Machine learning detection of Standard Model jets in collider events

Jinmian Li

第四届高能物理理论与实验融合发展研讨会 · 2025.9.20

Base on arXiv: 2506.18336

with Peng Li, Bingwei Long, Rao Zhang

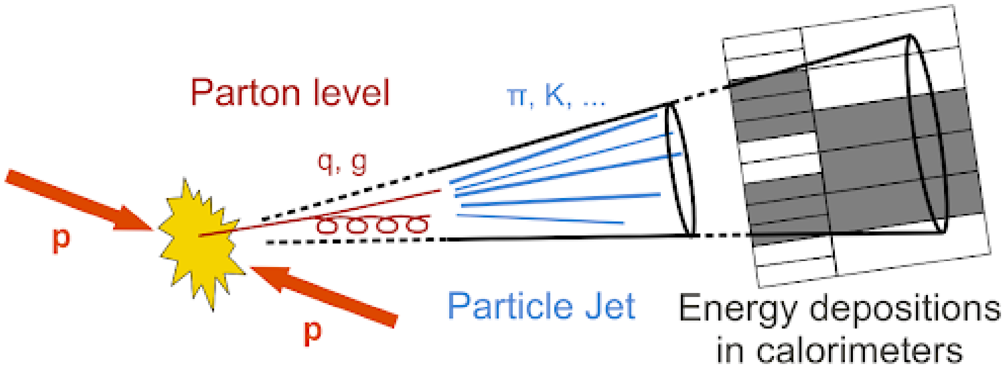
Outline

1 Jet at detector
  - Jet formation: parton shower and hadronization
  - Jet **tagging**: Machine Learning

2 Jet **detection** with machine learning
  - Event representation: event image
  - Data pre-processing
  - The network performance

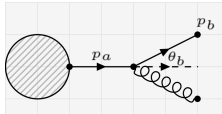
Jet in a detector

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

$$\sigma_{n+m} \sim \sigma_n \cdot \int dt_1 \cdots \int dt_m W(t_1) \cdots W(t_m)$$
$$\equiv \sigma_n \cdot \frac{1}{m!} (\int dt W(t))^m$$

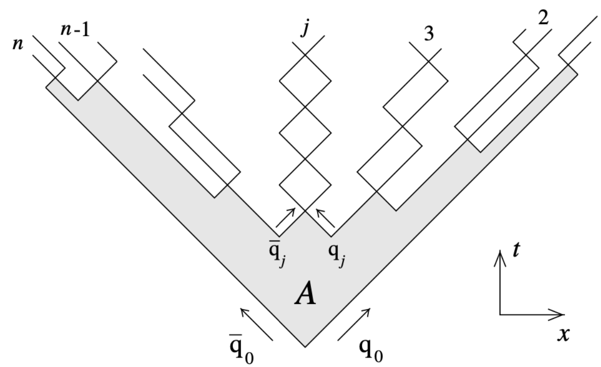
$\exp(-\int dt W(t))$ is Sudakov form factor = No emission probability

The probability for the next emission at t:

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

The Lund String Model:

