Probing Higgs exotic decay at the LHC with machine learning

Ke-Pan Xie (谢柯盼)
University of Nebraska-Lincoln

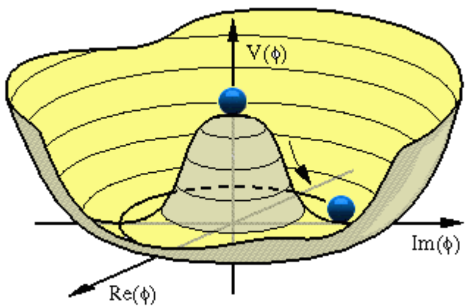
2021.11.25 @CLHCP2021, Nanjing (online)

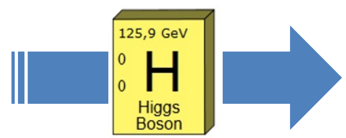
In collaboration with Sunghoon Jung, Zhen Liu and Lian-Tao Wang

[arXiv: 2109.03294]

• Introduction

Higgs: origin of mass & portal of new physics

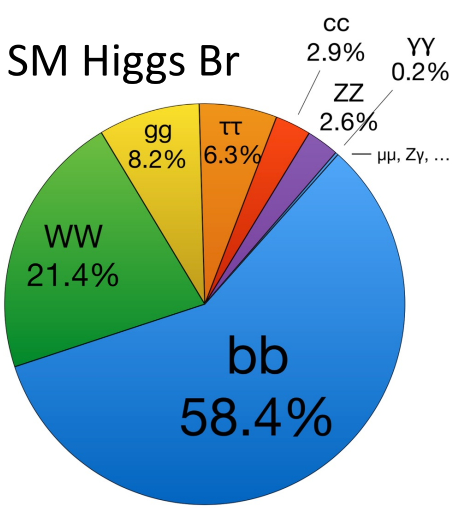




- Vacuum stability;
- First-order phase transition;
- New resonances;
- Dark matter;
- ……

Exotic decay as a probe for new physics

SM Higgs Br



- Portal coupling triggers exotic decay; See 1312.4992 for a complete review.

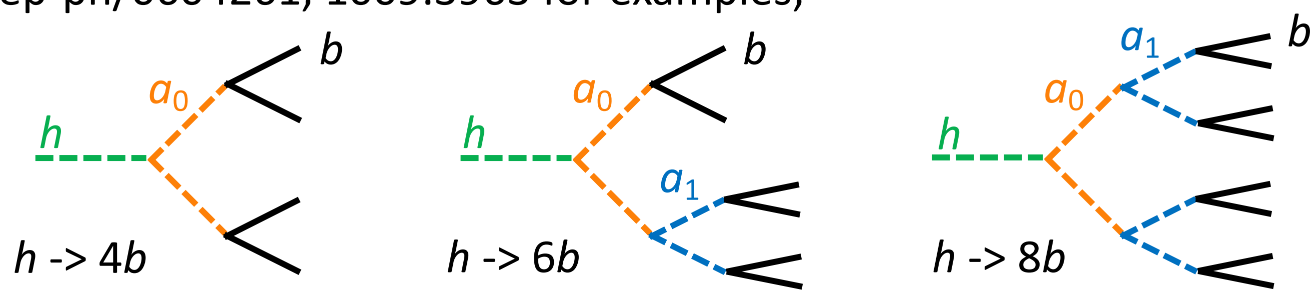
- The SM Higgs width is extremely small: 4.07 MeV;
- Even a small portal coupling can have considerable **exotic decay** branching ratio;

- For multiple BSM light particles, we might have cascade decays.

- # Higgs exotic decay to multiple b-jets

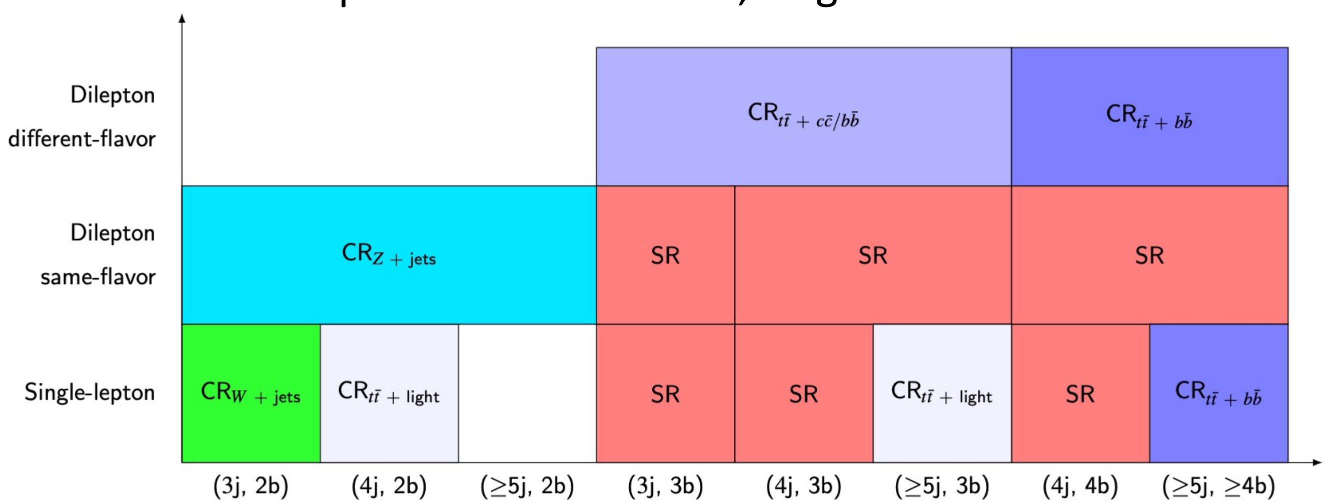
The cascade decays via light scalars

- Might exist in models with dark sector consists of multiple dark scalars, see hep-ph/0604261, 1009.3963 for examples;



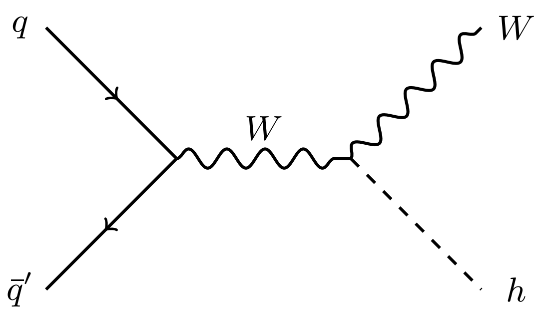
ATLAS's search for the 4*b* final state [1806.07355]

