University of Zurich[UZH]

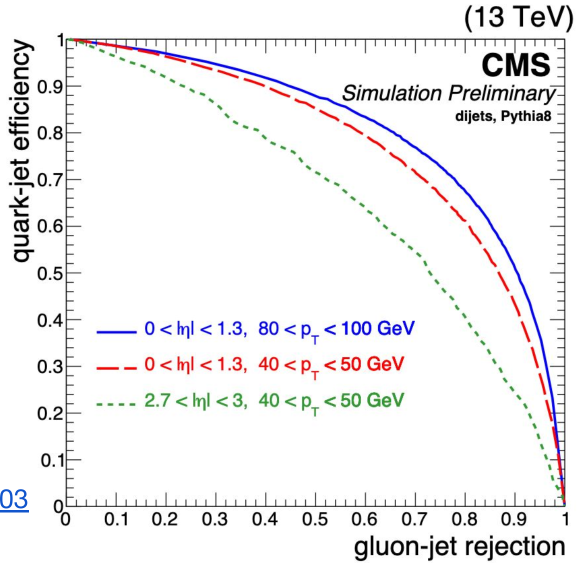
Quark-gluon tagging with point clouds

SNOWMASS EF02
Sep 3, 2020

Florencia Canelli
Vinicius M. Mikuni
Eduardo Ploerer

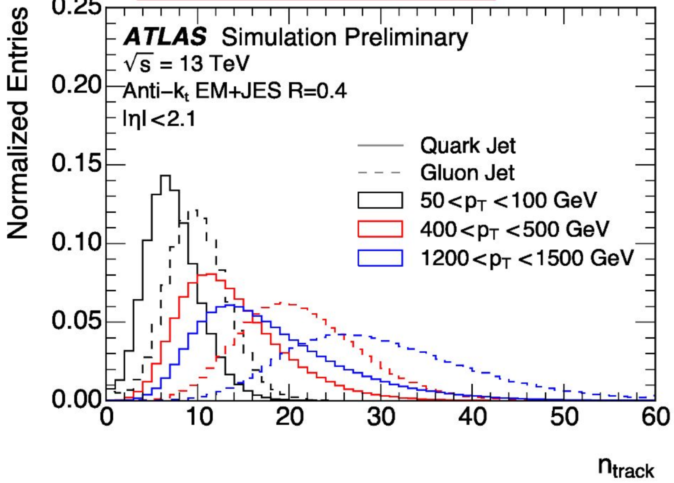
Introduction

- Methods to identify **quark** from **gluon** initiated jets have a long history in HEP
- Most of the current approaches take a small set of observables as a proxy for separation
- How to **efficiently** extract information from particles inside a jet?



