Distinguishing W' Signals at Hadron Colliders Using Neural Networks

Ting-Kuo Chen

Collaborators: Spencer Chang, Cheng-Wei Chiang

Department of Physics, NTU

October 5, 2020

Charged Resonance Searches

- As the mass limit of new physics (NP) charged bosons is pushed above TeV level $\rightarrow$ focus on high-energy hadron colliders.
- In this case, $\ell\nu$ channel is favorable.
  - Clean from QCD background.
  - Single final-state object $\rightarrow$ simple kinematic signature.
- If we consider exotic Higgs sectors, charged scalars are then also included $\rightarrow$ we are interested in the identification of the spin and coupling properties of possible NP bosons.

Challenges

- Challenges:
  - Missing longitudinal momentum.
  - Unrecognizable incidental partons $\rightarrow$ for *pp* colliders, this is even severed by the symmetry of the proton beams.
- Some ideas for workarounds:
  - Empirical fitting.
  - Derivative observables, e.g. p_T, η.
- In our study, we focus on 14 TeV LHC collisions, and explore the potential of neural network (NN) upon this problem.

Formulation and Method

- Instead of individual-event studies, we consider a 2D "global distribution" spanned by p_T^ℓ and η^ℓ $\rightarrow$ we can rephrase the problem as an image recognition problem[1].
- If we further include an extra QCD order to form one additional final-state jet, the system would possess 5 degrees of freedom (in the massless limit).
- Convolutional neural network (CNN) turns out to be a suitable candidate for this problem.

[1] This was proposed and used by Khosa, *et al.* (2019) in their study of WIMPs.

Formulation and Method

- We consider three simple effective models:
  - Vector/Axial (VA): W' with vector/axial-like couplings.
  - Chiral (CH): W' with LH/RH couplings.
  - Scalar (SC): $H^{\pm}(H)$ with Yukawa-like couplings.
- The following conditions are assumed, although it is straightforward to extend the study beyond them:
  - The pole mass is 1 TeV for all three models.
  - The couplings are universal to both the quark/lepton sectors, and to all generations.
  - Only the decay to $e\nu$ is studied.
  - The interference between the NP and the SM processes is neglected.

Formulation and Method

- Assuming an integrated luminosity of $\mathcal{L} = 60\ \text{fb}^{-1}$ (about half the expected annual luminosity of LHC Run-III), we define $B = \sigma_{\rm SM} \times \mathcal{L}$ in a specific phase space and form scenarios of different S/B or $S/\sqrt{B}$, S being the number of NP events $\rightarrow$ let CNN recognize histograms made from these events.
- For comparison, we propose a Bayesian hypothesis (BH) tests with the posteriors defined as the following:
  - $e\nu$ (LO): $P(D|H_k) = \prod_{m,n} p(h^D_{mn}, H^k_{mn})$
  - $e\nu + j$ (NLO): $P(D|H_k) = \prod_{m,n,ch} p(h^{D,ch}_{mn}, H^{k,ch}_{mn}),\ ch = 1, 2, 3$

  where we have assumed bin-wise Poisson likelihood models.

Theoretical Analysis

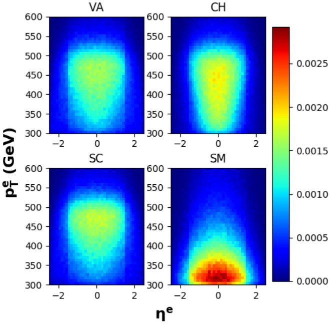
- First consider parton-level LO spin-0 and -1 processes. The differential p_T^e and η^e distributions are given by:

$$\frac{d\hat{\sigma}}{dp_T^e} = \begin{cases} y_H^4 \cdot J(p_T, p^2, m_H^2, \Gamma_H^2), \text{ for } H \\ (c_V^2 + c_A^2)\left(1 - \frac{2p_T^2}{p^2}\right) \cdot J(p_T, p^2, m_{W'}^2, \Gamma_{W'}^2), \text{ for } W' \end{cases}$$

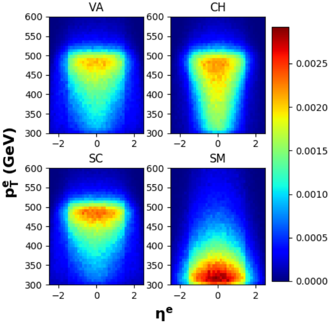
$$\frac{d\hat{\sigma}}{d\eta^e} = \begin{cases} y_H^4 \cdot F(\eta, p^2, m_H^2, \Gamma_H^2, E_1, E_2), \text{ for } H \\ (c_V^2 + c_A^2) \cdot G(\eta, p^2, m_H^2, \Gamma_H^2, E_1, E_2) \\ \quad + c_V^2 c_A^2 \cdot H(\eta, p^2, m_H^2, \Gamma_H^2, E_1, E_2), \text{ for } W' \end{cases}$$

$\rightarrow$ η^e allows us to probe different couplings of W'.

