



---

# Disentangling Boosted Higgs Boson Production Modes with Machine Learning

[arXiv:2009.05930](https://arxiv.org/abs/2009.05930)

---

**Yi-Lun (Alan) Chung**<sup>1</sup>, Shih-Chieh Hsu<sup>2</sup> and Benjamin Nachman<sup>3</sup>

National Tsing Hua University<sup>1</sup>

University of Washington<sup>2</sup>,

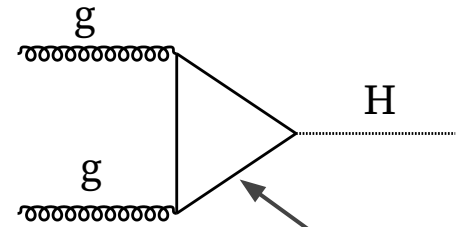
Lawrence Berkeley National Laboratory<sup>3</sup>

---

2021/ 02/ 16 @ IML Machine Learning Working Group

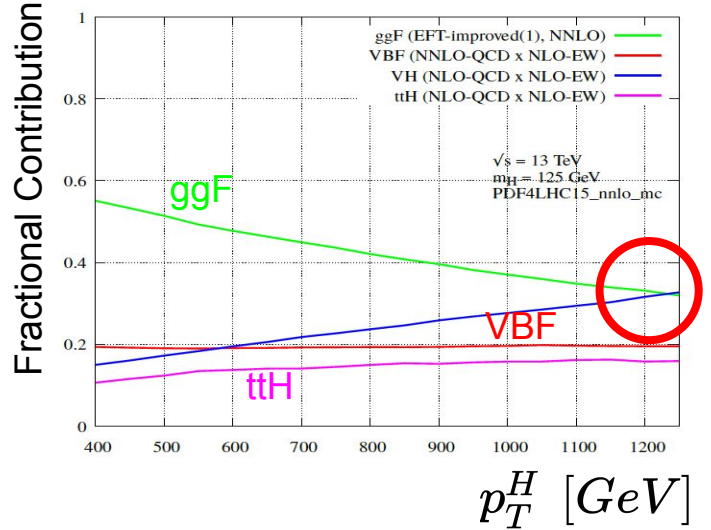
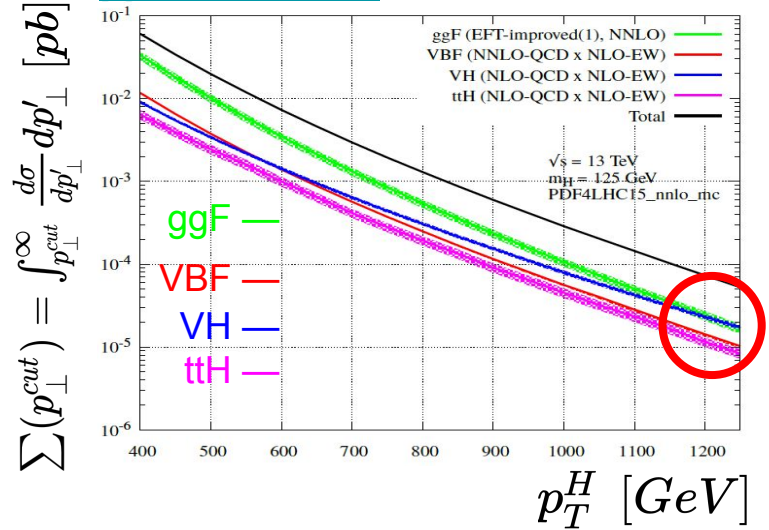
# Motivation

- High  $p_T$  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.



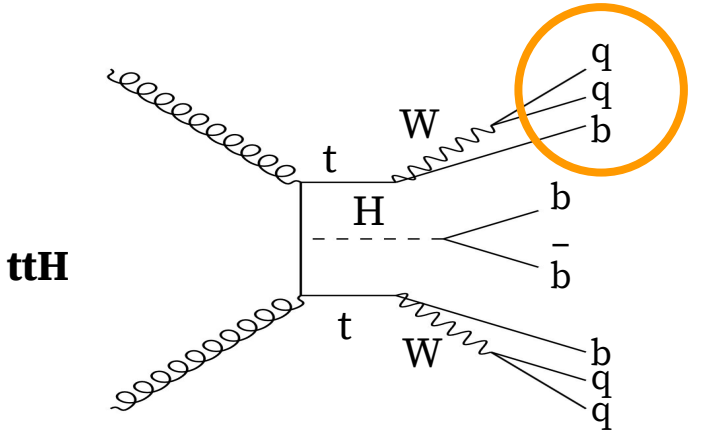
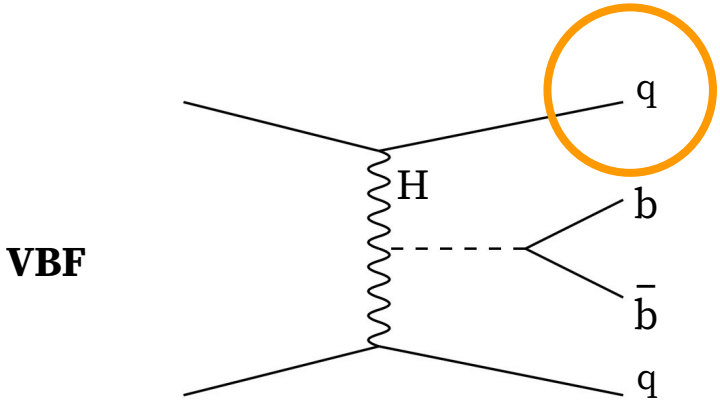
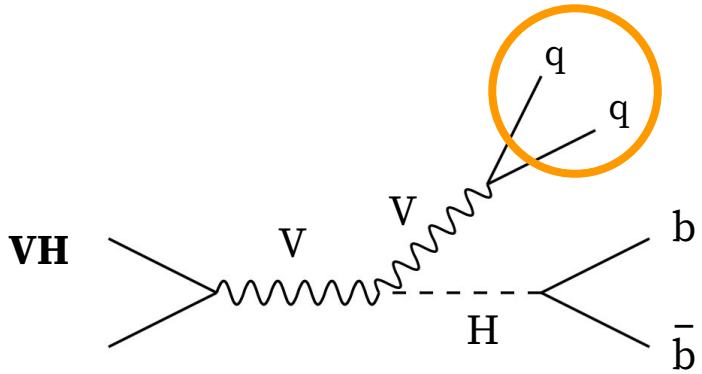
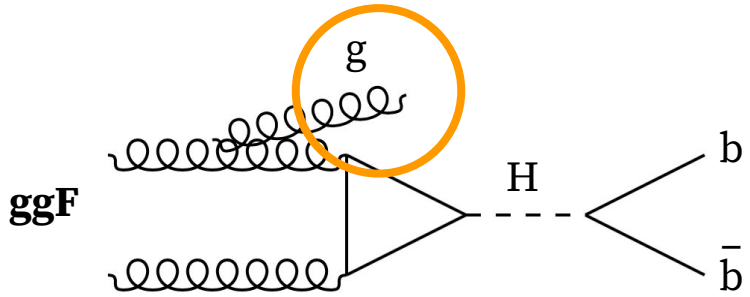
*BSM can be here*

[1] arXiv:2005.07762



# Leading Non-Higgs Jet

- The **leading non-Higgs jet** substructure and global event topology will be used to develop the best boosted Higgs classifier.



