Machine learning detection of Higgs and top quark jets in collider events

Jinmian Li

SICHUAN UNIVERSITY

29th Mini-workshop on the frontier of LHC

Fuzhou · 2024.12.14

Base on Phys.Rev.D 108 (2023) 11, 116002

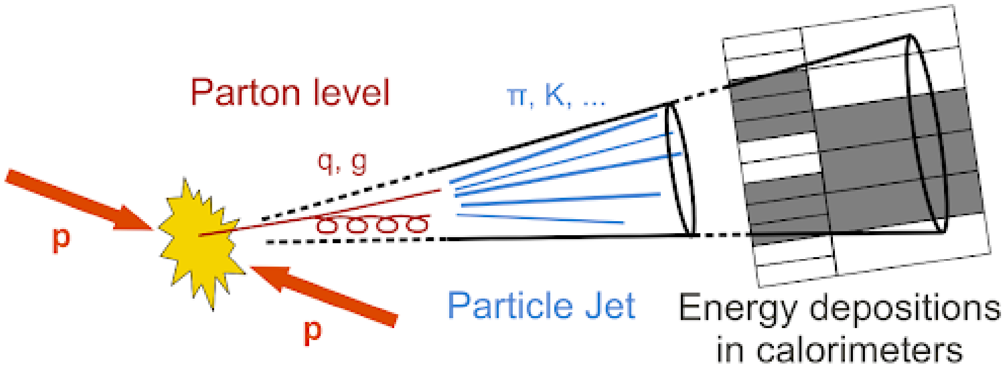
with Sang Kwan Choi, Cong Zhang, Rao Zhang

Outline

1. Jet at detector
    - Jet formation: parton shower and hadronization
    - Jet tagging - Machine Learning
2. Jet detection with machine learning
    - Event image: Mask RCNN
    - Data preprocessing
    - Performance of the Mask R-CNN for Higgs and top detection

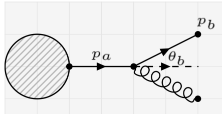
Jet in a detector

Jet is collimated spray of energetic detectable particles, that supposed to have the same origin.



At the LHC, most jets are originated from energetic quark and gluon.

Jet formation: parton shower and hadronization

For a collinear emission: 

$$\sigma_{n+1} \sim \sigma_n \int \frac{dp_a^2}{p_a^2} \int dz \frac{\alpha_s}{2\pi} \hat{P}(z) \equiv \sigma_n \int dt W(t)$$

For multiple emissions

$$\sigma_{n+m} \sim \sigma_n \cdot \int dt_1 \cdots \int dt_m W(t_1) \cdots W(t_m)$$

$$\equiv \sigma_n \cdot \frac{1}{m!} (\int dt W(t))^m$$

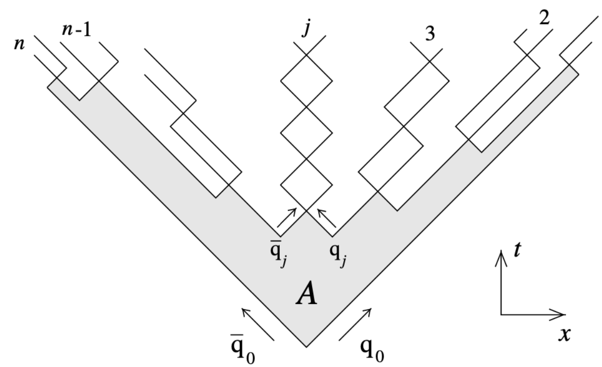
The probability for the next emission at t:

$$d\text{Prob}(t) = dt W(t) \exp(-\int_{t_0}^{t} dt W(t))$$

$\exp(-\int dt W(t))$ is Sudakov form factor = No emission probability

The Lund String Model:

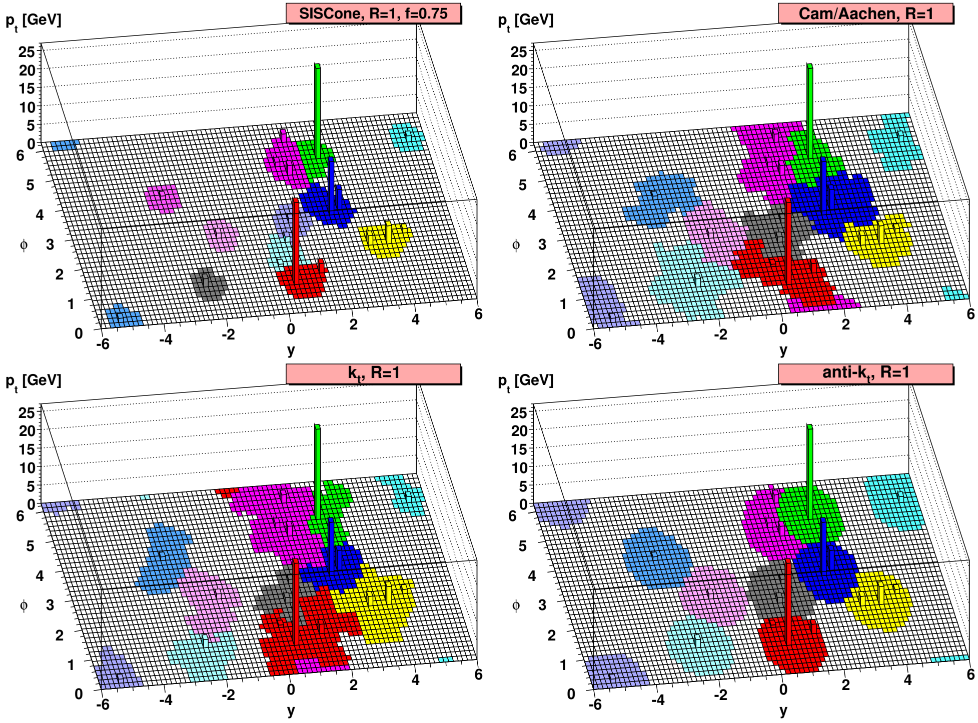
- String breaking probability
  $d\mathcal{P} \propto N \exp(bA) dA$
  N, b are free parameters
  A is the space-time area
- Different quark
  $\mathcal{P} \propto \exp(-\frac{\pi p_{\perp q}^2}{\kappa}) \exp(-\frac{\pi m_q^2}{\kappa})$



