PROCEEDINGS OF SCIENCE

Deep Learning Methods for Gamma/Hadron Separation in SWGO

M. Schneider,[a,*] **M. Pirke,**[a] **F. Leitl,**[a] **J. Glombitza**[a] **and C. van Eldik**[a] **for the SWGO Collaboration**

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

E-mail: martin.friedrich.schneider@fau.de, markus.pirke@fau.de

The Southern Wide-field Gamma-ray Observatory (SWGO) is a planned water Cherenkov-based observatory to be located in Pampa La Bola, Chile, providing continuous, wide-field observations of the gamma-ray sky. SWGO will provide a unique view of the wide southern hemisphere gamma-ray sky, complementing other very-high-energy observatories such as HAWC, CTAO, and LHAASO. A key challenge in ground-based gamma-ray astronomy is an effective gamma/hadron separation to suppress the dominant cosmic-ray background. In this contribution, we will present ongoing studies on advanced classification techniques for SWGO, specifically exploring deep learning approaches based on Graph Neural Networks and Transformer architectures. These methods, currently tested through simulations, offer promising advancements in gamma/hadron separation and event reconstruction.

39th International Cosmic Ray Conference (ICRC2025)
15–24 July 2025
Geneva, Switzerland

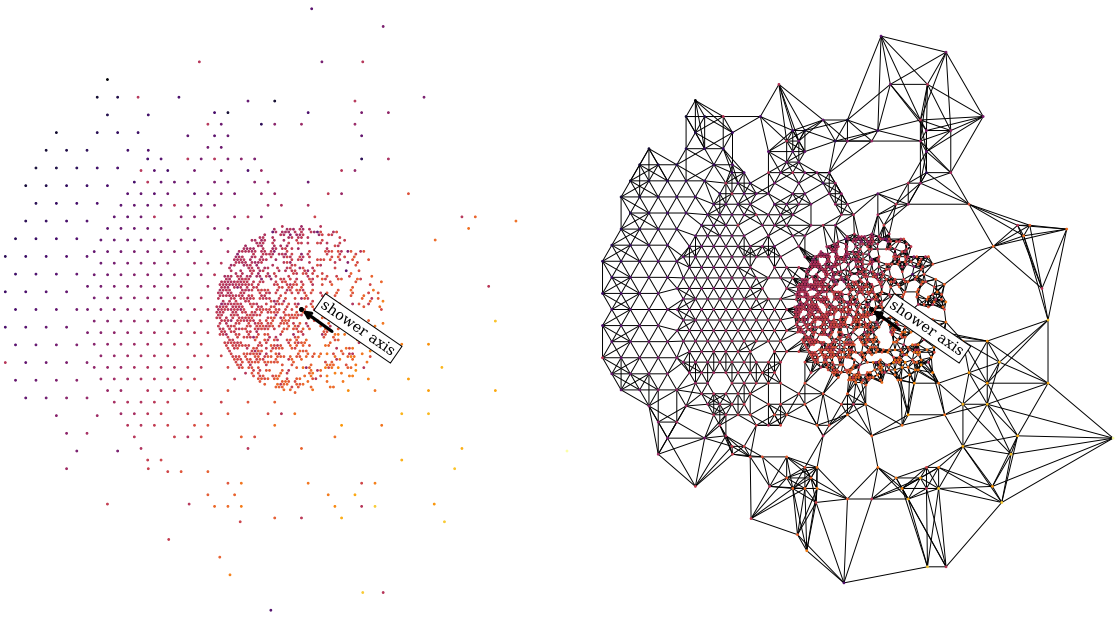
*Speaker

PoS(ICRC2025)836

1. Introduction

In recent years, ground-based instruments such as HAWC [1] and LHAASO [2] further enabled surveys of the gamma-ray sky as well as studies of the astronomy of cosmic phenomena at the highest energies. Following their success, the SWGO Collaboration aims to build a large field-of-view observatory in the southern hemisphere [3] to complement the future Cherenkov Telescope Array Observatory (CTAO) [4] and to follow up the H.E.S.S. [5] observations of the Galactic Plane at unprecedented gamma-ray energies. The sensitivity of these ground-based instruments strongly depends on their ability to suppress the hadronic background from air showers. Previous methods to distinguish cosmic-ray showers from gamma-ray induced showers used in IACTs and ground-based observatories heavily relied on combining high-level parameters through machine learning techniques [6–9]. The advent of Deep Neural Networks [10, 11] presented a new opportunity within physics [12] and astroparticle physics [13, 14] to improve upon previous methods, with various network architectures currently being explored.

