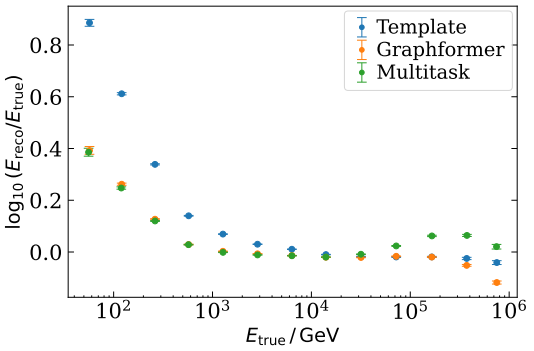
Energy reconstruction

For gamma-ray events, the multitask model performs slightly better than both the single-task Graphformer and the template method in energy resolution up to 10^4 GeV (Figure 19a). Beyond this energy, the performance becomes comparable to or slightly worse than the template, while the energy bias (Figure 19b) remains similar to the Graphformer up to 10^4 GeV and degrades thereafter.

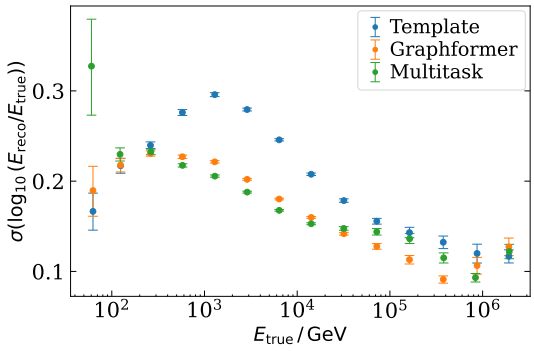
For hadronic events, the multitask model shows improved energy resolution at lower energies ($< 10^4$ GeV) but performs worse than the singletask Graphformer at higher energies, though still outperforming the classical template method (Figure 19c). The energy bias (Figure 19d) is only marginally better at low energies and deteriorates for intermediate and high energies.



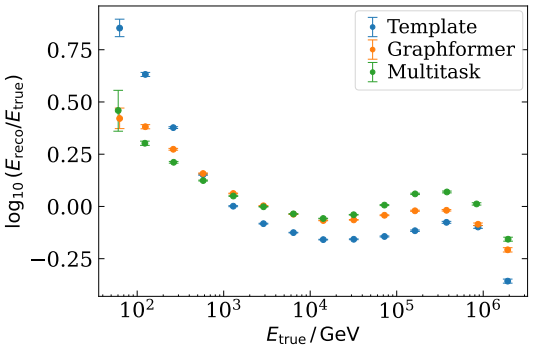
(a) Energy resolution for gamma events as a function of true energy.



(b) Energy bias for gamma events as a function of true energy.



(c) Energy resolution for hadron events as a function of true energy.



(d) Energy bias for hadron events as a function of true energy.

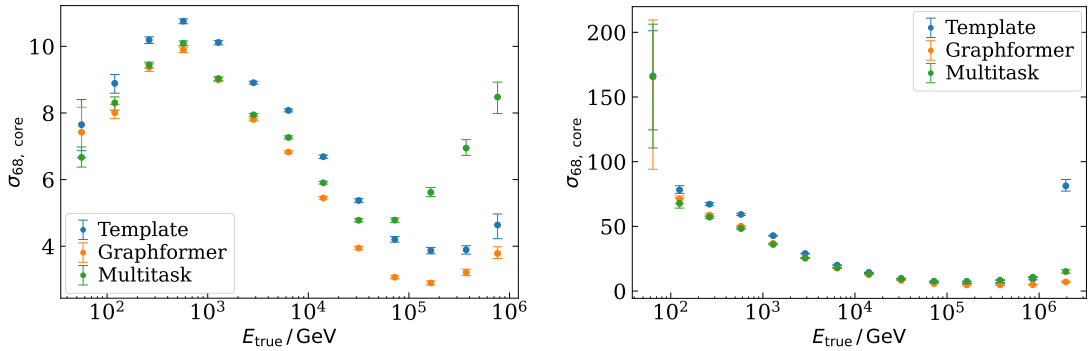
Figure 19: Gamma and hadron reconstruction for four-task multitasking (green) against template method (blue) and singletask Graphformer model (orange).

Core reconstruction

The results for core position reconstruction are shown in Figure 20a for γ-ray events and in Figure 20b for hadronic events.

For γ-ray events, the multitask model performs comparably to the singletask Graphformer at low energies but rapidly deteriorates at higher energies, indicating reduced generalization in this regime.