# Methods of Analysis

- Jet's clustering:  $R=0.8$ , anti-kt

- Higgs jet tagging via **ghost-association method**

[arXiv:1507.00508](https://arxiv.org/abs/1507.00508)

- High-level features:

-  $M_j, \eta_j, |\Delta\eta_{jj}|$ , girth  $\sum_{i \in J} \frac{p_{T,i}^j r_i^J}{p_T^J}$ , the central integrated jet shape  $\frac{1}{N} \sum_{i \in J} \frac{p_{T,i}^J (0 < r_i^J < 0.1)}{p_T^J}$

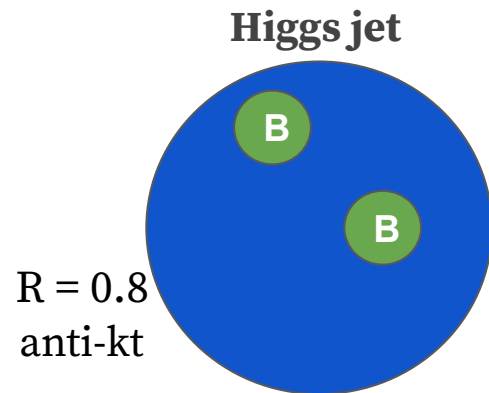
- Tool: **Boosted-Decision Tree (BDT)**

- Low-level features:

- **global full-event** images and **local leading non-Higgs jet's** images

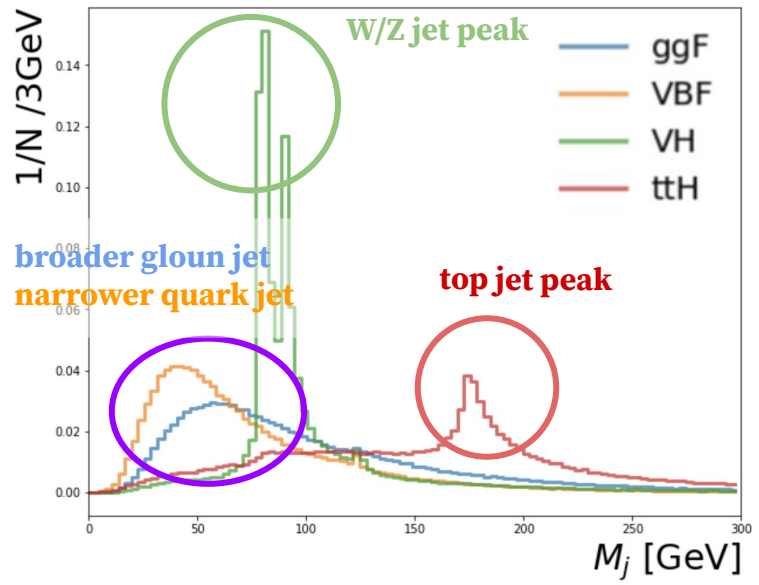
- Tool: The **two-stream convolutional neural network(2CNN)**

[arXiv:1807.10768](https://arxiv.org/abs/1807.10768)