- Use the *Wh* and *Zh* production channels; target on multi-*b* final state



• This talk

The *Wh* channel with boosted region



→ 1 lepton

Use the machine learning method to detect the Higgs jet

→ 1 fat jet containing 4, 6 or 8 *b*-jets

The selection cuts

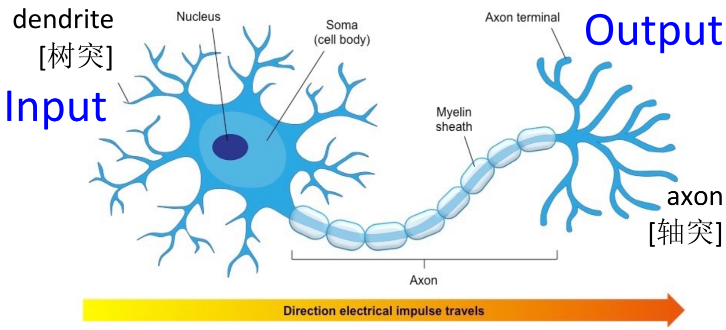
- The **fat jet**: anti-k_t, 200 GeV < p_T < 500 GeV, 100 GeV < m_J < 150 GeV;
- The signal cross sections have assume a 100% branching ratio;

Cross section	Higgs exotic decay			SM		
[Unit: fb]	$\ell^\pm\nu 4b$	$\ell^\pm\nu 6b$	$\ell^\pm\nu 8b$	$W^\pm$ + jets	$t\bar{t}$	$W^\pm h$
Boosted $\ell^\pm$	8.21	7.66	7.04	2.53×10^5	6.21×10^3	5.48
fat-jet	7.01	6.56	6.03	2.01×10^5	4.95×10^3	4.66
b-veto	6.17	5.80	5.35	1.96×10^5	2.17×10^3	4.07
Mass window	3.34	3.19	2.99	5.66×10^3	400	2.08
Efficiency	1.37%	1.36%	1.34%	0.96%	0.25%	1.31%

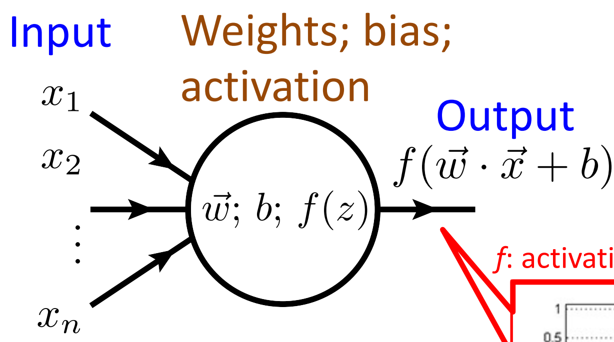
- Our goal: distinguish the Higgs jet from background QCD jets

- # Artificial neural networks

A neuron

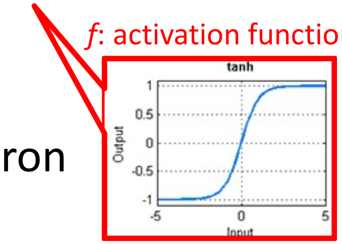


A boiological neuron

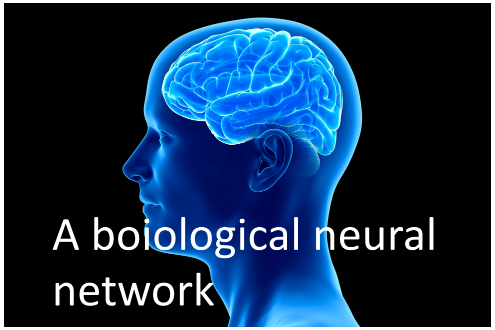


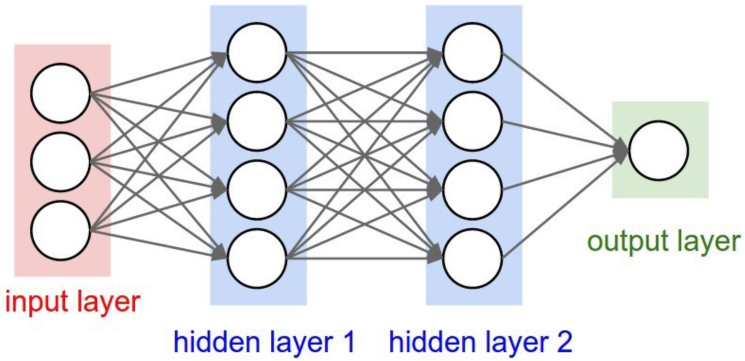
An artificial neuron

f: activation function



A neural network





An artificial neural network

Our wish

- Input the kinematic information of the fat-jet constituents;
- Output the probability of a given jet to be Higgs candidate or QCD jet.

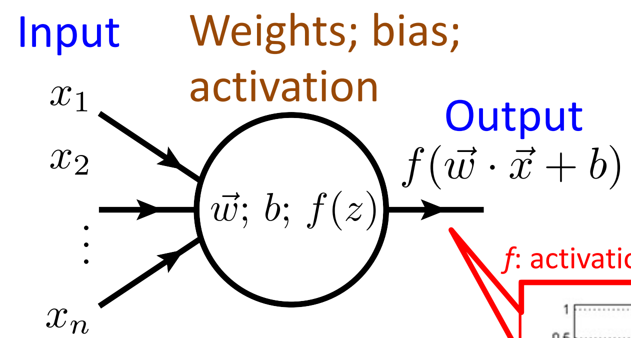
• Training artificial neural networks

The loss function

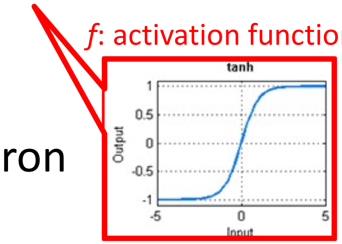
- Label a Higgs jet with *y* = 1, a QCD jet with *y* = 0;
- For a given jet input, the output of the network is a real number 0 < *r* < 1;
- For a given dataset with *N* mixing signal and background jets:

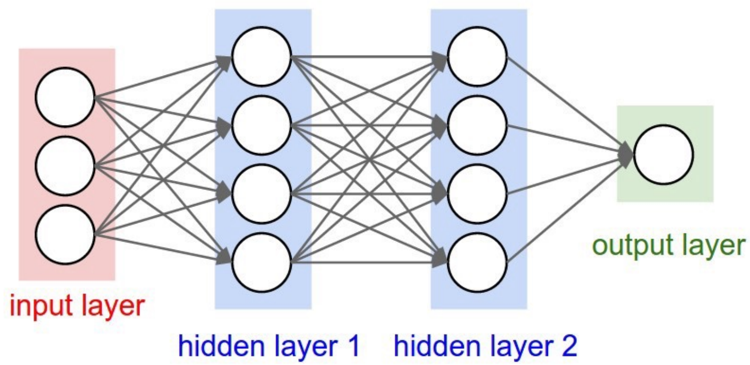
$$L(w,b) \equiv \frac{1}{2N}\sum_{i=1}^{N}(y_i - r_i)^2$$

- The smaller *L*(*w*, *b*) is, the better the network works.



