Note

Feng-Yang Hsieh

1 Higgs Production

We want to apply deep learning methods to distinguish vector boson fusion (VBF) from gluon-gluon fusion (GGF) and Higgs production at the LHC.

We want to apply the CWoLa methods, then can use the real data without knowing the true label.

2 Sample Preparation

2.1 Monte Carlo samples

We consider Standard Model (SM) di-photon Higgs events produced via GGF and VBF channels at a center-of-mass energy of $\sqrt{s} = 14$ TeV. The Higgs boson events are generated using `MadGraph 3.3.1` [1] for both GGF and VBF production. The Higgs decays into the di-photon final state, and the parton showering and hadronization are simulated using `Pythia 8.306` [2]. The detector simulation is conducted by `Delphes 3.4.2` [3]. Jet reconstruction is performed using `FastJet 3.3.2` [4] with the anti-k_t algorithm [5] and a jet radius of $R = 0.4$. These jets are required to have transverse momentum $p_{\mathrm{T}} > 25$ GeV.

The following `MadGraph` scripts generate Monte Carlo samples for each production channel.

GGF Higgs Sample Generation

```
generate p p > h QCD<=99 [QCD]
output GGF_Higgs
launch GGF_Higgs

shower=Pythia8
detector=Delphes
```

```
analysis=OFF
madspin=OFF
done

set run_card nevents 100000
set run_card ebeam1 7000.0
set run_card ebeam2 7000.0

set run_card use_syst False

set pythia8_card 25:onMode = off
set pythia8_card 25:onIfMatch = 22 22
done
```

VBF Higgs Sample Generation

```
define v = w+ w- z
generate p p > h j j $$v
output VBF_Higgs
launch VBF_Higgs

shower=Pythia8
detector=Delphes
analysis=OFF
madspin=OFF
done

set run_card nevents 100000
set run_card ebeam1 7000.0
set run_card ebeam2 7000.0

set run_card use_syst False

set pythia8_card 25:onMode = off
set pythia8_card 25:onIfMatch = 22 22
done
```

2.2 Event selection

The selection cuts after the `Delphes` simulation:

- n_γ cut: The number of photons should be at least 2.
- n_j cut: The number of jets should be at least 2.
- $m_{\gamma\gamma}$ cut: The invariant mass of two leading photons $m_{\gamma\gamma}$ are required 120 GeV $\leq m_{\gamma\gamma} \leq$ 130 GeV.

Table 1 summarizes the cutflow number at different selection cuts.

Table 1: Number of passing events and passing rates for GGF and VBF Higgs production at different selection cuts.

Cut	GGF	pass rate	VBF	pass rate
Total	100000	1	100000	1
n_γ cut	48286	0.48	53087	0.53
n_j cut	9302	0.09	42860	0.43
$m_{\gamma\gamma}$ cut	8864	0.09	40694	0.41

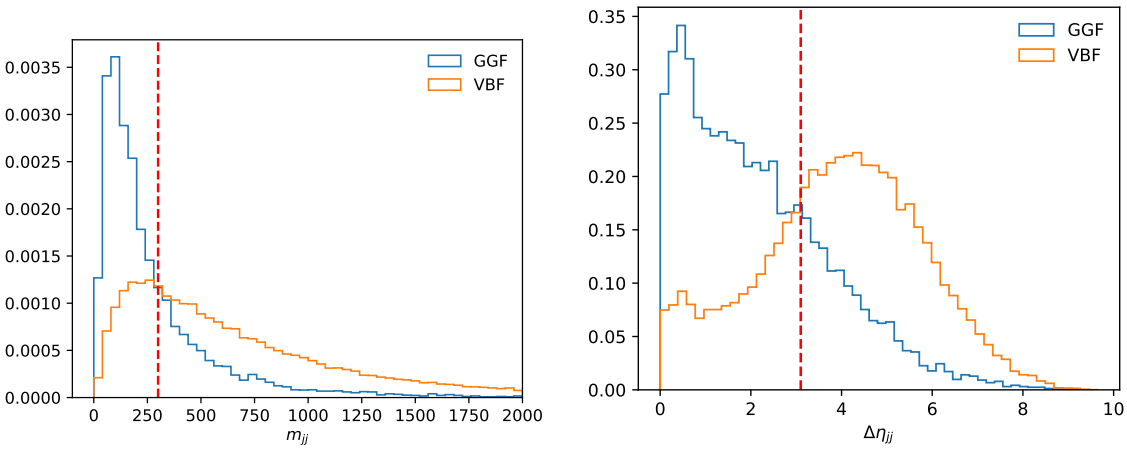
Figure 1 shows the distributions of m_{jj} (the invariant mass of the two leading jets) and $\Delta\eta_{jj}$ (the pseudorapidity difference between the two leading jets). The scatter plot of m_{jj} versus $\Delta\eta_{jj}$ is presented in Figure 2.