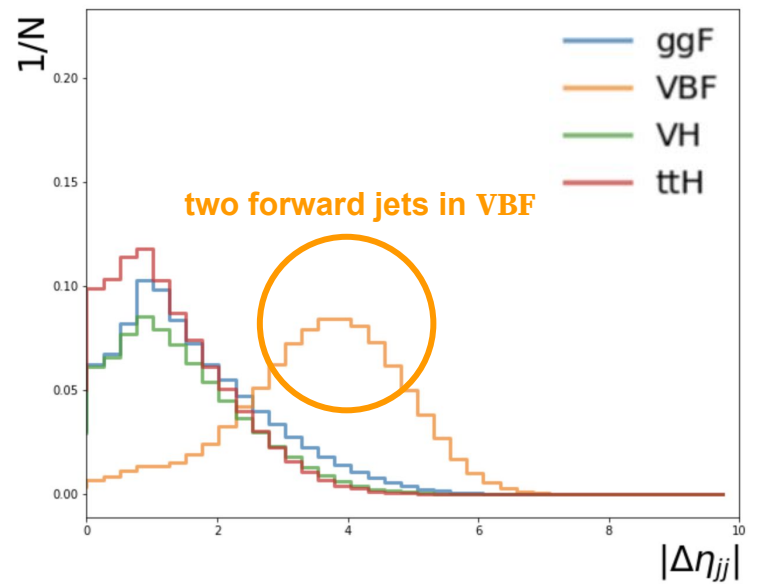


# High-Level Features for BDT

- characteristic high-level features of four Higgs production mechanisms

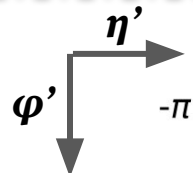


invariant mass of leading non-Higgs jet

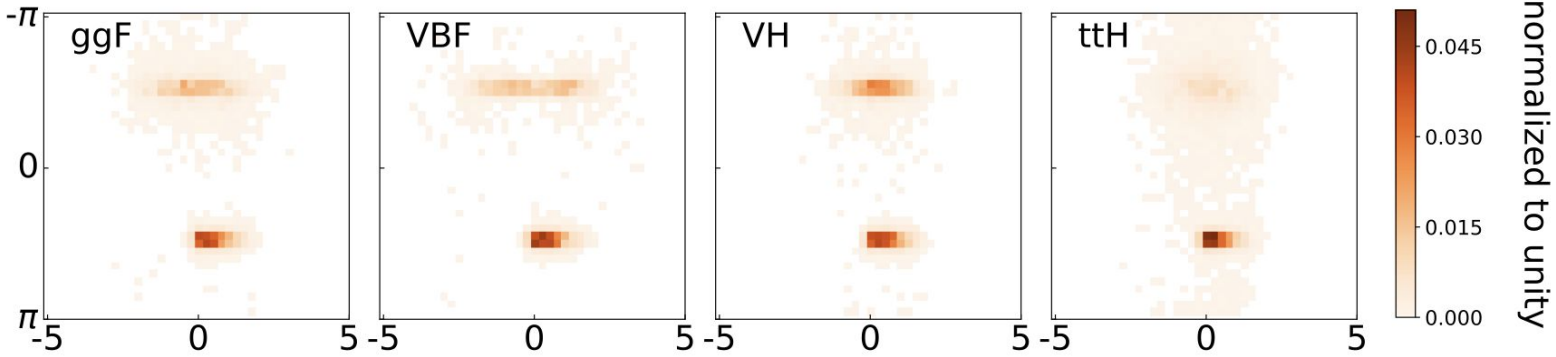


$|\Delta\eta|$  between leading non-Higgs and subleading non-Higgs jets

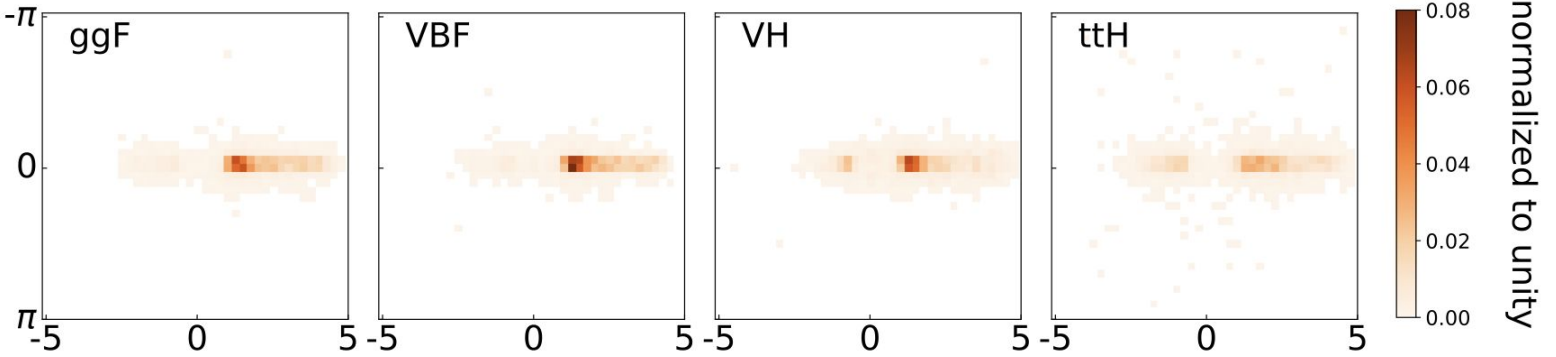
# Low-Level Features for 2CNN



Average global rotated full-event image



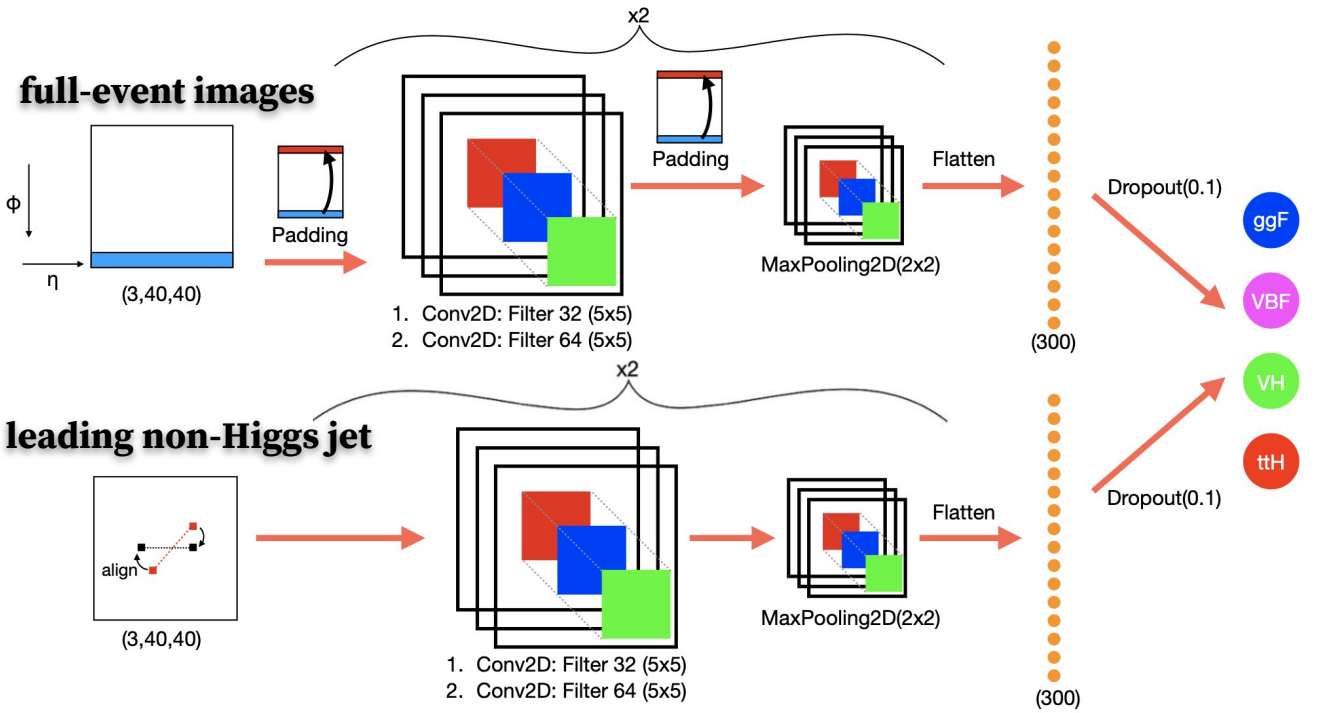
Average local leading non-Higgs jet image



# 2CNN Architecture

[arXiv:1807.10768](https://arxiv.org/abs/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

