Boosted object identification in searches in ATLAS and CMS

Pantelis Kontaxakis[a,*], on behalf of ATLAS and CMS Collaborations

[a]*National and Kapodistrian University of Athens, Greece*

E-mail: pantelis.kontaxakis@cern.ch

Reconstruction and identification of objects with high transverse momentum (p_T), like top quarks and W, Z, Higgs bosons, from their hadronic decays can play a significant role in both searches for new physics and measurements of Standard Model processes at LHC. Cut-based algorithms and recently advanced Machine Learning techniques have been employed to identify and classify hadronic decays of these highly Lorentz-boosted bosons and top quarks. An overview of some of the latest boosted object tagging techniques used by ATLAS and CMS is presented along with their performance on analyses for both experiments.

The Eighth Annual Conference on Large Hadron Collider Physics-LHCP2020
25-30 May, 2020
online

---

*Speaker

1. Introduction

A lot of new physics models predict signatures with high-p_T ("boosted") top quarks or gauge bosons (W/Z/Higgs). They are also present in interesting SM processes. Thus, reconstruction and identification of these boosted objects from their hadronic decays provide powerful handles for new physics searches at the Large Hadron Collider (LHC) at CERN. Robust identification of boosted jets plays an important role in order to increase the sensitivity in several analyses, but it is also very challenging due to the large background from the QCD jets which is difficult to distinguish. However, it can be suppressed by exploiting the jet substructure information and track differences in the radiation pattern between signal and background. These proceedings discuss some of the latest boosted object tagging techniques employed by analyses of data recorded by the ATLAS [1] and CMS [2] experiments.

2. Tagging Strategies

A variety of boosted jet tagging algorithms have been developed. The "*cut-based*" algorithms e.g. N-subjettiness and Energy Correlation Functions (ECF) [3, 4] use theory-inspired variables in order to discriminate signal from background. They are robust and easily interpretable and also serve as baseline for more advanced tools. "*Shallow ML*" techniques utilize high-level (HL) variables to train a Boosted Decision Tree (BDT) or Neural Network (NN) where a cut on output discriminant is implemented. "*Advanced ML*" algorithms are based on Deep Neural Network (DNN) architectures. They introduce HL or even low-level information to feed the training such that the DNN can learn the jet substructure by itself.

3. Validation in data (Top/W tagging)

Measurement of the t quark and W boson efficiency in data [5, 6] is performed in the single-μ sample ($t\bar{t}(1\ell)$ dominated) using the "Tag & Probe" method. For the Scale Factor (SF) extraction, a simultaneous fit on the mass of the selected probe jets is performed that pass or fail a given tagging condition. The SF uncertainty takes into account both statistical and systematic components (JES/JER, PDF, PS). For all the top and W taggers, the agreement between simulation and data is very good. Also, the SFs are consistent with unity and the systematic uncertainties are under control.

4. Boosted Event Shape Tagger (CMS)

4.1 Algorithm description

The boosted event shape tagger (BEST) [3] is a multi-classification algorithm able to discriminate between various heavy particles (t/W/Z/H) at the same time. It computes jet kinematics and shape variables in a hypothesized reference frame corresponding to the heavy particle mass. When boosting to the "correct" rest frame, jet constituents should be isotropic showing the expected N-prong structure. It uses NN trained with many kinematic variables as well as subjet b-tagging discriminants.

4.2 Search for vector-like quarks in CMS

The vector-like quarks (VLQs) transform under the same group of SM electroweak bosons leading to several decay modes, primarily to third generation quarks. This analysis [7] concentrates on the fully hadronic final state and it utilizes the new BEST algorithm, which has been developed exactly for these topologies. This strategy is complemented by a second, cut-based analysis which is optimized for a single decay mode (T → Wb). Figure 1 (right) shows that, using the new BEST tagger, stronger exclusion limits for a wider phase space of T/B decay achieved compared when using cut-based analysis Fig. 1 (left).
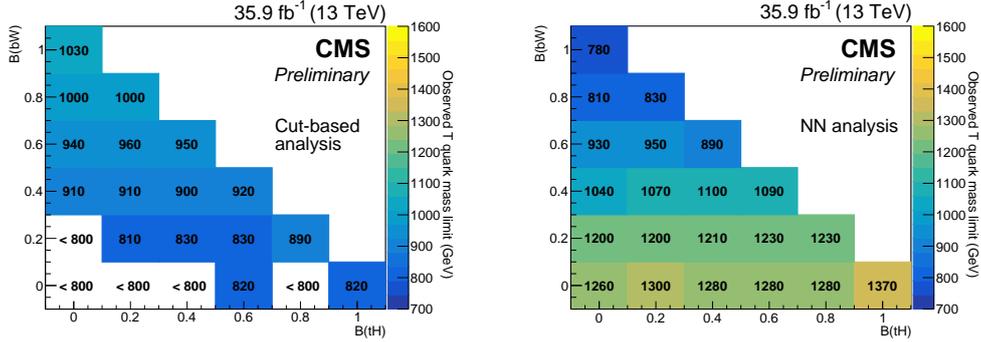

