GNN-based End-to-End Reconstruction in the CMS Phase 2 High-Granularity Calorimeter

Lead-PI: Matteo Cremonesi (CMU)
Co-PI: Manfred Paulini (CMU)
Co-PI: Lindsey Gray (FNAL)
Co-PI: Thomas Klijnsma (FNAL)
Research project term: September 1, 2023-August 31, 2024
Salary support: \$43,000 (total), \$35,000 (salary) \$9,000 (fringe)
Travel funds: \$5,000

Contents

1 Introduction

The High Luminosity LHC (HL-LHC) is expected to deliver an integrated luminosity of 3000 fb^{-1}, running with up to 200 simultaneous interactions per bunch crossing (pile-up) and with doses of 1.5 MGy (150 Mrads). This will pose significant challenges for radiation tolerance and event pile-up on detectors, especially for forward calorimetry. Since both the existing forward $PbWO_4$-based electromagnetic and the plastic scintillator based hadron calorimeters would exhibit significant performance degradation in these severe conditions [1, 2], the CMS Collaboration is designing a radiation-hard high-granularity calorimeter (HGCAL) to replace them as part of its HL-LHC upgrade program. The HGCAL will have unprecedented transverse and longitudinal segmentation and it will provide energy deposition measurements finely segmented in both space and time. The fine lateral and longitudinal granularity of the HGCAL requires the implementation of particle-flow calorimetry with precise feature extraction and provide excellent performance in the highest pile-up environments expected during the HL-LHC.

That being said, the HGCAL poses an exceptional challenge for particle-shower reconstruction. It is unclear whether traditional clustering algorithms can satisfy computing constraints while providing sufficient high level physics information that uses the improved detector technology. Performing particle reconstruction with machine-learning (ML) algorithms can provide a solution by making full use of hardware acceleration and advanced pattern-recognition techniques. More recently, an end-to-end ML reconstruction algorithm for the CMS Phase 2 HGCAL has been trained and tested in a multiparticle environment derived from simulated tau lepton decays with no pile-up [3], using the "Object Condensation" loss function [4]. Achieving accurate full end-to-end reconstruction with ML poses a formidable challenge, but compared to classical or hybrid (mix of classical and ML based) methods it presents a promising opportunity to fully exploit the detector's high resolution capabilities. Building on top of this progress, we propose to hire a new postdoc at Carnegie Mellon University (CMU), who will derive half of their support via the U.S. CMS operations program, to develop ML HGCAL reconstruction algorithms in the projected 200 pile-up environment, yielding the necessary physics performance while still satisfying the computing constraints.

2 Previous Work

Co-PIs Thomas Klijnsma and Lindsey Gray, together with other CMS collaborators, led the development of the first HGCAL ML reconstruction that used full-simulation CMS HGCAL data as input. Segmenting the event into individual particle showers presents a significant challenge due to the variability in the number of hits and particle showers per event, as well as the partial overlap of these showers. To tackle these obstacles, the Object Condensation loss function serves as an optimization target. This function facilitates instance segmentation by grouping hits from the same shower in close proximity within a latent space while pushing hits from different showers apart. Additionally, the Object Condensation loss function not only offers a starting point for clustering but also holds the potential to predict particle shower properties simultaneously. The ML model contains a graph neural network (GNN) with various "GravConv" [5] layers, trained using simulated di-tau production events with 0 pile-up. These events are obtained from the CMS detector simulation using Geant4 [6, 7] with enhanced tracking of each incident particle's history. This was obtained with a series of patches propagated to the CMSSW code, collectively referred to as “FineCalo”. FineCalo allows for more fine grained saving of the Geant4 truth information, resulting in a concise map of simulation-level energy deposits onto reconstructed energy deposits from HGCAL. Since the showers of incident particles may overlap entirely, and therefore cannot be feasibly reconstructed separately, this truth map is processed further to account for the limitations of the detector. To form the final ground truth for the training, simulated particles are merged together if they are not expected to be separable by the reconstruction. This is done by assessing the overlap with neighboring particles starting from the first hit it leaves in the detector. Based on parameters that can be tuned, adjacent hits in the same layer are collected and the spatial distribution of these hits is used to estimate a shower radius, taking into account the sensor sizes. If the circular projections of two showers on the front face of HGCAL overlap, the corresponding particles are merged. Each entry in the dataset consists of about 20k simulated detector hits, and each detector hit is 5-dimensional (energy, three spatial coordinates, and time). Every entry contains a different number of detector hits and particles and all hits are used as input to the algorithm.

3 Proposed Research

3.1 Deriving a 200 Pile-Up Training Dataset

The overarching goal of the project presented in this proposal is to develop GNNs for the HGCAL reconstruction that work with the HL-LHC projected pile-up. In order to achieve this goal the very first step is to derive a pile-up-aware training dataset.