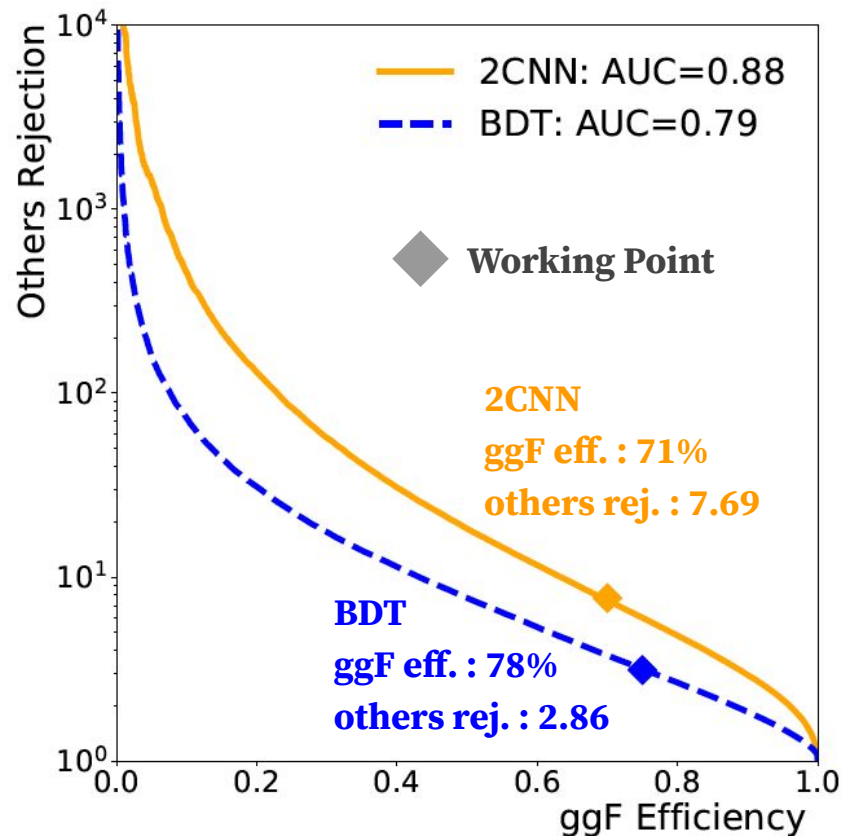
**Epochs:**  
100

**Total Parameters:**  
2,395,196

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

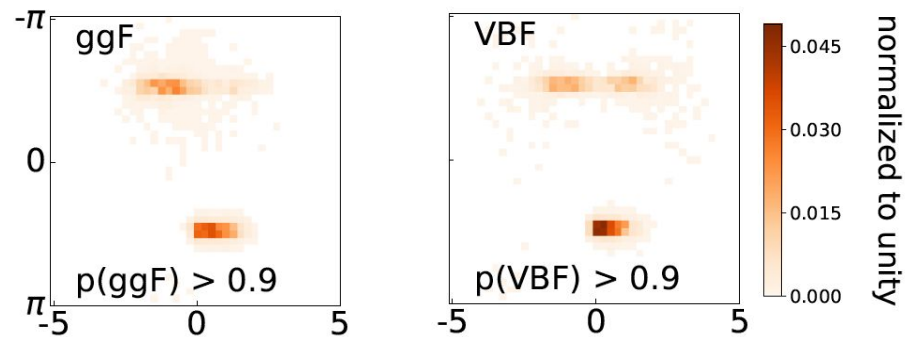
# Performance of Two Classifiers

- made by  $p(\text{ggF})$
- weighted ROC curves
  - signal: ggF
  - backgrounds: **VBF+VH+ttH**
  - events are weighted by corresponding cross section
- signal efficiency v.s. background rejection rate
  - **efficiency** =  $N[\text{survived events}] / N[\text{total events}]$
  - **rejection** =  $1/\text{efficiency}$
- **2CNN has better performance.**

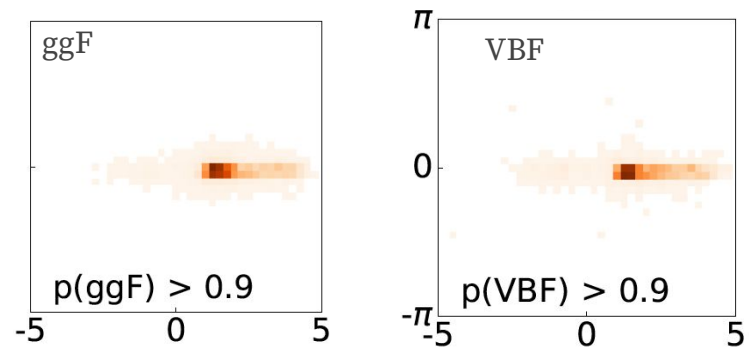


# Learning in Characteristic Images

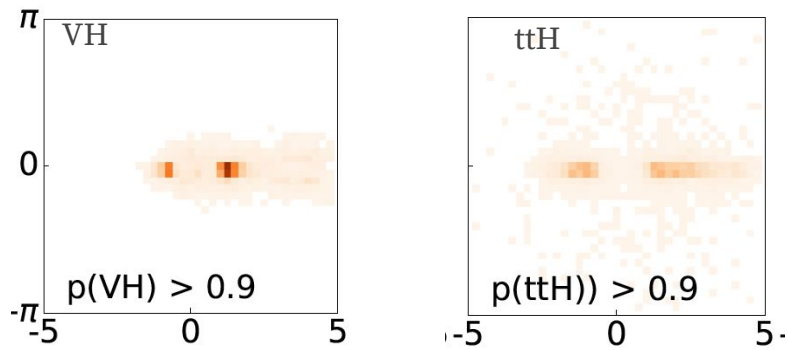
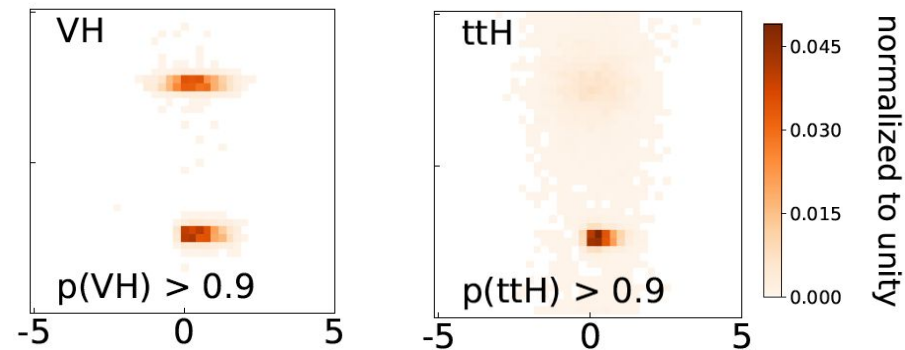
Average global full-event image



