Towards Machine-Learning Particle Flow with the ATLAS Detector at the LHC

Luca Clissa[1,2,*] on behalf of ATLAS Collaboration

[1]University of Bologna, Department of Physics and Astronomy "Augusto Righi"
[2]National Institute for Nuclear Physics (INFN), Bologna

Abstract. Particle flow reconstruction at colliders combines various detector subsystems (typically the calorimeter and tracker) to provide a combined event interpretation that utilizes the strength of each detector. The accurate association of redundant measurements of the same particle between detectors is the key challenge in this technique. This contribution describes recent progress in the ATLAS experiment towards utilizing machine-learning to improve particle flow in the ATLAS detector at the LHC. In particular, point-cloud techniques are utilized to associate measurements from the same particle, leading to reduced confusion compared to baseline techniques. Next steps towards further testing and implementation are also discussed.

ATL-SOFT-PROC-2025-018
23 January 2025

1 Introduction

Particle flow reconstruction represents a fundamental challenge in experimental particle physics, encompassing both the identification of particle species and the precise calibration of their energies from detector electronic readouts. This reconstruction process comprises three essential steps: the aggregation of tracker hits into particle trajectories, the clustering of calorimeter cell energies corresponding to individual particle interactions, and the association between these reconstructed tracks and calorimeter clusters.

In scenarios where particles traverse the detector in isolation, producing well-separated tracker hits and distinct calorimeter energy deposits, the reconstruction process is relatively straightforward. However, the task becomes significantly more complex in environments with multiple nearby particle trajectories and overlapping calorimeter deposits, particularly in high-luminosity hadron collider experiments. These challenging scenarios necessitate sophisticated approaches to disentangle the contributions of individual particles and accurately reconstruct their properties.

To address these challenges, the ATLAS experiment [1] has developed a dedicated particle flow algorithm that leverages the complementary information provided by the tracking and calorimeter systems. This algorithm aims at optimizing both particle identification and energy reconstruction by combining the superior angular and momentum resolution of the tracking system with the calorimeter's energy measurements. While effective, there remains space for improvement through the application of modern machine learning techniques.

*e-mail: luca.clissa2@unibo.it

2 Background and motivation

The current ATLAS particle flow algorithm [2] employs an iterative approach to particle reconstruction. It processes tracks in descending order of transverse momentum (p_T), associating each track with the nearest topologically-connected calorimeter cluster (topo-cluster) based on their angular separation ΔR[1]. For each track-cluster pair, the algorithm computes the expected calorimeter energy deposit from the track's momentum and compares it with the measured cluster energy. In cases of significant discrepancy, the algorithm searches for additional topo-clusters to account for the full particle energy. Once the association achieves satisfactory agreement, the algorithm subtracts the expected energy from the calorimeter cells, retaining residual energy deposits only if they exceed expected fluctuations. Significant remnants are then reconsidered in subsequent iterations, ensuring that energy not clearly associated with the current track can be attributed to other particles in the event.

While this approach effectively combines tracking and calorimetric information, it faces challenges in complex environments where multiple particles produce overlapping energy deposits. These limitations have motivated extensive research into improving reconstruction performance. A particularly promising direction involves machine learning solutions, which span a broad spectrum of complexity. At one end, ambitious end-to-end architectures attempt to handle the entire reconstruction pipeline [3], learning to simultaneously perform track reconstruction, calorimeter clustering, and energy calibration. At the other end, focused task-based solutions address specific reconstruction components independently, allowing for targeted optimizations while maintaining modularity with existing reconstruction frameworks. Among task-based approaches, image-based deep learning methods have shown particular potential. For instance, a recent study [4] demonstrated significant improvements over traditional calibration strategies in both particle identification and energy calibration. However, representing detector data as images presents two significant limitations. First, it introduces inherent inefficiencies in representing naturally sparse detector data. Second, and more critically, the image-based framework does not provide a natural mechanism for incorporating tracking information alongside calorimeter data.