2.3 Event image

The inputs for the neural networks are event images [6, 7, 8]. These images are constructed from events that pass the kinematic selection criteria described in section 2.2. Each event image has three channels corresponding to calorimeter towers, track, and photons. The following preprocessing steps are applied to all event constituents:

1. Translation: Compute the p_T-weighted center in the ϕ coordinates, then shift this point to the origin.
2. Flipping: Flip the highest p_T quadrant to the first quadrant.
3. Pixelation: Pixelate in a $\eta \in [-5, 5]$, $\phi \in [-\pi, \pi]$ box, with 40×40 pixels

Figure 3 shows the event images for GGF and VBF production modes.



(a) m_{jj} distribution (b) $\Delta\eta_{jj}$ distribution

Figure 1: Distributions of the invariant mass m_{jj} and pseudorapidity difference $\Delta\eta_{jj}$ of the two leading jets. Red dashed lines are selection cuts used to construct mixed datasets.

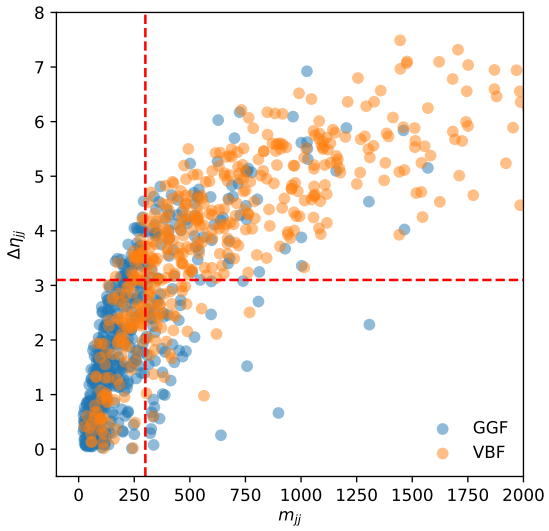
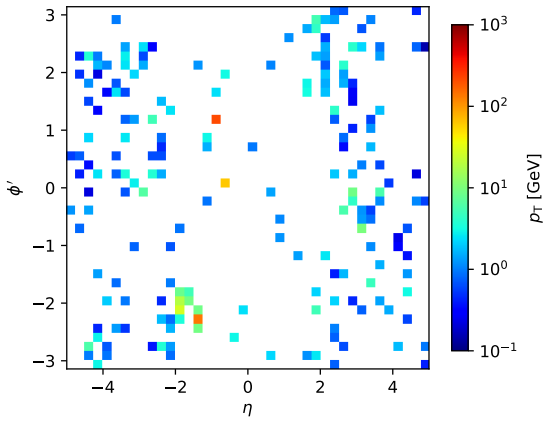
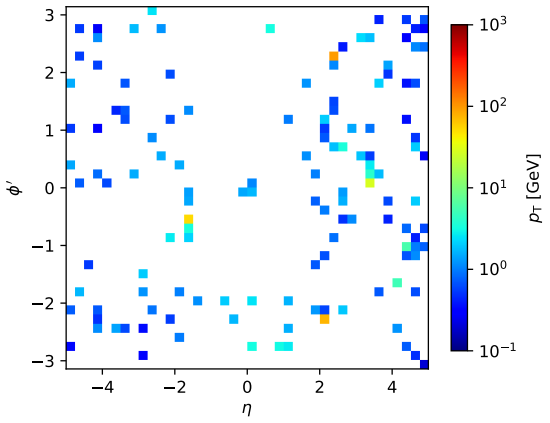


