Norfolk, Virginia, USA • May 8-12, 2023
CHEP 2023
Computing in High Energy & Nuclear Physics

Deep Learning Approaches for LHCb ECAL Reconstruction

Alexey Boldyrev[1,*], Denis Derkach[1], Fedor Ratnikov[1,2], Andrey Shevelev[1,2]
Ivan Spirin[1], Emil Akopyan[1], Ruslan Vorovchenko[1]

[1] *HSE University,* [2] *Yandex School of Data Analysis*
[*] *alexey.boldyrev@cern.ch*

NATIONAL RESEARCH UNIVERSITY
SCHOOL OF DATA ANALYSIS

Introduction

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 (Upgrade II conditions) imply 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.

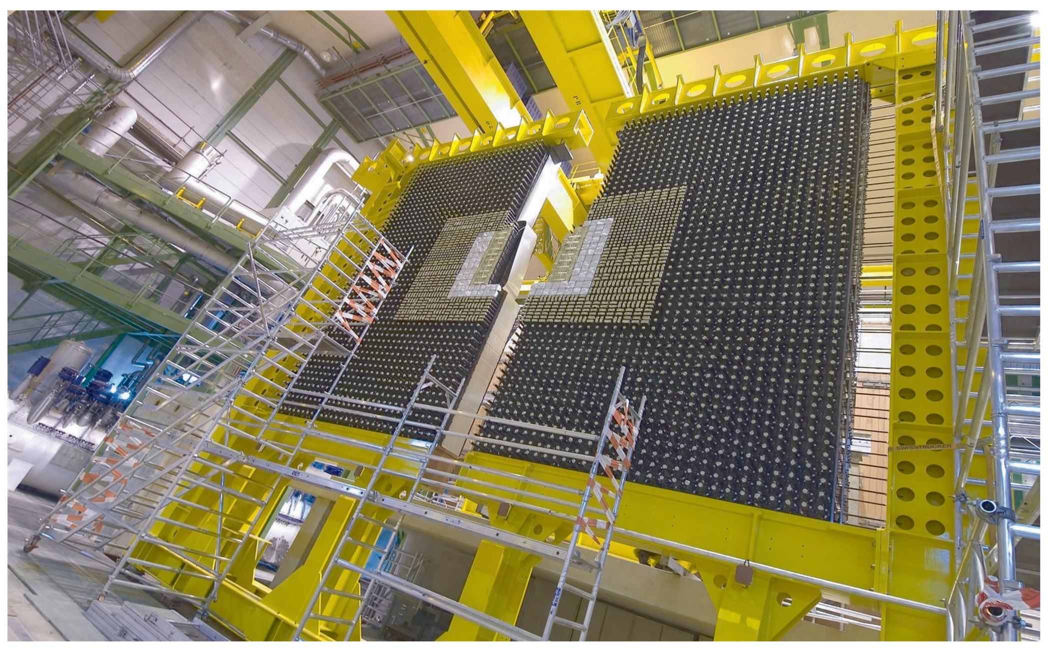
LHCb detector

LHCb is one of four major LHC experiments and provides:

- Precise tests for Standard Model verification;
- Detailed studies of Charm and Beauty physics;
- Precise measurements of CP violation effects.

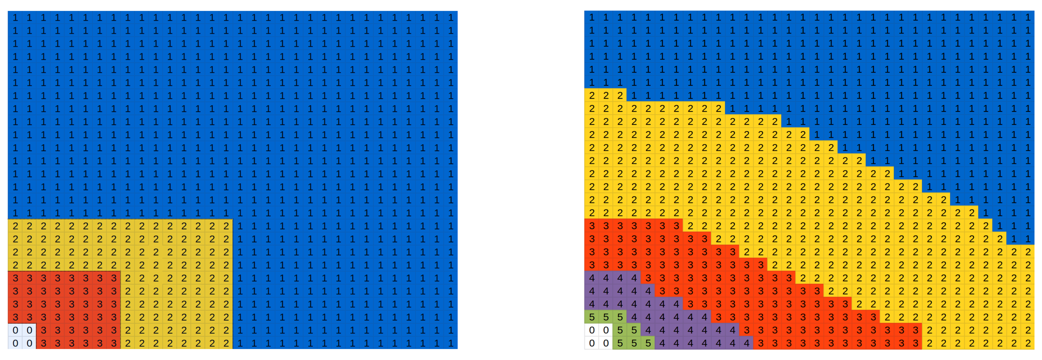
LHCb ECAL

The current ECAL is based on Shashlik-type modules of 3 granularities, and contains 1536/1792/2688 cells in its inner/middle/outer regions, respectively.