For hadronic events, the multitask model achieves slightly better performance at low energies, while at higher energies it becomes marginally worse than both the Graphformer and the template method.



(a) Core resolution for gamma events.

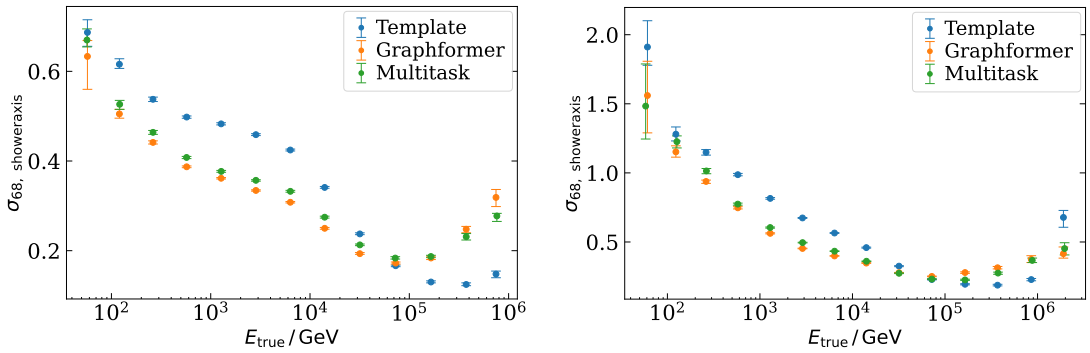
(b) Core resolution for hadron events.

Figure 20: Graphformer (orange) core reconstruction, evaluated for gamma (a) and proton (b) events separately, against template (blue) and multitasking (green).

Shower axis reconstruction

The results for shower-axis reconstruction are shown in Figure 21, with separate evaluations for γ-ray events (Figure 21a) and hadronic events (Figure 21b).

For both particle types, the multitask model performs nearly identically to the single-task Graphformer across all energy ranges. While the results remain stable and consistent, no significant improvement is observed from adding the additional tasks. This indicates that, for directional reconstruction, the model's capacity is already saturated in the single-task setting, and multitasking neither enhances nor degrades its ability to infer the shower axis



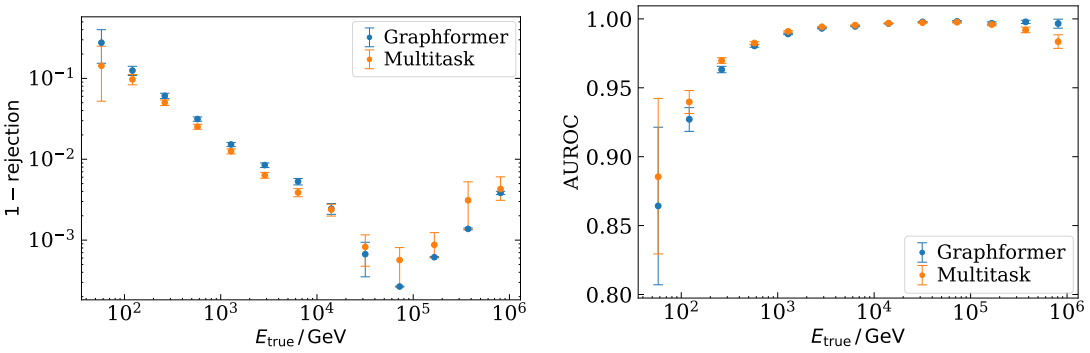
(a) Shower axis resolution for gamma events. **(b)** Shower axis resolution for hadron events.

Figure 21: Graphformer (orange) shower axis reconstruction, evaluated for gamma (a) and proton (b) events separately, against template (blue) and multitasking (green).

Gamma/hadron separation

The results for particle-type classification are presented in Figure 22, showing the background contamination (Figure 22a) and the area under the receiver operating characteristic curve (AUROC, Figure 22b).

The multitask model achieves a background rejection of about 90% at low energies and then exponentially increases to around 99.2% at $2 \cdot 10^4$ GeV, above this energy no background is left. The AUROC exhibits values of close to one for energies above 10^3 GeV, indicating strong discriminative power between electromagnetic and hadronic showers. However, the multitask configuration does not improve separation compared to singletask classification, suggesting that the shared learning between regression and classification objectives does not transfer effectively to the γ/hadron discrimination task.



(a) Background contamination as a function of true energy.

(b) AUROC curve for γ/hadron separation. Values close to 1 indicate excellent classification.