To address these limitations, we propose a point cloud data representation for particle identification and energy calibration. Point clouds offer several advantages: they provide a more natural encoding of detector data, enable efficient processing of sparse detector hits, and facilitate integration of heterogeneous detector information. This work focuses on developing and evaluating deep learning methods specifically designed for point cloud data, comparing their performance with both current reconstruction algorithms and image-based approaches.

3 Methods

Traditional image-based representations of detector data map individual detector electronic units to pixels in a two-dimensional image, with their readout values encoded as pixel intensities. For instance, calorimeter cells are represented as pixels with energy measurements determining the intensity level (see Figure 1a).

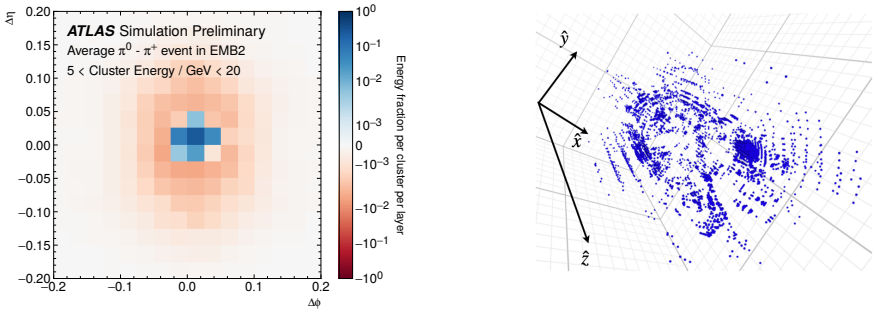
This representation presents three significant challenges. First, detector components typically exhibit varying spatial granularity, which is difficult to capture in a fixed-pixel grid where all elements must be uniformly sized and spaced. Second, particle detector data is

[1] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upwards. Polar coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln\tan(\theta/2)$ and is equal to the rapidity $y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right)$ in the relativistic limit. Angular distance is measured in units of $\Delta R \equiv \sqrt{(\Delta y)^2 + (\Delta\phi)^2}$.

inherently sparse, with only a small fraction of detector elements registering hits in any given event. The image-based approach necessitates maintaining the full detector grid, resulting in computationally inefficient representations dominated by empty pixels representing zero-suppressed cells. Third, and perhaps most critically, the framework lacks a natural mechanism for incorporating tracking information alongside calorimeter data. Indeed, tracker and calorimeter components have distinct geometries and granularities, which exacerbates the spatial representation problem. Hence, tracking information exists in a different geometric space than calorimeter data, making their reconciliation in a 2D image representation particularly challenging. Any attempt to combine these heterogeneous detector signals in a unified image framework requires artificial compromises that can lose important spatial correlations.

Point cloud representations address these limitations by modeling detector signals as points in a multi-dimensional space (3D or more). Each point corresponds to a hit detector element, characterized by its spatial coordinates, possibly augmented with additional features (dimensions) such as energy deposits, timing information, hit confidence metrics, and more (see Figure 1b). This approach offers several advantages: it naturally accommodates varying detector granularity, efficiently represents sparse data by including only relevant units for each event, and provides a unified framework for incorporating heterogeneous detector information. In particular, tracking and calorimeter data can be represented in their natural geometric forms while maintaining their spatial relationships in a common coordinate system.

Therefore, the point cloud representation is more suited to capture the genuine three-dimensional structure of particle interactions in the detector while maintaining the sparsity of the data. Also, it is possible to further augment this representation by including more features as additional dimensions. This provides a powerful and more natural setting for developing machine learning algorithms that can effectively process both tracking and calorimeter information.



(a) Image-based (calorimeter only) [4]

(b) Point cloud (whole detector) [5]

Figure 1: Examples of different detector data according to image-based and point cloud representations. The image-based approach (left) maps detector elements to a fixed grid, while the point cloud representation (right) preserves the natural geometry of particle interactions.