In this proceeding, we present two deep learning-based methods for gamma/hadron separation using Graph Neural Networks (GNNs) [15] and the transformer architecture [16]. We compare their performance on a simulated test array currently considered for SWGO [3] to a traditional machine learning method.



(a) Point cloud **(b)** Graph

Figure 1: Shower footprint of a simulated gamma-ray event with a true energy of 118.9 TeV and a zenith angle of 34.9°. The color shows the network's perception of the normalized arrival times, ranging from bright colors indicating early hits to dark colors indicating late hits. Left: Point cloud of tanks. Each feature vector of a triggered tank represents the input to the transformer. Right: Graph representation after kNN clustering. The feature graph forms the input to the GNN.

2. Monte Carlo simulations and preprocessing

The Monte Carlo (MC) simulations for protons and gamma-rays used were produced for the SWGO site located in Pampa La Bola, Chile, 4.770 m above sea level. They cover an energy range from 31.6 GeV to 1 PeV following a spectral index of -2 and zenith angles up to 65°. Additional details about the array and its tanks can be found in ref. [17]. The simulations do not include cosmic-ray-induced noise. A simulated high-energy gamma-ray event is shown in Figure 1, where the left panel depicts the point cloud of detector tanks that recorded hits, with colors representing the network's perceived arrival times.

As a baseline, we make use of a standard machine learning approach using a multi-layer perceptron (MLP). Historically speaking gamma/hadron separation has been performed by exploiting high-level parameters that somewhat characterized the nature of atmospheric air showers. They are often correlated with the muon content of the shower, the energy of the primary particle, the lateral shower development or the uniformity of the particle distribution at the ground. Since no single parameter dominates in their gamma/hadron separation capabilities over the whole phase space multiple parameters are combined through cuts to improve performance. This process can be optimized by using machine learning to combine these parameters efficiently with small neural networks suchs as MLPs. The specific MLP used here utilizes high-level parameters such as PINCness [9], LDFchi2 [9], and LCm [18] as well as reconstruction quality observables for angle fitting and template-based energy reconstruction [19]. To train our network, we split the simulation data 40%/60% into a training and test dataset. For training and validation, we require a multiplicity threshold of at least 20 detector hits, while the final evaluation uses a more realistic threshold of 65 hits [3]. For our simple machine learning model we bin our data in reconstructed energy, core location and zenith angle [1] and apply additional selection cuts based on reconstruction quality. Then, for each of these bins, a small MLP consisting of two hidden layers with 12 neurons each is trained. The reconstruction quality cuts help to establish a baseline quality of the observables that the MLP relies on, while training an MLP for each bin provides flexibility for different event classes.

3. GNN architecture

In previous work, GNNs have already been applied to gamma-ray astronomy [20] and water-Cherenkov-based detector arrays [21] by modeling events as signal graphs, where the nodes consist of the triggered stations and the edges connect stations in local proximity. Here we follow the same approach, constructing the graph through k-nearest neighbor clustering (kNN) on the positions. This is illustrated in Figure 1 on the right panel, where we use $k = 6$ and additionally include self-connections which are not shown explicitly. The GNN only uses low-level event information of triggered tanks, such as tank positions, and quantities measured by the PMT, such as integrated charge and arrival times. For the double-layer tanks used in this work, we simply add additional features to our nodes to describe the lower cell of the tank, so the feature vector of a tank will read as $\mathbf{x}_i = (x, y, q_{\text{upper}}, t_{\text{upper}}, q_{\text{lower}}, t_{\text{lower}})^T$. Our network, see Figure 2, relies on EdgeConvolutions

[1] We use eight logarithmic reconstructed energy bins from $10^{1.5}$ GeV to $10^{5.5}$ GeV, zenith angle bins of 0-30°, 30-45° and 45-52° and core location of 0-156 m, 156-400 m and 400-560 m

