Disentangling Boosted Higgs Boson Production Modes with Machine Learning

Yi-Lun (Alan) Chung[1], Shih-Chieh Hsu[2] and Benjamin Nachman[3]

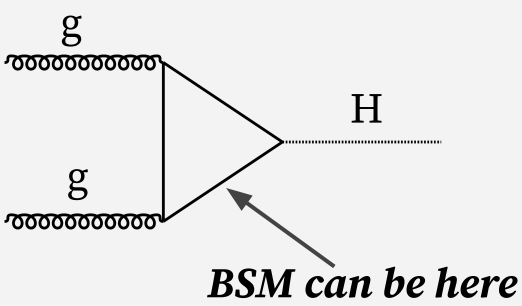
National Tsing Hua University[1]

University of Washington[2],

Lawrence Berkeley National Laboratory[3]

2020/ 7/ 20 @ BOOST 2020

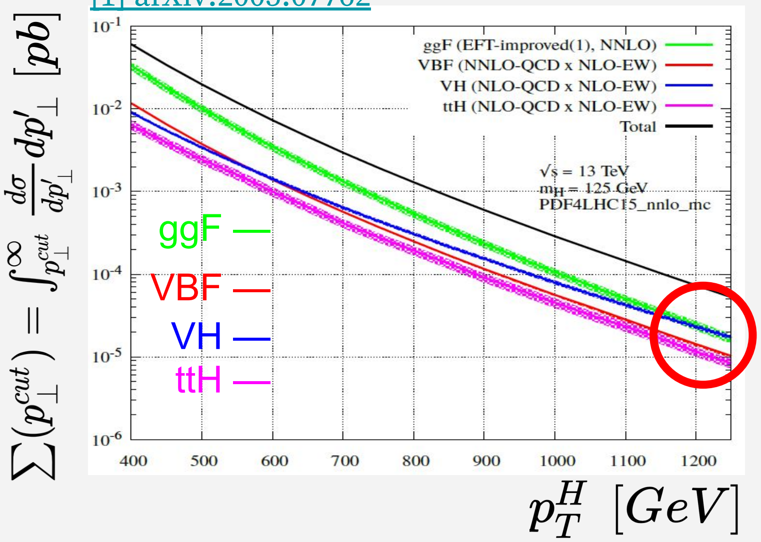
Motivation

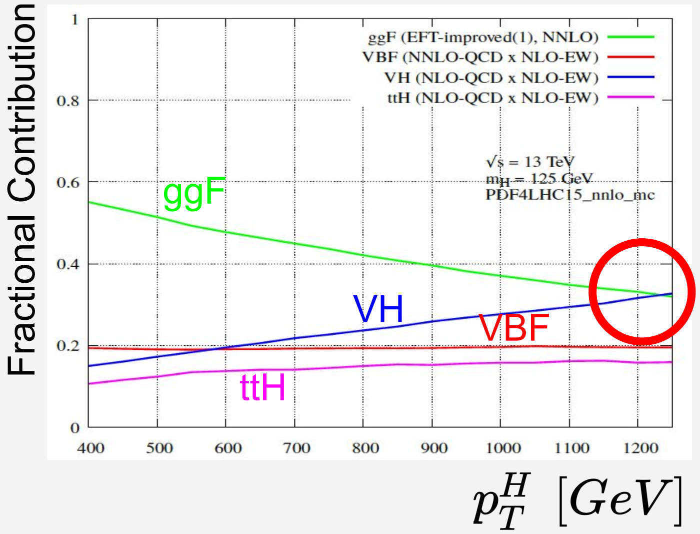


- High pT Higgs from
  - the SM Higgs, e.g. ggF
  - Beyond the Standard Model

- Many Higgs productions other than ggH could be substantial in the boosted region.

