A $W^\pm$ polarization analyzer from Deep Neural Networks

Taegyun Kim
Research Advisor: Dr. Adam Martin
Department of Physics, University of Notre Dame

arXiv:2102.05124
• Entering HL-LHC: bring out small number of event signals
• Precision testing to find potential BSM signatures
• This research is about building a tool and show possibility
Theoretical Motivation

Massive vector boson final states

\[ pp \rightarrow W^\pm W^{\mp} \]
\[ pp \rightarrow W^\pm Z \]
\[ pp \rightarrow ZZ \]

• Indirect approach of checking SM: polarization searches
  • Longitudinal vs. Transverse
• SM can predict polarization fraction
• Longitudinal polarization is sensitive to EWSB
• Some SMEFT operators can affect longitudinal fraction of a process
W polarization

Decay of W

- There is a limitation in leptonically decaying W
- Since W only interacts to the left handed particles, each polarization has distinct angular distribution
- Due to the deviation, it is possible to measure polarization fraction for diboson final states
- Large overlap in parton level distribution may suppress even by event tagging
Boosted $W$ Jet

Decay of $W$

- Quark becomes QCD jet
- Due to the boost, collimation of the jet deduces the angular distribution signature
- Possible subjet signature
- After boost $\theta^* \rightarrow$ opening angle (sensitive to $p_T$)
- At extreme high $p_T^W$, subjet signature can disappear

Fat jet $\Delta R \approx \frac{2m_W}{p_T^W}$
Machine Learning Motivation

Machine Learning in HEP

- The current most frequently used machine learning algorithm: Boosted Decision Tree (BDT) and Neural Network (NN)

- Major usage
  - **Classification** : PID, event identification
  - **Regression** : predict particle energy

- Recent researches on: quark vs. gluon, QCD vs. top, W vs. QCD

ML(NN) technique to distinguish hadronic W’s polarization to test on $W^\pm Z$ final state
Jet as an Image from Collider

Adjust for HEP using jet image

- In collider, images are created from outgoing particles
- Particles are plotted on pixelized $\eta - \phi$ plane and their color is determined from $p_T$

Convolutional Neural Network (CNN)

Image recognition

- Ordinary CNN structure: Convolution - Flatten - Dense

- The network is trained with simulated events (MadGraph + Pythia + Delphes) of boosted longitudinal and transverse $W$’s respectively for tagging purposes

- Depending on $p_T^W$, images are separated into 2 bins: [200,300] and [400,500] since for fat jet, $\Delta R \approx \frac{2m_W}{p_T^W}$
Testing on SM

Longitudinal fraction ($f_L$)

- Checking distribution can tell us how good the separation between two polarization
- Inhibits potential event by event tagging because of large overlap
- In order to apply for testing, we measure $f_L$ of randomly selected events
- Test on $WZ$ final state

<table>
<thead>
<tr>
<th>$p_T$ range</th>
<th>$\sigma(pp \rightarrow W^{\pm}(jj)Z(\ell\ell))$ (fb)</th>
<th>truth $\sigma_L/\sigma_{tot}$</th>
<th>predicted $f_L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$200$ GeV $\leq p_T \leq 300$ GeV</td>
<td>6.67</td>
<td>0.265</td>
<td>$0.259 \pm 0.013$</td>
</tr>
<tr>
<td>$400$ GeV $\leq p_T \leq 500$ GeV</td>
<td>0.35</td>
<td>0.304</td>
<td>$0.300 \pm 0.033$</td>
</tr>
</tbody>
</table>
SMEFT in Diboson Final States

SMEFT intro

- \( \mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_{D>4}^{\text{inf}} \frac{1}{\Lambda_{D-4}^{D-4}} c_j^{(D)} \mathcal{O}_j^{(D)} \)

- SMEFT extends the SM Lagrangian by gauge invariant higher dim \((D>4)\) operators

- We will investigate boosted \(W\) cases

Relevant operators (SILH) for diboson final states

\[
\mathcal{O}_W = \frac{ig}{2} \left(H^\dagger \sigma^a \overleftrightarrow{D^\mu H}\right) D^\nu W^a_{\mu\nu} \\
\mathcal{O}_B = \left(H^\dagger \sigma^a \overleftrightarrow{D^\mu H}\right) \partial^\nu B_{\mu\nu} \\
\mathcal{O}_{2W} = -\frac{1}{2} D^\mu W^a_{\mu\nu} D^\rho W^a_{\rho\nu} \\
\mathcal{O}_3W = \frac{1}{3!} g \epsilon_{abc} W^a_{\mu\nu} W^b_{\nu\rho} W^c_{\rho\mu} \\
\mathcal{O}_{HW} = ig \left(D^\mu H\right)^\dagger \sigma^a \left(D^\nu H\right) W^a_{\mu\nu} \\
\mathcal{O}_{HW} = ig' \left(D^\mu H\right)^\dagger \left(D^\nu H\right) B_{\mu\nu}
\]

