

Particle identification with the Belle II calorimeter using machine learning

Abtin Narimani Charan

Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany

E-mail: abtin.narimani.charan@desy.de

Abstract. I present an application of a convolutional neural network (CNN) to separate muons and pions in the Belle II electromagnetic calorimeter (ECL). The ECL is designed to measure the energy deposited by charged and neutral particles. It also provides important contributions to the particle identification (PID) system. Identification of low-momenta muons and pions in the ECL is crucial if they do not reach the outer muon detector. Track-seeded cluster energy images provide the maximal possible information. The shape of the energy depositions for muons and pions in the crystals around an extrapolated track at the entering point of the ECL is used together with crystal positions in $\theta - \phi$ plane and transverse momentum of the track to train a CNN. The CNN exploits the difference between the dispersed energy depositions from pion hadronic interactions and the more localized muon electromagnetic interactions. Using simulation, the performance of the CNN algorithm is compared with other PID methods at Belle II which are based on track-matched clustering information. The results show that the CNN PID method improves muon-pion separation in low momentum.

1. Introduction

The Belle II experiment at the SuperKEKB e^+e^- collider is a next generation of B factory and an upgrade of the Belle experiment. Among the main goals of the Belle II experiment are the search of New Physics and high precision measurements of the Standard Model parameters in the flavour sector [1]. Belle II intends to collect data samples with an integrated luminosity of 50 ab^{-1} at predominantly the $\Upsilon(4S)$ resonance. SuperKEKB's goal for instantaneous luminosity is $6 \times 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$, which is 40 times larger than that of KEKB. This luminosity can be achieved due to a significant decrease in the beam sizes by a factor of 20 at the interaction point based on the nano-beam collision scheme and by doubling the beam currents in both rings [2, 3]. With the above-mentioned upgrades, Belle II encounters higher beam background level. The physics reach of Belle II can be improved if a better muon-pion separation can be achieved. A better muon-pion separation for low-momenta tracks is advantageous for B physics with leptonic τ decays, since this can reduce fake muons at low momentum. It is crucial to rely on the information in the ECL because low-momentum muons ($p \lesssim 0.6 \text{ GeV}/c$) cannot reach the dedicated, outermost detector K_L^0 -and-muon detector (KLM).

2. The ECL

The ECL consists of 8736 thallium-doped caesium iodide CsI(Tl) crystals projected to the vicinity of the interaction point. It is immersed in a 1.5 T magnetic field and composed of a barrel and two endcaps with polar angle coverage of $12.4^\circ - 155.1^\circ$. The dimension of each

crystal is $\sim 5 \times 5 \text{ cm}^2$ with a length of 30 cm corresponding to $16.2 X_0$ (radiation length). The barrel has 6624 crystals positioned in 46 rings of crystals distributed in the θ -plane and each ring consists of 144 crystals in the ϕ -plane [3, 1]. In this study, tracks are extrapolated into the ECL. Then, at the entry point of the track into the ECL, a window of 7×7 crystals with the measured deposited energy therein is selected. Since the number of crystals in the barrel in each θ -plane is equal to 144, the 7×7 crystal images are symmetrical. Typical patterns of energy depositions for muons and pions are shown in figure 1. Muons and pions both deposit energy by ionization in the matter. Additionally, pions can undergo hadronic interactions. Therefore, crystal images of muons are generally more localized than pions.

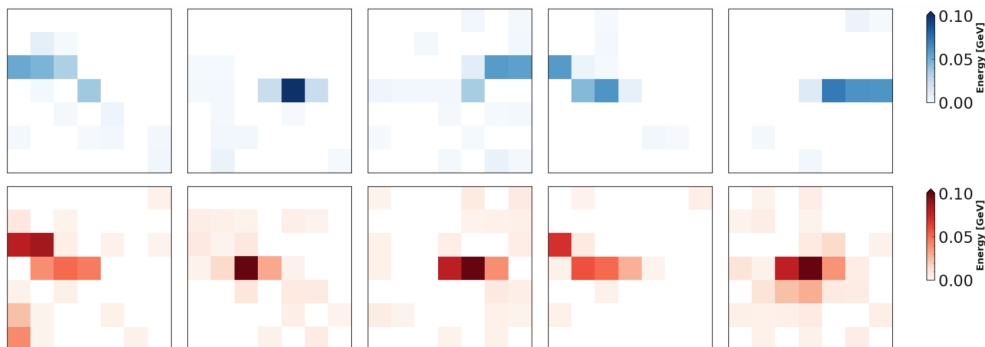


Figure 1. Patterns of energy depositions for muons (blue) and pions (red) inside the ECL barrel.