An artificial neuron





An artificial neural network

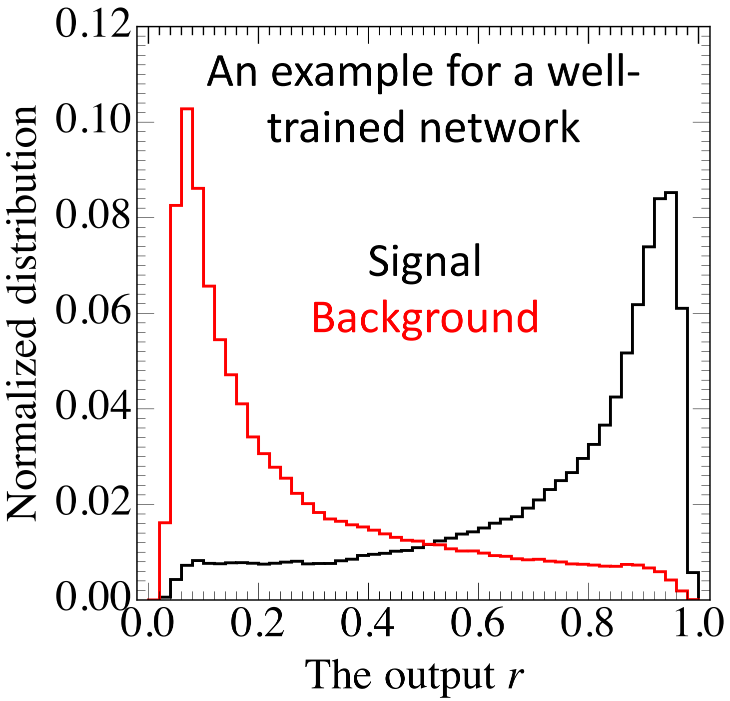
Training artificial neural networks

The loss function

- Label a Higgs jet with $y = 1$, a QCD jet with $y = 0$;
- For a given jet input, the output of the network is a real number $0 < r < 1$;
- For a given dataset with N mixing signal and background jets:

$$L(w,b) \equiv \frac{1}{2N}\sum_{i=1}^{N}(y_i - r_i)^2$$

- The smaller $L(w, b)$ is, the better the network works.



Training the network

- Varying (w, b) self-adaptively to find the minimum of $L(w, b)$.
- Millions or even billions of parameters.
- Train the network on MC events, and apply it to the real data.
- In pheno work: separate the MC events to two parts for training and testing

- # Applying to Higgs exotic decay

Variety based on our scenario

- One signal channel at a time: 3 neural outputs, representing signal and two backgrounds (W + jets, ttbar); $r_0 + r_1 + r_2 = 1$, probability interpretation.

Cross section	Higgs exotic decay			SM		
[Unit: fb]	$\ell^\pm\nu 4b$	$\ell^\pm\nu 6b$	$\ell^\pm\nu 8b$	$W^\pm$ + jets	$t\bar{t}$	$W^\pm h$
Boosted $\ell^\pm$	8.21	7.66	7.04	2.53×10^5	6.21×10^3	5.48
fat-jet	7.01	6.56	6.03	2.01×10^5	4.95×10^3	4.66
b-veto	6.17	5.80	5.35	1.96×10^5	2.17×10^3	4.07
Mass window	3.34	3.19	2.99	5.66×10^3	400	2.08
Efficiency	1.37%	1.36%	1.34%	0.96%	0.25%	1.31%

- We have different network structures to input the information of a jet.

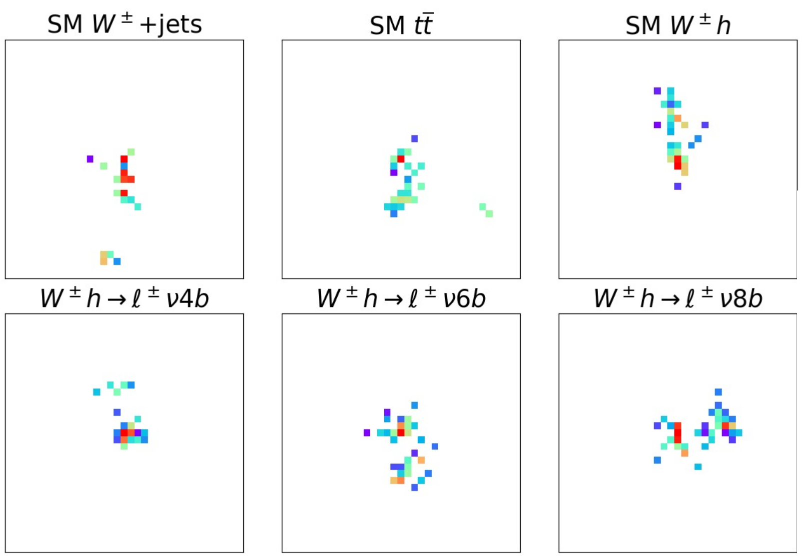
Three ways of inputting jet kinematic information

1. Computer vision: a jet as an image;
2. Natural language processing: a jet as a sentence;
3. Particle flow network: letting the neural network to find the best jet observables itself.

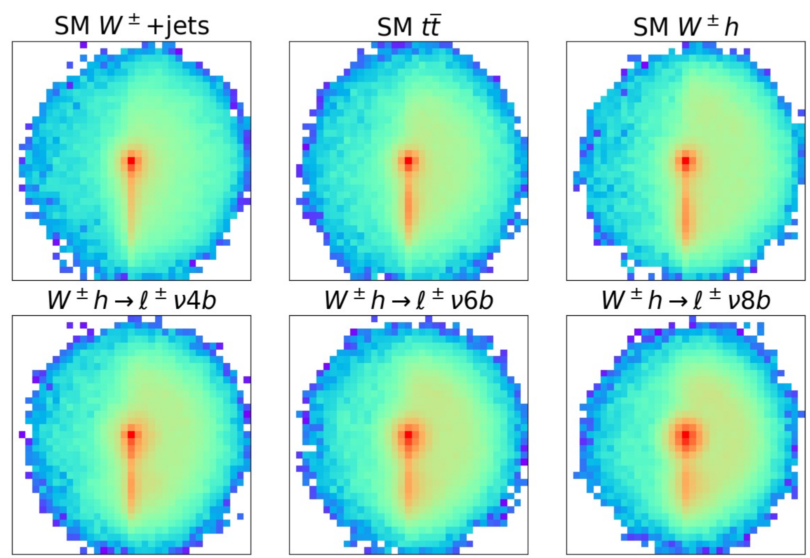
• Computer vision

A jet is an image

- Expand the constituents of the fat jet in the η-φ plane to form a 35x35 pixels jet image with the granularity of 0.1x0.1;
- Intensity of a pixel: p_T sum of the deposited particles; [Cogan *et al*, 1407.5675]
- Pre-processing: translation, rotation and reflection.



