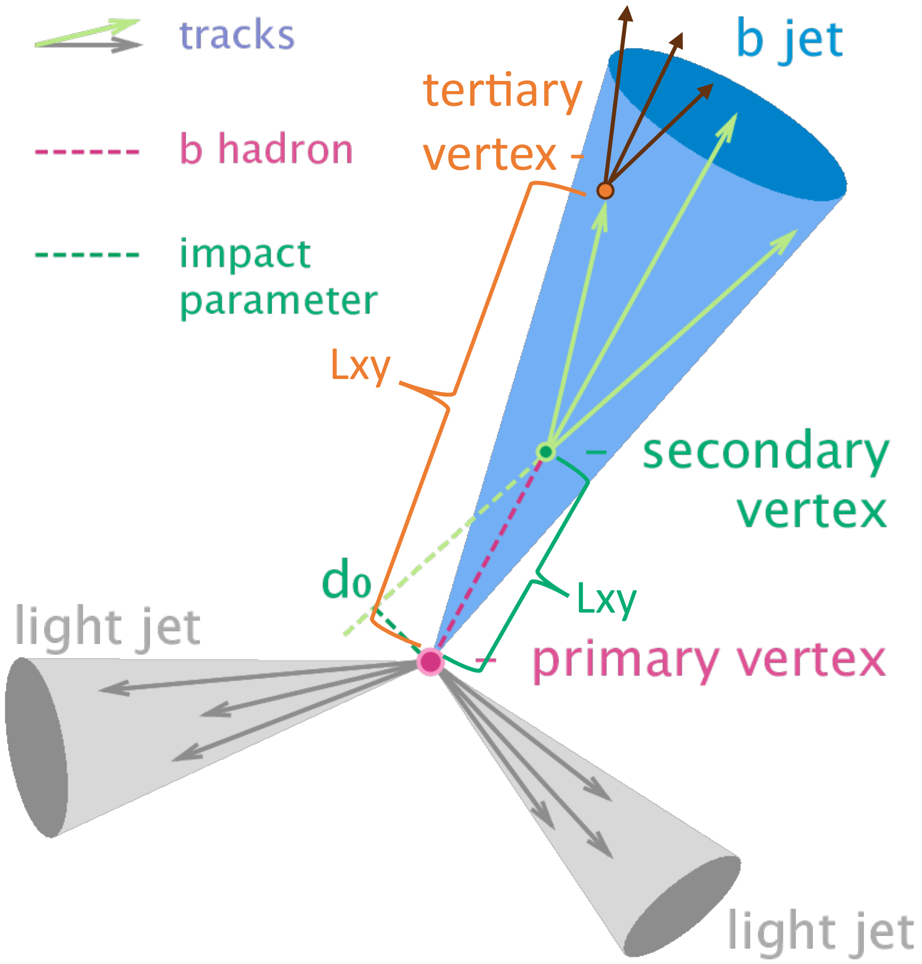
Hadronic Vertex Reconstruction with Transformers

ML4Jets
06/11/2023

N. Pond on behalf of
J. Barr, G. Facini, M. Hart, N. Pond, S. Rettie, T. Scanlon, S. Van Stroud

Introduction

- Vertex reconstruction widely useful in HEP
  - E.g, b-tagging
    - b-jets can have multiple displaced vertices
- Two main steps
  - Vertex finding
  - Vertex fitting
- Each jet can have varying numbers of tracks
  - Aim to reconstruct varying number of vertices
  - Can be difficult for machine learning (ML) based approaches
- We demonstrate a new ML based approach
  MaskFormer for Heavy Flavour Vertexing
- Performs both vertexing steps in unison, allowing
  - Multiple vertices to be identified
  - Truth properties to be regressed
- Preliminary proof-of-concept results



MaskFormer

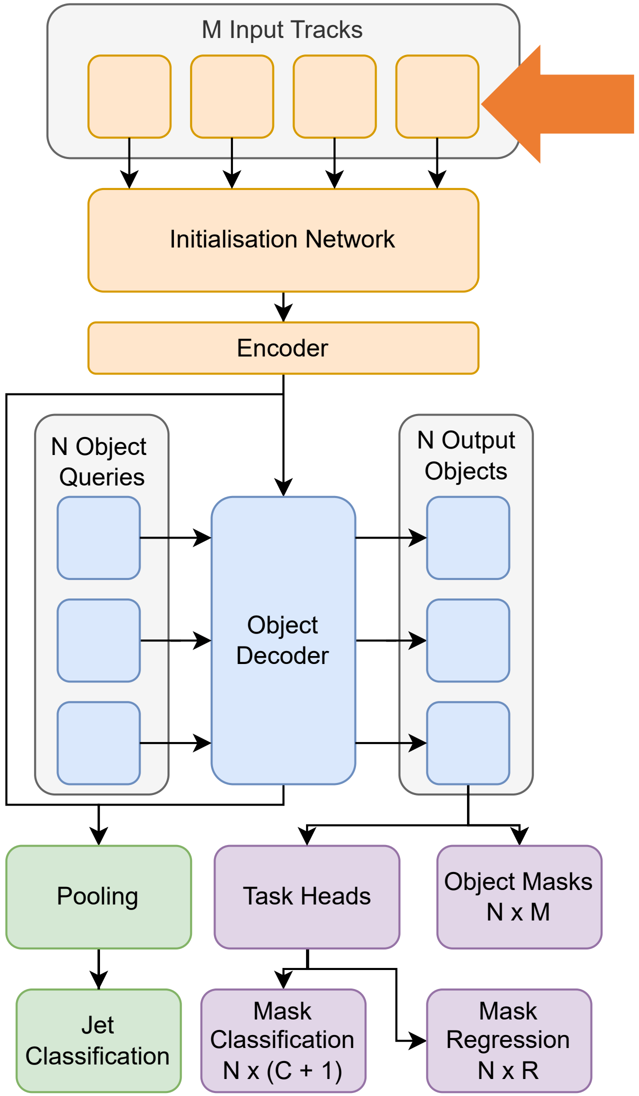
- Take inspiration from object detection in image processing
  - Utilise a *MaskFormer*
    - *Masked-attention Mask Transformer for Universal Image Segmentation*, 2022, B. Cheng et al., arXiv:2112.01527v3
  - Input pixels → Model → Output classified masks
- Vertex Reconstruction is *Many-To-Many* task
  - Input varying number of M tracks
  - Output varying number of N vertices
- We've modified MaskFormer to use
  - Sparse inputs
  - Sparse mask predictions
  - Output object regressions

