R&D of the Flavor-tag Method based on Machine Learning for High Energy Experiments

Osaka City U.[A] , NITEP Osaka City U[B] , RCNP Osaka U.[C] , IDS Osaka U.[D]

N.Kishida[A],M.Iwasaki[ABCD],Y.Nakashima[D],N.Takemura[D],H.Nagahara[D],T.Nakano[CD]

RCNP/IDS DNN Project

Analysis using Machine Learning

In high energy experiments, several Machine Learning (ML) methods are applied for the signal identification.

- "cut based" analysis (area division)
- Neural Network (NN, shallow NN)
- Boosted Decision Tree

These ML methods are based on the high-level feature data (high-level data, physics-inspired engineered) .

- Information of data may be degraded
- May not use all capabilities of data

Machine Learning using **low-level data** (pre-processing data), is expected to show better identification performance

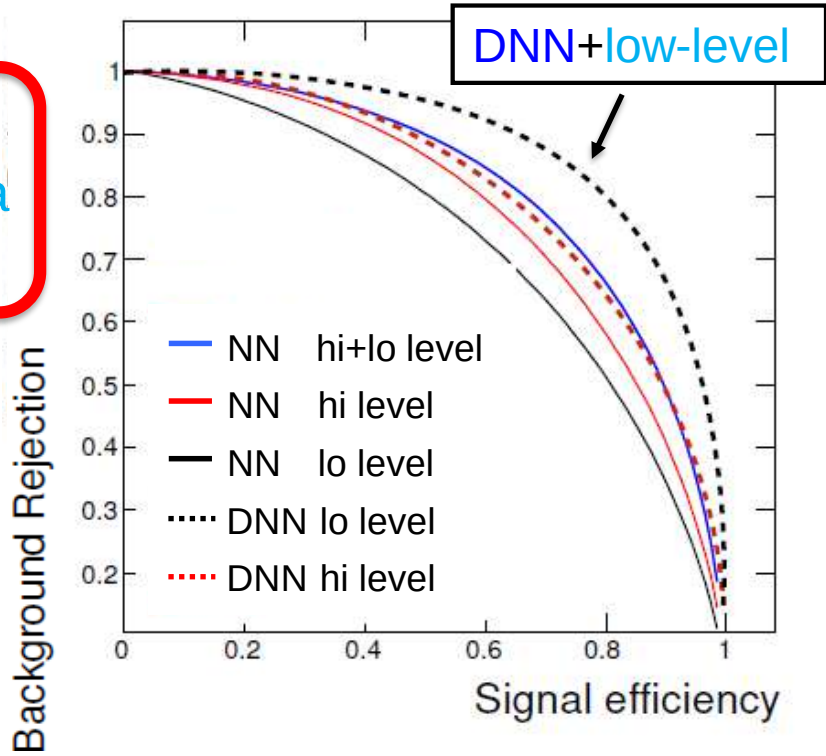
Analysis using Machine Learning

Deep learning (Deep Neural Network)

enables ML using **low-level data**.

Recent studies show the identification performance for the DNN with low-level data is better than (shallow) NN

b-jet identification in LHC



D. Guest, K. Cranmer, D. Whiteson, Annu. Rev. Nucl. Part. Sci. **68** 1-22 (2018)

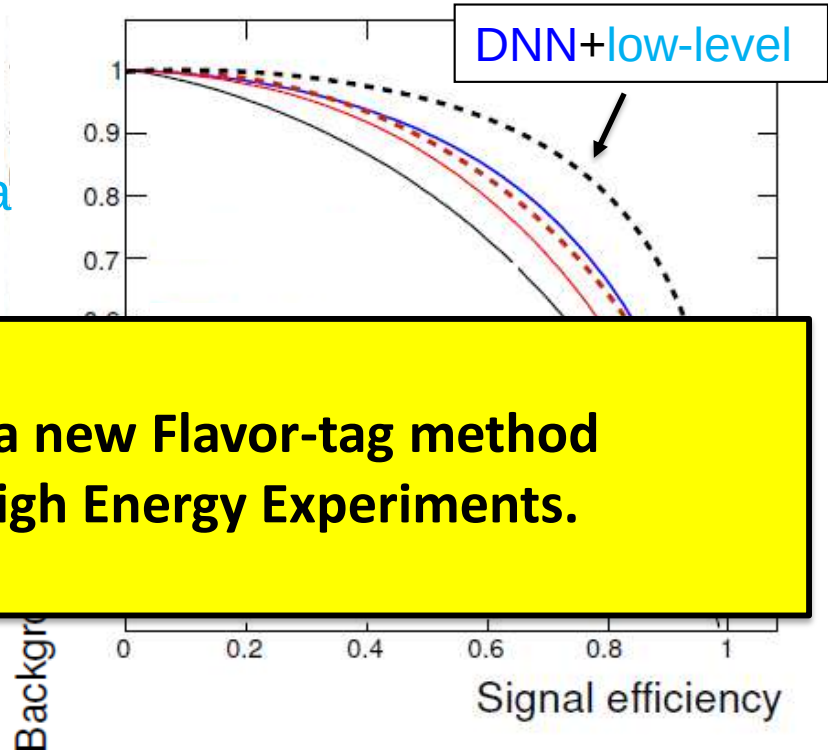
Analysis using Machine Learning

Deep learning (Deep Neural Network)

enables ML using **low-level data**.

Recent studies show the identification performance for the DNN with low-level data is better than (shallow) NN

b-jet identification in LHC



In this study, we have developed a new Flavor-tag method using DNN + low-level data for High Energy Experiments.

D. Guest, K. Cranmer, D. Whiteson, Annu. Rev. Nucl. Part. Sci. **68** 1-22 (2018)

New Analysis Method using DNN with Low-Level Data

In collaboration with information scientists,
we have developed a new analysis method
using DNN with low-level data for High Energy experiments

4-momentum and **production position of particles**
are used as low-level data for DNN input in the method

