

## Application of graph networks to a wide-field water-Cherenkov-based Gamma-ray Observatory

J. Glombitza , M. Schneider , F. Leitzl , S. Funk  and C. van Eldik 

*Erlangen Centre for Astroparticle Physics, Friedrich-Alexander-Universität Erlangen-Nürnberg,  
Nikolaus-Fiebiger-Str. 2, Erlangen 91058, Germany*

*E-mail:* [jonas.glombitza@fau.de](mailto:jonas.glombitza@fau.de), [martin.friedrich.schneider@fau.de](mailto:martin.friedrich.schneider@fau.de),  
[franziska.leitzl@fau.de](mailto:franziska.leitzl@fau.de)

**ABSTRACT:** With their wide field of view and high duty cycle, water-Cherenkov-based observatories are integral to studying the very high-energy gamma-ray sky. For gamma-ray observations, precise event reconstruction and highly effective background rejection are crucial and have been continuously improving in recent years. In this work, we investigate the application of graph neural networks (GNNs) to background rejection and energy reconstruction and benchmark their performance against state-of-the-art methods. In our simulation study, we find that GNNs outperform hand-designed classification algorithms and observables in background rejection and find an improved energy resolution compared to template-based methods.

**KEYWORDS:** gamma ray experiments, cosmic ray experiments, Machine learning

**ARXIV EPRINT:** [2411.16565](https://arxiv.org/abs/2411.16565)



---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Simulation of a WCD observatory in the southern hemisphere</b>	<b>2</b>
2.1	Simulated detector design	2
2.2	Data preparation and preprocessing	3
<b>3</b>	<b>Graph neural networks for SWGO</b>	<b>5</b>
3.1	Network design	5
3.2	Training	7
<b>4</b>	<b><math>\gamma</math>/hadron separation using GNNs</b>	<b>7</b>
4.1	Performance metrics	7
4.2	Separation performance using different features	8
4.3	Comparison to established classification observables	8
<b>5</b>	<b>Energy reconstruction using GNNs</b>	<b>12</b>
5.1	Energy dispersion	13
5.2	Energy bias and resolution	14
<b>6</b>	<b>Conclusion</b>	<b>16</b>
<b>A</b>	<b>Network architectures</b>	<b>17</b>
A.1	Task specific architectures	17
A.2	Kernel function and residual model	19
A.3	Parameters after hyperparameter search	19
A.4	Validation losses	20

---

## 1 Introduction

In the last decades, ground-based gamma-ray observatories such as WHIPPLE [1], HESS [2], MAGIC [3] and VERITAS [4] opened a new window to the non-thermal universe at very high energies. With these observatories, the search for the sources of cosmic rays — a mystery unsolved for more than 100 years — as well as astronomy of cosmic phenomena at the highest energies can be performed. To survey the gamma-ray sky at very-high energies, search for cosmic particle accelerators, and study diffuse gamma-ray emission within our galaxy, a large field-of-view observatory in the southern hemisphere is required to complement the future Cherenkov Telescope Array (CTA) [5]. Since 2019, the Southern Wide-field Gamma-ray Observatory [6] (SWGO) collaboration has been working on a next-generation gamma-ray observatory based on water-Cherenkov detectors (WCDs), which has been pioneered by Milagro [7, 8], and is also used in the currently operating HAWC [9] and LHAASO [10] observatories located in the northern hemisphere.

To maximize the sensitivity of any gamma-ray observatory, a powerful rejection of the dominating cosmic-ray background and precise reconstruction of the air shower properties induced by the impinging gamma ray is crucial. In the last decades, the reconstruction algorithms [11] have been continuously improved, e.g., using template-based methods [12, 13] and tree-based machine learning techniques applied to  $\gamma$ /hadron separation at Imaging Air Cherenkov Telescopes (IACTs) [14–16].

The latest progress in machine learning algorithms — based on deep neural networks (DNNs) — called *deep learning* [17], provides novel techniques for enhancing reconstruction algorithms in the physical sciences [17, 18]. Even the first application of neural networks to gamma-ray astronomy dates back to the 90s [19] and early 2000s [20]; to this time, they have not been capable due to their limited capacity to exploit the patterns in the recorded data in full detail. This has changed with *deep* networks that are able to analyze even tiny patterns in data with unprecedented precision. In turn, using DNNs, the event reconstruction capabilities for cosmic-ray research [21–23], neutrino [24, 25]–, and gamma astronomy [26–30] could be significantly improved, leading to new insights into the universe at the highest energies [31–33].

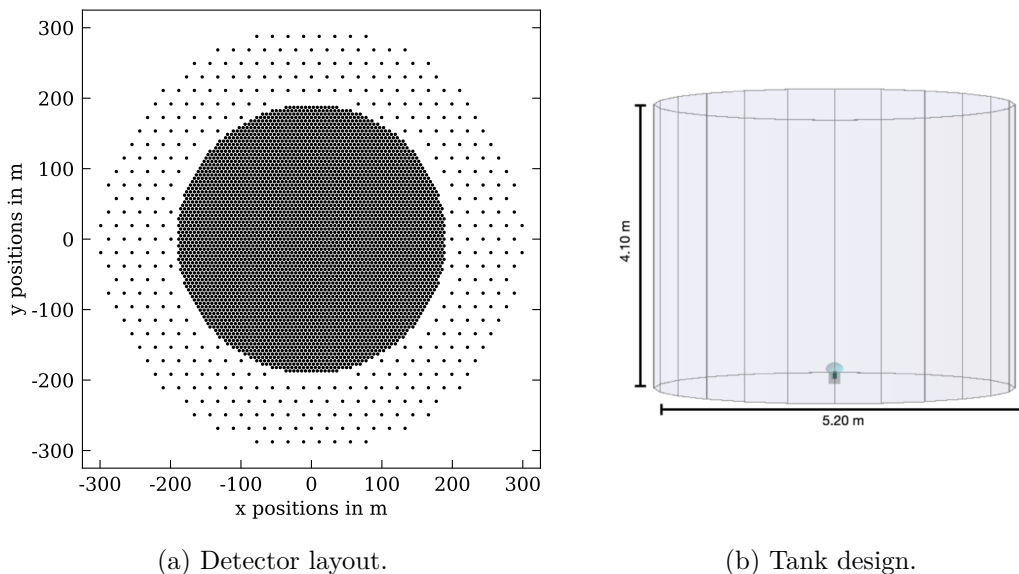
In this work, we investigate the application of DNNs to background rejection and energy reconstruction for a baseline design currently developed for SWGO. The graded layout considered in this work comprises around 4600 detector stations, and the number of triggered stations per event varies from tens to thousands but is small on average. Thus, we decided to use graph convolutional neural networks (GNNs) [34] to analyze the sparse signal patterns. The use of GNNs enables both the efficient analysis of signal patterns of different sizes and the exploitation of the advantages of convolutional neural networks, a central driving force in the recent success of deep learning [17]. With our GNN approach, we find significant improvements in hadronic background rejection compared to previous, hand-designed observables and classification algorithms and a promising energy reconstruction slightly surpassing the performance of currently state-of-the-art template-based reconstruction chains [13].

## 2 Simulation of a WCD observatory in the southern hemisphere

To prepare the data for our analysis, we use simulated events obtained from Monte-Carlo (MC) simulations for a baseline design currently developed within the SWGO collaboration. Similar to HAWC, SWGO uses a combination of standard simulation packages in use in the gamma-ray astronomy community, namely CORSIKA [35] for air shower simulations and GEANT4 [36] for the interaction of shower particles with the detector.

### 2.1 Simulated detector design

The stations in the considered design feature a large tank with a diameter of 5.20 m and a height of 4.10 m with an upward-looking 10-inch PMT located at the center of the tank floor, as shown in the right of figure 1. As such, it is very similar to the LHAASO and HAWC station designs. The studied layout comprises around 4600 detector units covering an area of roughly 280,000 m<sup>2</sup> and features a graded layout with two zones of different fill factors. Whereas in the central zone of a 188 m radius, the tanks are densely packed, reaching a fill factor of 80% to increase the sensitivity at low energies, the outer zone features a fill factor of 5% to increase the effective area at low cost.



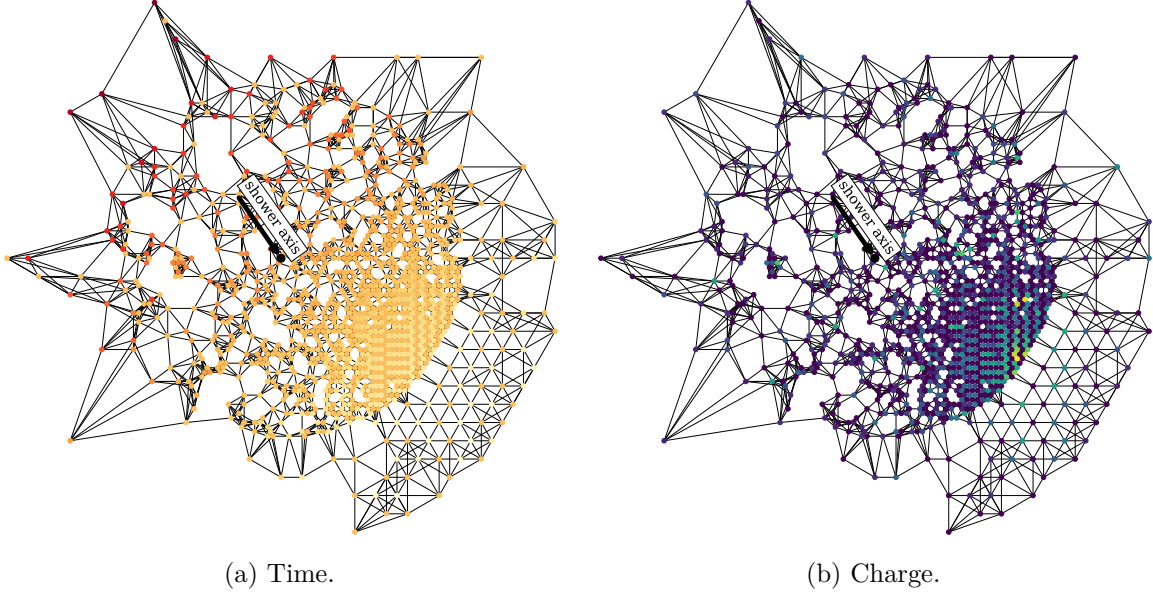
**Figure 1.** Detector configuration used in this work, currently investigated as a candidate design for SWGO [39]. Left: station layout featuring two zones with fill factors of 80% (central zone) and 5% (outer zone). Right: used water-Cherenkov tank design. The tank contains a single upward-looking 10-inch-PMT. Credit: SWGO collaboration [39], reproduced with permission.

