Machine-learning-based energy reconstruction for hadronic showers

Tianyi Yang[a], Yuexin Wang[b,c] and Manqi Ruan[b,c,*]

[a]Department of Physics, University of Science and Technology of China, Hefei, 230026, Anhui, China
[b]Institute of High Energy Physics, Chinese Academy of Sciences, Beijing, 100049, China
[c]University of Chinese Academy of Sciences(UCAS), Beijing, 100049, China

ARTICLE INFO

Keywords:
Hadronic energy regression
ParticleNet
Graph neural networks
Hit-level timing information
High-granularity calorimeters

ABSTRACT

Accurate hadronic energy reconstruction is essential for hadron calorimetry and plays a critical role in physics measurements at future Higgs factories, especially for final states involving jets. In this work, we develop a holistic machine-learning-based approach that reads in the hit-level information of high-granularity calorimeter for single-hadron energy reconstruction. Compared to the conventional method, this new holistic approach yields a 50 - 100% improvement on the energy resolution, while maintaining the same level of linearity over a fairly wide energy range. We also report the scaling behaviour of the regression performance with respect to training statistics and training epochs, which provides guidance for estimating the potential performance gain from larger samples and more complete training. Synchronized with the trend of 5-D calorimetry development, we also analyze the impact of hit-level timing information on the energy measurements. These results demonstrate that timing resolution below the $\mathcal{O}$(100 ps) scale achieves optimal results with ideal timing for hadron energy regression.

1. Introduction

Calorimetry is an essential component of collider detectors, providing the spatial and energetic information to reconstruct physics objects, including individual final-state particles such as photons and charged leptons, and composite objects such as hadronic tau candidates and jets [1, 2]. At the high-energy frontier, and especially at future Higgs factories, calorimetry is particularly important because the physics program focuses on precision measurements of the Z, W, Higgs bosons and top quarks. These massive Standard Model particles have extensive branching fractions into hadronic final states, making precise jet and hadronic-system reconstruction central to the physics performance [3, 4, 5]. In particle-flow-algorithm (PFA)-oriented detector concepts, high-granularity calorimetry is used not only to measure energy, but also to image and separate showers produced by individual incident particles, so that each particle can be measured in the most suitable subdetector system [1, 6]. This strategy enables high-precision reconstruction of hadronic systems, while the accuracy of hadronic-shower energy measurement remains a major contribution to the overall jet-energy resolution (JER) and boson mass resolution (BMR) [4, 6]. Therefore, improving hadronic energy resolution, a core figure of merit of calorimeter performance, is not only of intrinsic interest to calorimetry, but also has direct impact on the core physics program of future high-granularity collider detectors.

However, precise determination of the incident hadron energy from calorimeter signals remains intrinsically challenging. Hadronic showers involve coupled electromagnetic and hadronic components, invisible energy losses, delayed activity, leakage, and detector-response effects [2, 7, 8]. As a result, similar visible energy deposits may correspond to different underlying shower configurations, making the mapping from detector signal to true hadron energy non-linear and event dependent. This limits the precision of conventional energy-sum (ESUM)-based reconstruction: global calibration can correct the average response, but it cannot fully recover event-by-event shower fluctuations or topology-dependent invisible energy losses [9]. In contrast, high-granularity, PFA-oriented calorimeters provide richer spatial, energetic, and timing information that can be exploited for particle identification and incident-energy reconstruction.

Machine-learning methods are well suited to this task because they can learn non-linear, topology-dependent corrections from the full calorimeter response. In particular, ParticleNet treats detector hits as an unordered point cloud and uses dynamic graph convolutions to extract local and global shower features directly from the detector geometry [10, 11, 12]. This makes it a suitable architecture for high-granularity hadronic energy regression, where the spatial, energetic, detector-region, and timing information of all hits can be exploited simultaneously.

We propose a holistic approach to determine the hadronic shower energy, and realize it using ParticleNet at simulated samples with a benchmark detectors for the Circular Electron Positron Collider (CEPC), one of the proposed future Higgs factory. [13] The idea is inspired by the "learning from all particles" strategy, where reconstruction benefits from using complete low-level event information rather than heavily compressed high-level summaries [14]. Here, "holistic" means that the model reads the complete hit-level calorimeter response of a shower, including hit position, deposited energy, detector region, and timing information, and directly regresses the incident hadron energy from this full event representation. Its performance is compared with a globally recalibrated ESUM baseline in terms of energy

*Corresponding author
ruanmq@ihep.ac.cn (M. Ruan)
ORCID(s): 0000-0001-7553-9236 (M. Ruan)

resolution, response linearity, training stability, particle-species dependence, and timing dependence. The results show that the holistic approach can significantly improve hadronic energy resolution while maintaining controlled response linearity, demonstrating the potential of hit-level machine learning for precision calorimetry at future collider experiments.

2. Detector Concept and Dataset

2.1. High-granularity calorimeter reference design

This study is based on the calorimeter system of AURORA, a novel detector concept developed from the CEPC Conceptual Design Report (CDR) baseline detector [4, 13]. AURORA is designed for the goal of approaching a one-to-one correspondence between reconstructed particles and visible final-state particles. This objective is closely related to the particle-flow paradigm, in which charged particles are primarily measured by the tracking system, while photons and neutral hadrons are reconstructed from calorimeter information. Therefore, the calorimeter is required not only to measure deposited energy, but also to provide sufficiently detailed spatial and temporal information for particle separation, shower identification, and confusion reduction.

A key feature of AURORA is its use of high-granularity five-dimensional (5-D) calorimetry. In addition to recording the three-dimensional position and deposited energy of each calorimeter hit, timing information, or time-of-flight-related information, is also preserved at the hit level [13, 15]. Combined with advanced reconstruction and machine-learning algorithms, such a detector concept is expected to improve particle identification and jet reconstruction performance. In particular, the AURORA detector has demonstrated a BMR at the level of about 2.7%, providing a representative benchmark for future high-precision collider detector designs [4]. A schematic cross-section of the detector and the calorimeter layout is shown in Fig. 1.

The calorimeter system used in this work consists of an electromagnetic calorimeter (ECAL) followed by a hadronic calorimeter (HCAL), both implemented as high-granularity sampling calorimeters. The ECAL is a silicon–tungsten sampling calorimeter (SiWECAL), where silicon sensors provide fine transverse segmentation and tungsten plates serve as absorbers. In the geometry considered here, the ECAL has a transverse cell size of about 1.0 cm, corresponding to 10.17 mm in this detector model, and contains 29 sampling layers. The total tungsten absorber thickness is about 80 mm, corresponding to roughly 23 X_0(radiation length) in absorber depth. This fine granularity allows the ECAL to resolve compact electromagnetic subshowers and to provide detailed imaging information for early shower development.