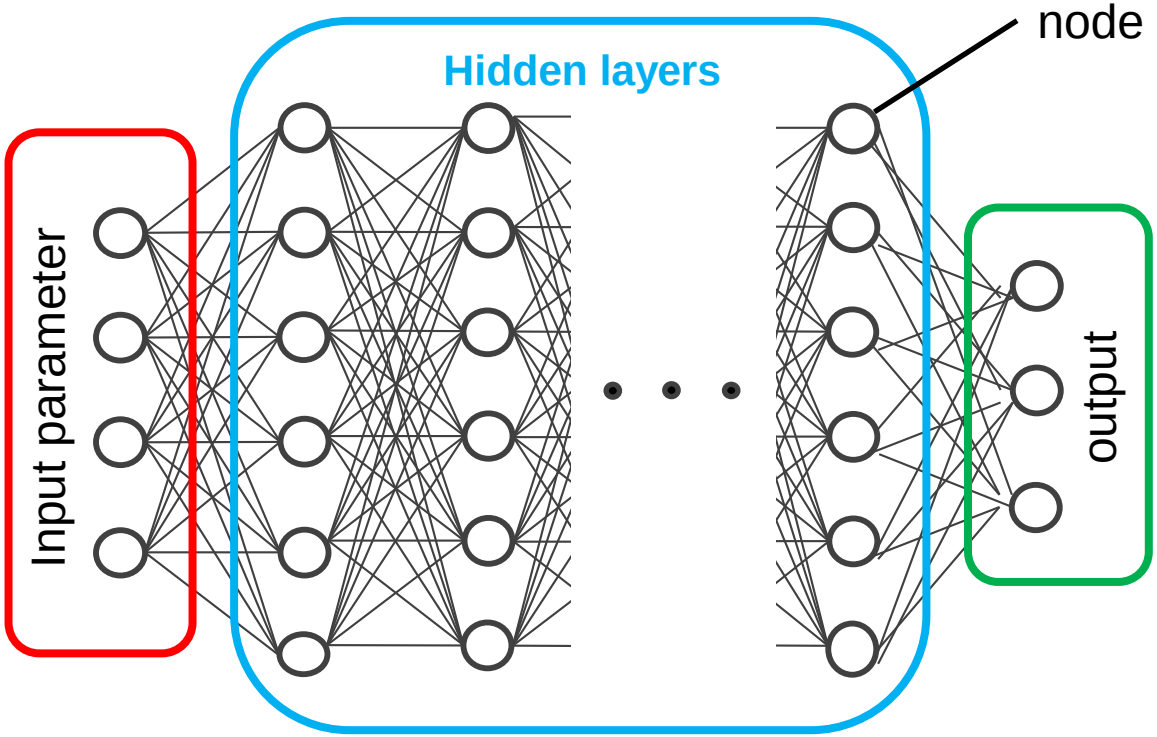
We apply the method to the Jet Flavor-tag in ILC

Neural Network(NN)

One of the machine learning technique, using a layered network structure, based on neuron (node) and its connection, expressed in a mathematical model.

NN is consist with input, hidden, and output layers.

NN with multiple hidden layers is called **DNN (Deep NN)**.



Input layer
Input data

Output layer
Output " ~ likeness"

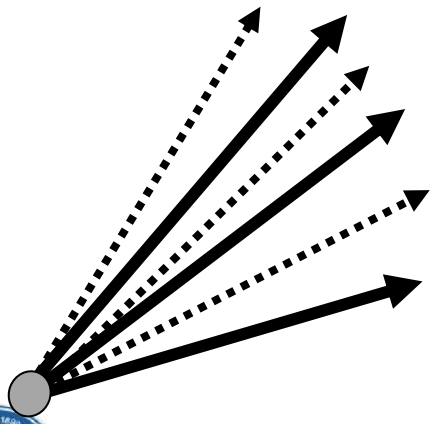
DNN application to the High Energy Experiments

In data science, the modern ML technology such as DNN has been developed for analysis of image, natural language, etc.

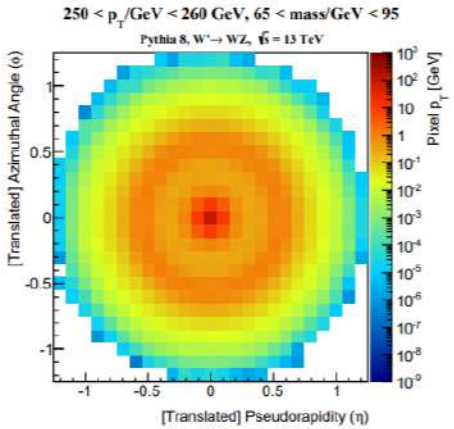
→ Many tools exist for image processing, natural language analysis..

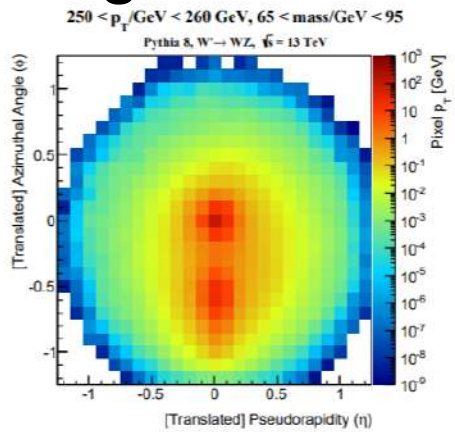
To apply DNN to the High Energy Experiment data processing, the HEP experimental data is converted to **"image data"** to use the modern DNN tools, e.g. CNN, in many cases.

Particle Jet



Jet as image





D. Guest, K. Cranmer, D. Whiteson, Annu. Rev. Nucl. Part. Sci. **68** 1-22 (2018)

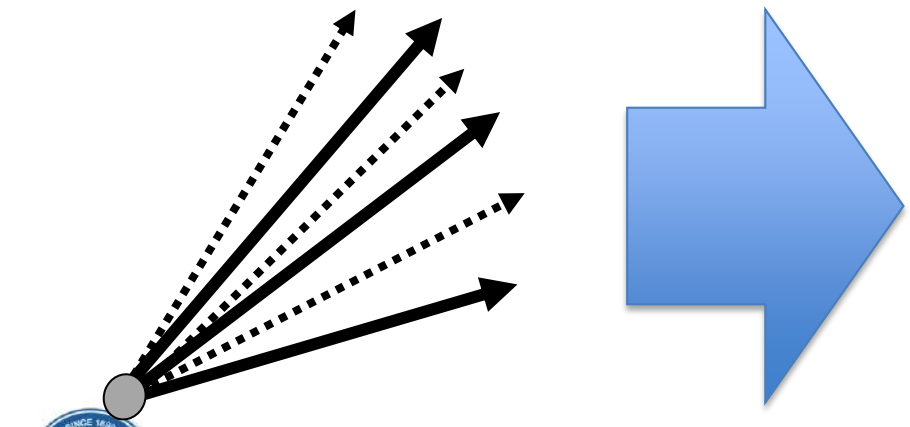
DNN application to the High Energy Experiments

Problems **If we convert the HEP data to the image data,**

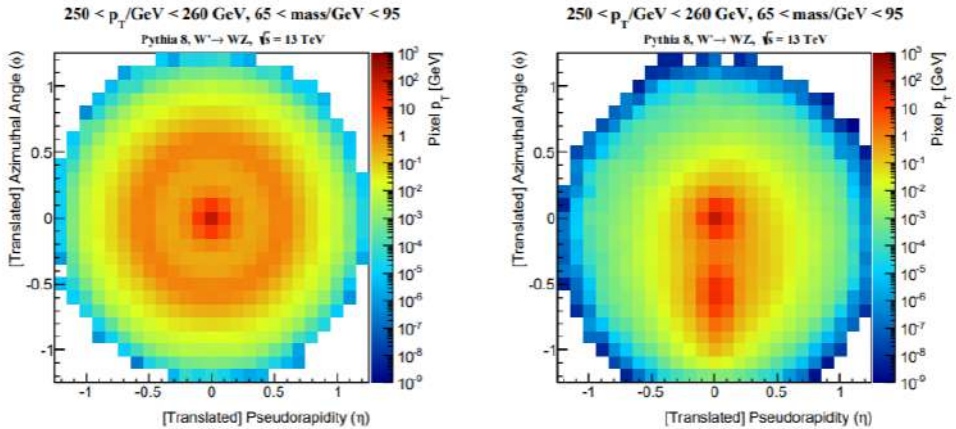
- **Lose position resolution**
- Arbitrary # of particles, with arbitrary direction, energy… produced