Jet tagging procudure

Jet sequential recombination algorithm

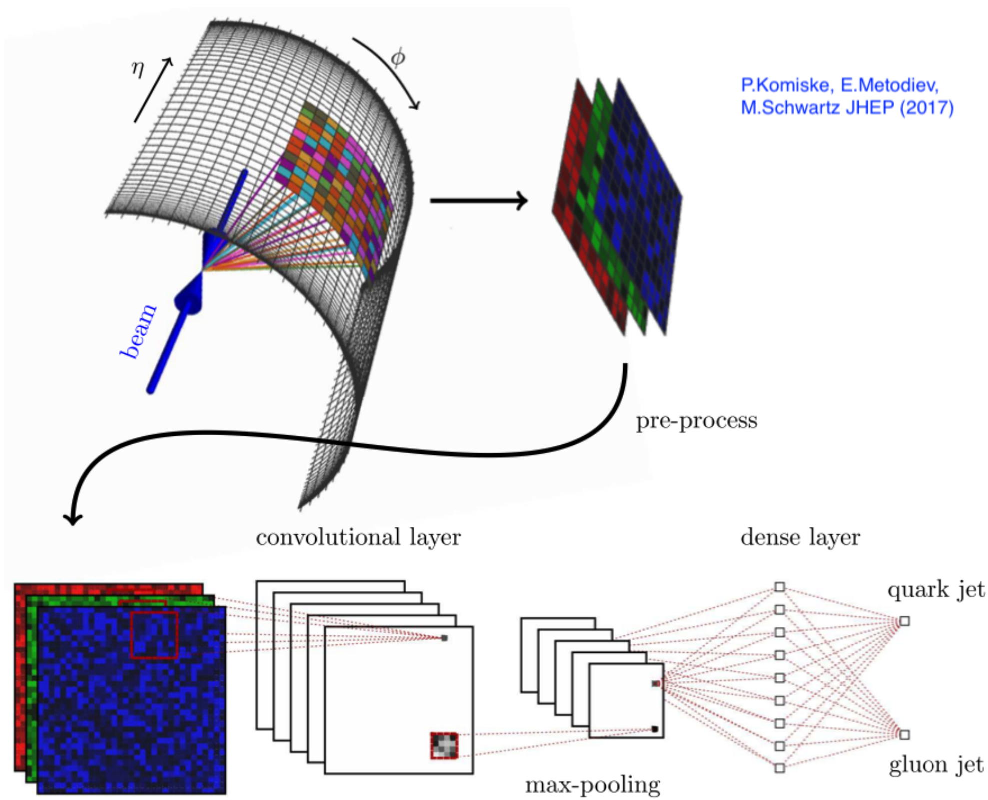
Different jet clustering algorithms use different definition of distance measure.



Figures from Gavin P. Salam Eur.Phys.J.C67:637-686,2010

Jet image representation

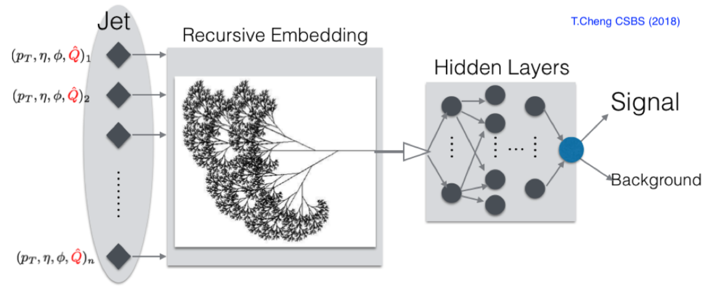
Each calorimeter cell as a pixel and the energy deposition as the intensity
A jet can be viewed as a digital image
Proceeded by 2-dimensional convolutional neural networks

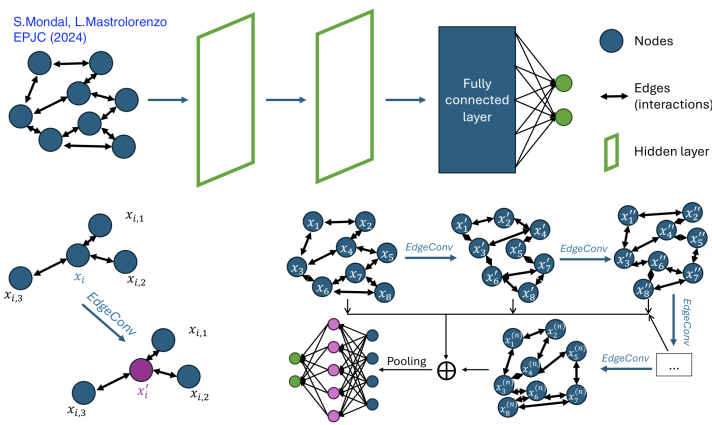


P.Komiske, E.Metodiev, M.Schwartz JHEP (2017)

