Particle identification algorithm of the RICH detector on STCF using CNN method

Authors: Lin, Miss Wanlin, zhou, Dr. xiaokang, Li, Dr. Teng, yao, Dr. zhipeng, Huang, Dr. Xing-Tao, zhou, Dr. xiaokang

Date: 2026-01-22T16:47:37+00:00

Abstract

This paper proposes a Convolutional Neural Network-based method to separate pions and kaons in the STCF barrel PID detector. The approach employs multi-modal learning, integrating dual-channel 2D images derived from Cherenkov photon hit channels and arrival times, along with 3D position and momentum data from the tracking system. The model achieves an overall accuracy of 97%, with a pion signal efficiency of 95% while maintaining a kaon misidentification rate below 2%. This work highlights the advantages of machine learning methods in high-energy physics PID tasks and offers valuable insights for future STCF detector design and optimization.

Full Text

Preamble

Particle identification algorithm of the RICH detector on STCF using CNN method∗ Wanlin Lin,1 Zhipeng Yao,2 Teng Li,2, † Xiaokang Zhou,1, ‡ and Xingtao Huang2 1Institute of Particle Physics and Key Laboratory of Quark and Lepton Physics (MOE), Central China Normal University, Wuhan, China 2Shandong University, Qingdao, China This paper proposes a Convolutional Neural Network-based method to separate pions and kaons in the STCF barrel PID detector. The approach employs multi-modal learning, integrating dual-channel 2D images derived from Cherenkov photon hit channels and arrival times, along with 3D position and momentum data from the tracking system. The model achieves an overall accuracy of 97%, with a pion signal efficiency of 95% while maintaining a kaon misidentification rate below 2%. This work highlights the advantages of machine learning methods in high-energy physics PID tasks and offers valuable insights for future STCF detector design and optimization.

Keywords: STCF Experiment, RICH Detector, Particle Identification, Convolutional Neural Network, Deep Learning

INTRODUCTION

Particle identification (PID) [1] is a basic experimen- tal technique in high-energy physics, aimed at separating charged and neutral particle species. The precision of PID of- ten determines the success of physics analyses. It purifies sig- nal samples, suppresses backgrounds, and controls system- atic uncertainties. Research in flavor physics, hadron spec- troscopy, and searches for rare processes all critically depend on accurate PID for final-state particles. Among various tech- niques, the Ring-Imaging Cherenkov (RICH) detector rep- resents a key solution for high-precision PID. With the pro- posed next-generation Super Tau-Charm Facility (STCF) [2], its design luminosity and physics goals demand PID perfor- mance with unprecedented standards—requiring broader mo- mentum coverage, higher efficiency and superior time resolu- tion.

As high-energy experiments progress towards higher lumi- nosity and more complex event environments, data generated by RICH detectors exhibit increased complexity, including overlapping rings from high-occupation events, noise, and edge effects. This trend presents a universal challenge to conventional reconstruction algorithms that rely on explicit parametric models. To address this fundamental challenge, we develop a general-purpose PID methodology based on deep learning. Departing from traditional steps such as ex- plicit ring fitting or geometrical reconstruction, our approach establishes an end-to-end, multi-modal deep learning frame- work. This framework employs a convolutional neural net- work (CNN) to simultaneously process raw RICH image pat- terns and auxiliary three-dimensional momentum and posi- tion information from the tracking system, thereby learning a direct mapping from complex Cherenkov radiation patterns to particle species. This data-driven multi-modal [3] fusion strategy leverages the full spectrum of detector information more effectively, automatically correcting for momentum de- ∗ Supported by the National Natural Science Foundation of China (No.12341504 and No.12475198) and the Fundamental Research Funds for the Central Universities. † Corresponding author, tengli@sdu.edu.cn ‡ Corresponding author, zhouxk@ccnu.edu.cn pendencies and adapting to geometric variations, thereby en- hancing robustness and accuracy under challenging condi- tions.

In the current development of the STCF RICH detec- tor, PID is primarily performed using a likelihood-based ap- proach that compares hypothesis-dependent responses, com- monly referred to as the Delta Log-Likelihood (DLL) method.

This technique has demonstrated strong PID capability on simulated samples; however, it suffers from a notable limita- tion in terms of computational efficiency. The construction of likelihood values requires repeated evaluations of probability density functions on the anode plane, together with frequent access to large databases of Cherenkov photon images. These operations introduce substantial computational overhead and significantly increase the runtime of the PID procedure.

As the STCF experiment is expected to handle large data volumes, it is therefore necessary to explore alternative PID strategies that can maintain comparable physics performance while reducing computational complexity. Such developments are not only motivated by technical considerations but also align with the increasing emphasis on green and sustain- able computing in large-scale scientific experiments.

In this context, machine learning-based approaches, and in particular convolutional neural networks, provide a promis- ing direction. By performing end-to-end inference directly from detector response patterns, CNN-based methods have the potential to replace multi-step reconstruction and likeli- hood evaluation with a more efficient and streamlined work- flow. The approach proposed in this work follows this strat- egy and aims to establish an efficient and robust PID frame- work for the STCF RICH detector. Moreover, the input repre- sentation can be flexibly adapted to different detector geome- tries, making the method readily extensible to other RICH- based particle identification systems.

