PROCEEDINGS OF SCIENCE

Machine learning in online / offline calibration and reconstruction at the LHC

[1,2]Christian Sonnabend[*†]

[1]*CERN,* [2]*Ruprecht-Karls Universität, Heidelberg*

E-mail: christian.sonnabend@cern.ch

These proceedings report on the status and use of machine learning algorithms and especially neural networks in online and offline reconstruction and calibration for the four major experiments at CERN. The paper provides an overview and specific examples, ranging from raw data readout to final offline calibrations.

Large Hadron Collider Physics
3rd-7th of June, 2024
Boston, MA, USA

[*]Speaker
[†]On behalf of the ALICE, ATLAS, CMS and LHCb collaboration

PoS(LHCP2024)151

1. Introduction

Machine learning, a wide-spread topic in many fields of research and industry, has gained popularity with the advent of neural networks (NN) and modern hardware accelerators such as GPUs, FPGAs and ASICs. The experiments at the LHC continue to make heavy use of the industry leading frameworks and model architectures in reconstruction and calibration of the data taken during Run 3 of the LHC (2021 - present). These proceedings will outline some of the applications, following the typical flow of data from cluster-level to track-level applications and finally calibrations, tagging and triggering.

2. Cluster-level applications

Cluster-level applications utilize direct detector output without incorporating global track information. These algorithms primarily access local cluster or digit data, along with the detector's local position and geometry.

A first application of neural networks on a digit level is applied in a novel approach to clusterization of the TPC readout by the ALICE collaboration. With its fully GPU based online reconstruction in Run 3 [1], ALICE is ideally suited for the application of parallelizable architectures, such as neural networks. Clusterization is performed with a cluster rejection and a regression NN around measured charge maxima. Fully connected and 3D convolutional architectures show promising results in terms of physics performance for clusterization efficiency and fake-rate (fig. 2.1) compared to the current GPU clusterizer based on the cellular automaton approach. With its local 3D input (row, pad and time dimension), the output of the clusterizer can even be optimised in training using real data and tracks alongside Monte-Carlo generated. However 3D information is computationally expensive and not always available. An approach to detect anomalous events on 2D calorimeter images is provided by CMS with its calorimeter Image Convolution Anomaly Detection Algorithm (CICADA) [2]. It is trained to produce low mean square error outputs for zero bias events (which it has been trained on) and large outputs for anomalous events. Its trigger rate coincides remarkably well with other triggers deployed in CMS such as single-τ or single-jet triggers with an energy greater than 180 GeV (fig. 2.2).

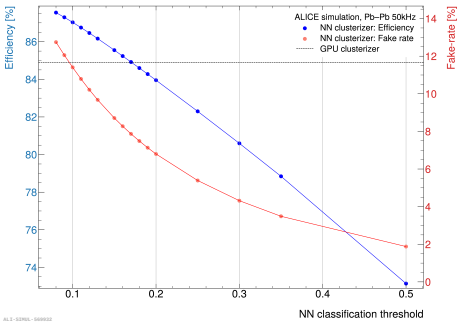


Figure 2.1: Reduction in fake-rate by the NN at same clusterization efficiency as the current GPU clusterizer (dashed line).

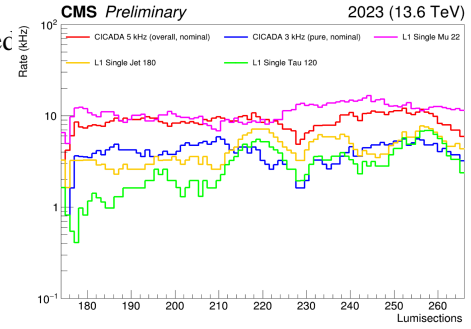


Figure 2.2: CICADA scores compared to other triggers deployed at CMS.

