Identification of boosted top quarks and $W$ bosons with Machine learning in ATLAS

Oliver Majerský (Comenius University in Bratislava)
On behalf of the ATLAS collaboration

BOOST 2017, Buffalo
To study boosted hadronic decays of top quarks and $W$ bosons in final state, need to distinguish from dominant QCD background.

Common practice – capture decay products within large-$R$ jet, use high-level observables to describe substructure – discriminate signal from QCD background.

Several observables employed in existing tagging techniques developed by ATLAS:

- More on the non-machine-learning tagging techniques in Nurifikri Norjoharuddeen’s talk
- Used not only in BSM searches, but also SM precision measurements – see Giulia Ucchielli’s talk

Run-II developments: taggers employing $p_T$-dependent cut on a pair of substructure variables (designed with constant signal efficiency across $p_T$):

- **Possible generalization:** combine many substructure variables in a machine-learning algorithm
Focus of this talk: combining high-level large-$R$ jet observables with machine learning algorithms for top/W tagging

- Boosted decision trees (BDT), deep neural networks (DNN)
- High-level substructure variables built from large-$R$ jet constituents – topo-clusters
- More in Peter Loch’s talk
- Exploit (non-linear) correlations between multiple observables
- BDT/DNN output classifier (right) with larger signal/background separation compared to individual variables
First MC-based studies of BDT/DNN for top/W tagging

ATL-PHYS-PUB-2017-004
Selection, simulation samples and signal definition

- Studies performed on Monte Carlo (MC) simulations
- Large-$R$ jets built from topo-clusters, trimmed anti-$k_t\ R = 1.0$
  \(f_{\text{cut}} = 0.05, \ R_{\text{sub}} = 0.2\)
- $p_T$-leading- and $p_T$-sub-leading jets used

Top tagging

- Pythia8 $Z' \rightarrow t\bar{t}$
- Truth jet $p_T \in [350, 2000]\text{GeV}$

W tagging

- Pythia8 $W' \rightarrow WZ$
- Truth jet $p_T \in [200, 2000]\text{GeV}$

QCD background

- Pythia8 dijets
- Truth jet $p_T \in [200, 2000]\text{GeV}$

Signal definition

- Match truth top/$W$ particle $\leftrightarrow$ truth jet $\leftrightarrow$ reco jet ($\Delta R < 0.75$)
- Truth top/$W$ decay particles merged within reco jet $\Delta R(q\bar{q}'b/q\bar{q}'; \text{reco jet}) < 0.75$
Optimization of the parameters of BDT and DNN for signal and background separation

- Reco jet $m_{\text{calo}} > 40$ GeV, $> 2$ constituents
- Training performed inclusively in $p_T$
- Equal number of signal and QCD background jets

<table>
<thead>
<tr>
<th>Topology</th>
<th># signal jets</th>
<th># QCD bckg. jets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top</td>
<td>$1 \times 10^6$</td>
<td>$1 \times 10^6$</td>
</tr>
<tr>
<td>$W$</td>
<td>$7 \times 10^5$</td>
<td>$7 \times 10^5$</td>
</tr>
</tbody>
</table>

Number of jets in training sets

- **Reweight $p_T$ distribution** of truth large-$R$ jets in signal and QCD dijets samples to a uniform distribution
  - Reduce $p_T$-based discrimination power, reduces topology dependence
  - DNN and BDT learn features mostly independent of jet $p_T$, though some variables correlated with $p_T$ to various extent
Check performance of BDT and DNN on a set of events independent from training set

- **Impose jet mass cut**
  - Examine QCD bckg. jets which have signal-like mass
  - Top tagging: $m_{\text{calo}} > 120$ GeV
  - $W$ tagging: $m_{\text{calo}} \in [60, 100]$ GeV

<table>
<thead>
<tr>
<th># top quark jets</th>
<th># $W$ boson jets</th>
<th># QCD jets</th>
</tr>
</thead>
<tbody>
<tr>
<td>$4 \times 10^5$</td>
<td>$3 \times 10^5$</td>
<td>$1 \times 10^6$</td>
</tr>
</tbody>
</table>