3. Charged particle identification at Belle II

There are two charged PID methods available in the Belle II analysis software framework [4]. The first method is standard PID which is based on a combination of measurements from different sub-detectors. For each charged particle hypothesis (i) in each PID system, a likelihood $\mathcal{L}_i^{\text{PID}}$ is determined. It is a function of the probability density function parameters for a given set of observables. These likelihoods can be used to construct a combined likelihood ratio for a particular sub-detector. A binary likelihood ratio between muon and pion is defined as $R_{\mu/\pi}^{\text{ECL}} = \frac{\mathcal{L}_{\mu}^{\text{ECL}}}{\mathcal{L}_{\mu}^{\text{ECL}} + \mathcal{L}_{\pi}^{\text{ECL}}}$ which is used as a benchmark for comparison in the study, and is indicated as *default*. The standard PID in the ECL (\mathcal{L}^{ECL}) defines a univariate likelihood as a function of E/p i.e., the ratio of reconstructed cluster energy over the momentum. The E/p distribution is not very powerful for muon-pion separation specifically for low momentum tracks (figure 2). The second method is based on boosted decision trees (BDTs) [5]. It uses the shower-shape information in the ECL combined with likelihood information from other sub-detectors to train BDTs.

4. Convolutional neural network (CNN)

A convolutional neural network (CNN) is a type of deep neural network which is used to recognize visual patterns from pixel images [6]. In this study, two CNNs are trained with 7×7 pixel images of muons and pions together with crystal positions and transverse momentum of the track in the laboratory frame (each pixel is a $\sim 5 \times 5 \text{ cm}^2$ crystal). Separate CNNs are trained for positive and negative charged tracks. This is due to the geometry of the ECL i.e. the direction of the crystals. Approximately 1 million single muon and pion candidates are generated with a flat distribution of transverse momentum between 0.2 and 1.0 GeV/c. Each track is first reconstructed in the inner tracking detectors, and then extrapolated into the ECL [7].

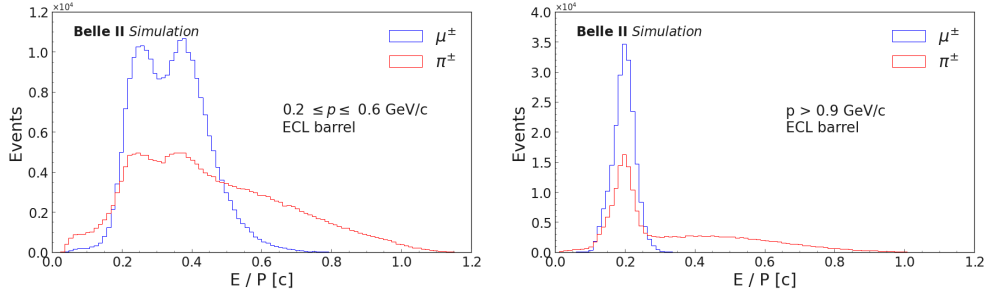


Figure 2. The distribution of the ratio of cluster energy over momentum (E/p) for simulated single track candidates of muons (blue) and pions (red) inside the ECL barrel for low (left) and high (right) momentum range.

4.1. Inputs, pre-processing, and training

There are two types of inputs for the CNN. One is the energy depositions in the 7×7 pixel images and the other is a set of inputs which are fed after convolutional layers including p_T , θ_{ID} , and ϕ_{ID} of the extrapolated tracks. The θ_{ID} and ϕ_{ID} represent integer numbers corresponding to the location of the crystal in the ECL. The energies in the pixel images are given as they are, i.e., without any scaling applied since they already have small values. However, there are very large values in a few pixel images of pions which have energies more than 1 GeV. This is due to pion inelastic interaction with nuclei producing protons. These large values are replaced with 1 GeV. Since there are very few of these pixels, this adjustment does not affect the standard deviation and mean of the energy depositions. A threshold value of 1 MeV on the energy depositions in the pixels is applied, so that pixels below this threshold are assigned zero energy. The p_T is already in the range of 0.2 – 1.0 GeV/c, therefore no scaling is applied. The θ_{ID} and ϕ_{ID} are used as categorical variables which are implemented as an embedding in the network. An embedding is a distributed representation for categorical variables where each category is mapped to a distinct vector that a neural network can learn during training. The number of simulated events is identical for signal and background. The equal size of the sample is important to avoid bias against a particular type of particle or charge.

