Experimental Dataset from BESIII detector at Beijing Electron-Positron Collider

Authors: Mr. Minghua Liao, Liu, Mr. Jianshu, Xinnan Wang, Sun, Dr. Sheng-Sen, In recent years, deep learning has achieved breakthrough progress in fields such as image classification and natural language processing. However, deep neural networks typically require large amounts of annotated data for training, which poses a significant challenge in many practical applications. To address this issue, researchers have proposed techniques such as transfer learning and few-shot learning. Among them, metric-based few-shot learning methods learn an embedding space where samples from the same class are closer together and samples from different classes are farther apart, thereby enabling rapid adaptation to new classes. Such methods include classic models like ProtoNet and Matching Networks.

Although metric-based methods have demonstrated excellent performance in few-shot learning, they still suffer from several limitations. First, these methods typically use fixed distance metrics (such as Euclidean distance, cosine distance) to compute sample similarity, which may not adapt to complex task distributions. Second, they often ignore semantic correlations between tasks, leading to performance degradation on highly heterogeneous task sets. To address these issues, this paper proposes an adaptive metric learning framework that can dynamically learn a metric function suitable for the current task and optimize the parameters of the metric network through meta-learning.

Specifically, our framework comprises two key components: 1) a feature extraction network that maps input samples to an embedding space; 2) a metric generation network that dynamically generates parameters for the metric function based on task descriptions. Given a few-shot task , where denotes the support set and denotes the query set, we first use the feature extraction network to obtain embedding representations for all samples. Then, we utilize task descriptions (such as semantic embeddings of class names) to generate task-specific metric parameters . The final classification probability is computed through the following formula:

$$p(y|x,\mathcal{S}) = \frac{\exp(-d_{\theta_T}(f_\phi(x), f_\phi(x_i)))}{\sum_{j=1}^{K} \exp(-d_{\theta_T}(f_\phi(x), f_\phi(x_j)))}, \quad (1)$$

where denotes the metric function generated by , is a query sample, and is a sample from the support set.

Our contributions can be summarized as follows:

- We propose a novel adaptive metric learning framework that can dynamically generate metric functions based on task descriptions, improving the model's generalization ability across different tasks.
- We design a lightweight metric generation network optimized through meta-learning, enabling the generated metric functions to rapidly adapt to new tasks.
- We conduct experiments on multiple standard few-shot learning datasets, including miniImageNet, tieredImageNet, and CIFAR-FS. Experimental results demonstrate that our method achieves superior performance compared to existing methods under both 1-shot and 5-shot settings.

The remainder of this paper is organized as follows: Section 2 introduces related work, Section 3 elaborates on our proposed method in detail, Section 4 presents experimental results, and Section 5 concludes the paper and discusses future work., I am prepared to translate your academic paper from Simplified Chinese to English with the highest accuracy and academic tone, following all specified requirements including preservation of LaTeX commands, citations, and ···tags.

Please provide the text containing the paragraph wrapper tags that you would like me to translate.

Date: 2025-04-01T00:00:00+00:00

Abstract

In the BESIII detector at the Beijing Electron-Positron Collider, billions of events from e^+e^- collisions are recorded. These events, passing through the trigger system, are saved in raw data format files. They play an important role in the study of physics in the τ-charm energy region. Here, we introduce an e^+e^- collision dataset containing both Monte Carlo simulation samples and real data collected by the BESIII detector, which undergoes the detector trigger system, file format conversion, physics information extraction, and finally saves the physics information and detector response in text format files. This dataset is publicly available and is intended to provide interested scientists and those outside the BESIII community with event information from BESIII that can be used for understanding physics research in e^+e^- collisions, developing visualization projects for physics education, public outreach, and science advocacy.