Average local leading non-Higgs jet image

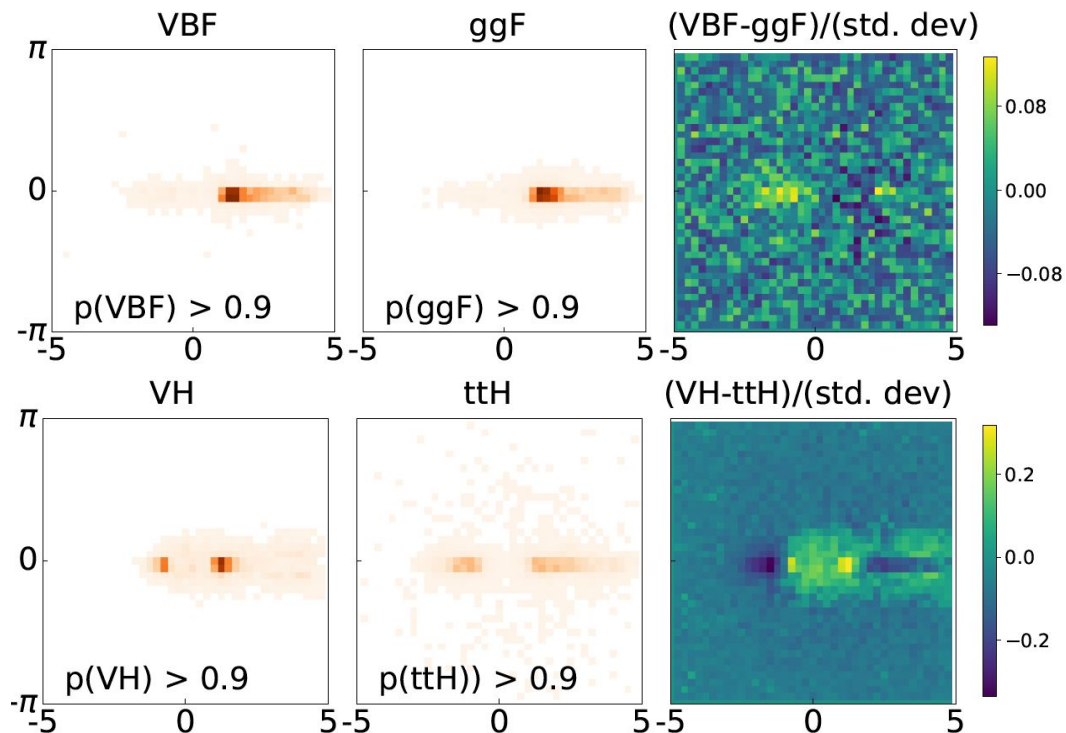


$\eta'$   
 $\phi'$



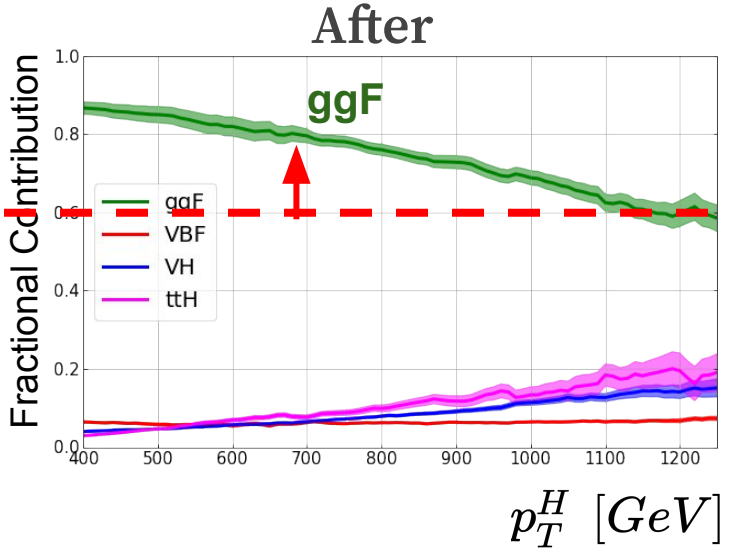
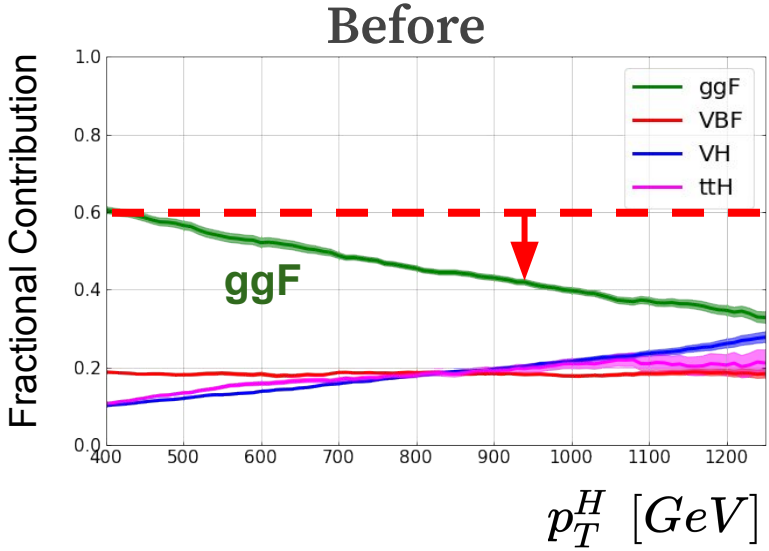
# Average difference between Jet Images

- leading non-Higgs jet images with high score



**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 leading non-Higgs 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 \rightarrow b\bar{b}\$ : How other Higgs production modes conspire to kill a rare signal at the LHC](#), arXiv:2005.10277
- [3] Joshua Lin, Marat Freytsis, Ian Moulton and Benjamin Nachman, [Boosting  \$H \rightarrow b\bar{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 sensitive to the Higgs boson production channel in  \$pp \rightarrow 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

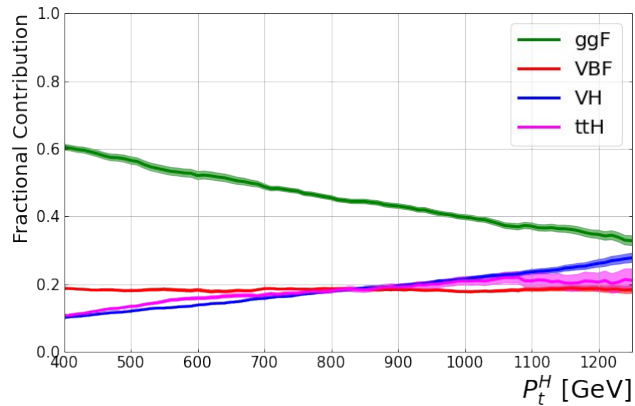
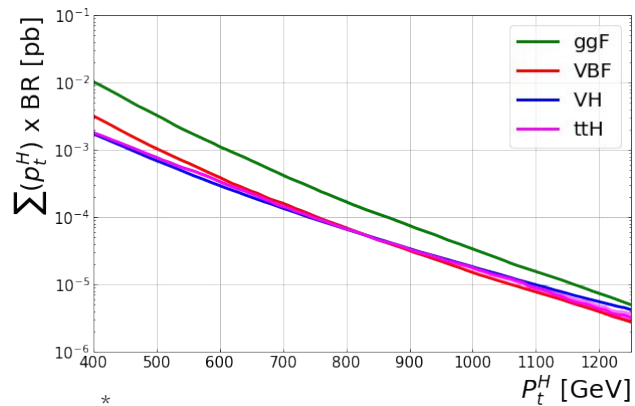
Preselection	
Soft Drop with $\beta = 0$ and $z_{cut} = 0.1$ in leading double-b jet	
invariant mass of leading double-b > 400 GeV	
$-6 < \rho = \log(m^2 / (pT)^2) < -2.1$	
$N_2 \leq 0.4$	