Graph and sequence representation

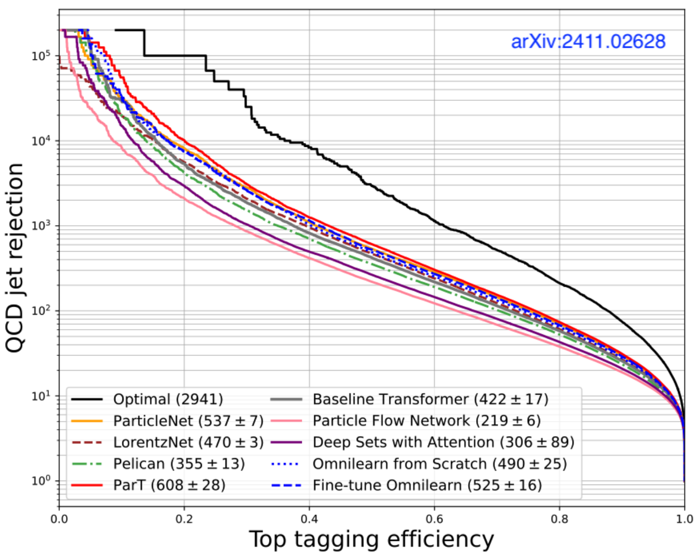
Sequences/trees formed through sequential parton showering and hadronization $\rightarrow$ recurrent neural networks, transformer network, recursive neural networks
Graphs/point clouds with the information encoded in the adjacency nodes and edges $\rightarrow$ graph neural networks





The state-of-art performance of NN

- The ParticleNet, Particle Transformer, LorentzNet and PELICAN are among the state-of-the-art methods, AUC values of over 0.98 for top tagging, without pileup effects.
- The momentum reconstruction component of PELICAN network can predict the p_T and mass of W boson with standard deviations of a few percent.



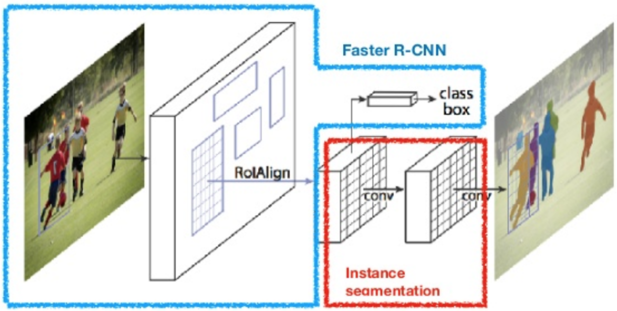
	Method	σ_{p_T} (%)	σ_m (%)	σ_ψ (centirad)
Without DELPHES	JH	0.70%	1.29%	0.162
	PELICAN	0.83%	1.21%	0.388
	PELICAN|JH	0.28%	0.60%	0.089
	PELICAN|FC	0.32%	0.76%	0.111
With DELPHES	JH	10.8 %	8.3 %	8.9
	PELICAN	5.6 %	3.2 %	4.2
	PELICAN|JH	3.8 %	2.9 %	2.7
	PELICAN|FC	4.4 %	3.1 %	3.0

Possible drawbacks

- Performance depends on jet clustering algorithms and their parameters.
- For new physics search, we do not know the jet mass, may be difficult to choose an appropriate jet size.
- Image presentation breaks Lorentz symmetry, the jet mass information is lost.
- May sensitive to pileup events.

Object detection and Semantic Segmentation

Semantic Segmentation is a computer vision task in which the goal is to categorize each pixel in an image into a class or object.



Benchmarks

Add a Result

These leaderboards are used to track progress in Semantic Segmentation

Trend	Dataset	Best Model	Paper	Code	Compare
	ADE20K	ONE-PEACE			See all
	NYU Depth v2	OmniVec2			See all
	Cityscapes test	VLTSeg			See all
	Cityscapes val	SERNet-Former			See all
	ADE20K val	BEiT-3			See all
	PASCAL Context	PlainSeg (EVA-02-L)			See all
	S3DIS Area5	OmniVec			See all
	S3DIS	PTv3 + PPT			See all
	PASCAL VOC 2012 test	DeepLabv3+ (Xception-65-JFT)			See all
	ScanNet	PTv3 + PPT			See all

Show all 148 benchmarks

Mask R-CNN (Mask Region-based Convolutional Neural Network) [K. He, et.al, arXiv:1703.06870] is a state-of-the-art deep learning model used for object detection and instance segmentation in computer vision. It builds upon the Faster R-CNN architecture by adding a branch for predicting object masks in addition to bounding boxes.

Data Preparation

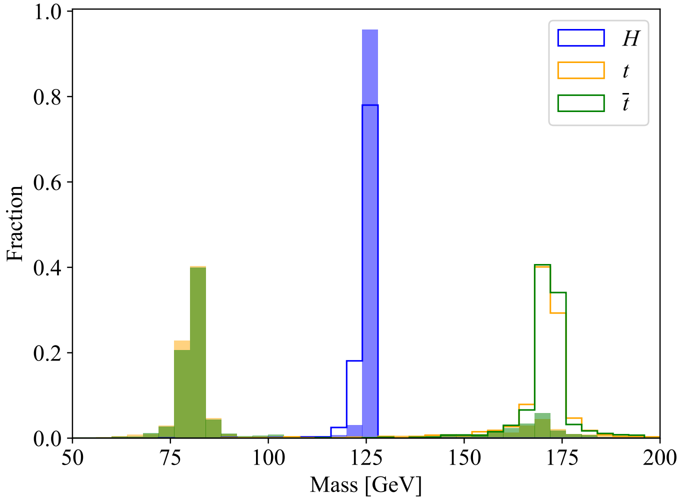
- Training samples: one million events of the $Ht\bar{t}$ process, $p_T(H) > 200$ GeV, $p_T(t) > 300$ GeV
- Test samples including $Ht\bar{t}$, $t\bar{t}t\bar{t}$, $HHt\bar{t}$ in SM, and $\tilde{\chi}_2^0\bar{\tilde{\chi}}_2^0$ with $\tilde{\chi}_2^0 \to H\tilde{\chi}_1^0$, $\tilde{t}\bar{\tilde{t}}$ with $\tilde{t} \to t\tilde{\chi}_1^0$ in SUSY model.
- Average number of 50 pileup events are superposed on each one of the hard events.
- Granulate the event image: the transpose momenta as grayscale on the $\eta - \phi$ plane, with pixel size $\Delta\eta \times \Delta\phi = 0.02 \times 0.02$.
- Could take photon layer, charged lepton layer, neutral and charged hadron layers. Not much improvement!

