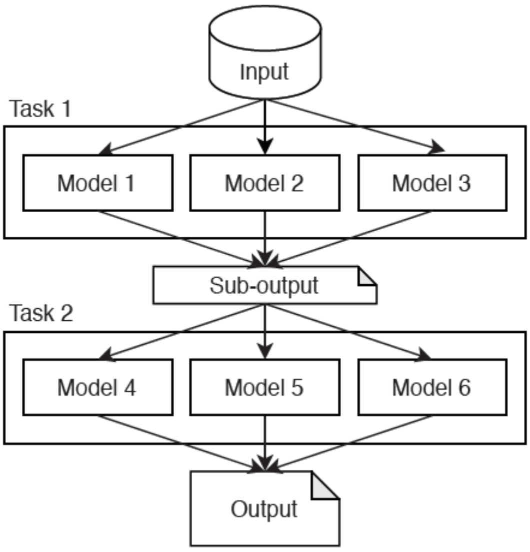
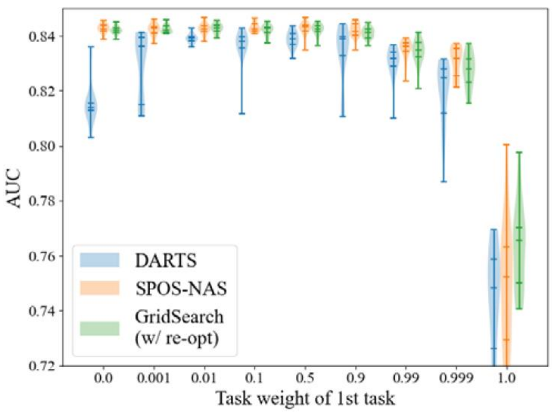
Report of the research project: Problem-solving methods using multiple AI models

R&D of Multi-AI



Regarding the "Multi-AI" we proposed, we developed a framework (right figure) that can connect and select multiple machine learning models [1]. The base technology is Neural Architecture Search (NAS), which optimizes the network architecture of machine learning models, and the MultiML framework developed in this research has three connection and selection methods: DARTS, SPOS-NAS, and ASNG-NAS.

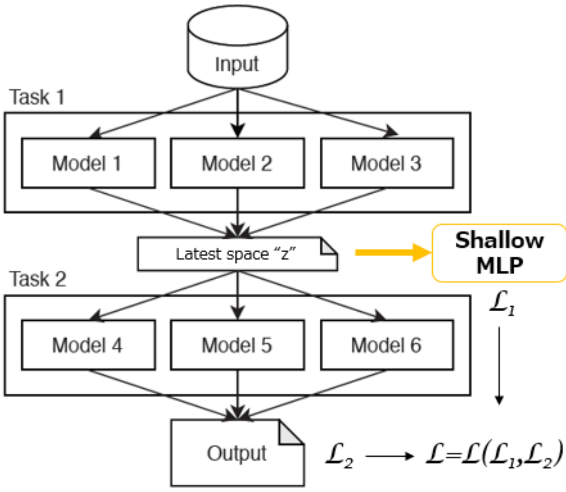
Using experimental particle physics simulation data, we developed this framework by benchmarking the problem of identifying Higgs and Z bosons that decay into tau pairs [2]. In this problem, two tasks are defined and connected. The former task is learning to calculate the momentum of the tau particle, and the latter task is learning to identify whether the tau particle pair is produced from the Higgs particle or the Z boson particle. The latter task uses the result of the former task as input and performs discrimination. First, we prepare some machine-learning models for each task. Then, the MultiML framework selects the optimal set from multiple combinations while learning individual models. This framework can be used simply for hyperparameter tuning, but it can also handle completely different machine-learning models as candidates. The right figure shows the classification performance (AUC value) of Grid Search (all the possible combinations are performed), and DARTS and SPOS-NAS in the MultiML framework as a function of the ratio of the loss function of the preceding task in the overall loss function. We can see that SPOS-NAS is comparable to the Grid Search, and DARTS is slightly worse in performance, but



sufficient results are obtained. We reported at the vCHEP2021 international conference in May 2021 [3].

[1] https://github.com/UTokyo-ICEPP/multiml
[2] https://github.com/UTokyo-ICEPP/multiml_htautau
[3] DOI: https://doi.org/10.1051/epjconf/202125103036, arXiv:2106.02301, Presentation by M. Saito, Event Classification with Multi-step Machine Learning, 25th International Conference on Computing in High-Energy and Nuclear Physics (vCHEP2021), May 18, 2021, https://indico.cern.ch/event/948465/

As a problem, there is information loss due to data transfer at the task connection. To solve this problem, the output of the previous task, in our benchmark case, the momentum is not passed directly to the subsequent task, but latent-space parameters are passed. The previous task is learned by adding another shallow network with latent parameters as input (right figure). This minimized information loss to subsequent tasks and restored overall performance. It was reported at the "Learning to Discover" international conference in April 2022 [4]. By using this framework, machine learning models developed for specific problems can be easily incorporated and optimized as part of a larger problem.



