Deep Learning Approaches for LHCb ECAL Reconstruction

Alexey Boldyrev[1,*], *Denis* Derkach[1], *Fedor* Ratnikov[1,2], and *Andrey* Shevelev[1,2]

[1]HSE University
[2]Yandex School of Data Analysis

Abstract. Calorimeters are a crucial component for most detectors mounted on modern colliders. Their tasks include identifying and measuring the energy of photons and neutral hadrons, recording energetic hadronic jets, and contributing to the identification of electrons, muons, and charged hadrons. To fulfill these many tasks while keeping costs reasonable, the calorimeter construction requires good and thoughtful balancing with other components of the detector. Much harder operation conditions during LHC's high luminosity Run 5 and beyond bring new technological and computational challenges. This requires optimization of technologies, layouts, readouts, reconstruction algorithms to achieve the best overall physics performance for the limited cost. In the traditional approach, the reconstruction of the physical objects in the calorimeter must be matched to the calorimetric showers simulation used. We present a deep learning-based approach to help utilize raw simulated calorimetric data of varying degrees of detail.

1 Introduction

The LHCb experiment is one of the major experiments conducted at the Large Hadron Collider (LHC). Its main objectives include conducting precise tests to verify the Standard Model, carrying out detailed studies of Charm and Beauty physics, and making accurate measurements of CP violation effects. The accurate identification of these processes and their properties assumes proper reconstruction techniques. The electromagnetic calorimeter is responsible for measuring the transverse energy of electrons, photons, and π^0 particles.

This measurement facilitates the selection of B decays with an electron or a photon in the final state. The LHCb Electromagnetic Calorimeter (ECAL) is composed of a 2-dimensional grid of modules constructed from lead absorber plates and scintillator tiles.

In current setup, these modules are arranged in three different regions with varying numbers of readout cells. The inner region, closest to the beam pipe, has 167 modules with 9 cells of 4x4 cm^2 area. The middle region consists of 448 modules with 4 cells of 6x6 cm^2. Lastly, the outer region contains 2688 modules consisting of a single cell of 12x12 cm^2. During the reconstruction phase, data is collected from the readout cells in each of these three regions.

Several technological options for the upgraded LHCb ECAL are foreseen. The most severe requirements for radiation tolerance can be met by SPACAL modules consisting of longitudinal fibres acting both as scintillator and light-transporting medium [1]. The use of new options for the upgraded calorimeter suggests challenges for reconstruction methods.

*e-mail: alexey.boldyrev@cern.ch

In this article, we propose a strategy for obtaining geometry agnostic inputs for machine learning based calorimeter reconstruction and discuss the applicability of such a strategy to the LHCb ECAL. These approaches primarily focus on demonstrating their feasibility rather than providing a specific reconstruction solution for the LHCb calorimeter data.

In recent years, the advancement of deep learning models has encouraged the use of image processing techniques in the field of high energy physics. This paper proposes a different reconstruction strategy that aims to simplify the processing of machine learning models to raw calorimeter data. It is based on the formulation of the current reconstruction algorithm but utilizes a set of deep learning structures, each trained to solve one step of the reconstruction process. With this approach, the same steps as the current algorithm are reproduced, but with the advantage of using efficient deep architectures that have constant cost functions. LHCb ECAL optimization is a good example for the general problem of global optimization of a complex physical detector.

A pipeline approach to LHCb ECAL optimization is developed and used to scan the parameter space in search of a desired subspace that satisfies the requirements of the LHCb experiment. Parameters for calorimeter optimization include technology, granularity, Molière radius, temporal resolution of the calorimetric modules, and module distribution in the ECAL region. The physical significance of the selected channel versus the detector cost can be considered as a metric of detector optimization performance. Assuming a differentiable implementation of surrogates, the pipeline approach allows a faster optimization cycle compared to black-box optimization. Recent improvements to the pipeline approach and its ability to determine the degree of modeling granularity required are discussed.

2 Previous approaches to reconstruction in the LHCb ECAL

Reconstruction of the properties of physical objects in the electromagnetic calorimeter depends on the technologies used, on the arrangement of sensitive elements and on the data-taking conditions in the experiment. Such properties are the energy, coordinate, and arrival time of the particle. Conventional parametric reconstruction obtains the properties of a physical object from the shape of clusters of calibrated energy deposits in a set of calorimeter cells selected by the seed finder algorithm. The reconstruction algorithm in the LHCb Calorimeter is based on Cellular Automaton algorithm [2]. This approach has been commonly used in calorimeters for high energy physics due to the detector's geometry [3]. While the current approach efficiently reconstructs clusters in an event, it requires multiple iterative calculations within cells data. This leads to the algorithm's complexity depending heavily on the number of clusters in the data.