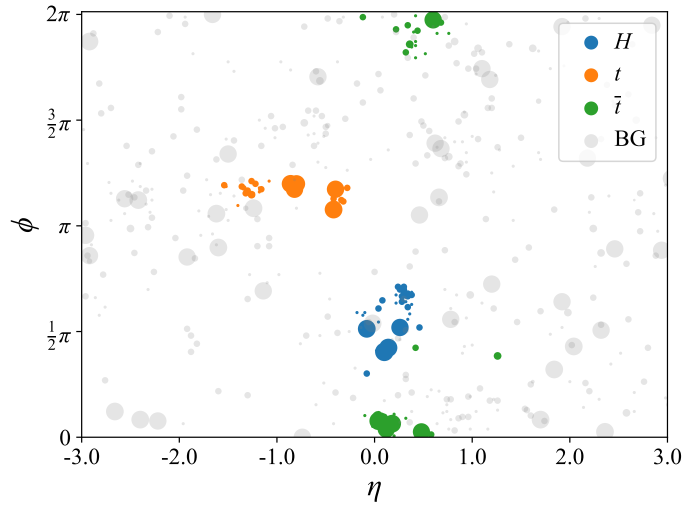
Data preparation

- **Periodicity**: To incorporate the periodicity, the ϕ range is chosen to be from 0 to 2.85π. The spatial size of input images is then 448×448 pixels.
- **IR safety**: Boost along the beam direction to the frame where p_z of the parton is zero, and discard the constituents with energy lower than 0.1 GeV or with angular separation to the parton greater than $\pi/2$. Then, among the remaining constituents, select those having at least 3 others in 20×20 pixels neighbourhood or with p_T greater than 5 GeV.

The top jet constituents

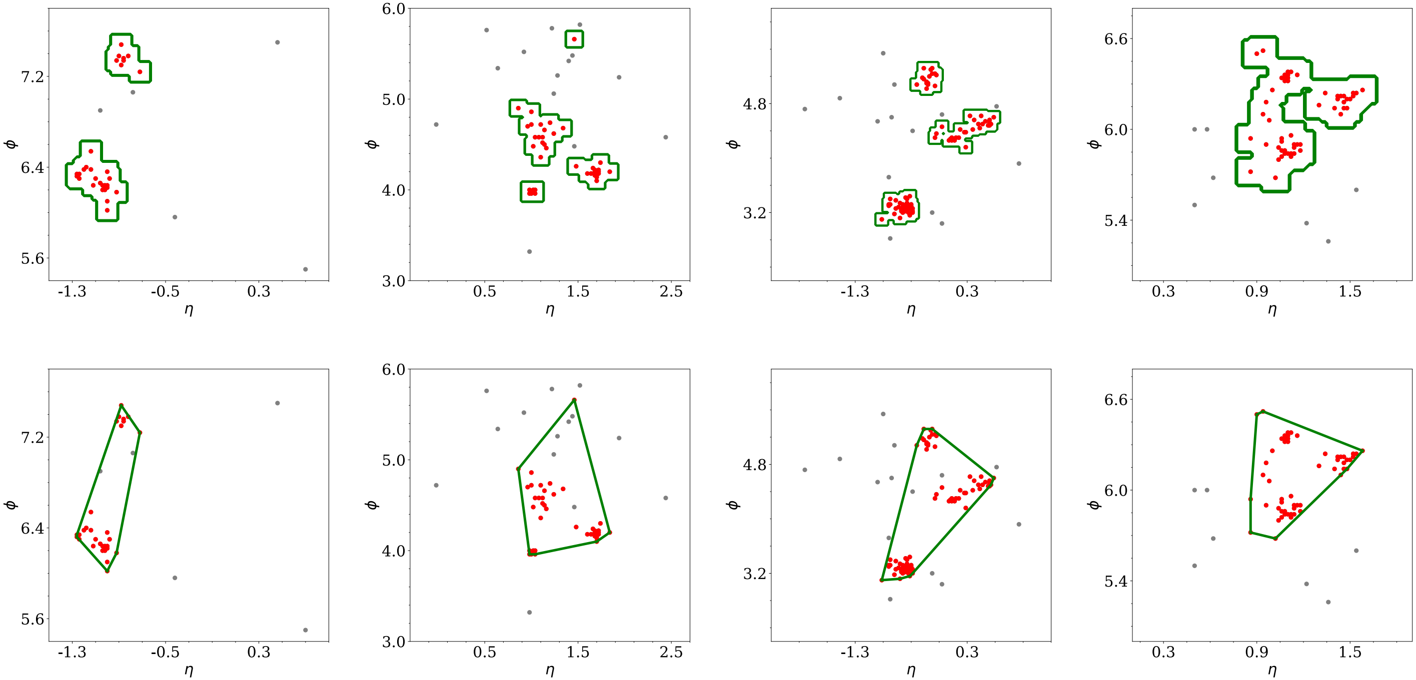
- Due to color confinement, some of the top quark final states could have multiple ancestors other than the top quark
- In MC simulation, final states have multiple ancestors are ranked according to their angular distances assigned to the top/anti-top categories in order until the reconstructed top/anti-top jet invariant mass exceeds $1.05m_t$