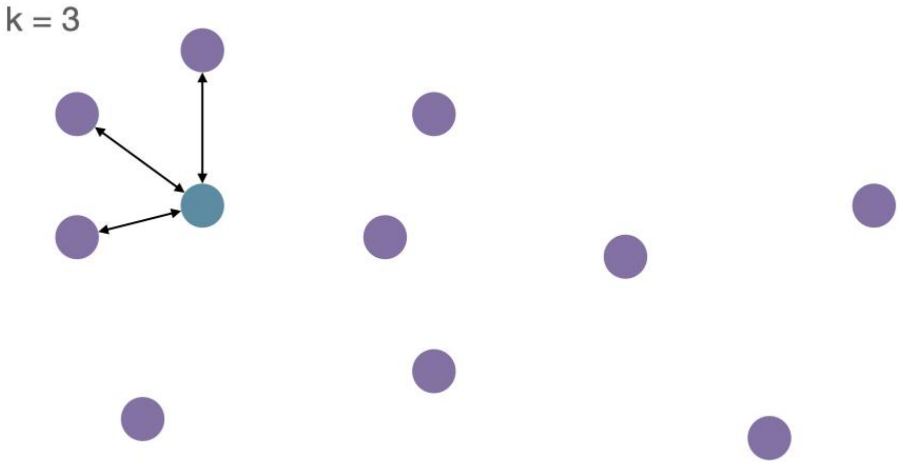
CMS-PAS-JME-16-003

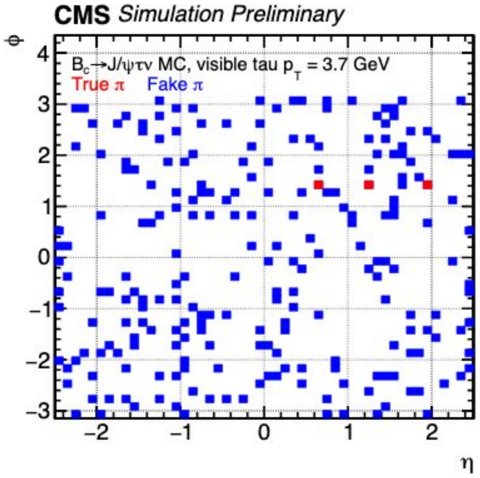
ATL-PHYS-PUB-2017-009



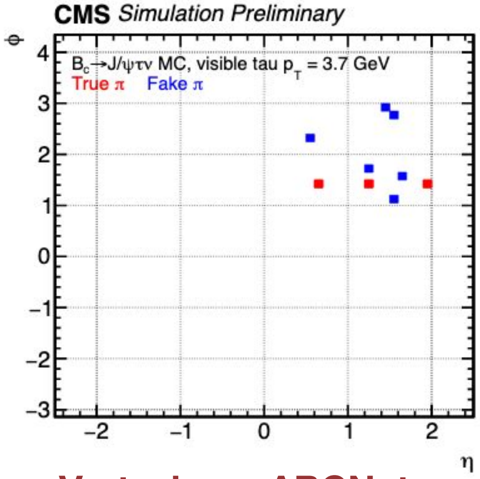
ABCNet: Easy as 1-2-3

- **ABCNet: A**ttention-**b**ased **c**loud **net**work
- Also used for low p_T tau reconstruction in CMS
- **Point clouds**: Permutation invariant set of objects
- Each particle represents a **node** in a graph
- Extract local information by combining the **k-nearest particles**





Initial event



Vertexing + ABCNet

ABCNet: Easy as 1-2-3

A woman is throwing a frisbee in a park.

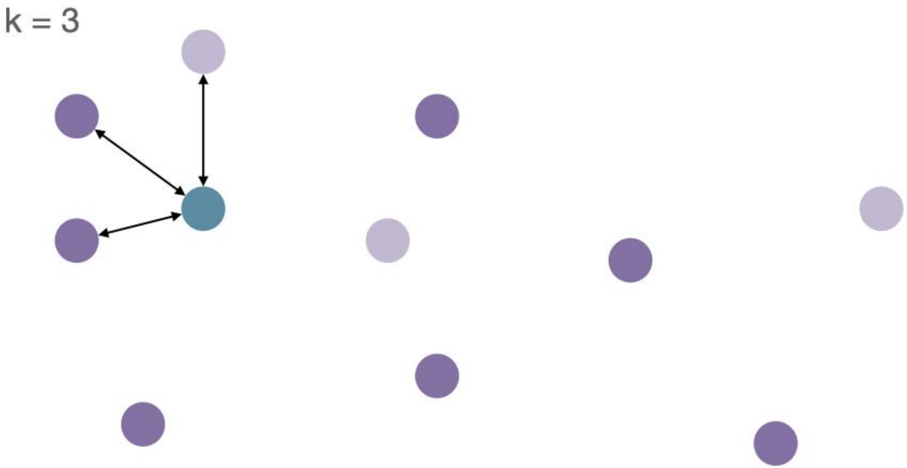
A dog is standing on a hardwood floor.

A stop sign is on a road with a mountain in the background.

A little girl sitting on a bed with a teddy bear.

A group of people sitting on a boat in the water.

A giraffe standing in a forest with trees in the background.

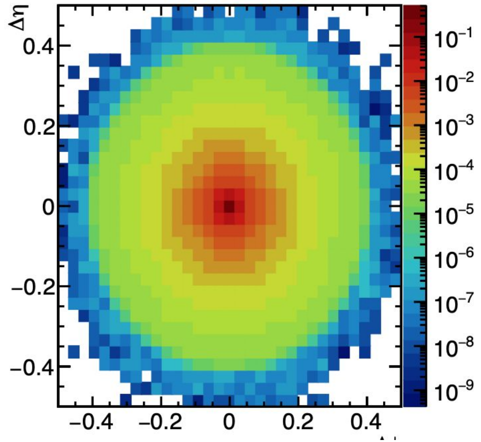


Attention: Let the **method learn** the **relevant parts** for the task at hand (like the **bold text** I'm using in this presentation)