Number of jets in testing sets

- **Reweight $p_T$ distribution** of truth large-$R$ jets in signal MC samples to distribution of QCD bckg.
  - Remove any bias due to different signal/background $p_T$ distribution
Optimize background rejection for a fixed signal efficiency

- Sequential addition of variables to BDT, starting from previously most performant JSS variables ($\tau_{32}^{WTA}, D_2^\beta=1$)
- At each stage pick combination with highest bckg. rejection
- Stop when no statistically significant gain in bckg. rejection

- BDT for $W$ tagging learns additional features from ECF variables beyond $C_2, D_2$ (which are ratios of powers of ECFs)
DNN variables (top tagging)

Compare performance for several groups of variables

- Iterative procedure used for BDT too computationally expensive + tedious hyperparameters optimization
- Create groups of variables instead, with varying features (scale-dependent, scale-less, ...), group 6 chosen

![Training input groups](chart)

- DNN benefits from ECF variables, but can’t construct $C_2$ & $D_2$ by itself
Performance of BDT and DNN top tagging (ATL-PHYS-PUB-2017-004)

Top tagging efficiency

- Background rejection

- DNN with DNN Obs.
- DNN with BDT Obs.
- BDT with BDT Obs.
- BDT with DNN Obs.

- $\tau_{32}$ tagger with $p_T$-dependent tag

- Fixed cut on jet mass: $m^{\text{calo}} > 120$ GeV

- $\tau_{32}$
- $p_T^{\text{truth}}$: [500,1000] GeV 

- $p_T^{\text{truth}}$: [1500,2000] GeV 

- $\sqrt{s} = 13$ TeV, $|\eta|^{\text{truth}} < 2.0$

- $\epsilon_{\text{sig}} = 50$

Oliver Majerský (on behalf of the ATLAS collaboration) | BOOST 2017
Performance of BDT and DNN $W$ tagging

- ROC curves for low-\(p_T\) and high-\(p_T\) regions
- $D_2$ tagger with \(p_T\)-dependent tag
- fixed jet mass window cut: $60 < m^{\text{calo}} < 100$ GeV
Due to training using fully-contained top/W definition
Top quarks and W bosons well separated from each other $\Rightarrow$ possibility of top vs. W discrimination
Lessons learned

- Both BDT and DNN offer substantial bckg. rejection improvement
  - But need comparisons vs. better non-machine-learning taggers
- Post-tag QCD jet mass distribution signal-like shaping
  - Majority of input variables strongly correlated with mass
  - Some analyses may find this undesirable

- Both learn very similar features – DNN doesn’t learn more from high-level features than BDT
Further studies and performance in data

ATLAS-CONF-2017-064
Updates to BDT and DNN

Re-optimization of input variables choice

- Use of **combined calo+track-assisted mass** ($m^{\text{comb}}$)

- linear combination of calorimeter and tracker-based masses weighted based on resolution

- See Joe Taenzer’s talk

- **Include jet $p_T$ and mass** among variable candidates for the **training**
  - Removed cuts: $m^{\text{calo}} > 120$ GeV (top) and $m^{\text{calo}} \in [60, 100]$ GeV ($W$)
  - Only $m^{\text{comb}} > 40$ GeV cut retained

- Based on feedback from ATLAS analyses, top tagging efficiency working point $50\% \rightarrow 80\%$
• ECF1, ECF2 dropped, since $p_T$ and $m_{comb}$ included
  - Similar features, but calibrated
• ECF3 replaced with a unitless $e_3 = ECF3/ECF1^3$
• $p_T$ has little effect unsurprisingly
  - scale-dependent variables already included
  - training on uniform-rewighted $p_T$ spectra
• $m_{comb}$ among the most promising variables for both top and $W$ tagging variables (though not very obvious for $W$ tagging BDT optimisation)
• Same variable substitutions for DNN
Updated DNN variables