→ **Huge amount of data is necessary for training**

Particle Jet



Jet as image



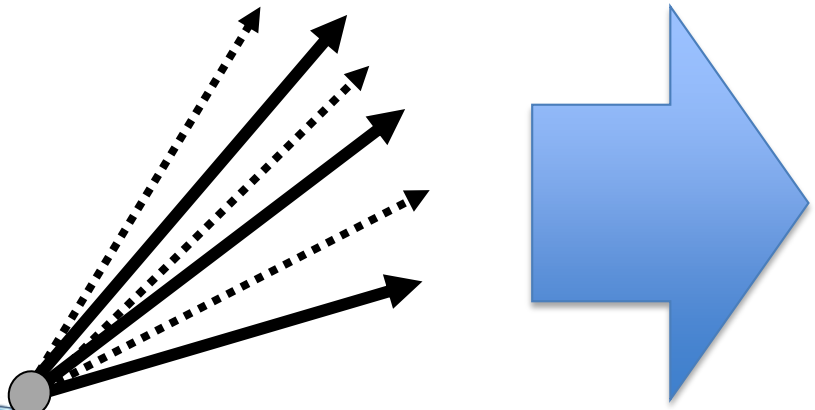
D. Guest, K. Cranmer, D. Whiteson, Annu. Rev. Nucl. Part. Sci. **68** 1-22 (2018)

New method to apply DNN to the HEP data processing

We have developed the new method to apply DNN to the HEP data analysis

We input the 4-momentum and position of particles, as low-level data, to DNN

Particle Jet



Jet as low-level data

Part1 (E, Px, Py, Pz, x, y, z)
Part2 (E, Px, Py, Pz, x, y, z)
Part3 (E, Px, Py, Pz, x, y, z)
……

Jet Flavor-Tag using DNN with low-level data in ILC

Acknowledgement
The MC data for this study is from T.Suehara of ILD group

$e^+ e^- \rightarrow q\bar{q}$ event with ILD DBD detector geometry

$\sqrt{s}$ = 91GeV, 500 k events (1 M jets)

Input parameters for DNN

We use the following parameters for DNN input

- 4-momentum of charged particles
- closest approach of charged particles

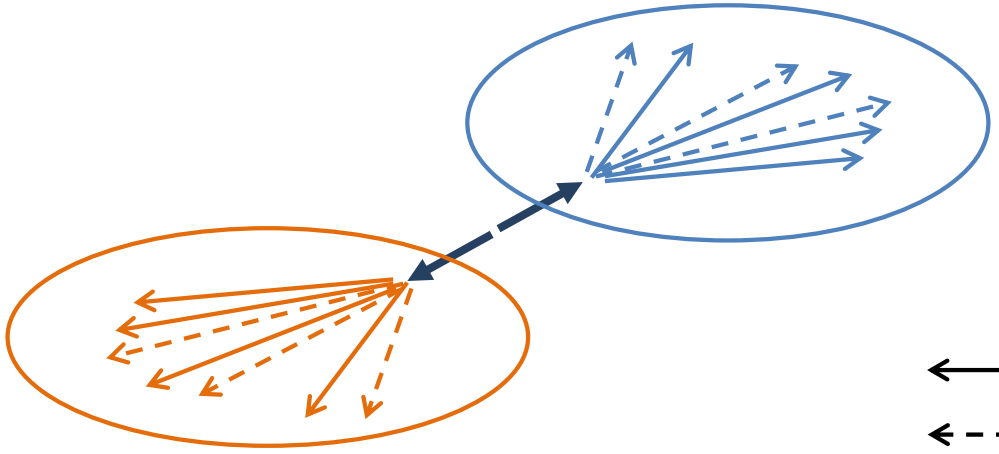
} low-level data

- features such as vertex information

→ high-level data

120 parameters, same parameter in

T.Suehara T. Tanabe NIM A808(2016) 109



$e^+e^- \rightarrow q\bar{q}$
jet event

← charged
<--- γ

Network design

We input several combinations of the High- and Low- level data to DNN

1. **High-level data**
2. **Low-level data**
3. **Low-level data + High-level data**

The network design parameter (hyper parameter), such as the number of hidden layers, is optimized for each cases.

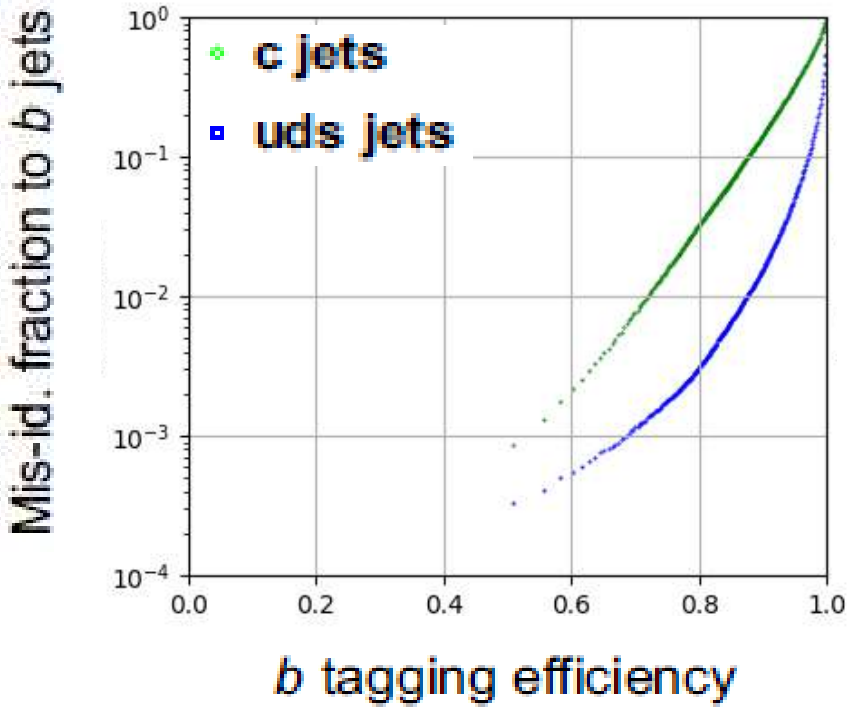
Input parameters	Number of hidden layer	Optimizer	Activation function
1. High-level	5	Adam	Relu
2. Low-level	6		
3. Low-level + High-level	9		