Earlier our group demonstrated the possibility of reconstructing the characteristics of a physical object in an electromagnetic calorimeter using a regressor based on gradient boosting on decision trees [4]. In this approach, each property of the physical object was trained in a separate regressor, which received as input data a two-dimensional array of energy deposits (as well as a 2d array of times measured in cells to determine the time of arrival of the particle). In the described approach, this 2d array has a fixed size of 5x5 cells. The position of this 2d array was determined by the cell with highest energy deposited in the event. This approach reconstructs properties of physical objects with comparable quality to that of the parametric conventional approach. On the one hand, the fixed size of the input data allows us to ease the scalability of the calculations and avoid bounds and bottlenecks of available computer resources. In addition, in this method it is possible to use raw counts from the photomultipliers instead of calibrated energy deposits in the conventional approach. On the other hand, the fixed size of the input data creates constraints for events near the boundaries between calorimeter regions with different granularity, as well as for events that do not fit into this 2d array. The importance of this problem increases when, under the conditions of the modernized calorimeter, the boundary between the regions no longer has a rectangular shape. Additional limitations common to both the conventional and the previously proposed Machine Learning (ML) based reconstruction methods apply to any other deviations from the geometric regularity of sensitive elements. For example, the SPACAL calorimetric module technology proposed for the LHCb ECAL modernization requires tilting the calorimetric modules by 3 degrees in both axes [5]. This requires a gap between modules with different tilts, which leads to deviations from the regular structure of the calorimeter cells.

3 Method

The following strategy is proposed to return to the regular cell structure of the calorimeter:

- Interpolation of the cells for equalization of granularity on both sides of the border (see Fig. 1);
- Rows or columns of sensitive elements skipped for engineering reasons. Interpolation of non-existing energy deposits in missing 'virtual' cells for rows or columns of sensitive elements skipped for engineering reasons;
- Recovering incomplete events at the edges of the calorimeter using classical or ML interpolation.

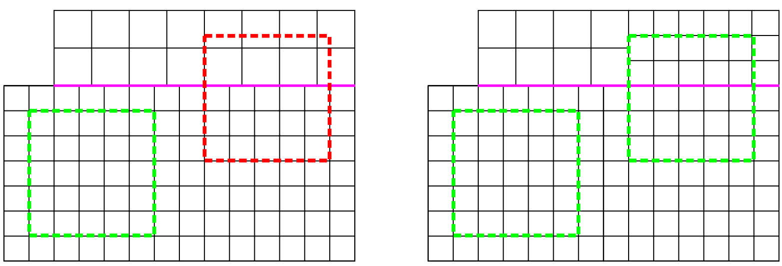

Figure 1. Left panel: the region under consideration while scanning regular cell structure (green dashed line) and irregular cell structure (red dashed line) in the border area between different granularities. Right panel: irregular cell structure in the border area is converted into regular cell structure using two-dimensional interpolation. The thick magenta line shows the border between the different granularities.

Two-dimensional linear, cubic, and nearest-neighbor interpolations were used to recover clusters near the boundaries of different granularities. The recovery of incomplete clusters in other cases is performed using fully-connected neural networks. Training is performed using Adam Optimizer [6]. MSE loss is used with an early stopping strategy. The architecture of the neural network approach is shown in Fig. 2.

4 Quality evaluation

The quality of calorimetric cluster inpainting is assessed using both standard image comparison methods such as peak signal-to-noise ratio (PSNR) and structural similarity index measures (SSIM) [7], and by reconstructing a physical object and then comparing its properties with MC Truth values. Image inpainting models can be evaluated with a large set of images without labeling. Evaluating is done by calculating the distance between restored samples and original ones and by comparison of corresponding distributions. Quality varies

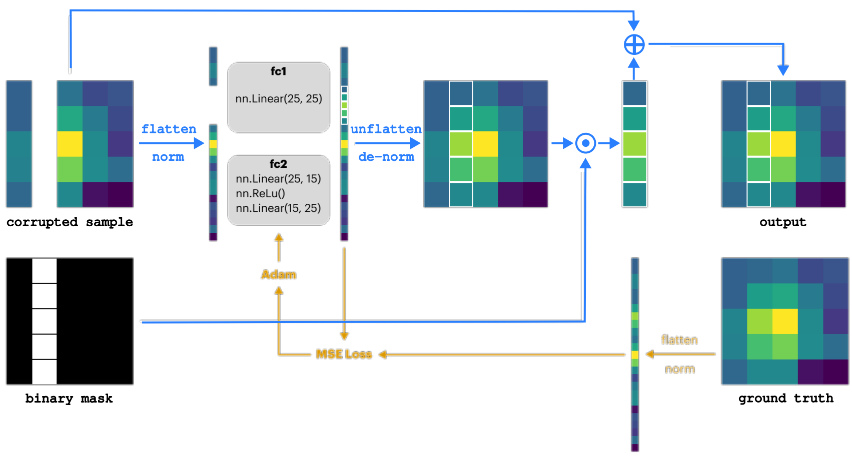

Figure 2. Cluster inpainting model architecture using fully-connected neural networks.

greatly based on mask scale, resulting in separate metrics for different mask scales. Additionally, advanced techniques are employed to create binary masks for training. The quality of a reconstructed physical object is determined by its position, energy and timing resolution.

5 Dataset

The data used is taken from the detailed simulation of the calorimeter responses obtained in the standalone simulation from the standard physical processes simulation framework GEANT4 [8] and is designed to optimize various proposed configurations of the LHCb electromagnetic calorimeter. A total of 100000 single photon events have been divided into 80/20 proportions for training and testing.