A single jet image is rather sparse

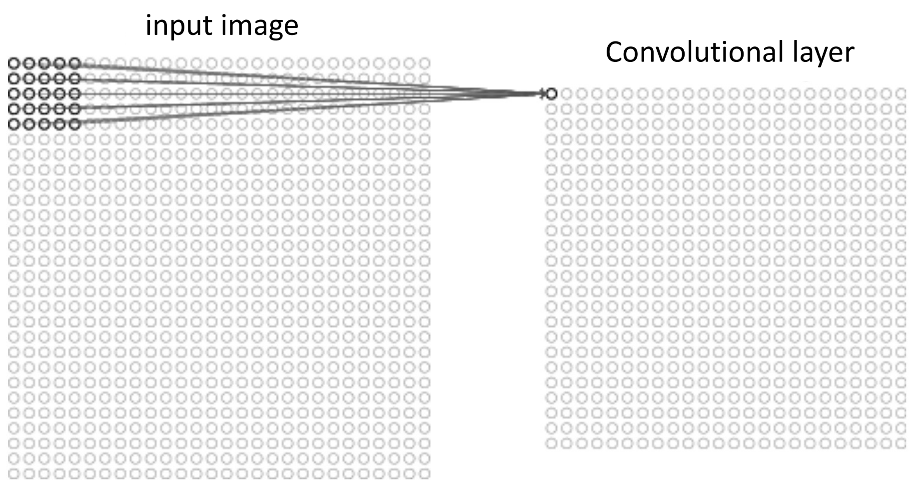


Average of 10,000 jet images

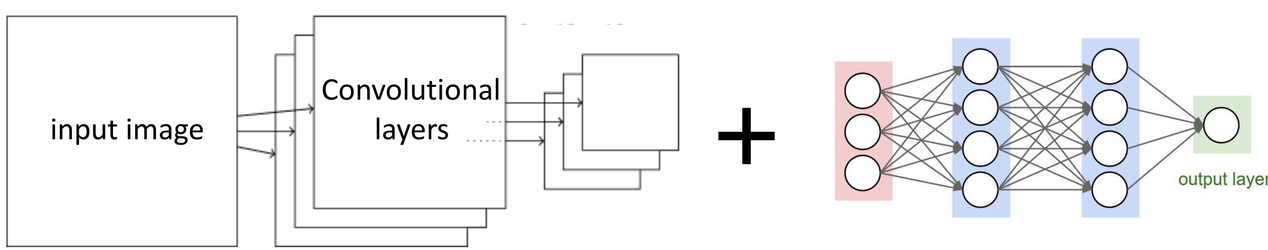
• Computer vision

Convolutional neural network (CNN)

- Mapping the image into the convolutional layers;



- Mapping into multiple channels, and finally to fully connected layers



- A 2-dim image is mapped into a triple vector (r_0 r_1 r_2) for probability interpretation.

• Natural language processing

A jet is a sentence

- Jet clustering history is a binary tree;
- Momenta are words, the jet is a sentence; [Louppe *et al*, 1702.00748]

Recursive neural net (RecNN)

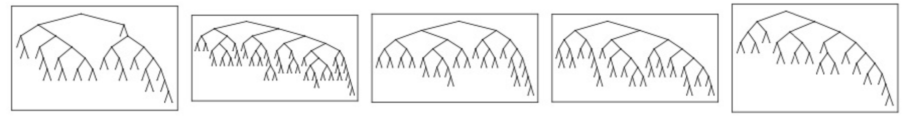
- Attach a vector ***u*** to each node (with momentum ***v***)

$$\mathbf{u}_k = \sigma(W_u \mathbf{v}_k + b_u),$$

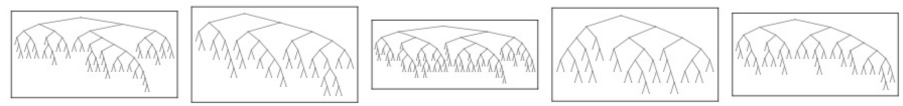
- Define the embedded vector recursively

$$\mathbf{h}_k = \begin{cases} \mathbf{u}_k, & \text{if } k \text{ is a leaf;} \\ \sigma\left(W_h \begin{bmatrix} \mathbf{h}_{k_L} \\ \mathbf{h}_{k_R} \\ \mathbf{u}_k \end{bmatrix} + b_h\right), & \text{otherwise,} \end{cases}$$

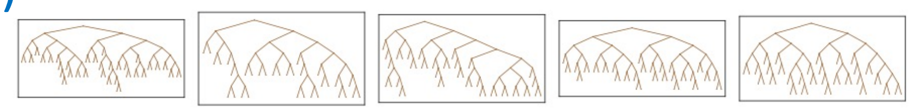
- The ***h*** of the tree root is the embedded vector.



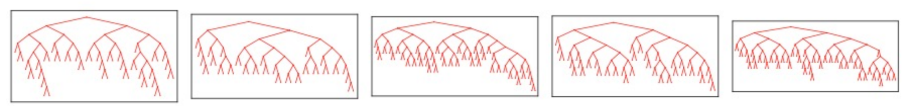
(a) The clustering history of fat-jets from SM $W^\pm$ + jets.



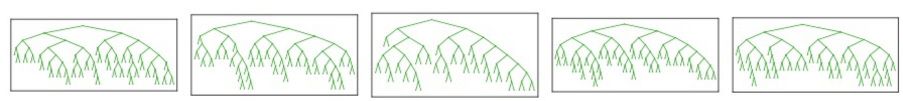
(b) The clustering history of fat-jets from SM $t\bar{t}$.



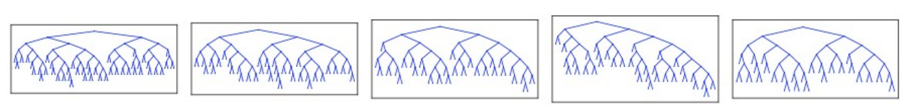
(c) The clustering history of fat-jets from SM $W^\pm h$.



(d) The clustering history of fat-jets from $\ell^\pm\nu 4b$.



(e) The clustering history of fat-jets from $\ell^\pm\nu 6b$.



(f) The clustering history of fat-jets from $\ell^\pm\nu 8b$.

• Natural language processing

A jet is a sentence

- Jet clustering history is a binary tree;
- Momenta are words, the jet is a sentence; [Louppe *et al*, 1702.00748]

Recursive neural net (RecNN)