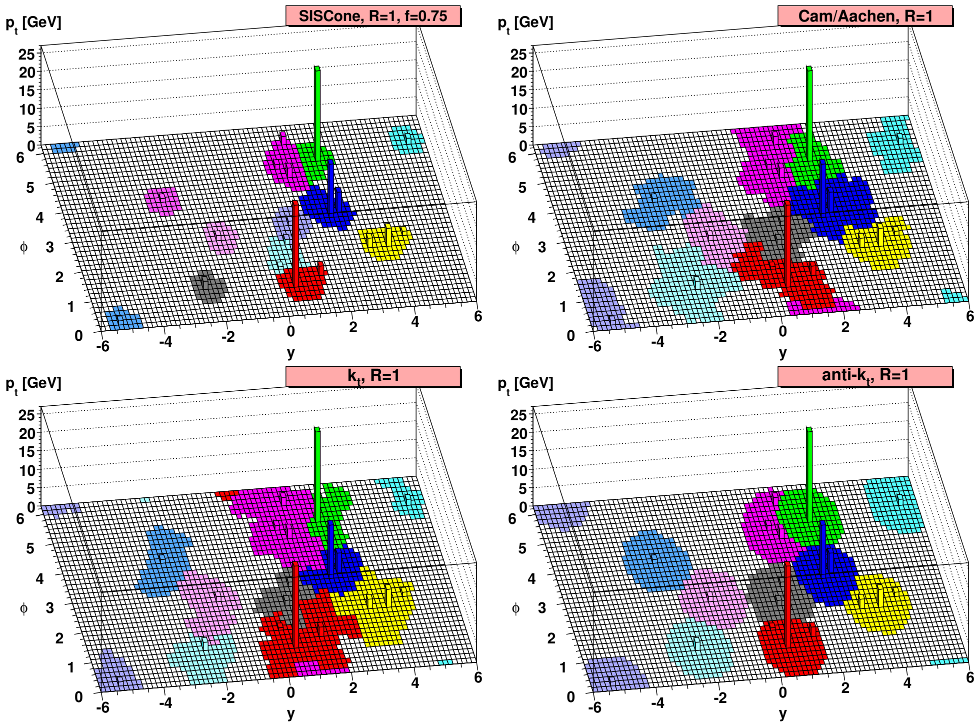
- String breaking probability
  $d\mathcal{P} \propto N \exp(bA) dA$
  N, b are free parameters
  A is the space-time area
- Excited quark
  $\mathcal{P} \propto \exp(-\frac{\pi p_{\perp q}^2}{\kappa}) \exp(-\frac{\pi m_q^2}{\kappa})$



Jet tagging procudure

Jet sequential recombination algorithm

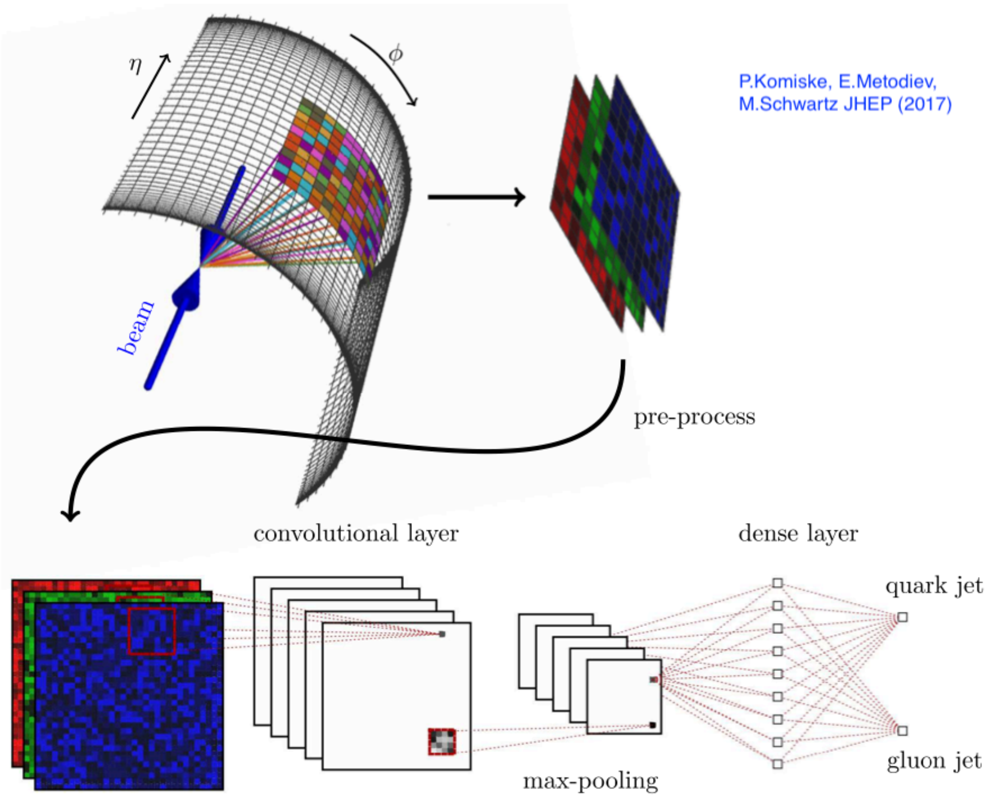
Different jet clustering algorithms use different definition of distance measure.