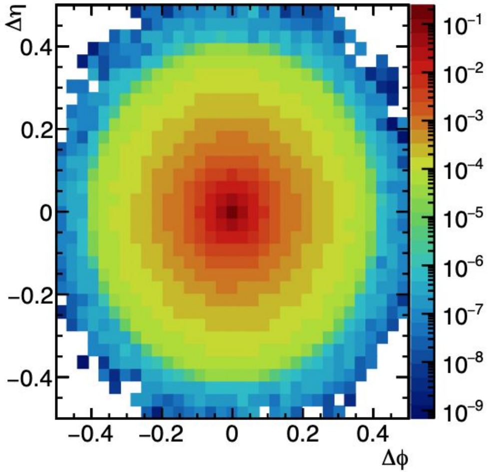
Results: QG separation

- Signal: Z(νν) + **(u,d,s)**
- Background: Z(νν) + **g**
- ak4 jets with **500< pT < 550 GeV** and **|y| < 1.7**
- Consider up to **100 particles**
- Use the same samples from [Energy Flow Networks]
- Input variables per particle:
  - Δη
  - Δϕ
  - Log pT
  - Log E
  - Log (pT/pT(jet))
  - Log (E/E(jet))
  - ΔR to jet axis
  - PID

Quark initiated



Gluon initiated



Results: QG separation

- **ParticleNet**: Point cloud approach, similar to ABCNet
- Improved background rejection with less training parameters

	Acc	AUC	$1/\epsilon_B\ (\epsilon_S = 0.5)$	$1/\epsilon_B\ (\epsilon_S = 0.3)$	Parameters
ResNeXt-50	0.821	0.960	30.9	80.8	1.46M
P-CNN	0.827	0.9002	34.7	91.0	348k
PFN	-	0.9005	34.7±0.4	-	82k
ParticleNet-Lite	0.835	0.9079	37.1	94.5	**26k**
ParticleNet	**0.840**	0.9116	39.8±0.2	98.6±1.3	366k
ABCNet	**0.840**	**0.9126**	**42.6±0.4**	**118.4±1.5**	230k

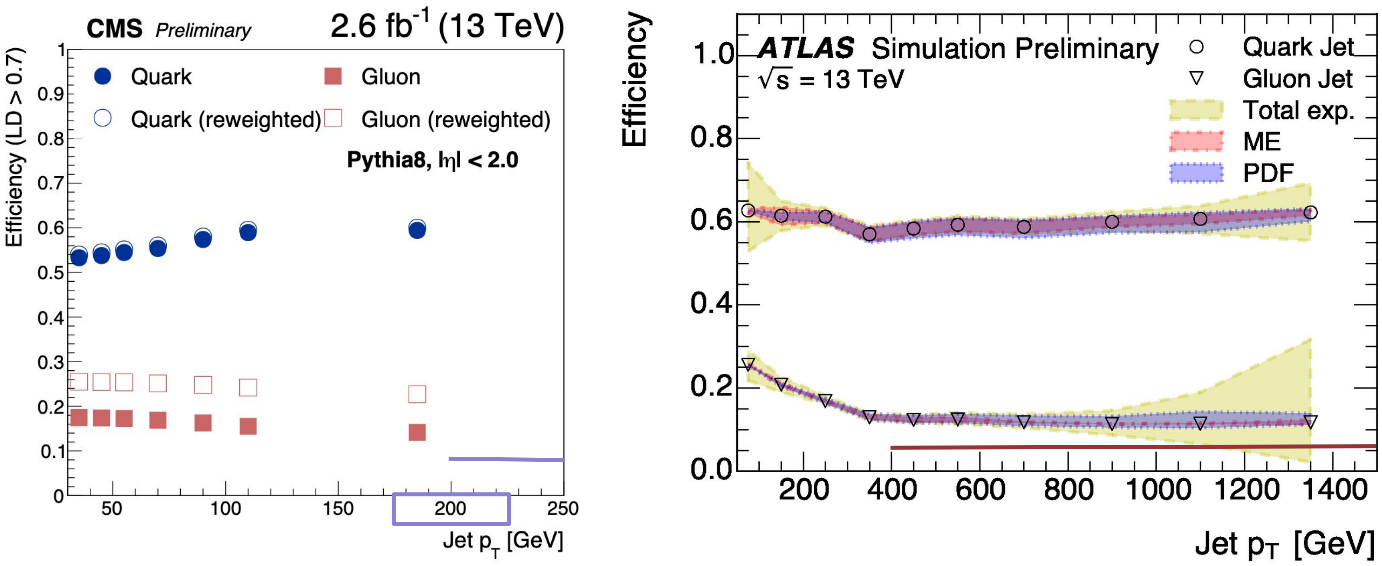
Detector level?