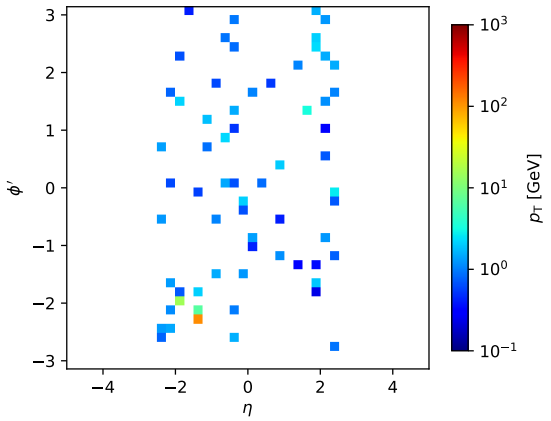
Figure 2: Scatter plot of m_{jj} versus $\Delta\eta_{jj}$. Red dashed lines are selection cuts used to construct mixed datasets.



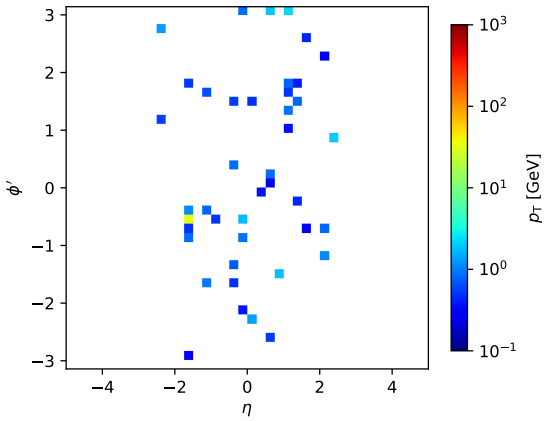
(a) GGF: Calorimeter Tower



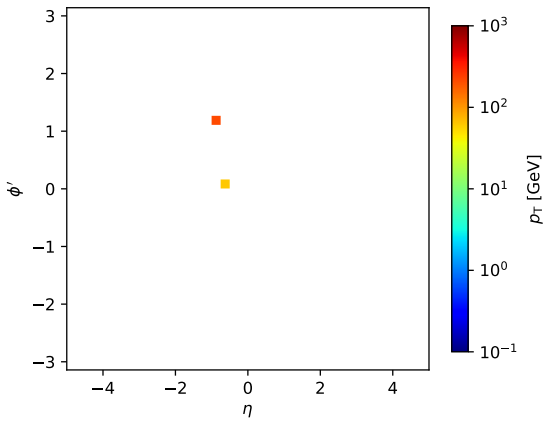
(b) VBF: Calorimeter Tower



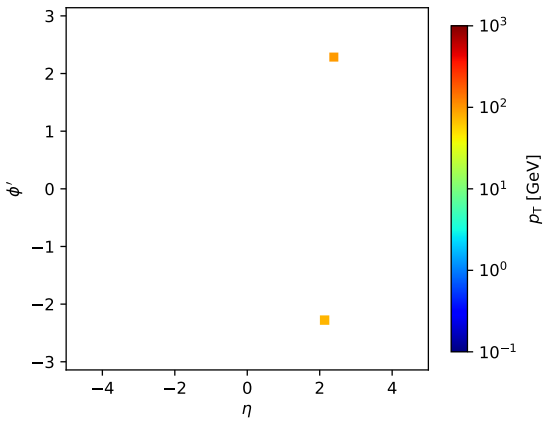
(c) GGF: Track



(d) VBF: Track



(e) GGF: Photon



(f) VBF: Photon

Figure 3: Event images for GGF and VBF production, separately shown for calorimeter towers, tracks, and photons.

References

[1] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations," *JHEP*, vol. 07, p. 079, 2014.

[2] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An introduction to PYTHIA 8.2," *Comput. Phys. Commun.*, vol. 191, pp. 159–177, 2015.

[3] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi, "DELPHES 3, A modular framework for fast simulation of a generic collider experiment," *JHEP*, vol. 02, p. 057, 2014.

[4] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual," *Eur. Phys. J. C*, vol. 72, p. 1896, 2012.

[5] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *JHEP*, vol. 04, p. 063, 2008.

[6] A. Butter *et al.*, "The Machine Learning landscape of top taggers," *SciPost Phys.*, vol. 7, p. 014, 2019.

[7] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman, "Jet-images — deep learning edition," *JHEP*, vol. 07, p. 069, 2016.

[8] G. Kasieczka, T. Plehn, M. Russell, and T. Schell, "Deep-learning Top Taggers or The End of QCD?," *JHEP*, vol. 05, p. 006, 2017.