The HCAL is a glass–steel sampling calorimeter (GSHCAL), designed to measure the hadronic component of the shower after the ECAL. It uses approximately 2×2 cm^2 transverse cells and 48 sampling layers, giving a total depth of about $6\,\lambda_I$ (nuclear interaction length). This depth is sufficient to contain a large fraction of hadronic showers

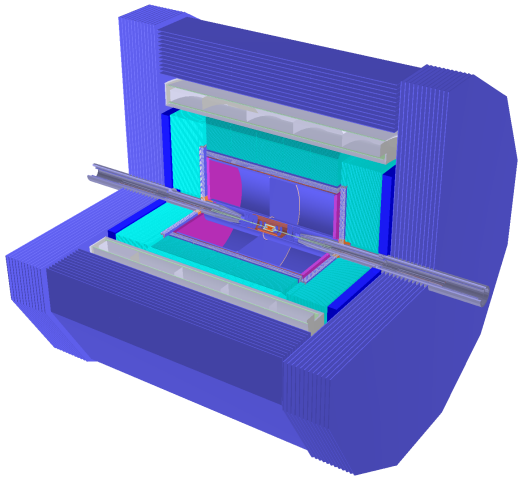


Figure 1: Schematic diagram of AURORA cross-section [4]

in the energy range considered in this study, while the fine segmentation preserves the transverse and longitudinal structure of the shower. Such information is essential for distinguishing compact electromagnetic components, diffuse hadronic activity, delayed components, and possible leakage-related patterns.

To align with the ongoing development of 5-D calorimetry [16, 17, 18], hit-level timing information is included in the detector response. Unless otherwise specified, a timing resolution of 100 ps is adopted as the realistic benchmark configuration in this work. This time resolution allows for the study of the performance of different methods on energy regression under relatively realistic conditions.

2.2. Simulated samples

The training and evaluation of the regression models are based on fully simulated single-hadron samples produced with the AURORA detector model. Single particles are generated with a particle-gun setup and propagated through the detector using a GEANT4-based full-simulation framework [19]. The reconstructed calorimeter hits are then used as the input to the regression model.

Unless otherwise stated, the π^+ sample is used as the primary benchmark, since charged pions are among the dominant constituents of jets and hadronic final states. The study is also repeated for K_L^0 and neutron samples to check the consistency of the regression strategy for different hadronic shower types. In all cases, separate ParticleNet regressors are trained with the same architecture and training configuration.

Two types of samples are used in this study. The first is an inclusive momentum-scan sample, where the incident momentum is randomly sampled in the range of 0–150 GeV. The particle-gun direction is randomized with uniform angular smearing in both θ and ϕ. For each particle species, this inclusive sample contains approximately 1×10^6 events and is used for model training and validation. It is split into training and validation subsets with a ratio of 8:2. This

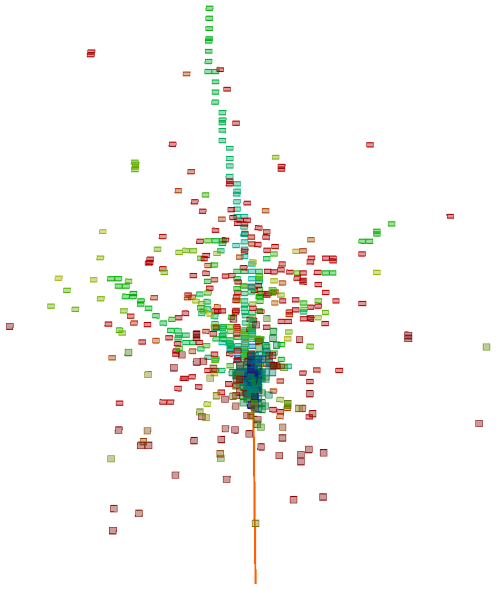


Figure 2: The energy deposition state of a π^+ at 30 GeV event within a calorimeter. The hits are color-coded by the time residual $T = T_0 - \mathrm{DisToVTX}/c$, where T_0 is the original reconstructed hit time, DisToVTX is the distance from the interaction vertex to the hit position, and c is the speed of light.

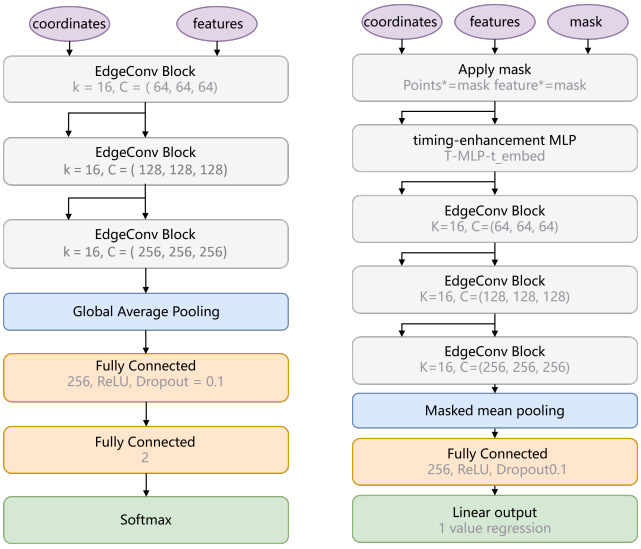


Figure 3: Comparison of (a) classification-head ParticleNet [10] and (b) the regression-head ParticleNet architecture used in this work.

sample provides broad coverage of the relevant momentum and angular phase space.

The second type consists of independent mono-momentum samples generated at several fixed momentum points,

$$p = 1,\ 2,\ 3,\ 5,\ 7,\ 10,\ 20,\ 30,\ 40,\ 50,\ 60,\ 70,\ 80,\ 100,\ 120,\ 140,\ 150,\ 160\ \mathrm{GeV}^{1}. \quad (1)$$

For each particle species and each fixed momentum point, approximately 1×10^4 events are generated. These mono-momentum samples are used exclusively as test samples. They allow a controlled evaluation of the momentum dependence of the energy resolution, response linearity, and possible non-Gaussian tails.

Momentum, rather than kinetic energy, is used as the scan variable in order to keep a consistent kinematic definition across different particle species. This choice facilitates direct comparisons among charged pions, neutral kaons, and neutrons.

3. Methodology

In this section, the hadronic energy regression problem is formulated and the reconstruction methods used in this study are described. A baseline method is introduced first, followed by the ParticleNet-based regression model.

[1]The points at 150 GeV and 160 GeV are included to probe the model behaviour near the upper boundary of the nominal training range and in the slightly extrapolated region beyond it.

3.1. Baseline method

The standard energy reconstruction method of a hadronic shower is obtained by summing the reconstructed energy deposits in all calorimeter cells,

$$E_{\mathrm{ESUM}} = \sum_{i \in \mathrm{hits}} E_i\,, \quad (2)$$

where E_i denotes the calibrated energy of the i-th cell [17]. In the case of a sampling calorimeter, the same global calibration constants as those used in the detector design study are applied. In this work, an additional global re-calibration is further introduced for the ESUM baseline. Specifically, after forming the summed calorimetric response, a single overall multiplicative correction factor is applied,

$$E_{\mathrm{ESUM_{re\text{-}calib}}} = \alpha\, E_{\mathrm{ESUM}}, \quad (3)$$

where the factor α is determined from calibration samples such that the overall response bias is minimized, i.e. the reconstructed-to-true energy ratio is brought as close as possible to unity over the studied energy range. Therefore, the "re-calib" reference used throughout this paper corresponds to a globally rescaled ESUM estimator rather than a more elaborate event-by-event correction scheme. It is used as a simple but transparent reference for quantifying the gain from hit-level topology learning. This reference still reflects the basic calorimetric response, but it does not explicitly correct for non-compensation, invisible energy, longitudinal leakage, or other topology-dependent shower fluctuations beyond the global normalization adjustment [9].

3.2. ParticleNet-based regression model

The machine-learning-based hadronic energy estimator is built using the ParticleNet architecture, adapted from its

original application in jet tagging to the calorimetric regression task [10]. The objective of the regression model is to use the full hit-level calorimeter information of a single-hadron shower to infer one definitive numerical value, namely the incident hadron energy, with minimal uncertainty with respect to the Monte Carlo truth label.