[4] Presentation by M. Saito, Study of model construction and the learning for hierarchical models, International Conference "Learning to Discover", April 27, 2022, https://indico.ijclab.in2p3.fr/event/5999/

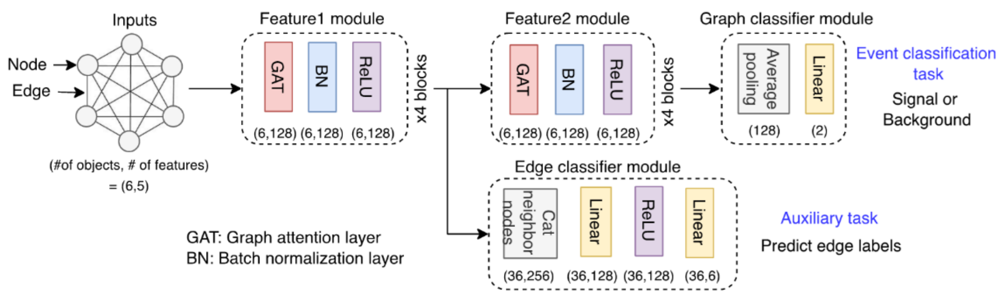
We also conducted applied research on the latest machine learning technology for data analysis of elementary particle experiments. We have worked on classification problems and generative models. The classification problem can be further subdivided into two, namely, 1) the identification of each "event" in our data, and 2) the identification of individual objects (particles) in the events. For event identification, we worked on event identification with a small amount of training data by transfer learning, event identification by combining Feynman diagrams and graph networks, and new particle search in anomaly detection using normalizing flow. For the particle identification, Higgs and large radius jet identification using Vision Transformer, and top-quark identification using graph network were studied. As a generative model, we developed a model of jet hadronization using the diffusion model. These are machine learning for specific problems, and in the future, it will be possible to use them as components of our data analysis using the MultiML framework. Each research is briefly reported below.

Event identification 1：event identification with a small amount of training data by transfer learning, and event identification by combining Feynman diagrams and graph networks

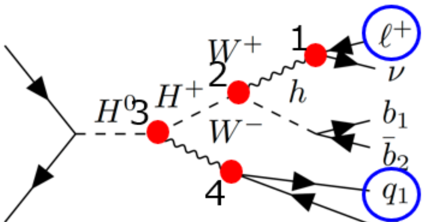
We investigated the possibility of transfer learning in data analysis of elementary particle experiments. We used a graph network that is compatible with the data structure of our experimental data. We assume that there are two event identification problems and they have similar final states. In this case, we found that one event identification can be sufficiently trained with a small amount of training data when it uses a pre-trained network of another event identification that is trained well before. We reported at the ISGC2022 international conference in March 2022 [5].

[5] DOI: https://doi.org/10.22323/1.415.0016, Presentation by T. Kishimoto, Application of transfer learning to event classification in collider physics, International Symposium on Grids & Clouds 2022 (ISGC 2022), March 21, 2022, https://indico4.twgrid.org/event/20/

Furthermore, as research to build a network that can understand the properties of elementary particle data, we developed a network that can focus on the decay



process of particles in our transfer learning studies. The developed network is shown at the bottom of the previous page. The fact that the Feynman diagram describing the event to be searched (signal event) is different from the Feynman diagram describing the background event (next figure and table) was incorporated

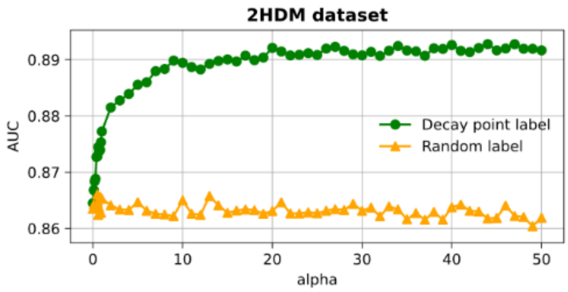


(a) 2HDM signal process

	l	v	b_1	b_2	q_1	q_2
l	0	1	3	3	4	4
v	1	0	3	3	4	4
b_1	3	3	0	1	4	4
b_2	3	3	1	0	4	4
q_1	4	4	4	4	0	1
q_2	4	4	4	4	1	0