The CNN includes two parts. In the first part, the 7×7 pixel images of muons and pions are used as inputs for a convolutional layer. During convolution, a window of 3×3 , called kernel, goes through each image. A padding and stride of (1, 1) is used for the images. The padding adds one pixel with value zero on the edge of the image which is beneficial to capture more information on the edges. The stride refers to the amount of kernel movement over the image. The feature map is set to 64. The second part of the training involves a feed-forward neural network (FNN). In this part, the results from the convolutional layer must be flattened and added on top of p_T , θ_{ID} , and ϕ_{ID} . The number of neurons in the first and second layer of FNN are 3295 and 128, respectively. The dropout layer with a value of 0.1 is used between the first and second layer of the FNN. An Adam optimizer is used during the training with a learning rate of 0.001. The loss used in this study is **CrossEntropy** which is suitable for binary classification problems. The number of epochs is set to 100. The model is saved based on the lowest validation loss value in the corresponding epoch. A schematic of the network is shown in figure 3.

4.2. Performance

The performance of the CNN is evaluated on a test dataset with an equal number of muons and pions. The test dataset is generated and reconstructed with the exact same conditions as the training dataset. The CNN is compared with two other methods (default and BDT). The performance of all methods for both charged tracks in a range of transverse momentum ($0.52 \leq$

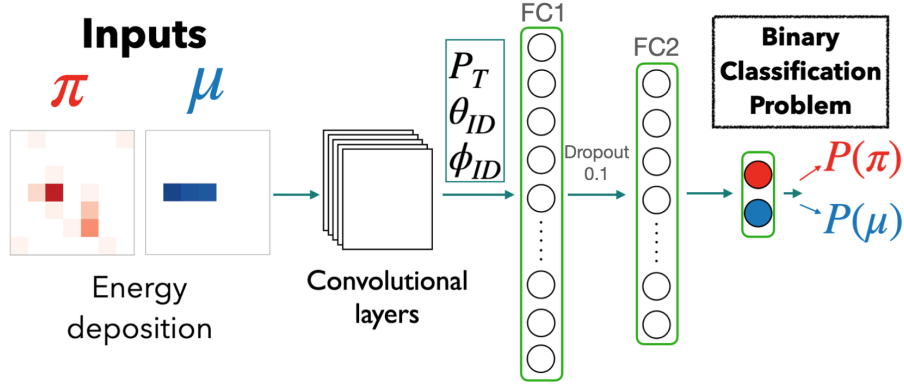


Figure 3. Neural network architecture. See text for details.

$p_T < 0.76$ GeV/c) are shown in figure 4. The μ efficiency is defined as the ratio of the number of correctly identified muons over the total number of muons. The π fake rate is defined as the ratio of the number of pions identified as muons over the total number of pions. The CNN method outperforms the other methods with auc (area under the curve) scores of 0.836 and 0.841 for positive and negative charged tracks, respectively.

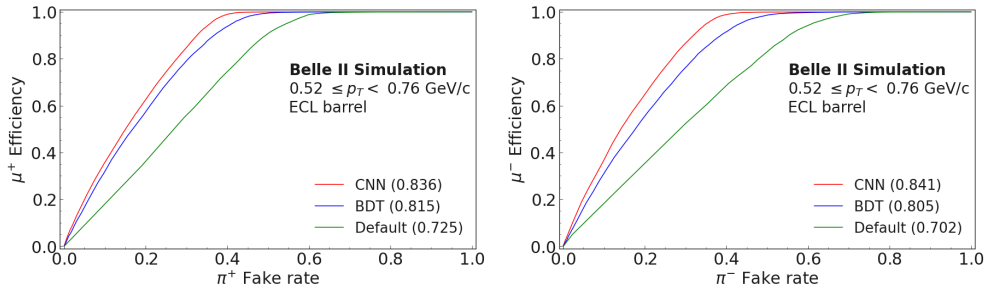


Figure 4. Comparison of CNN, BDT, and default PID. The left and right plots show positive and negative charged tracks, respectively. The value in front of each method shows the area under the curve (auc).

Since the CNN method does not rely on clustering, a comparison is made among all tracks and tracks with and without matched cluster. The comparison is shown in figure 5 in which the blue ROC curves, representing the CNN (all tracks), can be considered as an average between the red and green curves, which represent tracks with and without matched clusters, respectively. The large difference between red and green ones is due to statistics. The green roc curve is roughly 2.8% and 3.5% of the test dataset in the p_T range of 0.52 – 0.76 GeV/c for positive and negative charged tracks, respectively.

Due to the higher beam background level at Belle II, a higher minimal energy threshold for crystals may reduce the pile-up from beam background. In order to check the robustness of the CNN method against different levels of beam background, different energy thresholds for crystals of 0, 1, 2, 5, and 8 MeV are tested and the results are shown in figure 6 (zero means no threshold). Results show the CNN is robust against the choice of different energy thresholds.

