Particle Identification with the Belle II Calorimeter using Machine Learning.

DPG-Frühjahrstagung, Dortmund

Abtin Narimani Charan

18.03.2021
Belle II & Particle Identification (PID)

Electromagnetic Calorimeter (ECL)

Vertex Detectors (VXD)

Central Drift Chamber (CDC) tracking system

Magnet

Time-Of-Propagation counter (TOP, barrel)

Aerogel Ring-Imaging Cherenkov counter (ARICH, end-caps)

$e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$

$K_L$ and $\mu$ detector (KLM)

$e^- \rightarrow 7 \text{ GeV}$

$e^+ \rightarrow 4 \text{ GeV}$

$e^- \rightarrow e^- + e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$
Motivation for improving ECL PID

- Low-$P_T$ muons cannot reach the KLM:
  
  \[ 0.28 \lesssim P_T \lesssim 0.7 \text{ GeV}/c \]

- We need to rely on the information in the ECL.
Default PID in Belle II

**Belle II Simulation**
- ECL barrel
- $0.2 \leq P_T \leq 1.0$ GeV/c

**Cluster Energy / P**

![Cluster Energy / P](image)

**Belle II Simulation**
- ECL barrel
- $0.2 \leq P_T < 1.0$ GeV/c

**μ efficiency**

![μ efficiency](image)

**π fake rate**

![π fake rate](image)
Boosted decision trees (BDT) for PID

- It is based on multi-variate classification algorithm.
- BDTs are trained combining measurements from the ECL and the tracking system.
- There are observables that are defined to describe lateral shower shape development in the ECL which differs for $e$, $\mu$, and $\pi$.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Range</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E/p [c]$</td>
<td>–</td>
<td>Ratio of cluster energy over track momentum.</td>
</tr>
<tr>
<td>$E_{\text{cluster}}$ [GeV]</td>
<td>–</td>
<td>Cluster energy.</td>
</tr>
<tr>
<td>$E_1/E_9$</td>
<td>–</td>
<td>Ratio of the energy of the seed crystal over the energy sum of the 9 surrounding crystals.</td>
</tr>
<tr>
<td>$E_9/E_{21}$</td>
<td>–</td>
<td>Ratio of the energy sum of 9 crystals surrounding the seed over the energy sum of the 25 surrounding crystals (minus 4 corners).</td>
</tr>
<tr>
<td>$</td>
<td>Z_{40}</td>
<td>$</td>
</tr>
<tr>
<td>$</td>
<td>Z_{51}</td>
<td>$</td>
</tr>
<tr>
<td>$Z_{\text{MVA}}$</td>
<td>–</td>
<td>Score of BDT trained on 11 Zernike moments.</td>
</tr>
<tr>
<td>$\Delta L [\text{mm}]$</td>
<td>–</td>
<td>Projection on the extrapolated track direction of the distance between the track entry point in the ECL and the cluster centroid.</td>
</tr>
</tbody>
</table>
Convolutional Neural Network (CNN)

Energy deposition

$\pi$  $\mu$

$\theta_{ld}$ and $\phi_{ld}$ are crystal numbers.

$P_T$  $\theta_{ld}$  $\phi_{ld}$

3295 inputs

128 neurons

Binary Classification Problem

$P(\pi)$  $P(\mu)$
Energy deposition in crystals (MC samples)

Tracks are extrapolated in the calorimeter and I do not rely on clustering and shower shape variables.
Training samples and inputs

- $\mu^\pm$ and $\pi^\pm$ samples (1 track per event)
- The samples include beam background.
- Number of training samples: 865k
  
  Number of validation samples: 288k
CNN performance on MC samples with 1 track / event

\[ \mu \text{ Efficiency} = \frac{\text{Number of muons identified as muons}}{\text{Total number of muons}} \]

\[ \pi \text{ fake rate} = \frac{\text{Number of pions identified as muons}}{\text{Total number of pions}} \]

These plots includes tracks which have a cluster in the ECL assigned to it.

If you are interested in tracks without cluster, look at next slide.
CNN performance - comparison with tracks that has no cluster

total # of $\mu$: 144162

total # of $\pi$: 144162

$\mu$ with cluster: 95.7 %
$\pi$ with cluster: 84.4 %
$\mu$ without cluster: 4.3 %
$\pi$ without cluster: 15.6 %

$\mu$ with cluster: 99.6 %
$\pi$ with cluster: 94.3 %
$\mu$ without cluster: 0.4 %
$\pi$ without cluster: 5.7 %

$\mu$ with cluster: 99.9 %
$\pi$ with cluster: 95.8 %
$\mu$ without cluster: 0.1 %
$\pi$ without cluster: 4.2 %
Physics processes with low-$P_T$ particles

- $e^+e^- \rightarrow \mu^+\mu^-(\gamma_{ISR})$
- $e^+e^- \rightarrow \tau^+ [\rightarrow \pi^+\pi^-\pi^+] \bar{\nu}_\tau \quad \tau^- [\rightarrow 1 \text{ prong}] \nu_\tau$

![Belle II Simulation](image)

Abtin Narimani Charan - DPG 2021
CNN performance on physics processes (MC)

\[ \text{Efficiency} = \frac{\text{Number of muons identified as muons}}{\text{Total number of muons}} \]

\[ \text{Fake rate} = \frac{\text{Number of pions identified as muons}}{\text{Total number of pions}} \]
CNN performance - comparison in different MC samples

NOTE: These plots includes only $\mu^+$ and $\pi^+$. 

Abtin Narimani Charan - DPG 2021

18.03.2021
Summary

- PID for low-$P_T$ tracks in the Belle II detector can be improved using low-level information in the ECL.
- A Convolutional Neural Network (CNN) is trained using 7x7 pixel images in the ECL, based on energy deposition, $P_T$, $\theta_{ld}$, and $\phi_{ld}$.
- CNN method does not depend on clustering or shower shapes.
- CNN surpasses baseline and BDT method in the ECL:
  - For a $\pi^+$ fake rate of 0.2, $\mu^+$ efficiency is:
    - Baseline: 0.42 $\rightarrow$ BDT: 0.65 $\rightarrow$ **CNN: 0.71**