We simulated this detector layout with the established HAWCsim package of HAWC [37], which was modified for SWGO [38]. This enables a realistic simulation of the complete detector response, which includes photomultiplier tube (PMT) noise, but still lacks cosmic ray-induced noise hits. The simulation set amounts to roughly 440,000 proton events and 370,000 gamma-ray events with a threshold of  $> 25$  detector hits. This is slightly lower than the trigger threshold of 30 hits<sup>1</sup> investigated in the following, to expand generalization capacities of the trained network at the edges of the phase space. The data is split 70%/10%/20% into training, validation, and test data sets. The events cover an energy range between 31.6 GeV and 1 PeV following a spectral index of  $-2$ , and zenith angles up to  $65^\circ$ . We use the same spectral index for protons and gamma rays to make the classifier response independent of the energy spectrum, which could lead to increased systematic uncertainties when, in reality, the gamma source spectrum differs from the one used during training.

## 2.2 Data preparation and preprocessing

To exploit the air shower footprint for event reconstruction and  $\gamma$ /hadron separation, we use the arrival time distribution of the particles detected at ground level. Since exploring the detailed temporal structure of the PMT signal is computationally quite demanding, we make use of integrated signal, i.e., charge, and do not use the signal trace directly. Thus, we use the  $x$ ,  $y$  positions of the tanks, the arrival times  $t$  of the first Cherenkov photon hitting the PMT of each station, as well as the measured charge of each PMT as input into our DNN. Each point in this point cloud resembles the measurements of the PMT of a single

<sup>1</sup>A first approximation of the threshold realizable for the examined design assuming an available bandwidth of 2 Gb/s.



**Figure 2.** Graph representation of a simulated gamma-ray event with a true energy of 18.8 TeV and a zenith angle of  $39.7^\circ$  obtained after  $k$ nn clustering. Left: detected arrival time, yellow = early, red = late. Right: deposited charge, blue = low, yellow = high.

tank. The location of each point is given by the  $x$  and  $y$  coordinates of the tanks, and the arrival time  $t$  and signal charge  $q$  of the PMT form additional features of each point. Thus, each point holds  $2 + 2$  features for each single PMT.

### 2.2.1 Graph construction

The point cloud is used for the construction of the graph in the following way. We apply a  $k$ -next neighbor clustering ( $k$ nn) in the spatial dimension to define a graph using the  $N$  triggered stations of a given event, including self-connections. Each node in the graph will be connected to itself and its six nearest neighbors, which are determined based on the Euclidean distance between the points. The number of six neighbors is motivated by the triangular grid of the layout of SWGO, as each tank has six direct neighbors. This yields a directed graph with  $N$  nodes and  $7 \times N$  edges. An example of such a feature graph of an example event is shown in figure 2. The arrival times (left) and charges (right) are color-encoded.

### 2.2.2 Normalization

To stabilize the GNN training, we perform feature normalization of the positions, PMT charge, and time. For the timing information z-score normalization [18] is used  $t' = (t - \mu)/\sigma_{\text{std}}$  while a logarithmic re-scaling is performed for the charges  $q' = (\log_{10}(1 + q))/\sigma_{\text{std}}$ . Here,  $\mu$  denotes the mean and  $\sigma_{\text{std}}$  the standard deviation of the time and charge<sup>2</sup> distribution estimated over all events, respectively. For the station positions, we also perform z-score normalization.

<sup>2</sup>After applying the transform  $\log_{10}(q + 1)$ .

### 3 Graph neural networks for SWGO

The footprint of a typical gamma-ray shower at several TeV roughly covers an area of hundreds of  $\text{m}^2$ . For the currently discussed detector designs for SWGO [39], this translates into tens to thousands of detector stations, forming a set of triggered PMTs at ground level. Image-like data can be precisely and efficiently analyzed with deep learning techniques using convolutional neural networks (CNNs) [17, 18]. In CNNs, small filters with adaptive parameters exploit the image information by enforcing translational invariance and setting a prior on local correlations. However, the use of a filter with a fixed size in CNNs requires the analyzed data to be distributed on a homogeneous grid, which should have low sparsity to ensure efficient usage. The strong variation in the number of triggered stations and the graded layout makes using CNNs for analyzing the SWGO raw data unfeasible since either the algorithm training is inefficient or information needs to be dropped [27]. In the scope of this work, we model an SWGO event as a signal graph consisting of nodes and edges, where the nodes are formed by the triggered stations holding the information of the station’s PMTs as node features  $\mathbf{x}_i = (x, y, q, t)^T$ , and the edges connect stations with local proximity.

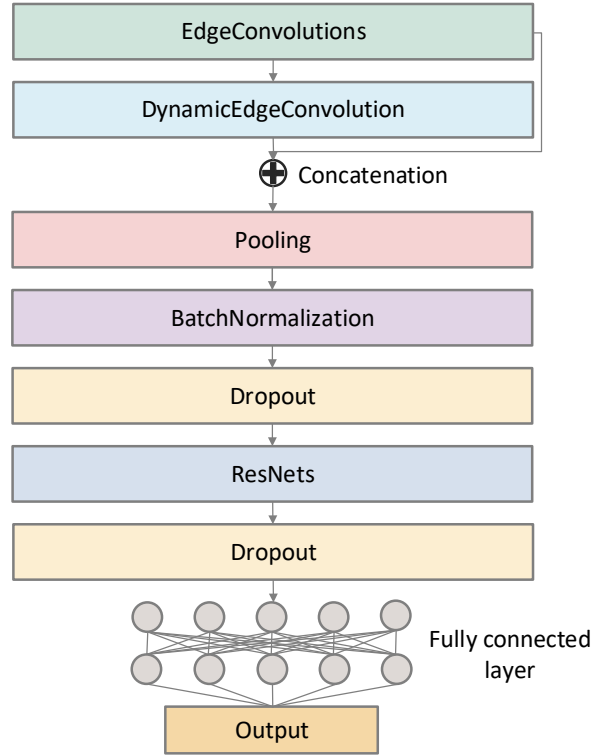
#### 3.1 Network design

We make use of EdgeConvolutions [40] in this work that were already successfully applied to various physics challenges [41–44]. The EdgeConvolution can be divided into graph construction, convolution, and aggregation. After the graph generation (see section 2.2), we regard the immediate neighborhood  $\mathbf{x}_j$  of each “central” node  $\mathbf{x}_i$ , which depends on a fixed number of  $k$ NN next neighbors. In this study, we use  $k = 6$ . The EdgeConvolution of nodes and aggregation of information is defined as:

$$\mathbf{x}'_i = \sigma \left( \bigoplus_{j \in \mathcal{N}(i)} h_{\Theta}(\mathbf{x}_i \parallel \mathbf{x}_j - \mathbf{x}_i) \right). \quad (3.1)$$

Each pair of  $(\mathbf{x}_j - \mathbf{x}_i)$ , in the neighborhood of each  $\mathbf{x}_i$ , is convolved leading to a number of  $f$  different edge features. Note that the  $\parallel$  indicates that the subtraction is not applied to the central node, i.e., the self-connection. This convolution is performed by the kernel function  $h_{\Theta}$  with output dimension  $f$ , to be chosen by the user. The kernel function combines the information of the feature vector of  $\mathbf{x}_i$  with the edge features of its neighboring nodes. Finally, the  $k$  edge features of the  $k$  neighbors will be aggregated using an aggregation function  $\bigoplus_{j \in \mathcal{N}(i)}$  and later passed through an activation function  $\sigma$ . Such aggregation functions combine, i.e., aggregate, information from neighbors of a node to produce a summary representation. Possible choices include mean, maximum, or the sum (used in this work), which means that the output remains independent of the order of the neighbors, conserving permutational invariance. This leads to the feature vector  $\mathbf{x}'_i$  after the edge convolution, which is located at the position of the central node. This convolutional operation is performed in parallel for each node  $\mathbf{x}_i$  of the graph.

Each EdgeConvolution layer disseminates information in its immediate vicinity, giving local information to the network. To exploit beyond local and global correlations across the whole graph, we stack multiple graph layers to ensure a large receptive field of view [18]. Furthermore, DynamicEdgeConvolutions [40] can be utilized, to extract global features, which



**Figure 3.** Sketch of the general network structure used for the  $\gamma$ /hadron separation and energy reconstruction. The network consists of a number of EdgeConvolution blocks, followed by a DynEdgeConvolution layer. After the concatenation of the layers, the result is propagated through batch normalization, dropout, ResNet blocks, and another dropout layer. Finally, a small, fully-connected layer is applied, which leads to the final output of the network.

perform a graph construction step during training by finding the  $k$ -closest nodes in the feature space. Since graph construction causes computational overhead, as it needs to be recalculated for each layer and sample during training, we use DynamicEdgeConvolutions in a limited fashion.

In this analysis, the general network structure for each task was inspired by the ParticleNet network described in ref. [41]. A sketch of the general network structure can be found in figure 3. The detailed network structure used for the  $\gamma$ /hadron separation and energy reconstruction can be found in the appendix A (see table 1 for the  $\gamma$ /hadron separation and table 2 for the energy reconstruction). In general, both networks consist of a number of consecutive EdgeConvolutions to gather local information and a DynamicEdgeConvolution block to extract global information from the air shower footprint. Depending on the reconstruction task, the number of EdgeConvolution and DynamicEdgeConvolution layers vary, as we found that different architectures exhibit slightly improved GNN performance optimized for the used layout configuration. Similar to ref. [41], the kernel function  $h_{\Theta}$  of the EdgeConvolution consists of three fully-connected layers where each is being followed by a batch normalization and an activation function, see table 3(a) for all the details. Next, a global pooling layer, performing pooling along the node dimension, is used to make the final part of the network



independent of the number of triggered stations. Finally, a set of task-specific fully-connected layers containing the output follows, completing the GNN structure.

### 3.2 Training

We performed extensive hyperparameter tuning for  $\gamma$ /hadron separation and energy reconstruction to converge to the final GNN architecture for each task. The best models for each configuration are chosen based on the lowest observed validation loss. All networks were implemented using PyTorch Geometric [45], and trained on single NVIDIA A40 or A100 GPUs, with training times ranging from 8 up to 24 hours depending on the hyperparameter configuration. We used binary cross-entropy for g/h separation and mean-squared-error as the loss function for energy reconstruction. During training, we applied a strategy that involved reducing the learning rate upon reaching a plateau and incorporated early stopping. Thus, the learning rate was lowered if the validation loss did not improve for five epochs, and training was stopped entirely if no improvement was observed after 11 epochs.

A hyperparameter search of 70 trainings is performed for each task-specific network, varying different parameters using a random search. As optimizer, we used Adam [46]. We vary parameters that affect the learning process, such as learning rate, weight decay, and decay factor,<sup>3</sup> dropout [47], and if batch normalization [48] is used, as well as parameters like the number of graph convolutional layers, the size of  $k$  in the clustering step of DynamicEdgeConvolution, the number of kernel features  $f$ , and the number of ResNet blocks<sup>4</sup> (see table 3(b) for the ResNet architecture). Note that dropout is applied once after the graph pooling and once at the end of the output of the last ResNet layer. A detailed table of the final parameters of the search is shown in table 4.