Result : Jet Flavor-tag with DNN in ILC

We apply DNN with low-level data input to the Jet Flavor-tag

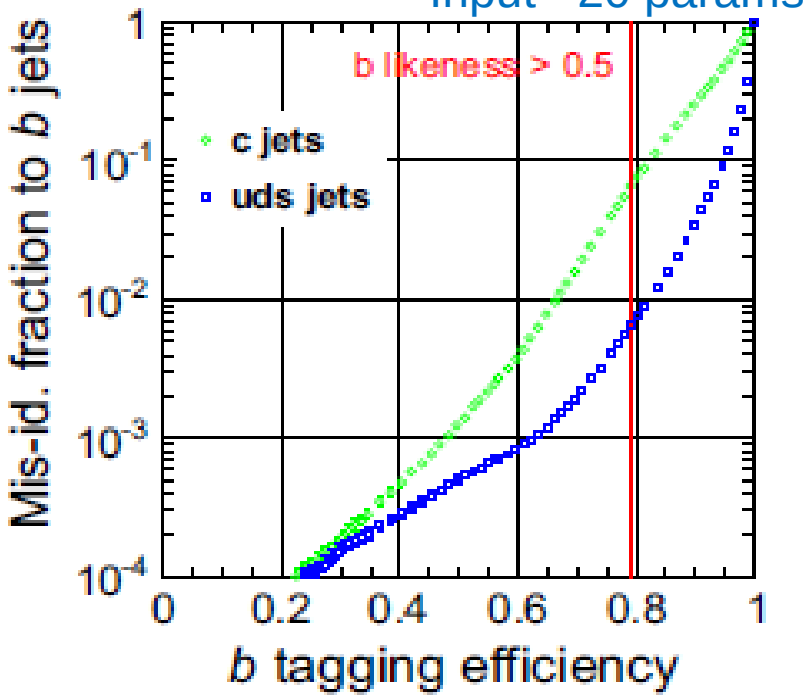
■ b-jet identification performance

DNN with Low+High level data



BDT with High-level data

Input ~20 params



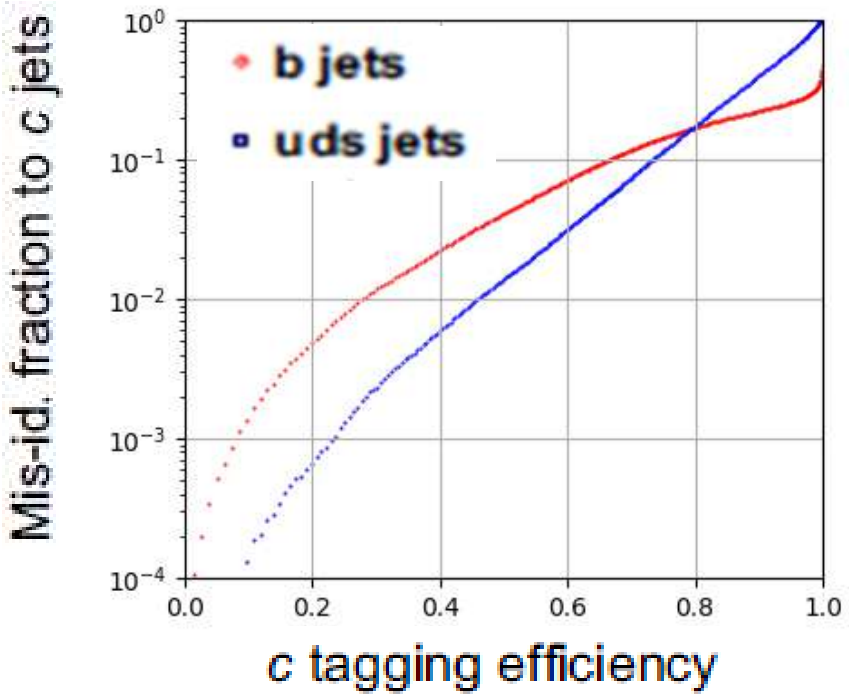
※ : T.Suehara T. Tanabe NIM A808(2016) 109

Result : Jet Flavor-tag with DNN in ILC

We apply DNN with low-level data input to the Jet Flavor-tag

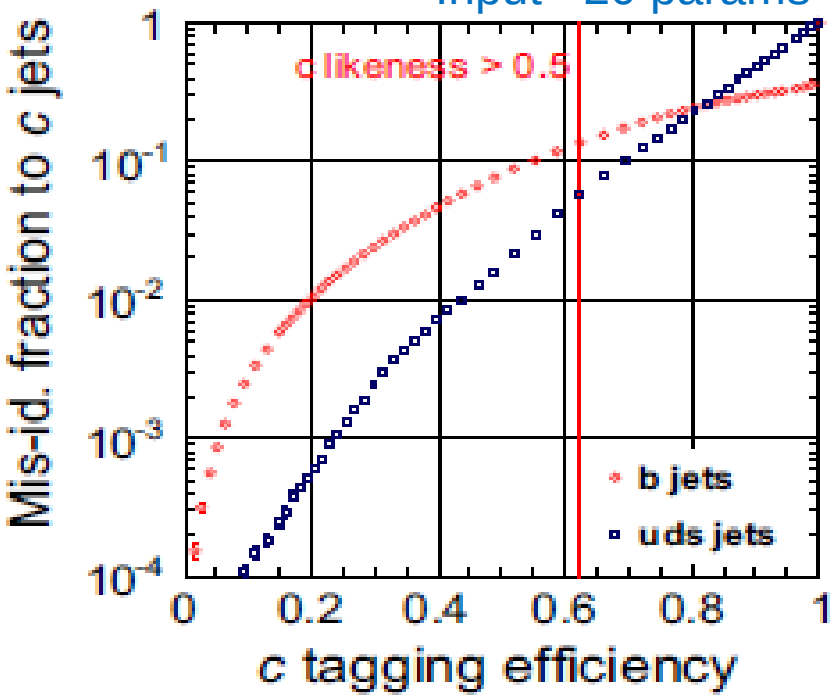
■ c-jet identification performance

DNN with Low+High level data



BDT with High-level data

Input ~20 params



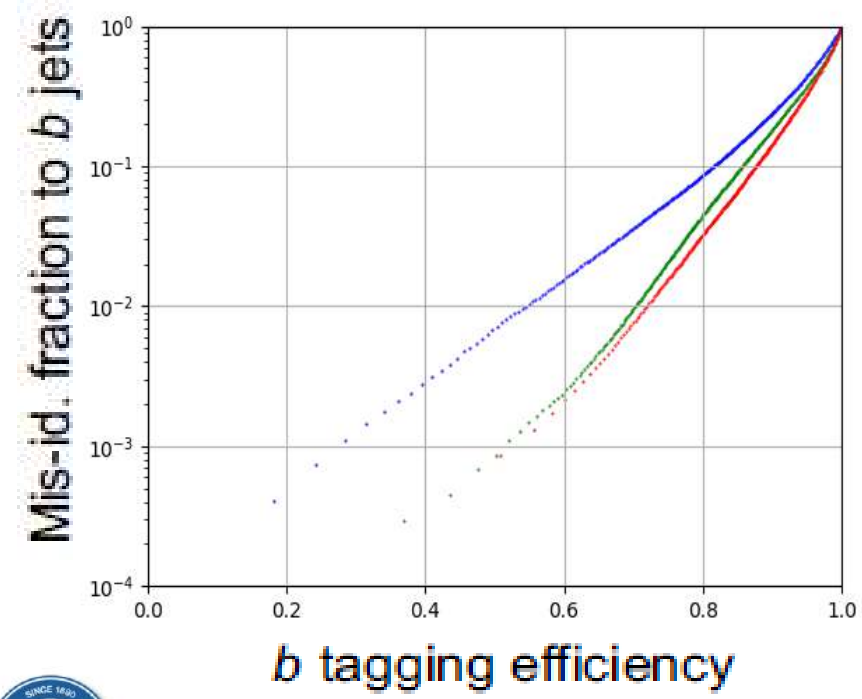
※ : T.Suehara T. Tanabe NIM A808(2016) 109