- Previous studies done on particle level
- How does the performance changes after detector effects?
- (Very) Preliminary studies on MC QCD events:
  - **MadGraph5 + Pythia8**
  - Detector simulation: **CMS** detector with **Geant4**
  - Select jets with $\mathbf{p_T > 200}$ **GeV** and $\mathbf{|\eta| < 2.4}$
- Same input features as before
- **Main difference**: Instead of gluon vs. quarks inclusively, change the problem to a multi-class classification problem
- **4 possible categories**:
  - Up
  - Down
  - Strange
  - Gluon
- Train on **700k jets**, split **equally** for each category
- **Output**: **Probability** of a jet belonging to a given class

Results: QG separation

P(quark) = P(u) + P(d) + P(s)

- Compare with CMS and ATLAS results
- For a rough comparison, take the point of **60%** quark efficiency for ABCNet
- **Gluon efficiency: 0.08 for [200 , 250] GeV and 0.06 for p_T > 400 GeV**
- Plots not **yet** public available



Results: strange vs. down

P'(s) = P(s)/(P(s) + P(d))

- Same charge particles with only a few differences
- Recently studied using Jet images. See Yuchiro's talk
- Differences:
  - Delphes instead of Geant4
  - Leading hard scattering quark **p_T > 200 GeV** and **$|\eta|$ < 0.05**

	AUC	ACC	R10	R50
Truth Cut1	0.65 (0.68)	0.61 (0.62)	31.9 (32.1)	3.6 (3.9)
Truth BDT3	0.67 (0.68)	0.62 (0.62)	37.3 (37.1)	3.7 (4.0)
Cut1	0.61 (0.63)	0.57 (0.59)	16.4 (17.9)	2.7 (3.0)
Cut1+	0.62 (0.63)	0.58 (0.60)	17.9 (18.8)	2.9 (3.1)
BDT3	0.61 (0.63)	0.59 (0.60)	16.0 (17.1)	2.9 (3.1)
BDT4	0.63 (0.63)	0.60 (0.60)	22.5 (16.6)	3.2 (3.2)
CNN3	0.62 (0.63)	0.59 (0.60)	17.9 (18.4)	3.0 (3.2)
CNN4	0.64 (0.64)	0.60 (0.60)	23.9 (18.8)	3.3 (3.2)
ABCNet	**0.68**	**0.63**	**30.0**	**3.8**

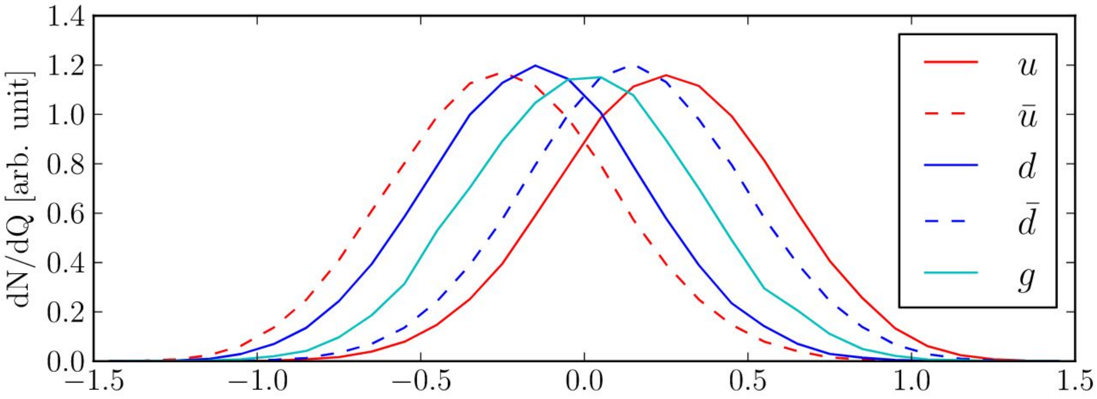
No detector simulation

Results: up vs. down

$P'(u) = P(u)/(P(u) + P(d))$

- Jet charge used for comparison (k=0.5)
- Distribution calculated on the **same samples** used to evaluate ABCNet

$$Q_\kappa^i = \frac{1}{(p_T^{\text{jet}})^\kappa} \sum_{j \in \text{jet}} Q_j (p_T^j)^\kappa$$