For each event, the calorimeter response is treated as an unordered set of hits. Each hit is described by a feature vector

$$\mathbf{f}_i = (x_i, y_i, z_i, E_i, t_i, S_i), \tag{4}$$

where (x_i, y_i, z_i) are the hit positions, E_i is the deposited energy, t_i is the hit time, and S_i is a binary subdetector label indicating whether the hit belongs to the ECAL or HCAL. Before being passed to the network, the input features are normalized according to the distributions obtained from the training sample, so that different feature channels have comparable numerical scales and the training procedure remains stable. The detailed preprocessing procedure is given in Appendix A.1.

ParticleNet operates directly on this hit-level point cloud. Through dynamic graph convolution layers, it learns local geometric correlations and global shower-topology features from the calorimeter hits. In this work, the original classification head is replaced by a regression head, so that the network output is a continuous estimate of the incident hadron energy. A schematic comparison between the original classification-head ParticleNet and the regression-head ParticleNet used in this work is shown in Fig. 3.

The regression mapping can therefore be written as

$$\mathcal{F}_{\mathrm{PN}} : \{(x_i, y_i, z_i, E_i, t_i, S_i)\}_{i=1}^{N_{\mathrm{hit}}} \longrightarrow \hat{E}, \tag{5}$$

where N_{hit} is the number of valid calorimeter hits in the event and $\hat{E}$ is the predicted incident energy. The training target is the corresponding Monte Carlo truth energy, denoted by E_{Truth}.

3.3. Evaluation metrics

For performance evaluation, the energy distribution is studied in bins of true momentum and particle type. In each bin, the central 90% interval of the reconstructed-energy distribution is used to measure energy resolution. The quantity mean90 is defined as the mean reconstructed energy within this selected central 90% interval, while RMS90 denotes the root-mean-square spread of the reconstructed energy around mean90 within the same interval. In this work, the energy resolution is defined as

$$\mathrm{Res.} = \frac{\mathrm{RMS90}}{\mathrm{mean90}}. \tag{6}$$

The relative improvement with respect to the recalibrated ESUM baseline is defined as

$$\mathrm{Impr.} = \left(\frac{(\mathrm{Res.})_{\mathrm{ESUM_{re\text{-}calib}}}}{(\mathrm{Res.})_{\mathrm{PN_{setup}}}} - 1 \right) \times 100\%, \tag{7}$$

where "setup" denotes the ParticleNet-based regressor under a given input or timing configuration. With this definition, a positive value indicates a reduction of the energy resolution relative to the recalibrated ESUM baseline. For example, Impr. = 100% means that the ParticleNet resolution is reduced to one half of the ESUM baseline resolution.

The response linearity is characterized by the relative deviation of the central reconstructed energy from the true energy,

$$\frac{\Delta E}{E} = \left(\frac{\mathrm{mean90} - E_{\mathrm{Truth}}}{E_{\mathrm{Truth}}} \right) \times 100\%. \tag{8}$$

A value of zero therefore corresponds to perfect linearity, while positive and negative values indicate over-response and under-response, respectively.

These metrics are computed in a consistent manner for the ESUM and holistic approach, enabling a direct and quantitative comparison of the regression performance.

4. Results

In this section, the performance of the ParticleNet-based hadronic energy regression is presented and compared with the baseline method introduced in Sec. 3.1. Unless stated otherwise, the timing-enabled ParticleNet results correspond to a representative timing resolution at the 100 ps; the impact of timing information and timing-resolution degradation is discussed separately in Sec. 4.4.

4.1. Overall regression performance

Fig. 4 presents the reconstructed energy distributions for the π^+ sample at four representative momentum points, namely 7, 30, 70, and 120 GeV. At all momenta, the ParticleNet-based regressor produces visibly narrower and more symmetric distributions than the recalibrated ESUM baseline, indicating a clear suppression of event-by-event shower fluctuations. This effect is already evident at 7 GeV, where the baseline distribution remains relatively broad, and becomes even more pronounced at 30, 70, and 120 GeV, where the ParticleNet predictions show substantially improved peak sharpness and reduced tail populations.

These distribution-level observations are consistent with the momentum-dependent performance trends summarized in Fig. 5. The visible narrowing of the reconstructed energy spectra at representative momenta translates directly into a lower Res., a sizable improvement relative to the recalibrated $\mathrm{ESUM_{re\text{-}calib}}$ baseline, and a reasonably stable response over the full momentum range.

The regression performance of the ParticleNet model is compared with the recalibrated ESUM baseline, $\mathrm{ESUM_{re\text{-}calib}}$. Fig. 5a shows the energy resolution, quantified by Res., as a function of the incident momentum. The ParticleNet estimator outperforms the recalibrated $\mathrm{ESUM_{re\text{-}calib}}$ baseline over the full momentum range considered. At 1 GeV, Res. is reduced from about 26% for $\mathrm{ESUM_{re\text{-}calib}}$ to about 13% for ParticleNet. With increasing momentum, the ParticleNet resolution improves further, reaching about 5.6% at 20 GeV and about 3.2% at 150 GeV, while the recalibrated $\mathrm{ESUM_{re\text{-}calib}}$ baseline remains at the level of about 7–8%