3. Track-level applications

Building on cluster-level data, tracking algorithms can provide final, calibrated tracks after the reconstruction is run. Several approaches are possible, including full NN-based track reconstruction, as investigated by LHCb for VELO-only and long tracks. It is based on a classification of edges and triplets (edge-edge connections), with edges corresponding to connections between pairs of hits [3]. The hits and edges can be represented as a graph and processed using Graph Neural Networks (GNNs), which provide good computational scalability as hit multiplicities increase. Compared to the default algorithm used in the Allen software framework, the GNN approach brings a similar efficiency with an improved reconstruction of electron tracks and a lower fake-rate (fig. 3.1).

Category	Metric	Allen	Etx4velo $x_{\text{triplet}} > 0.32$ $d^{xz}_{\text{max}} = 0.010$	Etx4velo $x_{\text{triplet}} > 0.36$ $d^{xz}_{\text{max}} = 0.020$
Long, no electrons ✓ In acceptance ✓ Reconstructible in the velo ✓ Reconstructible in the SciFi ✓ Not an electron	Efficiency	99.26%	99.28%	99.51%
	Clone rate	2.54%	0.96%	0.89%
	Hit efficiency	96.46%	98.73%	98.90%
	Hit Purity	99.78%	99.94%	99.94%
Long electrons ✓ In acceptance ✓ Reconstructible in the velo ✓ Reconstructible in the SciFi ✓ Electron	Efficiency	97.11%	98.80%	99.22%
	Clone rate	4,25%	7.42%	7.31%
	Hit efficiency	95.24%	96.54%	96.79%
	Hit purity	97.11%	98.46%	98.46%
Long, from strange ✓ In acceptance ✓ Reconstructible in the velo ✓ Decays from a strange *Good proxy for displaced tracks*	Efficiency	97.69%	97.50%	98.06%
	Clone rate	2.50%	0.92%	0.81%
	Hit efficiency	97.69%	98.22%	98.77%
	Hit purity	99.34%	99.68%	99.68%
X	Ghost rate	2.18%	0.76%	0.81%

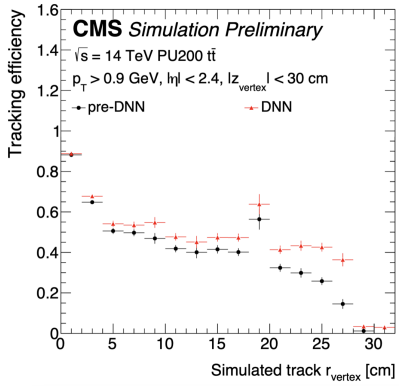
(a)

Category	Metric	Allen	Etx4velo $x_{\text{triplet}} > 0.32$ $d^{xz}_{\text{max}} = 0.010$	Etx4velo $x_{\text{triplet}} > 0.36$ $d^{xz}_{\text{max}} = 0.020$
Velo-only, no electrons ✓ In acceptance ✓ Reconstructible in the velo ✓ Not reconstructible in the SciFi ✓ Not an electron	Efficiency	96.84%	97.03%	97.86%
	Clone rate	3.84%	1.08%	1.02%
	Hit efficiency	93.89%	97.93%	98.32%
	Hit Purity	99.50%	99.84%	99.82%
Velo-only electrons ✓ In acceptance ✓ Reconstructible in the velo ✓ Not reconstructible in the SciFi ✓ Electron	Efficiency	67.81%	85.10%	86.69%
	Clone rate	10.27%	5.02%	4.97%
	Hit efficiency	79.21%	93.33%	93.88%
	Hit purity	97.35%	99.07%	98.99%
Velo-only, from strange ✓ In acceptance ✓ Not reconstructible in the velo ✓ Decays from a strange *Good proxy for displaced tracks*	Efficiency	93.53%	93.07%	96.05%
	Clone rate	5.60%	1.97%	1.77%
	Hit efficiency	90.05%	93.92%	96.05%
	Hit purity	99.36%	99.67%	99.64%

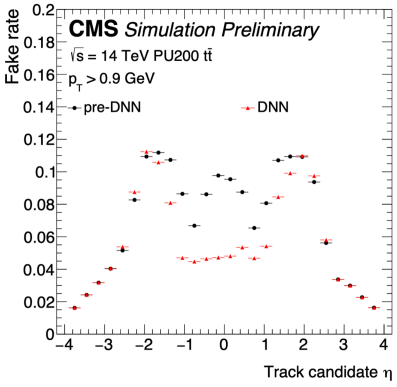
(b)

Figure 3.1: Efficiencies, clone rates and purities for long tracks (a) and Velo-only tracks (b).