Figure 1: Observed mass exclusion limits for each combination of T quark branching fractions, in the cut-based (left) and NN analysis (right) [7].

5. Particle based jet tagging: DeepAK8 (CMS)

5.1 Algorithm description

The DeepAK8 tagger [3] is an advanced multi-class classifier for top, W, Z, Higgs and QCD jets based on standard anti-k_T R=0.8 jets. It uses low-level inputs (Particle Flow candidates, secondary vertices) and it is solely based on 1D convolution networks utilizing residual networks and a kernel size 3. DeepAK8 can exploit substructure and flavor information at the same time while it is able to identify decay modes with different flavor content.

5.2 Top squark pair production in 1ℓ in CMS

This search [8] is carried out using events with a single isolated electron or muon, multiple jets and large transverse momentum imbalance. Events are categorized according to the number of identified hadronic top quarks, leading to significant improvement of $t\bar{t}(2\ell)$ background rejection, which is the main background of the analysis. By using the DeepAK8 top tagger for boosted tops and resolved top tagger for resolved tops, the sensitivity increases up to 30%.

6. TrackCalo Cluster (ATLAS)

6.1 Algorithm description

For low-moderate p_T, ATLAS is heavily relying on calorimeter inputs ("topological clusters"). However, for boosted regime, resolution of jet substructure variable gets worse due to the limited

angular resolution. For this, the "TrackCalo Cluster" (TCC) algorithm [9] has been developed which is a combination of improving energy resolution of calorimeter and excellent angular resolution of inner detector (ID) tracks in very high p_T regions. Using TCC, the background rejection for W tagging increases by a factor of 10 when combined with additional input variable to the tagger and using new optimal working point (WP).

6.2 Search for heavy diboson resonances in ATLAS

This search [10] targets final states in which one W or Z boson decays leptonically and the other W boson or Z boson decays hadronically. One of the main improvements in this analysis is the optimization of W/Z tagger using TCC jets. The mass window and upper cut on the discrimination variable D_2 are optimized for diboson resonance search, to maximize the sensitivity in each p_T bin. With TCC improvement the sensitivity in cross section upper limits at the highest mass point is improved by a factor of 3.

7. DNN-based boosted top tagger (ATLAS)

7.1 Algorithm description

This tagger [11] is a multivariate classifier for large-R jets based on DNN architecture. It utilizes multiple jet-level discriminants which are introduced as inputs for the training and it is able to separate top jets from QCD background. Using this new top tagger, the background rejection is improved by a factor of 2 or more with respect to the legacy top tagger Fig. 2 (left).

7.2 Search for $t\bar{t}$ resonances in ATLAS

This search [12] is sensitive to any narrow resonances decaying in $t\bar{t}$ and it is optimized for the fully hadronic final states. It searches for a bump against the smoothly falling SM background of the dijet mass. It utilizes the new DNN-based boosted top tagger to reduce more the background. Since in the analysis there are 2 top-tagged large-R jets, a smaller QCD background by a factor of 4 is expected, leading to a significant improvement in the sensitivity Fig. 2 (right).
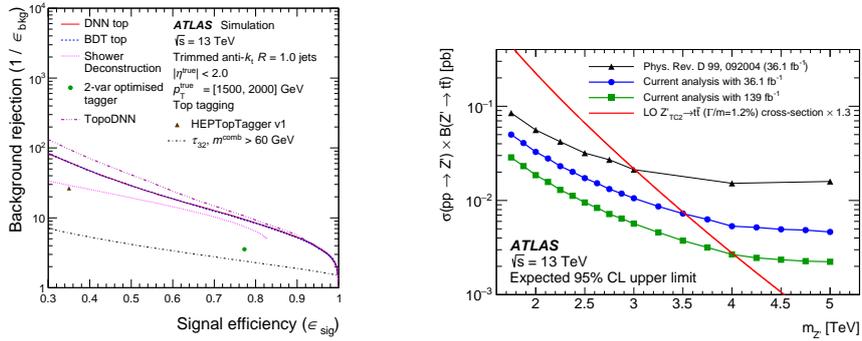

Figure 2: (Left) Performance comparison of the top-quark taggers in high-p_T^{true} used by ATLAS [11]. (Right) Comparison of expected upper limits on the cross-section times branching fraction of Z'$_{TC2 \rightarrow t\bar{t}}$ as a function of the Z'$_{TC2}$ mass, in search for $t\bar{t}$ resonances in ATLAS [12].