For the energy reconstruction, we apply a minor modification of each convolution block by using a residual connection in contrast to the  $\gamma$ /hadron separation model. As required by ResNet, the input of each convolution block is projected via a fully connected layer prior to the addition of the output of the convolutional layer to ensure similar dimensions. Finally, a ReLU activation is applied. The energy of each predicted event is given in  $\log_{10}(E/\text{GeV})$ . More information on the architecture and the validation losses can be found in the appendix A.

## 4 $\gamma$ /hadron separation using GNNs

In the following section, we investigate the performance of the GNN in terms of  $\gamma$ /hadron separation. Our trained GNN model output is a score that indicates the likelihood of the event to be a gamma ray (output close to 1) or a proton (output close to zero).<sup>5</sup>

### 4.1 Performance metrics

For estimating the performance of the trained GNN classifier, we study the false positive rate  $\epsilon_p$  (fraction of protons relative to the simulated number of protons misclassified as

<sup>3</sup>The decay factor is used to reduce the learning rate after a certain number of epochs without improvements in the validation loss.

<sup>4</sup>ResNet blocks refer to a stack of two fully-connected layers with a shortcut connection that is utilized to perform an identity mapping by simply adding the inputs to the outputs of the stacked layers. This learning of residual functions has been proven to ease the training of DNNs [49].

<sup>5</sup>Two example score distributions (protons = blue, gamma = red) are shown in figure 7.



gamma rays), in the following referred to as *background contamination*, as a function of reconstructed energy [51]. We study the energy dependence of the background contamination for a fixed gamma-ray efficiency  $\epsilon_\gamma$ , i.e., the surviving fractions of gamma rays, which is usually set to be  $\approx 80\%$  [52].

We also make use of the *quality factor*, which combines proton contamination and gamma-ray efficiency and is defined via:

$$Q = \frac{\epsilon_\gamma}{\sqrt{\epsilon_p}}. \quad (4.1)$$

Assuming that each bin contains a sufficiently large number of events,  $Q$  is proportional to the discovery Poisson significance improvement of a source. By changing the score threshold of the classifier, i.e., the score at which an event is tagged as a gamma ray, the quality factor will be maximized<sup>6</sup> to optimize sensitivity [7].

## 4.2 Separation performance using different features

As a first step, we study the impact of different inputs: charge, arrival time, and the positions of triggered stations, i.e., the signal pattern induced by the shower footprint, for the network architecture. To study the  $\gamma$ /hadron separation performance using only positional information, the network is trained using two positional features  $x$  and  $y$  and a feature representing the PMT signal, which was set to random uniform noise, with unit variance and zero mean. Additionally, the performance of adding either the timing or charge information is investigated by using this information as a feature in place of the noise. The performance is shown in figure 4 using the events of the test data set. Error bars denote statistical uncertainties and are estimated following ref. [50]. Utilizing only positional information (red markers), i.e., the position of triggered stations, the GNN has only weak  $\gamma$ /hadron separation capabilities. Adding time as a feature improves  $\gamma$ /hadron separation performance, with a stronger gain at higher energies. Using only charge instead of time information yields excellent separation performance, which is not surprising, given that mostly signal-based classifiers [7, 37, 52–54] primarily focusing on detecting signal fluctuations in the shower footprint, i.e., hadronic sub showers, were used for  $\gamma$ /hadron separation in the past. The best result is obtained by using both timing and charge information as input, which slightly improves the performance over the whole energy range, resulting in an overall background contamination ranging from 5% at low energies to below 0.1% at the highest energies.

## 4.3 Comparison to established classification observables

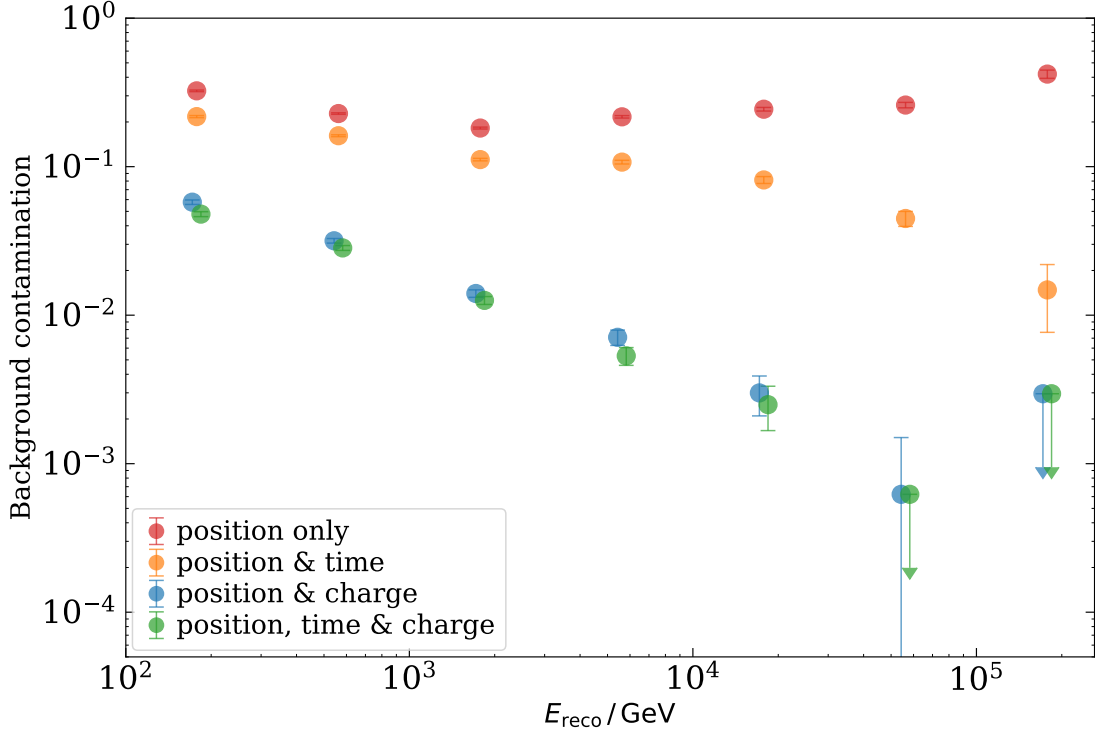
In the following subsection, we compare the GNN performance to two established observables that have proven their effectiveness and are currently used within HAWC [52].

### 4.3.1 Definition of baselines observables

The  $\gamma$ /hadron separation utilized in HAWC relies on these two distinct parameters, called LIC and PINCness, to distinguish between cosmic-ray and gamma-ray events. The LIC parameter

---

<sup>6</sup>An example is shown in figure 7, alongside the quality factor as a function of the threshold on the GNN output for the low energy and high energy regime.



**Figure 4.** Comparison of the background rejection performance, when training the GNN using different inputs. The residual background contamination using all available features is shown in green. Orange and blue show the network performance with positions and time or positions and charge used as input features. In red, the performance of the network utilizing only positional information is shown. Error bars denote statistical uncertainties following ref. [50].

is the logarithm of the inverse of the so-called compactness [7], originally developed for Milagro:

$$\text{LIC} = \log_{10} \left( \frac{CxPE_{40}}{n_{\text{hit}}} \right). \quad (4.2)$$

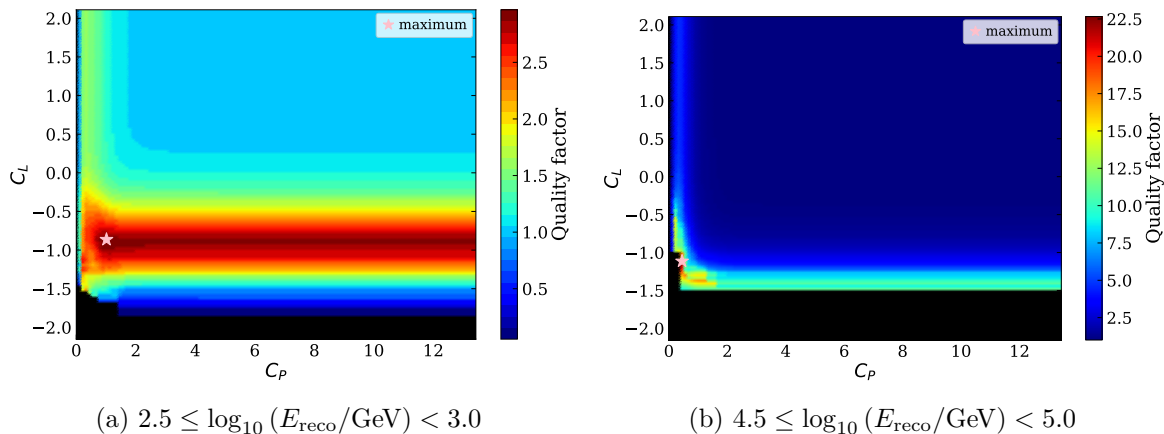
Here,  $CxPE_{40}$  denotes the largest charge measured in a PMT at least 40 m away from the shower core, and  $n_{\text{hit}}$  is the number of tanks hit in the event (our example design features only a single PMT per tank).

The second parameter, PINCness [37], makes use of the lateral distribution function to separate gamma rays from cosmic rays. It is defined as

$$\text{PINCness} = \frac{1}{N} \sum_{i=0}^N \frac{(\log_{10}(q_i) - \langle \log_{10}(q_i) \rangle)^2}{\sigma^2}, \quad (4.3)$$

where  $q_i$  is the charge of the  $i$ -th PMT that was triggered in the event and  $\sigma$  is the uncertainty on the charge. It is calculated by fitting Gaussian distributions to the logarithm of average charges for different rings centered on the shower core and then fitting the best-fit values to a quadratic function.

