Particle flow algorithm with Transformer-based algorithm for highly-granular calorimeters

Stefan Barbu – Ecole Polytechnique – France
Taikan Suehara, Tatsuki Murata – ICEPP – The University of Tokyo – Japan

Introduction

Particle flow algorithms (PFA) reconstruct individual particles by optimally combining information from tracking detectors and calorimeters to achieve precise four-momentum measurements. This task is challenging due to overlapping energy deposits in highly granular calorimeters. Traditional approaches using hand-crafted rules struggle with the complex, high-multiplicity environments and irregular geometries of modern collider detectors.

Graph Neural Networks (GNNs), particularly GravNet [1], address these challenges by representing detector hits as nodes in a learned graph structure. GravNet's gravity-inspired approach dynamically constructs edges based on learned spatial embeddings, enabling effective information propagation across irregular detector geometries and capturing complex spatial relationships inherent in particle shower development. As our baseline, we use the GravNet-based particle flow algorithm presented in [2], where the GravNet framework is applied to the dual task of track-cluster matching and energy regression for particle flow reconstruction.

Motivation & Objective

This work explores the integration of **transformer**-based **self-attention** mechanisms into the GravNet architecture to address these limitations. Self-attention enables each detector hit to directly attend to all other hits in an event, capturing global relationships that purely local message-passing might overlook.
Our objectives are to develop a hybrid architecture that combines GravNet's efficient local feature extraction with transformers global reasoning capabilities, and analyze the computational trade-offs between accuracy gains and increased model complexity.

Proposed Architectures

We worked on two models, a Geometric Self-Attention GravNet (GSA-GravNet) and a Transformer based (Tr-model) model. Both models keep the same pre-processing and post-processing as the original GravNet model.

Proposed architectures

1. **GSA-GravNet**
   - Keep GravNet's learned graph structure
   - Change message passing layer by multi-head local self attention
   - Add distance between the vertices as a bias in the attention mechanism
   - Increase the size of the graphs from the 40 closest neighboors to 80 or 120

2. **Transformer model**
   - Change the GravNet Block by a Transformer Encoder block
   - Global self-attention between all events
   - Five times longer computation time

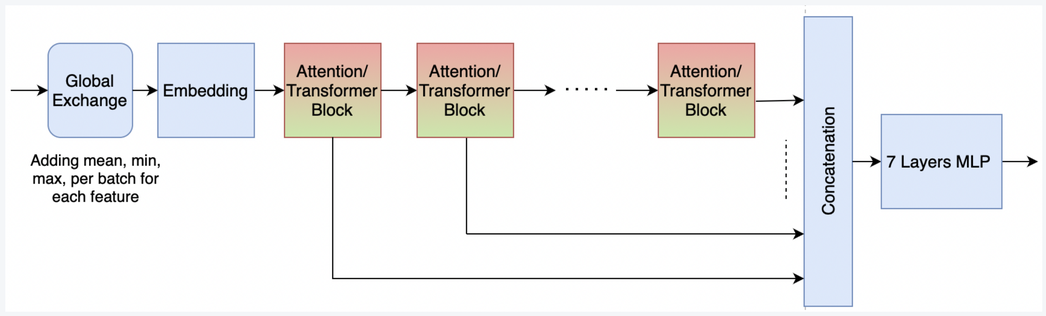

Figure 1: Illustration of the global architecture for both models.

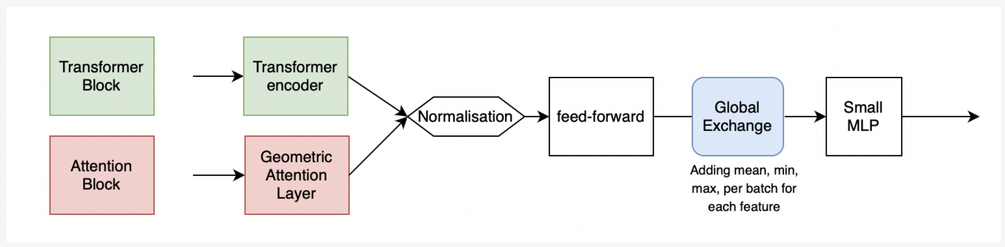

Figure 2: Transformer and Attention block.