Such applications demonstrate the power of NNs but are computationally expensive. A higher level approach can be taken by classifying which pre-built track segments should be connected to track candidates, as investigated by the CMS collaboration for the inner and outer tracker layers, to build T5 (5-cluster) from T3 (3-cluster) track segments [4]. Improved tracking efficiency (fig. 3.2a) with a reduced fake-rate (fig. 3.2b) is found compared to tracking without the NN for all primary vertex positions and particularly for the central rapidity region ($\eta \leq 2$).



(a)

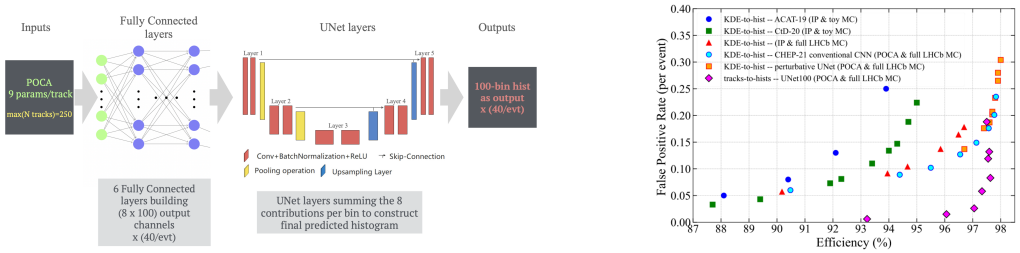


(b)

Figure 3.2: Comparison of tracking efficiency (a) and fake-rate (b) with and without the NN.

Once final tracks have been built, NNs can further improve primary vertex (PV) finding by processing the tracks in an event, as shown by a collaborative effort between ATLAS and LHCb [5]. A first

approach was taken by performing a kernel density estimation (KDE) on the number of tracks crossing the beamline along the z-direction and a convolutional NN identifying PV candidates (fig. 3.3). It was found that fully connected NNs outperform the KDE approach using additional tracking information with at most 250 tracks and 9 parameters per track. This is transformed to a latent space encoding of (8 x 100) output channels x 40 outputs / event which is further processed by a U-Net architecture. The final output are 100 x 40 channels per event which represent primary vertex candidates in bins of 100 μm steps along the z-direction with a total coverage of z $\in$ [-100 mm, 300 mm]. Ultimately, this represents a full tracks-to-hist classification with NNs, achieving significantly improved false positive rates compared to the previous KDE approach.

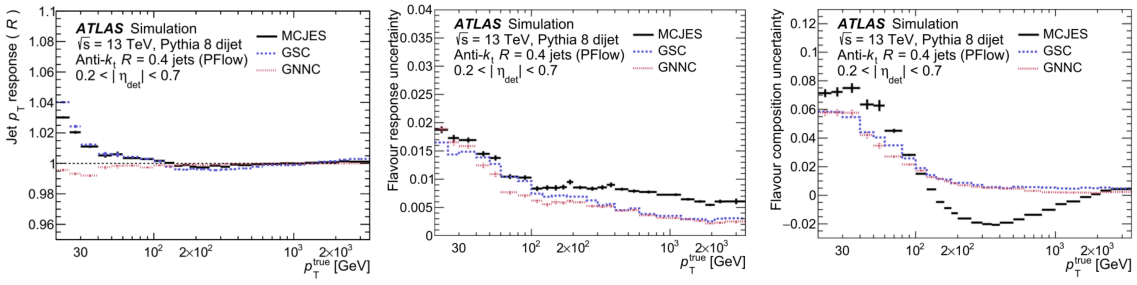


(a) **(b)**

Figure 3.3: NN architecture for PV finding (a) and corresponding efficiency vs. false-positive rate for different architectures and approaches (b).