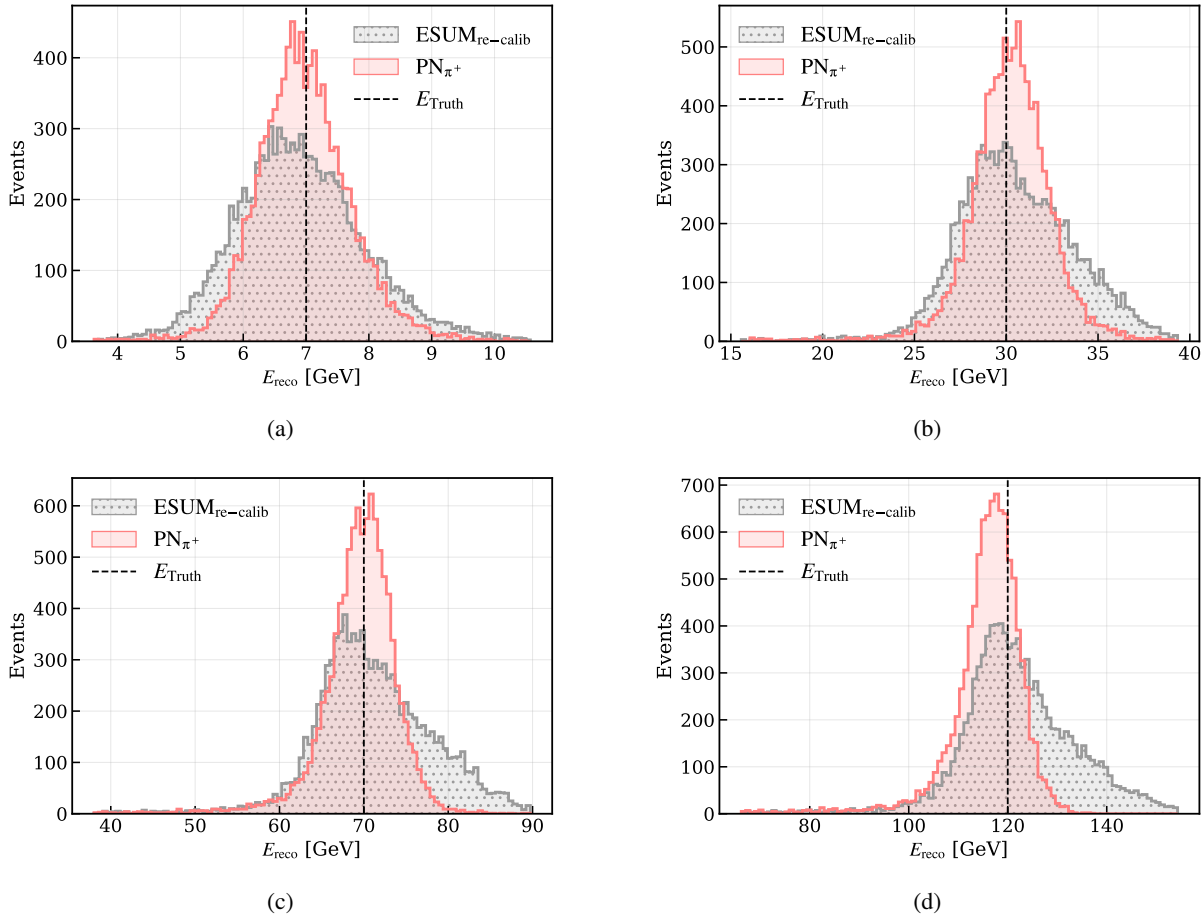


Figure 4: Energy distribution for π^+ at (*a*)7 GeV, (*b*)30 GeV, (*c*)70 GeV, (*d*)120 GeV. The vertical dashed line indicates the true particle energy.

over most of the high-momentum region. This behavior indicates that the network is able to exploit shower-topology information beyond what is accessible to a simple summed-energy estimator.

The upper panel of Fig. 5b presents Impr. with respect to ESUM$_{\text{re-calib}}$. A substantial gain is observed over the entire momentum range. The improvement is already close to 100% at 1 GeV, remains at the level of about 50–70% over the few-GeV to few-tens-of-GeV region, and then increases again at higher momentum, reaching about 78% around 60–120 GeV and exceeding 100% at 150 GeV. This trend suggests that the ParticleNet regressor is effective both in suppressing the large stochastic fluctuations at low energy and in capturing the topology-dependent corrections that become increasingly important at higher energy.

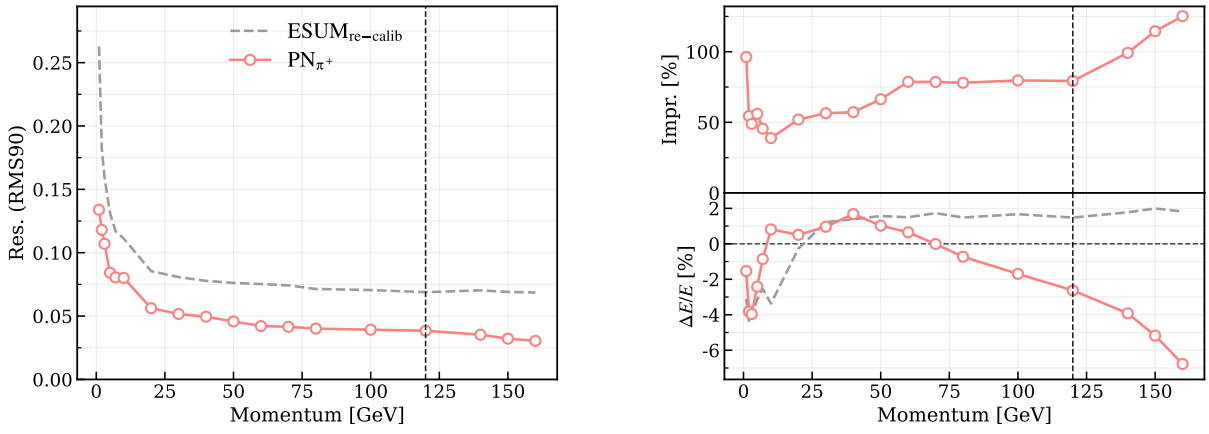
The lower panel of Fig. 5b shows the response linearity, expressed as $\Delta E/E$. Up to the 120 GeV benchmark scale marked by the vertical dashed line, the response bias remains within about 5%. This indicates that, compared with a simple globally rescaled calorimetric baseline, the learned regressor not only provides a better resolution, but also preserves reasonably well-controlled linearity.

It is also worth noting that the training range of the model is 0–150 GeV, while the benchmark evaluation still includes the highest points at 150 GeV and even 160 GeV. The fact that the regression performance remains at a comparatively good level at these upper-end test points suggests that the model has a certain degree of generalization capability, including limited extrapolation beyond the nominal training range.

4.2. Training-statistics and epoch scaling

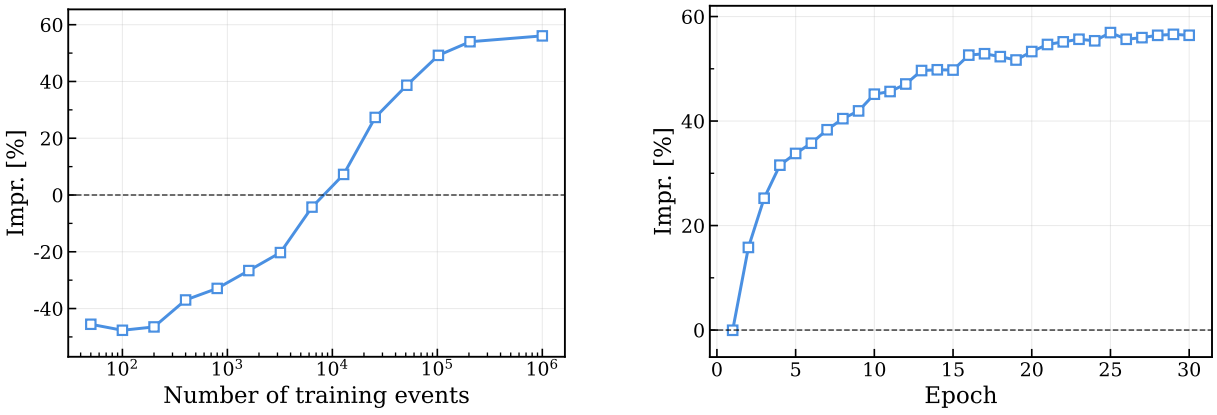
The training-statistics and epoch dependence of the regression performance are summarized in Fig. 6 for the π^+ sample. Such scaling studies are useful for checking whether a topology-learning model is limited by training statistics, optimization time, or model capacity [20].

The left panel shows the dependence of Impr. on the number of training events. When only a small number of events is used, the learned regressor does not yet outperform the ESUM baseline, indicating that the topology-dependent correction cannot be reliably learned from very limited statistics. As the training statistics increase, the improvement rises monotonically and then gradually approaches a plateau. This trend suggests that the model continues to benefit from larger samples, while the nominal training statistics at the



(a) (b)

Figure 5: (a) Res. and (b) Impr. (upper) and $\Delta E/E$ (lower) versus true momentum for π^+. The vertical dashed line marks 120 GeV, which is taken as the representative upper energy scale for typical CEPC benchmark studies.



(a) (b)

Figure 6: Dependence of the π^+ regression performance on (a) the number of training events and (b) the training epoch.

level of 10^6 events already lie close to the saturation region for the present configuration.