3.1 Experimental setup

To establish a baseline for evaluating point cloud deep learning methods, we begin with a simplified Monte Carlo simulation dataset. This approach allows us to validate the methods in

a controlled environment before addressing more complex scenarios such as multijet events. The dataset comprises approximately 20 million single pion events, including both charged ($\pi^{\pm}$) and neutral (π^0) pions, simulated using the full ATLAS detector simulation based on Geant4 [6]. Pions are generated with uniform distributions in azimuthal angle ϕ, pseudo-rapidity ($|\eta| < 3$), and the logarithm of true pion energy ranging from 0.2 to 2,000 GeV.

For training and evaluation, we select approximately 4 million events (3.5M for training and 500k for validation) meeting quality criteria and having exactly one track per event. These events are roughly equally distributed between charged and neutral pions. This selection ensures a clean topology for initial method development while maintaining sufficient statistics for robust training of deep learning models.

Starting from these data, we compare point cloud versus image-based approaches on two distinct learning tasks: *i)* particle identification and *ii)* energy calibration. For the **identification task**, we implement a *binary classification* to distinguish between neutral and charged pions using *only calorimeter information*, processing individual topo-clusters independently. The **energy calibration task**, instead, is framed as a *regression problem with two distinct configurations*. In the first configuration, the models predict calibrated energy using only calorimeter information from single topo-clusters. The second configuration incorporates both calorimeter and tracking information, considering one track and all associated topo-clusters within a ΔR cone. This dual approach to energy calibration allows us to evaluate the potential benefits of including tracking information while maintaining a controlled comparison with calorimeter-only methods.

Given the above settings, we evaluate several deep learning architectures designed to process either point cloud or image-based representations of detector data. We focus here on the most successful architectures studied in Ref. [5] (see also Ref. [4] for further details).

The Deep Sets model [7] provides a theoretical foundation for processing unordered collections of detector signals. It implements permutation-invariant functions through a per-point mapping Φ to a latent space, followed by a symmetric aggregation function. Each point is characterized by features including cell energy, sampling layer, and angular positions relative to the cluster center.

The Graph Neural Network (GNN) [8] extends this framework by explicitly modeling relationships among detector elements. The architecture consists of four GNN blocks using multi-layer perceptrons and permutation-invariant aggregation functions. Calorimeter cells are represented as nodes connected by edges based on their geometric proximity, with features encoding energy deposits and spatial information.

For comparison with previous approaches, we implement a Convolutional Neural Network (CNN) that processes calorimeter information as image data. The CNN architecture handles the varying granularity of different calorimeter layers by processing the finely-segmented electromagnetic layer separately, while combining other layers with similar granularity. Starting from these representations, the network applies 3 convolutional blocks followed by fully connected network layers.

Additionally, a baseline Dense Neural Network (DNN) with three hidden layers serves as a reference model for the energy regression task. The DNN processes high-level features derived from both tracking and calorimeter information, though without access to the detailed cell-level information available to the point cloud models.

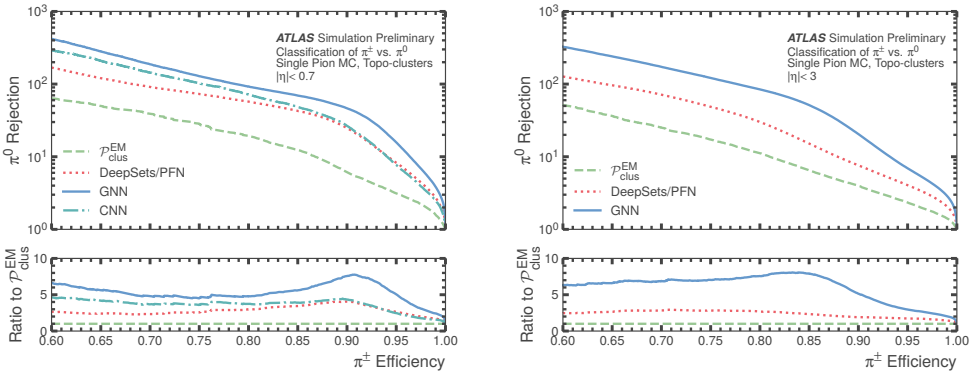
4 Results

We evaluate our methods against established baselines for both classification and regression tasks. For classification, we compare against the standard ATLAS electromagnetic cluster

classification probability ($P_{\text{clus}}^{\text{EM}}$). For regression, we benchmark against both the electromagnetic (EM) scale and local cell weighting (LCW) calibrations. Detailed descriptions of these baseline methods can be found in [5].