- **Clustering Algorithm** For each event, our models predicts virtual coordinates in the learned space and a scalar, β used to define condensation points. The clustering algorithm uses these outputs to create clusters locally around the condensation points.
- **Object Condensation [3] loss function** performs both **clustering** and **energy regression** simultaneously. This loss function learns a condensation process where particles are attracted to cluster centers through a learned potential field. It consists of three components: an attractive potential (L_V) that handles the clustering dynamics, a β learning term (L_β) that optimizes the condensation parameters, and an energy regression term (L_E) that processes energy predictions.

$$\mathcal{L} = \lambda_\beta \mathbf{L}_\beta + \lambda_V \mathbf{L_V} + \lambda_E \mathbf{L_E}$$

Datasets

Data Generation
- ILD [4] full simulation with SiW-ECAL and AHCAL
- 10 Tau per event : 10 GeV, random directions emanating from the center of the detector
- Good mixture of decay products (pions, elections, muons, photons)
- ECAL: 5 x 5 mm^2, 30 layers/ HCAL: 30 x 30 mm^2, 48 layers

Dataset size
- 80k training/ 10k validation/ 10k test events

Results

Performance Metrics

To evaluate the clustering performance of the models, we use two metrics: efficiency and purity, which are defined as follows:

$$\text{Efficiency} = \frac{E_{match}}{E_{true}} \quad \text{and} \quad \text{Purity} = \frac{E_{\text{match}}}{E_{reco}}$$

with E_{match} being the correctly clustered energy, E_{true} the Monte Carlo truth ennergy and E_{reco} the reconstruced cluster energy.

Efficiency and Purity of cluster reconstruction in Tr-model

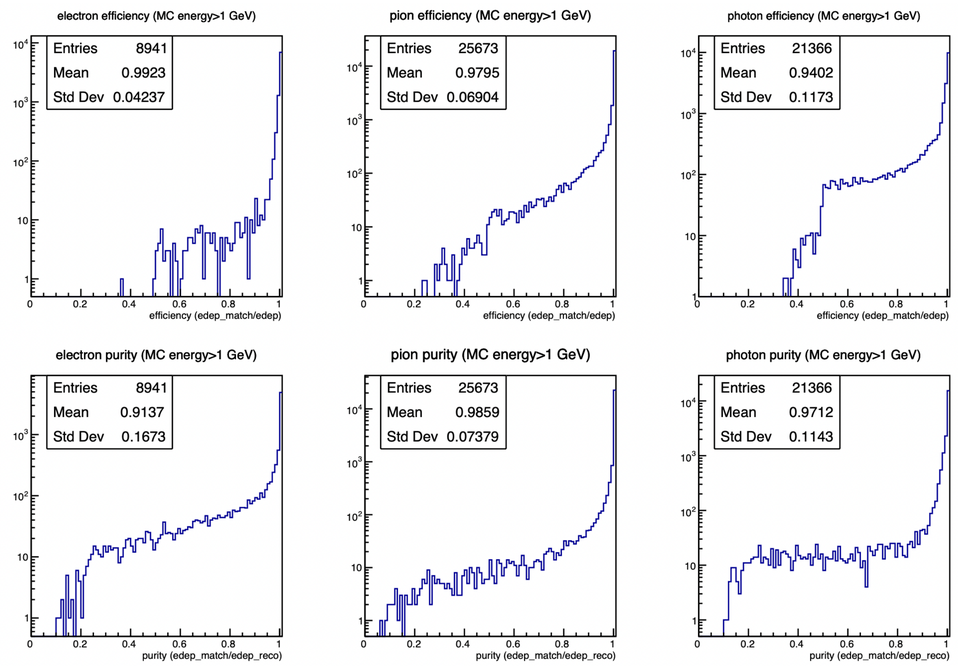

Figure 3: graph of efficency and purity in transformer model

Metric	Tr-model	GSA-GravNet	GravNet	Improvement (Tr-model-GravNet)
Number of parameters (M)	2	0.9	0.4	
Electron Efficiency/Purity	99.2 / 91.4	98.2 / 91.5	98.9 / 94.5	+0.3 / -3.1
Pion Efficiency/Purity	98.0 / 98.6	95.7 / 98.4	95.9 / 99.0	+2.1 / -0.4
Photon Efficiency/Purity	94.0 / 97.1	93.3 / 97.1	97.1 / 98.6	-3.1 / -1.5

Table 1: Efficiency/Purity Comparison.