Given that simple box cuts on these two parameters are widely used to combine both parameters for  $\gamma$ /hadron separation, we optimize them in the same way as ref. [55] to define



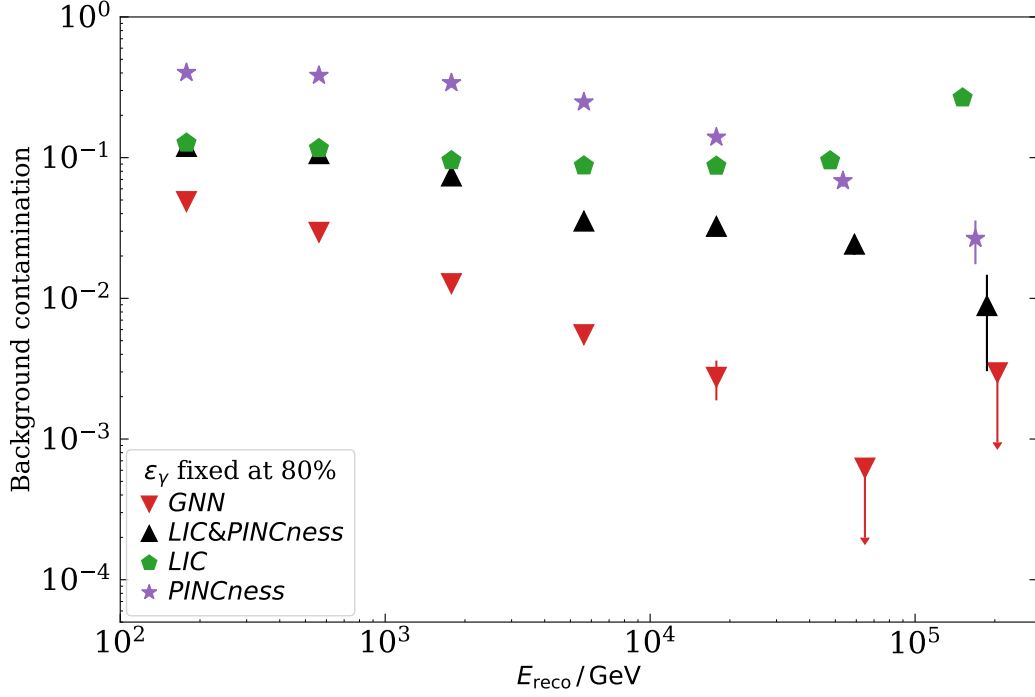
**Figure 5.** The quality factor shown as a function of a cut on PINcness ( $C_P$ ) and LIC ( $C_L$ ) for two different energy bins. The maximum of the quality factor is indicated by a pink star. The thresholds  $C_P$  and  $C_L$  at this point provide the optimal  $\gamma$ /hadron separation in terms of maximizing the sensitivity as estimated by the quality factor and are used in the following.

a baseline to compare to. We find the optimal cuts for the LIC and PINcness parameters by optimizing the quality factor for every bin in our dataset, which consists of seven logarithmic reconstructed energy bins from  $10^{2.0}$  GeV to  $10^{5.5}$  GeV with a step size of half a decade. Figure 5 shows the best cuts obtained for one low-energy bin and one high-energy bin.

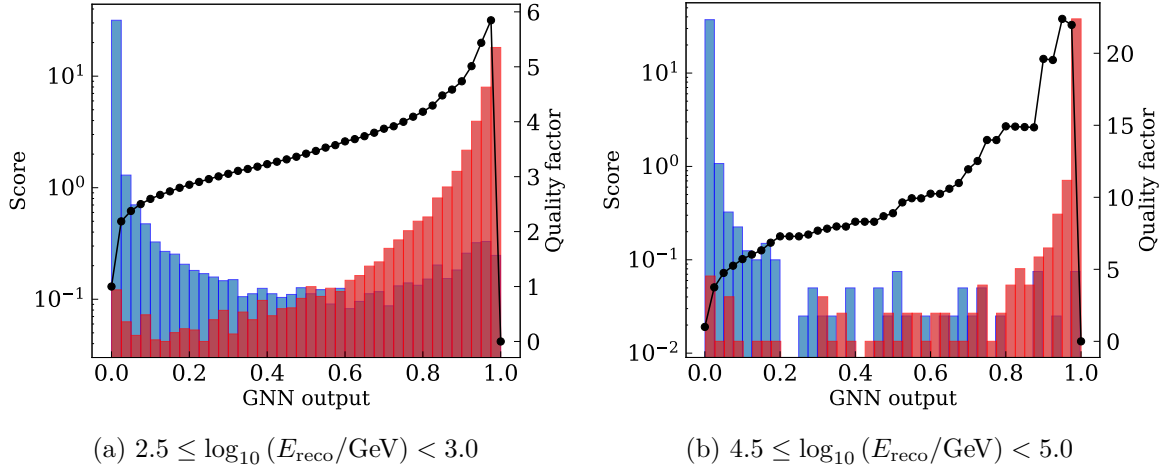
### 4.3.2 Performance comparisons

To compare the GNN to an established baseline, we investigated the background contamination for the different  $\gamma$ /hadron separators as a function of energy in figure 6. To enable a direct comparison, the gamma-ray efficiency was fixed to 80%. Note that we did not apply any quality cuts on the data. Statistical uncertainties are estimated following ref. [50]. If no background event is present in an energy bin, an upper limit is estimated by assuming a single background event and depicted by a down-facing arrow. As can be seen, the GNN (red triangles) features a more efficient background rejection than LIC (green pentagons) and PINcness (purple stars) across the full studied range and even outperforms the combination of both observables (black triangles). Whereas at low energies, the rejection is improved by around a factor of two, at higher energies, we find an improvement by one order of magnitude.

**Quality factor.** Since the quality of the  $\gamma$ /hadron separation directly propagates into the sensitivity that depends on  $\epsilon_p$  and  $\epsilon_\gamma$ , the quality factor can be used to optimize source detections in the Gaussian limit. In each bin, we optimized the quality factor (requiring a gamma-ray efficiency of at least 50% to retain sufficient signal) for the different methods respectively. Example plots for a low-energy and high-energy bin are shown in figure 7 for the GNN, where the score distribution of the protons (blue histogram) and gamma rays (red histogram) are shown alongside the quality factor (black line). In case there are no surviving background events, we instead estimate a lower limit of the quality factor with a single background event and indicate the limits with down-facing arrows. To examine the effect of using the GNN on source sensitivity, we show in figure 8 the background contamination

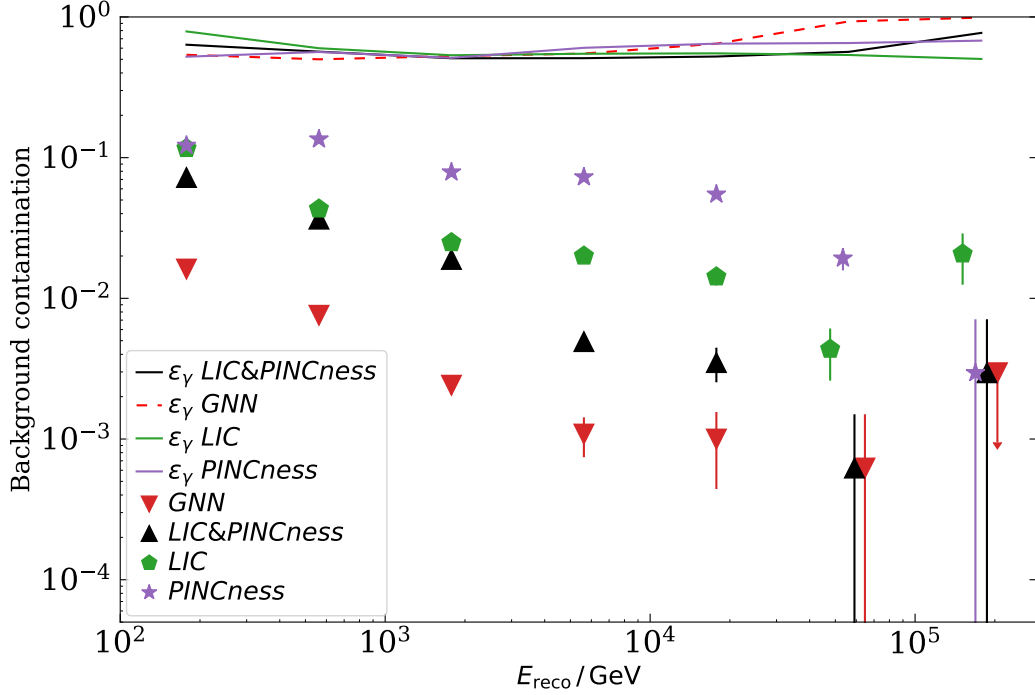


**Figure 6.**  $\gamma$ /hadron separation performance comparing the GNN (red triangles) to PINcness (purple stars), LIC (green pentagons), and their combination (black triangles). The gamma-ray efficiency was fixed to 80% to enable a fair comparison. The data points in the last two bins shifted horizontally to improve readability. Error bars denote statistical uncertainties. Upper limits are estimated by assuming a single background event and are depicted by downward-facing arrows.



**Figure 7.** The GNN output for protons (blue) and gamma rays (red) using events in two different energy bins. The quality factor  $q$  is shown in black here as a function of the score threshold (x-axis) of the GNN classifier.

(colored markers) for the investigated approaches and the gamma-ray efficiency (colored lines) as a function of energy. In terms of standalone parameters, LIC, shown in green, generally outperforms PINcness, shown in purple, in the low and medium energy range. We find that our GNN, shown in red, outperforms the combination of LIC & PINcness,



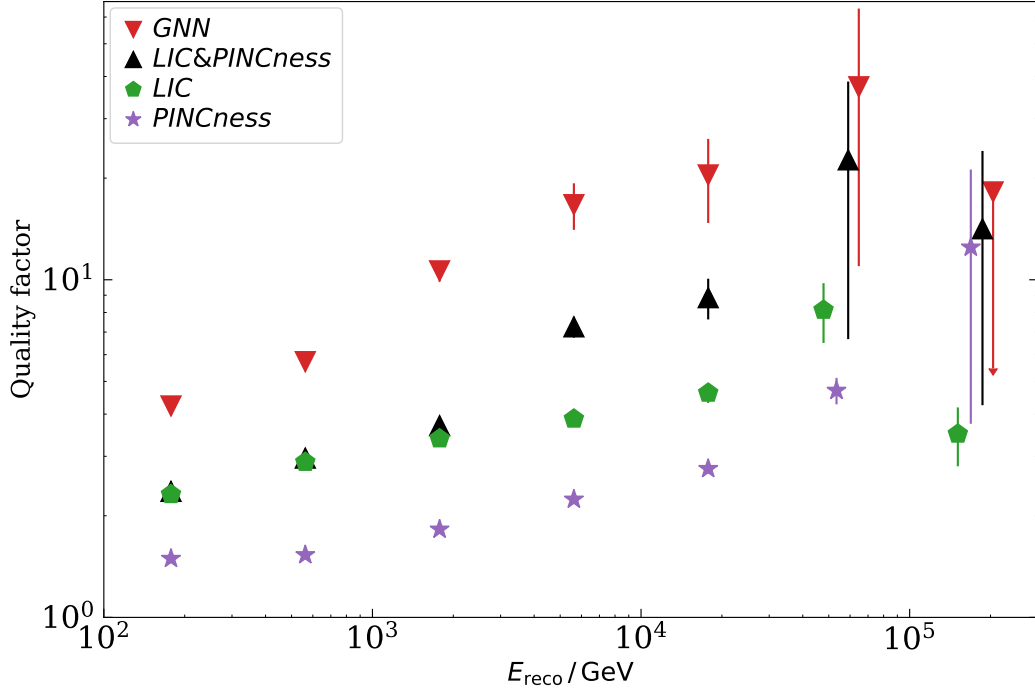
**Figure 8.**  $\gamma$ /hadron separation performance comparing the GNN to PINCness, LIC, and their combination without any quality cuts applied. The cuts were chosen to maximize the quality factor for a gamma-ray efficiency (shown as solid lines) of at least 50%. The data points in the last two bins shifted horizontally to improve readability. Error bars denote statistical uncertainties. Downward-facing arrows indicate upper limits for bins if no background remains. No quality cuts have been applied.

depicted in black, by a factor of three at low energies and reaches a factor of five up to eight at medium and high energies.