The right panel shows the dependence of Impr. on the training epoch. The improvement increases rapidly during the first few epochs and then evolves more slowly toward a stable value. After about 20–30 epochs, the performance becomes nearly saturated, and no significant degradation is observed at later epochs. Therefore, the nominal choice of 30 training epochs is sufficient for the present setup and does not show an obvious sign of early overfitting.

These two scaling studies provide a useful consistency check for the regression results. They indicate that the adopted training sample size, at the 10^6 level, and the nominal 30-epoch training schedule are reasonable choices under the current model and dataset configuration.

4.3. Dependence on particle type

The performance on additional particle species is shown in Fig. 7, where separate ParticleNet regressors are trained and evaluated for neutrons and K_L^0 using the same architecture and training configuration as the charged-pion benchmark.

For neutrons, shown in Figs. 7a and 7b, the ParticleNet regressor gives an even stronger improvement over the conventional ESUM$_{\text{re-calib}}$ baseline in terms of energy resolution. For momenta above about 3 GeV, the performance is broadly similar to the charged-pion case, with a clear reduction of Res. and a relative improvement typically at the level of 50–80%. In the very-low-momentum region below about 3 GeV, the relative improvement becomes particularly large, reaching up to roughly a factor of three compared with the ESUM baseline. However, this large resolution gain is accompanied by a sizable response bias, and the sign of the bias is opposite to that of the conventional energy-sum estimator. A plausible reason is that, at such low momenta, the neutron rest mass and threshold-related effects are no longer negligible compared with the visible shower energy,

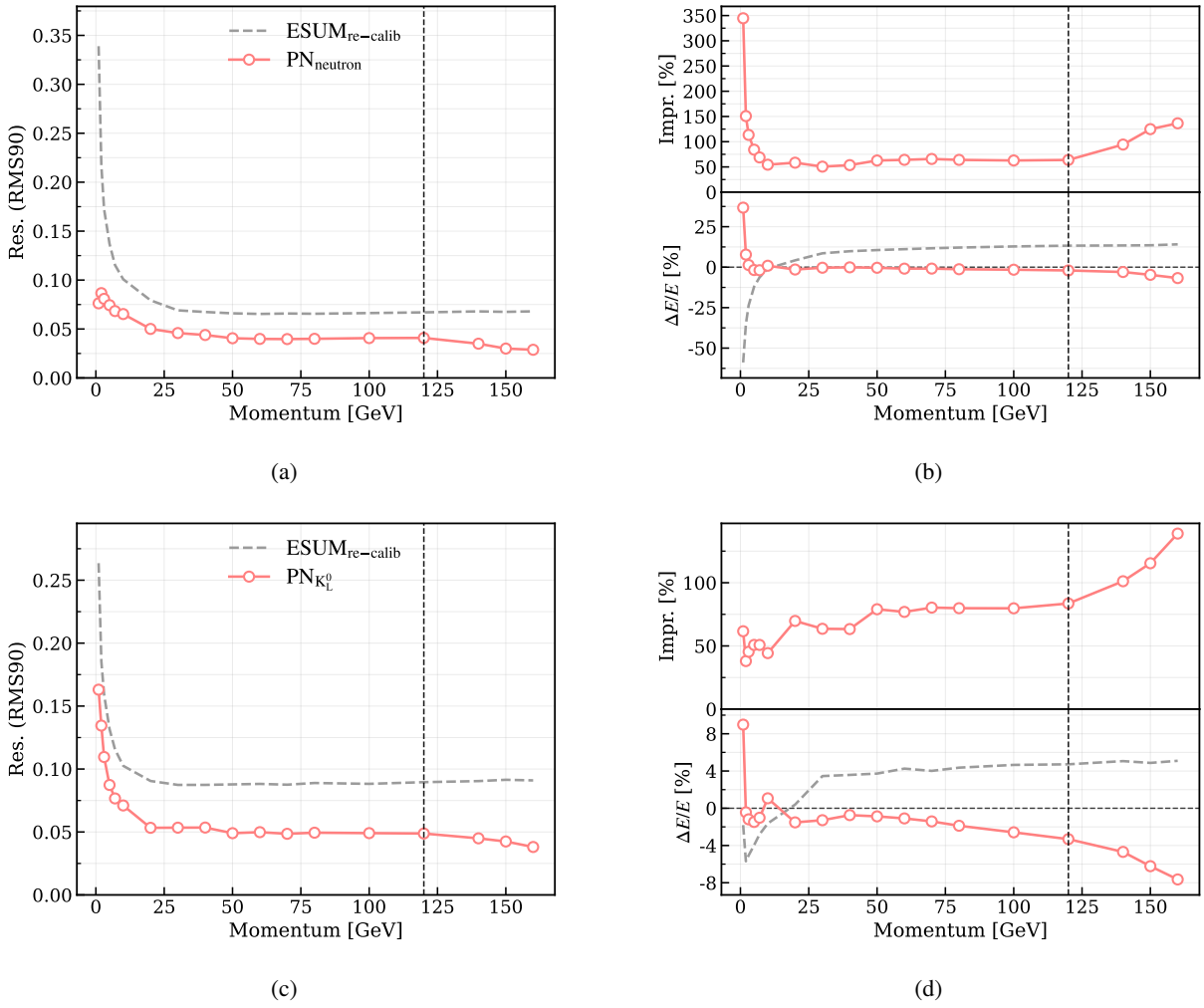


Figure 7: Dependence of the regression performance on particle type. (a) Res. versus true momentum for neutron; (b) Impr. (upper) and $\Delta E/E$ (lower) versus true momentum for neutron; (c) Res. versus true momentum for K^0_L; (d) Impr. (upper) and $\Delta E/E$ (lower) versus true momentum for K^0_L.

making the relation between incident momentum, deposited energy, and reconstructed response more nonlinear.

For K^0_L, as shown in Figs. 7c and 7d, the ParticleNet estimator also consistently improves the energy resolution relative to ESUM$_{\text{re-calib}}$. The improvement remains substantial over most of the tested momentum range, with a magnitude comparable to that observed for charged pions and neutrons above the very-low-momentum region. The response linearity is also generally controlled, although some residual bias remains at high momenta.

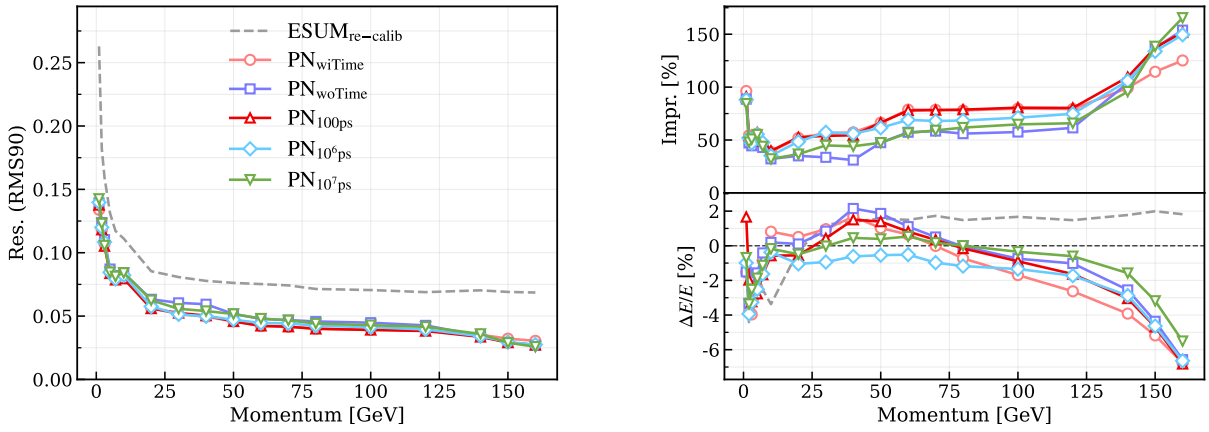
Overall, these results indicate that the same hit-level regression strategy remains effective for neutral hadrons. The strongest improvement is observed for neutrons, especially at low momentum, while the remaining response-bias behaviour suggests that very-low-energy neutral-hadron reconstruction may require additional calibration or dedicated treatment.

