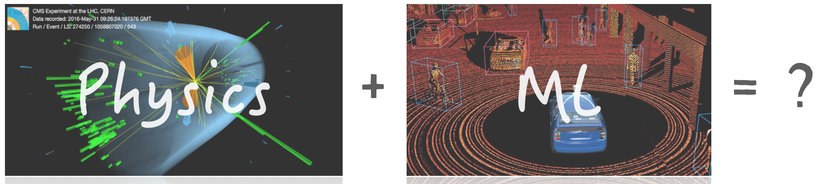
ParticleNeXt: Pushing the Limit of Jet Tagging With Graph Neural Networks

Huilin Qu (CERN)

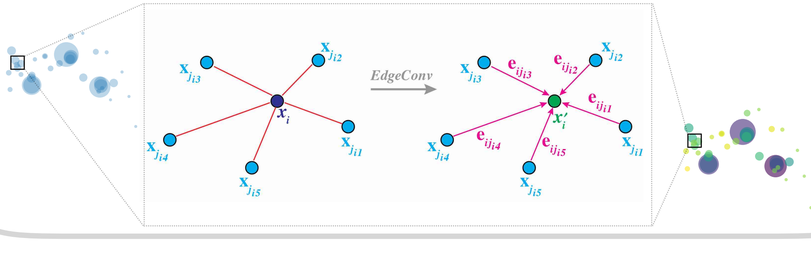
Introduction

Jet tagging: crucial for physics at high-energy colliders

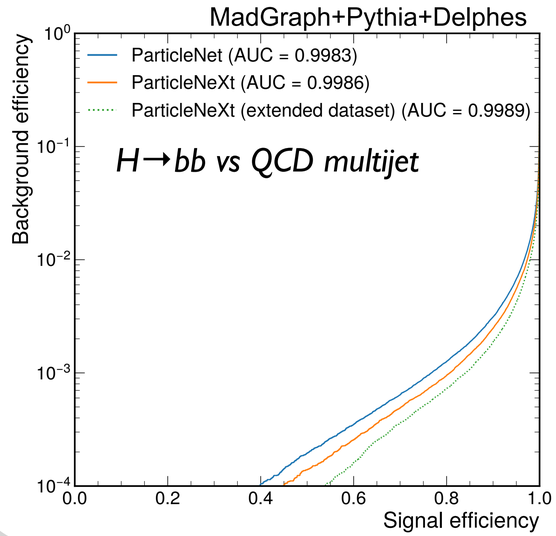


= ?

State-of-the-art: ParticleNet [Qu and Gouskos (2020)]



ParticleNeXt: Performance



- *Significant performance improvement: +30% bkg. rejection*
- *Also: crucial to use large training samples for deep taggers*
  - *another +30% when training set extended from 2M to 10M*

	Accuracy	AUC	$1/\varepsilon_b$ at $\varepsilon_s = 70\%$
ParticleNet	0.983	0.9983	1562 ± 24
ParticleNeXt	0.985	0.9986	2045 ± 29
ParticleNeXt (extended dataset)	**0.986**	**0.9989**	**2770**

	Parameters	Inference time (CPU)	Inference time (GPU)	Training time (GPU)
ParticleNet	366k	23 ms	0.30 ms	1.0 ms
ParticleNeXt	560k	30 ms	0.54 ms	1.7 ms

ParticleNeXt: next-generation of ParticleNet

Pairwise Features

- Addition of (explicit) pairwise features on the edges

ParticleNet

$e_{ij} = \text{MLP}(x_i, x_j)$

ParticleNeXt

$e_{ij} = \text{MLP}(x_i, x_j, x_{ij})$

- Examples of pairwise features:

$\Delta_{ij}^2 \equiv (y_i - y_j)^2 + (\phi_i - \phi_j)^2, \quad m^2 \equiv (p_i + p_j)^2,$

$k_T \equiv \min(p_{T,i}, p_{T,j})\,\Delta_{ij}, \quad z \equiv \dfrac{\min(p_{T,i}, p_{T,j})}{p_{T,i} + p_{T,j}}$

(use the logarithm to improve stability of the training)

Attentive Pooling

- Use attention-based pooling to increase the expressive power
  - for both the **local neighborhood pooling**, and the final **global pooling**

ParticleNet

$z_i = \text{mean}_j(e_{ij})$

ParticleNeXt

$\text{attn}_{ij} = \text{MLP}(e_{ij})$

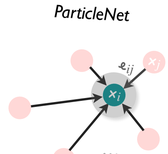
$w_{ij} = \text{softmax}_j(\text{attn}_{ij})$

$z_i = \sum_j (w_{ij}\, e_{ij})$

Multi-Scale Aggregation

- Multi-scale aggregation to better capture both short- and long-range correlations
  - perform local aggregation for the 4, 8, 16 and 32 nearest neighbors (with different attentive pooling) and combine the 4 aggregated representations with a MLP

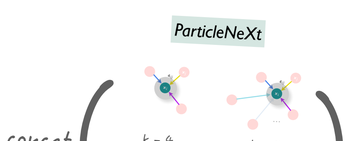
ParticleNet



$k = 16$

$x_i' = z_i$

ParticleNeXt



$x_i' = \text{MLP}(z_i^{\text{concat}})$