6 Results

Several classical ML and Deep Learning (DL) models are probed: Boosted Decision Trees [9], Feed-Forward Networks [10], and Convolutional neural networks. Without a pile-up, position, energy and timing of the photons are estimated consistently with current LHCb ECAL design.

Figure 3 shows that classical interpolation methods can be applied to recover missing information in a calorimetric cluster when the irregularity is at its periphery. Table 1 shows that fully-connected deep learning models outperform classical interpolation methods for all selected metrics and can be applied including for the recovery of defects in the calorimetric cluster when the irregularity passes through its center.

7 Conclusions

Automated reconstruction in the calorimeter using machine learning models accelerates the R&D process and ensures uniformity and consistency of results. Interpolation at the bound-

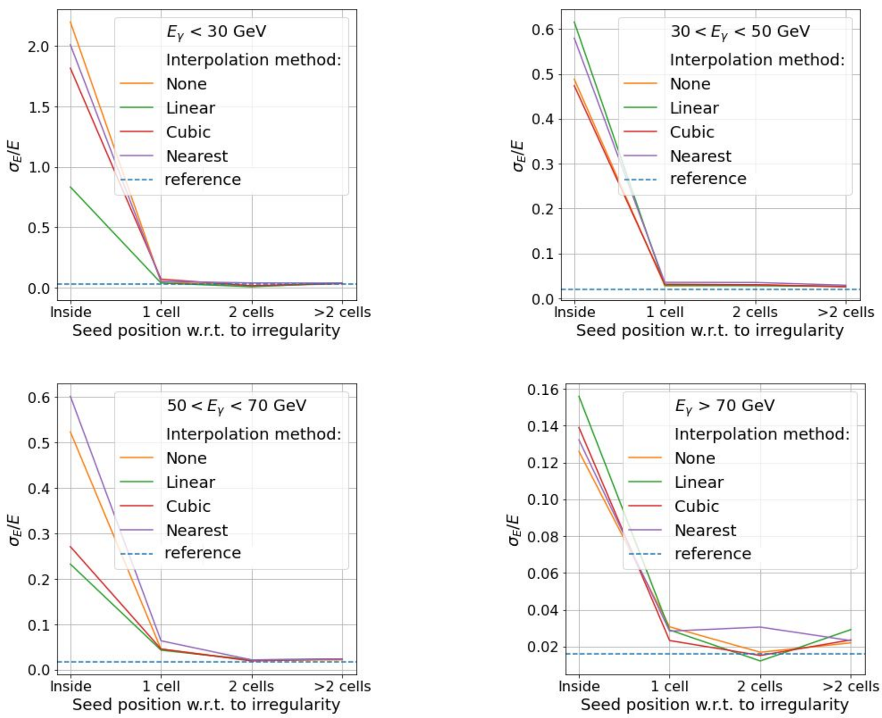

Figure 3. Energy resolution for different methods of interpolation of the missing information in clusters depending on the position of irregularity and cluster energy.

Model		Energy RMSE↓	PSNR↑	SSIM↑
Interpolation	Nearest-neighbor	223545	85.3	0.74
	Cubic	208187	91.2	0.75
	Linear	159946	92.8	0.79
Deep Learning	FC-1	58343	82.8	0.78
	FC-2	**28928**	**96.9**	**0.94**

Table 1. A comparison of classical interpolation methods and DL interpolation for recovering of the calorimetric cluster with missing information. The metrics are averaged over the position and type of the irregularity.

aries between different granularities and at the technological gaps of the electromagnetic calorimeter improves the physical quality metrics.

References

[1] S. Kholodenko (LHCb ECAL UII working group), PoS **PANIC2021**, 100 (2022)

[2] V. Breton, N. Brun, P. Perret (2001)

[3] V. Breton, H. Fonvieille, P. Grenier, C. Guicheney, J. Jousset, Y. Roblin, F. Tamin, Nucl. Instrum. Meth. A **362**, 478 (1995)

[4] A. Boldyrev, D. Derkach, F. Ratnikov, A. Shevelev, JINST **15**, C09030 (2020), 2005.07700

[5] L. An, E. Auffray, F. Betti, F. Dall'Omo, D. Gascon, A. Golutvin, Y. Guz, S. Kholodenko, L. Martinazzoli, J. Mazorra De Cos et al., Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **1045**, 167629 (2023)

[6] D.P. Kingma, J. Ba, *Adam: A method for stochastic optimization* (2017), 1412.6980

[7] A. Hore, D. Ziou, *Image quality metrics: PSNR vs. SSIM*, in *2010 20th international conference on pattern recognition* (IEEE, 2010), pp. 2366–2369

[8] S. Agostinelli et al. (GEANT4), Nucl. Instrum. Meth. A **506**, 250 (2003)

[9] T. Chen, C. Guestrin (2016), 1603.02754

[10] D. Svozil, V. Kvasnicka, J. Pospichal, Chemometrics and Intelligent Laboratory Systems **39**, 43 (1997)