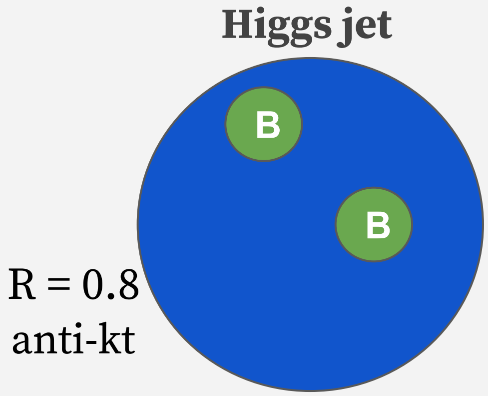
[1] arXiv:2005.07762





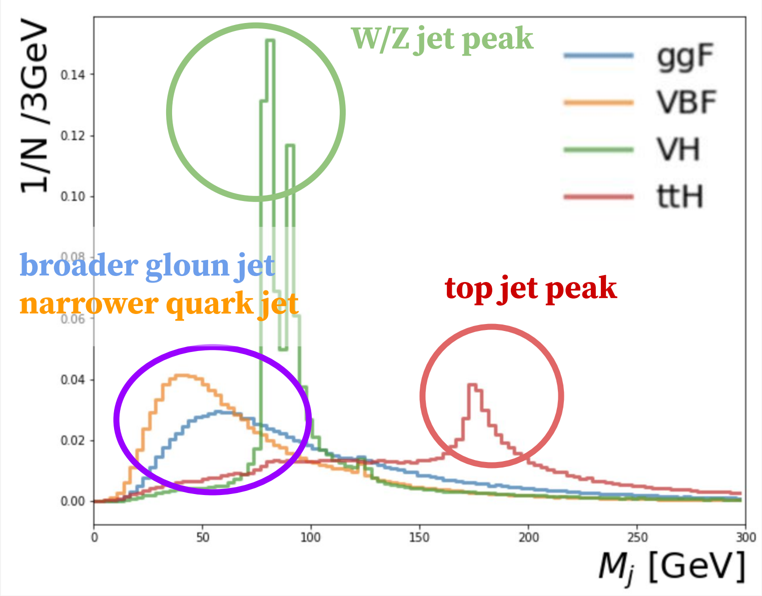
Methods of Analysis

- Jet's clustering: R=0.8, anti-kt
- Higgs jet tagging via **ghost-association method**
  arXiv:1507.00508
- High-level features:
  - M_j , η_j , $|\Delta\eta_{jj}|$, girth $\sum_{i\in J} \frac{p^j_{T,i} r^J_i}{p^J_T}$, the central integrated jet shape $\frac{1}{N}\sum^N_{i\in J} \frac{p^J_{T,i}(0<r^J_i<0.1)}{p^J_T}$
  - Tool: **Boosted-Decision Tree (BDT)**
- Low-level features:
  - **global full-event** images and **local non-Higgs-leading jet's** images
  - Tool: The **two-stream convolutional neural network(2CNN)**
  arXiv:1807.10768