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

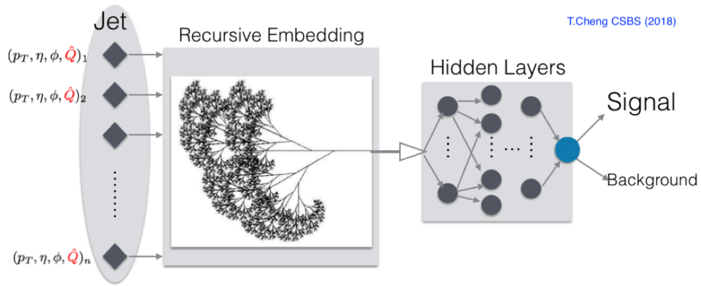
Jet image representation

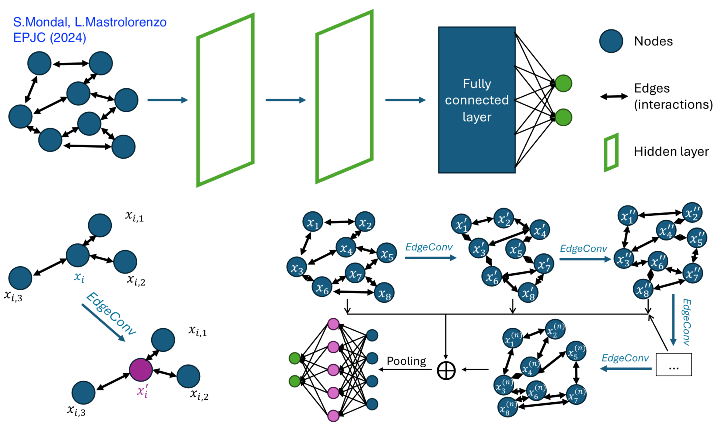
Each calorimeter cell as a pixel and the energy deposition as the intensity
A jet can be viewed as a digital image
Proceeded by 2-dimensional convolutional neural networks



Graph and sequence representation

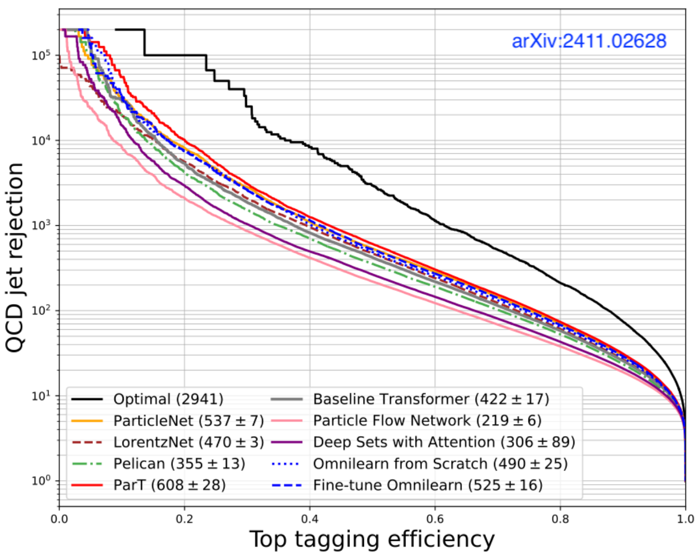
Sequences/trees formed through sequential parton showering and hadronization → recurrent neural networks, transformer network, recursive neural networks
Graphs/point clouds with the information encoded in the adjacency nodes and edges → graph neural networks





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.



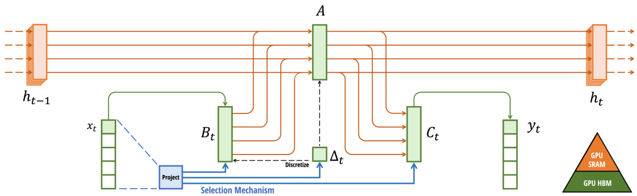
	Method	σ_{p_T} (%)	σ_m (%)	σ_ψ (centirad)
Without DELPHES	JH	0.70%	1.29%	0.162
	PELICAN	0.83%	1.21%	0.388
	PELICAN|JH	0.28%	0.60%	0.089
	PELICAN|FC	0.32%	0.76%	0.111
With DELPHES	JH	10.8 %	8.3 %	8.9
	PELICAN	5.6 %	3.2 %	4.2
	PELICAN|JH	3.8 %	2.9 %	2.7
	PELICAN|FC	4.4 %	3.1 %	3.0