II. THE SUPER TAU-CHARM FACILITY A. The Super Tau-Charm Facility The STCF is an electron-positron collider proposed in China, with a center-of-mass energy (CME) ranging from 2 to 7 GeV, and a peak luminosity of 0.5 × 1035cm−2s−1.

The STCF detector is designed to provide broad solid-angle coverage, minimal noise, high detection efficiency, excellent resolution, and superior PID capabilities. The STCF will provide a unique platform for exploring the asymmetry of matter-antimatter (charge-parity violation), in-depth studies of the internal structure of hadrons and the nature of non- perturbative strong interactions, as well as searching for ex- otic hadrons and physics beyond the Standard Model. The conceptual layout of the STCF detector system is shown in Fig. 1 [Figure 1: see original paper]. Extending from the interaction point, the STCF de- tector includes a tracking system, which consists of an inner tracker (ITK) and a Main Drift Chamber (MDC); a PID sys- tem, an electromagnetic calorimeter (EMC), a superconduct- ing solenoid (SCS) and a muon detector (MUD).

Fig. 2 [Figure 2: see original paper]. Momentum distributions of final-state charged particles from various processes at STCF consists of several major components, including the core soft- ware, application software and external libraries. The core software [5] is based on the lightweight and highly flexible SNiPER framework, which provides basic functionalities such as the event loop control, event data management, de- tector description management and database interface, etc. It also coordinates the execution of various tasks in both serial and parallel modes, enabling efficient large-scale data pro- cessing. The simulation software handles event generation, detector geometry description, particle transport and interac- tions, and digitization—combining background and detector response to produce final outputs. The reconstruction soft- ware performs charged particle tracking, photon reconstruc- tion, and PID.

In this study, the proposed CNN-based PID algorithm is implemented within the OSCAR framework as a reconstruc- tion module. The algorithm design and performance evalua- tion are presented in the following sections.

Fig. 1. Schematic layout of the STCF detector concept.

The PID system, located between the MDC and the EMC systems, serves as the primary detector at STCF for hadron discrimination, playing a critical role in effective background suppression. Accurate PID is essential, forming the cornerstone of most physics research programs at STCF. As il- lustrated in Fig. 2, the momentum distribution of final-state particles from various processes covers a wide range, ex- tending from less than 50 MeV/c momenta up to 3.3 GeV/c.

Consequently, the STCF experiment imposes stringent per- formance requirements for the PID system: for π/K separa- tion, a pion (kaon) identification efficiency exceeding 97% is required for momenta below 2 GeV/c , while maintaining a kaon (pion) misidentification rate below 2%.

B. Offline Software for STCF The offline software is one of the key components of mod- ern high-energy physics experiments. For STCF, a dedi- cated offline data processing system, OSCAR (the offline software of the Super Tau-Charm Facility [4], has been de- veloped to support the detector simulation, digitization, cal- ibration, reconstruction and physics analysis. Fig. 3O [Figure 3: see original paper]SCAR Fig. 3. Architecture of the OSCAR C. RICH Detector The PID system comprises a barrel section and two end- cap sections. For the barrel section, two detector options are currently under evaluation: a RICH detector and a Barrel Time-of-Flight Internal Total Reflection Cherenkov (BTOF) detector. This paper focuses on the RICH detector option. As shown in Fig. 4 [Figure 4: see original paper], the RICH consists of 12 independent mod- ules. Using the line connecting the centers of the endcaps as the z-axis, the detector provides nearly full azimuthal cover- age (360°) and operates within a polar angular range of 36° to 144°. Each module is housed in a light-tight aluminum en- closure and incorporates integrated gas purification and cool- ing systems. The physical dimensions of each RICH module measure 2700 mm in length, 450 mm in width, and 150 mm in height.

When a high-energy charged particle traverses a uniform transparent medium at a velocity exceeding the phase veloc- ity of light in that medium (i.e., v > c/n), it emits faint elec- tromagnetic radiation known as Cherenkov radiation [6]. De- tectors specifically designed to detect and record this radia- tion are termed Cherenkov detectors. It should be noted that the STCF RICH design differs from implementations in other experiments, such as the LHCb experiment [7], by not em- ploying spherical mirrors to generate circular Cherenkov ring images.

Fig. 5 [Figure 5: see original paper]. The RICH detector structure. In this version, sion 2.6.2 to simulate the incidence of single kaon and pion particles.

In the current simulation setup, only electronic noise is considered as background, while physical back- ground processes are not included.

Cherenkov photon yield of the RICH detector is 15. The corresponding number distribution of Cherenkov photons at the RICH detector is shown in Fig. 6 [Figure 6: see original paper]. A small peak ap- pears in the photon count between 0 and 5, caused by unsatu- rated Cherenkov photons from very low-momentum particles.

When the particle momentum exceeds 1 GeV/c, the emitted Cherenkov photons gradually saturate at about 15 photons, slightly shifting the overall statistical peak to the left.

Fig. 4. The conceptual design of the RICH detector.