[\[3\]arXiv:1807.10768](#)  
[\[4\]arXiv:1709.05543](#)

## Training/Testing Information:

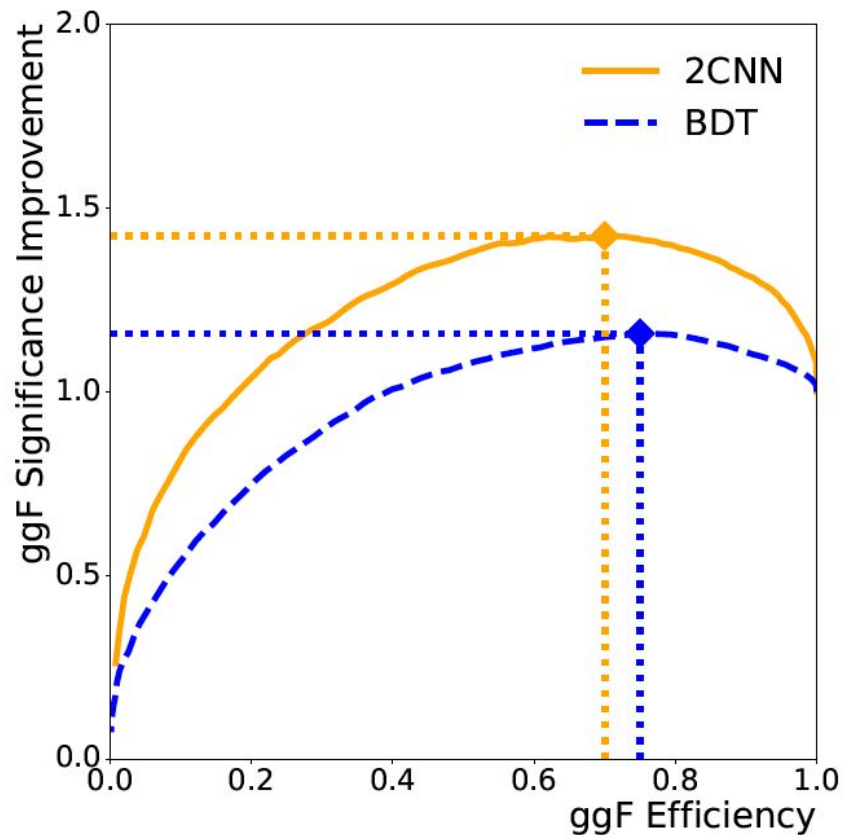
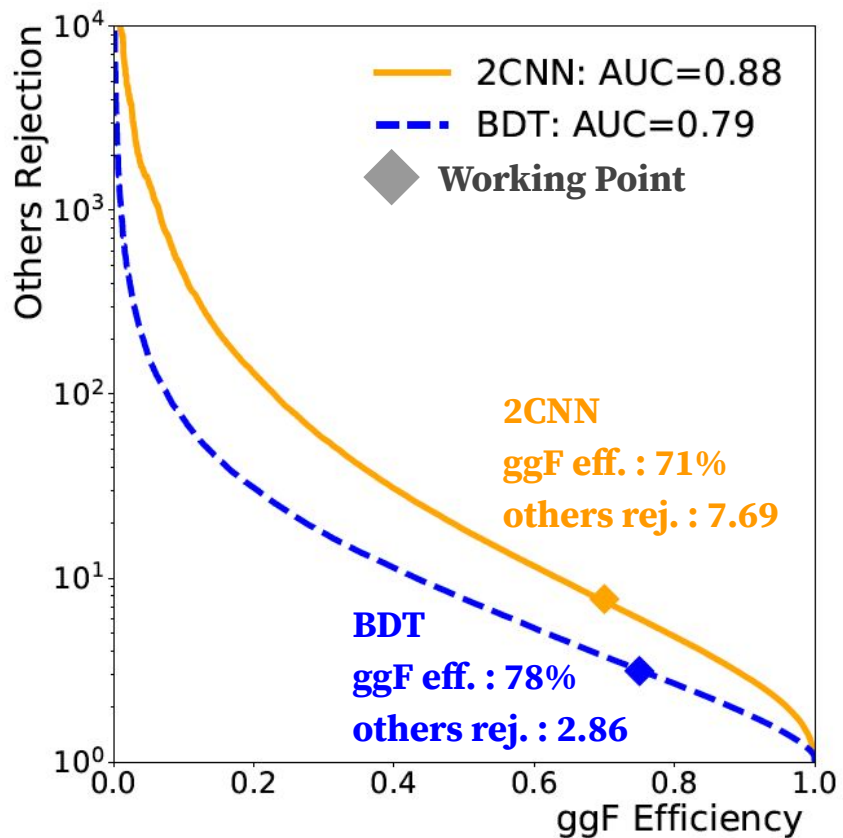
	ggH	VBF	VH	ttH
# of training	170,000	170,000	170,000	170,000
# of test	229,000	229,000	229,000	229,000

## Cumulative cross section and Fractional Distribution:

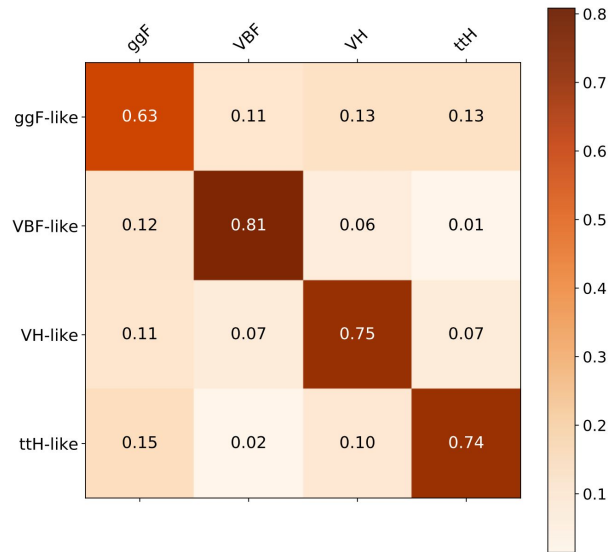


$$* \sum(p_{\perp}^{cut}) = \int_{p_{\perp}^{cut}}^{\infty} \frac{d\sigma}{dp'_{\perp}} dp'_{\perp}$$