Figure 22: γ/hadron separation performed by the singletask Graphformer model (blue) compared with multitasking (orange).

Summary and comparison

Compared to the singletask configuration, the four-task multitask model performs comparably for energy and core reconstruction at low to intermediate energies, but its performance degrades at higher energies. The shower axis reconstruction and γ/hadron separation show nearly identical results to the singletask Graphformer model across all energy ranges.

The three-task gamma-only model, however, consistently outperforms the four-task setup in nearly all metrics. It benefits from a more homogeneous training sample and task focus, whereas the four-task model was trained on a smaller dataset, limited to 6000 gamma files due to computational constraints. Future work should re-evaluate this setup using multi-GPU training or sequential batch loading to enable larger, more balanced datasets and potentially recover lost performance.

Overall, multitask learning remains a promising strategy for integrated event reconstruction at SWGO, but in its current implementation, the three-task gamma-only configuration provides the best compromise between accuracy, stability, and training efficiency.

7 Conclusion and outlook

In this thesis, the performance of multitask learning for event reconstruction in the Southern Wide-field Gamma-ray Observatory (SWGO) was systematically investigated using the Graphformer architecture, a hybrid model combining local GNN-based feature extraction with global Transformer attention mechanisms. Using Monte Carlo simulations of γ-ray and hadronic air showers, the model was trained to predict four key event quantities: energy, arrival direction, core position, and particle type.

The multitask model trained on three tasks using gamma events achieved the best overall resolution, outperforming the template-likelihood baseline in both energy and core reconstruction while performing comparably to the single-task Graphformer. For angular resolution, the model exceeded the template performance up to 10^5 GeV and showed improved results compared to the singletask configuration at higher energies. The four-task mixed-particle model, which additionally included gamma/hadron classification, performed slightly worse overall, primarily due to reduced gamma statistics and computational constraints during training.

Future improvements should focus on optimized loss weighting strategies. Adaptive methods such as uncertainty-based weighting (Kendall et al., 2018) could replace manually tuned static weights, ensuring that all tasks contribute effectively during training. Additionally, larger and more diverse simulation datasets will further enhance model robustness and generalization. This can be achieved through multi-GPU training setups and sequential data loading, allowing for more extensive parameter exploration.

Acronyms

AGN Active galactic nucleus.

BCE Binary Cross Entropy.

CORSIKA COsmic Ray SImulations for KAscade.

CTAO Cherenkov Telescope Array Observatory.

EAS Extensive Air Showers.

Fermi-LAT Fermi Large Area Telescope.

GEANT4 GEometry ANd Tracking Toolkit.

GNN Graph Neural Network.

GRB gamma-ray burst.

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

HAWC High-Altitude Water Cherenkov.

IACT Imaging Atmospheric Cherenkov Telescope.

kNN k-nearest neighbor.

LHAASO Large High Altitude Air Shower Observatory.

MAE Mean Absolute Error.

MAGIC Major Atmospheric Gamma Imaging Cherenkov.

MC Monte Carlo.

ML Machine Learning.

MSE Mean Squared Error.

PMT Photomultiplier Tube.

PWN Pulsar Wind Nebula.

SGD Stochastic Gradient Descent.

SNR Supernova remnant.

SWGO Southern Wide-field Gamma-ray Observatory.

WCD Water-Cherenkov Detector.

Bibliography

Abeysekara, A. U., A. Albert, R. Alfaro, C. Alvarez, et al. (July 2023). "The High-Altitude Water Cherenkov (HAWC) observatory in México: The primary detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1052, p. 168253. ISSN: 0168-9002. DOI: 10.1016/j.nima.2023.168253. URL: https://www.sciencedirect.com/science/article/pii/S0168900223002437 (visited on 10/24/2025).

Agostinelli, S., J. Allison, K. Amako, J. Apostolakis, et al. (July 2003). "Geant4—a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3, pp. 250–303. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(03)01368-8. URL: https://www.sciencedirect.com/science/article/pii/S0168900203013688 (visited on 10/26/2025).

Aharonian, F., A. G. Akhperjanian, A. R. Bazer-Bachi, M. Beilicke, et al. (Oct. 2006). "Observations of the Crab nebula with HESS". en. In: *Astronomy & Astrophysics* 457.3. Publisher: EDP Sciences, pp. 899–915. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361:20065351. URL: https://www.aanda.org/articles/aa/abs/2006/39/aa5351-06/aa5351-06.html (visited on 10/24/2025).