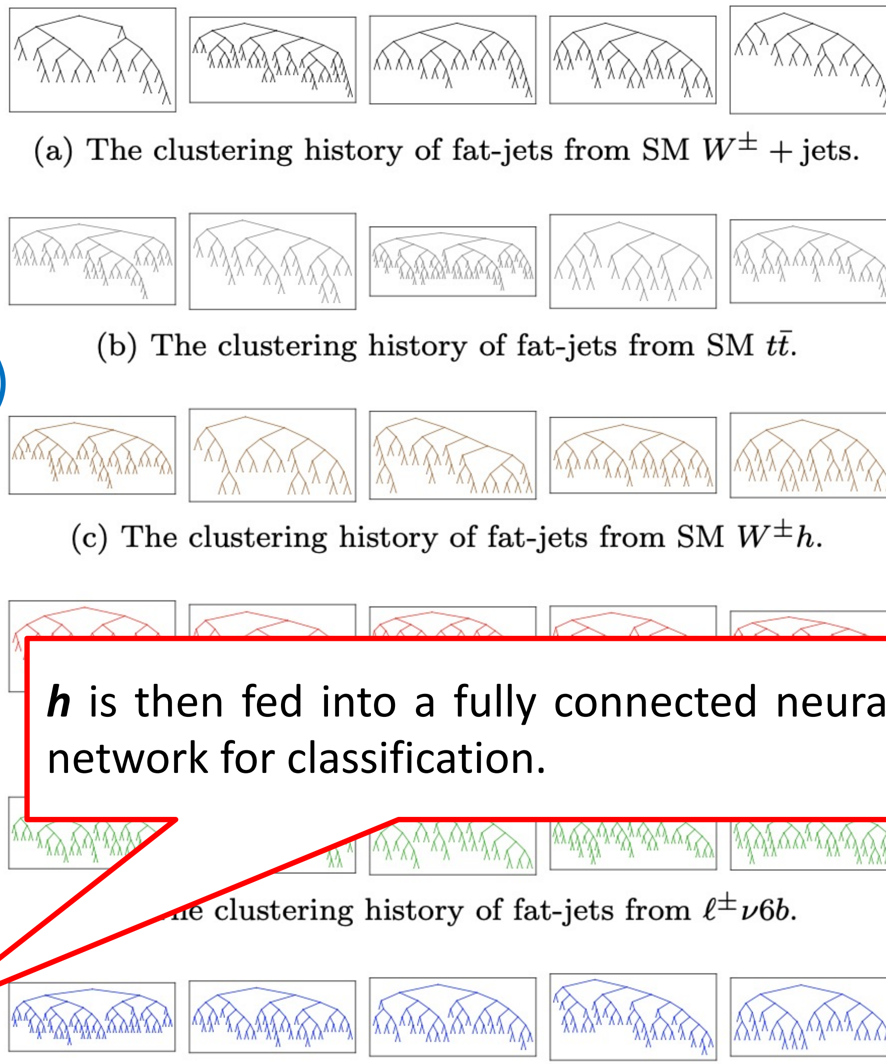
- Attach a vector ***u*** to each node (with momentum ***v***)

$$\mathbf{u}_k = \sigma(W_u \mathbf{v}_k + b_u),$$

- Define the embedded vector recursively

$$\mathbf{h}_k = \begin{cases} \mathbf{u}_k, & \text{if } k \text{ is a leaf;} \\ \sigma\left(W_h \begin{bmatrix} \mathbf{h}_{k_L} \\ \mathbf{h}_{k_R} \\ \mathbf{u}_k \end{bmatrix} + b_h \right), & \text{otherwise,} \end{cases}$$

- The ***h*** of the tree root is the embedded vector.



(a) The clustering history of fat-jets from SM $W^{\pm}$ + jets.

(b) The clustering history of fat-jets from SM $t\bar{t}$.

(c) The clustering history of fat-jets from SM $W^{\pm}h$.

...he clustering history of fat-jets from $\ell^{\pm}\nu 6b$.

(f) The clustering history of fat-jets from $\ell^{\pm}\nu 8b$.

h is then fed into a fully connected neural network for classification.

• Particle flow network (PFN)

The high-level observables of a jet

- The observables of a jet constituent (single particle) $p = \{\xi_1, \xi_2, \cdots, \xi_d\}$
- The high-level observables can be generally written as $\mathcal{O} = F\left(\sum_i \Phi(p_i)\right)$
- Here F and Φ are functions depending on the observable O.

Observable $\mathcal{O}$		**Map** Φ	**Function** F
Mass	m	p^μ	$F(x^\mu) = \sqrt{x^\mu x_\mu}$
Multiplicity	M	1	$F(x) = x$
Track Mass	m_{track}	$p^\mu \mathbb{I}_{\text{track}}$	$F(x^\mu) = \sqrt{x^\mu x_\mu}$
Track Multiplicity	M_{track}	$\mathbb{I}_{\text{track}}$	$F(x) = x$
Jet Charge [72]	$\mathcal{Q}_\kappa$	$(p_T, Q\, p_T^\kappa)$	$F(x,y) = y/x^\kappa$
Eventropy [74]	$z \ln z$	$(p_T, p_T \ln p_T)$	$F(x,y) = y/x - \ln x$
Momentum Dispersion [93]	p_T^D	(p_T, p_T^2)	$F(x,y) = \sqrt{y/x^2}$
C parameter [94]	C	$(\lvert\vec{p}\rvert, \vec{p} \otimes \vec{p}/\lvert\vec{p}\rvert)$	$F(x,Y) = \frac{3}{2x^2}[(\operatorname{Tr} Y)^2 - \operatorname{Tr} Y^2]$

Table from 1810.05165

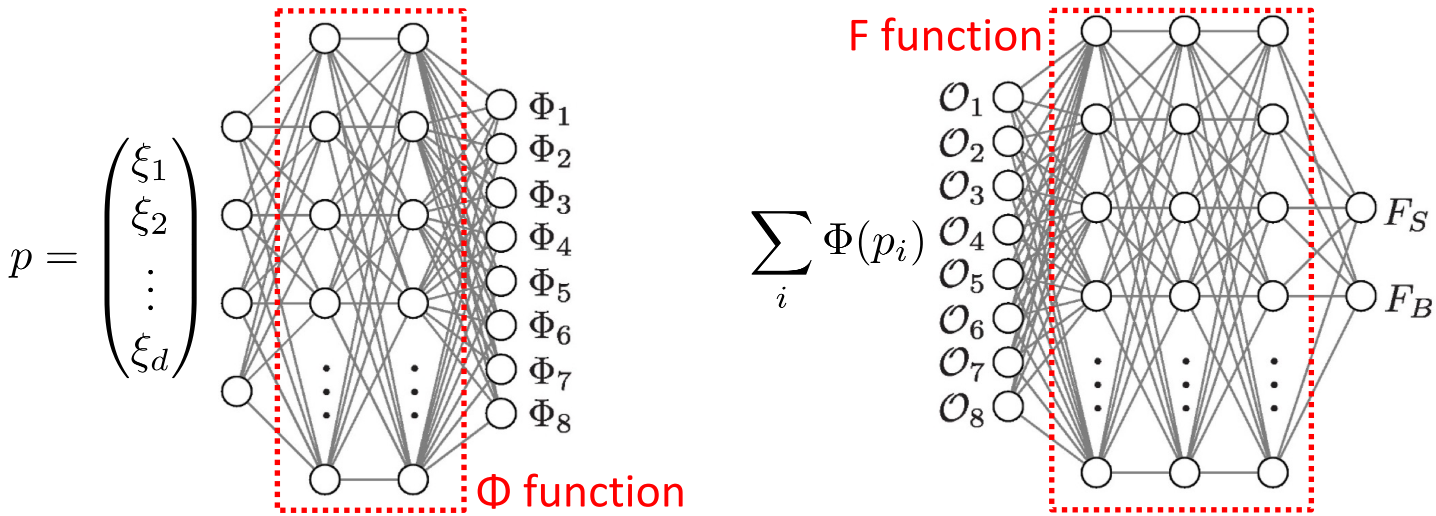
Let the neural network find the correct observable itself!

- Use neural layers to represent F and Φ;
- Train the network to find the best F and Φ. [Komiske *et al*, 1810.05165]

• Particle flow network (PFN)

The high-level observables of a jet

- The observables of a jet constituent (single particle) $p = \{\xi_1, \xi_2, \cdots, \xi_d\}$
- The high-level observables can be generally written as $\mathcal{O} = F\left(\sum_i \Phi(p_i)\right)$
- Here F and Φ are functions depending on the observable O.