Possible drawbacks

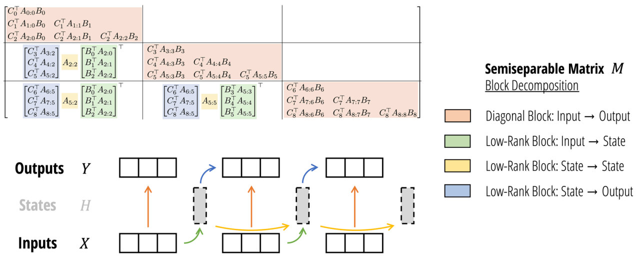
- Performance depends on jet clustering algorithms and their parameters.
- For new physics search, we do not know the jet mass, may be difficult to choose an appropriate jet size.
- Image presentation breaks Lorentz symmetry, the jet mass information is lost.
- May sensitive to pileup events.

Object detection and Instance Segmentation

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



Mamba: Linear-Time Sequence Modeling with Selective State Spaces
Albert Gu*, Tri Dao*
Paper: https://arxiv.org/abs/2312.00752



Mamba is a revolutionary State Space Model architecture that excels in both efficiency and long-range modeling. Linear-time complexity, a hardware-aware design inspired by FlashAttention, unique selection mechanism allows it to intelligently compress context and selectively capture long-range dependencies.

模型名称	核心架构/技术	突出特点/主要贡献
InternImage	大核CNN / Deformable Conv v3	使用大至51×51的可变形卷积核，实现了超越Transformer的全局建模能力，同时保持了CNN的效率。
Mask2Former	Transformer (Masked Attention)	将分割任务统一为“Mask-classification”（掩码分类）的查询式框架，在全景、实例和语义分割任务上都达到顶尖水平。
K-Net (K-Team)	Transformer (Dynamic Kernels)	引入动态卷积核学习，统一了实例分割和语义分割，在多个基准上持续霸榜。
OneFormer	Transformer (Task-conditioned)	提出一个单一的、任务引导的Transformer框架，能够处理语义、实例和全景分割，实现了SOTA性能。
U-Mamba / VM-UNet	Mamba (SSM) + U-Net	将Mamba（状态空间模型）引入U-Net架构，高效捕捉长距离依赖，在医学影像分割上表现尤为突出。
ConvNeXt V2	现代CNN	从头开始现代化标准CNN（ResNet），在设计和训练上借鉴了Transformer的成功经验，性能强大且高效。
SegMamba	Mamba (SSM)	专为3D医学图像分割设计，通过三向Mamba块有效捕捉3D空间上下文，性能优于许多Transformer模型。
ViT-Adapter	Vision Transformer (Adapter)	在预训练的Vision Transformer中插入轻量级的适配器模块，使其能够高效地适应下游的分割任务。

Objects and Methods

- Treat the entire event as a single image in the pseudorapidity-azimuth ($\eta - \phi$) plane.
- Employ a vision-adapted Mamba network (**VMamba-V2**) as the core feature extractor
- Simultaneous instance segmentation, classification, and kinematic regression for H, t, W/Z, b and q/g jets.
- Recognize their subsequent **decay products**.

Data Preparation

- Simulate the **SM processes**:

$$jjjj,\ b\bar{b}WW,\ Hb\bar{b},\ HHWW,\ Ht\bar{t},\ tb,\ t\bar{t},\ t\bar{t}W,\ t\bar{t}Z,\ t\bar{t}t\bar{t}$$
$$WH,\ WW,\ WZ,\ ZH,\ ZZ,\ ZZZZ$$

- Transverse momentum (p_T) cuts of $p_T^{b/W/Z/H} > 200\,\mathrm{GeV}$ and $p_T^t > 300\,\mathrm{GeV}$ were applied at parton level.
- Heavy particles were required to decay hadronically: $t \to Wb$, $H \to b\bar{b},\ W/Z \to jj$.
- 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 $\Delta\eta \times \Delta\phi = 0.02 \times 0.02$.