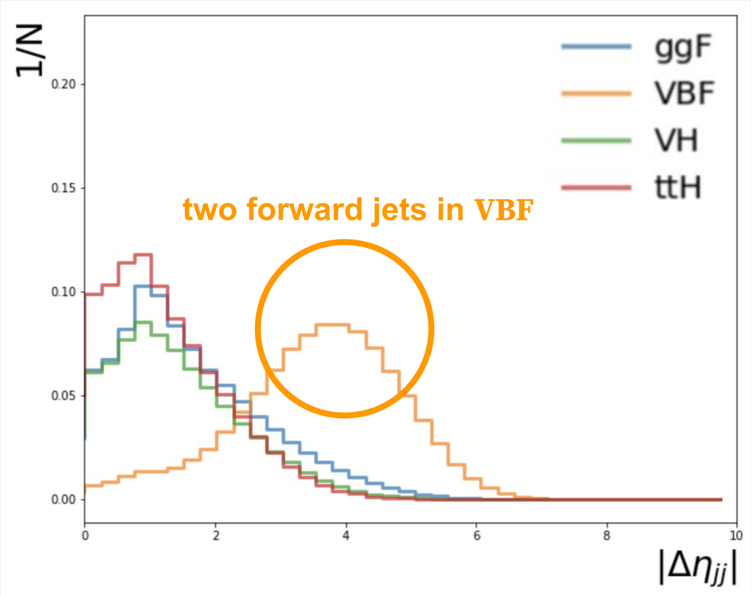


High-Level Features for BDT

- characteristic high-level features of four Higgs production mechanisms

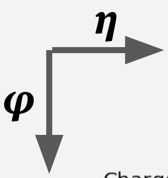


invariant mass of non-Higgs-leading jet



$|\Delta\eta|$ between non-Higgs-leading and non-Higgs-subleading jets

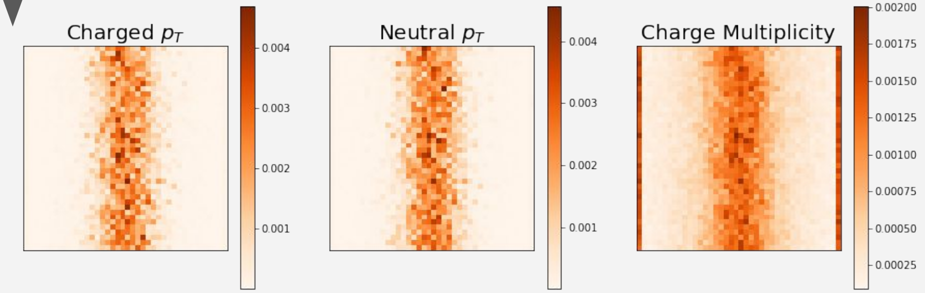
Low-Level Features



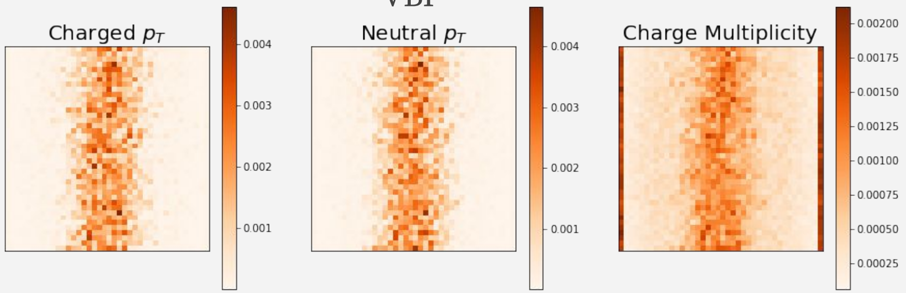
Global

average full-event images

ggF



VBF

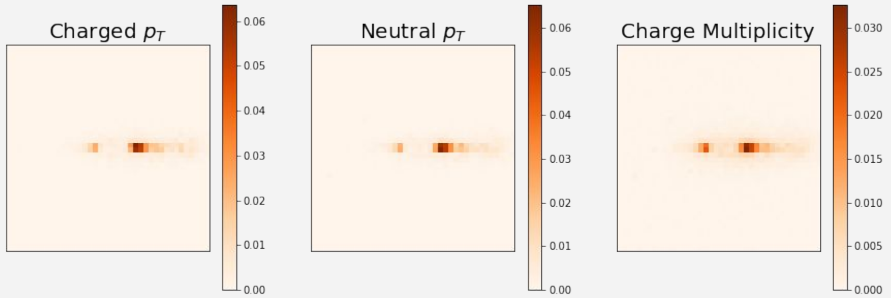


hard to find features by eyes!

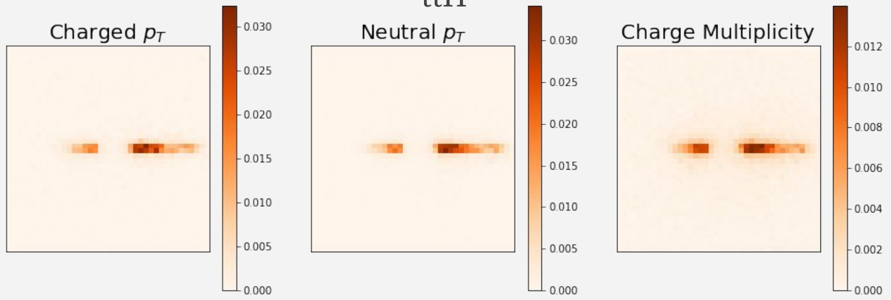
Local

average non-Higgs-leading jet images

VH



ttH

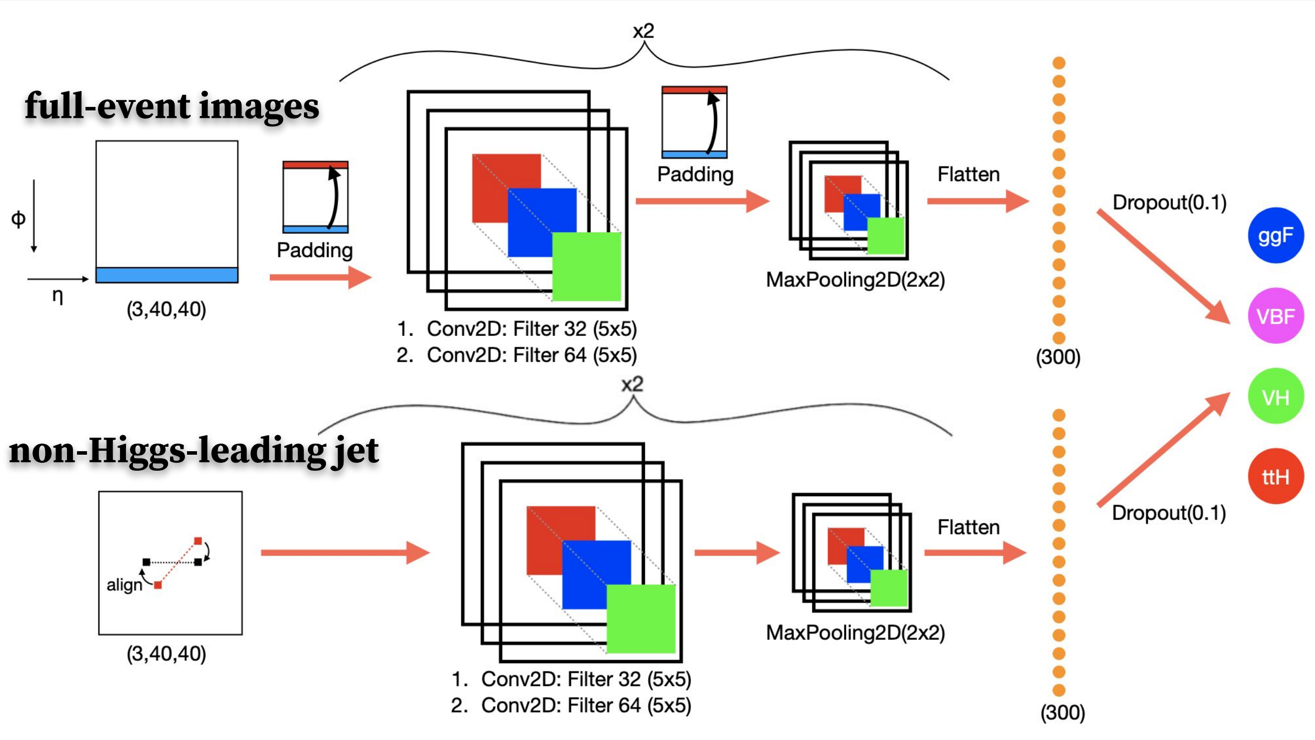


clear jet's substructures for VH and ttH

2CNN Architecture

arXiv:1807.10768

- first stream acting on global information
- second stream acting on local information