5. Summary and outlook

This study shows that by using patterns of energy depositions in the ECL, muon-pion separation can be improved for low-momenta tracks. The pion fake rate is 4.2% and 6.7% larger at a typical

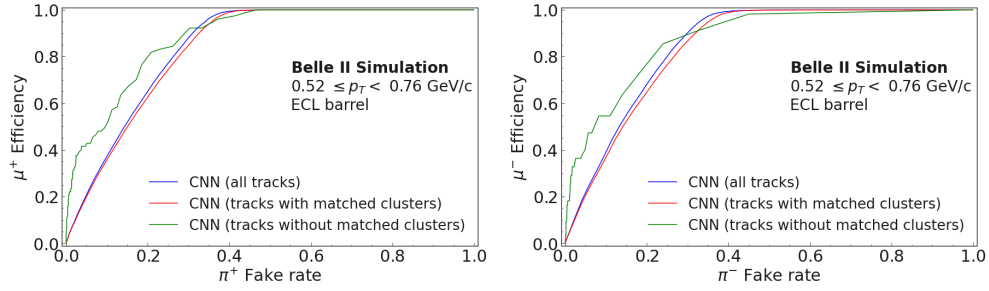


Figure 5. Comparison of CNN performance for all tracks, tracks with and without matched cluster. The left and right plots show positive and negative charged tracks, respectively.

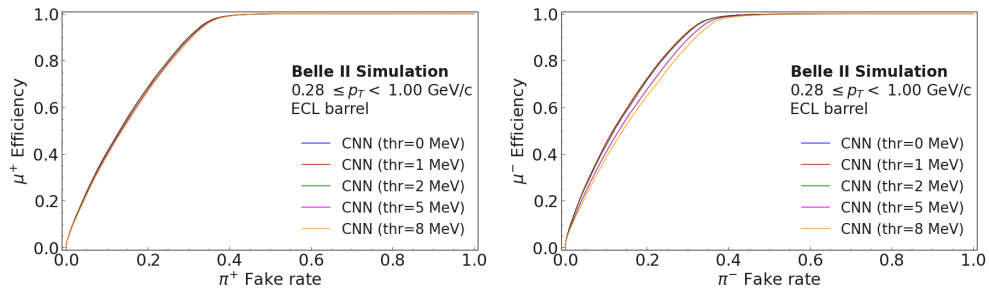


Figure 6. Comparison of CNN performance for different thresholds. The left and right plots show positive and negative charged tracks, respectively.

working point of 90% muon identification efficiency for positive and negative charged tracks, respectively. The CNN method is available in the Belle II analysis software framework and will be integrated as part of the standard Belle II reconstruction software. This study can be extended to include additional low-level ECL crystal information, e.g., pulse-shape discrimination [8, 9] which is useful to separate hadronic and electromagnetic interactions. In order to validate the CNN method on data, clean samples of muons and pions are selected using $e^+e^- \rightarrow \mu^+\mu^-\gamma$ and $D^{*+} \rightarrow D^0 [\rightarrow K^+\pi^-] \pi^+$, respectively. These results are underway.

References

- [1] Kou E et al. 2019 The Belle II Physics Book *Prog. Theor. Exp. Phys.* **2019** 123C01
- [2] Akai K, Furukawa K and Koiso H (SuperKEKB) 2018 SuperKEKB Collider *Nucl. Instrum. Meth. A* **907**, 188–99
- [3] Abe T et al. 2010 *Belle II Technical Design Report* 1011.0352
- [4] Kuhr T, Pulvermacher C, Ritter M, Hauth T and Braun N 2019 The Belle II Core Software *Comput. Softw. Big Sci.* **3** 1
- [5] Milesi M, Tan J and Urquijo P 2020 Lepton identification in Belle II using observables from the electromagnetic calorimeter and precision trackers *EPJ Web Conf.* **245** 06023
- [6] Bishop C M 2006 *Pattern Recognition and Machine Learning* (Singapore: Springer Science+Business Media, LLC)
- [7] Agostinelli S et al. (GEANT4) 2003 GEANT4—a simulation toolkit *Nucl. Instrum. Meth. A* **506** 250–303
- [8] Longo S and Roney J M 2018 Hadronic vs. electromagnetic pulse shape discrimination in CsI(Tl) for high energy physics experiments *JINST* **13** P03018
- [9] Longo S et al. 2020 CsI(Tl) pulse shape discrimination with the Belle II electromagnetic calorimeter as a novel method to improve particle identification at electron-positron colliders *Nucl. Instrum. Meth. A* **982** 164562