4. Tagging, triggering, PID and calibrations

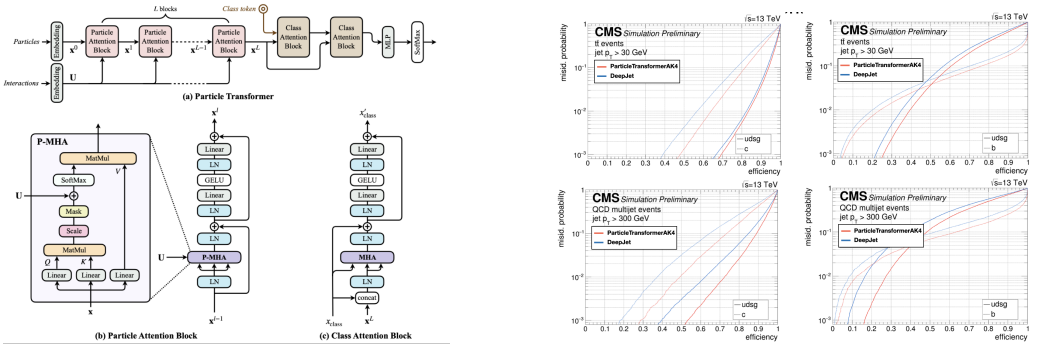
An application of fully connected NNs for the jet p_{T} calibration is performed by ATLAS with a global neural network based calibration (GNNC) for small-R jets [6]. Sequential, multiplicative corrections (global sequential calibration, GSC) based on 6 informative jet variables are replaced with a neural network approach with 13 inputs. Especially at low transverse momenta, the neural network approach shows significant improvements of the jet-p_{T} response compared to the Monte-Carlo ground truth. This can be explained as the GNNC approach is not bound to decorrelated corrections but can take multi-variable correlations across a high dimensional input-space into account (fig. 4.1). This also leads to slight improvements in the flavour response and composition uncertainty of the jet.



(a) **(b)** **(c)**

Figure 4.1: Jet-p_{T} response (a), flavour response uncertainty (b) and flavour composition uncertainty (4.1c) against the MC truth p_{T} of the jet for the GNNC and GSC approach.

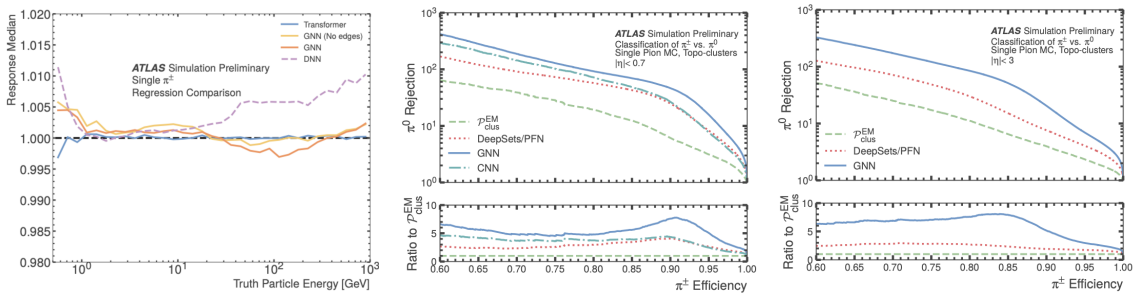
A recent architecture in research and industry, the transformer model, finds application in the CMS ParticleTransformerAK4 [7], which is tuned for anti-k_T jets with a radius parameter of 0.4. Its performance surpasses the heavy flavour jet tagging capabilities of the previous DeepJet DNN architecture in all p_T intervals and for all flavours (charm, beauty and up-down-strange-gluon). The transformer model is particularly well suited for jet tagging as it can take into account the full jet information in a single pass and is not bound to a fixed input size. This allows for a more flexible and versatile approach to jet tagging and calibration. The augmented attention mechanism for flavour jet tagging in the ParticleTransformerAK4 model is shown to be particularly powerful in the low p_T regime (fig. 4.2).