The working principle and imaging process of the RICH detector are illustrated in Fig. 5. When a charged particle tra- verses the radiator liquid (C6F14) and its surrounding trans- parent quartz at a constant velocity, Cherenkov radiation is generated. The particle and the Cherenkov photons propa- gate forward by 100 mm before reaching a thick gas elec- tron multiplier (THGEM) coated with a CsI layer [8]. Upon striking the CsI photocathode, the photons release photoelec- trons. These photoelectrons, under the influence of an elec- tric field, initiate an electron avalanche, thereby amplifying the initial signal. Positioned 2 mm below the THGEM is a MicroMEGAS (MM) [9] detector. It forms a sub-millimeter gas gap between a mesh electrode and a readout electrode, creating a region with a very high electric field for further avalanche amplification. The signal is ultimately collected by the anode after this amplification stage [10].

We employed a particle gun based on OSCAR ver- Fig. 6. Distribution of the number of Cherenkov photons detected in the RICH detector, obtained from particle-gun simulations using OSCAR version 2.6.2.

Electronic noise is introduced at the digitization stage to emulate the intrinsic fluctuations of the readout system. Noise hits are generated independently for each readout channel ac- cording to a uniform random process, with a nominal noise rate of approximately 100 Hz per channel within a readout time window of 2 µs. The noise hit time and charge are ran- domly sampled within the corresponding readout ranges and are labeled separately from signal-induced Cherenkov photon hits. To suppress noise contributions and retain hits consistent with Cherenkov photon signals, a timing selection is applied during data preparation. Only hits with arrival times within 0–100 ns are retained for further analysis. This selection sig- nificantly reduces electronic noise while preserving the dom- inant fraction of signal photons.

Disregarding the influence of the superconducting solenoid on charged particles and leveraging the detector' s symmetri- cal structure, we select data from a single side for training purposes. Additionally, we exclusively utilized single-track, single-module events for the learning process. The data phase space covered momenta range from 0.3 to 2.4 GeV/c and polar angles spanning 36 ∘

to 90 ◦ . The anode plane, measur- ing 2.7 m × 1.5 m, is partitioned into pixels of 5 mm × 5 mm in size to identify the hit channels of photons. Based on these photon hit channels, we can construct a two-dimensional im- age with dimensions of 90 × 540 pixels. As an illustrative example, Fig. 7 [Figure 7: see original paper] shows cases of both normal and oblique inci- dence for a 1 GeV/c kaon particle. The Cherenkov rings are observed to exhibit shapes resembling ellipses and hyperbo- las, respectively, under these incidence conditions.

Fig. 7. Particle' s Cherenkov images in a RICH module, with mo- mentum = 1 GeV, the bright spot at the center represents the hit of the kaon, while the two halos correspond to the two types of radia- tors: C6F14 and quartz. When the incident occurs directly, the two halos overlap. The top one is a normal incident $\theta = 90 \circ$, while the bottom one is from oblique incidence $\theta = 50 \circ$.

III. CNN-BASED PARTICLE IDENTIFICATION METHOD The spatial topology of images generated by particle in- teractions in the detector encodes rich physical information, serving as the foundation for PID. The characteristic patterns within these images often demonstrate high nonlinearity and complexity, rendering them a highly valuable research sub- ject. As a deep learning architecture specifically designed for processing grid-structured data (e.g., images), CNNs inher- ently excel at automatically learning multi-level feature rep- resentations. CNNs can directly extract hierarchical features from raw data, spanning from local edges to global seman- tics, thereby providing a powerful data-driven approach for in-depth exploration of discriminative information in images generated by the detector.

A. Convolutional Neural Networks CNN are well-suited for processing single and multi- channel images, including both 2D and 3D inputs [11]. The architecture of CNNs primarily comprises several key com- ponents: The initial layer is the input layer, which receives the raw image data in Euclidean space, typically represented as a two-dimensional matrix. The convolutional layer stands as the cornerstone of CNN architecture, enabling translation invariance and the ability to capture local features of the im- age without being affected by their positions. Different con- volutional kernels are usually employed to convolve the input data and extract various features. The subsequent activation layer incorporates non-linear functions, enabling the network to handle complex image data better. Following this, the pool- ing layer reduces the size of the input image, thus decreasing computational complexity. The final component is the fully connected layer, which maps the extracted features to the fi- nal output. In the context of binary classification tasks, this layer yields a two-dimensional output vector.

With the deepening understanding of CNNs, their applica- tions in particle physics have been progressively expanding.

For instance, a CNN-based PID algorithm developed for the endcap PID detector (DTOF) at the STCF has demonstrated superior performance compared to traditional methods [12].

Similarly, on the LHCb RICH detector [13], the CNN algo- rithm achieves performance comparable to conventional ap- proaches, while its computational simplicity renders it partic- ularly suitable for high-speed online reconstruction through parallel processing. In the General Antiparticle Spectrome- ter(GAPS) experiment [14], CNNs are employed to identify correlations between particle arrival position and deposited energy. Furthermore, in the MicroBooNE experiment [15], a multi-PID network utilizing CNN architecture was developed to accept dual-image inputs, thereby addressing inefficiencies in vertex reconstruction and particle clustering. Collectively, these examples showcase the remarkable capability of CNNs in processing detector images and uncovering implicit corre- lations among physical quantities.

