Deep learning techniques for energy clustering in the CMS Electromagnetic Calorimeter

POLINA SIMKINA
ON BEHALF OF THE CMS COLLABORATION

SuperClustering in ECAL

- **Rechits**
  - Reconstructed energy deposits in the PbWO4 crystals of the calorimeter (rechits) left by particles.
  - Rechits are gathered together around the crystal with the highest deposited energy to form clusters.
  - Each cluster represents a single particle.
  - Or several overlapping particles.
  - The energy of the initial particle can be reconstructed from the SuperCluster.

Mustache Algorithm

- The algorithm currently used in CMS for reconstruction of SuperClusters.
- Purely geometrical approach:
  - All the clusters falling into the specified “mustache” shape are considered as part of the SuperCluster. The size of the area depends on energy and position of the seed.
  - “Mustache” shape due to the CMS magnetic field (spread along φ).
- High efficiency: the algorithm is able to gather even low-energy clusters.
- Downside: suffers because of pileup (PU) and noise contamination.
- Energy regression is further applied that can correct PU and noise on average.

Model Architecture

  - Maintains the efficiency while improving PU and noise rejection.
  - Graph NN are able to aggregate the information between the neighbors.

Dataset and Particle ID

- Dataset for the training:
  - Electrons and photons generated uniformly in \( p_T = [1, 100] \) GeV.
  - PU uniformly distributed between \([55, 75]\) interactions.
  - Windows opened around all the clusters with \( E_T > 1 \) GeV (seeds).
- Model Input:
  - Cluster information \((E, E_T, \eta, \phi, z, \text{number of crystals, …})\), list of rechits, summary window features \((\text{max, min, mean of the crystal variables})\).
  - Same network to identify the flavor of the particle.
  - Extra dataset: sample containing jets.
  - Goal: classify jets/electrons/photons.
  - Transfer Learning was used to re-train only the ID part of the network to avoid the performance degradation for electrons/photons.

Results: Energy Resolution

- Resolution of the reconstructed uncorrected SuperCluster energy \(E_{\text{Raw}}\) divided by the true energy deposits in ECAL \(E_{\text{Sim}}\) versus:
  - the transverse energy of the gen-level particle \(E_{T,\text{Gen}}\) (left)
  - the gen-level particle position \(|\eta_{\text{Gen}}|\) (center)
  - the number of simulated PU interactions (right)
- The resolution is computed as half of the difference between the 84% quantile and the 16% quantile (one σ) of the \(E_{\text{Raw}}/E_{\text{Sim}}\) distribution in each bin.

Results: Particle Classification

- Particle classification performance (DeepSC model) for jet vs. photon (left) and photon vs. electron samples (right).
- Only ECAL variables are used.
- High performance for jet vs. photon discrimination.

Significantly improved resolution, particularly for low \(E_T\) signals and at high PU.

DeepSuperCluster Model