Albert, J., E. Aliu, H. Anderhub, P. Antoranz, et al. (Apr. 2008). "Implementation of the Random Forest method for the Imaging Atmospheric Cherenkov Telescope MAGIC". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 588.3, pp. 424–432. ISSN: 0168-9002. DOI: 10.1016/j.nima.2007.11.068. URL: https://www.sciencedirect.com/science/article/pii/S0168900207024059 (visited on 10/24/2025).

Alvarez-Muñiz, J., Z. Cao, U.F. Katz, P. Mertsch, et al. (2024). *"Cosmic Rays", Review of Particle Physics 2024*. Particle Data Group.

Bose, D., V. R. Chitnis, P. Majumdar, and B. S. Acharya (Jan. 2022). "Ground-based gamma-ray astronomy: history and development of techniques". en. In: *The European Physical Journal Special Topics* 231.1, pp. 3–26. ISSN: 1951-6401. DOI: 10.1140/epjs/s11734-021-00396-3. URL: https://doi.org/10.1140/epjs/s11734-021-00396-3 (visited on 09/28/2025).

Collaboration, LHAASO (Jan. 2021). *Performance of LHAASO-WCDA and Observation of Crab Nebula as a Standard Candle*. DOI: 10.48550/arXiv.2101.03508. URL: http://arxiv.org/abs/2101.03508 (visited on 10/24/2025).

Collaboration, SWGO, P. Abreu, R. Alfaro, A. Alfonso, et al. (June 2025). *Science Prospects for the Southern Wide-field Gamma-ray Observatory: SWGO*. DOI: 10.48550/arXiv.2506.01786. URL: http://arxiv.org/abs/2506.01786 (visited on 10/01/2025).

Crawshaw, Michael (Sept. 2020). *Multi-Task Learning with Deep Neural Networks: A Survey*. DOI: 10.48550/arXiv.2009.09796. URL: http://arxiv.org/abs/2009.09796 (visited on 08/19/2025).

Elharrouss, Omar, Yasir Mahmood, Yassine Bechqito, Mohamed Adel Serhani, et al. (Apr. 2025). *Loss Functions in Deep Learning: A Comprehensive Review*. DOI: 10.48550/arXiv.2504.04242. URL: http://arxiv.org/abs/2504.04242 (visited on 10/19/2025).

Glombitza, Jonas, Martin Schneider, Franziska Leitl, Stefan Funk, and Christopher van Eldik (Feb. 2025). "Application of Graph Networks to a wide-field Water-Cherenkov-based Gamma-Ray Observatory". In: *Journal of Cosmology and Astroparticle Physics* 2025.02, p. 066. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2025/02/066. URL: http://arxiv.org/abs/2411.16565 (visited on 10/03/2025).

Hamilton, William L. (2020). *Graph Representation Learning*. en. Synthesis Lectures on Artificial Intelligence and Machine Learning. Cham: Springer International Publishing. ISBN: 978-3-031-01588-5. DOI: 10.1007/978-3-031-01588-5. URL: https://link.springer.com/10.1007/978-3-031-01588-5 (visited on 10/24/2025).

Heck, D., J. Knapp, J-N. Capdevielle, G. Schatz, and T. Thouw (1998). *CORSIKA: A Monte Carlo Code to Simulate Extensive Air Showers*. Tech. rep., pp. 1–90. URL: https://in2p3.hal.science/in2p3-00005094 (visited on 10/26/2025).

Hofmann, Werner and Roberta Zanin (May 2023). *The Cherenkov Telescope Array*. arXiv:2305.12888 [astro-ph]. DOI: 10.48550/arXiv.2305.12888. URL: http://arxiv.org/abs/2305.12888 (visited on 10/24/2025).

Holder, J., R. W. Atkins, H. M. Badran, G. Blaylock, et al. (July 2006). "The first VERITAS telescope". In: *Astroparticle Physics* 25.6, pp. 391–401. ISSN: 0927-6505. DOI: 10.1016/j.astropartphys.2006.04.002. URL: https://www.sciencedirect.com/science/article/pii/S092765050600051X (visited on 10/24/2025).

Joshi, Vikas, Jim Hinton, Harm Schoorlemmer, Rubén López-Coto, and Robert Parsons (Jan. 2019). "A template-based γ-ray reconstruction method for air shower arrays". en. In: *Journal of Cosmology and Astroparticle Physics* 2019.01, p. 012. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2019/01/012. URL: https://doi.org/10.1088/1475-7516/2019/01/012 (visited on 10/24/2025).