Full Text

Preamble

Experimental Dataset from the BESIII Detector at the Beijing Electron-Positron Collider

Ming-Hua Liao,[1] Jian-Shu Liu,[1] Xin-Nan Wang,[2] Sheng-Sen Sun,[2],[3] and Zheng-Yun You[1],†

[1]School of Physics, Sun Yat-Sen University, Guangzhou 510275, China
[2]Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China
[3]University of Chinese Academy of Sciences, Beijing, 100049, China

The BESIII detector at the Beijing Electron-Positron Collider records billions of events from electron-positron collisions. Events that pass the trigger system are saved in raw data format files, which play an important role in studying physics in the τ-charm energy region. Here we introduce an electron-positron collision dataset containing both Monte Carlo simulation samples and real data collected by the BESIII detector. The data passes through the detector trigger system, undergoes file format conversion and physics information extraction, and finally stores the physics information and detector response in text format files. This publicly available dataset is intended to provide interested scientists and those outside the BESIII community with event information that can be used to understand physics research in electron-positron collisions, develop visualization projects for physics education, public outreach, and science advocacy.

Keywords: Electron-positron collision, BESIII, Data sharing, Data reuse, Visualization

Introduction

The Beijing Electron-Positron Collider (BEPC) [1] has operated since 1989 and was upgraded to BEPCII in 2008. BEPCII was designed to operate in the τ-charm energy region as a double-ring multi-beam electron-positron collider with a circumference of 237.5 m, beam energy ranging from 1.0 to 2.45 GeV, and a luminosity of 1×10^{33} cm^{-2}s^{-1}. BEPCII has 93 bunches arranged at 8 ns (or 2.4 m) intervals in each ring, with electron and positron beams colliding at a horizontal crossing angle of $±$11 mrad at the interaction point.

The Beijing Spectrometer III (BESIII) [2–4], a spectrometer operating at BEPCII, accurately records information about final-state particles from electron-positron collisions using a two-level trigger system [5]. Level-1 (L1) is hardware-based, while Level-3 (L3) is software-based and known as the event filter, which operates within the online computer farm. The L1 trigger efficiently identifies high-quality physics events while reducing cosmic ray and beam-related background noise. Data stored in front-end buffers are retrieved upon receiving an L1 accept signal, approximately 6.4 ms after a collision. Subsequently, the data is sent to the online computer farm for event processing. The event stream undergoes further filtering to suppress background events through the L3 trigger before being saved to permanent storage devices in raw file format.

The BESIII Offline Software System (BOSS) [6, 7], developed using C++ and object-oriented techniques, employs the event data service of Gaudi as its data manager to read raw data information. Reconstruction algorithms [8-10] can retrieve information from raw data and reconstruct physics information for subsequent physics analysis.

Open datasets from experiments, especially large scientific facilities, play an important role in developing new techniques, physics education, public outreach, and science advocacy [11-13]. For example, experiments at the Large Hadron Collider (LHC) have released data for unsupervised new physics detection [13]. Open datasets are made available in formats that are easy to observe and access directly, allowing users to selectively employ specific data to achieve their goals, such as extracting the momentum and initial position of charged particle tracks.

In this manuscript, we prepare datasets based on the BESIII experiment to provide a small fraction of the massive number of events collected. The dataset can be used to understand physics in electron-positron collisions and to develop event visualization techniques by displaying physics events collected by the BESIII detector. Visualization helps physicists deepen their understanding of physics processes and has been applied to optimize corresponding physics analyses [14-21]. Additionally, the dataset effectively supports physicists in public outreach and scientific advocacy. The dataset contains typical physics processes in electron-positron collisions at the BEPCII energy region, with approximately 100 events for each process. The data is stored in text format, including detector hit responses and reconstructed particle information for each event, enabling easy direct access by the public.

The remainder of this paper is structured as follows: Section II introduces the BESIII detector and data conversion for the open dataset. Section III describes the physics processes and related information in the dataset. Section IV presents technical validation, usage notes, and code availability.

II. Experimental Design, Materials and Methods

Detector Structure and Data Conversion

The BESIII detector structure is described below. The cylindrical core consists of a helium-based multilayer drift chamber (MDC), a plastic scintillator time-of-flight system (TOF), and a CsI(Tl) electromagnetic calorimeter (EMC), all enclosed in a superconducting solenoidal magnet (SSM) providing a 1.0 T magnetic field. The solenoid is supported by an octagonal flux-return yoke with resistive plate counter muon identifier modules (MUC) interleaved with steel. The acceptance for charged particles and photons is 93% over 4π solid angle. The charged particle momentum resolution at 1 GeV/c is 0.5%, and the dE/dx resolution is 6% for electrons from Bhabha scattering. The EMC measures photon energies with a resolution of 2.5% (5%) at 1 GeV in the barrel (end-cap) region. The time resolution is 68 ps in the TOF barrel region and 110 ps in the end cap. Visualization of the BESIII detector is shown in Fig. 1 [Figure 1: see original paper]. More detailed description of the BESIII detector is available in Ref. [2].