4.1 Particle identification

The classification performance is evaluated comparing π^0 rejection versus $\pi^\pm$ efficiency. Results are analyzed in two pseudo-rapidity ranges: the central barrel region ($|\eta| < 0.7$) and the extended range ($|\eta| < 3$). As shown in Figure 2, all machine learning approaches significantly outperform the $P_{\text{clus}}^{\text{EM}}$ baseline. At a fixed $\pi^\pm$ efficiency of 90%, the improvement in π^0 rejection ranges from a factor of 2 to 8 (Table 1). The GNN architecture demonstrates particularly strong performance, achieving the highest rejection across all efficiencies in both η ranges.



(a) $|\eta| < 0.7$ (b) $|\eta| < 3$

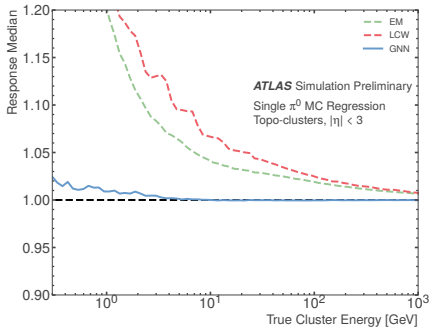
Figure 2: Comparison of π^0 rejection versus $\pi^\pm$ efficiency for different classification models. Results are shown for the central barrel ($|\eta| < 0.7$, left) and the extended range ($|\eta| < 3$, right). The lower panels show the ratio to the $P_{\text{clus}}^{\text{EM}}$ baseline performance [5].

Table 1: Neutral pion rejection at fixed charged pion efficiency of 90% for various classification models. The CNN was only trained on the central barrel region.

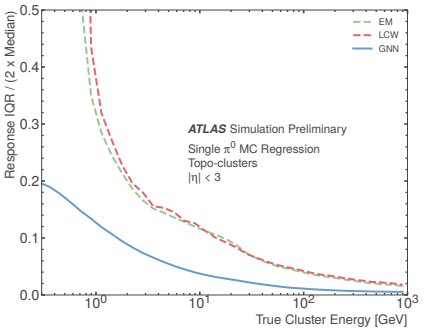
Model	Rejection @ 90% Efficiency	
	$\|\eta\| < 0.7$	$\|\eta\| < 3$
CNN	26.584	-
GNN	46.419	20.500
Deep Sets	24.814	7.608
$P_{\text{clus}}^{\text{EM}}$	6.123	3.977

4.2 Energy regression

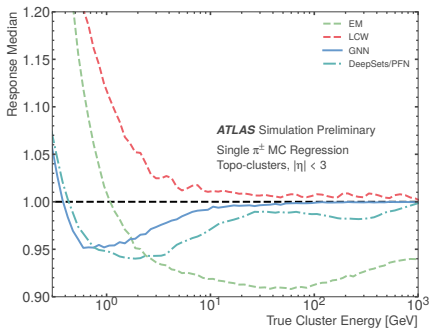
For the regression task, we evaluate performance using the energy response ratio $R = \frac{E_{\text{predicted}}}{E_{\text{true}}}$. Two key metrics are considered: the median response, which indicates calibration accuracy,