Label matrix based on decay chains

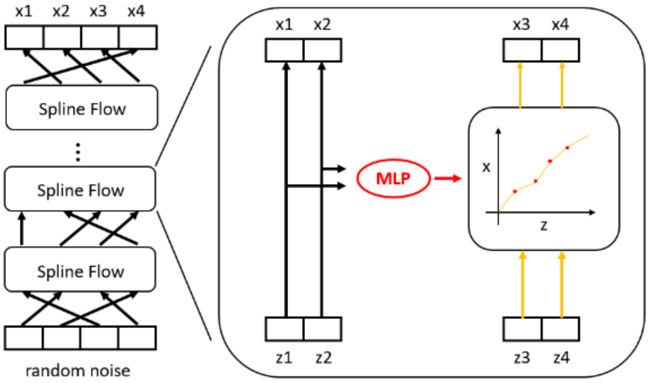
into the learning as edges in the graph network to improve the classification performance. As shown in the right figure, this method (green) improved performance by about 3%. This study was selected at the NeurIPS 2022 workshop on Machine Learning and the Physical Sciences and was presented as a poster in December 2022 [6].

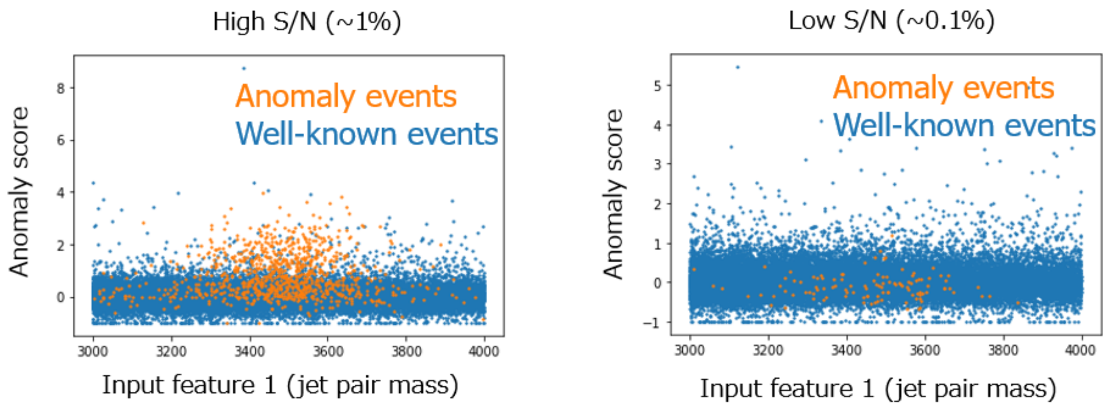


[6] https://ml4physicalsciences.github.io/2022/files/NeurIPS_ML4PS_2022_42.pdf, arXiv: 2212.08759, Presentation by T. Kishimoto, Decay-aware neural network for event classification in collider physics, NeurIPS 2022 Machine Learning and the Physical Sciences Workshop, December 4, 2022, https://ml4physicalsciences.github.io/2022/

Event identification 2 : new particle search in anomaly detection using normalizing flow

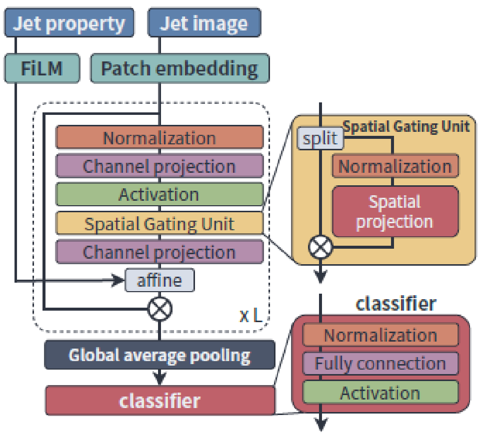
So-called "anomaly detection" using deep learning has a good possibility of being applied to search for new particles (= anomalies) in our experimental data. In this study, we developed a method to detect anomalous events (= new particles) in data by modeling the distribution for the standard model of elementary particles (= ordinarily known particles) by estimating density using normalizing flow. This method requires accurate density estimation of the distribution, but some previous studies were not able to accurately estimate the distribution depending on the shape of signal events that are detected as anomalies. To flexibly fit distributions that have sharp changes locally, we constructed models using various normalizing flow models (right figure) to improve the sensitivity of the anomaly detection. However, contrary to our initial expectations, we found that flexible modeling like normalizing flow may not contribute significantly to the performance of new particle search. In the figures shown below, the signal (= Anomaly events) and background events (= Well-known events) are shown as the feature variable (jet pair mass) on the horizontal axis and the anomaly score on the vertical axis. If the S/N ratio is about 1%, it is possible to detect anomalies with the developed methods, but if it becomes a level of 0.1%, it will be very difficult.