Defination of jet constituents for quarks

- Due to color **confinement**, some of the quark fragmentation final states could have multiple ancestors
- Classify final-state hadrons into **two categories** based on their ancestral traceability: **U** (unambiguously mapped to a single parent quark), **C** (momentum inherited from a group of color-connected ancestor quarks)
- The momentum of the k-th **initial-state** quark is

$$p_k = \sum_{u \in \mathbf{U}_k} p_u + \sum_{c \in \mathbf{C}} \alpha_{kc} p_c$$

- Determination of α corresponds to a Mixed-Integer Linear Programming problem:

$$\hat{\alpha} = \arg\min_{\alpha} \mathcal{L}(\alpha), \quad \text{subject to } \sum_{k=1}^{K} \alpha_{kj} = 1, \ \forall j \in \mathbf{C}, \ \alpha_{kj} \in 0, 1$$

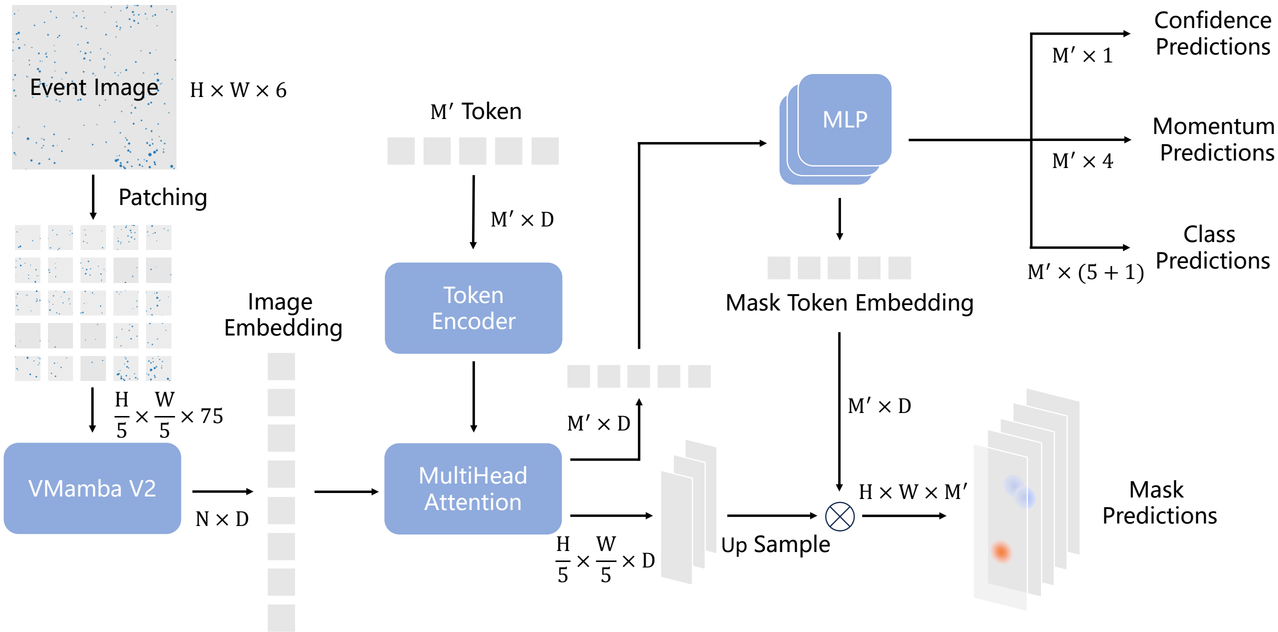
with

$$\mathcal{L}(\alpha) = \sum_{k=1}^{k=K} \Delta p_k \equiv \sum_{k=1}^{k=K} |p_{q_k} - p_k(\alpha_{kc})|$$

The inputs to the network

- Grid covers $\eta \in [-\pi, \pi]$ and $\phi \in [0, 2\pi]$ and has a fine granularity of 0.02×0.02.
- **Base features** include four-momenta components (p_x, p_y, p_z, E), the absolute charge $|Q|$ and the transverse impact parameter d_0
- **Instance Segmentation Mask:** 1). ensure IRC safety (constituent particle with transverse momentum $p_T > 0.1$ GeV, pseudorapidity $|\eta| < \pi$ and at least 4 other assigned constituents are present within a 20-pixel radius.) 2). 2D Gaussian distributed mask with $\sigma = 2$ pixel. 3). each of the initial masks is individually normalized by its own maximum value.
- **Kinematic Regression Targets:** Ancestor particle's four-momentum: p_T, y, ϕ, m.