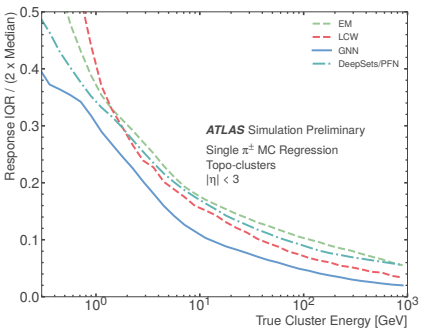
(a) π^0 Median Energy Response



(b) π^0 Interquartile Range (IQR)



(c) $\pi^{\pm}$ Median Energy Response



(d) $\pi^{\pm}$ Interquartile Range (IQR)

Figure 3: Energy response versus true cluster energy using only calorimeter data as input [5]. Median response (left) and normalized energy resolution, IQR, (right) are reported for π^0 (top) and $\pi^{\pm}$ (bottom).

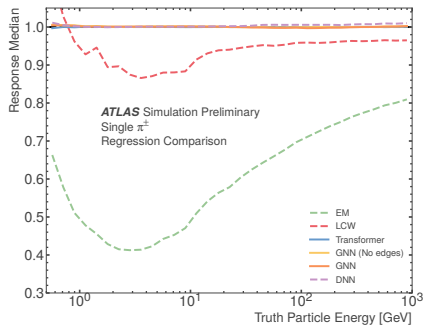
and a measure of energy resolution defined by $IQR = R_{\text{median}} \pm \sigma_{\text{R}}$. This is intended to capture the central 68% of the energy response distribution, and is plotted as normalized by twice the median to get rid of scale effects (Figures 3b and 3d).

Using **only calorimeter information**, the GNN shows significant improvement over both EM and LCW baselines across the entire energy spectrum. The median response stays closer to unity, while the normalized IQR shows improved resolution. The Deep Sets approach also outperforms the baselines, particularly for charged pions at low energies. These improvements effectively address long-standing calibration challenges at extreme energy values [4].

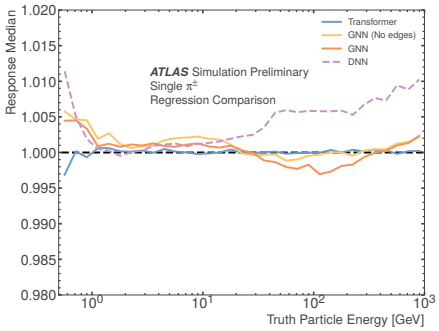
Including **tracking information** further enhances calibration performance (Figure 4). All point cloud methods demonstrate improved accuracy compared to both the baseline calibrations and the image-based DNN approach. This advantage becomes particularly pronounced at higher energies, where the point cloud methods maintain better energy resolution.

5 Conclusions and outlook

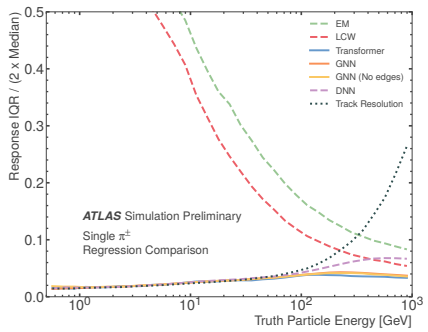
We have presented a review of deep learning approaches for particle identification and energy calibration in the ATLAS detector using point cloud representations. These techniques



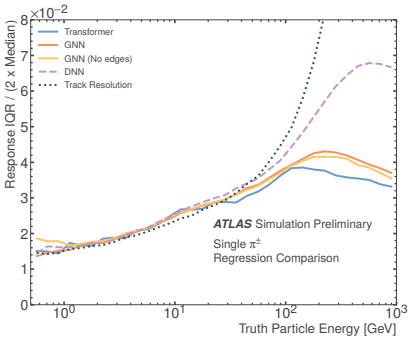
(a) With EM and LCW Baselines



(b) Without EM and LCW Baselines



(c) With EM and LCW Baselines



(d) Without EM and LCW Baselines

Figure 4: Energy response versus true particle energy using both track and calorimeter data as input [5]. Point-cloud techniques are compared against standard EM and LCW baselines. Median response (top) and normalized energy resolution, IQR, (bottom) are reported with (left) and without EM and LCW baselines (right).

demonstrate significant improvements over both traditional calibration techniques and previous image-based machine learning approaches.