Activation Fn. for Dense Layer :
ReLU

Activation Fn. for CNN Layer :
ReLU

Activation Fn. for last Layer :
Softmax

Loss:
categorical crossentropy

Optimizers:
Adadelta

Batch size:
512

Epochs:
100

Total Parameters:
2,395,196

Training/valdation/Test:
170,000/ 25,000/ 229,000
for each production

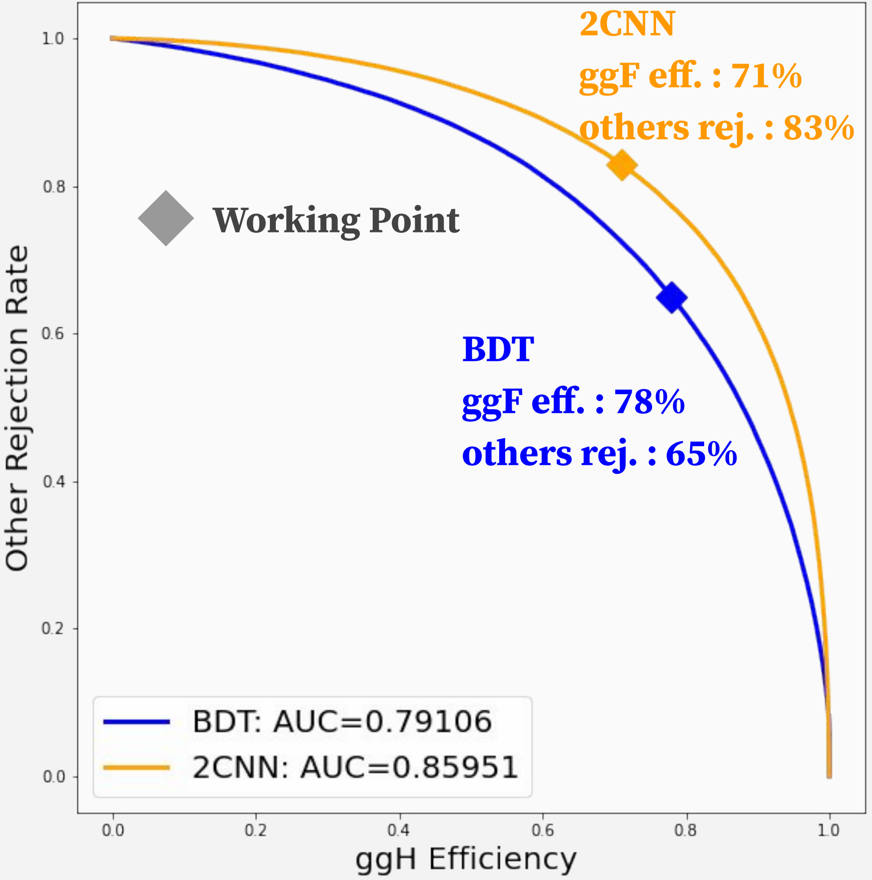
Performance of Two Classifiers

- made by p(ggF)

- weighted ROC curves
  - signal: ggF
  - backgrounds: **VBF+VH+ttH**
  - events are weighted by corresponding cross section