Let the neural network find the correct observable itself!

- Use neural layers to represent F and Φ;
- Train the network to find the best F and Φ.[Komiske *et al*, 1810.05165]

• The results

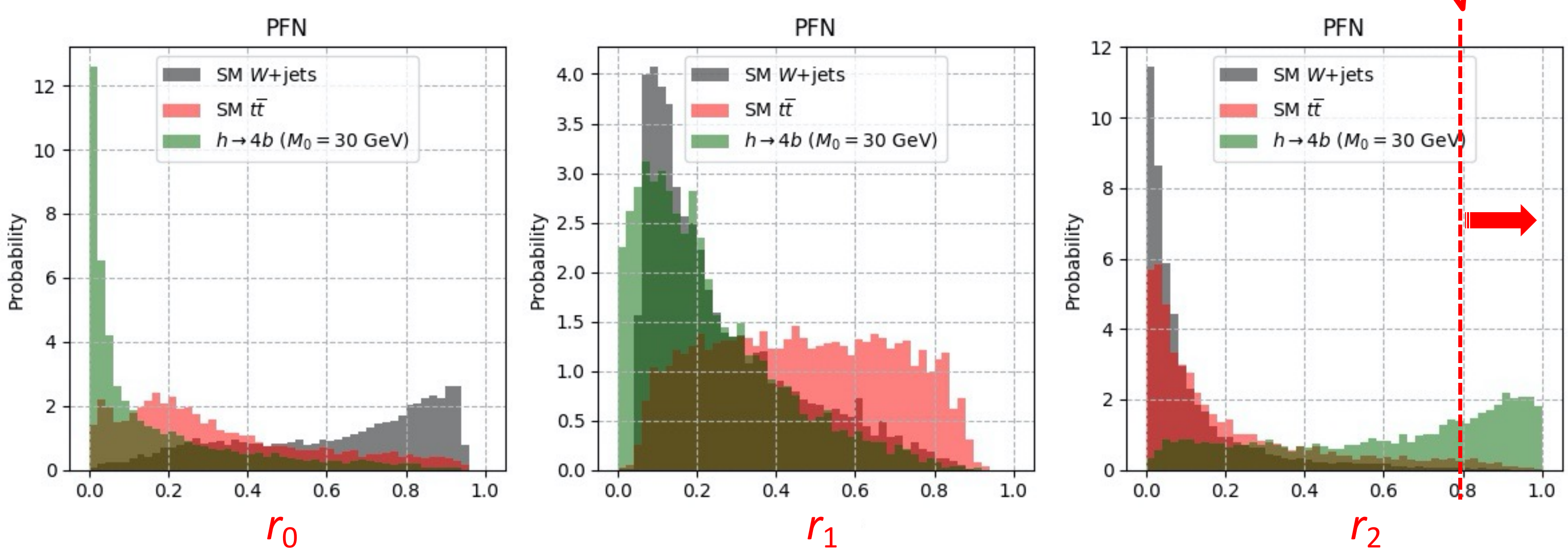
Summary for the networks

1. Computer vision: CNN;
2. Natural language processing: RecNN;
3. Particle flow network: PFN.

Make a cut on the neuron output and get the signal & (weighted) background efficiencies

The training results

- Use particle flow network (PFN) as an illustration;
- Signals has a peak for the r_2 neuron around 1;

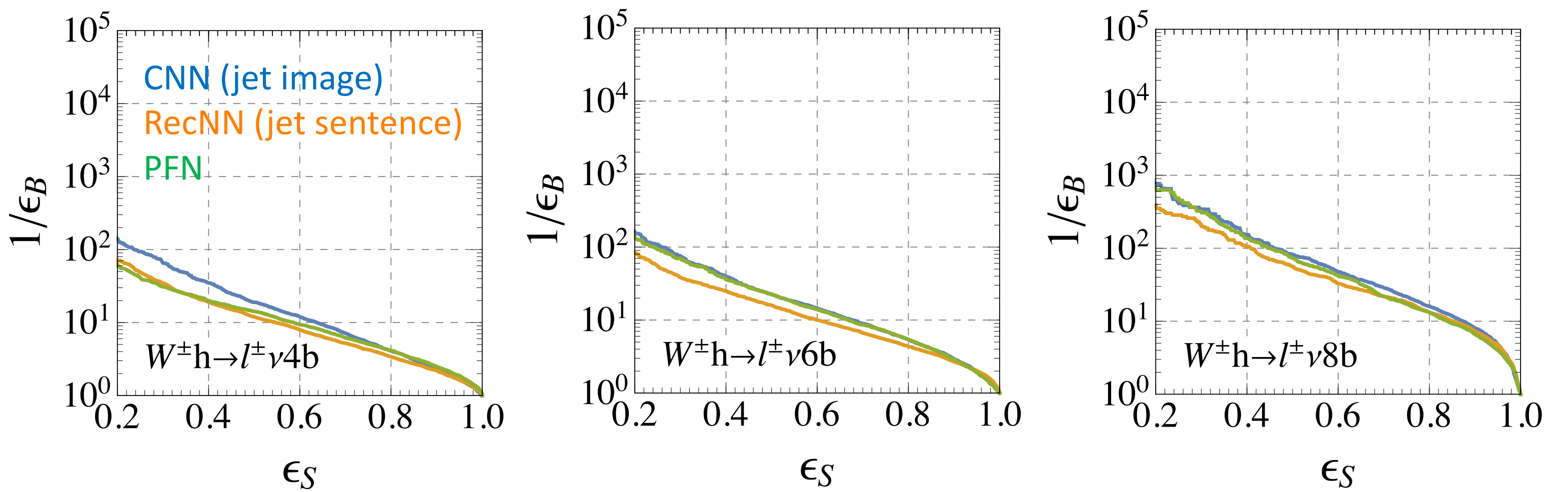


- Based on output we can construct the (weighted) efficiency curves.

• The results

Signal efficiencies versus background rejections

- Benchmark: M(a_0) = 30 GeV; M(a_1) = 12 GeV



Features

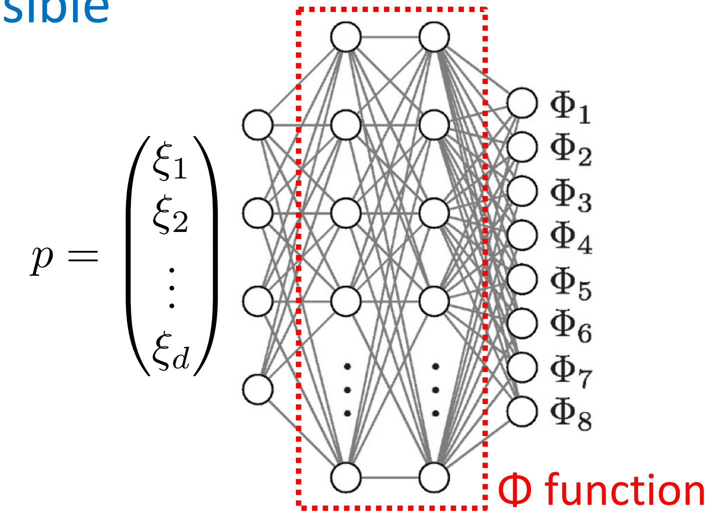
- For a given signal channel, all three neural networks have similar performance, implying the kinematic information has been efficiently learned.