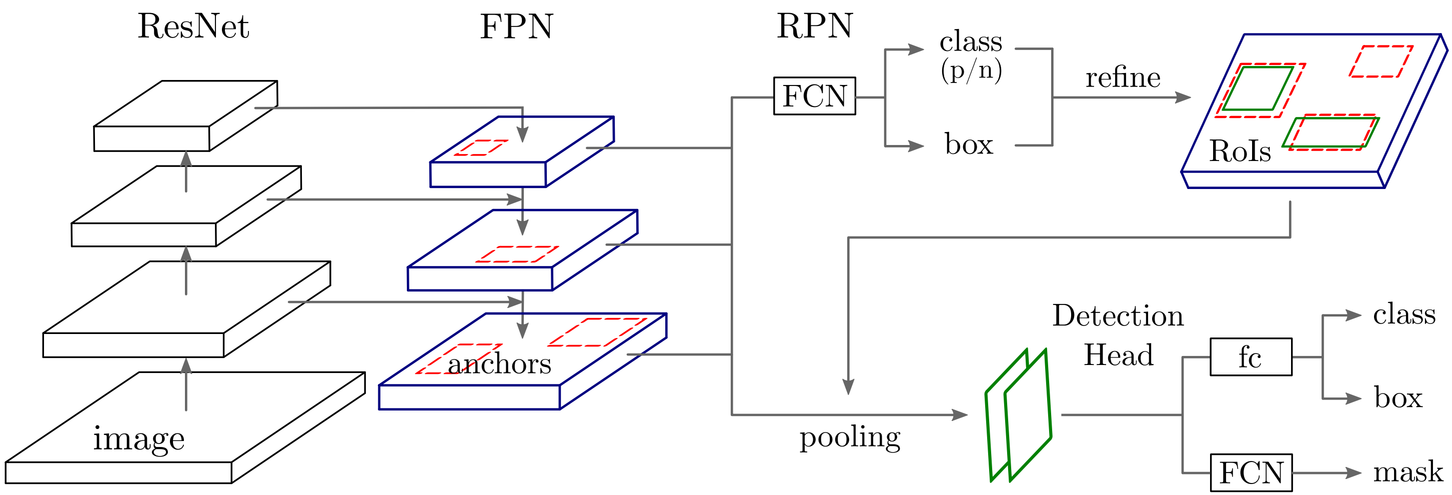


Mask of a Higgs and top jets: two definitions

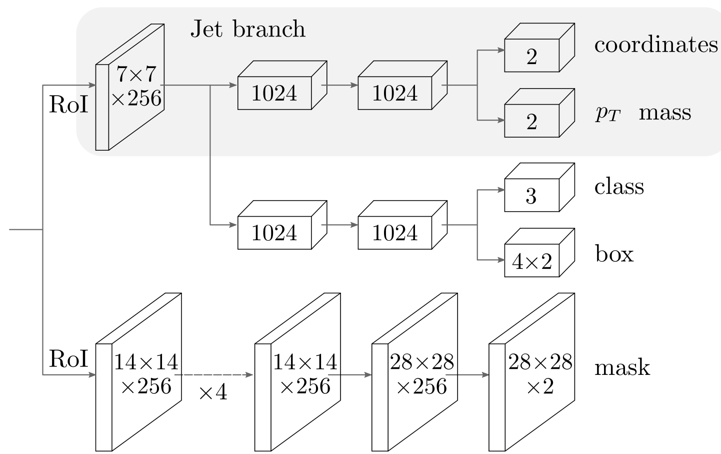
- Convex hull covering all selected constituents
- Enlargement expanding each selected pixel into the area of 9×9 pixels with the selected one at the center