	AUC	$1/\varepsilon_{down}$ for $\varepsilon_{up} = 0.1$	$1/\varepsilon_{down}$ for $\varepsilon_{up} = 0.5$
Jet charge	0.74	52.5	6.0
ABCNet	**0.77**	**103.2**	**7.0**

Conclusion

- Graph neural networks are becoming more common in HEP
- In this presentation, a particular approach treating the data as a set of **permutation invariant points** was shown
- **Q/G separation** improves compared to different methods
- **Expand the concept**: **separate** each **light flavour** component
- (Very) preliminary results using **detector level information** are **promising**

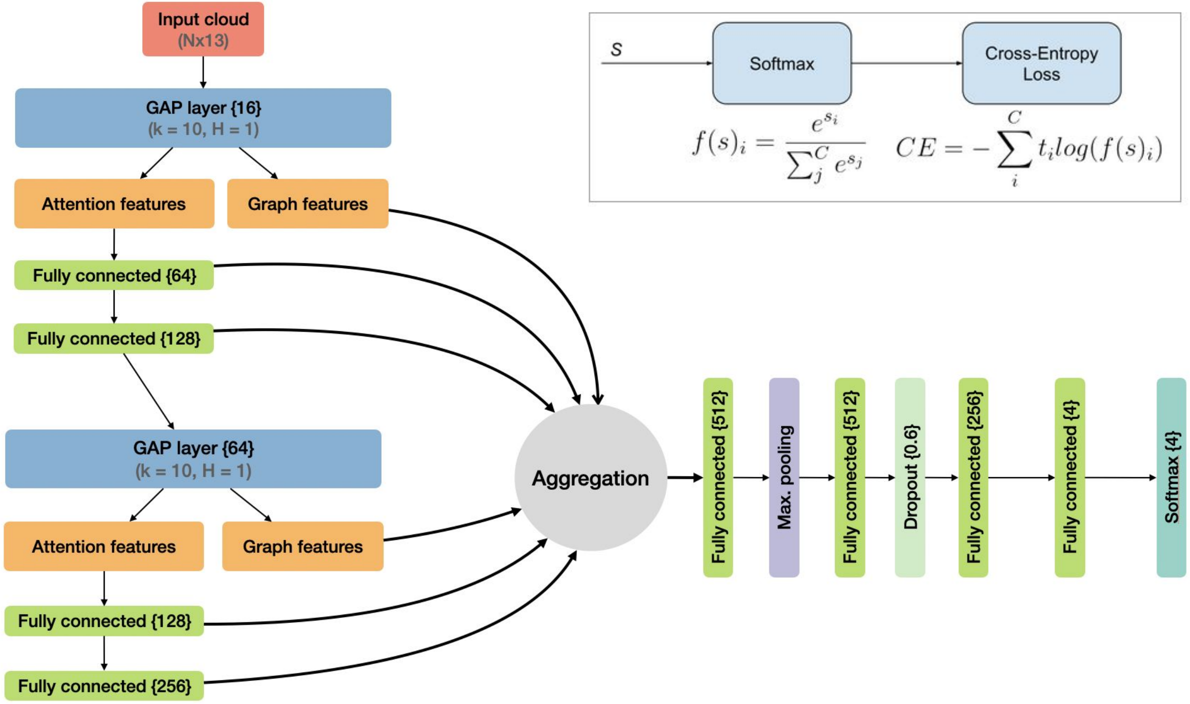
	Accuracy	AUC	$1/\varepsilon_{down}$ for $\varepsilon_{up} = 0.1$	$1/\varepsilon_{down}$ for $\varepsilon_{up} = 0.5$
Up vs. strange	0.77	0.85	229	14
Up vs. gluon	0.81	0.89	614	28
Down vs. gluon	0.80	0.88	397	22
Strange vs. gluon	0.80	0.88	452	24

Thanks!