The comparison of the optimized quality factor as a function of energy is shown in figure 9. The error bars denote statistical uncertainties and were estimated following ref. [50]. For bins without background contamination, indicated by a down-facing array below the marker, a lower limit is estimated, assuming a single background event. It can be seen that the GNN (red triangles) outperforms the classical observables (black triangles, green pentagons, and purple stars) over the whole energy range. This demonstrates that the GNN efficiently extracts additional information from the sampled shower footprint, facilitating a significant gain in the signal-to-noise ratio. Overall, the GNN offers a two to three times higher quality factor over the whole range and continues to improve the background rejection at very high energies. However, the exact extent of the  $\gamma$ /hadron separation performance is somewhat unclear above 50 TeV given the limited statistics in these bins.

## 5 Energy reconstruction using GNNs

In general, we compare the performance of the energy reconstruction for events with zenith angle  $\theta \leq 30^\circ$  and  $30^\circ < \theta \leq 45^\circ$ . Showers at larger zenith angles have a much broader spatial spread across the array and have to travel a longer way through the atmosphere, which reduces the number of particles reaching the detector.

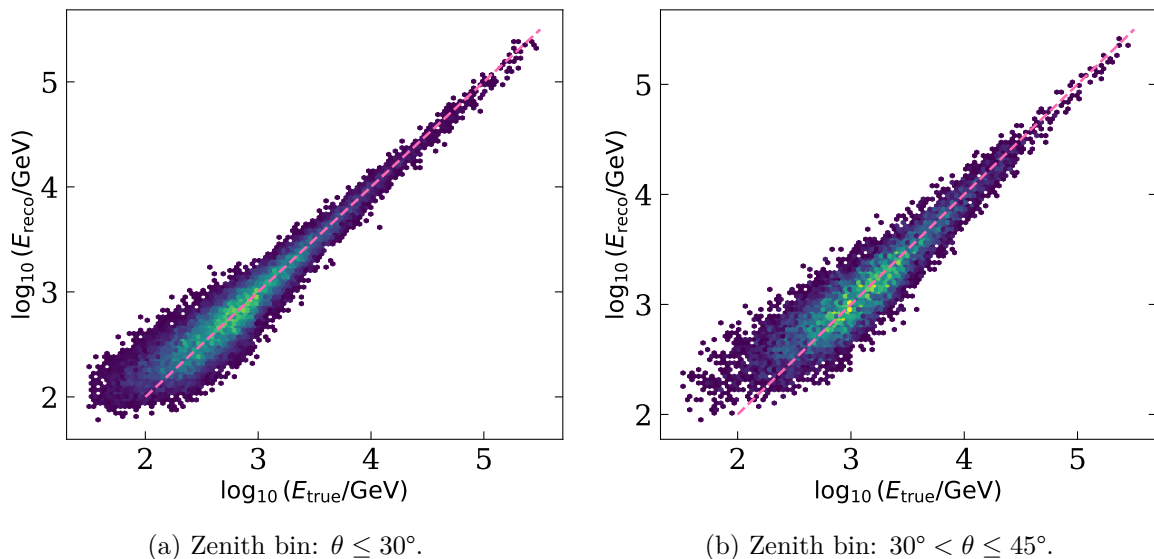


**Figure 9.** Comparison of the optimized quality factor for GNN (red), PINCness (purple), LIC (green), and their combination (black) without any quality cuts applied. Cuts were chosen to ensure at least 50% gamma-ray efficiency. Upper limits are estimated assuming a single background event in the bin. Due to limited statistics resulting in large errors, the data points were shifted to avoid overlap.

**Data selection.** For the energy reconstruction, we apply a dedicated selection, which comprises cuts on the zenith angle  $\theta \leq 45^\circ$ , the core position of the shower within the array ( $r_{\text{array}} < 300 \text{ m}$ ), and  $n_{\text{hit}} > 30$ . We also apply a cut on the score of the reconstructed events, which essentially acts as a quality cut for well-reconstructed gamma rays. The cut value was taken from the previous  $\gamma$ /hadron separation. In the energy reconstruction we only reconstruct gamma rays.

### 5.1 Energy dispersion

In figure 10, we compare the energy dispersion for events at smaller zenith angles figure 10(a) with ones at larger zenith angles figure 10(b). We plot the reconstructed energy  $\log_{10}(E_{\text{reco}}/\text{GeV})$  of the GNN versus its initial true energy  $\log_{10}(E_{\text{true}}/\text{GeV})$ . In an ideal case, the events in the plot align along the diagonal where  $\log_{10}(E_{\text{reco}}/\text{GeV}) = \log_{10}(E_{\text{true}}/\text{GeV})$ . In both cases, we find that while there is a slight shift towards higher reconstructed energies, the distribution of events aligns well along this diagonal. Within each zenith bin, we find that the distribution at smaller energies deviates slightly more from the diagonal than at higher energies. This presumably stems from higher energy showers triggering more tanks, providing more information to the network. Additionally, lower energies are more affected by upward fluctuations because they are close to the energy threshold of the detector. We find that the overall distribution in the higher zenith angle bin is slightly broader than the one at lower zenith angles. This is caused by the fact that the shower has to travel a longer



**Figure 10.** Energy dispersion matrix for two different zenith angle bins. The diagonal shows where  $\log_{10}(E_{\text{reco}}) = \log_{10}(E_{\text{true}})$ .

distance through the atmosphere, leading to increased fluctuations in the shower development. We find that there are more low-energy events at lower zenith angles than at higher angles. Low-energy showers at high zenith angles are less likely to reach the array due to atmospheric absorption effects, thus changing the energy threshold.

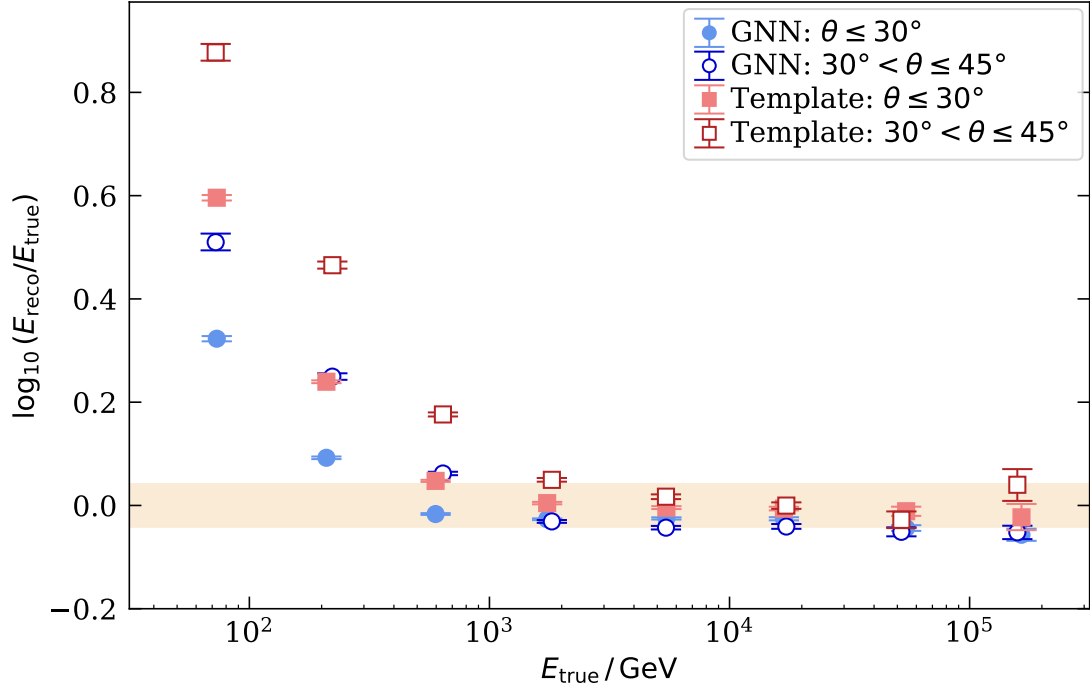
## 5.2 Energy bias and resolution

To judge the performance of the GNN, we investigate the energy bias and energy resolution, which we define as the mean and standard deviation of the distribution  $\log_{10}(E_{\text{reco}}/E_{\text{true}})$  and its uncertainties which we obtain by bootstrapping. We compare the GNN results to the performance that can be attained using the current state-of-the-art in [13] and [51], which relies on a template-based likelihood fitting procedure.

In figure 11, the energy bias and in figure 12, the energy resolution is shown as a function of the energy. We show the results of the GNN-based energy reconstruction as circles compared to the performance that can be attained for the standard template method as squares for different zenith angle bins, filled markers for the  $\theta \leq 30^\circ$  and open markers for  $30^\circ < \theta \leq 45^\circ$ . The energy bias is stable over multiple orders of magnitude from 600 GeV to around 200 TeV. Where the bias for both methods stays roughly within 10% on a linear scale, this region is marked in yellow. At lower energies below 600 GeV, reconstruction is biased. Since low-energetic events are less likely to be detected, only events with upwards fluctuations, i.e. showers that feature a shower maximum close to the detector, are measured. This shifts the bias to higher energies, which affects the standard method more than the GNN. The bias for the smaller zenith angle bin generally performs better compared to higher angles, which can again be explained by common shower physics.