(a) **(b)**

Figure 4.2: Architecture (a) and performance (b) of the ParticleTransformerAK4 model for jet tagging.

A comparison of different model architectures was performed for point-cloud based data by ATLAS for pion identification [8]. The transformer architecture here shows a significantly improved performance compared to any other architectures which further signifies its versatility and performance (fig. 4.3).

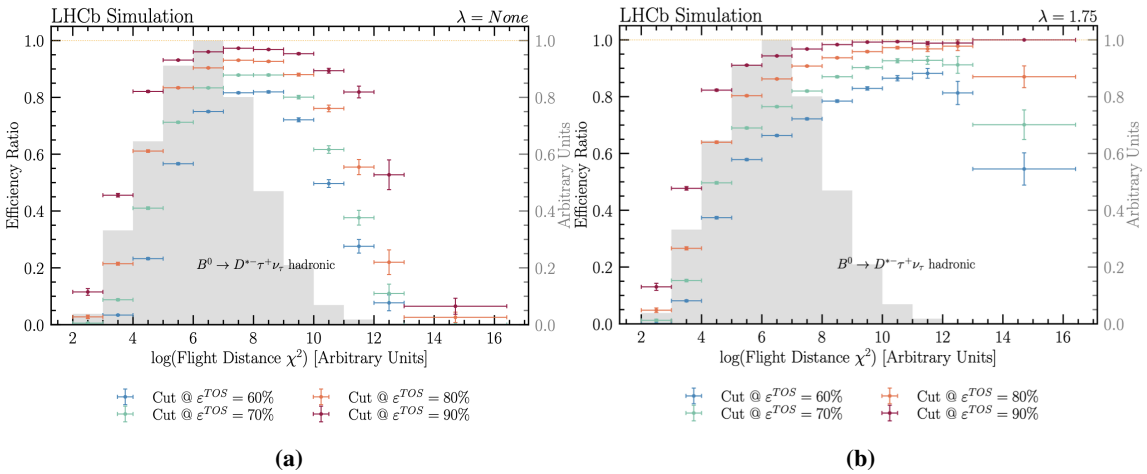


(a) **(b)** **(c)**

Figure 4.3: Study and performance of different model architectures for pion identification at $|\eta| < 0.7$ (a), $|\eta| < 3$ (b) and energy calibration (c) for point cloud based data in ATLAS.

Utilizing jet-constituent correlations, GNNs have also provided great utility for Lund-plane based W tagging including an adversarial NN for jet mass decoupling [9]. Lund plane variables spanned by individual two-particle contributions can be represented as a graph and lead to improvements in background rejection and jet energy calibration by the GNN.

In order to improve the robustness and interpretability (e.g. monotonicity) Lipschitz neural networks are utilized by LHCb in various applications of NN based tagging and triggering. A key characteristic of these neural networks is their stability, where an additional, linear term in the loss function ensures that all weights are constrained below a specified Lipschitz constant. This greatly improves the stability of the prediction [10]. The application of such Lipschitz networks shows significant improvements for inclusive heavy-flavour triggers and particularly for particles with long flight paths (fig. 4.4).



(a) **(b)**

Figure 4.4: Comparison of triggering efficiency ratio for the hadronic decay channel $B^0 \rightarrow D^{*-}\tau^+\nu_\tau$ hadronic with a regular NN (a) and a Lipschitz NN (b).

For offline calibrations, fully connected NNs have shown great success for multidimensional corrections to the Bethe-Bloch function for particle identification in the ALICE TPC [11]. Based on clean V0 selection, this approach is fully data-driven and improves the previous single-dimension, multiplicative corrections for electrons, pions, kaons and protons (fig. 4.5). Six track parameters are used as an input to the model and two values are returned (mean correction factor and σ-estimation). This approach is fully utilized and deployed within the collaboration and the Hyperloop train system for offline data analysis in Run 3.