4.4. Impact of timing measurements

To fulfill the physics requirements of time-of-flight (ToF) measurements, which are essential for particle identification, AURORA adopts a 5-D calorimetry concept in which the time of each calorimeter hit is recorded together with its spatial position and deposited energy. As illustrated in Fig. 2, hadronic showers exhibit a complex hit-time structure in the calorimeter. This motivates the use of hit-level timing information not only for ToF-related particle identification, but also as a possible additional input for hadronic energy reconstruction.

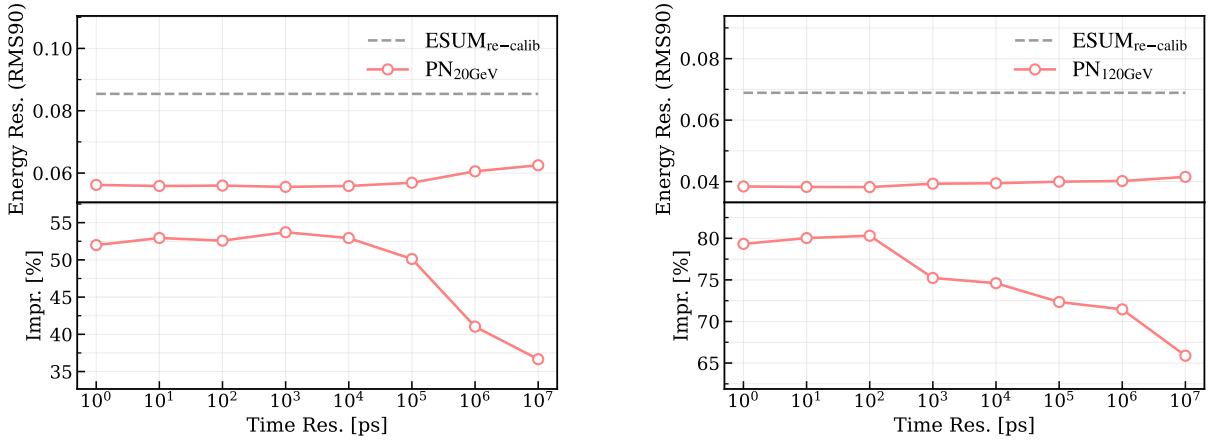
In this study, we therefore compare ParticleNet-based regressors trained with and without hit-time information. We further examine the effect of finite timing resolution by applying different timing-smearing configurations, in order to quantify how sensitive the energy-regression performance is to the precision of the time measurement.

Fig. 8 first shows that the availability of timing information itself is important. Compared with PN$_{\text{woTime}}$, the



(a) (b)

Figure 8: Momentum dependence of the π^+ energy-regression performance under different timing configurations. (a) Energy resolution, defined as Res. = RMS90/mean90, and (b) relative improvement Impr. and response bias $\Delta E/E$ are shown as functions of the true momentum. The curve labeled ESUM$_{\text{re-calib}}$ denotes the globally recalibrated ESUM baseline. PN$_{\text{wiTime}}$ denotes the ParticleNet configuration with timing information included, while PN$_{\text{woTime}}$ denotes the configuration in which the timing channel is removed. PN$_{100\,\text{ps}}$, PN$_{10^6\,\text{ps}}$, and PN$_{10^7\,\text{ps}}$ denote ParticleNet configurations with finite or strongly degraded hit-time resolutions, respectively.



(a) (b)

Figure 9: Dependence of the ParticleNet regression performance on timing resolution for π^+ at fixed momenta: (a) 20 GeV and (b) 120 GeV. The comparison illustrates that the high-energy region is more sensitive to the degradation of timing resolution.

timing-enabled ParticleNet configurations generally achieve lower Res. and larger Impr. over a broad momentum range. This demonstrates that the network does not rely only on the spatial and energy-deposition pattern; it also extracts useful information from the time structure of the shower. The main conclusion from this comparison is that timing information provides a sizable and physically meaningful contribution to hadronic energy regression.

A second question is how precise the timing measurement needs to be. The curves with finite timing smearing in Fig. 8 indicate that the performance is best preserved when the timing resolution remains around the $\mathcal{O}$(100 ps) scale. The PN$_{100\text{ps}}$ curve stays close to the timing-enabled reference and consistently outperforms the cases with severely degraded timing information. When the timing resolution is degraded to extremely coarse values, such as 10^6–10^7 ps, the timing variable loses much of its discriminating power and the performance approaches that of a model with little useful timing information.

The timing-resolution scan in Fig. 9 further shows that this dependence is energy dependent. At 20 GeV, the degradation of timing resolution changes the energy resolution and relative improvement only moderately, indicating that the low- and intermediate-energy performance is less sensitive to the exact timing precision once timing information is

available. By contrast, at 120 GeV, the performance changes more visibly as the timing resolution becomes worse: the energy resolution deteriorates and the improvement over the ESUM baseline decreases. This suggests that high-energy hadronic showers are more sensitive to the timing resolution than low-energy showers.

This behaviour is physically reasonable. At higher energies, hadronic showers are typically more extended and contain richer late-time components, so the time profile carries more information about the separation between prompt electromagnetic subshowers and delayed hadronic activity. If the timing resolution is too coarse, this separation is blurred, and the network loses part of the information needed for topology-dependent energy correction. At lower energies, the shower is more compact and the reconstruction is more strongly affected by stochastic fluctuations and threshold effects, so the exact timing precision plays a comparatively smaller role.

5. Discussion

5.1. Impact on BMR

Although the present study is performed at the isolated single-hadron level, its implications extend naturally to jet-level reconstruction at future Higgs factories. In particle-flow-oriented detectors, hadrons carry a large fraction of the visible jet energy, while neutral hadrons remain one of the main limitations of calorimeter-based reconstruction [1, 6]. Therefore, an improvement in single-hadron energy resolution can potentially contribute to better jet-energy resolution and BMR.

As a rough estimate, taking the CEPC/AURORA BMR of about 2.7% as a reference [4], the level of single-hadron resolution improvement observed in this work could in principle correspond to a BMR reduction to around 2.2%–2.3%, or an improvement at the level of roughly 10%–15%. This estimate should not be regarded as a validated end-to-end result, since the final jet-level performance also depends on jet clustering, particle-flow confusion, neutral-hadron identification, shower overlap, detector acceptance, and reconstruction details [1]. Nevertheless, the result indicates that topology-aware hadronic energy regression may provide a useful ingredient for improving the global performance of high-granularity calorimeter systems.