# Possible Scenarios with SMEFT

<table>
<thead>
<tr>
<th>SM</th>
<th>$p_T$ range</th>
<th>$\sigma(pp \to W^{\pm}(jj)Z(\ell\ell))$ (fb)</th>
<th>truth $\sigma_L/\sigma_{tot}$</th>
<th>predicted $f_L$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>200 GeV $\leq p_T \leq$ 300 GeV</td>
<td>6.67</td>
<td>0.265</td>
<td>$0.259 \pm 0.013$</td>
</tr>
<tr>
<td></td>
<td>400 GeV $\leq p_T \leq$ 500 GeV</td>
<td>0.35</td>
<td>0.304</td>
<td>$0.300 \pm 0.033$</td>
</tr>
</tbody>
</table>

1. Shift longitudinal fraction with cross section shift

<table>
<thead>
<tr>
<th>$O_W$</th>
<th>$p_T$ range</th>
<th>$\sigma(pp \to W^{\pm}Z)$ (fb)</th>
<th>truth $\sigma_L/\sigma_{tot}$</th>
<th>predicted $f_L$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>200 GeV $\leq p_T \leq$ 300 GeV</td>
<td>6.93</td>
<td>0.311</td>
<td>$0.297 \pm 0.010$</td>
</tr>
<tr>
<td></td>
<td>400 GeV $\leq p_T \leq$ 500 GeV</td>
<td>0.42</td>
<td>0.439</td>
<td>$0.391 \pm 0.033$</td>
</tr>
<tr>
<td>$O_{3W}$</td>
<td>200 GeV $\leq p_T \leq$ 300 GeV</td>
<td>6.58</td>
<td>0.258</td>
<td>$0.254 \pm 0.011$</td>
</tr>
<tr>
<td></td>
<td>400 GeV $\leq p_T \leq$ 500 GeV</td>
<td>0.50</td>
<td>0.198</td>
<td>$0.181 \pm 0.043$</td>
</tr>
</tbody>
</table>

2. Shift longitudinal fraction without cross section shift

<table>
<thead>
<tr>
<th>SM + $O_W + O_{3W}$</th>
<th>$p_T$ range</th>
<th>$\sigma(pp \to W^{\pm}Z)$ (fb)</th>
<th>truth $\sigma_L/\sigma_{tot}$</th>
<th>predicted $f_L$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>200 GeV $\leq p_T \leq$ 300 GeV</td>
<td>6.68</td>
<td>0.202</td>
<td>$0.207 \pm 0.011$</td>
</tr>
<tr>
<td></td>
<td>400 GeV $\leq p_T \leq$ 400 GeV</td>
<td>0.34</td>
<td>0.285</td>
<td>$0.282 \pm 0.044$</td>
</tr>
</tbody>
</table>
Conclusion/Discussion

- Simple CNN can be used to tag $W^\pm$ polarization though event by event tagging is suppressed.

- Ensemble analysis using network’s output average values can help to predict $f_L$.

- Network prediction can catch small $f_L$ deviations originated from dim 6 operators.

- If cross section changes, polarization measurement can clear out degeneracies between EFT operators.

- Possible applicability on $Z$ jets.

- Potential limits:
  - $W^\pm$ vs. $Z$ vs. QCD is not perfectly separable.
  - Cuts that can cause polarization interference.
Thank you
References


(5) “scikit-hep/pyjet: 1.6.0 (version 1.6.0),”
Backup slides
• Checking distribution can tell us how good the separation between Logi and trans is.

• Inhibits potential event by event tagging since accuracy is ~ 60%

• Ensemble distribution checking to find longitudinal fraction \((f_L)\)
Simpler Method

Network output average method

- Template fitting method depends on finding “sweet spot” for $f_L$
  - number of bins
  - find minimum $\chi^2(f_L)$
- Simplify by treating output distribution as probability distribution

\[ \int x \, dx \left( D_u(x) = f_L D_L(x) + (1 - f_L) D_T(x) \right) \]

\[ \langle x_u \rangle = f_L \langle x_L \rangle + (1 - f_L) \langle x_T \rangle \]

\[ f_L = \frac{\langle x_u \rangle - \langle x_T \rangle}{\langle x_L \rangle - \langle x_T \rangle} \]

Confirmed that both yield the same result
Jet Images

Network friendly form

Bring out subjet signature

1. Identify jet with clustering algorithm
2. Check if clustered jet lies under $p_T$ bin range
3. Select jets with correct angular position
4. Recluster to identify subjets

https://github.com/scikit-hep/pyjet
Reduce image discrepancies by putting into consistent orientation

1. Translate to centralize the highest $p_T$ subjet
2. Rotate so that the second highest $p_T$ subjet below the highest
3. Reflect
4. Pixelize
5. Normalize
• Checking distribution can tell us how good the separation between two polarization