8. Correlation with Jet mass

Jet substructure variables and ML taggers are commonly correlated with jets, leading to changes of the background mass shapes after applying selection on these taggers ("mass sculpting"). Depending on the analysis, this may not be desirable. For this reason, both ATLAS [13] and CMS [3, 14] have explored numerous mass decorrelation techniques.

8.1 Direct search for H→cc in CMS

This is a direct search for the SM Higgs boson produced in association with vector bosons decaying to a charm quark [15]. Since the signal is extracted by fitting the mass of the Higgs candidate, decorrelation between the tagger and the mass of the jet is important for robust background estimation. Thus, for the boosted-jet topology, a single jet with distance parameter R=1.5 is used to reconstruct H→cc decay utilizing the new DeepAK15 Mass Decorrelated (MD) algorithm, the first cc-tagger at the LHC.

9. Conclusions

Boosted jet identification at the LHC is a key ingredient for the searches beyond SM. Advanced algorithms based on the combination of multiple subdetectors or machine learning techniques provide the ability to fully exploit the particle information provided by ATLAS and CMS detectors that leads to significant gain compared to the traditional approaches. This improvement translates also into significant gains for important analyses (BSM, Higgs, etc).

References

[1] ATLAS Collaboration, *The ATLAS experiment at the CERN Large Hadron Collider*, JINST 3 S08003 (2008).

[2] CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST 3 S08004 (2008).

[3] CMS Collaboration, *Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques*, JINST 15 P06005 (2020).

[4] ATLAS Collaboration, *Performance of top-quark and W-boson tagging with ATLAS in Run 2 of the LHC*, Eur. Phys. J. C 79 (2019) 375.

[5] CMS Collaboration, *W and top tagging scale factors for Run 2 data*, CMS-DP-2020-025 (2020).

[6] ATLAS Collaboration, *A W/Z-boson tagger using Track-CaloCluster jets with ATLAS*, ATL-PHYS-PUB-2020-008 (2020).

[7] CMS Collaboration, *Search for Pair Production of Vector-Like Quarks in the Fully Hadronic Channel*, CMS-PAS-B2G-18-005 (2019).

[8] CMS Collaboration, *Search for direct top squark pair production in events with one lepton, jets and missing transverse energy at 13 TeV*, CMS-PAS-SUS-19-009 (2019).

[9] ATLAS Collaboration, *Improving jet substructure performance in ATLAS using Track-CaloClusters*, ATL-PHYS-PUB-2017-015 (2017).

[10] ATLAS Collaboration, *Search for heavy diboson resonances in semileptonic final states in pp collisions at $\sqrt{s}$=13 TeV with the ATLAS detector*, arXiv:2004.14636 [hep-ex].

[11] ATLAS Collaboration, *Performance of top-quark and W-boson tagging with ATLAS in Run 2 of the LHC*, Eur. Phys. J. C 79 (2019) 375.

[12] ATLAS Collaboration, *Search for $t\bar{t}$ resonances in fully hadronic final states in pp collisions at $\sqrt{s}$=13 TeV with the ATLAS detector*, arXiv:2005.05138 [hep-ex].

[13] ATLAS Collaboration, *Performance of mass-decorrelated jet substructure observables for hadronic two-body decay tagging in ATLAS*, ATL-PHYS-PUB-2018-014 (2018).

[14] CMS Collaboration, *Identification of highly Lorentz-boosted heavy particles using graph neural networks and new mass decorrelation techniques*, CMS-DP-2020-002 (2020).

[15] CMS Collaboration, *Search for the standard model Higgs boson decaying to charm quarks*, CMS-PAS-HIG-18-031 (2019).

[16] ATLAS Collaboration, *Reconstruction and identification of boosted di-τ systems in a search for Higgs boson pairs using 13 TeV proton-proton collision data in ATLAS*, arXiv:2007.14811 [hep-ex].