Backup

Training details

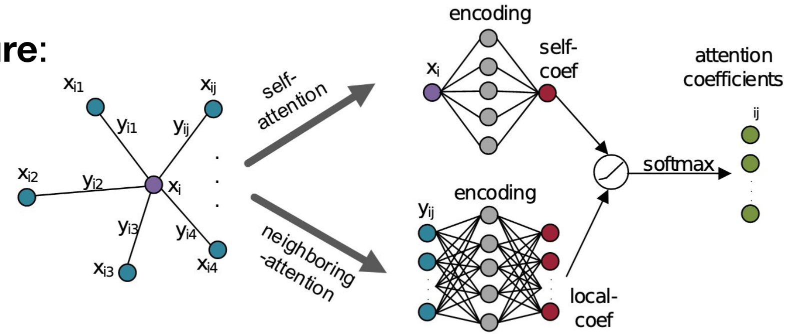
- **Adam** optimizer
- **Mini-batch** size: 64
- **Loss**: Categorical cross-entropy
- **Learning rate**: 1e-2, decreasing by a factor of 10 every 10 epochs
- **Early stopping**: 10 consecutive epochs without improvement



$$f(s)_i = \frac{e^{s_i}}{\sum_j^C e^{s_j}} \qquad CE = -\sum_i^C t_i log(f(s)_i)$$