**Training input groups**

<table>
<thead>
<tr>
<th>Group 1</th>
<th>$C_2, D_2, \tau_{21}, \tau_{32}$, $m_{comb}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 2</td>
<td>$C_2, D_2, \tau_{21}, \tau_{32}$, $m_{comb}$, $p_T$</td>
</tr>
<tr>
<td>Group 3</td>
<td>$C_2, D_2, \tau_{21}, \tau_{32}$, $\sqrt{d_{12}}, \sqrt{d_{23}}, Q_W$</td>
</tr>
<tr>
<td>Group 4</td>
<td>$\tau_1, \tau_2, \tau_3, e_3, m_{comb}, p_T$</td>
</tr>
<tr>
<td>Group 5</td>
<td>$C_2, D_2, \tau_{21}, \tau_{32}, \sqrt{d_{12}}, \sqrt{d_{23}}, Q_W, m_{comb}$</td>
</tr>
<tr>
<td>Group 6</td>
<td>$\tau_1, \tau_2, \tau_3, e_3, m_{comb}, p_T, \sqrt{d_{12}}, \sqrt{d_{23}}, Q_W$</td>
</tr>
<tr>
<td>Group 7</td>
<td>$C_2, D_2, \tau_{21}, \tau_{32}, \sqrt{d_{12}}, \sqrt{d_{23}}, Q_W, m_{comb}, p_T$</td>
</tr>
<tr>
<td>Group 8</td>
<td>$\tau_1, \tau_2, \tau_3, \tau_{21}, \sqrt{d_{12}}, \sqrt{d_{23}}, Q_W, C_2, D_2, e_3, m_{comb}, p_T$</td>
</tr>
<tr>
<td>Group 9</td>
<td>$\tau_1, \tau_2, \tau_3, \tau_{21}, \sqrt{d_{12}}, \sqrt{d_{23}}, Q_W, C_2, D_2, e_3, m_{comb}, p_T$</td>
</tr>
</tbody>
</table>

- $m_{comb}$ important variable for rejection

**Relative background rejection ($1/e_{\text{bkg}}$)**

- **Group 9 chosen**

**ATLAS Simulation Preliminary**

- $\bar{s} = 13$ TeV, DNN Top Tagging
- $\epsilon_{\text{rel}} = 80\%$
- $m_{\text{comb}} > 40$ GeV, $|\eta^{\text{truth}}| < 2.0$

**ATLAS Simulation Preliminary**

- $\bar{s} = 13$ TeV, DNN W Tagging
- $\epsilon_{\text{rel}} = 50\%$
- $m_{\text{comb}} > 40$ GeV, $|\eta^{\text{truth}}| < 2.0$

• $m_{\text{comb}}$ important variable for rejection
BDT and DNN performance – \( W \) tagging

**ATLAS Simulation Preliminary**

- **\( D_2 \) tagger with \( p_T \)-dependent tag and fixed \( m_{\text{comb}} \) cut for reference**
- **Optimized 50 % WP \( D_2 + m_{\text{comb}} \) \( p_T \)-dependent tagger**
- $\tau_{32}$ tagger with $p_T$-dependent tag and fixed $m^{\text{comb}}$ cut for reference
- Optimized 80% WP $\tau_{32}^{WTA} + \sqrt{d_{23}}$ $p_T$-dependent tagger
- DNN & BDT also outperform advanced algorithms Shower Deconstruction and HepTopTagger
Performance in data

- Measure signal efficiency and background rejection in data
- Full ATLAS 2015+2016 dataset, $\int L \, dt = (36.1 - 36.7) \text{ fb}^{-1}$

Signal efficiency

- $t\bar{t}$ with single leptonic top decay
- Top tagging efficiency $\Delta R(b\text{-jet, large-}\text{R jet}) < 1.0$
- $W$ tagging efficiency $\Delta R(b\text{-jet, large-}\text{R jet}) > 1.0$

Background rejection

- Dijet events (rejection studies from jet $p_T > 450$ GeV)
- Photon + jets events (rejection studies from jet $p_T > 200$ GeV)

See also Nurфикри Norжохаруддеең’е talk talk for more details
Modelling of DNN classifier (top tagging) (ATLAS-CONF-2017-064)