Although a variety of CNN architectures have demon- strated strong performance in PID tasks based on high-energy physics detector data, practical deployment in the STCF ex- periment requires careful consideration of computational ef- ficiency, model complexity, and training stability. In this con- text, we aim to adopt a network architecture that achieves a well-balanced trade-off between classification accuracy and computational cost. Among the numerous CNN architectures proposed in recent years, the EfficientNet [16] family has attracted significant attention due to its favorable balance be- tween performance and efficiency. Unlike conventional ap- proaches that independently scale network depth, width, or input resolution, EfficientNet introduces a compound scaling strategy that jointly adjusts these three dimensions within a unified framework. This design enables the network to achieve higher classification accuracy with fewer parame- ters and reduced computational complexity. As illustrated in Fig. 8 [Figure 8: see original paper], EfficientNetV1 exhibits superior performance on the ImageNet [17] benchmark compared to mainstream architectures such as ResNet and DenseNet, under comparable pa- rameter counts and floating-point operation budgets.

In addition, EfficientNet models show strong transfer learning capability, allowing effective reuse of pretrained weights obtained from large-scale natural image datasets.

This property significantly accelerates convergence and im- proves training stability. As further summarized in Table. 1, among the EfficientNet family, EfficientNetV2-S maintains high classification accuracy and strong generalization perfor- mance with a relatively small number of parameters. Taking into account model performance, computational efficiency, and scalability to large detector datasets, EfficientNetV2-S is therefore selected as the backbone network for the PID task in this work.

B. A Hybrid CNN for RICH-Based PID To fully utilize the response information from the RICH detector, we developed a hybrid neural network architecture, as detailed in Table 2 , that processes both the raw detector hit patterns and auxiliary kinematic parameters of indicent par- ticles. This design enables the model to learn from comple- mentary data sources: the spatial distribution of Cherenkov photon hits and the physical constraints provided by particle kinematics. two distinct input types The network processes two distinct input types: • Dual-Channel Image Input: The RICH detector data is represented as sparse matrices encoding two distinct physical quantities: channel 1 contains hit identifiers, and channel 2 contains timing information. These are combined to form dual-channel images of dimension.

A third channel is created by replicating channel 1, re- sulting in final input images of shape 90 × 540 × 3. • Physics Kinematics Input: The three components of momentum (px, py, pz) and the three components of position of the incident charged particle (xx, xy, xz) are used, which are provided by the tracking system.

Image Feature Extraction with EfficientNetV2-S EfficientNetV2-S [18] is a lightweight variant of the second-generation EfficientNet CNNs. To further acceler- ate training, a transfer learning strategy is adopted, and the model is initialized with parameters pretrained on the Ima- geNet [17] dataset [19].

In terms of network architecture, fused-MBConv blocks are employed in the shallow stages (stage 2–4) to improve training speed, while classical MB- Conv blocks [16] are used in the middle and later stages (stage 5–7) to enhance feature representation. As the network depth increases, the spatial resolution of feature maps is gradually reduced, accompanied by an increase in channel dimensions, enabling a hierarchical transition from low-level local fea- tures to high-level abstract representations. This hierarchical feature extraction is well-suited for capturing complex spatial patterns and adapts naturally to photon hit images from the RICH detector. Finally, we removed its top fully connected layer at stage 8, preserving its robust feature extraction ca- pacity, and a 1 × 1 convolution followed by global average pooling (GAP) is applied to produce fixed-length high-level feature vectors.

3. Kinematic Branch and Feature Fusion

The kinematic input is processed by a lightweight fully connected branch consisting of two linear layers with an inter- mediate batch normalization step. Rather than directly con- catenating the raw six-dimensional kinematic variables with high-dimensional image features, this mapping projects the kinematic information into a higher-dimensional latent space.

This transformation serves two purposes: it reduces the scale mismatch between low-dimensional physical parameters and image-derived feature vectors, and it allows the network to learn nonlinear combinations of kinematic variables that are more compatible with image-based representations.

The output features from the image backbone and the kine- matic branch are concatenated along the feature dimension to form a unified representation. This fused feature vector is then processed by a fully connected layer with 256 units and ReLU activation, enabling joint learning of correlations be- tween image-based and kinematic information. To improve generalization and suppress overfitting at the high-level fea- ture stage, a dropout layer [20] with a rate of 0.2 is applied after the fusion layer. The final classification layer consists of a fully connected layer with two output nodes and a soft- max activation function, producing normalized probabilities for the π/K hypotheses.

4. Training Setup

The network is trained using the Adam [21] optimizer with a fixed learning rate of 1 × 10−4 and the categorical cross- entropy loss function [22]. The Adam optimizer combines adaptive learning rate estimation with momentum-based updates, providing stable convergence for networks with het- erogeneous inputs and multiple branches. This property is particularly suitable for the present architecture, where im- age features and kinematic parameters contribute differently to the loss landscape.

The categorical cross-entropy loss is defined as Eq. (1) L(y, p) = −[y log(p) + (1 − y) log(1 − p)] where y denotes the label and p is the predicted probability for the positive class. This loss function directly optimizes the probabilistic output of the classifier and is well-suited for two-class PID tasks.