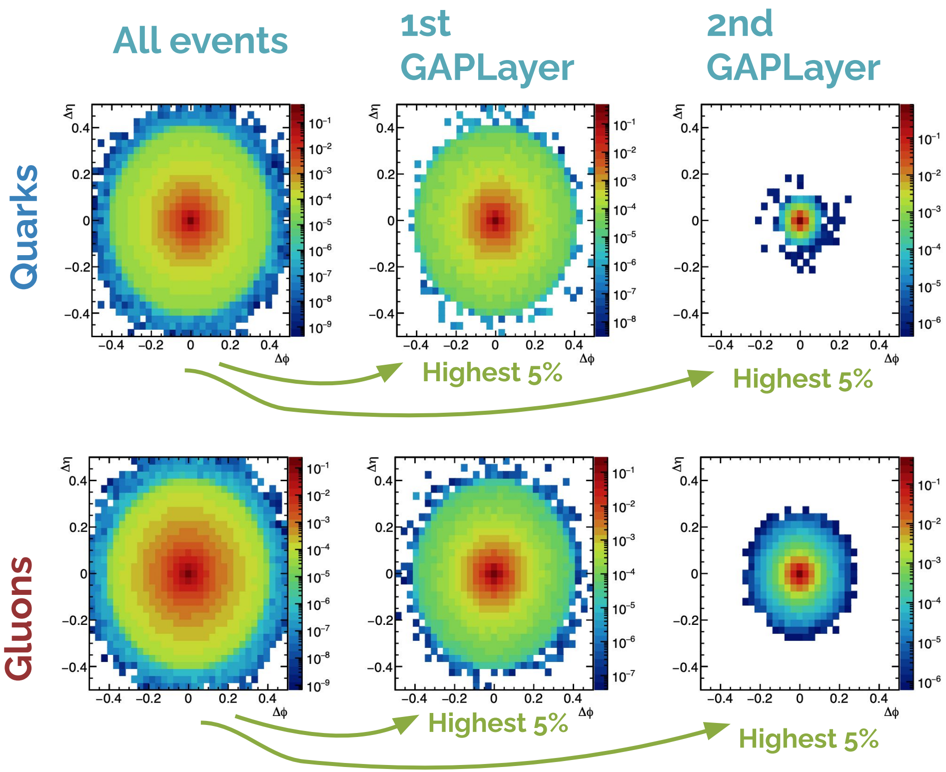
GAPLayer

- The core component of ABCNet are Graph attention pooling layers [GAPLayers]
- **Nodes**: $\mathbf{x}_i$
- **Edge features**: $\mathbf{y}_{ij} = x_i - x_{ij}$
- **Encode** the **nodes** (and **edges**) by passing it to a **2 layers NN** with **output size F** and **1**
- **Self-attention:** encoded nodes $\mathbf{x'}_i$
- **Local-attention:** encoded edge $\mathbf{y'}_{ij}$
- **Merge** all the **coefficients** and pass the result to a **nonlinear function**
  - $\mathbf{c}_{ij} = \text{LeakyRelu}(x'_i + y'_{ij})$
- **Align** $\mathbf{c}_{ij}$ with softmax
  - $\mathbf{c'}_{ij} = S(c_{ij})$
- **Each node** $\mathbf{x}_i$ receives **1 attention feature**:
  - $\mathbf{a}_i = \text{Relu}(\sum_j c'_{ij}\, y'_{ij})$
- GAPLayer **outputs**:
  - **Graph features**: y'_{ij}
  - **Attention features**: a_i



Visualizing important particles

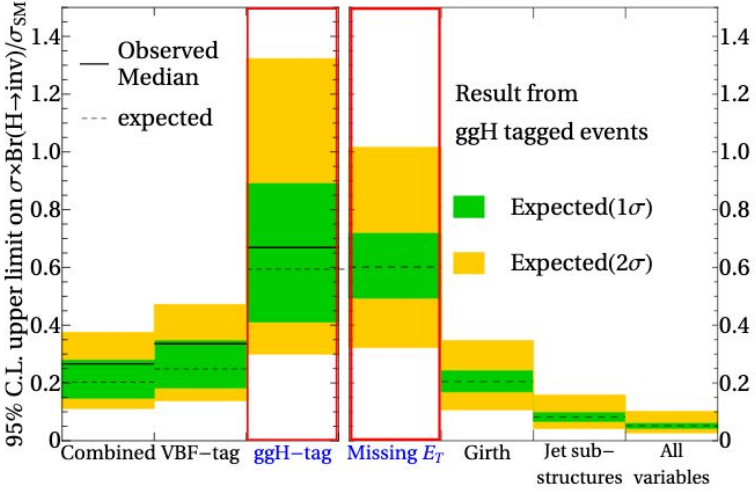
- Identify the particles on each jet containing the largest attention coefficients



- Can we **look** at what **ABCNet** is **learning**?
    - Look at the **self-coefficients**
    - Only plot the **5% particles** inside a jet with the **highest self-coefficients**
- Particle importance spams a **broader region in gluon jets compared to quark initiated jets**

Applications

- [Higgs to invisible limits with ggF]
- Large background from W+jets limit the sensitivity of the ggF contribution to Higgs->Invisible
- Requiring an ISR jet together with the higgs gives an additional handle to separate the 2 components
- In the central region, ISR from ggF is mostly gluon initiated, while in W+jets it's mostly quark initiated



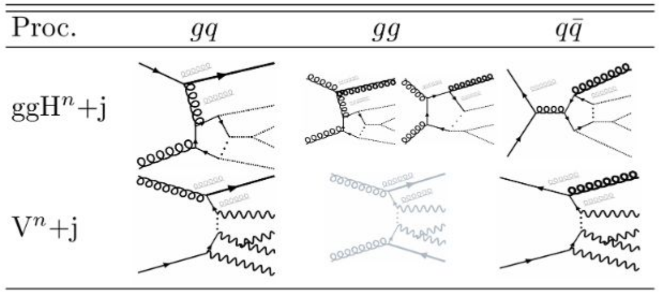
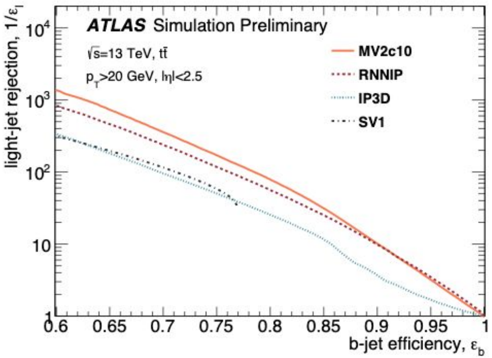


FIG. 1. Leading diagrams (bold) of the (multi) Higgs productions from gluon fusion (ggH^n+jets), against the corresponding irr. (multi) EWVB backgrounds (V^n+jets) with additional ISR(s) for 3 parton initial states (gq,gg,$q\bar{q}$).

B/C tagging

- Commonly used with machine learning exploiting secondary vertices and displaced tracks
- See more in [Javier's talk]

Light-quark ROC



c-quark ROC