• Inhibits potential event by event tagging because of large overlap
  • Putting decision threshold would contain large contamination

• Ensemble distribution checking to find longitudinal fraction \( f_L \)
Kinematic Cut Effect

\[
\frac{1}{\sigma} \frac{d\sigma}{d\cos\theta^*} = \frac{3}{8} (1 - \cos\theta^*)^2 f_L + \frac{3}{8} (1 + \cos\theta^*)^2 f_R + \frac{3}{4} \sin^2\theta^* f_0,
\]

\[
\frac{1}{\sigma} \frac{d\sigma}{d\cos\theta^* d\phi^*} = \frac{3}{16\pi} [(1 + \cos^2\theta^*) + A_0 \frac{1}{2} (1 - 3\cos^2\theta^*) + A_1 \sin 2\theta^* \cos \phi^*
+ A_2 \frac{1}{2} \sin^2\theta^* \cos 2\phi^* + A_3 \sin \theta^* \cos \phi^* + A_4 \cos \theta^*],
\]

- Integrating over $\phi^*$ will give the same result but kinematic cut can change
Kinematic Cut Effect

\[
\frac{1}{\sigma} \frac{d\sigma}{d \cos \theta^*}
\]

- \( p_T, p_T^l \)
- \( p_T, p_T^l, p_T^m \)
- \( p_T^l, p_T^l, p_T^m, \eta \)

\[
\frac{1}{\sigma} \frac{d\sigma}{d \cos \theta^*}
\]

- No cut
- \( p_T > 20 \text{ GeV} \)
Kinematic Cut Effect

Figure 9: Normalised azimuthal angle distributions for a set of different selection cuts imposed on final-state leptons and jets for $W^+ + 1$ jet production at 7 TeV.
W vs. Z

Jet charge

Additional observable: $Q_\kappa = \frac{1}{(p_T, j)^\kappa} \sum_{i \in J} q_i \times (p_T^i)^\kappa$

- Depending on $\kappa$, separation may change.
- Need to find optimal value of $\kappa$
- Input is $p_T$ and $Q_K$ depth=2 image
Preparing Samples

Training / Validation

Longitudinal  \[ pp \to \phi \to W^\pm W^{\mp} \]

Transverse  \[ pp \to W^\pm j \]

- **MadGraph + Pythia + Delphes**
- We separate into \( p_T \) bins of \( W \) jet: [200,300] and [400,500]
- To make sure the quality of sample, we plotted \( W \) decay in parton level

Why asymmetric?

[arXiv: 1204.6427v1]
**Analysis**

**Template fit method**

- Consider each pure polarization histogram as “template” that can be applied to the unknown sample

- Fit quality is determined by $\chi^2$ distance test

$$D_u(x_i) = f_LD_L(x_i) + (1 - f_L)D_T(x_i)$$

$$\chi^2(f_L) = \sum_{i=1}^{B} \frac{(O_i - N_s(f_LL_i + (1 - f_L)T_i))^2}{N_s(f_LL_i + (1 - f_L)T_i)}$$

![Graph showing reduced $\chi^2$ vs $f_L$.]
Test on Unknown Samples

SM testing using average method

\[ pp \rightarrow W^{\pm}Z \]

<table>
<thead>
<tr>
<th>( p_T ) range</th>
<th>truth ( \frac{\sigma_L}{\sigma_{tot}} )</th>
<th>predicted ( f_L )</th>
</tr>
</thead>
<tbody>
<tr>
<td>[200,300]</td>
<td>0.265</td>
<td>0.259 ± ?</td>
</tr>
<tr>
<td>[400,500]</td>
<td>0.304</td>
<td>0.300 ± ?</td>
</tr>
</tbody>
</table>

- Output average method can predict well for both \( p_T \) bins
- Estimate error on our prediction can tell us the precision
- Truth value is calculated from MadGraph
Uncertainty

Small experiments

- From large test set, we randomly select subset (\(N\) number of events) to obtain \(f_L\)

- \(N\) is determined from expected number of events at particular luminosity

- At current LHC luminosity \(\sim 2000\) events at low \(p_T\) and \(200\) events at high \(p_T\)

- At High Lumi LHC \(\sim 20k\) events at low \(p_T\) and \(2k\) events at high \(p_T\)

- By iterating the process, we can obtain average value with standard deviation

<table>
<thead>
<tr>
<th></th>
<th>300 fb(^{-1})</th>
<th>3000 fb(^{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>[200,300]</td>
<td>0.044</td>
<td>0.010</td>
</tr>
<tr>
<td>[400,500]</td>
<td>0.130</td>
<td>0.033</td>
</tr>
</tbody>
</table>
1. Previous attempts from ATLAS collaboration to measure polarization with leptonic final states

   • Leptonic final state: small branching ratio

   • Complication in $\nu$ reconstruction

2. If we can use hadronic $W$, we gain more statistics but need to deal with hadronic jets