In figure 12, we compare the energy resolution for the standard method as squares and the GNN method in circles for different zenith angle bins. Again, we show filled markers for



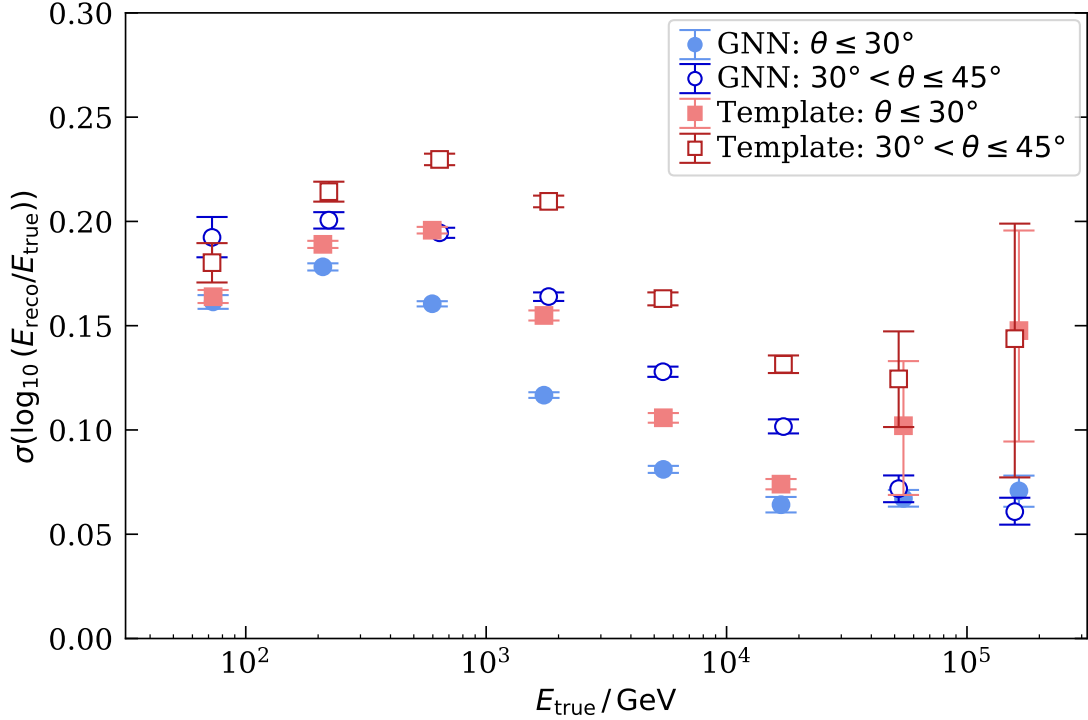


**Figure 11.** Comparison of the energy bias using the current state-of-the-art template approach [13] and [51], which uses a likelihood fitting procedure (square markers) and the GNN (circles), as a function of energy for two zenith bins with filled markers for the  $\theta \leq 30^\circ$  bin and open markers for the  $30^\circ < \theta \leq 45^\circ$  bin. The bias is estimated as the mean of the distribution of  $\log_{10}(E_{\text{reco}}/E_{\text{true}})$ . The yellow marked area shows the  $\pm 10\%$  region around 0 for the energy bias in linear.

the  $\theta \leq 30^\circ$  bin and open markers for the  $30^\circ < \theta \leq 45^\circ$  bin. For energies below 200 GeV for the GNN, and below around 600 GeV for the template method, we find that the energy resolution is getting slightly better with decreasing energy. The reason for this is that the cut on the number of tanks hit acts as a quality cut for events in that region. As it is harder for low-energy showers to reach the threshold for reconstruction, more events are removed, artificially boosting the resolution in this region with the cost of a larger bias (compare the trend below 600 GeV in figure 11 and figure 12). Above 600 GeV, this threshold no longer influences the results. Beyond 200 GeV across the entire energy range, we find the GNN resolution is consistently surpassing what is currently achievable with the standard method. At the highest energy above 100 TeV, the resolution seems to worsen slightly, presumably due to the size of the shower footprint getting larger than the size of the array.

For the GNN resolution, we reach a resolution of around 16% (linear scale) at around 17 TeV for zenith angles  $\theta \leq 30^\circ$  and a resolution of around 26% at the same energy for higher zenith angles.

Overall, we find a stable energy bias over many orders of magnitude for both the template method and the GNN and a very good energy resolution, especially medium energies, with the GNN showing evident improvements over the standard state-of-the-art method and improving over the performance requirement in the science proposal of SWGO [56].



**Figure 12.** Comparison of the energy resolution using the current state-of-the-art template approach [13] and [51], which uses a likelihood fitting procedure (square markers) and the GNN (circles), as a function of energy for two zenith bins with filled markers for the  $\theta \leq 30^\circ$  bin and open markers for the  $30^\circ < \theta \leq 45^\circ$  bin. We calculate the resolution as the standard deviation of the distribution of  $\log_{10}(E_{\text{reco}}/E_{\text{true}})$ .

## 6 Conclusion

In this work, we developed a novel deep-learning-based strategy for event reconstruction and  $\gamma$ /hadron separation in the context of SWGO, a next-generation gamma-ray observatory in the southern hemisphere currently in the research and development phase. By modeling air shower footprints detected by a simulated water-Cherenkov detector array as graphs, we exploit the spatio-temporal patterns comprising the measured integrated signals and the timing information using graph neural networks (GNNs). In the context of our simulation study, we examined the graph-based algorithms on energy reconstruction and  $\gamma$ /hadron separation for a detector design currently studied within the SWGO collaboration.

We presented the first comprehensive study of the application of graph neural networks to a water-Cherenkov-based gamma-ray observatory. The obtained performance within this study surpasses the performance expected in the SWGO white paper [56]. We found that without the need for any quality cuts, the GNN provides a strong background rejection, outperforming hand-design variables currently in use within HAWC over the whole energy range. These improvements are achieved by combining the spatial structure of the footprint, which is particularly efficient in discriminating showers at high energies, with the charge and timing information. Interestingly, the timing information, which has not been investigated in detail for  $\gamma$ /hadron separation, appears to have additional separation power.

Furthermore, we found that GNNs offer an accurate energy reconstruction that provides reliable energy estimates from 600 GeV over the full studied energy range. The GNN reconstruction offers improvements to the template-based approach [51], which is currently state-of-the-art, demonstrating an improved resolution, particularly at mid and high energies. Due to the flexible nature of the proposed GNN architecture, changes in the layout and tank design, e.g., adding second chambers to enable the tagging of single muons, can be simply integrated into the presented algorithm.

Future work will focus on additional algorithm improvements, i.e., performing different clustering within the central and outer zones, utilizing attention mechanisms, and investigating its performance under realistic operation conditions, including cosmic ray noise. This aims to exploit the full information contained in the detected air shower footprint, ultimately providing improved capabilities to survey the gamma-ray sky at very high energies.

## Acknowledgments

We thank the SWGO collaboration for allowing us to use SWGO simulations and to make use of the SWGO reconstruction software for this publication and the use of the common shared software framework (AERIE), kindly provided by HAWC. We would like to particularly express our gratitude to R. Turner and X. Wang for their work on the PINCness observable in the SWGO reconstruction chain. The authors gratefully acknowledge the scientific support and HPC resources provided by the Erlangen National High Performance Computing Center (NHR@FAU) of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) under the NHR project b129dc. NHR funding is provided by federal and Bavarian state authorities. NHR@FAU hardware is partially funded by the German Research Foundation (DFG) – 440719683.

## A Network architectures

### A.1 Task specific architectures

Architecture for $\gamma$ /hadron separation					
	Layer	features	setting	input shape	output shape
$3 \times \left\{ \begin{array}{l} \text{EdgeConv} \\ \text{ReLU} \end{array} \right.$	EdgeConv	$n_{\text{feat}}$	$\square_j$ : mean	$n_{\text{nodes}} \times 4$	$n_{\text{nodes}} \times n_{\text{feat}}$
	ReLU	—	—	$n_{\text{nodes}} \times n_{\text{feat}}$	$n_{\text{nodes}} \times n_{\text{feat}}$
	EdgeConv	$n_{\text{feat}}$	$\square_j$ : mean	$n_{\text{nodes}} \times n_{\text{feat}}$	$n_{\text{nodes}} \times n_{\text{feat}}$
	ReLU	—	—	$n_{\text{nodes}} \times n_{\text{feat}}$	$n_{\text{nodes}} \times n_{\text{feat}}$
	DynEdgeConv	$n_{\text{feat}}$	$\square_j$ : mean	$n_{\text{nodes}} \times n_{\text{feat}}$	$n_{\text{nodes}} \times n_{\text{feat}}$
	Concat	—	layer outputs	$4 \cdot (n_{\text{nodes}} \times n_{\text{feat}})$	$4 \cdot (n_{\text{nodes}} \times n_{\text{feat}})$
	MaxPooling	—	—	$4 \cdot (n_{\text{nodes}} \times n_{\text{feat}})$	$4 \cdot (n_{\text{nodes}} \times n_{\text{feat}})$
	Dropout	—	$p = 0.13$	$4 \cdot (n_{\text{nodes}} \times n_{\text{feat}})$	$4 \cdot (n_{\text{nodes}} \times n_{\text{feat}})$
	ResNet	$n_{\text{feat}}$	—	$4 \cdot (n_{\text{nodes}} \times n_{\text{feat}})$	$n_{\text{feat}}$

Architecture for $\gamma$ /hadron separation					
	Layer	features	setting	input shape	output shape
$4 \times \{$	ResNet	$n_{\text{feat}}$	—	$n_{\text{feat}}$	$n_{\text{feat}}$
	Dropout	—	$p = 0.13$	$n_{\text{feat}}$	$n_{\text{feat}}$
	Linear	2	—	$n_{\text{feat}}$	2
	SoftMax	—	—	$n_{\text{feat}}$	2

**Table 1.** Network architectures for the separation task.  $n_{\text{nodes}}$  describes the number of nodes in the input graph, which depend on the number of triggered tanks, and  $n_{\text{feat}}$  the number of features. We further use a bottleneck layer in the first ResNet layer to adjust the dimensionality.

Architecture for energy reconstruction					
	Layer	features	setting	input shape	output shape
$3 \times \left\{ \begin{array}{l} \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{array} \right.$	EdgeConv	$n_{\text{feat}}$	$\square_j$ : mean	$n_{\text{nodes}} \times 4$	$n_{\text{nodes}} \times n_{\text{feat}}$
	Linear	$n_{\text{feat}}$	—	$n_{\text{nodes}} \times 4$	$n_{\text{nodes}} \times n_{\text{feat}}$
	Addition	—	—	$n_{\text{nodes}} \times 4$	$n_{\text{nodes}} \times n_{\text{feat}}$
	ReLU	—	—	$n_{\text{nodes}} \times n_{\text{feat}}$	$n_{\text{nodes}} \times n_{\text{feat}}$
	EdgeConv	$n_{\text{feat}}$	$\square_j$ : mean	$n_{\text{nodes}} \times n_{\text{feat}}$	$n_{\text{nodes}} \times n_{\text{feat}}$
	Linear	$n_{\text{feat}}$	—	$n_{\text{nodes}} \times n_{\text{feat}}$	$n_{\text{nodes}} \times n_{\text{feat}}$
	Addition	—	—	$n_{\text{nodes}} \times n_{\text{feat}}$	$n_{\text{nodes}} \times n_{\text{feat}}$
	ReLU	—	—	$n_{\text{nodes}} \times n_{\text{feat}}$	$n_{\text{nodes}} \times n_{\text{feat}}$
	DynEdgeConv	$n_{\text{feat}}$	$\square_j$ : mean	$n_{\text{nodes}} \times n_{\text{feat}}$	$n_{\text{nodes}} \times n_{\text{feat}} \cdot 2$
	Linear	$n_{\text{feat}}$	—	$n_{\text{nodes}} \times n_{\text{feat}}$	$n_{\text{nodes}} \times n_{\text{feat}} \cdot 2$
	Addition	—	—	$n_{\text{nodes}} \times n_{\text{feat}}$	$n_{\text{nodes}} \times n_{\text{feat}} \cdot 2$
	ReLU	—	—	$n_{\text{nodes}} \times n_{\text{feat}} \cdot 2$	$n_{\text{nodes}} \times n_{\text{feat}} \cdot 2$
	Concat	—	layer outputs	$5 \cdot (n_{\text{nodes}} \times n_{\text{feat}})$	$5 \cdot (n_{\text{nodes}} \times n_{\text{feat}})$
	MaxPooling	—	—	$5 \cdot (n_{\text{nodes}} \times n_{\text{feat}})$	$5 \cdot (n_{\text{nodes}} \times n_{\text{feat}})$
	Batchnorm	—	momentum = 0.1	$5 \cdot (n_{\text{nodes}} \times n_{\text{feat}})$	$5 \cdot (n_{\text{nodes}} \times n_{\text{feat}})$
	Dropout	—	$p = 0.24$	$5 \cdot (n_{\text{nodes}} \times n_{\text{feat}})$	$5 \cdot (n_{\text{nodes}} \times n_{\text{feat}})$
	ResNet	$n_{\text{feat}}$	—	$5 \cdot (n_{\text{nodes}} \times n_{\text{feat}})$	$n_{\text{feat}}$
$2 \times \{$	ResNet	$n_{\text{feat}}$	—	$n_{\text{feat}}$	$n_{\text{feat}}$
	Dropout	—	$p = 0.24$	$n_{\text{feat}}$	$n_{\text{feat}}$
	Linear	1	—	$n_{\text{feat}}$	1