A recent milestone has been reached with the production of the first dataset incorporating pile-up. To achieve this, a dedicated pile-up library was created initially, which involved generating minimum bias events using the FineCalo modifications in the CMSSW simulation code. Following that, di-tau production events were simulated, also incorporating the FineCalo modifications. The standard CMSSW pile-up mix-in tools were employed to create events with tau decays and 30 pile-up collisions drawn from the minimum bias library. For a realistic reconstruction algorithm, it is important that simulated particle showers from pile-up events are treated on par with those from the main event. To this end, a custom algorithm was implemented to extract and save additional truth information from the pile-up events, and to merge this supplementary data with the corresponding information from the main event.

The production of this dataset is a significant achievement, as it involved overcoming numerous challenging technical obstacles. However, despite this success, further improvements are necessary to effectively train the reconstruction algorithm for future pile-up scenarios. Overcoming the current limitations will be a responsibility of the postdoc supported in the context of this proposal. The workflow used to create the pile-up library currently relies on custom adaptations to CMSSW that are not yet integrated into the mainstream, which increases the complexity to use it. Also, the number of events generated is 40k, which is not enough to reliably represent the wide range of event topologies expected at the HL-LHC. One of the first tasks of the postdoc would be to streamline the workflow in order to improve its scalability and allow for the efficient production of a larger pile-up library. Ideally such a workflow will be integrated into CMSSW. The choice of di-tau production has proved to be very effective so far since it provides a wide variety of decay products with locally dense environments, which is essential when validating reconstruction algorithms. At the same time, in order to produce a reconstruction algorithm that performs under real conditions, more physics processes are needed. The postdoc will work on the generation of single photon production, single pion production, and top quark pair production events using FineCalo. The current dataset contains 40000 events, each with 30 pile-up collisions, which is enough for a proof-of-concept training, but to prove the viability of GNNs for a wider range of event topologies, larger datasets are needed. The postdoc will work on producing a larger dataset, that will derive from the mixing of events from all the aforementioned physics processes with 200 pile-up collisions from the larger minimum bias library, as opposed to the current 30 pile-up collisions.

3.2 Optimizing the Loss Function

First attempts at training the GNN on the proof-of-concept 30 pile-up dataset failed, due to saturating the GPU memory when computing the Object Condensation loss function. This is most likely caused by an implementation of the loss function in Python that does not consider memory constraints but rather optimizes for runtime. Addressing this issue will not only force us to rework the loss function implementation, but also to rethink and improve the loss function itself. Currently the Object Condensation loss function is used, which is essentially a combination of multiple loss functions: one attractive-repulsive loss, which pushes hits belonging to the same cluster closer together in a latent space; one “beta loss term”, which assigns per hit a likelihood that it is a “condensation point”, which means it can be used as a clustering ansatz; and a property loss term, which may be used to predict properties like energy or particle ID per cluster. All these different components can be optimized. The postdoc will try first-order improvements such as different scaling behaviors, but also more involved optimizations such as the “modified differential multiplier method” [8] or the “influencer loss” [9]. This gradual improvement process, which requires translating physics intuitions about potential fields into differentiable loss functions, will be essential for future HGCAL reconstruction efforts using GNNs. In Table 1 a detailed timeline of the goals is outlined. The different tasks are split into the different quarters of the year.

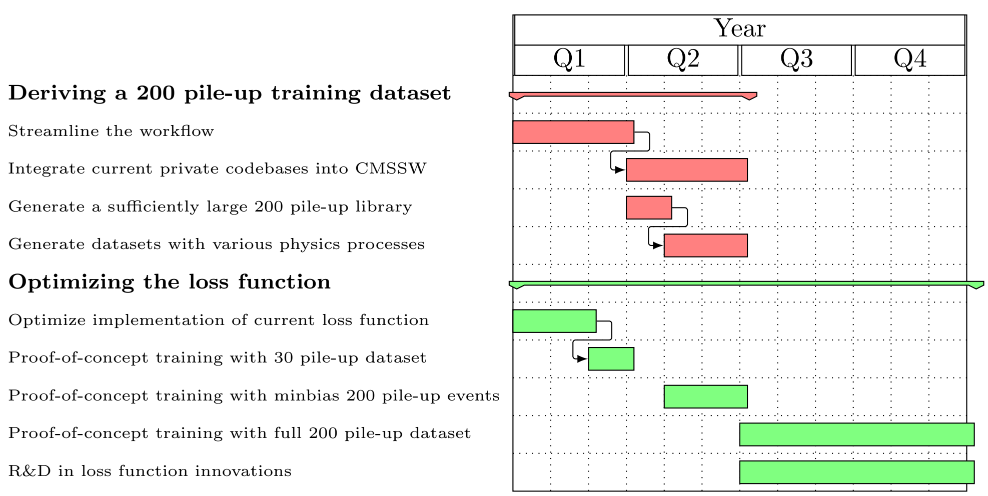

Table 1: Quarter by quarter outline of the proposal.