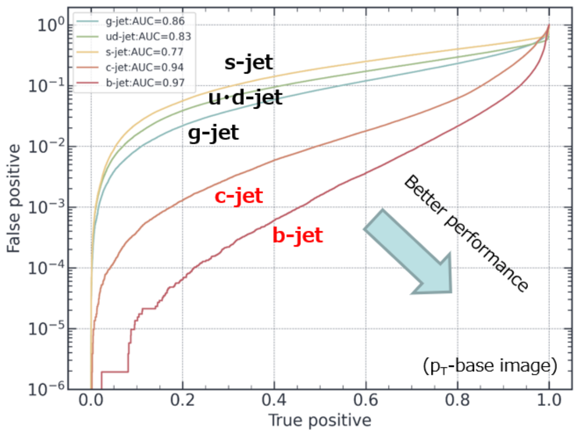




Particle Identification 1: Higgs and large radius jet identification using Vision Transformer

We applied image recognition methods such as Vision Transformer, MLP Mixer, and gMLP to the jet flavor classification problem. We constructed a general-purpose network that can learn more types of jets simultaneously (bottom left figure) and achieved the same level of discrimination power as existing methods or even better (bottom right figure). We challenged the identification of charm quarks, which will be important in future research on the Higgs boson, and the identification of storage quarks, which have even lighter mass. We reported at ISGC2022 in March 2022 [7]. In addition, we studied the identification of jets that spread over a wider area, so-called large-radius jets, to capture internal structures that are made by particles inside a jet. In this study, initially, we used methods specialized for converting our experiment data into images, but we realized that this approach limited performance due to the sparsity of our data. We moved to non-image Transformer networks instead of images.





[7] DOI: https://doi.org/10.22323/1.415.0031, Presentation by M. Morinaga, Study for jet flavor tagging by using machine learning, Symposium on Grids & Clouds 2022 (ISGC 2022), March 24, 2022, https://indico4.twgrid.org/event/20/

Particle Identification 2: top-quark identification using graph network

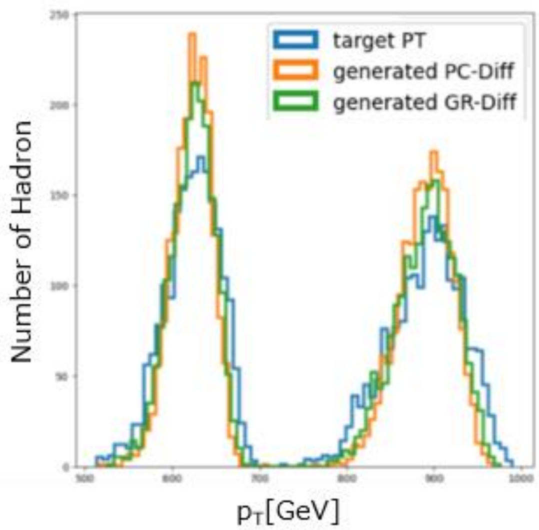
In parallel with the R&D mentioned above, we studied a network that can utilize the invariant properties of elementary particles for the particle-flavor classification problem. For the problem of top-quark identification, we conducted basic research on networks that guarantee Lorentz invariance and Lorentz equivalence using graph networks. We obtained hints such that introducing an invariant quantity for a system parallel to the beam axis, which is not Lorentz invariant, leads to performance improvement even in a network with relatively small parameters.

Generative model: modeling of jet hadronization using diffusion model

The generation process of jets, especially the process of hadronization in the low-energy region where the perturbation theory fails, has not yet been successfully described by fundamental equations (quantum chromodynamics). We considered a method to reproduce this process using machine learning and conducted research using a method called the diffusion model.

Our model used kinematic information of a single quark or gluon as input. As shown in the below figure (flow-chart), this input is first mapped in the latent space and then multiple particles are generated with a diffusion model.

The figure on the right shows the momentum distribution of the generated particles, and we can see that the "targets" are reproduced to some extent.



Prospects

Participating in Beyond AI not only improved our team's research and technical capabilities in machine learning and deep learning but also visibilities in our research community. All the studies we reported have room for improvement and we will continue to study them. In addition, we plan to actively conduct international joint research using the personal connections that we have established through this research. During the period of this research, we were not able to sufficiently work on the application using the “MultiML” framework and the improvement of the framework itself. In the future, we would like to develop it as a framework that can be used easily by users as one of the “differentiable computational model frameworks”.

Hosted or co-sponsored domestic and international workshops related to machine learning

1. Domestic workshop ML@HEP (hosted) July 8-9, 2022, The University of Tokyo, https://indico.cern.ch/event/1162214/
2. International workshop ML at HEP (co-sponsored) Feb 23-24, 2023, KEK (Tsukuba), https://kds.kek.jp/event/44830/