Result : Jet Flavor-tag with DNN in ILC

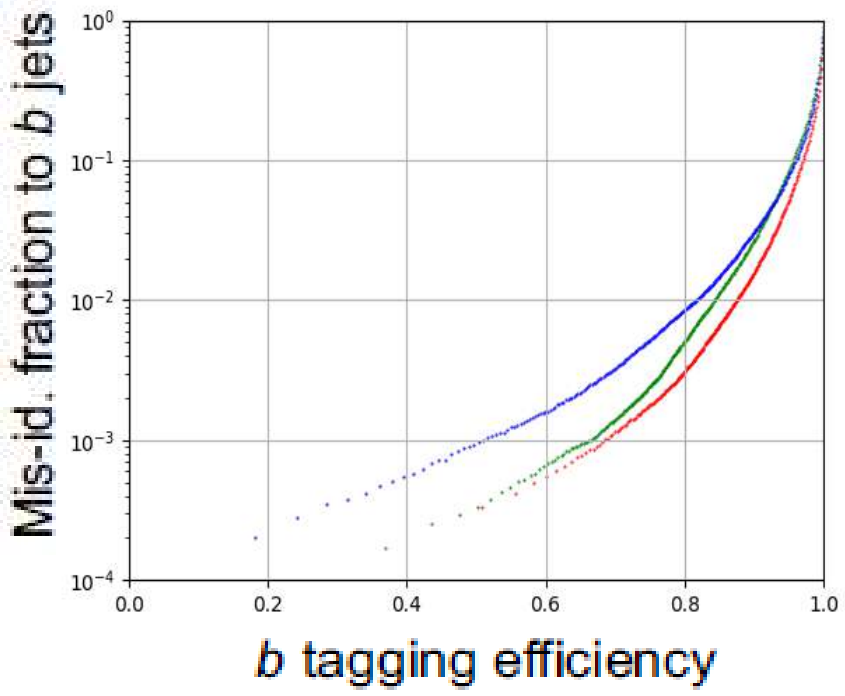
■ **b-jet** ID performance comparison

: 1. DNN High
: 2. DNN Low
: 3. DNN Low+High

for c-jet



for uds-jet

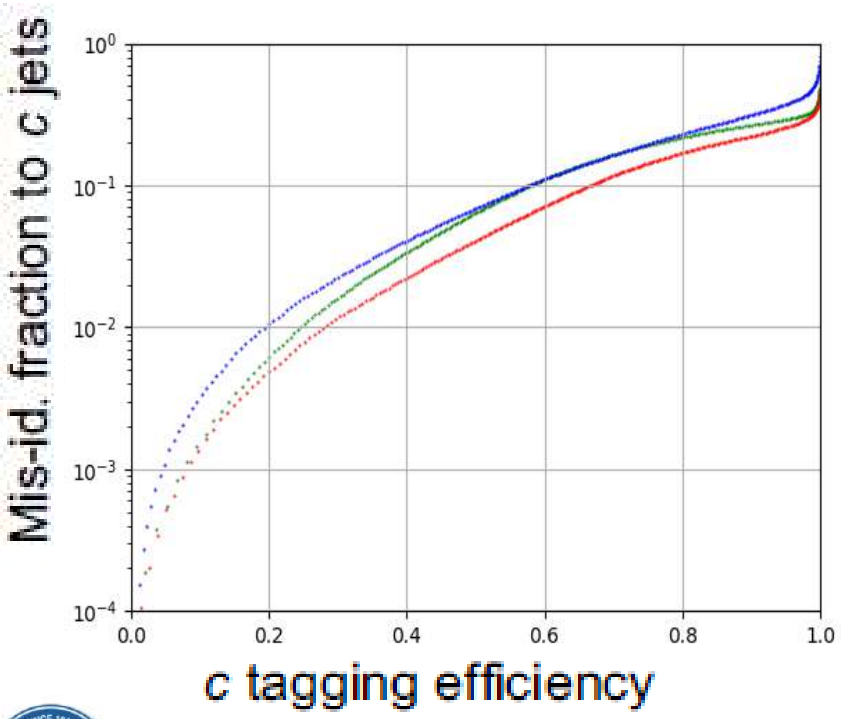


Result : Jet Flavor-tag with DNN in ILC

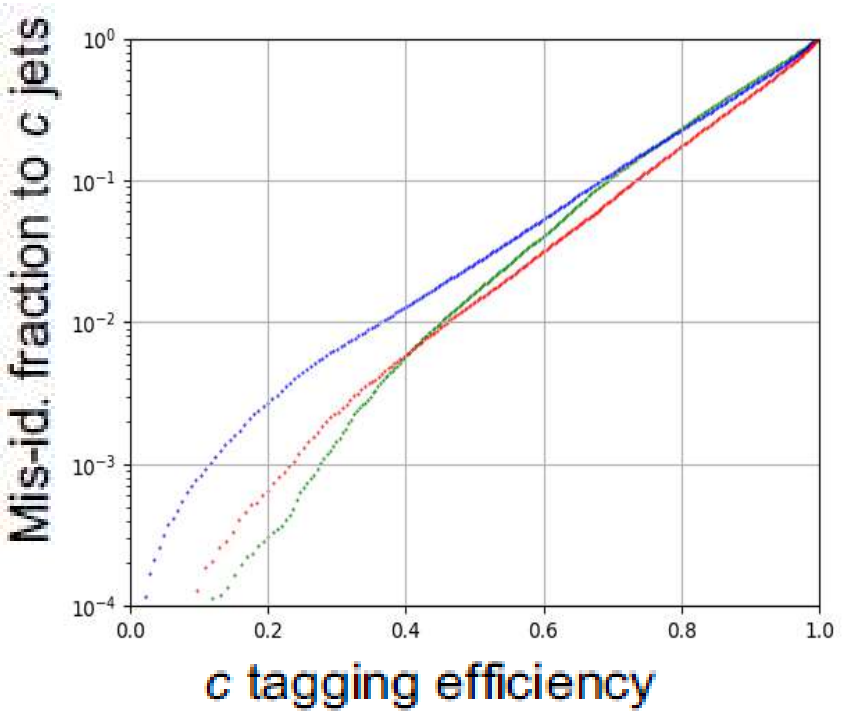
■ **c-jet** ID performance comparison

: 1. DNN High
: 2. DNN Low
: 3. DNN Low+High

for b-jet



for uds-jet



Result : Jet Flavor-tag with DNN in ILC

■ Summary of flavor-tag performance for each jet

efficiency	jet	Mis-id fraction			
		DNN 1 high	DNN 2 low	DNN 3 High + low	BDT with high
b_tag efficiency ~0.8	c-jet	0.04	0.09	0.03	0.08
	uds-jet	0.005	0.009	0.003	0.008
c_tag efficiency ~0.6	b-jet	0.11	0.11	0.07	0.10
	uds-jet	0.04	0.05	0.03	0.06