Normalized 2D LO Distribution



(a)



(b)

Figure 1: LO p_T^e vs. η^e distributions for $\Gamma \approx$ (a) 100 and (b) 10 GeV. The resolutions for these and the upcoming plots are all 40×40.

Challenge for NLO Processes

- There are 5 degrees of freedom in a 3-body massless system $\rightarrow$ which observables should be used?
- We propose 3 schemes:
  - Physics Relation (Scheme 1): p_T^e vs. η^e, p_T^j vs. η^j, $\Delta\phi_{e\nu}$ vs. $\Delta\phi_{j\nu}$.
  - Principal Component Analysis (Scheme 2): p_T^e vs. $\not{E}_T$, η^e vs. η^j, $\Delta\phi_{e\nu}$ vs. $\Delta\phi_{j\nu}$
  - Common Axis (Scheme 3): p_T^e vs. $\not{E}_T$, p_T^e vs. η^e, p_T^e vs. $\Delta\phi_{ej}$.

$\rightarrow$ It turns out that the results are quite consistent.

- We only study $\Gamma \approx 10$ GeV as the training outcomes are similar for different widths.

Training Samples

- We use $S + B$ number of events in every single sample histogram for each significance scenario.

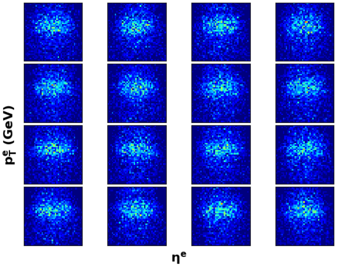


Figure 2: Examples of LO VA sample histograms for $S/B = 1.0$ with $\Gamma \approx 10$ GeV.

CNN Structure

- For LO processes, we only have 1 color channel; for NLO processes, we have 3 color channels.
- The aim is to find the simplest model that is able to produce the same level of results as BH test does.

	LO	NLO
Input	40 × 40 images	
	p_T^e vs. η^e	RGB Color Schemes Scheme 1: p_T^e vs. η^e,p_T^j vs. η^j,$\Delta\phi_{eE}$ vs. $\Delta\phi_{jE}$ Scheme 2: p_T^e vs. $\not{E}_T$,η^e vs. η^j,$\Delta\phi_{eE}$ vs. $\Delta\phi_{jE}$ Scheme 3: p_T^e vs. $\not{E}_T$, p_T^e vs. $\not{E}_T$, p_T^e vs. $\Delta\phi_{ej}$
Layers	batch normalization layer convolutional 2D layer: 3-32[a] max pooling 2D layer: 2-2[b] convolutional 2D layer: 3-32 max pooling 2D layer: 2-2 flatten layer dense layer: 128[c]	
Layer settings	hidden layer activation = `relu` output layer activation = `softmax`	
Compilation	loss = `categorical_crossentropy` optimizer = `adam` metric = `accuracy`	

[a] This means that the filter kernel dimension is 3 × 3, and that there are 32 nodes in the convolutional layer.

[b] This means that the max pooling kernel dimension is 2 × 2, and that each stride is 2 pixels.

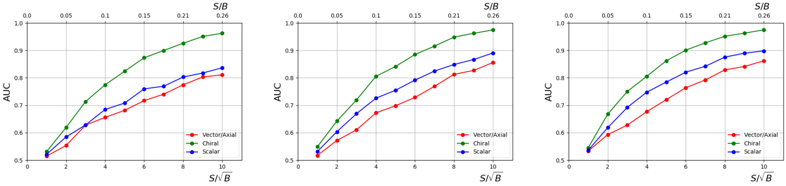
[c] This means that there are 128 nodes in the dense layer.

Figure 3: CNN structure.

Training Specifications

- For each effective model (including the SM), we have roughly $700K$ events.
- For each S/B or $S/\sqrt{B}$ scenario, we use the events to generate roughly $15K$ sample histograms.
- The sample histograms are split into training, validation, and testing sets with the ratio $0.64 : 0.16 : 0.20$.

LO Results



(a) (b) (c)

Figure 4: LO low-significance training results for $\Gamma \approx$ (a) 100, (b) 10, and (c) 1 GeV.

LO Results

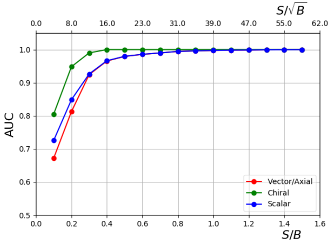
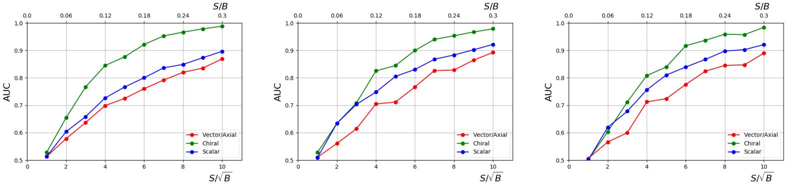


Figure 5: LO high-significance training results for $\Gamma \approx 10$ GeV.