Fig. 8. Performance comparison of EfficientNet models and other mainstream CNN architectures on the ImageNet dataset. [16] (a) Top- 1 accuracy as a function of floating-point operations (FLOPs), highlighting the computational efficiency of EfficientNet under comparable inference cost. (b) Top-1 accuracy as a function of the number of parameters, illustrating the favorable accuracy-model size trade-off achieved by the EfficientNet family.

Table 1 . Performance comparison of EfficientNet family models on multiple benchmark datasets. [18] Model EfficientNet-B7 EfficientNetV2-S EfficientNetV2-M EfficientNetV2-L Params (M) ImageNet Acc. (%) CIFAR-10 (%) CIFAR-100 (%) Flowers (%) Cars (%) 98.7 ± 0.04 99.0 ± 0.08 99.1 ± 0.03 91.5 ± 0.11 92.2 ± 0.08 92.3 ± 0.13 97.9 ± 0.03 93.8 ± 0.11 98.5 ± 0.08 94.6 ± 0.10 98.8 ± 0.05 95.1 ± 0.10 Stage Operation Type Stride Output Channels Layers Image Backbone Path: Input Image (90$×540)$Table2.NetworkArchitectureDesign3\times3ConvolutionFused - MBConv13\times3Fused{-}MBConv43\times3Fused{-}MBConv43\times3MBConv43\times3MBConv63\times3MBConv63\times31\times$ Convolution + GAP Auxiliary Path: Particle Kinematics Input Vector (6-dimensional) Fully Connected Mapping Batch Normalization Fully Connected Mapping Multimodal Fusion and Classification Head Concatenation + Fully Connected π/K Probability Output Early stopping with a patience of 13 epochs is applied to mitigate overfitting. During training, a custom data generator is employed to handle the sparse representation of RICH de- tector data and to convert it into dense tensors on-the-fly dur- ing batch processing, ensuring efficient memory usage and stable training behavior.

IV. MODEL TRAINING AND PERFORMANCE EVALUATION A. Experimental Setup All datasets used in this study are derived from simulated events generated with OSCAR version 2.6.2. The dataset contains more than 50 million events of four charged parti- cle species (K $+$, K $-$, $\pi+$, $\pi-$), which are classified into two categories for $\pi/$K identification, with pions ($\pi+$ and $\pi-$) la- belled as 1 (positive), while kaons (K $+$ and K $-$) are labeled as 0 (negative). Considering the increased discrimination difficulty at high momentum, the event statistics in this region are enhanced to ensure adequate training performance. The dataset is divided into a training set of 49 million events, a validation set of 300,000 events, and an independently gen- erated test set of 230,000 events. The test set is uniformly distributed over the full momentum range to provide an unbiased performance evaluation. The model training was con- ducted on a single NVIDIA A800 GPU.

B. Overall Performance In the issue of $\pi/$K identification, the machine learning evaluation metrics can be directly related to physically mean- ingful quantities. The model performance is evaluated using metrics that correspond directly to physical quantities. The true positive rate (TPR, Eq. (2)) represents the pion signal ef- ficiency, defined as the probability that a true pion is correctly identified as a pion. and the false positive rate (FPR, Eq. (3)) represents the kaon misidentification probability, defined as the probability that a true kaon is incorrectly classified as a pion. All results presented in this section are obtained using a fixed discrimination threshold of 0.5.

Here, TP and FN denote the numbers of true pions correctly and incorrectly identified as pions, respectively, while FP and TN denote the numbers of true kaons incorrectly and cor- rectly classified as pions.

For the full independent test sample, the model achieves a pion signal efficiency of 96.82% and a kaon misidentification rate of 3.18%. These results indicate a balanced $\pi/$K separa- tion performance under a fixed operating threshold. Overall classification accuracy of 96.79%. These results indicate a balanced $\pi/$K separation performance under a fixed operat- ing threshold. Since the majority of final-state particles in STCF have energies concentrated below 2 GeV/c, the model performance is further evaluated in the momentum region p $<$ 2 GeV/c. In this region, the pion signal efficiency the pion signal efficiency increases to 97.29%, the kaon misiden- tification rate is reduced to 2.20%, and the accuracy reached 97.56%.

We further evaluate its performance across momentum and polar angle. As momentum increases, signal efficiency and background misidentification converge due to reduced Cherenkov angle separation at higher velocities (Fig. 9 [Figure 9: see original paper]). Per- formance remains stable over polar angle, with a slight de- cline at small angles likely caused by detector edge effects (Fig. 10 [Figure 10: see original paper]).

Fig. 9. Signal efficiency and background misidentification rate as a function of momentum.

Fig. 10. Signal efficiency and background misidentification rate as a function of the polar angle.

The overall discriminative power of the model is summa- rized by the receiver operating characteristic (ROC) curve, which plots TPR against FPR across all possible classifica- tion thresholds. As shown in Fig. 11 [Figure 11: see original paper], the model achieves an area under the curve (AUC) of 0.995, indicating excellent sep- aration capability between pions and kaons over the full kinematic range. The underlying separation is further visualized in Fig. 12 [Figure 12: see original paper], which displays the distribution of model discrim- inant scores for true pions and kaons. The two distributions are well separated with minimal overlap, explaining the high AUC value. Unless otherwise stated, a fixed discrimination threshold of 0.5 is used to report the following performance metrics. tent with those obtained using a fixed discrimination thresh- old of 0.5, as discussed in the previous section.