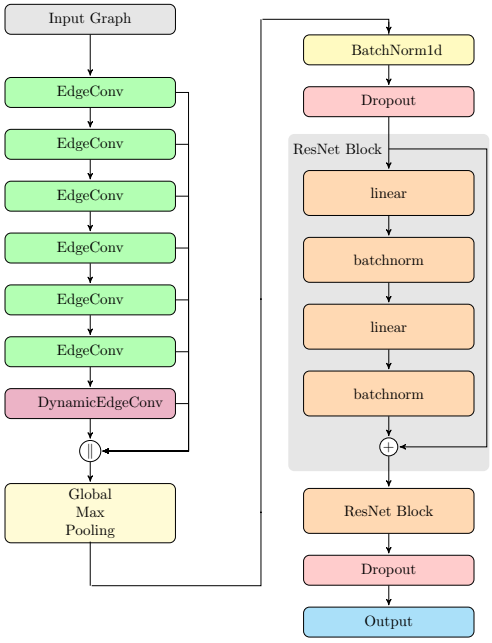


Figure 2: Sketch of the network structure of the GNN used [21]. It consists of multiple EdgeConvolution blocks, followed by a DynamicEdgeConvolution layer. The outputs of each layer are concatenated and fed (after pooling) through some regularization layers. Finally, ResNet blocks with residual connections [23] are utilized to extract features from the nodes of the feature graphs, which are then employed for the classification.

and DynamicEdgeConvolutions [22] to extract information from the feature graphs. Since our graphs were generated through kNN clustering, the EdgeConvolution will have a strong prior on local correlations. Through iterative application of EdgeConvolutions, the network will be able to access global information similar to the concept of the receptive field of view [12] in regular Convolutional Neural Networks (CNNs). We also make use of a DynamicEdgeConvolution which, through graph construction, finds the k-closest nodes in the feature space, before concatenating and pooling along the node dimension. Finally, a series of fully-connected layers is applied to the now node-independent output, ending with the binary classification output. For further details, we refer to our previous analysis [21] of this type of network.

4. Transformer architecture

Another approach to exploit low-level data is to use the well-established transformer architecture [16]. Initially applied in the context of natural language processing, the transformer architecture was quickly adapted for use in computer vision and image recognition [24]. In this work, we have adapted the vision transformer (ViT) architecture [24] to air shower footprints, enabling a direct comparison with the GNN architecture and extending over [25, 26]. Instead of using latent attention our approach uses full self-attention exclusively on the triggered elements. We uses the same feature vectors as for the GNN but processes them by feeding them into the network as a sequence. The first layer in our architecture, shown in Figure 3, is the embedding layer, where the feature vectors are linearly transformed into a higher-dimensional space. This is followed up by a

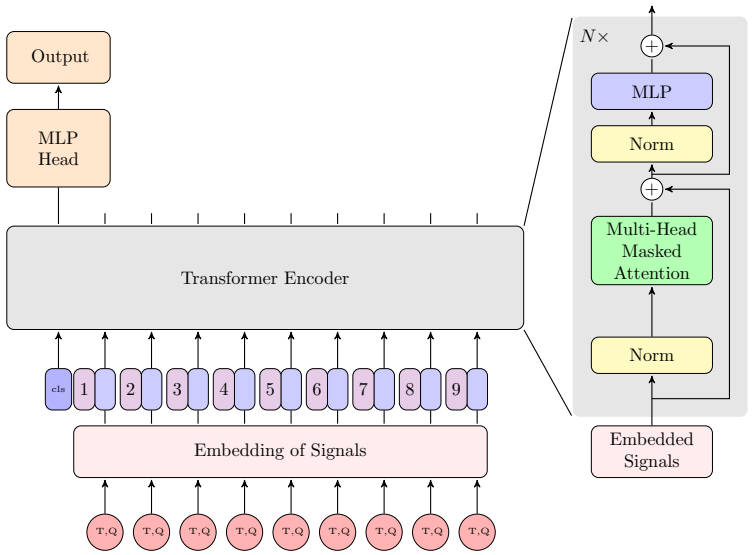


Figure 3: Sketch of the general transformer architecture for gamma/hadron separation. A sequence containing charge and time information is linearly embedded, and positional embeddings are added. The resulting sequence is fed with the addition of a CLS token to the transformer block. Ultimately, only the first token is used in the final layer.

learnable positional encoding, which maps the position vectors into the embedding vector space via a linear transformation. We add a classification (CLS) token at the start of the sequence allowing it to gather all relevant global correlations after multiple attention operations. This enables sequence length independent information extraction. This forms the input to the transformer block, which only differs from the ViT [24] in the masking operation.

Event sequences, i.e. the number of triggered tanks, can vary greatly in length, a characteristic amplified by the simulation's spectral index of -2, which produces a few very long events alongside many short ones. To create uniform batches for processing, we pad shorter sequences with zeros. These are masked out during self-attention but processing additional token creates significant computational overhead nonetheless. To minimize this inefficiency, we group events into "buckets" of similar lengths as in [27], thereby reducing the amount of padding needed.

