Boosted Higgs Tagging at ATLAS and CMS

Michael Kagan

SLAC

HH Workshop at Fermilab
September 5, 2018
Why Boosted Higgs?

• High $p_T$ Higgs bosons from:
  – Tails of SM $hh$ distribution
  – Anomalous coupling enhancement
  – New heavy resonances

• Standard jet finding + $b$-tagging approaches fail when Higgs decay products merge due to Lorentz boost

• Boosted jet approaches: large-$R$ jets + jet substructure
Higgs Tagging: B-tagging + Substructure

• Ingredients
  – Large-R jet, Jet Mass and substructure
  – B-hadron identification

• How to identify multiple b-jets in this dense environment?
  – How to identify regions of interest / seed axes to run b-tagging algorithms?

• How to combine jet substructure with b-tagging?
### Higgs Tagging Methods

<table>
<thead>
<tr>
<th>ATLAS</th>
<th>CMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Large-R jet</td>
<td></td>
</tr>
<tr>
<td>Large-R jet soft-drop subjets</td>
<td></td>
</tr>
<tr>
<td>Fixed radius track jets</td>
<td></td>
</tr>
<tr>
<td>Variable radius track jets</td>
<td></td>
</tr>
<tr>
<td>Exclusive $k_T$ subjets</td>
<td>$\tau$-axis guided double tagging</td>
</tr>
<tr>
<td>Center of mass</td>
<td></td>
</tr>
<tr>
<td>ML driven improvements</td>
<td>ML driven improvements</td>
</tr>
</tbody>
</table>

- Note: differences in mass window and jet $p_T$ ranges, make direct comparisons not possible
Large-R Jet and Subjet Based Higgs Tagging

- Apply standard b-tagging algorithms to:
  - Entire large-R jet as region of interest
  - Subjets from clustering large-R jet constituents as regions of interest

- Pros:
  - Internal structure of jet, including impact of neutrals, used to guide RoI finding

- Cons
  - Sensitive to grooming algorithm
  - For large-R jet tagging, lose the benefit of the two b-hadron RoI hypothesis to guide b-tagging
  - Harder to calibrate subjet independent of large-R jet
Large-R Jet and Subjet Based H-Tagging

CMS Simulation

13 TeV, 2016

Tagging efficiency (H→b̅b)

AK8 jet
50 < m < 200 GeV
300 < p_T < 500 GeV

AK8 jet
50 < m < 200 GeV
1200 < p_T < 1800 GeV

Note:
Will come back to blue line later
Data/MC for Subjet Tagging

- b-jet calibration in muon enriched subjets of large-R jets in multijet events
- Light-jet calibration using “negative tagged” subjets
Track Jet based Boosted Higgs Tagging

- Jet clustering on tracks using small-R to resolve fine features
- Match track jets to larger-R jets
- Pros
  - Excellent track resolution allows for small-R clustering
  - Pileup insensitive
  - Independent of large-R jet
    - Insensitive to grooming algorithm
    - $b$-tagging calibration independent of large-R jet
- Cons:
  - No access to neutral particle information in finding direction
H-Tagging with Fixed Radius Track Jets

**ATLAS Simulation Preliminary**

- $p_T > 250$ GeV, $76$ GeV $< m_{\text{calo}}^{\text{jet}} < 146$ GeV
- Double b-tag
- Asymm. b-tag (70% wp)
- Single b-tag
- Leading subjet b-tag

**Higgs-jet efficiency**

**Multi-jet rejection**

**ATLAS Simulation Preliminary**

- MV2c10 b-tagging at 77% WP
- 1 b-tag, Loose $m_{\text{calo}}^{\text{jet}}$ window
- 2 b-tags, No $m_{\text{calo}}^{\text{jet}}$ selection
- 2 b-tags, Loose $m_{\text{calo}}^{\text{jet}}$ window
- 2 b-tags, Tight $m_{\text{calo}}^{\text{jet}}$ window, $D_2$ sel.

**Hadronic top rejection**

**Simulation Preliminary**

- MV2c10 b-tagging at 77% WP
- 1 b-tag, Loose $m_{\text{calo}}^{\text{jet}}$ window
- 2 b-tags, No $m_{\text{calo}}^{\text{jet}}$ selection
- 2 b-tags, Loose $m_{\text{calo}}^{\text{jet}}$ window
- 2 b-tags, Tight $m_{\text{calo}}^{\text{jet}}$ window, $D_2$ sel.
Performance in Data

- **Left:** Single jet $b$-tagging calibration in dilepton $t\bar{t}$ events

- **Right:** Muon matched large-$R$ jets in multijet events for double $b$-tagging validation in $\text{gluon} \to \text{bb}$ enriched sample
  - Useful to check for topological effects or “cross talk” between among the two $b$-tagged jets
VR Track Jet Based b-tagging

- At high Higgs jet $p_T$, fixed radius track-jets will merge
  - One option: use single and double track-jet selections

- Shrink track jet $R$ with $p_T$
  - $R \rightarrow R_{eff}(p_T) = \frac{\rho}{p_T}$

- Pro:
  - Resolve and tag multiple b-jet region of interests up to higher Higgs jet $p_T$

- Cons:
  - Need to be careful about overlapping track jets
VR Track Jet Based b-tagging

- At high Higgs jet $p_T$, fixed radius track-jets will merge
  - Necessitates using single and double track-jet selections

- Shrink track jet R with $p_T$

- Pro:
  - Resolve and tag multiple b-jet region of interests up to higher Higgs jet $p_T$

- Cons:
  - Need to be careful about overlapping track jets
Exclusive subjet guided b-tagging

- Identify two exclusive regions of interest using exclusive-\(k_T\) declustering of large-R jet
  - ATLAS: standard b-tagging on ExKt subjets
  - CMS: Tagger using RoIs based on \(\tau\) axes (found with exclusive-\(k_T\))

- Alternatively boost into center of mass frame and do exclusive clustering
  - E.g. with EECambridge algorithm, \(\min y_{ij} = 2(1-\cos \theta_{ij})\)
Note: Trainings are based on ttbar samples, and no specific training has yet been performed using jets in boosted topologies.
• Associate tracks and vertices to $\tau$-axes

• Combine in BDT: b-tagging information from both subjets and information relating the two subjets
CMS DoubleB Scale factors

- Double-b calibration using muon matched large-R jets in multi-jet events
- Top jet mistag rate calibration in boosted top events
- Z-bosons identified with tagger in Higgs+ISR analysis

arXiv:1712.07158
ML improvements to single b-jet tagging

From L. Gouskos, BOOST 2018
DeepDoubleB

Inputs (differences wrt “slim” jet case):
- Increased num of particles/SV (larger R)
- Reduced num of features (less mass sculpting)
- Drop Neutral PF candidates

From L. Gouskos, BOOST2018
Conclusion

- Wide array of Higgs tagging approaches utilizing b-tagging and substructure in different ways

- Exciting new algorithms being deployed, while active development continues!
  - Including dedicated double b-tagging and utilizing modern machine learning methods
  - Need to understand if such approaches expose us to new and potentially larger modeling uncertainties.

- Methods to calibration / validate these methods are becoming fairly well established
  - Do we need to calibrate samples other than bb? What about b, cc, c, l?
  - Are there topology specific effects?