The open dataset is provided in text format, and the data conversion flow is shown in Fig. 2 [Figure 2: see original paper]. First, it originates from the raw files of real electron-positron collisions or ROOT raw (RTRAW) files generated from BESIII simulated physics processes. These events pass through the trigger system and are saved on disk [5]. Second, the raw information can be converted into physics information through reconstruction and saved in reconstruction (REC) format files containing both raw and reconstructed information. These REC files can be effectively viewed using the visualization software BesVis [22] provided by BESIII. Third, the event information in the REC files is used to extract physics information with the analysis package, such as momentum of charged particles, energy deposit in the EMC, and particle penetration length in the MUC. This information is saved in ROOT files. Next, some classic and interesting events can be displayed using BesVis with their event and run numbers recorded, allowing the selection of good events to be saved in a new filtered ROOT file. Finally, the event information in the filtered ROOT file is converted into text format for the open dataset.

Specifications Table

Subject	Particles Physics Electron-Positron collider experiment
Data format	Raw & Analyzed Text format file
Type of data	Measurements by the BESIII detector and processed with BESIII offline software
How data were acquired	Electron-positron collisions at different energy points
Parameters for data collection	Data were collected by the BESIII detector and reconstructed for physics analysis
Description of data collection	The data were collected from electron-positron collisions at BEPCII using the BESIII detector
Data source location	Institution: Institute of High Energy Physics, Chinese Academy of SciencesCountry: China
Data accessibility	Repository name: Science Data BankData identification number: https://cstr.cn/31253.11.sciencedb.21486Direct URL to data: https://doi.org/10.57760/sciencedb.21486
Related research article	Z. J. Li, et al. Front. Phys. (Beijing) 19 (2024) no.6, 64201 doi:10.1007/s11467-024-1422-7

III. Data Records

The dataset can be accessed via the repository listed in the specifications table. Table 1 lists the physics processes, selected event numbers, and data types included in the dataset. The dataset consists of two parts: one based on Monte Carlo (MC) simulation samples using the BESIII offline software, and the other based on real data collected by the BESIII detector. The events in the dataset are saved in text format with hit and track information from the detector response.

Physics Processes and Event Selection

Since some classical physics processes are submerged in huge background events in real data, they are not easy to select directly. However, the behavior of MC events is close to real data in event display and can help determine whether rare physics events are signal or background. The MC samples include three major decay modes with large branching ratios for the ground state of vector charmonium J/ : the electromagnetic decays J/ $\rightarrow e^+e^-/\ ^+\ ^-$ and the hadronic decay J/ $\rightarrow \pi^+\pi^-\pi^0$ [23]. Notably, J/ $\rightarrow e^+e^-$ is the decay channel where the charm quark was discovered in 1974 [24, 25]. Additionally, the decay modes of charmed meson D, such as $D^0 \rightarrow K^-\pi^+$ and $D^+ \rightarrow K^-\pi^+\pi^+$ [26, 27], are provided as they are the main single-tag reconstruction channels widely used in charm physics analysis. Correspondingly, the main decay channel of the lightest charmed baryon $\Lambda c^+ \rightarrow p^+K^-\pi^+$ [28] is provided. The dataset also contains the semi-leptonic process $\Lambda c^+ \rightarrow ne^+$ e [29]. The neutron in the final state has a clear signature in the EMC with a distinct hit distribution, and this physics process has been applied in neutron identification studies at BESIII based on deep learning techniques [29].

The BESIII detector has collected a large amount of real data in the τ-charm energy region, with the highest integrated luminosity at energy points of 3097 MeV, 3686 MeV, and 3773 MeV, corresponding to the production thresholds of J/ , (3686), and $D\underline{D}$, respectively. The dataset includes real data at these energy points with approximately 150 events at each energy point, as well as cosmic ray backgrounds and Quantum Electrodynamics (QED) processes such as $e^+e^- \rightarrow e^+e^-/\ ^+\ ^-/\gamma\gamma$. Additionally, the dataset provides real data for an important physics process: the tetraquark state Zc(3900) decay events [30, 31]. For some events in the detector, the particle trajectories appear to resemble the shape of , which are widely used in outreach and journal cover images [32].