LHCb calorimeter wall. Image credit: CERN.

Several technological options for the upgraded calorimetry are foreseen. The most severe requirements for radiation tolerance can be met by SpaCal modules consist of longitudinal fibres acting both as scintillator and light-transporting medium.



Baseline configuration (left) and the first level of optimisation of the cell sizes for LHC Upgrade II conditions (right). Upper right quarter of the calorimeter wall is shown.

Figures of merit

The performance of the calorimeter consists of:

- Radiation tolerance to sustain the expected lifetime span;
- Energy and spatial resolution for good photon reconstruction and electron identification;
- High granularity and longitudinal segmentation to facilitate better precision, both spatial and in energy, which in turn improves reconstruction algorithms;
- Timing resolution enough to facilitate pileup suppression in high-occupancy areas as well as better matching of separate signal components.

However, the ultimate goal for the optimization process is to achieve the dependency of the physics performance on the cost of the configuration of the detector under study.

Models in the optimization cycle

Model	Primary implementation	Differentiable impl.
Detector response	GEANT4	Surrogate model
Detector reconstruction	N/A	ML regressor
Performance metric aggregator	Physics significance vs. cost	Physics quality of the objects

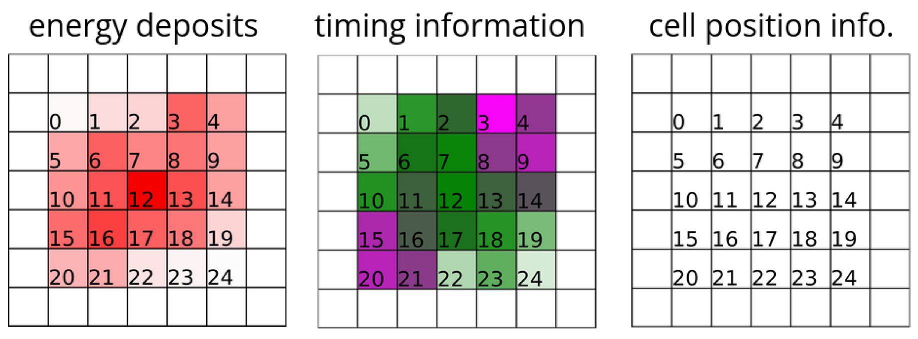
Models are required to be differentiable with respect to θ_{model} in vicinity of the current optimization point. For non-differentiable models the differentiable machine learning (ML) based surrogates can be used.

Conclusions

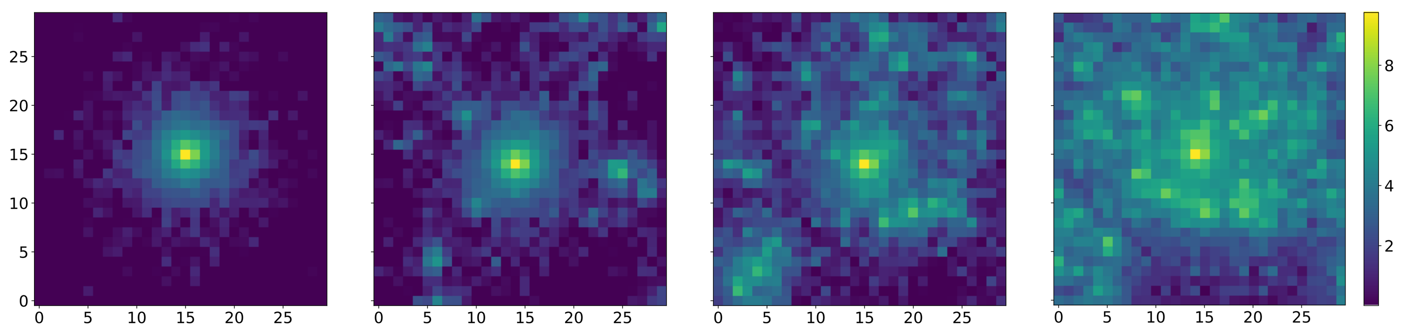
- LHCb ECAL is a good use case for the generic problem of comprehensive optimization of the complex physics detector;
- The R&D process requires time consuming computation steps to evaluate physics performance for different detector techniques and configurations;
- Automatic training using machine learning models speeds up the turnover for the performance studies and ensures consistency and uniformity of obtained results.