In both cases, the performance degradation appears predominantly in the high-momentum and small-polar-angle regions, indicat- ing that the underlying behavior is governed by intrinsic de- tector and physics effects rather than the choice of threshold.

Compared to the fixed-threshold scenario, the imposed misidentification constraint of 2% represents a relatively stringent requirement. As a consequence, the efficiency loss becomes more pronounced in the kinematic regions where the intrinsic π/K separation power of the RICH detector is lim- ited. The localized performance degradation can be attributed primarily to intrinsic detector and physics effects rather than limited training statistics. At high momentum, the Cherenkov angles of pions and kaons become increasingly similar, lead- ing to a reduced intrinsic separation power of the RICH de- tector. In the small polar angle region close to the beam direc- tion, geometrical effects further limit the effective Cherenkov photon distribution on the detector plane, making the discrim- ination task more challenging. Under a fixed misidentifica- tion constraint, these effects naturally result in a partial loss of signal efficiency.

Overall, the observed efficiency variations are consistent with the expected behavior of RICH-based particle identifi- cation systems operating near their intrinsic separation limits.

Despite these localized effects, the proposed method satisfies the core PID performance requirements of the STCF over the full kinematic range relevant for physics analyses.

D. Effect of training sample size at fixed momentum The size of the training dataset is a key factor affect- ing the performance of machine learning models, particu- larly in detector-based analyses where generalization depends strongly on the statistical coverage of the data. For PID based on RICH detector images, variations in particle momentum, incident angle, and detector response lead to non-uniform fea- ture distributions, and limited training samples may result in insufficient learning in certain regions of phase space. Study- ing the relationship between training sample size and clas- sification accuracy provides a systematic way to assess the model’ s dependence on available statistics. Such an analysis helps distinguish whether the observed performance is lim- ited by the model architecture or by the amount of training data, and offers insight into the model’ s generalization be- havior under different statistical conditions.

To study the dependence of the model performance on the training sample size, we evaluate the π/K identification ac- curacy at a fixed particle momentum of 1 GeV/c in the RICH detector. The classification accuracy is used here as a repre- sentative metric to characterize the convergence behavior of the model. As shown in Fig. 14 [Figure 14: see original paper], the identification accuracy increases rapidly with the size of the training sample in the low-statistics region, indicating that the model benefits sig- nificantly from additional training data. When the training sample size reaches approximately 8 × 105 events, the accuracy gradually saturates, and further increases in the sample Fig. 11. Model ROC curve. The curve illustrates the trade-off be- tween pion signal efficiency and kaon misidentification rate.

Fig. 12. Distribution of model discriminant scores for pions and kaons. Histograms show the output scores for pion and kaon test samples.

C. Compliance with STCF RICH PID Requirements By dynamically adjusting the decision threshold, the model can tune the background misidentification rate or the signal efficiency to meet specific experimental needs. In accordance with the STCF specifications, we constrain the kaon misiden- tification rate to be below 2% by adjusting the discrimina- tion threshold. As shown in Figure 13 [Figure 13: see original paper], the pion signal effi- ciency exceeds 97% across most momentum and polar an- gle regions, despite a slight decline observed in the high- momentum, small-angle region, indicating room for further optimization. It is worth noting that the efficiency trends ob- served under this adaptive-threshold setting are fully consis- Fig. 13. Pion signal efficiency at different momenta and polar angles, with a kaon misidentification rate controlled to be below 2% size result in only marginal performance improvement. This behavior suggests that the model has sufficiently learned the discriminating features of the detector response under fixed- momentum conditions. For comparison, the performance of a traditional RICH PID method based on conventional ring reconstruction and likelihood-based discrimination are also shown in Fig. 14. The proposed model surpasses the tradi- tional method when the training sample size exceeds approxi- mately 2$×$105 events, demonstrating its superior capability in exploiting the available detector information. This study pro- vides a quantitative reference for the required training statis- tics in practical applications of the proposed method.

Fig. 14. π/K identification accuracy as a function of the training sample size at a fixed momentum of 1 GeV/c. The accuracy in- creases with the training statistics and gradually saturates when the sample size reaches the order of 106.

Impact of angular coverage and photon statistics The photon yield is a core physical quantity that determines the performance of the RICH detector. Regions exhibiting a decline in efficiency are typically directly linked to an insufficient photon count, which leads to a degraded signal-to-noise ratio and blurred reconstruction of the pixel ring. In this sub- section, we will investigate the relationship between PID per- formance and Cherenkov photon yield.

We analyzed data from two OSCAR versions: version 2.5.0 with an average photon yield of 10, and version 2.6.0 with an average yield of 15. Models were trained using events at 1 GeV and a polar angle of 90°. As shown in the Fig. 15 [Figure 15: see original paper], the version with higher photon yield achieves greater accu- racy. Furthermore, the accuracy steadily improves with in- creasing training set size, which aligns with the findings from the previous section. This observation provides a promising direction for future detector upgrades.