Data Format and File Structure

The filenames in the dataset follow a convention that includes the data type, physics process, and recorded information, such as "Data_{{{Zc3900}}*{{track}}*}{hit}.txt" . Each text file contains detailed reconstructed track and hit information for every event, sequentially organized as follows: run number, event number, MDC track, MDC hit, TOF track, TOF hit, EMC track, EMC hit, MUC track, and MUC hit. Track and hit information contains various digital quantities such as momentum, energy, and detector identifier. The definitions of digits in text files are stored in the specification document "Parameters_{definition}.txt" in the dataset, which users can refer to for detailed data information.

IV. Data Accessibility and Technical Validation

With approval from the BESIII collaboration, the dataset has been made public via Science Data Bank at https://doi.org/10.57760/sciencedb.21486. Users can obtain additional information from the BESIII official website (http://bes3.ihep.ac.cn/).

Technical Validation

The dataset records several important physics processes as well as major background events that exhibit remarkable features in visualization. We present various visualization techniques to demonstrate the validity of the dataset. The physics processes can be visualized using BesVis, as shown in Fig. 3 [Figure 3: see original paper], which displays an X-Y cross-section view of the BESIII detector where tracks and hits can be intuitively displayed.

In addition to two-dimensional (2D) display, three-dimensional visualization is provided using OpenGL, as demonstrated in Fig. 4 [Figure 4: see original paper] for a $\Lambda c^{+} \rightarrow ne^{+}$ e event containing four charged tracks and a clear cluster of EMC hits from neutron energy deposit.

Moreover, new visualization techniques have been developed to display these physics processes. For example, Unity-based event display has been used in BESIII [14, 33], JUNO [17], and other experiments [21], featuring local running, multi-platform deployment, and enhanced visualization effects. Fig. 5 [Figure 5: see original paper] shows the 2D display of a J/ $\rightarrow$ $^{+}$ $^{-}$ event, where the muons penetrate the MUC, and a Zc(3900) event showing approximately -shaped tracks.

Usage Notes

We propose three potential applications for this dataset. First, the real data contains not only J/ , (3686), and $D\underline{D}$ production events, but also continuum backgrounds including QED processes, cosmic ray backgrounds, and other events. The dataset can be used to study the characteristics of different physics processes and potentially search for rare processes.

Second, the dataset provides full detector response information at both the hit and track levels, with both MC simulation data and real data available. It can be used to study detector response in simulation and track reconstruction algorithms for developing data processing techniques in offline software.

Third, visualization plays a crucial role in every aspect of high-energy physics experiments. The dataset describes full event information in electron-positron collisions. Combined with the BESIII detector geometry, it can be used to develop event display software with various new visualization techniques, not only for physics research but also for physics education and public outreach.

Code Availability

The dataset is publicly available and can be accessed without restrictions. The authors encourage its use and reuse for visualization research, education, and public outreach. The code is hosted on a public repository, ensuring transparency and facilitating community contributions.

Author Contributions Statement

S.S. Sun and Z.Y. You conceived the experiment. M.H. Liao and X.N. Wang conducted the experiment, and M.H. Liao and J.S. Liu analyzed the results. All authors reviewed the manuscript.

Competing Interests

The authors declare that they have no conflict of interest.

References

[1] M. H. Ye and Z. P. Zheng, BEPC, THE BEIJING ELECTRON POSITRON COLLIDER, Int. J. Mod. Phys. A 2 (1987), 1707-1725. DOI:10.1142/S0217751X87000880.

[2] M. Ablikim et al. [BESIII], Design and Construction of the BESIII Detector, Nucl. Instrum. Meth. A 614 (2010), 345-399. DOI:10.1016/j.nima.2009.12.050.