**Table 2.** Network architectures for the energy reconstruction task.  $n_{\text{nodes}}$  describes the number of nodes in the input graph, which depend on the number of triggered tanks, and  $n_{\text{feat}}$  the number of features. The “Addition” denotes a residual connection that connects consecutive convolutional and linear layers. We further use a bottleneck layer in the first ResNet layer to adjust the dimensionality.

## A.2 Kernel function and residual model

Kernel network $h_{\Theta}$		
layer	features	settings
Linear	$n_{\text{feat}}$	no bias
Batchnorm	—	momentum = 0.9
Activation	—	—
Linear	$n_{\text{feat}}$	no bias
Batchnorm	—	momentum = 0.9
Activation	—	—
Linear	$n_{\text{feat}}$	no bias
Batchnorm	—	momentum = 0.9
Activation	—	—

**Table 3 (a).** Architecture of the kernel function  $h_{\Theta}$ , where  $n_{\text{feat}}$  is to be chosen by the user. For  $\gamma$ /hadron separation, we used SiLU as an activation function and for the energy reconstruction ReLU.

Residual Module “ResNet”		
layer	features	settings
Linear	$n_{\text{feat}}$	no bias
SiLU	—	—
Batchnorm	—	momentum = 0.1
Linear	$n_{\text{feat}}$	no bias
SiLU	—	—
Batchnorm	—	momentum = 0.1
Add	—	—
SiLU	—	—

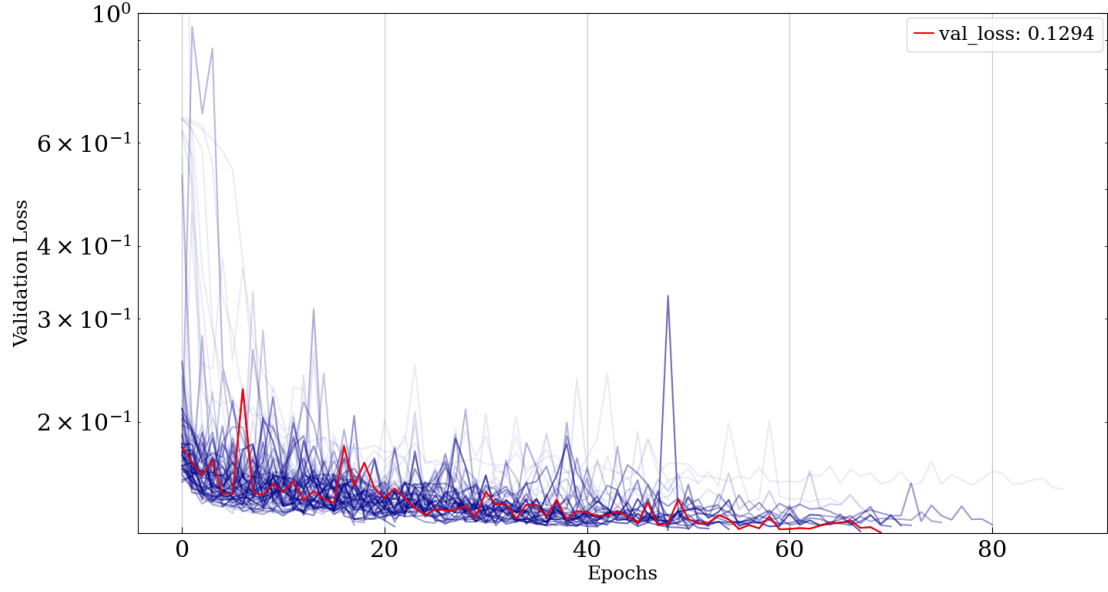
**Table 3 (b).** Details of our used ResNet modules.

## A.3 Parameters after hyperparameter search

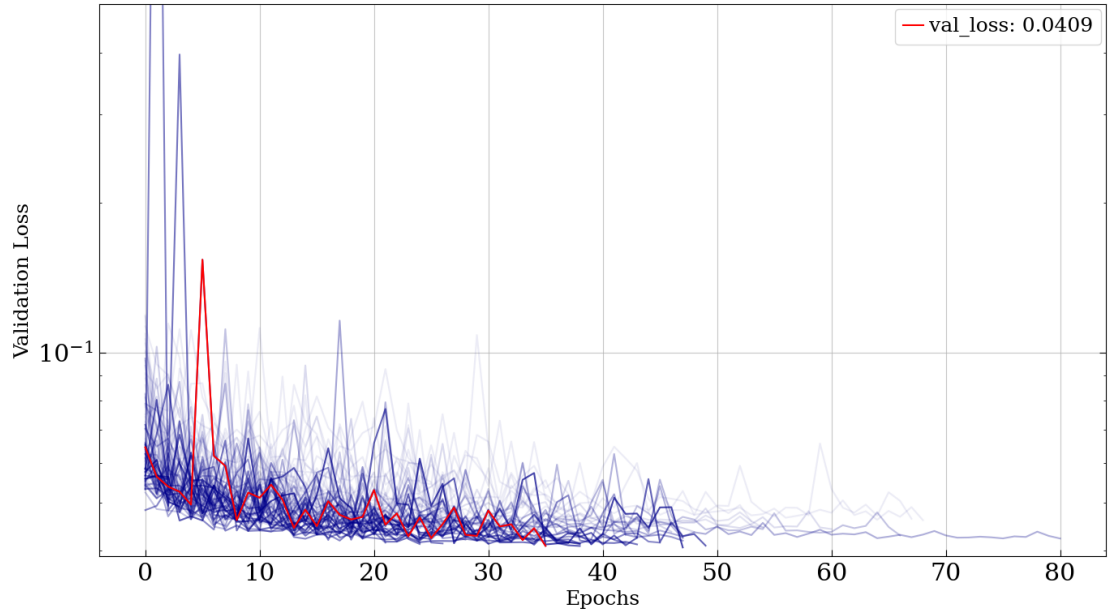
Parameters	$\gamma$ /hadron separation	Energy reconstruction	range
Learning rate	0.005	0.010	[5e-5, 1.5e-2]
Decay factor	0.39	0.29	[0.2, 0.8]
Batchsize	135	105	[96, 150]
Weight decay	0 (fixed)	8.5e-4	[5e-5, 1e-3]
$n_{\text{EdgeConv}}$	3	3	[1, 4]
$n_{\text{DynEdgeConv}}$	1	1	[1, 2]
$n_{\text{feat}}$	57	124	[16, 128]
$n_{\text{ResNet}}$	5	3	[1, 5]
Batchnorm	False	True	[True, False]
Dropout	0.13	0.24	[0, 0.5]
$n_{k\text{NN}, \text{DynEdgeConv}}$	16 (fixed)	13	[8, 16]

**Table 4.** Found parameters after the hyperparameter search for  $\gamma$ /hadron separation and energy reconstruction.  $n_{\text{EdgeConv}}$  and  $n_{\text{DynEdgeConv}}$  denote the number of EdgeConvolution and DynamicEdgeConvolution layers, and  $n_{\text{ResNet}}$  the number of ResNet blocks.  $n_{\text{feat}}$  is the number of features and  $n_{k\text{NN}, \text{DynEdgeConv}}$  the number of  $k\text{NN}$  neighbors for the DynamicEdgeConvolution.

#### A.4 Validation losses



**Figure 13.** History of the validation losses during the hyperparameter search for  $\gamma$ /hadron separation. The best performing model is shown in red.



**Figure 14.** History of the validation losses during the hyperparameter search for energy reconstruction. The best performing model is shown in red.