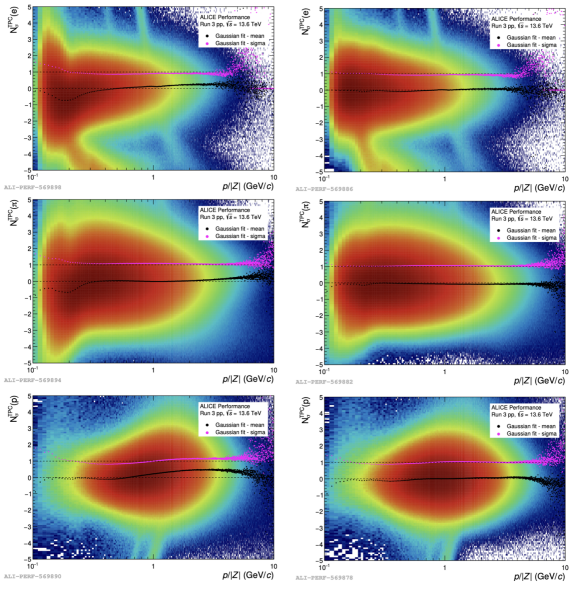


Figure 4.5: Improvement of the Nσ distributions for cleanly selected electrons, pions, and protons from pure Bethe-Bloch ratios (left) and with NN corrections (right).

5. Conclusion

In summary, the LHC experiments have made great strides in the application of machine learning algorithms and neural networks in online and offline reconstruction and calibration. The applications range from cluster-level to track-level applications and finally to calibrations, tagging and triggering. The use of neural networks with a variety of architectures has shown significant improvements in terms of physics performance and computational scalability to larger object multiplicities. The experiments continue to push the boundaries of machine learning applications in high-energy physics and are well positioned to take advantage of the latest developments in the field.

References

[1] David Rohr. Usage of GPUs in ALICE Online and Offline processing during LHC Run 3. *EPJ Web Conf.*, 251:04026, 2021. doi: 10.1051/epjconf/202125104026.

[2] CMS collaboration. Level-1 Trigger Calorimeter Image Convolutional Anomaly Detection Algorithm. . URL https://cds.cern.ch/record/2879816?ln=en.

[3] Anthony Correia, et al. GNN-based pipeline for track finding in the Velo at LHCb. URL https://indico.cern.ch/event/1252748/contributions/5521484/attachments/2731094/4748485/etx4velo_ctd2023.pdf.

[4] CMS collaboration. Improved Performance of Line Segment Tracking Using Machine Learning. . URL https://cds.cern.ch/record/2872904?ln=en.

[5] Simon Akar, Mohamed Elashri, Rocky Bala Garg, Elliott Kauffman, Michael Peters, Henry Schreiner, Michael Sokoloff, William Tepe, and Lauren Tompkins. Advances in developing deep neural networks for finding primary vertices in proton-proton collisions at the lhc, 2023.

[6] ATLAS collaboration. New techniques for jet calibration with the ATLAS detector. URL https://doi.org/10.1140/epjc/s10052-023-11837-9.

[7] Transformer models for heavy flavor jet identification. 2022. URL https://cds.cern.ch/record/2839920.

[8] Point Cloud Deep Learning Methods for Pion Reconstruction in the ATLAS Experiment. Technical report, CERN, Geneva, 2022. URL https://cds.cern.ch/record/2825379. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2022-040.

[9] Tagging boosted W bosons applying machine learning to the Lund Jet Plane. Technical report, CERN, Geneva, 2023. URL https://cds.cern.ch/record/2864131. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2023-017.

[10] Blaise Delaney, Nicole Schulte, Gregory Ciezarek, Niklas Nolte, Mike Williams, and Johannes Albrecht. Applications of lipschitz neural networks to the run 3 lhcb trigger system, 2023.

[11] Christian Sonnabend. Neural network regression for particle identification with the ALICE TPC detector in Run 3, 2022. URL https://cds.cern.ch/record/2856252. Presented 28 Nov 2022.