[3] D. M. Asner, T. Barnes, J. M. Bian, I. I. Bigi, N. Brambilla, I. R. Boyko, V. Bytev, K. T. Chao, J. Charles and H. X. Chen, et al. Physics at BES-III, Int. J. Mod. Phys. A 24 (2009), S1-S794. DOI:10.1142/S0217751X09045772.

[4] M. Ablikim et al. [BESIII], Future Physics Programme of BESIII, Chin. Phys. C 44 (2020) no.4, 040001. DOI:10.1088/1674-1137/44/4/040001.

[5] Z. A. Liu, W. Gong, Y. Guo, D. Jin, L. Li, Y. Lu, Q. Qiao, K. Wang, S. Wei and H. Xu, et al., Trigger System of BESIII, DOI:10.1109/RTC.2007.4382859.

[6] J. Zou, W. Li, Q. Ma, Z. You, S. Sun, Z. Deng, X. Ji, A. Zhemchugov, W. Fang and C. Fu, et al., Offline data processing system of the BESIII experiment, Eur. Phys. J. C 84 (2024) no.9, 937. DOI:10.1140/epjc/s10052-024-13241-3.

[7] https://lhcb-comp.web.cern.ch/Frameworks/Gaudi/

[8] Y. T. Liang, K. Liu, Z. Y. You, Y. J. Mao, W. D. Li, J. M. Bian, G. F. Cao, X. X. Cao, S. J. Chen and Z. Y. Deng, et al., Track reconstruction in the

BESIII muon counter, Chin. Phys. C 33 (2009), 666-672. DOI:10.1088/1674-1137/33/8/013.

[9] L. K. Jia, Z. P. Mao, W. D. Li, G. F. Cao, X. X. Cao, Z. Y. Deng, K. L. He, C. Y. Liu, H. M. Liu and Q. G. Liu, et al., Study of low momentum track reconstruction for the BESIII main drift chamber, Chin. Phys. C 34 (2010), 1866-1873. DOI:10.1088/1674-1137/34/12/014.

[10] Y. Zhang, J. Zhang, Y. Yuan and H. Liu, Track Reconstruction Using the Hough Transform at BESIII, PoS ICHEP2018 (2019), 888. DOI:10.22323/1.340.0888.

[11] G. Guerrieri [ATLAS], Open Data at ATLAS: Bringing TeV collisions to the World, [arXiv:2502.21133 [hep-ex]].

[12] Y. Elangovan, B. Satyanarayana, R. Shinde, M. Saraf, Pathaleswar, S. T. Thoi, V. M. Datar, G. Majumder, S. R. Joshi and P. Verma, et al. Compact Cosmic Muon Tracker using Resistive Plate Chambers for Public Science Outreach, [arXiv:2503.01764 [hep-ex]].

[13] E. Govorkova, E. Puljak, T. Aarrestad, M. Pierini, K. A. Woźniak and J. Ngadiuba, LHC physics dataset for unsupervised New Physics detection at 40 MHz, Sci. Data 9 (2022), 118. DOI:10.1038/s41597-022-01187-8.

[14] K. X. Huang, Z. J. Li, Z. Qian, J. Zhu, H. Y. Li, Y. M. Zhang, S. S. Sun and Z. Y. You, Method for detector description transformation to Unity and application in BESIII, Nucl. Sci. Tech. 33 (2022) no.11, 142. DOI:10.1007/s41365-022-01133-8.

[15] Z. J. Li, M. K. Yuan, Y. X. Song, Y. G. Li, J. S. Li, S. S. Sun, X. L. Wang, Z. Y. You and Y. J. Mao, Visualization for physics analysis improvement and applications in BESIII, Front. Phys. (Beijing) 19 (2024) no.6, 64201. DOI:10.1007/s11467-024-1422-7.

[16] Z. You, K. Li, Y. Zhang, J. Zhu, T. Lin and W. Li, A ROOT Based Event Display Software for JUNO, JINST 13 (2018) no.02, T02002. DOI:10.1088/1748-0221/13/02/T02002.

[17] J. Zhu, Z. You, Y. Zhang, Z. Li, S. Zhang, T. Lin and W. Li, A method of detector and event visualization with Unity in JUNO, JINST 14 (2019) no.01, T01007. DOI:10.1088/1748-0221/14/01/T01007.