The Network Architecture



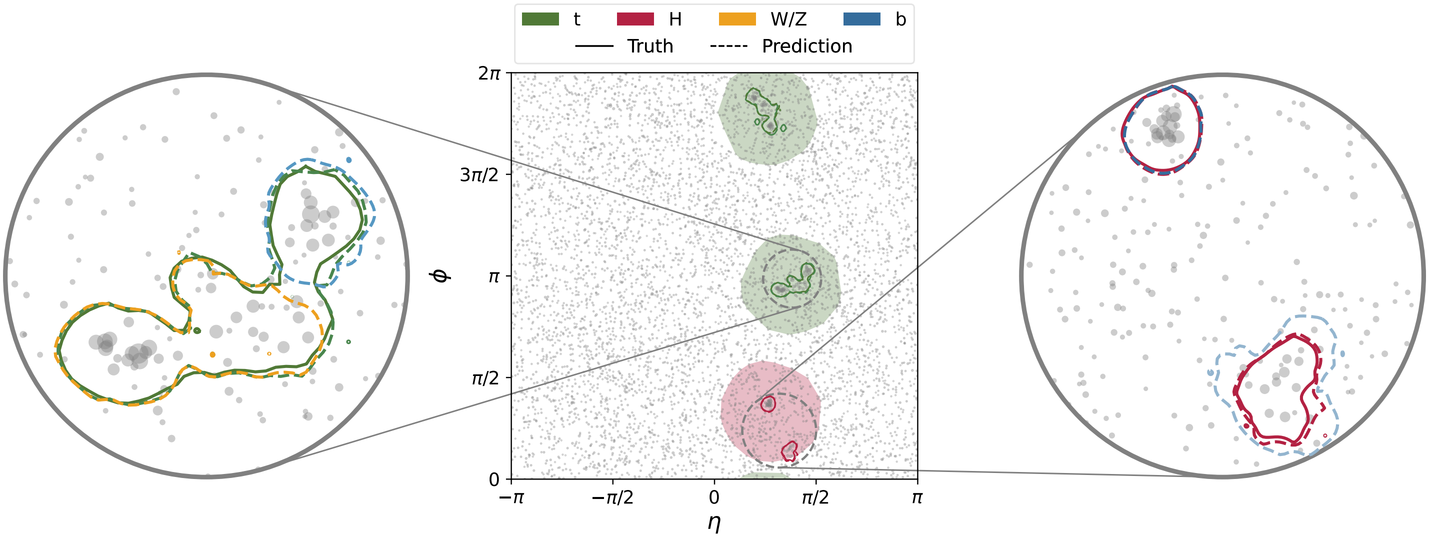
1. The input event image processed by a "Patching" module. **2.** The VMamba V2 backbone processes these patches to produce an image embedding feature. **3.** A Multi-Head Attention module fuses the global event embedding with query tokens. **4.** Output tokens are then processed by two parallel prediction heads.

Training the network

- The model consists of **4.75 million** trainable parameters
- The model was trained for 120 epochs on two NVIDIA RTX 3090 (24GB) GPUs, a process that took approximately **194 hours**.
- On a single RTX 3090 GPU, the average inference time per event is approximately **10.6 milliseconds**.
- Each event sample in every epoch was randomly overlaid with **pileup** interactions.
- To improve the model's rotational invariance and leverage the detector's azimuthal symmetry, we apply a random **azimuthal shift** during training.