5. Gamma/Hadron separation performance for one of the current test arrays

In the following, we examine the background suppression power of both networks and compare their performance for the task of gamma/hadron separation with that of the standard MLP. The background suppression, i.e., the false positive rate, is defined as the fraction of simulated protons that are incorrectly classified as gamma-rays.

To compare these methods, we look at both the energy and zenith angle dependence of the background suppression and fix the surviving fractions of gamma-rays to 80%. Statistical uncertainties on the background suppression are estimated following ref. [28]. For bins that do not contain misclassified proton events, an upper limit is estimated by assuming a single background

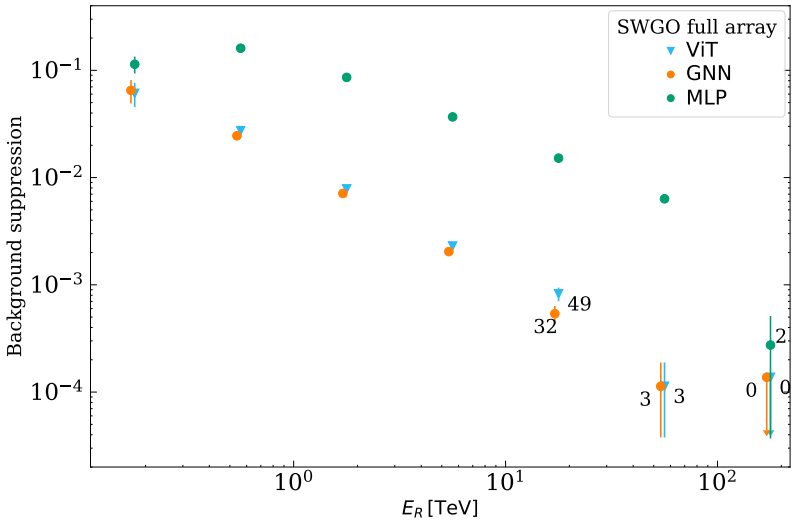


Figure 4: Comparison of different gamma/hadron separation methods on the SWGO full array. Background suppression is plotted against reconstructed energy with the vision transformer shown in light blue, the GNN shown in orange, and the MLP shown in green. The orange datapoints are slightly shifted to the left for visibility, and we annotate the total number of misclassified protons for the low statistics data points.

event and indicated by a down-facing arrow. Comparing the energy dependence of the background suppression for both models, we find that the deep learning methods outperform the standard machine learning approach of the MLP by a significant margin, see Figure 4. Across most of the phase space we observe an improvment in rejection power of about one order of magnitude. Out of the two deep learning methods, the GNN is slightly outperforming the transformer network in the medium energy regime. In this region, about 10% more background events will be left in the dataset using the transformer network compared to the GNN. In the last two energy bins, the performance of both models is identical on this dataset. Looking at the zenith angle dependence of the background suppression, we observe a similar pattern. The deep learning methods perform strictly better than the classical machine learning approach. We find that the GNN offers marginally better performance than the transformer, with approximately 15% more background events remaining when using it, see 5. Across the zenith angle range, the performance is relatively consistent, equalizing again for the low-statistic at high zenith angles.

6. Summary and outlook

In this proceeding, we compared two deep-learning-based networks for gamma/hadron separation in the context of SWGO, a next-generation gamma-ray observatory currently in the R&D phase in the southern hemisphere. Both the transformer networks and the graph neural network have been successfully applied to air-shower footprints detected by a simulated water-Cherenkov detector array, utilizing the charge and timing information of triggered PMTs. Confirming previous results [21], the investigated deep learning methods have already been proven to surpass high-level observables for Water-Cherenkov detectors, we found that these novel methods strongly outperform standard machine learning techniques based on hand-designed observables. Between the two deep

