Novel Method of Jet Clustering for Future Energy Frontier Electron-Positron Collider

Masakazu Kurata (University of Tokyo)
kurata@icepp.s.u-tokyo.ac.jp

August 31, 2020

Letter of Interest for SnowMass2021: CompF

1 Introduction

In physics analyses of energy frontier electron-positron experiments, it is most important to reconstruct physics processes correctly from particles measured by detectors.

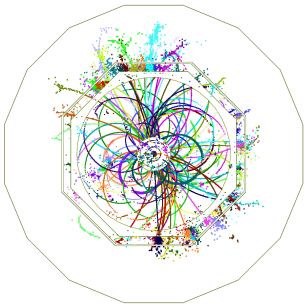

Figure 1: Event display of $e^+e^-\to t\bar{t}H$ process of ILC simulation

As the number of produced quarks increases, it is difficult to reconstruct physics processes because huge number of particles are produced by hadronization. Figure 1 shows one example of an event display of eight-quark event produced by $e^+e^-\to t\bar{t}H \to (bq\bar{q})(\bar{b}q\bar{q})(b\bar{b})$ process with the International Linear Collider (ILC) simulation. We have to reconstruct ttH process from such extremely complicated event.

Such kinds of multi-quark production events have significant sensitivity to new physics beyond the Standard Model (BSM). However, mis-assignment of the particles to their parent quarks is a big obstacle and strongly limits the sensitivity to BSM. For example, Figure 2 shows reconstructed mass distribution of Higgs boson in dominant double Higgs events, $e^+e^-\to ZHH\to(q\bar{q})(b\bar{b})(b\bar{b})$, of ILC simulation at center of mass energy of $\sqrt{s} = 500$ GeV. Due to misclustering of present jet clustering (denotes as Durham) [1][2], mass distribution of reconstructed Higgs boson (dashed line) is smeared drastically compared to mass distribution when all the reconstructed particles measured by a detector are assigned to their parent quarks (solid line). We will be able to obtain about 40% better result of Higgs trilinear coupling λ_{HHH} measurement precision if we can reconstruct Higgs mass distribution with the resolution of solid line. More than 20% deviation of λ_{HHH} from the Standard Model will lead to verification of electroweak baryogenesis [3]. In this situation, extension of the Standard Model, such as Two Higgs Doublet Model (2HDM), is necessary. Our goal is to construct general purpose novel jet clustering. Hence application of the jet clustering spreads wide range of physics analyses in electron-positron collider experiments. Therefore, it is worth tackling development of novel jet clustering method.

2 Jet Clustering Using Deep Learning

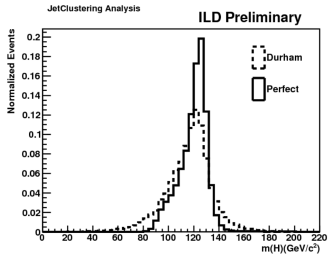

Figure 2: Comparison of Higgs mass distribution between present jet clustering (Durham) and perfect particle assignment to parent quarks (Perfect).

The remarkable progress in computer science including deep learning has been made and, as a result, has shown high performance overwhelming other methods for various tasks in speech, images, and natural language fields. Jet clustering seems very similar task as assigning correct 'color' to reconstructed particles or determining jet 'area' in certain phase space. Such kinds of tasks have been well studied and various techniques have been established in image recognition field. 'Automatic colorization' or 'semantic segmentation' algorithms using Convolutional Neural Network (CNN) can realize to solve such kinds of tasks. We can apply CNN by creating 'maps' of physics variables for each particle in certain phase space and treating them as images. By using an encoder/decoder type CNN as shown in Fig. 3, we will be able to extract features indispensable for event reconstruction effectively and automatically.

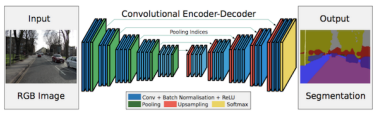

Figure 3: Schematic picture of the network architecture of encoder-decoder type CNN (refered from [4]).

From our research so far, it is found that the following points are different from the knowledge in semantic segmentation algorithm, and detailed study is necessary to improve performance of event reconstruction.

(a) In an image, there will be same number of clusters as quarks produced in an event. Therefore, it is necessary to identify the boundaries between adjacent clusters. In semantic segmentation algorithm, the boundary between the same objects is not specified when the same objects are adjacent.

(b) The boundaries of jet clusters are unclear. In image recognition, sharp edge of an object is a powerful tool and CNN will learn the difference of shapes between objects automatically by using edge information. If the distances between clusters are close, event reconstruction may not work well due to blurred boundaries. In addition, particles which locates cluster boundaries tend to have small energy and quark assignment of such kinds of particles will fail. It is necessary to develop an algorithm that defines jet cluster boundaries more effectively. Now, we reach almost same performance as Durham jet clustering using CNN. To exceed the performance of Durham jet clustering, more detailed study is necessary.

References

[1] S. Catani *et al.*, Phys. Lett. B 269, 432-438 (1991)

[2] Matteo Cacciari, Gavin P. Salam, Gregory Soyez, arXiv:1111.6097v1 [hep-ph](2011)

[3] Cheng-WeiChiang, Kaori Fuyuto, Eibun Senaha, Phys. Lett. B 762, 315 (2016)

[4] Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla, arXiv:1511.00561 [cs](2015)