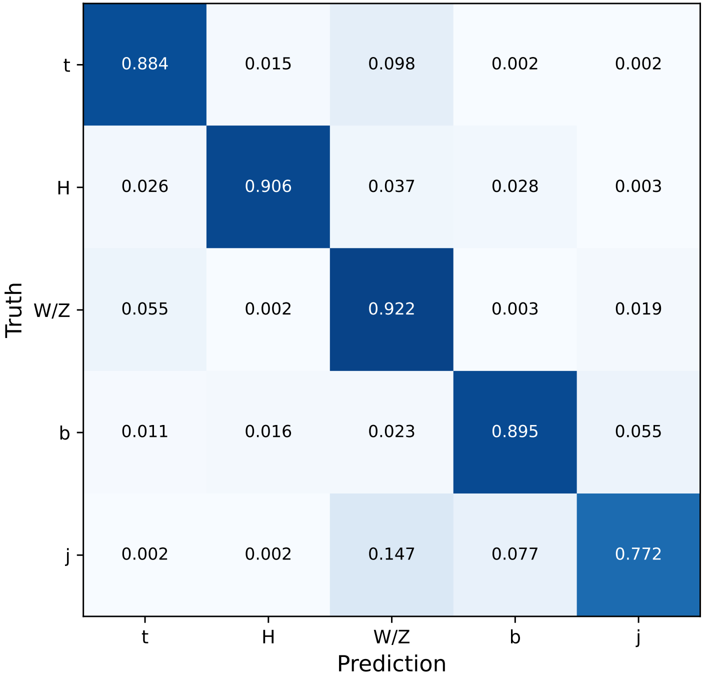
`https://github.com/scu-heplab/seg-any-sm-jet`.

The network performance



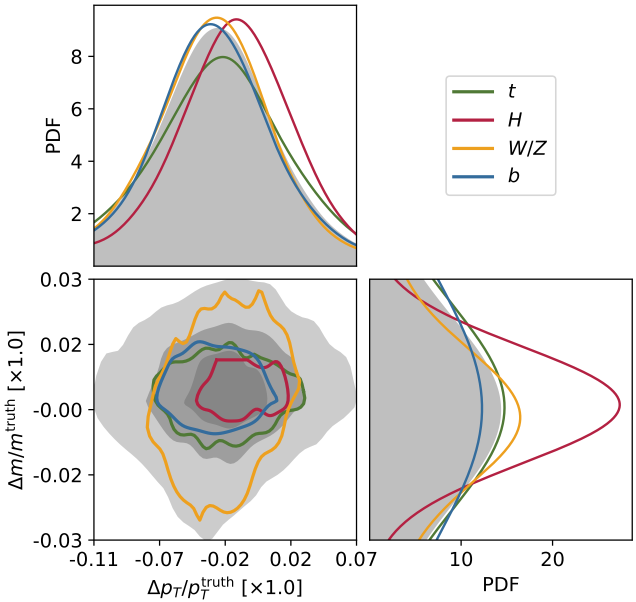
- Multi-level jet identification: recognizing **primary** heavy particles and their **decay products**.
- Correctly associates these non-adjacent jets, grouping them to reconstruct the parent Higgs boson.
- The shaded regions delineate the jet areas found using the anti-k_T algorithm with cone size parameter $R = 0.8$.

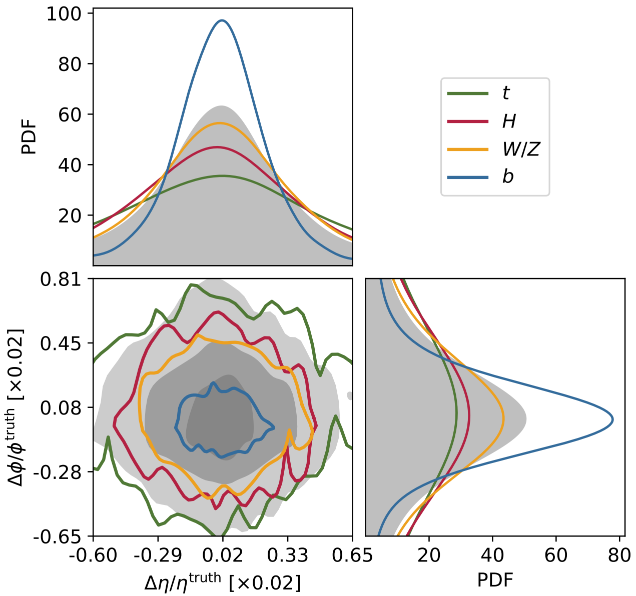
The Confusion Matrix



- The **diagonal** elements represent the recall (true positive rate) for each category, indicating the fraction of jets correctly classified.
- The highest recall is achieved for H- and W/Z-jets, their constituents can be unambiguously identified from the Monte Carlo truth.
- The tendency for top-jets to be misidentified as W/Z-jets, due to soft or merged b-jet.