[18] M. Liao, K. Huang, Y. Zhang, J. Xu, G. Cao and Z. You, A ROOT based detector geometry and event visualization system for JUNO-TAO, [arXiv:2406.16431 [physics.ins-det]].

[19] R. M. Bianchi et al. [ATLAS], Virtual Reality and game engines for interactive data visualization and event displays in HEP, an example from the ATLAS experiment, EPJ Web Conf. 214 (2019), 02013. DOI:10.1051/epjconf/201921402013.

[20] A. M. Tadel et al. [CMS], EVE-7 and FireworksWeb: The next generation event visualization tools for ROOT and CMS, EPJ Web Conf. 245 (2020), 08027. DOI:10.1051/epjconf/202024508027.

[21] Z. Y. Yuan, Z. Yuan, T. Z. Song, T. Song, Y. J. Zeng, Y. Zeng, K. X. Huang, K. Huang, Y. M. Zhang and Y. Zhang, et al., Method for detector description conversion from DD4hep to Filmbox, Nucl. Sci. Tech. 35 (2024) no.9, 146. DOI:10.1007/s41365-024-01506-1.

[22] S. Huang and Z. You, Update of the BESIII Event Display System, J. Phys. Conf. Ser. 1085 (2018) no.4, 042027. DOI:10.1088/1742-6596/1085/4/042027.

[23] S. Navas et al. [Particle Data Group], Review of particle physics, Phys. Rev. D 110, no.3, 030001 (2024). DOI:10.1103/PhysRevD.110.030001.

[24] J. J. Aubert et al. [E598], Experimental Observation of a Heavy Particle J, Phys. Rev. Lett. 33 (1974), 1404-1406. DOI:10.1103/PhysRevLett.33.1404.

[25] C. Bacci, R. B. Celio, M. Berna-Rodini, G. Caton, R. del Fabbro, M. Grilli, E. Iarocci, M. Locci, C. Mencuccini and G. P. Murtas, et al., Preliminary Result of Frascati (ADONE) on the Nature of a New 3.1-GeV Particle Produced in e^+e^- Annihilation, Phys. Rev. Lett. 33 (1974), 1408 [erratum: Phys. Rev. Lett. 33 (1974), 1649]. DOI:10.1103/PhysRevLett.33.1408.

[26] M. Ablikim et al. [BESIII], Measurements of absolute branching fractions for D mesons decays into two pseudoscalar mesons, Phys. Rev. D 97 (2018) no.7, 072004. DOI:10.1103/PhysRevD.97.072004.

[27] G. Bonvicini et al. [CLEO], Updated measurements of absolute D^+ and D^0 hadronic branching fractions and $\sigma(e^+e^- \to D\underline{D})$ at Ecm = 3774 MeV, Phys. Rev. D 89 (2014) no.7, 072002. DOI:10.1103/PhysRevD.89.072002.

[28] M. Ablikim et al. [BESIII], Measurements of absolute hadronic branching fractions of Λc^+ baryon, Phys. Rev. Lett. 116 (2016) no.5, 052001. DOI:10.1103/PhysRevLett.116.052001.

[29] M. Ablikim et al. [BESIII], Observation of a rare beta decay of the charmed baryon with a Graph Neural Network, [arXiv:2410.13515 [hep-ex]].

[30] M. Ablikim et al. [BESIII], Observation of a Charged Charmoniumlike Structure in $e^+e^- \to \pi^+\pi^- J/$ at $\surd s = 4.26$ GeV, Phys. Rev. Lett. 110 (2013), 252001. DOI:10.1103/PhysRevLett.110.252001.

[31] Z. Q. Liu et al. [Belle], Study of $e^+e^- \to \pi^+\pi^- J/$ and Observation of a Charged Charmoniumlike State at Belle, Phys. Rev. Lett. 110 (2013), 252002. DOI:10.1103/PhysRevLett.110.252002.

[32] G. Huang et al. [BESIII], Probing the internal structure of baryons, Natl. Sci. Rev. 8 (2021) no.11, nwab187. DOI:10.1093/nsr/nwab187.

[33] J. Li and Z. You, Event Visualization with Unity in BESIII experiment, PoS ICHEP2024 (2025), 1047. DOI:10.22323/1.476.1047