- MC prediction normalized to data (dijets & photon+jets – first subtract small signal contribution from data)
- No systematics for dijets and photon+jets
- No systematics on input variables (no scale, resolution uncertainties)
  - Have for some observables but not all + lack of proper knowledge of correlations
  - Rather start with calibration to data via scale factors
- Modelling uncertainties large in the $t\bar{t}$ topology
Performance in data – top tagging

BDT top, DNN bottom, left signal efficiency ($t\bar{t}$), middle bckg. rejection (dijets), right bckg. rejection ($\gamma$+jet)
Performance in data – $W$ tagging

BDT top, DNN bottom, left signal efficiency ($t\bar{t}$), middle bckg. rejection (dijets), right bckg. rejection ($\gamma$+jet)

ATLAS Preliminary
$E=13$ TeV, 36.1 fb$^{-1}$
lepton+jets selection
$W$ tagger ($c_{\text{tag}} = 50\%$): BDT

ATLAS Preliminary
$E=13$ TeV, 36.7 fb$^{-1}$
Trimmer anti-$k_t$, $R=1.0$
Dijet Selection
$W$ tagger ($c_{\text{tag}} = 50\%$): BDT

ATLAS Preliminary
$E=13$ TeV, 36.1 fb$^{-1}$
Trimmer anti-$k_t$, $R=1.0$
$\gamma$ + jet selection
$W$ tagger ($c_{\text{tag}} = 50\%$): BDT

Data - Sig.
Pythia8 dijet
Herwig++ dijet
Stat. uncert.

Data - Sig.
$\gamma +$ jet
Stat. uncert.
Robustness against pile-up

- Dependence on mean ML classifier vs average # interactions per bunch crossing $\mu$
  
  ![Graph showing dependence on mean ML classifier vs average # interactions per bunch crossing $\mu$]

- No systematic uncertainties included
- Higher dependence on $\mu$ observed for $W$ tagging
- Possible discrepancy in slope modelling in $t\bar{t}$ (pending systematics)
Conclusion

- Combining high-level inputs in DNN and BDT does improve background rejection.
- DNN doesn’t seem to learn more features from high-level inputs compared to BDT.
- Signal efficiency measurement:
  - Modelling in agreement with data within uncertainties.
  - $t\bar{t}$ modelling uncertainty dominant from examined uncertainties (but do not have uncertainties on inputs).
- Background rejection measurement:
  - First ATLAS measurement of tagging background rejection in gamma+jets.
  - Observe potential mismodelling of Herwig++ in dijets (pending systematic uncertainties).
- Robustness against pile-up:
  - Possible mismodelling and trend observed.
  - Further investigation and evaluation of uncertainties pending.
Backup
DNN variables (W tagging)

- DNN learns new features from more variables than BDT
- But not a large increase in bckg. rejection
- Group 5 chosen

ATLAS Simulation Preliminary
\( \sqrt{s} = 13 \) TeV, DNN W Tagging
\( \epsilon_{\text{rel}}^\text{sig} = 50\% \)
\( p_T^{\text{truth}}: [200,2000] \) GeV
\( m_{\text{calo}}^{\text{truth}} > 40 \) GeV, \( |\eta|^{\text{truth}} < 2.0 \)

<table>
<thead>
<tr>
<th>Observable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7 (BDT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( ECF_1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( ECF_2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( ECF_3 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( C_2 )</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>( D_2 )</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>( \tau_1 )</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>( \tau_2 )</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>( \tau_{21} )</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>( R_{2}^{\text{FW}} )</td>
<td></td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>( S )</td>
<td></td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>( P )</td>
<td></td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>( D )</td>
<td></td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>( a_3 )</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>( A )</td>
<td></td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>( T_{\text{MIN}} )</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>( T_{\text{MAJ}} )</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>( Z_{\text{CUT}} )</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>( \mu_{12} )</td>
<td></td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>( \sqrt{d_{12}} )</td>
<td></td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>( KtDR )</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
<td>o</td>
<td></td>
</tr>
</tbody>
</table>