Deep Learning reconstruction

Simulated GEANT4 response is the 5x5 array of cells. They are used as base features for Spatial and Energy regressors. Time regressor uses weighted energy deposits.

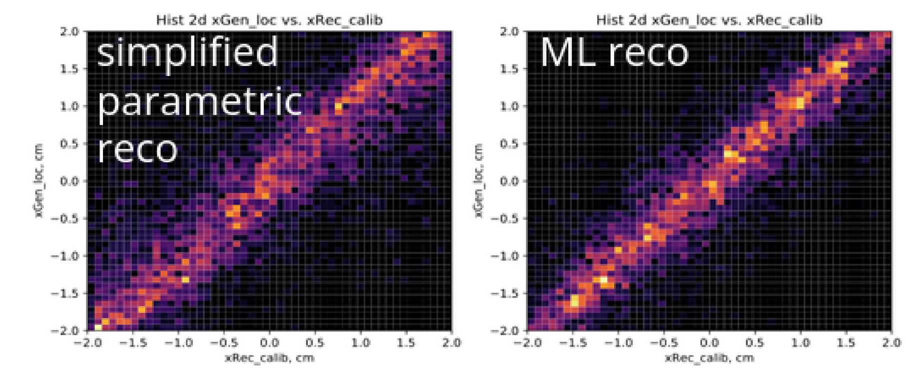


Basic features for the regressors inputs. The 5x5 dimension is due to the typical size of a calorimetric cluster and an additional cell layer to estimate background contributions.

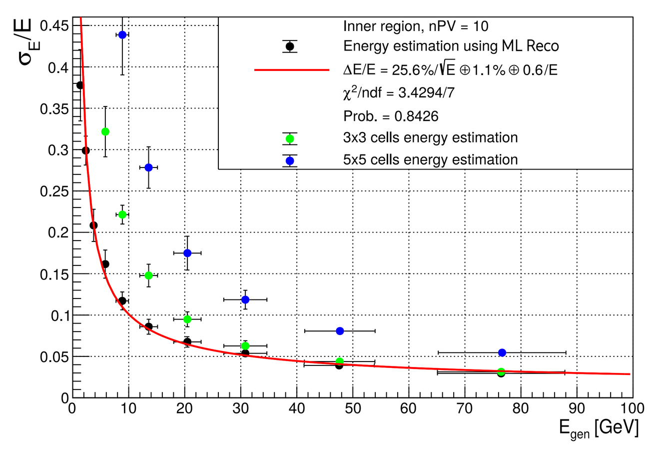


The calorimetric clusters in the matrix of 30x30 cells. The same signal cluster is enclosed by the background clusters for different pile-up conditions. From left to right: without pile-up, nPV = 5, nPV = 20, nPV = 50. The colour represents ln(E/MeV) for every cell. Standalone GEANT4 simulation used.

Several classical ML and DL models are probed: Boosted Decision Trees, Feed-Forward Networks, Convolutional NNs. Without pile-up position, energy and time are estimated consistently with LHCb ECAL design. At increased pile-up ML reconstruction still shows meaningful estimation while parametric reconstruction require fine-tuning.