- **Using DNN, we get better Jet Flavor-tag performance**
- **DNN with low-level is comparable to BDT with high-level**
- **DNN with Low + High gives the highest performance**

Summary

we have developed a new Flavor-tag method using DNN + low-level data for high energy experiments

- We use DNN with Low-level data and/or High-level data.
- As the Low-level data, we input 4-momentum and closest approach of particles to DNN
- In the ILC Jet Flavor-tag, DNN with simple Low-level input shows the comparable performance of BDT with ~20 High-level input
- DNN with High- + Low-level data gives the best identification performance

ありがとうございました。

Simulation data

ILC MC simulation data

- $\sqrt{s} =$ 91 [GeV]
- Event generation $e^+e^- \rightarrow qq$ (q = u,d,s,c,b)
- bb event : 1,000,000、 cc event : 1,000,000、 uds event : 1,000,000
- ILD : DBD detector geometry

Identify $b\bar{b}$,$c\bar{c}$,uds jets

- Use the different data set for training and test(validation)

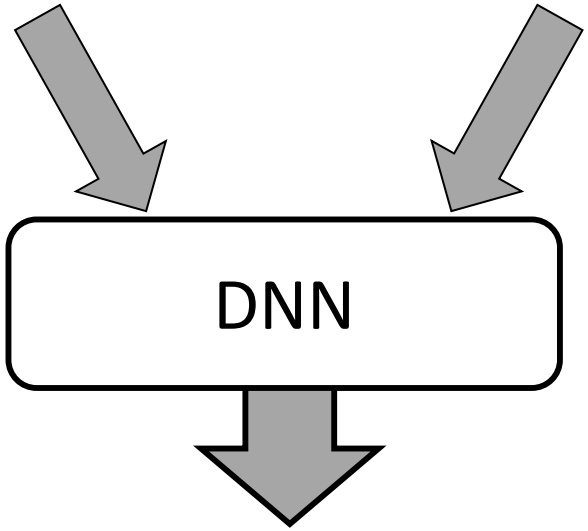
Network design

low-level-data

Part1 (E, Px, Py, Pz, x, y, z)
Part2 (E, Px, Py, Pz, x, y, z)
Part3 (E, Px, Py, Pz, x, y, z)
……

high-level-data

data1, data2, data3, …..