V. ONNX-BASED MODEL DEPLOYMENT AND INTEGRATION WITH OSCAR Currently, a robust ecosystem has formed around the Open Neural Network eXchange (ONNX) [23] standard. The same ONNX model file can be executed across diverse platforms such as servers, mobile devices, and web browsers via ONNX Runtime. The trained model has been successfully deployed into the STCF OSCAR software framework using ONNX technology, enabling efficient invocation through a C++ in- ference interface. This completes the closed loop from algo- event). To further evaluate the achievable acceleration, ad- ditional inference benchmarks were carried out using the exported ONNX model on an NVIDIA A800 GPU. When GPU acceleration is enabled, the inference latency is reduced to below 1 ms per event, providing more than an order-of- magnitude speedup. Additionally, this indicates that CNN- based methods can provide a potential implementation solu- tion for ultra-low latency fast event reconstruction in the soft- ware trigger system based on GPU.

Together, these tests confirmed the reliability of the end- to-end pipeline, from raw signals to identification results, demonstrating that the CNN-based approach not only meets the computational requirements of the STCF offline recon- struction workflow but also possesses strong potential for fu- ture large-scale data processing.

Table 3 . Average inference time per event for different PID methods within the OSCAR framework.

Hardware

Method

CNN-based PID CPU(Xeon Silver 4314) CNN-based PID GPU(NVIDIA A800) CPU(Xeon Silver 4314) Time per event [ms] VI. CONCLUSION In this study, we developed and validated a CNN-based multi-modal learning framework for PID in the barrel RICH detector of the STCF. By integrating two-dimensional images constructed from Cherenkov photon hit channels and arrival times with three-dimensional tracking information, the model effectively captures complex, nonlinear patterns in particle- induced signatures.

The proposed method achieves an overall accuracy of 96.95% on an independent test set and satisfies the STCF per- formance requirements for π/K separation, delivering a pion signal efficiency of 97.3% at momenta below 2 GeV/c while maintaining a kaon misidentification rate under 2%. Sys- tematic evaluations across momentum and polar angle phase spaces reveal stable performance over most regions, with a slight decline observed only at high momentum and small polar angles—a behavior attributable to detector edge effects and reduced Cherenkov angle separation at higher velocities.

Beyond the specific STCF configuration, the proposed framework is largely detector-agnostic and can be readily adapted to other RICH-based PID systems. Since the model operates directly on raw or minimally processed photon hit information, its input representation and network structure can be flexibly adjusted to accommodate different detector geometries, radiator materials, and readout schemes. This makes the approach broadly applicable to current and future experiments employing RICH detectors.

Fig. 15. Test accuracy versus training set size for detectors with dif- ferent photon count versions. Increasing the training set size consis- tently improves model performance, and versions with higher pho- ton counts achieve higher accuracy. rithm development to experimental deployment, ensuring ex- cellent inference speed and scalability. The specific deploy- ment workflow is as follows:

Model Conversion: The trained model was exported in the TensorFlow-recommended SavedModel format and then converted to the standard ONNX format using the tf2onnx tool. This step serializes the model architecture and parame- ters into a portable format.

Inference Engine Integration: The ONNX Runtime [24] C++ library was linked into the RICH reconstruction module of OSCAR.

Dedicated Interface Development: A CNNModel wrap- per class was implemented to manage the ONNX session life- cycle, handle memory allocation, and perform synchronous inference. It provides a Predict interface that accepts prepro- cessed tensor data and returns classification probabilities.

Reconstruction Algorithm Implementation: RICHCnnAlg algorithm module was developed and in- tegrated into the standard event processing loop of the OSCAR framework. Within its execute method, this module accomplishes the following: extracts track parameters and RICH photon hit positions/times from the event; organizes the photon hit information into a three-dimensional image tensor (540$×90×3)*accordingtothedetectorgeometry*; *combinesimageandnon*−*imagefeaturestoperformforwardpropagationbycallingtheCNNModel*; *writestheinferenceresults*($/K identification probabilities) into a new data object for use in subsequent reconstruction stages.

Finally, the functionality and stability were tested using simulated data. The average inference time per event for different PID methods is summarized in Table 3. The CPU- based tests were performed within the OSCAR framework using a dual-socket Intel Xeon Silver 4314 platform (32 phys- ical cores, 2.40 GHz base frequency) in single-thread exe- cution mode. Under this configuration, the proposed CNN- based PID algorithm achieves an average processing time of approximately 60 ms per event, which is already slightly faster than the traditional method algorithm (173 ms per ACKNOWLEDGEMENTS Shandong University, Lanzhou University and University of Science and Technology of China.

The authors acknowledge the computational resources pro- vided by the High Performance Computing platforms of [1] Christian Lippmann, Particle identification, Nuclear Instru- ments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equip- ment, Volume 666, 2012, Pages 148-172, ISSN 0168-9002, DOI:10.1016/j.nima.2011.03.009 [2] M. Achasov, X. C. Ai, R. Aliberti, L. P. An, Q. An, X. Z. Bai, Y. Bai, O. Bakina, A. Barnyakov and V. Blinov, et al..,STCF conceptual design report (Volume 1): Physics & detector. Front. Phys. (Beijing), 19: no.1, 14701 (2024) .