- The AUCs still steadily improve, and reach nearly perfect identification rates for $S/B \gtrsim 0.8$.
- CH class is always the easiest to be identified $\rightarrow$ bottleneck: VA vs. SC.

NLO Results



(a) (b) (c)

Figure 6: NLO low-significance training results for $\Gamma \approx 10$ GeV, using scheme (a) 1, (b) 2, and (c) 3.

NLO Results

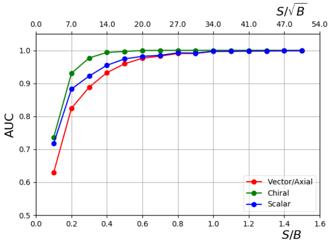
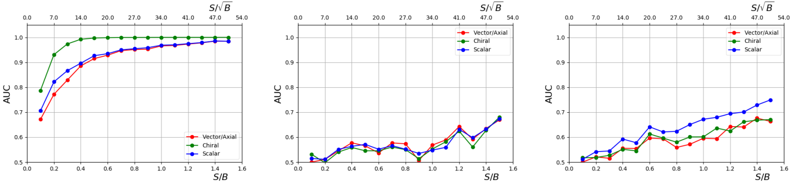


Figure 7: NLO high-significance training results for $\Gamma \approx 10$ GeV, using scheme 3.

- The AUCs reach nearly perfect identification rates for $S/B \gtrsim 1.0$.

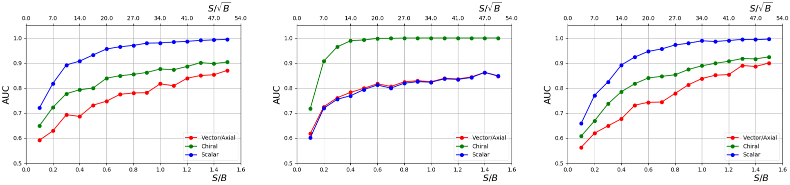
NLO Results



(a) (b) (c)

Figure 8: NLO high-significance training results for $\Gamma \approx 10$ GeV, using (a) p_T^e vs.η^e, (b) p_T^j vs.η^j, and (c) $\Delta\phi_{e\nu}$ vs.$\Delta\phi_{j\nu}$.

NLO Results



(a) (b) (c)

Figure 9: NLO high-significance training results for $\Gamma \approx 10$ GeV, using (a) p_T^e vs.$\not{E}_T$, (b) η^e vs.η^j, and (c) p_T^e vs.$\Delta\phi_{ej}$.

NLO Results

- Different variable pairs have different importances, but using all of them does lead to better results.
- p_T^e vs. η^e plays the most important role, as the angular and coupling information should mostly be preserved in e.
- p_T^e vs. η^e and η^e vs. η^j are best at identifying the CH class.
- p_T^e vs. $\Delta\phi_{ej}$ and p_T^e vs. $\not{E}_T$ are best at identifying the SC class.
- VA is always the most difficult to be identified.

Comparison with BH Tests

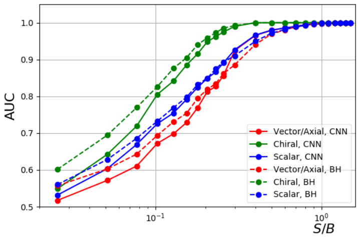


Figure 10: LO results using CNN (solid) and BH test (dashed).

- At $S/B \leq 0.3$, the BH test outperforms the CNN.
- Above that threshold, the CNN then becomes competitive with the BH test.

Comparison with BH Tests

- There are a few issues about a typical BH test:
  - It is highly sensitive to small expected distributions, and cannot tolerate 0 expectation values. Preprocessing such as symmetrization, extrapolation, and interpolation might solve the problem, but is not guaranteed.
  - Such problems become more complicated when the resonance mass gets higher, or when the analysis dimension increases.
  - Other than the efforts needed to optimize the network, these concerns are tolerable for a typical NN
- Mathematically, the best results can be obtained by performing a maximum likelihood test in the multi-dimensional space $\rightarrow$ this is technically challenging when the dimension becomes greater than 2.

Comparison with BH Tests

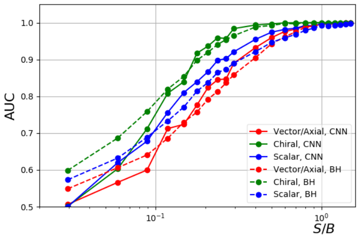


Figure 11: NLO results using CNN (solid) and BH test (dashed).

- At $S/B \leq 0.2$, the BH test and CNN are competitive with each other.
- Above that threshold, the CNN then outperforms the BH test.

Summary

- It is possible to study the spin and coupling properties of hypothesis charged bosons through its leptonic decay channel which involves missing energy at hadron colliders.
- These properties can be studied using $2D$ kinematic distributions.
- Neural networks can classify the effective models with roughly the same efficiencies as the Bayesian hypothesis tests do, and even better in some versions of higher dimensional studies.