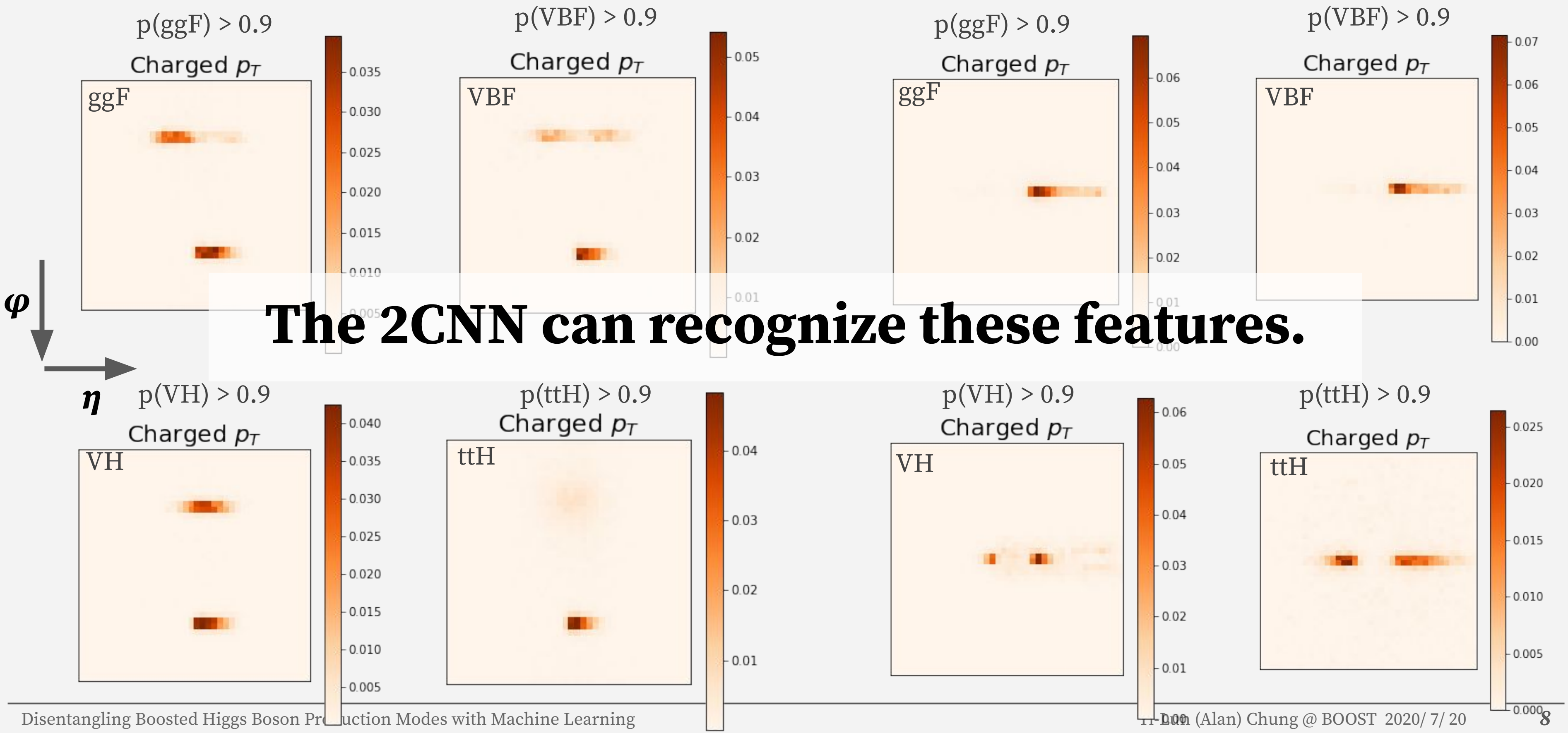
- signal efficiency v.s. background rejection rate
  - **efficiency** = N[survived events] / N[total events]
  - **rejection** = 1 - N[survived events] / N[total events]

- **2CNN has better performance.**



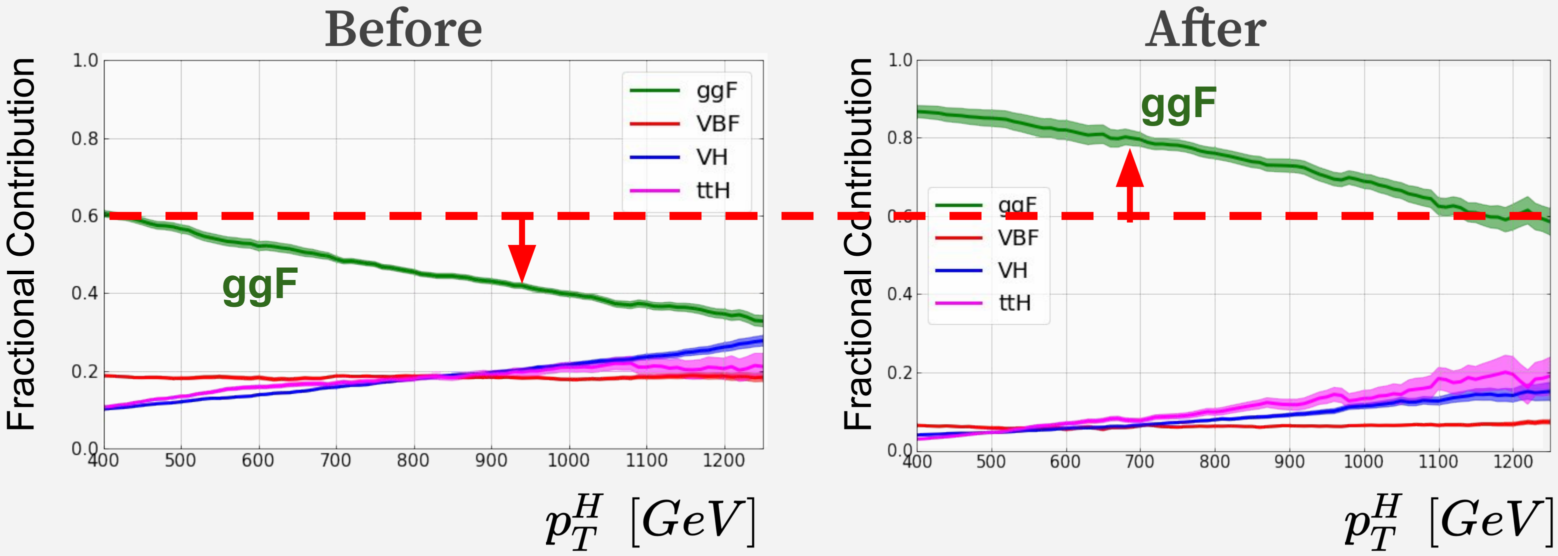
Learning in Characteristic Images

Average global full-event image | **Average local non-Higgs-leading jet image**



The 2CNN can recognize these features.