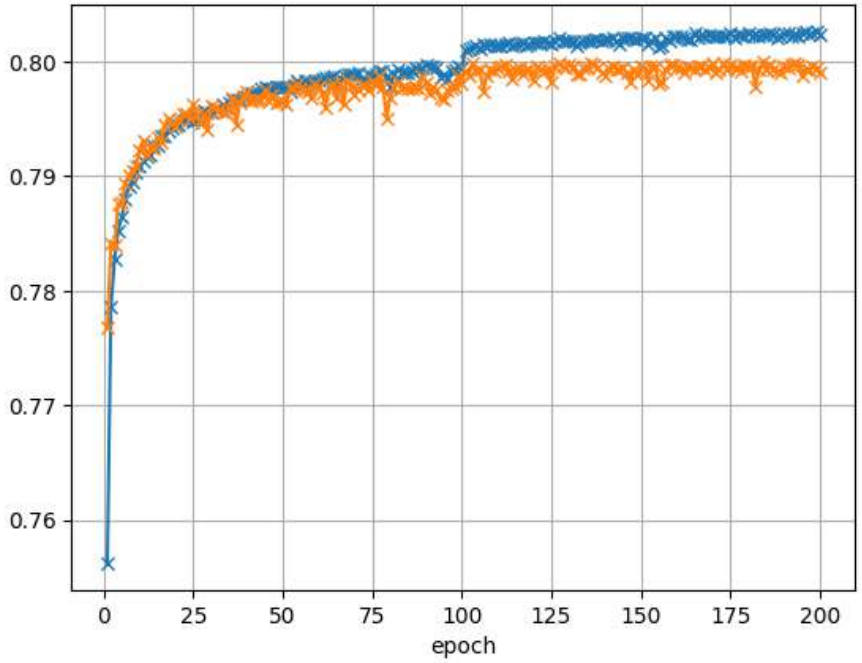


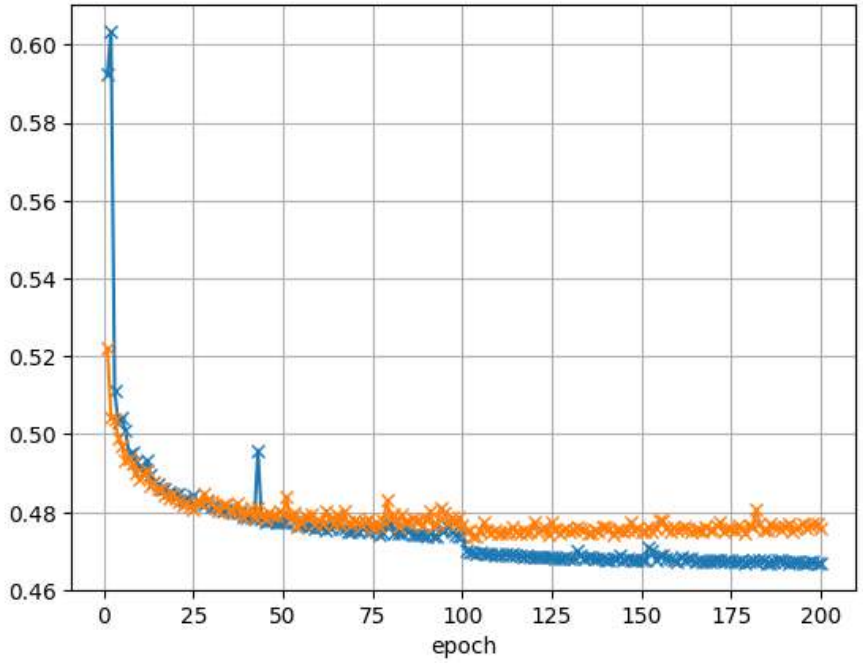
We use Chainer
to implement the DNN

likeness as output

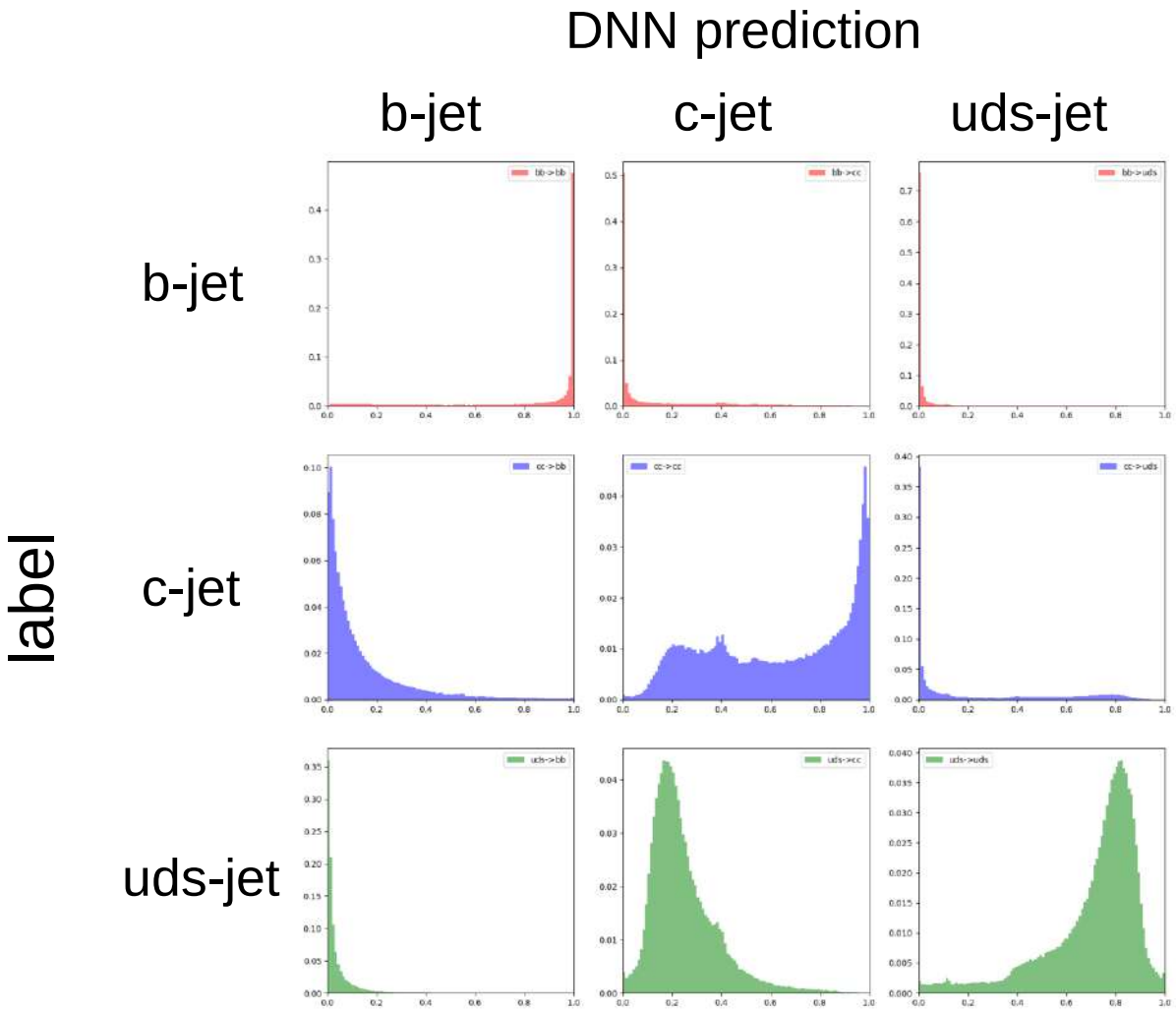
- b-jet likeness
- c-jet likeness
- uds-jet likeness