Jung, Alexander (2022). *Machine Learning: The Basics*. en. Machine Learning: Foundations, Methodologies, and Applications. Singapore: Springer Nature. ISBN: 978-981-16-8193-6. DOI: 10.1007/978-981-16-8193-6. URL: https://link.springer.com/10.1007/978-981-16-8193-6 (visited on 10/16/2025).

Kendall, Alex, Yarin Gal, and Roberto Cipolla (Apr. 2018). *Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics*. DOI: 10.48550/arXiv.1705.07115. URL: http://arxiv.org/abs/1705.07115 (visited on 08/19/2025).

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton (2012). "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems*. Vol. 25. Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html (visited on 10/27/2025).

Leitl, Franziska (2025). "Event Reconstruction for Particle Detector Arrays using Template-Based and Machine Learning Methods". PhD thesis, forthcoming. PhD thesis. Friedrich-Alexander-Universität Erlangen Nürnberg.

Matthews, J. (Jan. 2005). "A Heitler model of extensive air showers". In: *Astroparticle Physics* 22.5, pp. 387–397. ISSN: 0927-6505. DOI: 10.1016/j.astropartphys.2004.09.003. URL: https://www.sciencedirect.com/science/article/pii/S0927650504001598 (visited on 10/24/2025).

Parsons, R. D. and J. A. Hinton (Apr. 2014). "A Monte Carlo template based analysis for air-Cherenkov arrays". In: *Astroparticle Physics* 56, pp. 26–34. ISSN: 0927-6505. DOI: 10.1016/j.astropartphys.2014.03.002. URL: https://www.sciencedirect.com/science/article/pii/S0927650514000231 (visited on 10/24/2025).

Robino, Matteo (2025). "Optimization of high energy event reconstruction algorithms for the SWGO experiment". Master's thesis. MA thesis. Università degli Studi di Torino. URL: https://www.swgo.org/SWGOWiki/lib/exe/fetch.php?media=undefined:tesi_robino_matteo.pdf.

Schneider, Martin, Markus Pirke, Franziska Leitl, Jonas Glombitza, et al. (Sept. 2025). "Deep Learning Methods for Gamma/Hadron Separation in SWGO". en. In: *Proceedings of 39th International Cosmic Ray Conference — PoS(ICRC2025)*. Vol. 501. Conference Name: 39th International Cosmic Ray Conference. SISSA Medialab, p. 836. DOI: 10.22323/1.501.0836. URL: https://pos.sissa.it/501/836/ (visited on 10/05/2025).

Thompson, David J. and Colleen A. Wilson-Hodge (2023). "Fermi Gamma-ray Space Telescope". In: arXiv:2210.12875 [astro-ph], pp. 1–31. DOI: 10.1007/978-981-16-4544-0_58-1. URL: http://arxiv.org/abs/2210.12875 (visited on 10/24/2025).

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, et al. (2023). *Attention Is All You Need*. arXiv: 1706.03762 [cs.CL]. URL: https://arxiv.org/abs/1706.03762.

Wang, Qi, Yue Ma, Kun Zhao, and Yingjie Tian (Apr. 2022). "A Comprehensive Survey of Loss Functions in Machine Learning". en. In: *Annals of Data Science* 9.2, pp. 187–212. ISSN: 2198-5812. DOI: 10.1007/s40745-020-00253-5. URL: https://doi.org/10.1007/s40745-020-00253-5 (visited on 10/19/2025).

Wang, Yue, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, et al. (June 2019). *Dynamic Graph CNN for Learning on Point Clouds*. arXiv:1801.07829 [cs]. DOI: 10.48550/arXiv.1801.07829. URL: http://arxiv.org/abs/1801.07829 (visited on 10/26/2025).

Declaration of Originality

I, Leoni Kim Decker, student registration number: 22664497, hereby confirm that I completed the submitted work independently and without the unauthorized assistance of third parties and without the use of undisclosed and, in particular, unauthorized aids. This work has not been previously submitted in its current form or in a similar form to any other examination authorities and has not been accepted as part of an examination by any other examination authority.

Where the wording has been taken from other people's work or ideas, this has been properly acknowledged and referenced. This also applies to drawings, sketches, diagrams and sources from the Internet.

In particular, I am aware that the use of artificial intelligence is forbidden unless its use an aid has been expressly permitted by the examiner. This applies in particular to chatbots (especially ChatGPT) and such programs in general that can complete the tasks of the examination or parts thereof on my behalf.

Any infringements of the above rules constitute fraud or attempted fraud and shall lead to the examination being graded "fail" ("nicht bestanden").

Place, Date	Signature