- Increasing the *b*-multiplicity enhances the performance.

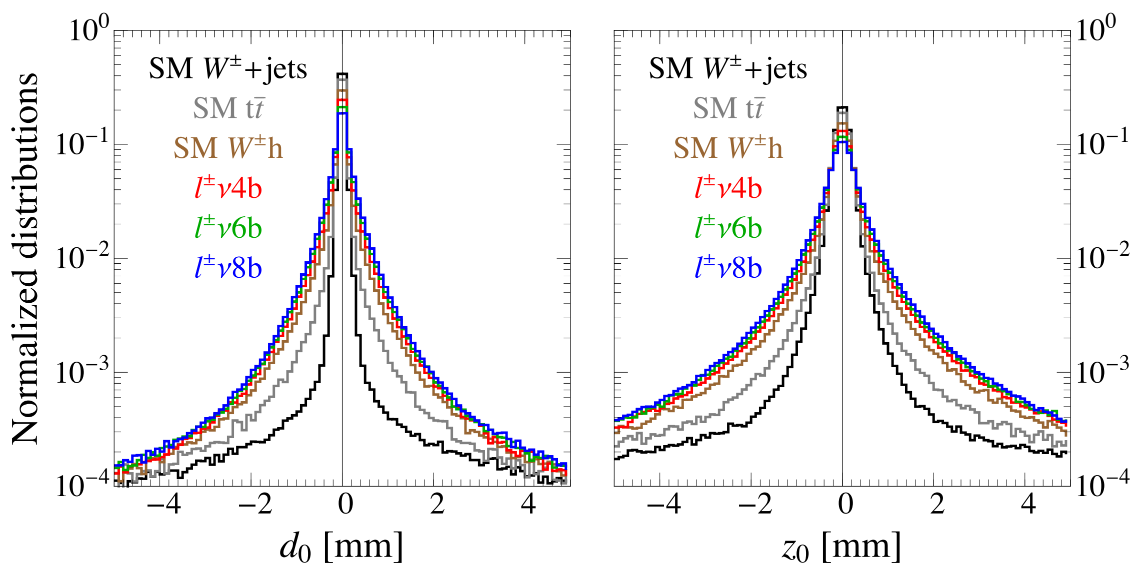
• Extending the PFN

The particle flow neural net is extensible

- The single particle observables can be extended to include particle ID and impact parameters from tracker.



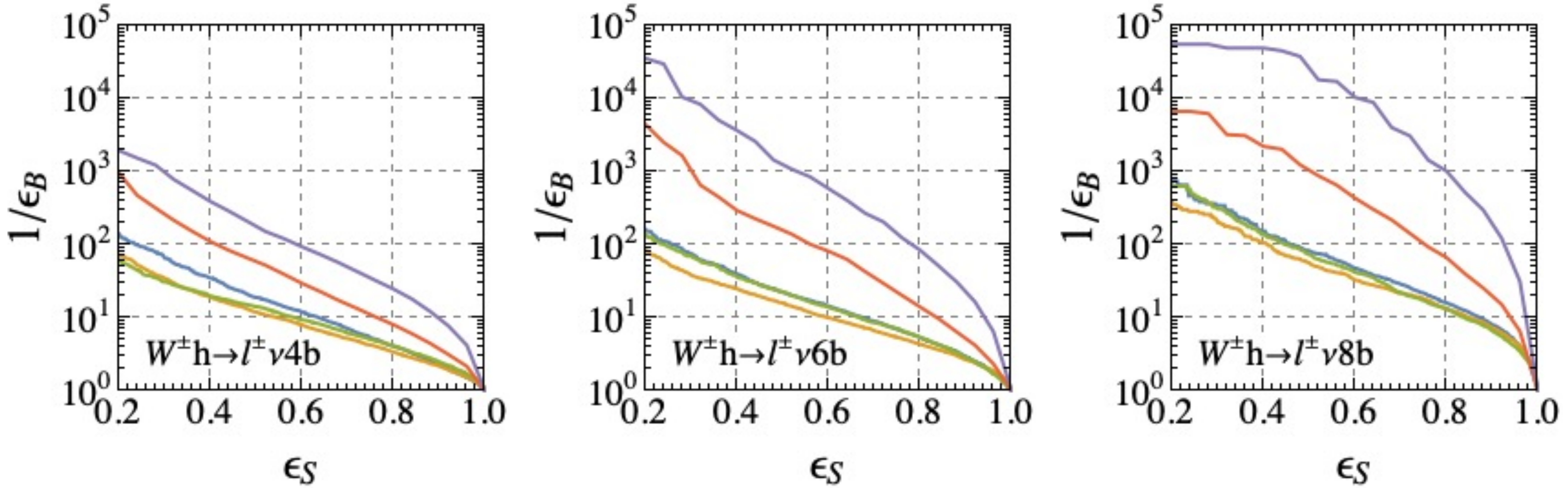
- The impact parameters are especially helpful for a b-rich final state.



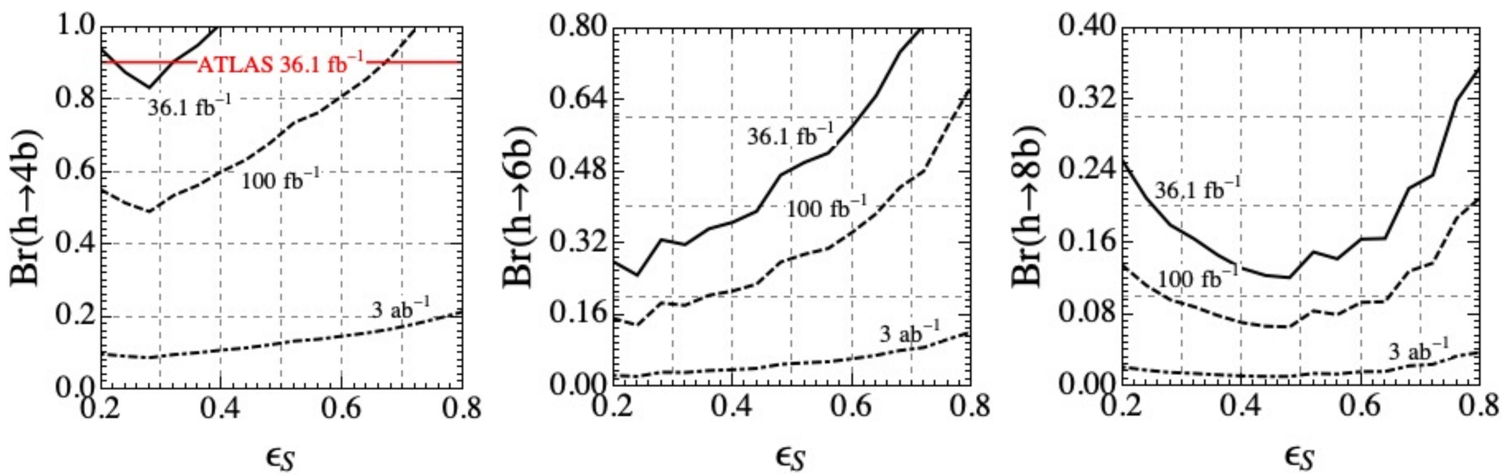
- # Extending the PFN

CNN (jet image)
RecNN (jet sentence)
PFN (momentum only)
PFN (with PID information)
PFN (with impact parameters)