Such an improvement is also relevant for Higgs physics. Better hadronic reconstruction can enhance the separation of hadronic W, Z, and Higgs decay modes, improve the stability of multi-jet reconstruction, and reduce migrations among different hadronic final-state categories [3, 13, 15]. The present study should therefore be viewed as a first step toward evaluating how machine-learning-based hadronic calorimetry may contribute to precision Higgs and electroweak measurements.

5.2. Future work

The present study provides a proof of principle for a holistic hadronic-energy reconstruction approach that learns directly from all available calorimeter hits. It demonstrates that hit-level spatial, energy, detector-region, and timing information can be combined by a topology-aware regression model to improve single-hadron energy resolution. Meanwhile, the current analysis is still based on isolated single-particle samples and simulation-only detector responses. Further studies are therefore needed before its full detector- and physics-level impact can be established.

Several directions are particularly important:

- **Test-beam data.** The approach should be validated with test-beam data and tested with training–test samples generated using different Geant4 physics lists, detector-response models, and calibration assumptions. These studies are needed to quantify possible data–Monte Carlo discrepancies and assess the robustness of the learned correction [8, 21].
- **Detector-geometry optimization.** Since the regression performance depends on the available shower information, geometry scans of segmentation, absorber depth, ECAL–HCAL configuration, and timing capability could help guide high-granularity calorimeter design [1, 6, 18].
- **ToF reconstruction.** Dedicated studies of time-of-flight-related reconstruction are needed beyond using hit time as an auxiliary input, especially for particle identification and delayed hadronic shower components [4, 16, 17, 22].
- **Architecture optimization.** Alternative graph-based, transformer-based, equivariant, or lightweight architectures should be compared with ParticleNet, while model compression and knowledge distillation may improve inference efficiency [7, 10, 11, 12, 23].

Finally, the approach should be embedded into a complete particle-flow reconstruction chain. Only an end-to-end study with realistic jet events and benchmark physics channels can determine how the single-hadron improvement propagates to jet-level observables, BMR, and final physics sensitivities [1, 6].

6. Summary

A holistic approach for single-hadron energy reconstruction has been studied for a high-granularity calorimeter using hit-level calorimeter information from the standard simulation and reconstruction chain. The calorimeter response is represented as a hit-level point cloud, and the network is trained to regress the incident hadron energy directly.

Compared with the conventional globally recalibrated ESUM baseline, the proposed holistic approach improves the hadronic-shower energy resolution by about 50–100% over a broad momentum range, while maintaining response linearity at a comparable level, typically within $\mathcal{O}(1\%)$ to a few percent. This demonstrates that the network can exploit hit-level shower information beyond a simple energy sum without introducing a significant response bias. In addition,

the scaling behaviour of the performance improvement with respect to training statistics and training epochs is reported. These studies provide useful diagnostics for monitoring the machine-learning behaviour, checking whether the model is limited by statistics or training time, and assessing the stability of the nominal training configuration.

The timing study shows that timing information is an important input for hadronic energy regression. The difference between timing-enabled and timing-disabled models is significant, indicating that the time structure of hadronic showers provides information beyond spatial topology and deposited energy alone. Moreover, the high-energy region is more sensitive to timing-resolution degradation than the low-energy region. A timing resolution at the $\mathcal{O}$(100 ps) scale is therefore desirable for retaining most of the ParticleNet improvement observed in the ideal-timing configuration.

Overall, the results support the use of geometry-aware machine-learning methods for precision hadronic calorimetry in high-granularity detectors. They also indicate that the combination of spatial imaging and realistic timing information provides a suitable basis for further developments toward more complete event reconstruction.

Acknowledgements

This work was supported by National Key Program for S&T Research and Development under contract number 2024YFA1610603 and and the National Natural Science Foundationof China under Grant No.W2441004. The authors would like to thank the CEPC software and detector groups for providing the simulation and reconstruction framework used in this study. The authors also thank Mr. Chao Yang and Mr. Jianfeng Jiang for helpful discussions.

A. Input representation and training setup

A.1. Input representation and truth definition

For each event, the calorimeter response is represented as a variable-length set of calorimeter hits stored in the ROOT ntuples. Each hit i is described by a low-level feature vector

$$\mathbf{f}_i = (x_{i,k}) = (x_i, y_i, z_i, E_i, t_i, S_i), \tag{9}$$

where (x_i, y_i, z_i) denotes the three-dimensional hit position, E_i the deposited energy, t_i the reconstructed hit time, and S_i the subdetector label, indicating whether the hit belongs to the ECAL or HCAL. Thus, each event is treated as a hit-level point cloud rather than as a fixed-size image or a single scalar calorimetric response.

Each feature channel is preprocessed by an affine standardization followed by range clipping to ensure numerical stability. Let $x_{i,k}$ denote the raw value of the k-th feature for the i-th hit. The standardized feature $\widetilde{x}_{i,k}$ is computed as

$$\widetilde{x}_{i,k} = \frac{x_{i,k} - \mu_k}{\sigma_k}, \tag{10}$$

where μ_k and σ_k are the mean and standard deviation of the k-th feature variable, calculated over the training dataset. To suppress pathological outliers, non-coordinate features, such as energy and time, are further clipped to a fixed range $[c_{\min}, c_{\max}]$:

$$x'_{i,k} = \min\big(\max(\widetilde{x}_{i,k},\, c_{\min}),\, c_{\max}\big). \tag{11}$$

For spatial coordinates, no clipping is applied:

$$x'_{i,k} = \widetilde{x}_{i,k}, \tag{12}$$

so that the full geometric extent of the shower is preserved.

Since the number of hits varies from event to event, batching is handled by padding or truncation. For each batch, events are padded with zeros or truncated to a maximum length $N_{\max}$, and a binary mask is constructed to mark valid hits. The network input is organized into a point tensor and a feature tensor,

$$\mathbf{P} \in \mathbb{R}^{B\times 3\times N_{\max}}, \qquad \mathbf{X} \in \mathbb{R}^{B\times (C-3)\times N_{\max}}, \tag{13}$$

together with a mask

$$\mathbf{M} \in \mathbb{R}^{B\times 1\times N_{\max}}, \tag{14}$$

where B is the batch size and C is the total number of per-hit feature channels. This representation preserves the full 3D imaging information and, optionally, the timing information at the hit level, while allowing efficient mini-batch training.

The regression target is read directly from dedicated truth-level branches in the ntuples. In this work, the primary target for hadronic response is defined at generator level as the energy of the primary hadron at the entrance of the calorimeter system, denoted by E_{Truth}.

A.2. Training setup and loss choice

The network is trained with an asymmetric relative Huber loss. For the i-th event, the relative energy residual is defined as

$$r_i = \frac{\hat{E}_i - E_{\mathrm{Truth},i}}{E_{\mathrm{Truth},i} + \epsilon}, \tag{15}$$

where $\hat{E}_i$ is the predicted energy, $E_{\mathrm{Truth},i}$ is the truth energy, and ϵ is a small constant introduced for numerical stability. The Huber threshold is chosen according to the sign of the residual,

$$\delta_i = \begin{cases} \delta_{\mathrm{low}}, & r_i < 0, \\ \delta_{\mathrm{high}}, & r_i \geq 0, \end{cases} \tag{16}$$

with $\delta_{\mathrm{low}} = 0.08$ and $\delta_{\mathrm{high}} = 0.15$ in the default training setup.