## References

- [1] J. Kildea et al., *The Whipple Observatory 10 m gamma-ray telescope, 1997–2006*, *Astropart. Phys.* **28** (2007) 182 [[INSPIRE](#)].
- [2] H.E.S.S. collaboration, *Observations of the Crab Nebula with H.E.S.S.*, *Astron. Astrophys.* **457** (2006) 899 [[astro-ph/0607333](#)] [[INSPIRE](#)].
- [3] MAGIC collaboration, *Performance of the MAGIC stereo system obtained with Crab Nebula data*, *Astropart. Phys.* **35** (2012) 435 [[arXiv:1108.1477](#)] [[INSPIRE](#)].
- [4] J. Holder et al., *The first VERITAS telescope*, *Astropart. Phys.* **25** (2006) 391 [[astro-ph/0604119](#)] [[INSPIRE](#)].
- [5] B.S. Acharya et al., *Introducing the CTA concept*, *Astropart. Phys.* **43** (2013) 3 [[INSPIRE](#)].
- [6] P. Abreu et al., *The Southern Wide-Field Gamma-Ray Observatory (SWG0): A Next-Generation Ground-Based Survey Instrument for VHE Gamma-Ray Astronomy*, [arXiv:1907.07737](#) [[INSPIRE](#)].
- [7] MILAGRO collaboration, *Observation of TeV gamma-rays from the Crab nebula with MILAGRO using a new background rejection technique*, *Astrophys. J.* **595** (2003) 803 [[astro-ph/0305308](#)] [[INSPIRE](#)].
- [8] MILAGRO collaboration, *Milagrito: A TeV air shower array*, *Nucl. Instrum. Meth. A* **449** (2000) 478 [[astro-ph/9912456](#)] [[INSPIRE](#)].
- [9] A.U. Abeysekara et al., *The High-Altitude Water Cherenkov (HAWC) observatory in México: The primary detector*, *Nucl. Instrum. Meth. A* **1052** (2023) 168253 [[arXiv:2304.00730](#)] [[INSPIRE](#)].
- [10] LHAASO collaboration, *Performance of LHAASO-WCDA and Observation of Crab Nebula as a Standard Candle*, [arXiv:2101.03508](#) [[INSPIRE](#)].
- [11] A.J. Smith, *HAWC: Design, Operation, Reconstruction and Analysis*, *PoS ICRC2015* (2016) 966 [[arXiv:1508.05826](#)] [[INSPIRE](#)].
- [12] R.D. Parsons and J.A. Hinton, *A Monte Carlo Template based analysis for Air-Cherenkov Arrays*, *Astropart. Phys.* **56** (2014) 26 [[arXiv:1403.2993](#)] [[INSPIRE](#)].
- [13] V. Joshi et al., *A template-based  $\gamma$ -ray reconstruction method for air shower arrays*, *JCAP* **01** (2019) 012 [[arXiv:1809.07227](#)] [[INSPIRE](#)].
- [14] S. Ohm, C. van Eldik and K. Egberts, *Gamma-Hadron Separation in Very-High-Energy gamma-ray astronomy using a multivariate analysis method*, *Astropart. Phys.* **31** (2009) 383 [[arXiv:0904.1136](#)] [[INSPIRE](#)].
- [15] M. Krause, E. Pueschel and G. Maier, *Improved  $\gamma$ /hadron separation for the detection of faint  $\gamma$ -ray sources using boosted decision trees*, *Astropart. Phys.* **89** (2017) 1 [[arXiv:1701.06928](#)] [[INSPIRE](#)].
- [16] J. Albert et al., *Implementation of the Random Forest Method for the Imaging Atmospheric Cherenkov Telescope MAGIC*, *Nucl. Instrum. Meth. A* **588** (2008) 424 [[arXiv:0709.3719](#)] [[INSPIRE](#)].
- [17] Y. LeCun, Y. Bengio and G. Hinton, *Deep learning*, *Nature* **521** (2015) 436 [[INSPIRE](#)].
- [18] M. Erdmann, J. Glombitza, G. Kasieczka and U. Klemradt, *Deep Learning for Physics Research*, World Scientific (2021) [[DOI:10.1142/12294](#)] [[INSPIRE](#)].



- [19] S. Westerhoff et al., *Separating gamma and hadron induced cosmic ray air showers with feed forward neural networks using the charged particle information*, *Astropart. Phys.* **4** (1995) 119 [INSPIRE].
- [20] P. Boinee et al., *Neural networks for gamma-hadron separation in MAGIC*, in the proceedings of the 6th International Symposium on Frontiers of Fundamental Physics, Udine, Italy, 26–29 September 2004, Springer, Dordrecht, Netherlands (2006), pp. 297–302 [DOI:10.1007/1-4020-4339-2\_41] [astro-ph/0503539] [INSPIRE].
- [21] M. Erdmann, J. Glombitza and D. Walz, *A deep learning-based reconstruction of cosmic ray-induced air showers*, *Astropart. Phys.* **97** (2018) 46 [arXiv:1708.00647] [INSPIRE].
- [22] PIERRE AUGER collaboration, *Deep-learning based reconstruction of the shower maximum  $X_{\max}$  using the water-Cherenkov detectors of the Pierre Auger Observatory*, *2021 JINST* **16** P07019 [arXiv:2101.02946] [INSPIRE].
- [23] PIERRE AUGER collaboration, *Extraction of the muon signals recorded with the surface detector of the Pierre Auger Observatory using recurrent neural networks*, *2021 JINST* **16** P07016 [arXiv:2103.11983] [INSPIRE].
- [24] R. Abbasi et al., *A Convolutional Neural Network based Cascade Reconstruction for the IceCube Neutrino Observatory*, *2021 JINST* **16** P07041 [arXiv:2101.11589] [INSPIRE].
- [25] S. Aiello et al., *Event reconstruction for KM3NeT/ORCA using convolutional neural networks*, *2020 JINST* **15** P10005 [arXiv:2004.08254] [INSPIRE].
- [26] C.R. Bom et al., *Bayesian Deep Learning for Shower Parameter Reconstruction in Water Cherenkov Detectors*, *PoS ICRC2021* (2021) 739 [INSPIRE].
- [27] F. Assunção et al., *Automatic Design of Artificial Neural Networks for Gamma-Ray Detection*, *IEEE Access* **7** (2019) 110531.
- [28] B.S. González et al., *Tackling the muon identification in water Cherenkov detectors problem for the future Southern Wide-field Gamma-ray Observatory by means of Machine Learning*, *Neural Comput. Appl.* **34** (2022) 5715 [arXiv:2101.11924] [INSPIRE].
- [29] R. Conceição et al., *Muon identification in a compact single-layered water Cherenkov detector and gamma/hadron discrimination using machine learning techniques*, *Eur. Phys. J. C* **81** (2021) 542 [arXiv:2101.10109] [INSPIRE].
- [30] I. Watson, *Deep Learning for the HAWC Observatory*, *PoS ICRC2023* (2023) 927 [INSPIRE].
- [31] ICECUBE collaboration, *Observation of high-energy neutrinos from the Galactic plane*, *Science* **380** (2023) 1338 [arXiv:2307.04427] [INSPIRE].
- [32] PIERRE AUGER collaboration, *Inference of the Mass Composition of Cosmic Rays with Energies from  $10^{18.5}$  to  $10^{20}$  eV Using the Pierre Auger Observatory and Deep Learning*, *Phys. Rev. Lett.* **134** (2025) 021001 [arXiv:2406.06315] [INSPIRE].
- [33] PIERRE AUGER collaboration, *Measurement of the depth of maximum of air-shower profiles with energies between  $10^{18.5}$  and  $10^{20}$  eV using the surface detector of the Pierre Auger Observatory and deep learning*, *Phys. Rev. D* **111** (2025) 022003 [arXiv:2406.06319] [INSPIRE].
- [34] M.M. Bronstein et al., *Geometric Deep Learning: Going beyond Euclidean data*, *IEEE Signal Process. Mag.* **34** (2017) 18 [arXiv:1611.08097] [INSPIRE].
- [35] D. Heck et al., *CORSIKA: A Monte Carlo Code to Simulate Extensive Air Showers*, in2p3-00005094, version 1 (1999), <https://hal.in2p3.fr/in2p3-00005094>.
- [36] GEANT4 collaboration, *GEANT4 — A Simulation Toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250 [INSPIRE].

- [37] A.U. Abeysekara et al., *Observation of the Crab Nebula with the HAWC Gamma-Ray Observatory*, *Astrophys. J.* **843** (2017) 39 [[arXiv:1701.01778](#)] [[INSPIRE](#)].
- [38] H. Schoorlemmer, R. Conceição and A.J. Smith, *Simulating the performance of the Southern Wide-view Gamma-ray Observatory*, *PoS ICRC2021* (2021) 903 [[INSPIRE](#)].
- [39] R. Conceição, *The Southern Wide-field Gamma-ray Observatory*, *PoS ICRC2023* (2023) 963 [[arXiv:2309.04577](#)] [[INSPIRE](#)].
- [40] Y. Wang et al., *Dynamic Graph CNN for Learning on Point Clouds*, [arXiv:1801.07829](#) [[INSPIRE](#)].
- [41] H. Qu and L. Gouskos, *ParticleNet: Jet Tagging via Particle Clouds*, *Phys. Rev. D* **101** (2020) 056019 [[arXiv:1902.08570](#)] [[INSPIRE](#)].
- [42] R. Abbasi et al., *Graph Neural Networks for low-energy event classification & reconstruction in IceCube*, *2022 JINST* **17** P11003 [[arXiv:2209.03042](#)] [[INSPIRE](#)].
- [43] T. Bister et al., *Identification of patterns in cosmic-ray arrival directions using dynamic graph convolutional neural networks*, *Astropart. Phys.* **126** (2021) 102527 [[arXiv:2003.13038](#)] [[INSPIRE](#)].
- [44] J. Glombitza, V. Joshi, B. Bruno and S. Funk, *Application of graph networks to background rejection in Imaging Air Cherenkov Telescopes*, *JCAP* **11** (2023) 008 [[arXiv:2305.08674](#)] [[INSPIRE](#)].
- [45] M. Fey and J.E. Lenssen, *Fast Graph Representation Learning with PyTorch Geometric*, [arXiv:1903.02428](#) [[INSPIRE](#)].
- [46] D.P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, [arXiv:1412.6980](#) [[INSPIRE](#)].
- [47] N. Srivastava et al., *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*, *J. Mach. Learn. Res.* **15** (2014) 1929 [[INSPIRE](#)].
- [48] S. Ioffe and C. Szegedy, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, [arXiv:1502.03167](#) [[INSPIRE](#)].
- [49] K. He, X. Zhang, S. Ren and J. Sun, *Deep Residual Learning for Image Recognition*, [arXiv:1512.03385](#) [[DOI:10.1109/CVPR.2016.90](#)] [[INSPIRE](#)].
- [50] T. Ullrich and Z. Xu, *Treatment of Errors in Efficiency Calculations*, [physics/0701199](#).
- [51] F. Leitzl, V. Joshi and S. Funk, *Status of the SWGO air shower reconstruction using a template-based likelihood method*, *PoS ICRC2023* (2023) 593 [[INSPIRE](#)].
- [52] A. Albert et al., *Performance of the HAWC Observatory and TeV Gamma-Ray Measurements of the Crab Nebula with Improved Extensive Air Shower Reconstruction Algorithms*, *Astrophys. J.* **972** (2024) 144 [[arXiv:2405.06050](#)] [[INSPIRE](#)].
- [53] Z. Hampel-Arias and S. Westerhoff, *Gamma Hadron Separation using Pairwise Compactness Method with HAWC*, *PoS ICRC2015* (2016) 1001 [[arXiv:1508.04047](#)] [[INSPIRE](#)].
- [54] R. Conceição, L. Gibilisco, M. Pimenta and B. Tomé, *Gamma/hadron discrimination at high energies through the azimuthal fluctuations of air shower particle distributions at the ground*, *JCAP* **10** (2022) 086 [[arXiv:2204.12337](#)] [[INSPIRE](#)].
- [55] R. Alfaro et al., *Gamma/hadron separation with the HAWC observatory*, *Nucl. Instrum. Meth. A* **1039** (2022) 166984 [[arXiv:2205.12188](#)] [[INSPIRE](#)].
- [56] A. Albert et al., *Science Case for a Wide Field-of-View Very-High-Energy Gamma-Ray Observatory in the Southern Hemisphere*, [arXiv:1902.08429](#) [[INSPIRE](#)].