Oliver Majerský (on behalf of the ATLAS collaboration) | BOOST 2017
BDT and DNN input/output correlations – top quark jets

Oliver Majerský (on behalf of the ATLAS collaboration) | BOOST 2017

\[ \sqrt{s} = 13 \text{ TeV}, \text{ Top Jet} \]
\[ p_T^{\text{truth}} = [1000, 1500] \text{ GeV} \]
\[ m_{\text{calo}} > 40 \text{ GeV}, |\eta|^{\text{truth}} < 2.0 \]

**ATLAS Simulation Preliminary**
BDT and DNN input/output correlations – $W$ boson jets

<table>
<thead>
<tr>
<th>Linear Correlation Factor</th>
<th>ATLAS Simulation Preliminary</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sqrt{s} = 13$ TeV, $W$ Jet</td>
<td>$p_{T}^{\text{truth}} = [1000, 1500]$ GeV</td>
</tr>
<tr>
<td>$m_{\text{calo}} &gt; 40$ GeV, $</td>
<td>\eta</td>
</tr>
</tbody>
</table>

Table: BDT-DNN Correlations

<table>
<thead>
<tr>
<th>Variable</th>
<th>BDT-W</th>
<th>BDT-Top</th>
<th>DNN-W</th>
<th>DNN-Top</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{tDR}$</td>
<td>77</td>
<td>74</td>
<td>70</td>
<td>67</td>
</tr>
<tr>
<td>$\mu_{12}$</td>
<td>-8</td>
<td>-12</td>
<td>-7</td>
<td>-5</td>
</tr>
<tr>
<td>$Z_{cut}$</td>
<td>-12</td>
<td>-27</td>
<td>-18</td>
<td>-12</td>
</tr>
<tr>
<td>$T_{MIN}$</td>
<td>-11</td>
<td>-4</td>
<td>-6</td>
<td>-2</td>
</tr>
<tr>
<td>$T_{MAJ}$</td>
<td>12</td>
<td>3</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>$Q_{W}$</td>
<td>70</td>
<td>34</td>
<td>29</td>
<td>5</td>
</tr>
<tr>
<td>$\tau_{22}$</td>
<td>8</td>
<td>11</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>$\tau_{21}$</td>
<td>6</td>
<td>18</td>
<td>19</td>
<td>9</td>
</tr>
<tr>
<td>$\tau_{20}$</td>
<td>51</td>
<td>25</td>
<td>53</td>
<td>19</td>
</tr>
<tr>
<td>$\tau_{02}$</td>
<td>66</td>
<td>31</td>
<td>42</td>
<td>30</td>
</tr>
<tr>
<td>$\tau_{10}$</td>
<td>83</td>
<td>25</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>$S$</td>
<td>-5</td>
<td>1</td>
<td>68</td>
<td>27</td>
</tr>
<tr>
<td>$V_{d_{12}}$</td>
<td>67</td>
<td>74</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>$V_{d_{20}}$</td>
<td>-14</td>
<td>-4</td>
<td>61</td>
<td>43</td>
</tr>
<tr>
<td>$R_{FW}^{P}$</td>
<td>15</td>
<td>5</td>
<td>60</td>
<td>35</td>
</tr>
<tr>
<td>$D_{FW}$</td>
<td>3</td>
<td>27</td>
<td>30</td>
<td>7</td>
</tr>
<tr>
<td>$ECF_{3}$</td>
<td>74</td>
<td>50</td>
<td>26</td>
<td>11</td>
</tr>
<tr>
<td>$ECF_{2}$</td>
<td>80</td>
<td>33</td>
<td>47</td>
<td>30</td>
</tr>
<tr>
<td>$ECF_{1}$</td>
<td>93</td>
<td>34</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>$D$</td>
<td>-6</td>
<td>-4</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>$A$</td>
<td>-4</td>
<td>-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_{0}$</td>
<td>63</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### ATLAS Simulation Preliminary