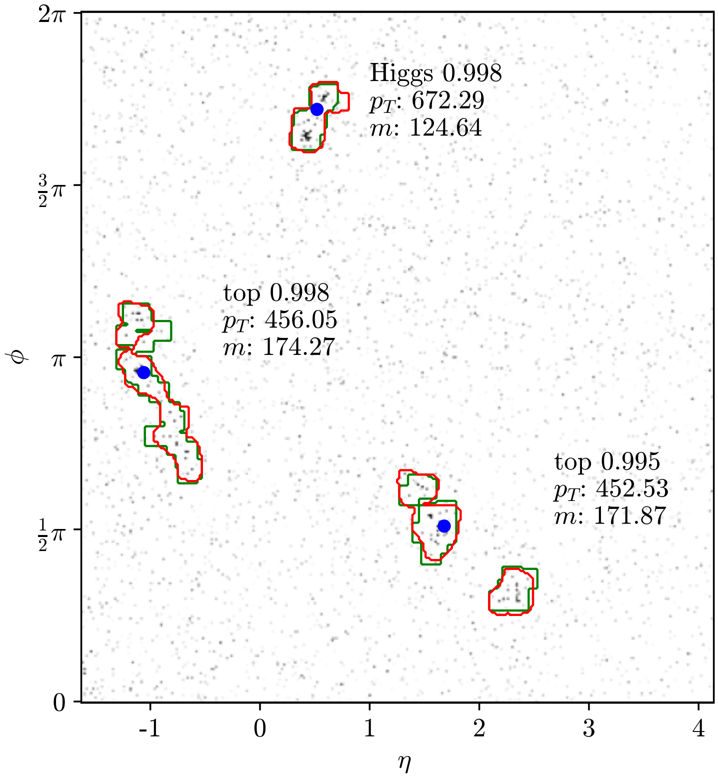
Modified Mask RCNN

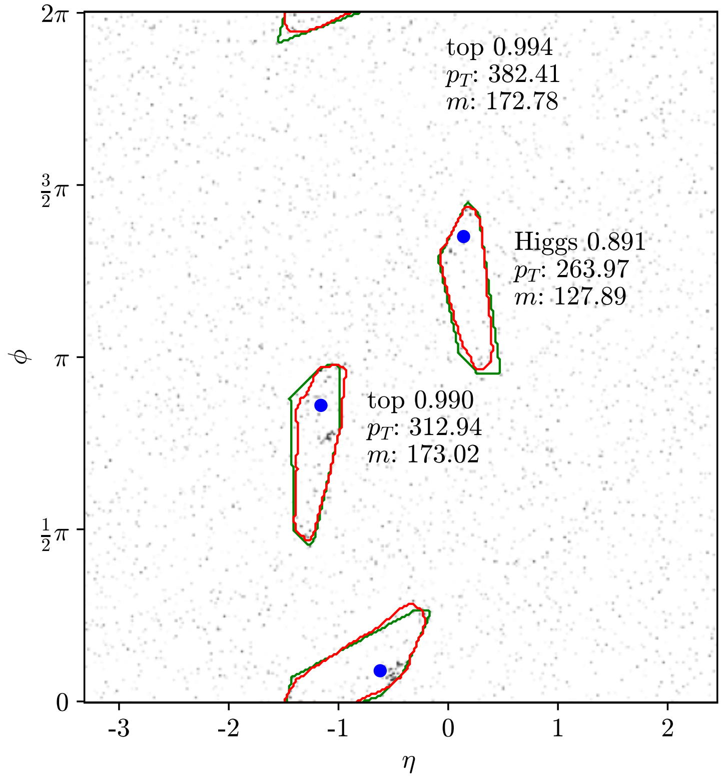


- Backbone architecture for multi-scale feature extraction
- RPN for region proposal generation
- Detection head for classification, box regression, mask segmentation
- An additional branch for predicting the mass, p_T , and coordinates of partons

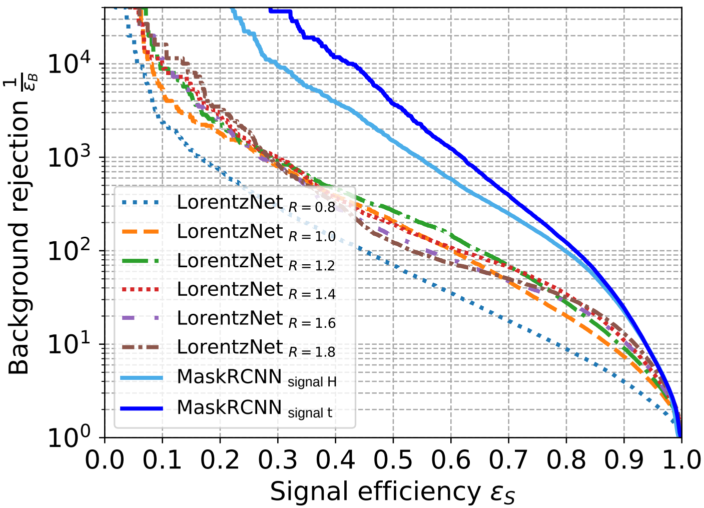


Higgs and top tagging





Performance of the Mask R-CNN



- Retrain these networks on Higgs and top quark jets from our $Ht\bar{t}$ event sample (with pileup events)
- Top quark jet is taken as the signal and Higgs jet is the background
- Performance of LorentzNet degrades when the cone size becomes too small

	LorentzNet						Mask R-CNN	
	$R = 0.8$	$R = 1.0$	$R = 1.2$	$R = 1.4$	$R = 1.6$	$R = 1.8$	signal Higgs	signal top
AUC	0.920	0.953	0.960	0.964	0.962	0.966	0.9723	0.9754

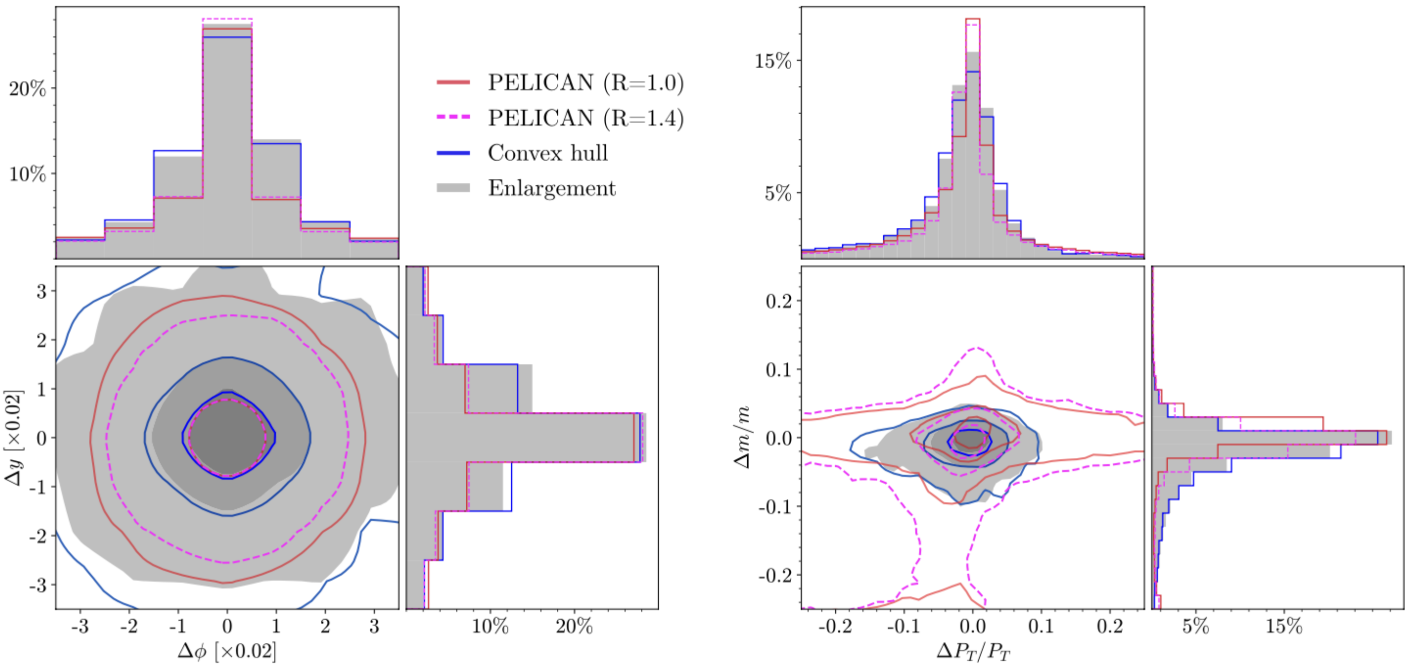
Effects of pileup events

- Two versions of enlarged Mask R-CNN: network@PU50 is trained on events with 50 pileups, network@PU200 is obtained by further training the network@PU50 on 300 thousand events with 200 pileups
- AUC of Mask R-CNN is barely changing for events with pileups smaller than 50