Energy regression display from Tr-model

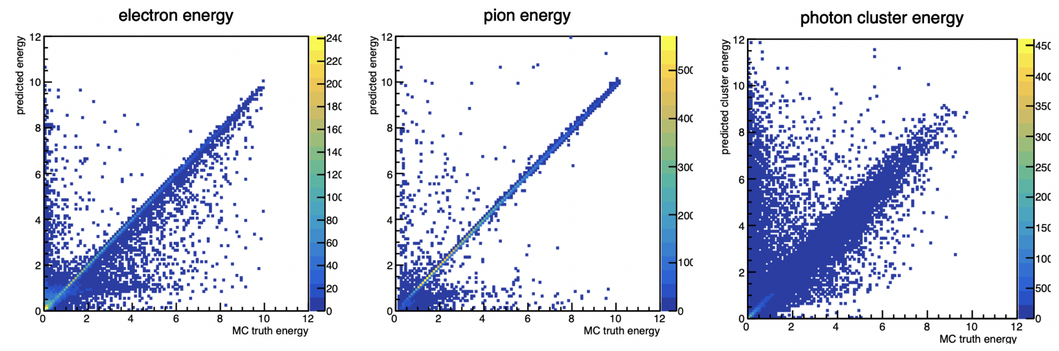

Figure 4: Graph of the energy predicted by the transformer model in function of Monte Carlo truth energy

Energy regression performances on photons

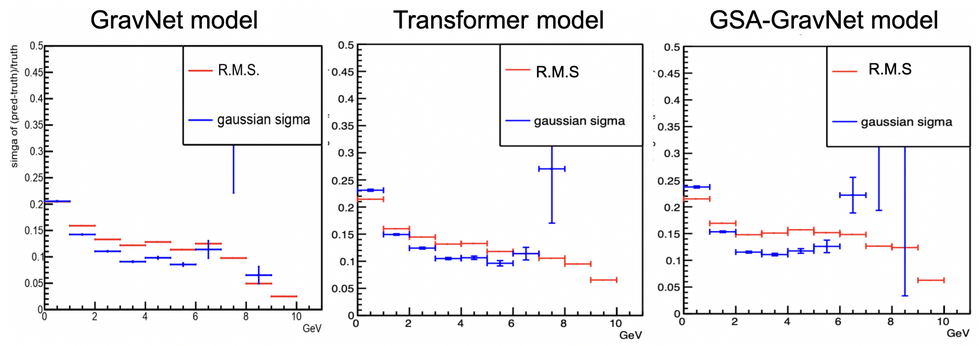

Figure 5: Gaussian sigma and Root Mean Square of energy regression on photons with the models from left to right: GravNet, Tr-model and GSA-GravNet

Conclusion & Perspective

The comparaison of the efficiency and purity between our models and GravNet shows:

- **under-performance of the GSA-GravNet**
- **improvement in the efficiency of pions and electrons in the Tr-model**

The energy regression is promising, especially with the Tr-model, who slightly under performs GravNet. It should be noted that the presented results have potential for improvement, given that the tuning was not as thorough as that performed for the GravNet model.

References

[1] Shah Rukh Qasim et al. "Learning representations of irregular particle-detector geometry with distance-weighted graph networks". In: *Eur. Phys. J. C* 79.7 (2019), p. 608.
[2] Taikan Suehara et al. "High Level Reconstruction with Deep Learning using ILD Full Simulation". In: *PoS* ICHEP2024 (2025). arXiv:2410.08772, p. 1019.
[3] CMS Collaboration. "Multi-particle reconstruction in the High Granularity Calorimeter using object condensation and graph neural networks". In: *EPJ Web Conf.* 251 (2021), p. 03072.
[4] ILD Concept Group. "International Large Detector: Interim Design Report". In: *arXiv:2003.01116* (2020).