Learning curve for high-level data

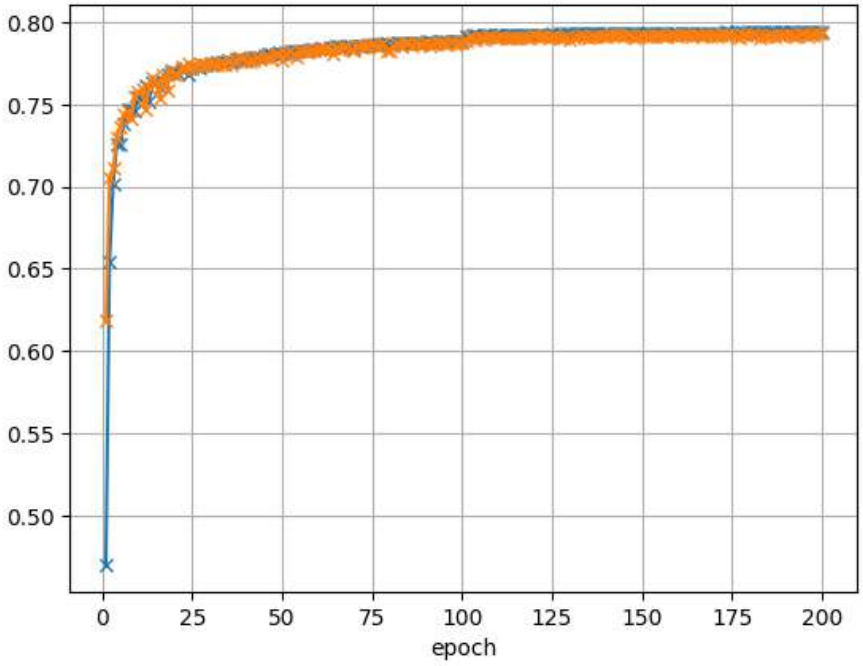


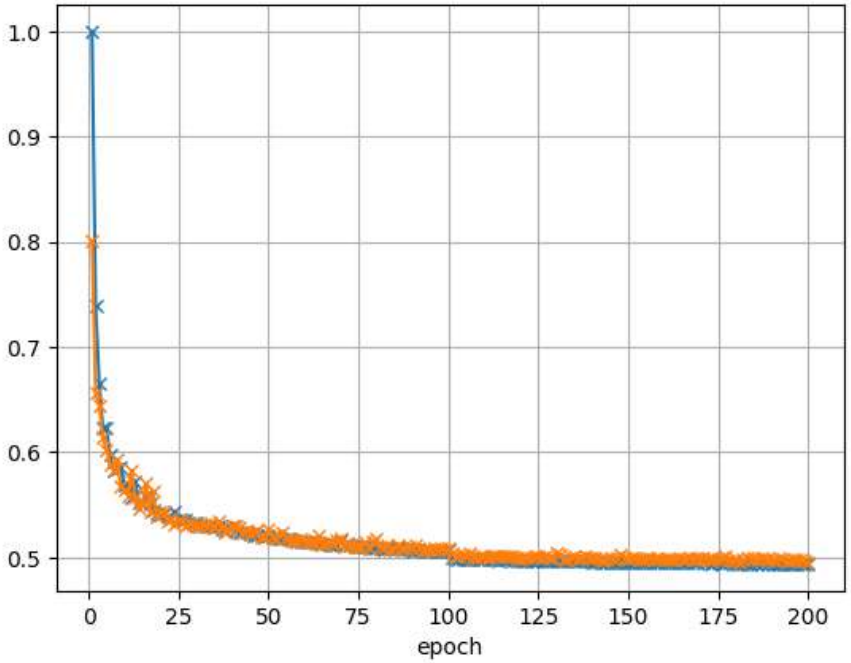


DNN output for high-level data



Learning curve for low-level data





DNN output for low-level data

DNN prediction

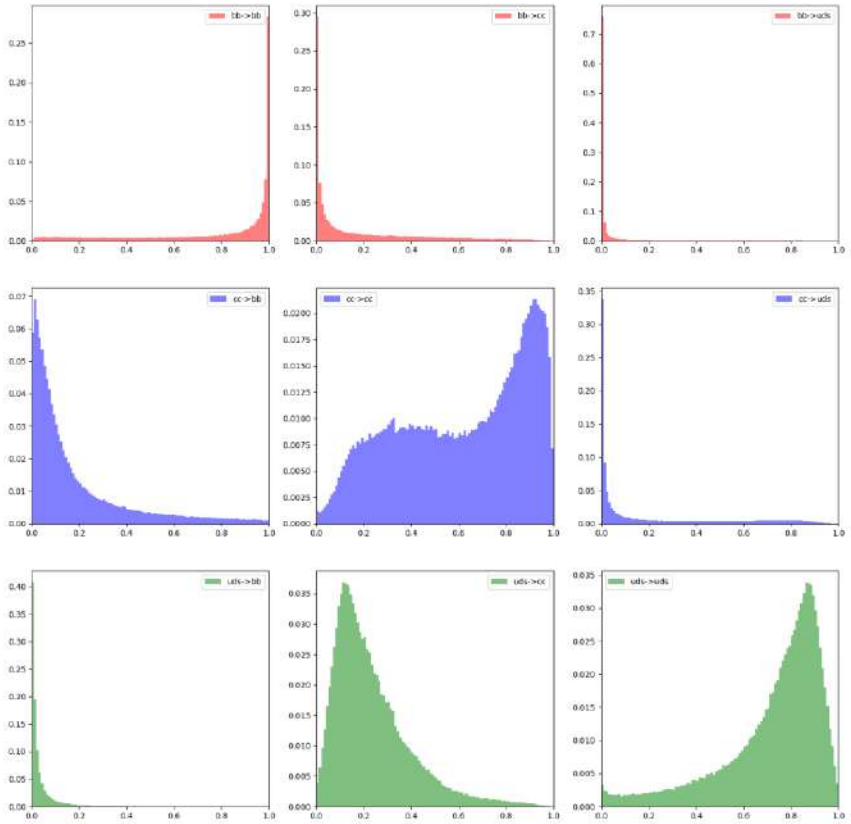
b-jet c-jet uds-jet

label

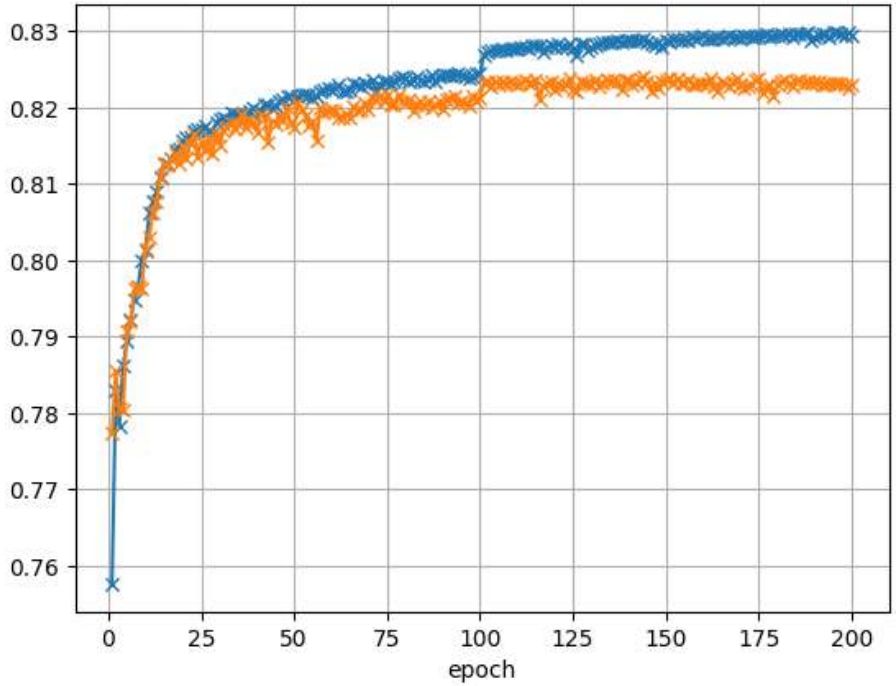
b-jet

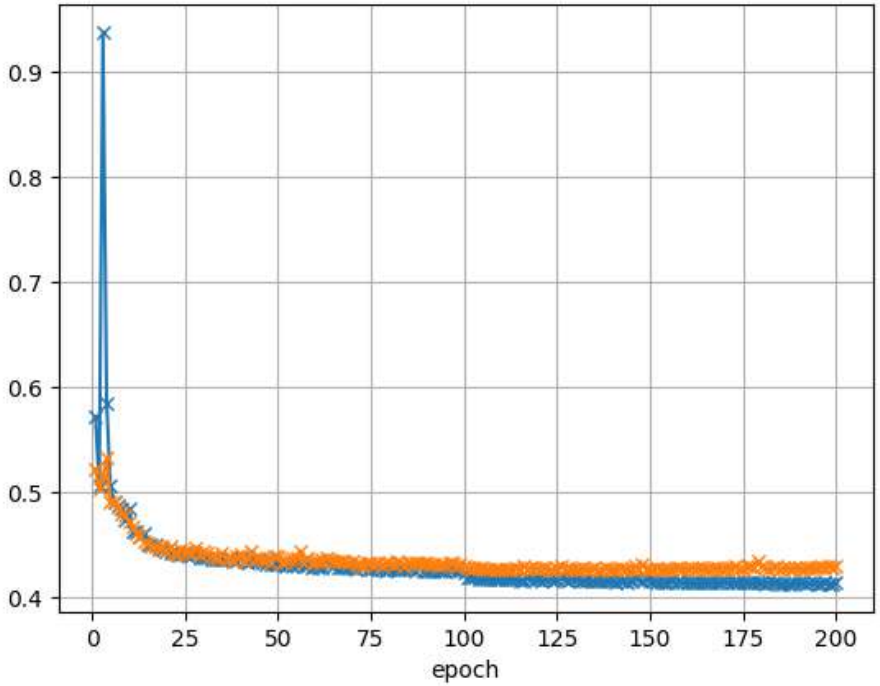
c-jet

uds-jet



Learning curve for Low− + High− level data





DNN output for Low- + High- level data