The Momentum Regression





RMSD	20%	40%				60%
		t	H	W/Z	b	
$(\Delta p_T/p_T^{\text{truth}},\ \Delta m/m^{\text{truth}})$	0.02914	0.0392	0.02322	0.04411	0.03939	0.06266
$(\Delta \eta/\eta^{\text{truth}},\ \Delta \phi/\phi^{\text{truth}})$	0.00401	0.01410	0.00964	0.00819	0.00404	0.01262

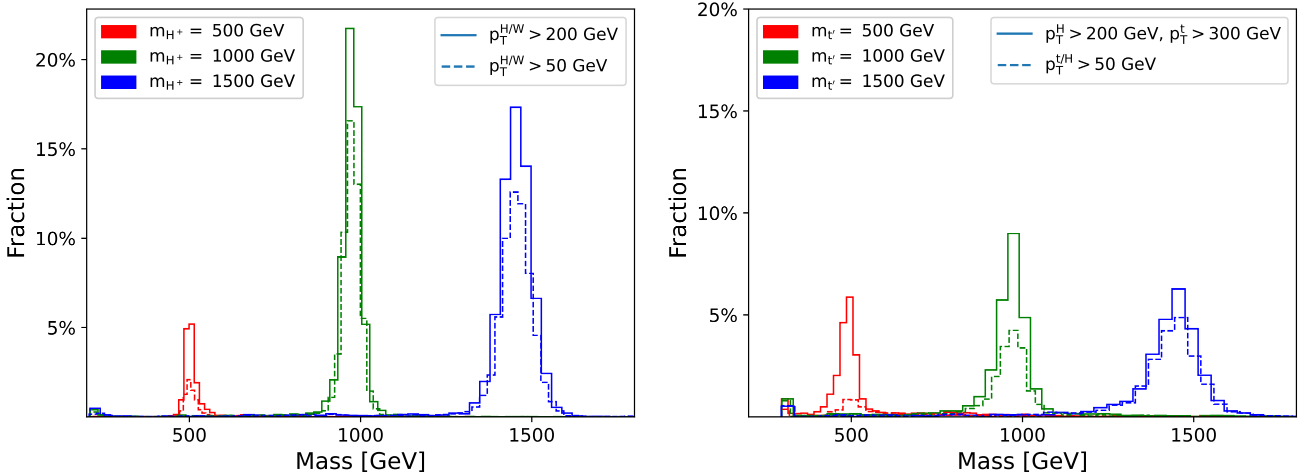
Effects of pileup events

Using the **Average Precision** at an IoU threshold of 0.5 (AP^{50})

$\langle\mu\rangle$	5	20	50	100	150	200
t	0.537	0.527	0.505	0.47	0.426	0.364
H	0.482	0.472	0.458	0.429	0.392	0.341
W/Z	0.604	0.591	0.569	0.519	0.466	0.401
b	0.588	0.58	0.568	0.539	0.517	0.487

- Increasing pileup uniformly degrades the classification performance across all jet classes
- Jets with complex internal structures, like hadronically decaying top- and W/Z-jets, are more susceptible.
- The b-jet exhibits the highest resilience to pileup

Reconstruction of BSM resonances



- H^+ from $pp \to H^+(\to HW)$, t' from $pp \to Zt'(\to tH)$
- The model can generalize to successfully identify and reconstruct **unseen** BSM signals (which did not seen in training).
- The nominal-cut samples consistently yield **narrower** peaks, indicating better mass resolution.
- H^+ resonance is reconstructed with much better resolution than the t'

Conclusion

- Our model unifies **instance segmentation**, **classification**, and **kinematic regression** into a single multi-task learning system, allowing for the simultaneous identification of both **primary** jets and their internal **sub-jet** structures.
- VMamba benefited from efficiency and long-range modeling capabilities.
- Pileup mitigation can be implemented intrinsically.
- Applicable to events of **general** processes.

Future prospects:

- Improve the light-flavor, gluon jet reconstruction?
- The tracker has better angular resolution than the electromagnetic/hadronic calorimeter. To incorporate the particle level information in event image?