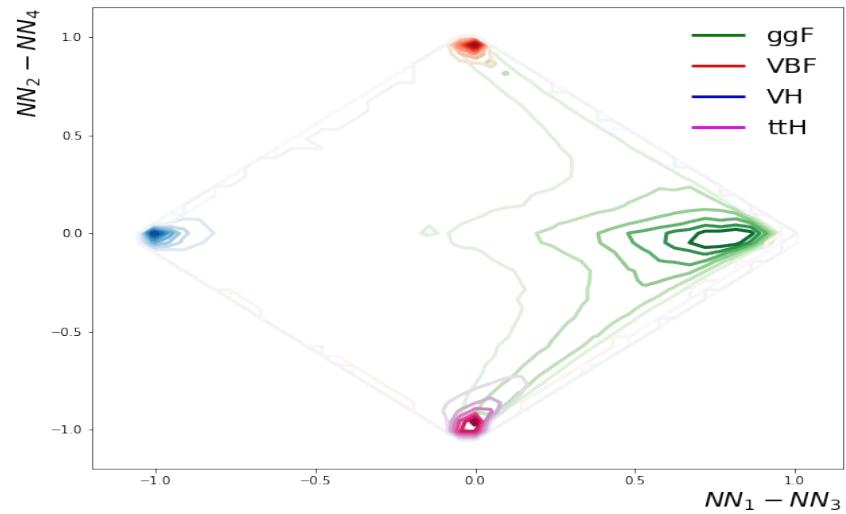
\* These two plots are passed preselection and included decay branching ratio



# Output Scores in Different Angles

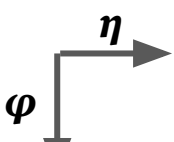


Confusion Matrix

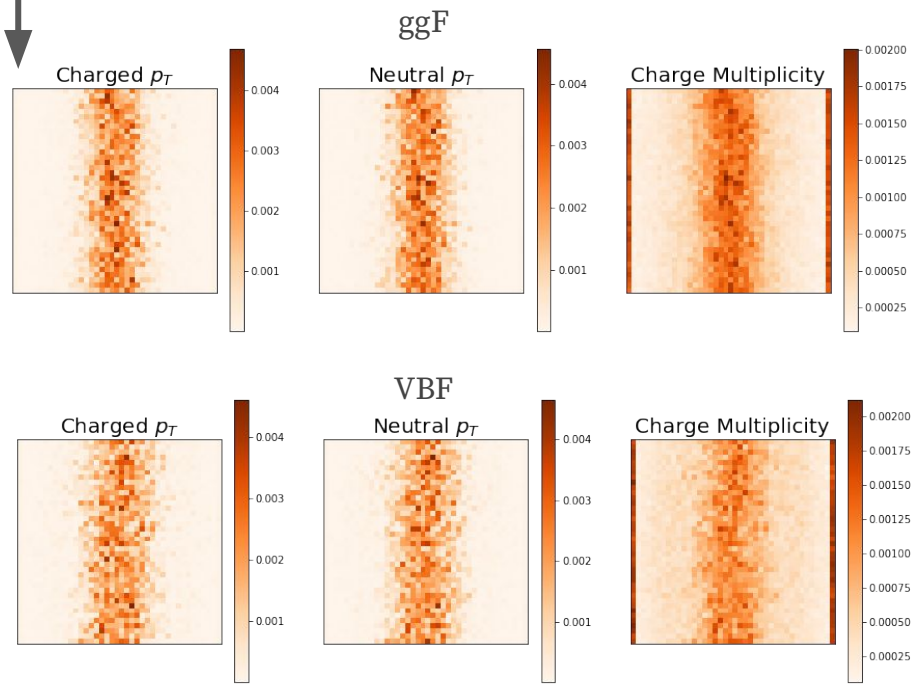


Visualization of differences between Neuron Network scores

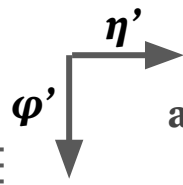
# Low-Level Features



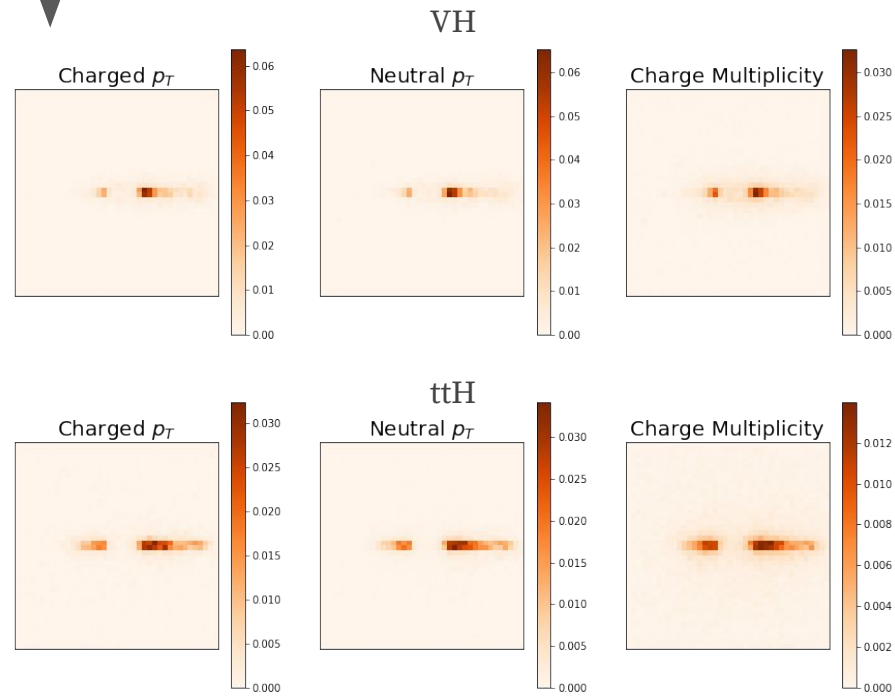
**Global**  
average full-event images



hard to find features by eyes!