The per-event loss is then defined as

$$\rho_{\delta_i}(r_i) = \begin{cases} \dfrac{1}{2} r_i^2, & |r_i| \leq \delta_i, \\ \delta_i \left(|r_i| - \dfrac{1}{2}\delta_i \right), & |r_i| > \delta_i. \end{cases} \tag{17}$$

The loss minimized over a mini-batch of N events is

$$\mathcal{L}_{\text{relHuber}} = \frac{1}{N}\sum_{i=1}^{N}\rho_{\delta_i}(r_i). \tag{18}$$

The Huber loss combines a quadratic behaviour for small residuals with a linear tail for larger deviations, thereby reducing the impact of outliers while preserving sensitivity in the core region [24]. A learning-rate schedule based on the validation loss are employed to mitigate overtraining.

References

[1] Thomson, M., 2009. Particle flow calorimetry and the pandorapfa algorithm. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 611, 25–40. URL: https://www.sciencedirect.com/science/article/pii/S0168900209017264, doi:https://doi.org/10.1016/j.nima.2009.09.009.

[2] Lee, S., Livan, M., Wigmans, R., 2018. On the limits of the hadronic energy resolution of calorimeters. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 882, 148–157. URL: https://www.sciencedirect.com/science/article/pii/S0168900217311762, doi:https://doi.org/10.1016/j.nima.2017.10.087.

[3] Zhao, H., Zhu, Y.F., Fu, C.D., Yu, D., Ruan, M.Q., 2019. The Higgs Signatures at the CEPC CDR Baseline. Chin. Phys. C 43, 023001. doi:10.1088/1674-1137/43/2/023001, arXiv:1806.04992.

[4] Wang, Y., Liang, H., Zhu, Y., Che, Y., Xia, X., Qu, H., Zhou, C., Zhuang, X., Ruan, M., 2025. One-to-one correspondence reconstruction at the electron-positron higgs factory. Computer Physics Communications 314, 109661. URL: https://www.sciencedirect.com/science/article/pii/S0010465525001638, doi:https://doi.org/10.1016/j.cpc.2025.109661.

[5] Benedikt, M., et al., 2019. FCC-ee: The Lepton Collider. Technical Report 2. CERN. Geneva. URL: https://cds.cern.ch/record/2651299, doi:10.1140/epjst/e2019-900045-4.

[6] Sefkow, F., White, A., Kawagoe, K., Pöschl, R., Repond, J., 2016. Experimental Tests of Particle Flow Calorimetry. Rev. Mod. Phys. 88, 015003. doi:10.1103/RevModPhys.88.015003, arXiv:1507.05893.

[7] Akchurin, N., Cowden, C., Damgov, J., Hussain, A., Kunori, S., 2021. On the use of neural networks for energy reconstruction in high-granularity calorimeters. Journal of Instrumentation 16, P12036. URL: https://doi.org/10.1088/1748-0221/16/12/P12036, doi:10.1088/1748-0221/16/12/P12036.

[8] Aamir, M., et al., 2024. Using graph neural networks to reconstruct charged pion showers in the cms high granularity calorimeter. Journal of Instrumentation 19, P11025. URL: https://doi.org/10.1088/1748-0221/19/11/P11025, doi:10.1088/1748-0221/19/11/P11025.

[9] C Adloff, J.B., et al., 2012. Hadronic energy resolution of a highly granular scintillator-steel hadron calorimeter using software compensation techniques. Journal of Instrumentation 7, P09017. URL: https://doi.org/10.1088/1748-0221/7/09/P09017, doi:10.1088/1748-0221/7/09/P09017.

[10] Qu, H., Gouskos, L., 2020. Jet tagging via particle clouds. Physical Review D 101. URL: http://dx.doi.org/10.1103/PhysRevD.101.056019, doi:10.1103/physrevd.101.056019.

[11] Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M., 2019. Dynamic graph cnn for learning on point clouds. ACM Trans. Graph. 38. URL: https://doi.org/10.1145/3326362, doi:10.1145/3326362.

[12] Shlomi, J., Battaglia, P., Vlimant, J.R., 2020. Graph neural networks in particle physics. Machine Learning: Science and Technology 2, 021001. URL: https://doi.org/10.1088/2632-2153/abbf9a, doi:10.1088/2632-2153/abbf9a.

[13] Group, T., 2018. Cepc conceptual design report: Volume 2 - physics and detector doi:10.48550/arXiv.1811.10545.

[14] Zhu, Y., Wang, Y., Liang, H., Che, Y., Wang, H., Zhou, C., Qu, H., Ruan, M., 2025. Learning from all particles in high-energy collisions arXiv:2506.11783.

[15] Adhya, S.P., et al. (CEPC Study Group), 2025. CEPC Technical Design Report - Reference Detector arXiv:2510.05260.

[16] Che, Yuzhi, Boudry, Vincent, Videau, Henri, He, Muchen, Ruan, Manqi, 2023. Cluster time measurement with cepc calorimeter. Eur. Phys. J. C 83, 93. URL: https://doi.org/10.1140/epjc/s10052-023-11221-7, doi:10.1140/epjc/s10052-023-11221-7.

[17] Graf, C., Simon, F., 2022. Time-assisted energy reconstruction in a highly-granular hadronic calorimeter. Journal of Instrumentation 17, P08027. URL: https://doi.org/10.1088/1748-0221/17/08/P08027, doi:10.1088/1748-0221/17/08/P08027.

[18] White, A., et al., 2023. Design, construction and commissioning of a technological prototype of a highly granular sipm-on-tile scintillator-steel hadronic calorimeter. Journal of Instrumentation 18, P11018. URL: https://doi.org/10.1088/1748-0221/18/11/P11018, doi:10.1088/1748-0221/18/11/P11018.

[19] Agostinelli, S., et al. (GEANT4), 2003. GEANT4 - A Simulation Toolkit. Nucl. Instrum. Meth. A 506, 250–303. doi:10.1016/S0168-9002(03)01368-8.

[20] Belayneh, D., Carminati, F., Farbin, A., Hooberman, B., Khattak, G., Liu, M., Liu, J., Olivito, D., Barin Pacela, V., Pierini, M., Schwing, A., Spiropulu, M., Vallecorsa, S., Vlimant, J.R., Wei, W., Zhang, M., 2020. Calorimetry with Deep Learning: Particle Simulation and Reconstruction for Collider Physics. Eur. Phys. J. C 80, 688. URL: https://cds.cern.ch/record/2706000, doi:10.1140/epjc/s10052-020-8251-9, arXiv:1912.06794. 26 pages, 38 figures. Corrected typos and added additional references in v2. Extended Acknowledgements section in v3.

[21] Adloff, C., et al., 2013. Validation of geant4 monte carlo models with a highly granular scintillator-steel hadron calorimeter. Journal of Instrumentation 8, P07005. URL: https://doi.org/10.1088/1748-0221/8/07/P07005, doi:10.1088/1748-0221/8/07/P07005.

[22] Lai, S., et al., 2024. Software compensation for highly granular calorimeters using machine learning. Journal of Instrumentation 19, P04037. URL: https://doi.org/10.1088/1748-0221/19/04/P04037, doi:10.1088/1748-0221/19/04/P04037.

[23] Hinton, G., Vinyals, O., Dean, J., 2015. Distilling the knowledge in a neural network. URL: https://arxiv.org/abs/1503.02531, arXiv:1503.02531.

[24] Meyer, G.P., 2021. An alternative probabilistic interpretation of the huber loss, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5257–5265. doi:10.1109/CVPR46437.2021.00522.