DOI:10.1007/s11467-023-1333-z [3] T. Baltrušaitis, C. Ahuja and L. -P. Morency, Multimodal Ma- chine Learning: A Survey and Taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence. 41, no. 2, 423- 443, (2019) DOI:10.1109/TPAMI.2018.2798607 [4] Ai, X., Huang, X., Li, T., Qi, B., & Qin, X, Design and devel- opment of STCF offline software. Modern Physics Letters A, 39(40), 2440006, (2024). DOI:10.1142/S0217732324400066 [5] W. H. Huang, H. Li, H. Zhou, T. Li, Q. Y. Li and X. T. Huang, Design and development of the core software for STCF offline data processing. JINST 18, no.03, P03004 (2023).

DOI:10.1088/1748-0221/18/03/P03004 [6] Hubert, Pierre. Cherenkov radiation. No. CEA-R-451. CEA Saclay, 91-Gif-sur-Yvette (France), 1955. [7] Gambetta, The LHCb RICH detectors: Operations and perfor- mance, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associ- ated Equipment, Volume 952, 2020, 161882, ISSN 0168-9002, DOI:10.1016/j.nima.2019.02.009 [8] J. Agarwala, M. Alexeev, C. D. R. Azevedo, F. Bradamante, A. Bressan, M. Büchele, M. Chiosso, C. Chatterjee, A. Ci- cuttin and P. Ciliberti, et al., The MPGD-Based Pho- ton Detectors for the upgrade of COMPASS RICH-1 and beyond. Nucl. (2019) DOI:10.1016/j.nima.2018.10.092 Instrum. Meth. A 936, 416-419, [9] Y. Giomataris, P. Rebourgeard, J. P. Robert and G. Charpak.

MICROMEGAS: A High granularity position sensitive gaseous detector for high particle flux environments. Nucl.

Instrum. Meth. A 376, 29-35, (1996) DOI:10.1016/0168- 9002(96)00175-1 [10] B. L. Hou, L. Zhao, J. J. Qin, Y. Q. Qi, J. M. Li, Z. Y. Yang, S. B. Liu and Q. An, Prototype of the readout electronics for the RICH PID detector in the STCF. Nucl. Sci. Tech. 33 no.6, 80, (2022). DOI:10.1007/s41365-022-01056-4 [11] Younesi, Abolfazl, et al., A comprehensive survey of convolutions in deep learning: Applications, challenges, and future trends. IEEE Access 12 (2024): 41180-41218.

DOI:10.1109/ACCESS.2024.3376441 [12] Z. Yao, T. Li and X. Huang, Pion/Kaon Identification at STCF DTOF Based on Classical/Quantum Convolu- tional Neural Network. EPJ Web Conf. 295, 09030 (2024).

DOI:10.1051/epjconf/202429509030 [13] M. P. Blago, Deep learning particle identification in LHCb J. Phys. Conf. Ser.2438, no.1, 012076 (2023) RICH.

DOI:10.1088/1742-6596/2438/1/012076 [14] M. Yamatani, Y. Nakagami, H. Fuke, A. Kawachi, M. Kozai, Y. Shimizu and T. Yoshida, New Particle Identification Approach with Convolutional Neural Networks in GAPS.

DOI:10.57350/jesa.9 [15] P. Abratenko et al. Neural Network Chamber. Phys. Rev. D 103, DOI:10.1103/PhysRevD.103.092003 [MicroBooNE], A Convolutional Identification the MicroBooNE Liquid Argon Time Projection 092003 (2021) for Multiple Particle no.9, [16] Mingxing Tan and Quoc V. Le, EfficientNet: Rethink- ing Model Scaling for Convolutional Neural Networks.

International Conference on Machine Learning, DOI:10.48550/arXiv.1905.11946 [17] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Im- ageNet classification with deep convolutional neural networks.

Commun. ACM 60, 84-90, (2017). DOI:10.1145/3065386 [18] Tan, M., & Le, Q. Efficientnetv2:

Smaller mod- In International conference on 10096-10106, els and faster machine DOI:10.48550/arXiv.2104.00298 learning. PMLR 139, training. [19] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. Bert:

Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1, 4171- 4186, (2019). DOI:10.18653/v1/N19-1423 [20] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn.

Res. 15, 1 1929–1958. (2014). DOI:10.5555/2627435.2670313 [21] Diederik P. Kingma and Jimmy Ba, Adam: A Method for Stochastic Optimization. (2017) 10.48550/arXiv.1412.6980 [22] De Boer, P. T., Kroese, D. P., Mannor, S., & Rubinstein, R. Y. A tutorial on the cross-entropy method. Annals of operations research, 134(1), 19-67, (2005). DOI:10.1007/s10479-005-5724- [23] ONNX. Onnx: Open neural network exchange[EB/OL]. 2025. https://github.com/onnx/onnx [24] Microsoft. Onnx runtime:

A cross-platform inference and training machine-learning accelerator[EB/OL]. 2025. https://github.com/microsoft/onnxruntime/releases/tag/v1.17.3.

Note: Figure translations are in progress. See original paper for figures.