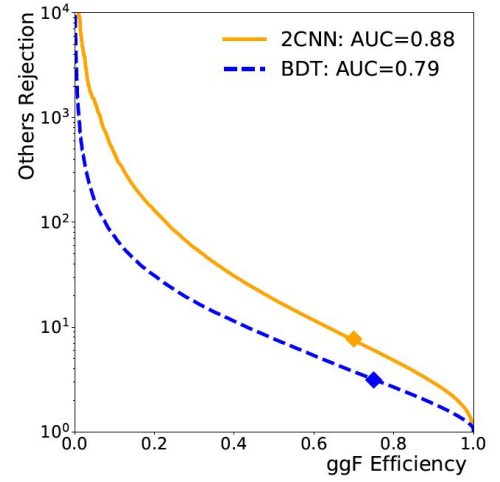
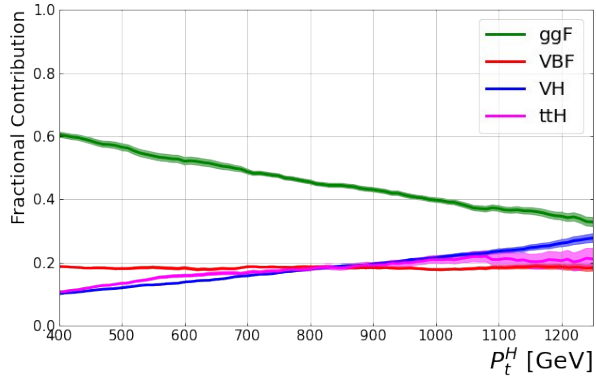
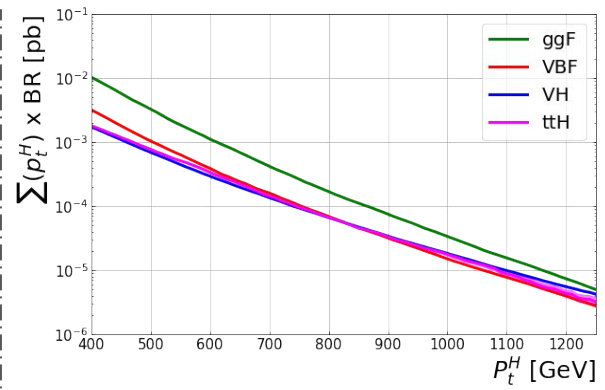
**Local**  
average leading non-Higgs jet images



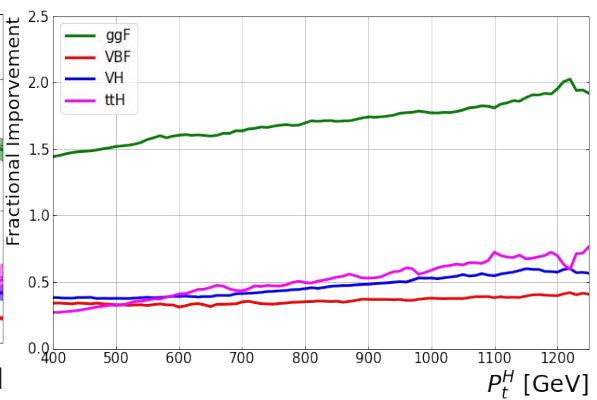
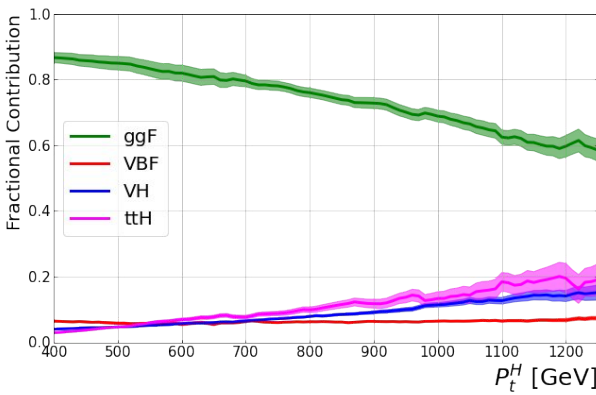
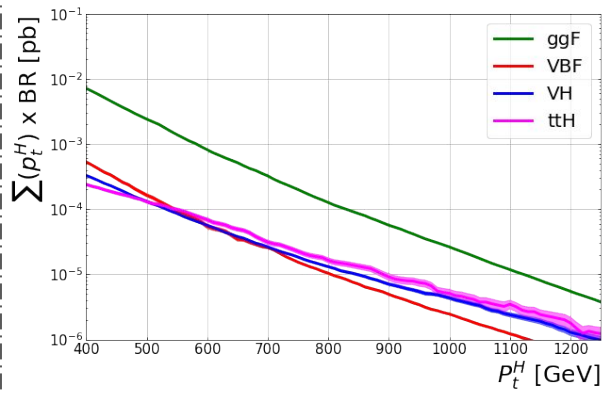
clear jet's substructures for VH and ttH

# 2CNN Performance

With preselection and decay



Applied 2CNN

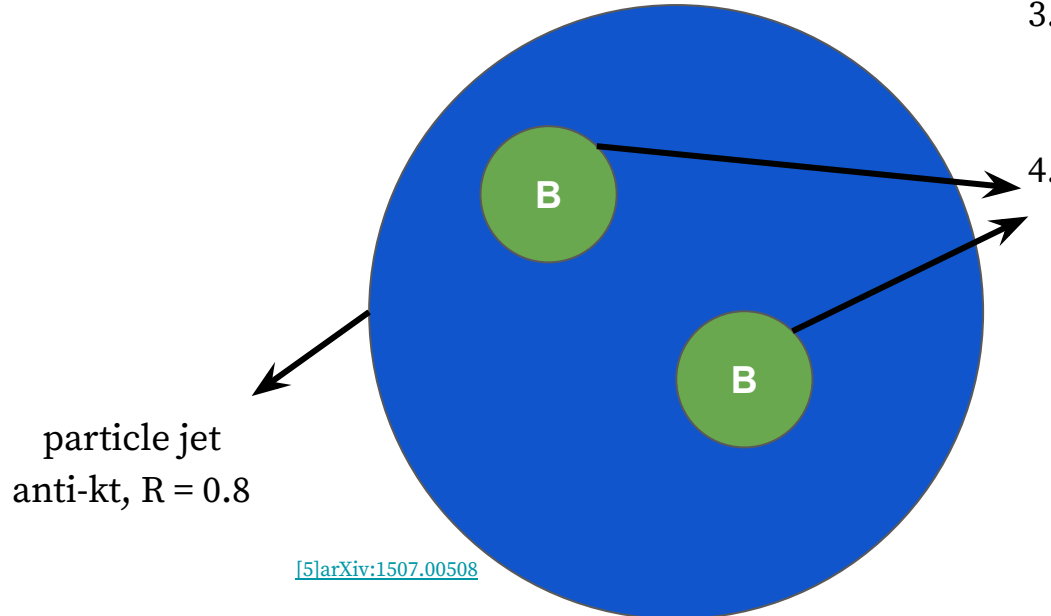


\* 
$$\sum(p_{\perp}^{cut}) = \int_{p_{\perp}^{cut}}^{\infty} \frac{d\sigma}{dp'_{\perp}} dp'_{\perp}$$

# Higgs-Jet-Tagging Method

Higgs jet is recognized by double b-tagging due to the hadronic 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  $p_t > 5$  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.**