The point cloud methods achieve up to eight times better neutral pion rejection at fixed charged pion efficiency compared to standard techniques, while maintaining superior energy response and resolution across a wide range of energies.

Importantly, incorporating tracking information alongside calorimeter data yields further improvements in energy calibration performance, especially at higher energies.

Building on these promising results, we are extending our approach to address the segmentation step of the particle flow reconstruction. This extension focuses on the challenging task of cell-to-track matching, where the goal is to identify calorimeter cells whose energy deposits primarily originate from a specific particle track. The method processes events by considering one "focus track" at a time, with all calorimeter cells within a specified ΔR cone, employing the specialized PointNet [9] architecture for cell assignment.

Initial results in simple topologies, such as ρ or Δ decay scenarios, where events typically contain a single track, demonstrate promising performance consistent with our previous findings. However, the generalization to more complex environments, particularly dijet events

with multiple overlapping tracks and calorimeter deposits, presents additional challenges that require further investigation. These challenges highlight the importance of developing robust methods that can maintain performance across varying event complexities while preserving the advantages demonstrated in simpler topologies.

Future work will focus on improving the robustness of our methods in complex environments and exploring additional architectural innovations to better handle multiple overlapping particle signatures. The insights gained from this work provide valuable direction for the development of next-generation particle flow algorithms that can fully leverage the complementary strengths of tracking and calorimeter information.

Acknowledgements

This research was partially funded by the PNRR - M4C2 - Investimento 1.3, as part of the Partenariato Esteso PE00000013, titled "FAIR - Future Artificial Intelligence Research", specifically Spoke 8, "Pervasive AI", funded by the European Commission under the NextGenerationEU program.

References

[1] ATLAS Collaboration, The ATLAS Experiment at the CERN Large Hadron Collider, JINST **3**, S08003 (2008). 10.1088/1748-0221/3/08/S08003

[2] ATLAS Collaboration, Jet reconstruction and performance using particle flow with the ATLAS Detector, Eur. Phys. J. C **77**, 466 (2017), 1703.10485. 10.1140/epjc/s10052-017-5031-2

[3] F.A. Di Bello, E. Dreyer, S. Ganguly, E. Gross, L. Heinrich, A. Ivina, M. Kado, N. Kakati, L. Santi, J. Shlomi et al., Reconstructing particles in jets using set transformer and hypergraph prediction networks, The European Physical Journal C **83**, 596 (2023). 10.1140/epjc/s10052-023-11677-7

[4] ATLAS Collaboration, Deep Learning for Pion Identification and Energy Calibration with the ATLAS Detector, ATL-PHYS-PUB-2020-018 (2020), https://cds.cern.ch/record/2724632

[5] ATLAS Collaboration, Point Cloud Deep Learning Methods for Pion Reconstruction in the ATLAS Experiment, ATL-PHYS-PUB-2022-040 (2022), https://cds.cern.ch/record/2825379

[6] ATLAS Collaboration, Software and computing for Run 3 of the ATLAS experiment at the LHC (2024), 2404.06335.

[7] M. Zaheer, S. Kottur, S. Ravanbhakhsh, B. Póczos, R. Salakhutdinov, A.J. Smola, Deep Sets, in *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Curran Associates Inc., Red Hook, NY, USA, 2017), NIPS'17, p. 3394–3404, ISBN 9781510860964, 10.5555/3294996.3295098

[8] F. Scarselli, M. Gori, A.C. Tsoi, M. Hagenbuchner, G. Monfardini, The graph neural network model, IEEE Trans. Neural Networks **20**, 61 (2009). 10.1109/TNN.2008.2005605

[9] C.R. Qi, H. Su, K. Mo, L.J. Guibas, Pointnet: Deep learning on point sets for 3d classification and segmentation, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 77–85, 10.1109/CVPR.2017.16