Result After 2CNN



These two plots are passed preselection and included decay branching ratio

The 2CNN highly increases the ggF fraction in whole pT range!

Conclusion

1. By using 2CNN method, we can provide exceptionally clear separation for boosted Higgs bosons produced via ggF at the LHC.

2. Our architecture has **4-class outputs** and contains one stream acting on **global event information**, and the other stream acting on **local non-Higgs-leading jet information**.

3. This state-of-art architecture makes us not only use fully information in the event but also enable us to understand the physics of four Higgs productions that neural network learned.

4. The approach in this study additionally has the potential to **improve the precision for other Higgs production modes** in extreme regions of phase space.

Thank you for your attentation.

Reference

[1] the gluon-fusion Working Group, Precise predictions for boosted Higgs production, arXiv:2005.07762

[2] Pagani, Davide and Shao, Hua-Sheng and Zaro, Marco, RIP H →bb: How other Higgs production modes conspire to kill a rare signal at the LHC, arXiv:2005.10277

[3] Joshua Lin, Marat Freytsis, Ian Moult and Benjamin Nachman, Boosting H→ b b with Machine Learning, 10.1007/JHEP10(2018)101

[4]CMS Collaboration, Inclusive search for a highly boosted Higgs boson decaying to a bottom quark-antiquark pair, arXiv:1709.05543

[5]Andy Buckley, Chris Pollard, QCD-aware partonic jet clustering for truth-jet flavour labelling, arXiv:1507.00508

[6] the gluon-fusion Working Group, Recommended predictions for the boosted-Higgs cross section

[7] Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector, 10.23731/CYRM-2017-002

[8] Report from Working Group 2 on the Physics of the HL-LHC, and Perspectives at the HE-LHC, Higgs Physics at the HL-LHC and HE-LHC

[9] Jessie Shelton, TASI Lectures on Jet Substructure, arXiv:1302.0260v2

[10] N Belyaev, R Konoplich and K Prokofiev, Study of kinematic observables sentitive to the Higgs boson production channel in pp→Hjj process, J.Phys.Conf.Ser. 934 (2017)

Backup

Monte Carlo Samples

Simulation Setup:

Simulation Setup	
$\sqrt{s}$	13 GeV
Parton Generator	MG5_aMC@NLO
PDF set	PDF4LHC15_nnlo_mc
Central scale	$\mu_R=\mu_F=\sum m_T/2$
Parton Shower	Pythia8 with FxFx merging

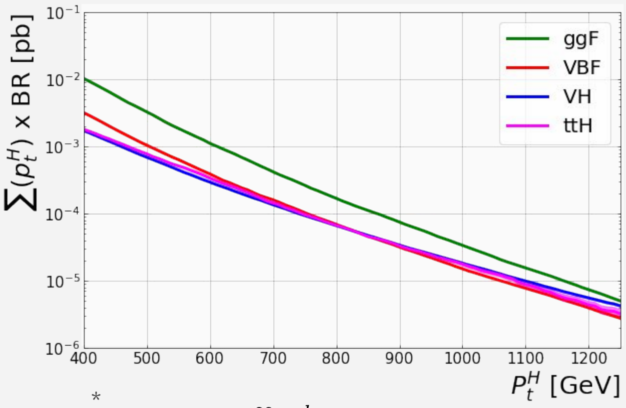
Preselection
Soft Drop with $\beta = 0$ and $z_{cut} = 0.1$ in leading double-b jet
invariant mass of leading double-b > 400 GeV
$-6 < \rho = log(m^2/(pT)^2) < -2.1$
$N_2 \leq 0.4$

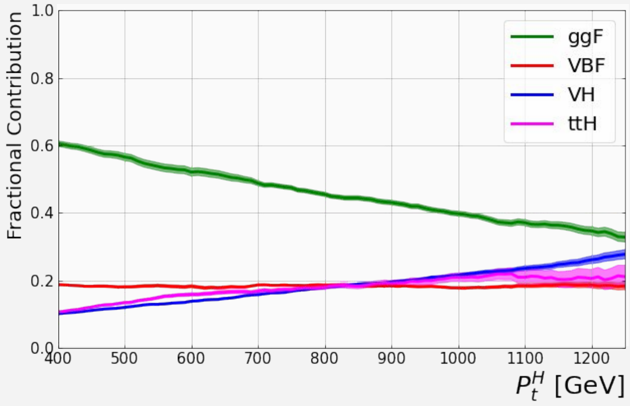
[3]arXiv:1807.10768
[4]arXiv:1709.05543

Training/Testing Information:

	ggH	VBF	VH	ttH
# of training	170,000	170,000	170,000	170,000
# of test	229,000	229,000	229,000	229,000

Cumulative cross section and Fractional Distribution:





* $\sum(p_\perp^{cut}) = \int_{p_\perp^{cut}}^{\infty} \frac{d\sigma}{dp'_\perp} dp'_\perp$