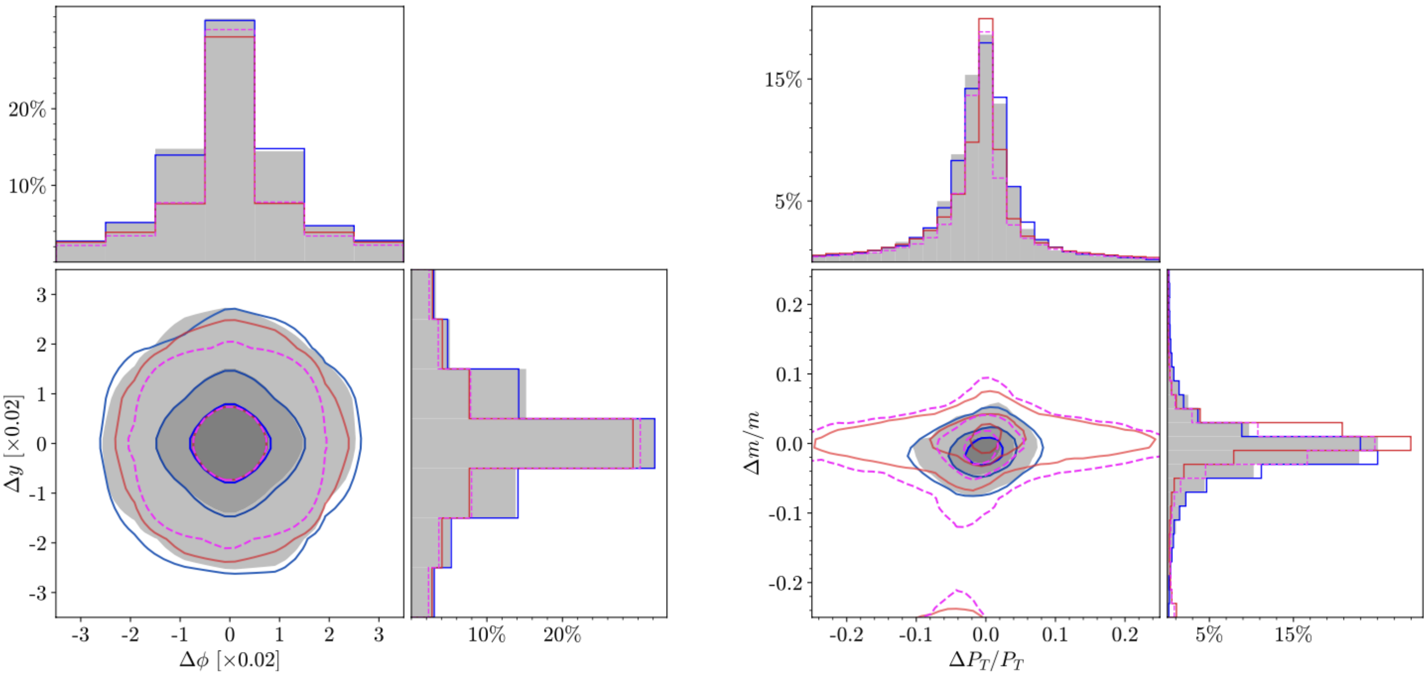
		PU5	PU10	PU20	PU30	PU50
network@PU50	signal Higgs	0.9724	0.9728	0.9729	0.9732	0.9723
	signal top	0.9743	0.9746	0.9751	0.9756	0.9754
network@PU200	signal Higgs	0.9609	0.9620	0.9636	0.9656	0.9678
	signal top	0.9684	0.9690	0.9701	0.9710	0.9723

PU80	PU100	PU120	PU150	PU180	PU200
0.9670	0.9589	0.9446	0.9016	0.8051	0.7037
0.9718	0.9659	0.9547	0.9228	0.8732	0.8323
0.9691	0.9696	0.9702	0.9705	0.9697	0.9691
0.9737	0.9741	0.9744	0.9744	0.9743	0.9740

Performance of the Mask R-CNN (Higgs momentum)



Performance of the Mask R-CNN (Top momenum)



Detector effects

- The energy of final state particles are Gaussian smeared with standard deviation varying from 1% to 20% of the total energy
- Sort the events by the distance of $(\Delta p_T/p_T, \Delta m/m)$ from the origin $(0, 0)$
- Calculate root mean squared distance (RMSD) for given percentages of most accurately predicted Higgs and Top jets.

- Higgs jet

	w/o DS	$0.01E$	$0.02E$	$0.04E$	$0.08E$	$0.1E$	$0.12E$	$0.14E$	$0.18E$	$0.2E$
10%	0.0098	0.0098	0.0101	0.0110	0.0129	0.0133	0.0143	0.0152	0.0167	0.0173
20%	0.0150	0.0150	0.0153	0.0165	0.0189	0.0197	0.0214	0.0225	0.0250	0.0260
30%	0.0198	0.0203	0.0202	0.0219	0.0242	0.0258	0.0277	0.0291	0.0325	0.0341
40%	0.0251	0.0261	0.0259	0.0279	0.0302	0.0322	0.0343	0.0362	0.0404	0.0424
50%	0.0322	0.0337	0.0333	0.0354	0.0372	0.0400	0.0421	0.0444	0.0496	0.0520