The extended PFN



The branching ratio limits

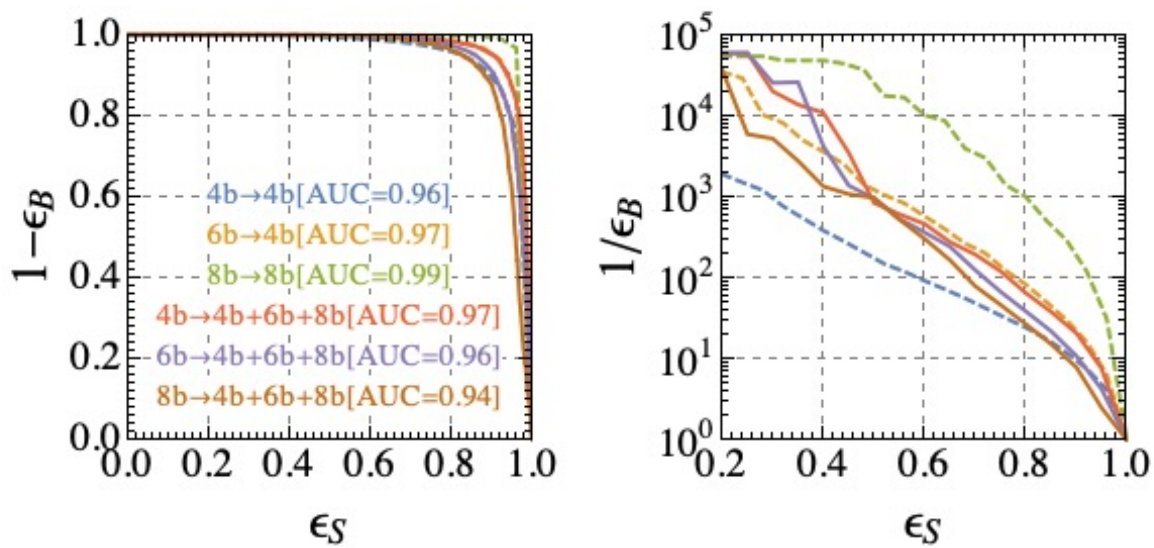


• Possibility of a h -> n b tagger

The universality

- We can apply a PFN trained on one channel to another channel;
- Results are insensitive to the b-multiplicity;
- We might be able to have a universal tagger for h to n b-jets.

Classification accuracies		$M_0 = 30$ GeV, $M_1 = 12$ GeV		
Tested on \ Trained on		$\ell^\pm\nu 4b$	$\ell^\pm\nu 6b$	$\ell^\pm\nu 8b$
$M_0 = 30$ GeV, $M_1 = 12$ GeV	$\ell^\pm\nu 4b$	73.1%	69.7%	68.1%
	$\ell^\pm\nu 6b$	77.0%	76.5%	74.9%
	$\ell^\pm\nu 8b$	79.4%	79.9%	79.4%
	$4b + 6b + 8b$	76.4%	74.7%	73.6%



• Conclusion

We have trained different neural networks to learn the fat jets from

1. *Wh* (with *h* decaying to 4*b*, 6*b* or 8*b*);

2. *W* + jets;

3. *tt*bar,

To distinguish signal from backgrounds.

What we learn from machine learning:

- All of CNN, RecNN and PFN can learn the kinematic information of jet constituents efficiently and yield similar performances;
- PFN can be extended to include more information such as PID and tracks, thus work much better;
- A universal tagger for *h* to *n* *b*-jet is possible.

Thank you!

Nanjing Museum
Photographed on April 14, 2018

• Backup

To suppress the background, we require the final state to have exactly one charged lepton with

$$p_T^\ell > 25 \text{ GeV}, \quad |\eta^\ell| < 2.5,$$
$$p_T^{\ell+\not{E}_T} > 200 \text{ GeV}, \quad M_T < 100 \text{ GeV}, \tag{1}$$

where the transverse mass is defined as

$$M_T = \sqrt{2p_T^\ell \not{E}_T (1 - \cos\Delta\phi)}, \tag{2}$$

with $\Delta\phi$ being the azimuthal angle difference between $\ell^\pm$ and $\not{E}_T$. We also demand at least one fat-jet reconstructed by the anti-k_t algorithm with $\Delta R = 1.5$ and

$$200 \text{ GeV} < p_T^J < 500 \text{ GeV}, \quad |\eta^J| < 2.5 \,. \tag{3}$$

The fat-jets are trimmed by $R_{\rm cut} = 0.3$ and $f_{\rm cut} = 0.05$ [36]. Next, the small-R jets are clustered using anti-k_t algorithm with $R = 0.4$, and b-tagged ones within

$$p_T^b > 25 \text{ GeV}, \quad |\eta^b| < 2.5 \,, \tag{4}$$

are vetoed to suppress the $t\bar{t}$ background. Finally, we require the mass of the leading fat-jet to be in the Higgs mass window

$$100 \text{ GeV} < m_J < 150 \text{ GeV}, \tag{5}$$

- Backup

Classification accuracies		SM	$M_0 = 30$ GeV $M_1 = 12$ GeV		
Tested on \ Trained on		$h \to b\bar{b}$	$\ell^\pm \nu 4b$	$\ell^\pm \nu 6b$	$\ell^\pm \nu 8b$
SM	$h \to b\bar{b}$	67.1%	61.4%	58.1%	56.5%
$M_0 = 30$ GeV $M_1 = 12$ GeV	$\ell^\pm \nu 4b$	69.3%	73.1%	69.7%	68.1%
	$\ell^\pm \nu 6b$	72.3%	77.0%	76.5%	74.9%
	$\ell^\pm \nu 8b$	74.4%	79.4%	79.9%	79.4%
	$4b + 6b + 8b$	–	76.4%	74.7%	73.6%

Classification accuracies		SM	$M_0 = 50$ GeV $M_1 = 20$ GeV		
Tested on \ Trained on		$h \to b\bar{b}$	$\ell^\pm \nu 4b$	$\ell^\pm \nu 6b$	$\ell^\pm \nu 8b$
SM	$h \to b\bar{b}$	67.1%	56.3%	55.8%	52.3%
$M_0 = 50$ GeV $M_1 = 20$ GeV	$\ell^\pm \nu 4b$	62.4%	72.9%	73.9%	70.7%
	$\ell^\pm \nu 6b$	64.6%	76.8%	77.3%	76.6%
	$\ell^\pm \nu 8b$	66.5%	79.4%	80.2%	80.1%