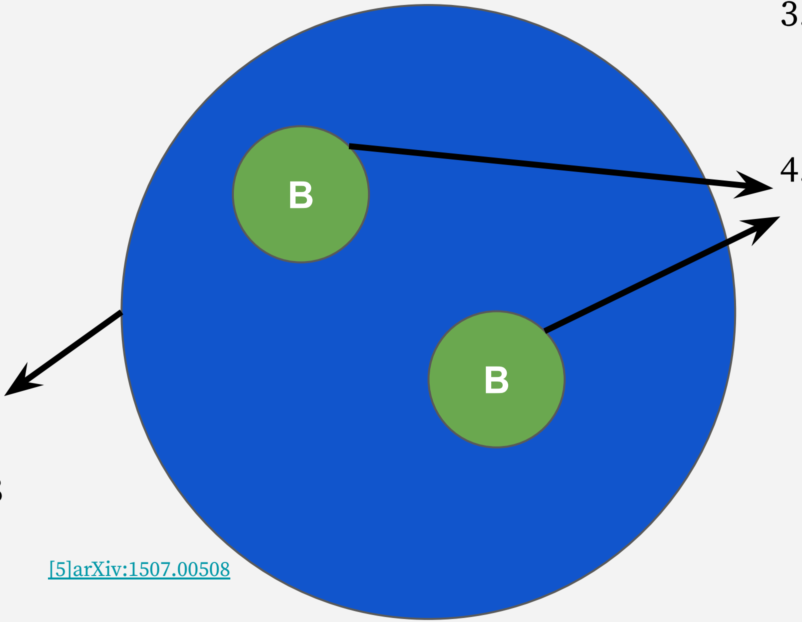
* These two plots are passed preselection and included decay branching ratio

Higgs-Jet-Tagging Method

Higgs jet is recognized by double b-tagging due to the hardronic Higgs decay.

Double-B Hadrons-tagging via **ghost-association [5]** method is used to do double b-tagging in this study.

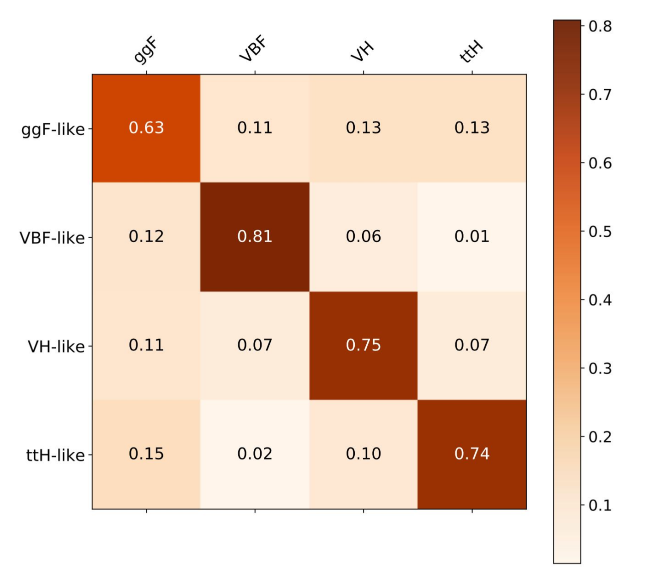
1. Find B hadrons which pt > 5 GeV before they decay.
2. Multiply infinitesimal value to B hadrons, it is **ghosted B hadrons**.
3. Adding this ghosted B hadrons into the final state list and cluster the jets
4. **If large R(=0.8) jet contains two ghost-associated B hadrons, it will be tagged to the Higgs jet**.