- **\( \sqrt{s} = 13 \) TeV, QCD Jet**
- \( p_T^{\text{truth}} = [1000, 1500] \) GeV
- \( m_{\text{calo}} > 40 \) GeV, \(|\eta^{\text{truth}}| < 2.0\)

### Plots

- **Linear Correlation Factor**
- **BDT and DNN input/output correlations – QCD jets**

---

**Table:**

<table>
<thead>
<tr>
<th>LinCor</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNN-W</td>
<td>33</td>
</tr>
<tr>
<td>BDT-W</td>
<td>281</td>
</tr>
<tr>
<td>DNN-Top</td>
<td>23</td>
</tr>
<tr>
<td>BDT-Top</td>
<td>21</td>
</tr>
</tbody>
</table>

**KdDR**

- \( \mu_{12} \)
- \( Z_{cut}^{\text{cut}} \)
- \( T_{\text{MIN}}^{\text{MIN}} \)
- \( T_{\text{MAJ}}^{\text{MAJ}} \)
- \( Q_W^{W} \)
- \( \tau_{32}^{32} \)
- \( c_{21}^{3} \)
- \( c_{3}^{3} \)
- \( \tau_{2}^{2} \)
- \( \tau_{1}^{1} \)

**Other Variables:**

- \( S \)
- \( \bar{V}_{d,2}^{d,2} \)
- \( P \)
- \( R_{FW}^{FW} \)
- \( D_{a}^{a} \)
- \( E_{CF}^{3} \)
- \( E_{CF}^{2} \)
- \( E_{CF}^{1} \)

---

**Notes:**

- **BDT-Top**
- **DNN-Top**
- **KdDR**
- **Additional Data**

---

**References:**

- **ATL-PHYS-PUB-2017-004**

---

**Author:**

Oliver Majerský (on behalf of the ATLAS collaboration) | BOOST 2017
• BDT vs DNN, top tagging

ATLAS Preliminary $\sqrt{s} = 13$ TeV, $L = 36.1$–$36.7$ fb$^{-1}$

- $t\bar{t}$ (Data - non-top bkg)
- $t\bar{t}$ (top, MC)
- Dijets (Data)
- Dijets (MC)
- $\gamma$+jets (Data)
- $\gamma$+jets (MC)

ATLAS Preliminary $\sqrt{s} = 13$ TeV, $L = 36.1$–$36.7$ fb$^{-1}$

- $t\bar{t}$ (Data - non-top bkg)
- $t\bar{t}$ (top, MC)
- Dijets (Data)
- Dijets (MC)
- $\gamma$+jets (Data)
- $\gamma$+jets (MC)
Robustness against pile-up, full comparison (ATLAS-CONF-2017-064)

- BDT vs DNN, W tagging

**ATLAS** Preliminary \( \sqrt{s} = 13 \) TeV, \( L = 36.1-36.7 \) fb\(^{-1} \)

BDT vs DNN, W tagging

\[
\begin{align*}
\text{BDT} & \quad \text{WBDT} \\
0.25 & \quad 0.3 \\
0.35 & \quad 0.4 \\
0.4 & \quad 0.45
\end{align*}
\]

\[
\begin{align*}
\mu & \quad 10, 15, 20, 25, 30, 35, 40
\end{align*}
\]

\[
\begin{align*}
\text{p} & \quad 10, 15, 20, 25, 30, 35, 40
\end{align*}
\]

\[
\begin{align*}
\text{µ} & \quad 0.22, 0.24, 0.26, 0.28, 0.3, 0.32, 0.34
\end{align*}
\]

\[
\begin{align*}
\text{µ} & \quad 0.65, 0.6, 0.55, 0.5, 0.45, 0.4, 0.35, 0.3, 0.25, 0.2
\end{align*}
\]

\[
\begin{align*}
\text{µ} & \quad 0.66, 0.68, 0.7, 0.72, 0.74
\end{align*}
\]

Preliminary ATLAS \( = 13 \) TeV, \( L = 36.1-36.7 \) fb\(^{-1} \)

W (Data - non-tt, MC) W (tt)

Dijets (Data) Dijets (MC)

+jets (Data) +jets (MC)