Dependence of the reconstructed coordinate on the true coordinate under pail-up conditions. The color from violet (dark) to yellow (bright) represents the normalised counts of the events from 0.0 to 1.0, respectively.



Energy resolution as a function of photon energy for the ECAL inner modules (nPV = 10). The fit (red line) corresponds to the ML reco data (black circles). Energy is measured in GeV. Standalone GEANT4 simulation.

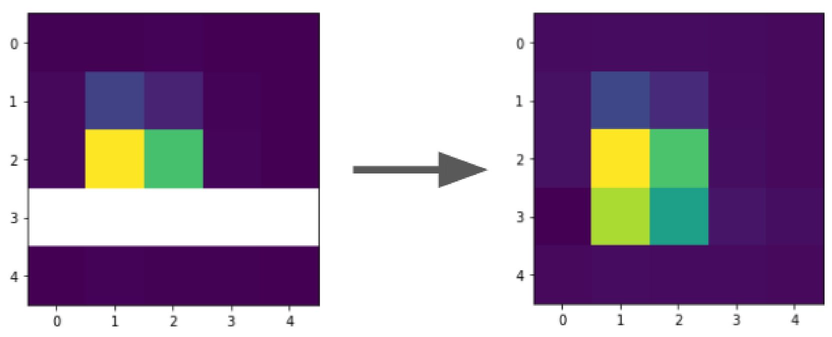
Creation of geometry agnostic inputs for large models

Another area of application of deep learning is to compensate for the irregular arrangement of detector sensitive elements. Examples of such irregularities include:

- Boundaries between regions with different granularity. Interpolation of the cells for equalization of granularity on both sides of the border can be applied.
- Rows or columns of sensitive elements skipped for engineering reasons. Interpolation of non-existing energy deposits in missing 'virtual' cells can be used here.
- Edges of the calorimeter. As in the preceding, classical or ML interpolation can recover events at the edges.

The creation of geometry agnostic inputs in reconstruction algorithms allows us to greatly simplify the architecture of the models. This can be done with small additional models that preprocess the data upstream of the larger reconstruction models.

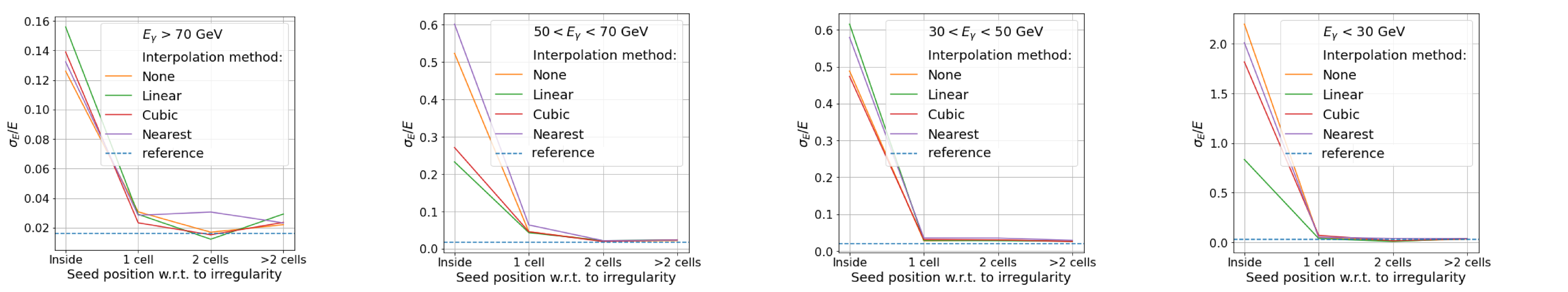
We have compared several interpolation methods and fully connected neural networks to recover missing row or column inside the matrix of cells. The averaged metrics by the position of the missing row (column) of cells are presented in the table. The best model turned out to be fully-connected neural network with two linear layers and ReLU activation (FC-2).



An example of calorimetric cluster with missing row of cells (left) and recovered cluster (right).

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**

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.



A comparison of methods of interpolation of the missing information in clusters depending on the position of irregularity and cluster energy.