4 Mentoring Plan

CMU is one of the only three HGCAL module assembly centers in the US, responsible for producing about 10k standard modules for the hadronic section of HGCAL. The local CMS group built at CMU a class 1000 cleanroom and it commissioned all the major pieces of equipment needed for the module production. This will provide the postdoc with the unique opportunity to connect the work done on the reconstruction software to the hardware components of the HGCAL. Furthermore, the members of the CMS group at CMU are all involved in projects that use GNNs for reconstruction, an expertise that is directly applicable to the project described in this proposal.

Lead-PI Matteo Cremonesi has given significant contributions to CMS computing, such as leading the Computing Operation Production & Reprocessing Group as well as co-founding the Columnar Object Framework For Effective Analysis (COFFEA) project. Furthermore he led the development of the first GNN for missing transverse momentum (MET) reconstruction, which outperformed existing MET reconstruction algorithms by bringing a 10% improvement in resolution. This experience will prove valuable when supervising the postdoc in the usage of similar algorithms for the HGCAL reconstruction.

Co-PI Manfred Paulini is a senior member of the CMU faculty as well as the leader of the CMS group at CMU. He is one of the leading HGCAL experts in the US-CMS HL-LHC upgrade project, where he serves as Level-3 Manager for Modules for the endcap upgrade. He is also leading an effort that uses GNNs for particle reconstruction in the electromagnetic calorimeter (ECAL), developing expertise with the same type of architecture to be used in the context of this project.

Co-PIs Lindsey Gray and Thomas Klijnsma, together with Jan Kieseler, are the leaders of the HGCAL reconstruction project and they led the development of the first GNN for HGCAL reconstruction, which is the stepping stone toward the research described in this proposal. Their direct involvement will not only provide credibility to the project but most importantly it will offer to the postdoc an invaluable source of knowledge, ensuring their success. Under their direction the postdoc will be formally included in the HGCAL reconstruction project.

Furthermore Matteo Cremonesi and Lindsey Gray have a long history of collaborating on a multitude of computing projects. For example, Lindsey Gray was the other co-founder of COFFEA and the current Project Manager. He also started with Matteo Cremonesi the effort on the development of GNNs for MET reconstruction, defining the model architecture and being involved as a consultant. The postdoc will certainly benefit from this well established collaboration that has always resulted in the creation of a lively group dynamic.

The HGCAL reconstruction is reviewed by the CMS ML4RECO group. The postdoc is expected to attend the ML4RECO meetings and to frequently report progress. Besides taking part in official CMS meetings, the postdoc will meet with the PIs at least on a weekly basis. In these meetings the discussion will also be focused on career planning. At CMU, postdocs have access to multiple resources devoted to career development, both in academia and in industry. The local diversity and inclusion (DEI) committee has just started an initiative to eliminate barriers faced by students and postdocs when approaching a non-academic job search, promoting the creation of a network of alumni who successfully transitioned to an industry career. The postdoc will be included in this initiative. At the same time, the PIs will work with the postdoc in identifying key conferences and meetings where to participate and ensure the postdoc receives excellent visibility both locally and across the scientific community, in the US and internationally, which is key to building a successful academic career. Finally, CMU provides opportunities to participate in and lead public outreach events, which are important for both career paths.

5 Conclusions

The HGCAL is the single most important project for the HL-LHC CMS detector upgrade. The ability to reconstruct particles inside the HGCAL with high precision is key to fully take advantage of the capabilities of the new detector and have an impact on the physics reach. At the same time, with the demanding HL-LHC pile-up conditions it will be challenging for particle-shower reconstruction algorithms to satisfy computing constraints while exploiting the full physics potential of the improved detector technology. Reconstructing particles with GNNs can provide the ideal solution. In this proposal, support for 50% of the remuneration of a postdoc is asked in order to develop GNNs for the HGCAL reconstruction that work with 200 pile-up. The postdoc will build on top of the work already done in this direction. They will also rely on the guidance of the PIs, who are all experts in GNNs for particle reconstruction (including HGCAL reconstruction), and on the resources offered by CMU, which is a HGCAL module assembly center and ranks among the best US schools in ML/AI.

References

[1] D. Contardo, M. Klute, J. . Mans, L. Silvestris, and J. Butler, Technical Proposal for the Phase-II Upgrade of the CMS Detector (2015).

[2] CMS Collaboration, The Phase-2 Upgrade of the CMS Endcap Calorimeter (2017).

[3] S Bhattacharya *et al.*, J. Phys.: Conf. Ser. **2438** 012090 (2023)

[4] J. Kieseler, Eur. Phys. J. C **80**, 886 (2020).

[5] S. R. Qasim, K. Long, J. Kieseler, M. Pierini and R. Nawaz, 2021 EPJ Web Conf. 251 03072

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

[7] J. Allison *et al.*, Nucl. Instrum. Meth. A **835**, 186 (2016)

[8] J. Platt & A. Barr (1987). "Constrained differential optimization." In Neural Information Processing Systems.

[9] P. Calafiura, "End-to-End Geometric Representation Learning for Track Reconstruction." Presentation at CHEP2023.