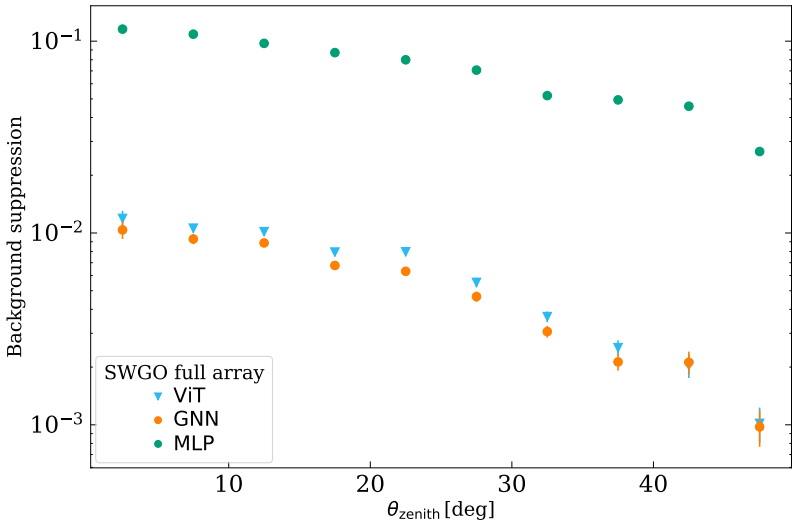


Figure 5: Comparison of different gamma/hadron separation methods on the SWGO full array. Background suppression is plotted against reconstructed zenith angle with the vision transformer shown in light blue, the GNN shown in orange, and the MLP shown in green.

learning architectures we tested the GNN currently slightly outperformed the transformer architecture on gamma/hadron separation on our dataset. The GNN's superior performance may stem from its inductive bias for modeling local graph connectivity, which is crucial for capturing the intrinsic spatial relationships of particle showers, unlike the transformer's more general attention mechanism. By combining both architectures one could utilize the GNN's strength in modeling local graph structures together with the Transformer's capability to capture long-range dependencies and global relationships within the data. The success of deep-learning-based strategies on simulated data highlights the need to explore their capabilities and test their robustness on gamma-ray observation data.

Acknowledgments

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.

References

[1] A. U. Abeysekara *et al.* (HAWC Collaboration), Nucl. Instrum. Meth. A **1052** (2023) .
[2] F. Aharonian *et al.* (LHAASO collaboration), [ArXiv:2101.03508], 2021.
[3] P. Abreu *et al.* (SWGO Collaboration), [ArXiv:2506.01786], 2025.
[4] B. S. Acharya *et al.* (CTA Collaboration), Astropart. Phys. **43** (2013) 3-8.
[5] F. Aharonian *et al.* (H.E.S.S. Collaboration), A&A **457** (2006) 899-915.

[6] S. Ohm, C. van Eldik, K. Egberts, Astropart. Phys. **31** (2009) 383-391.
[7] M. Krause *et al.*, Astropart. Phys. **89** (2017) 1-9.
[8] J. Albert *et al.* (MAGIC Collaboration), Nucl. Instrum. Meth. A **588** (2008) 424-432.
[9] R. Alfaro *et al.* (HAWC Collaboration), Nucl. Instrum. Meth. A **1039** (2022) .
[10] Y. LeCun *et al.*, Nature **521** (2015) 436-444.
[11] I. Goodfellow *et al.*, MIT Press, 2016.
[12] M. Erdmann *et al.*, World Scientific, 2021.
[13] R. Abbasi *et al.* (IceCube Collaboration), Science **380** (2023) 1338-1343.
[14] A. Abdul Halim *et al.* (Pierre Auger Collaboration), Phys. Rev. Lett. **134** (2025) .
[15] M. M. Bronstein *et al.*, IEEE Signal Process. Mag. **34** (2017) 18-42.
[16] A. Vaswani *et al.*, Adv. Neural Inf. Process. Syst. **30** (2017) .
[17] J. Glombitza *et al.* (SWGO Collaboration), PoS ICRC2025 (2025)
[18] R. Conceição *et al.*, J. Cosmol. Astropart. Phys. **2022** (2022) 86.
[19] F. Leitl *et al.* (SWGO Collaboration), PoS ICRC2023 (2023) 593.
[20] J. Glombitza *et al.*, JCAP **2023** (2023) 8.
[21] J. Glombitza *et al.*, J. Cosmol. Astropart. Phys. **2025** (2025) 66.
[22] Y. Wang *et al.*, [ArXiv:1801.07829], 2018.
[23] K. He *et al.*, [ArXiv:1512.03385], 2015.
[24] A. Dosovitskiy *et al.*, [ArXiv:2010.11929], 2021.
[25] I. Watson (HAWC Collaboration), PoS ICRC2023 (2023) 927.
[26] I. Watson *et al.* (SWGO Collaboration), PoS ICRC2025 (2025)
[27] H. Bukhari *et al.*, [ArXiv:2310.15674], 2023.
[28] T. Ullrich and Z. Xu, arXiv:0701199v1, 2012.

PoS(ICRC2025)836