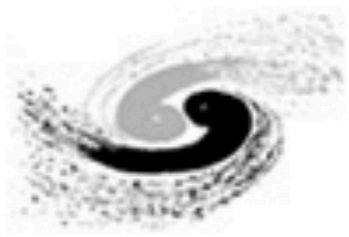
W tagging ML 项目进展

徐达

2023-01-05



中国科学院高能物理研究所
Institute of High Energy Physics Chinese Academy of Sciences

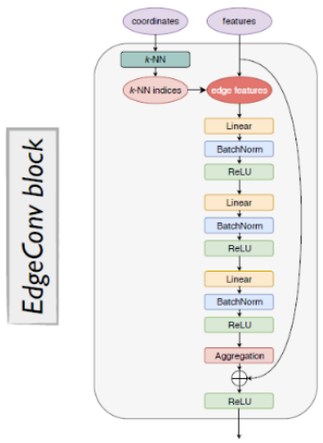
项目综述

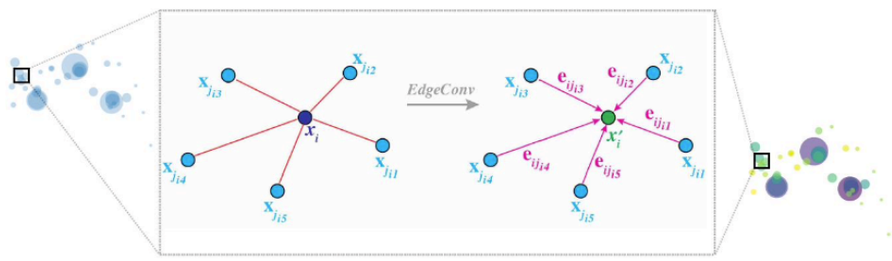
- 目标：基于图形神经网络的机器学习方法（ParticleNet, ParticleTransformer, PointNet etc），显著提升对撞机实验中 W tagging的鉴别效率，助力HL-LHC上 di-Higgs 的首次观测。
- 人员：李刚、徐达、王书栋（学生 0.5FTE）（+可能新招一位学生（0.5FTE））
- 预期成果：在国际会议上报告研究结果一次，发表文章一篇。
- 需求：总共约8万，包含单人一次国际会议、一次国内会议 约 2万，学生劳务费 约6万 （3000*12*3*0.5 ）。

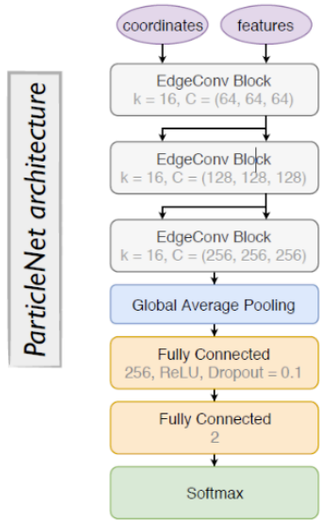
The architecture of ParticleNet

H. Qu and L. Gouskos [*Phys.Rev.D 101 (2020) 5, 056019*]

- EdgeConv can be stacked to form a deep network
  - learning both local and global structures, in a hierarchical way







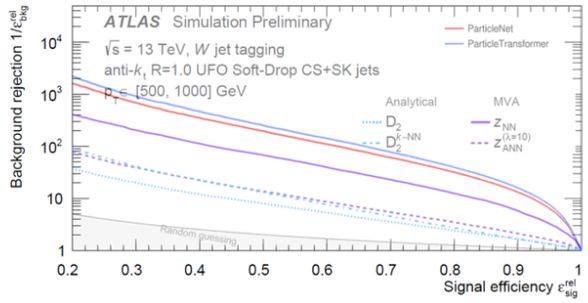
项目进展

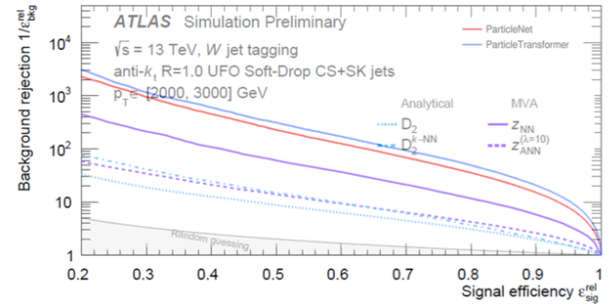
- 迄今已数次在ATLAS jet tagging组进行汇报讨论
- 使用ATLAS官方样本，基于ParticleNet/Transformer获得初步SigEff vs BkgRej结果 — 优于ATLAS现有结果
- 仍需进一步优化与完善，包括：
  - Consider more input variables: IP, charge, n_track
  - Consider mass decorrelation
  - How to improve W/Z classification
  - Try others PFN/EFN..

Classification Results

- **Compare with other UFO jet taggers (in different p_T region)**
  - Train on whole training sample set, test on different testing sample sets with different p_T regions
  - p_T region divided by using truth jet p_T: `fjet_ungroomed_truthJet_pt`

figures from: ATL-PHYS-PUB-2021-029



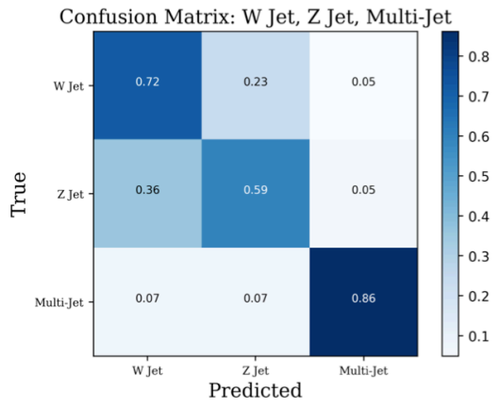


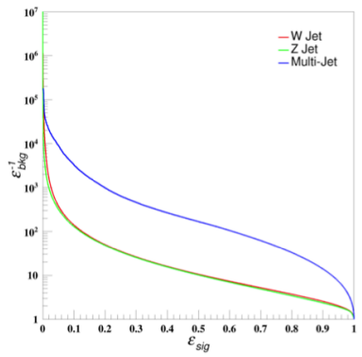
Background rejection ($1 / \varepsilon_{bkg}$) as a function of W-jet signal efficiency (ε_{sig}), in the pT range [500, 1000] GeV (left) and [2000, 3000] GeV (right).
Analytical (cyan lines): cut-based tagger (cuts on D_2, D_2: energy-correlation function ratio); MVA(violet lines): DNN tagger;
ParticleNet(red line); ParticleTransformer(azure line)

by Shudong W.

2022/12/8 shudong.wang@cern.ch 15

- **W/Z/QCD 3-category classification with ParticleNet**





- No great discrimination between W and Z
- Similar classification between W/Z and multi-jet (to previous page)

总体计划

- 基于ATLAS实验，完成基于PN/PT优化的W/Z tagging，使其成为成熟可供合作组使用的tagger。
- 基于ATLAS实验，通过新型机器学习方法研究HH—WW的全强子末态，计划基于PN/PT，并尝试结合PointNet。
- Potential possibility：研究PointNet开展物理研究。
  - 难点：使用PointNet需获得大量底层hit信息，目前尚无。