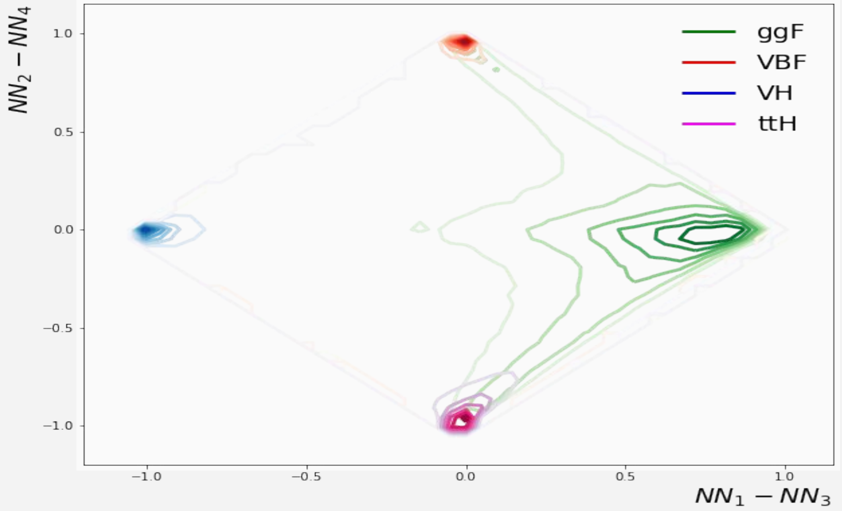
particle jet
anti-kt, R = 0.8

[5]arXiv:1507.00508

Output Scores in Different Angles



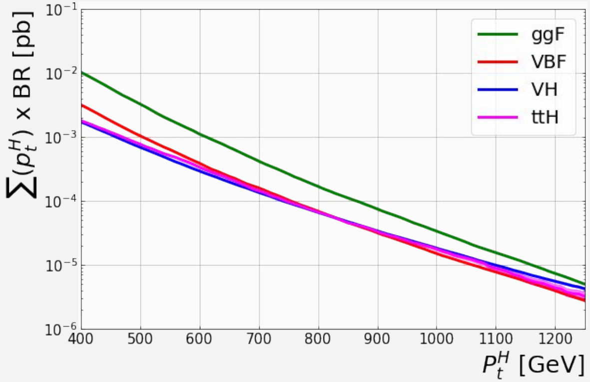
Confusion Matrix

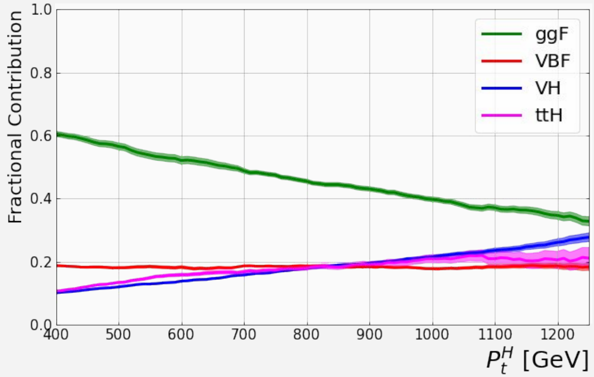


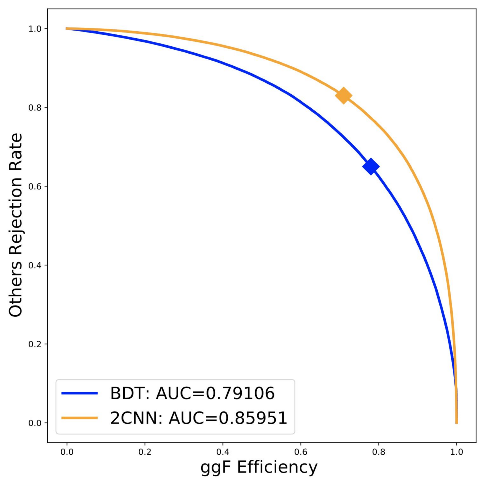
Visualization of differences between Neuron Network scores

2CNN Performance

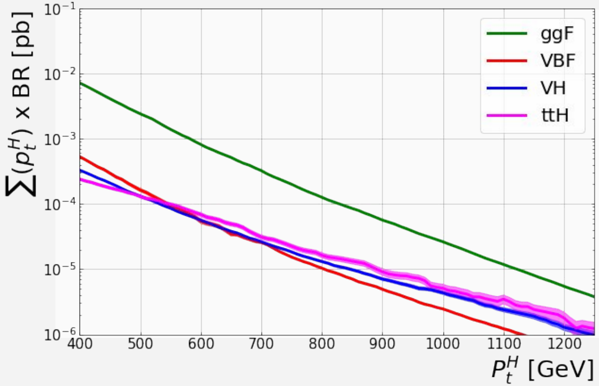
With preselection and dacay

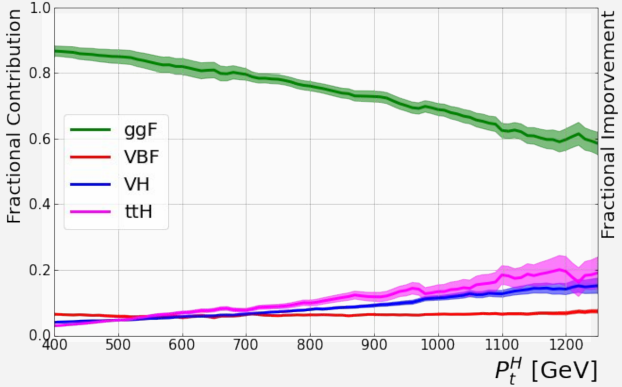


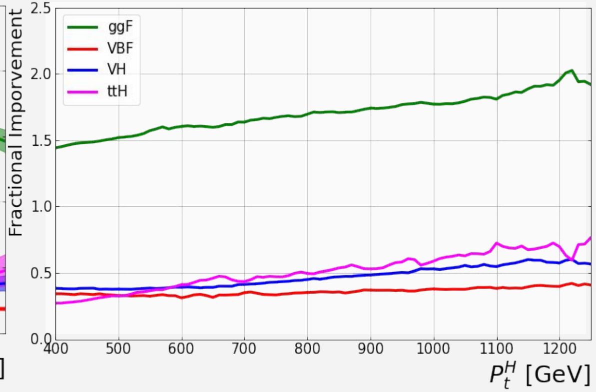




Applied 2CNN







* $\sum(p_\perp^{cut}) = \int_{p_\perp^{cut}}^{\infty} \frac{d\sigma}{dp'_\perp} dp'_\perp$