- Top jet

	w/o DS	$0.01E$	$0.02E$	$0.04E$	$0.08E$	$0.1E$	$0.12E$	$0.14E$	$0.18E$	$0.2E$
10%	0.0099	0.0095	0.0100	0.0101	0.0105	0.0101	0.0108	0.0107	0.0112	0.0113
20%	0.0147	0.0144	0.0149	0.0151	0.0154	0.0152	0.0160	0.0162	0.0167	0.0170
30%	0.0192	0.0188	0.0194	0.0196	0.0201	0.0200	0.0210	0.0212	0.0219	0.0224
40%	0.0239	0.0235	0.0243	0.0246	0.0252	0.0251	0.0261	0.0265	0.0275	0.0282
50%	0.0295	0.0292	0.0302	0.0306	0.0312	0.0310	0.0324	0.0330	0.0342	0.0351

Applications to other processes

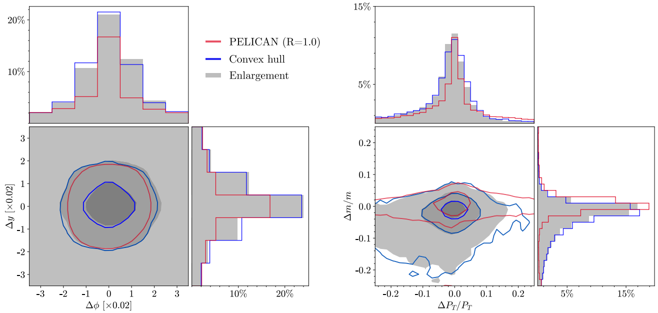
Higgs and Top in $HHt\bar{t}$

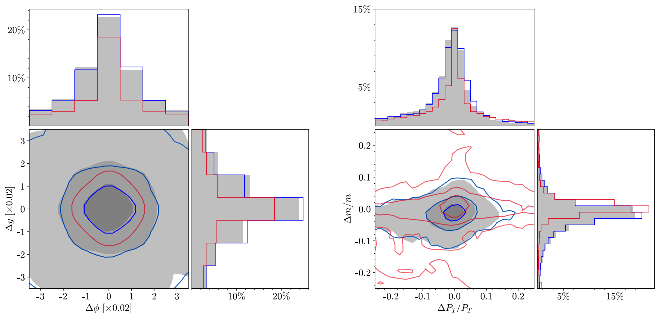
Higgs in $\tilde{\chi}_2\tilde{\chi}_2$

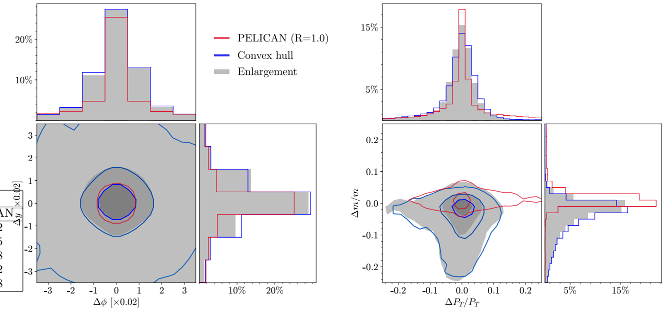
Top in $\tilde{t}\bar{\tilde{t}}$

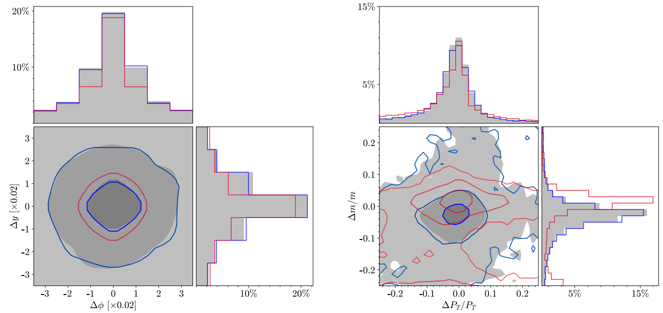
Top in $t\bar{t}t\bar{t}$

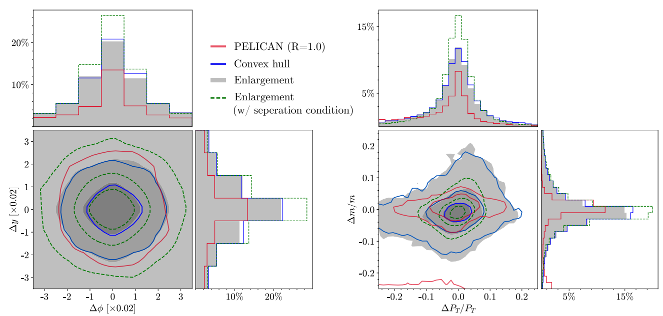
RMSD	Higgs of $HHt\bar{t}$		top of $HHt\bar{t}$		$\tilde{\chi}_2^0\tilde{\chi}_2^0$		$\tilde{t}\bar{\tilde{t}}$		$t\bar{t}t\bar{t}$	
	Enlarge	PELICAN	Enlarge	PELICAN	Enlarge	PELICAN	Enlarge	PELICAN	Enlarge	PELICAN
10%	0.0133	0.0119	0.0127	0.0110	0.0133	0.0080	0.0184	0.0160	0.0143	0.0152
20%	0.0210	0.0221	0.0197	0.0198	0.0207	0.0139	0.0271	0.0261	0.0221	0.0315
30%	0.0292	0.0416	0.0268	0.0342	0.0284	0.0236	0.0364	0.0399	0.0306	0.0938
40%	0.0405	0.0804	0.0359	0.0616	0.0376	0.0504	0.0489	0.0698	0.0413	0.1712
50%	0.0593	0.1497	0.0485	0.1108	0.0508	0.1227	0.0731	0.1254	0.0561	0.2438











Conclusion

- MASK R-CNN is powerful in detecting jets with substructure: no artificial parameter, able to detect jets with arbitrary size, has advantage of smaller jet area thus more precise reconstruction.
- MASK R-CNN predicts jet area and momentum for a wide class of processes.
- Pileup mitigation can be implemented intrinsically.

Future prospects:

- Detect jet with unknown mass?
- The tracker has better angular resolution than the electromagnetic/hadronic calorimeter. To incorporate the particle level information in event image?