M Input pixels

N Classified Masks

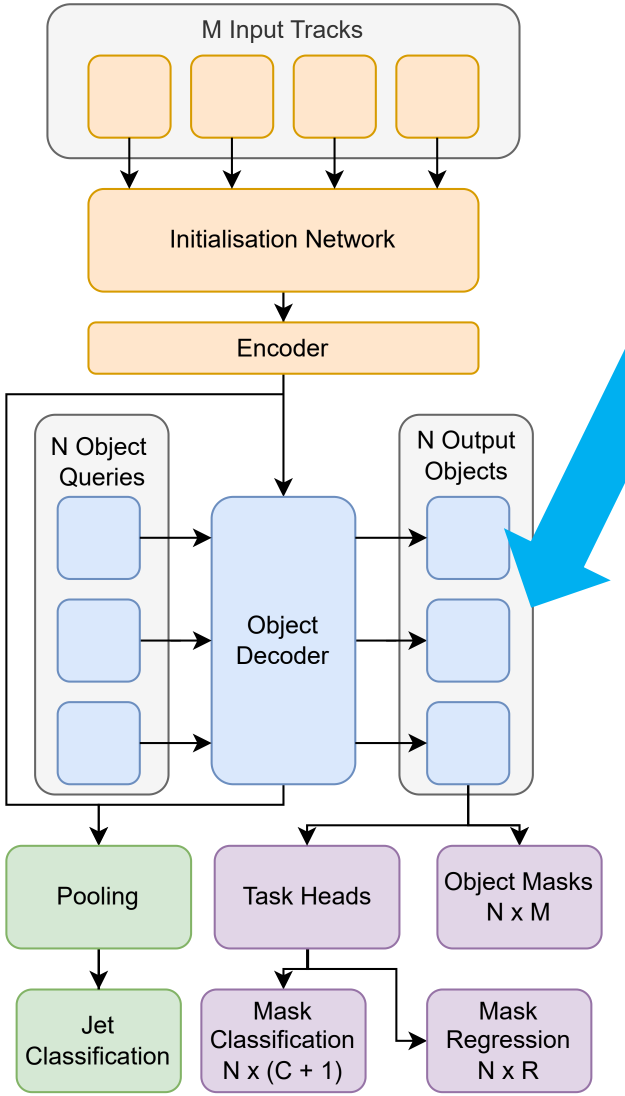
Meta AI : segment-anything.com

MaskFormer in HEP



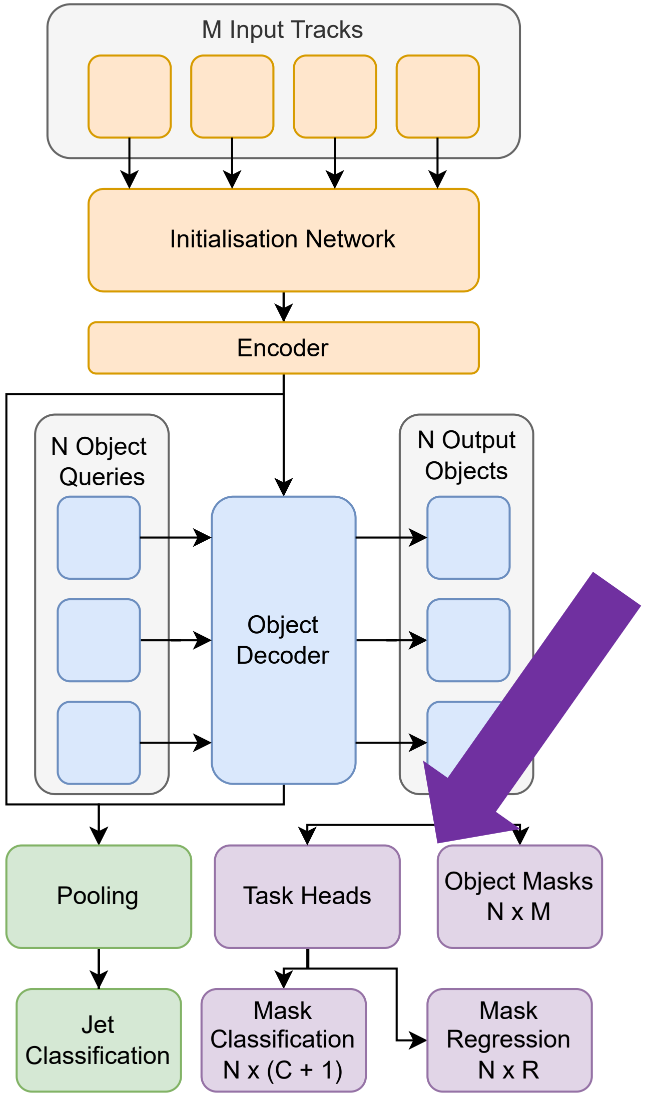
- Utilise M tracks per jet as inputs

MaskFormer in HEP



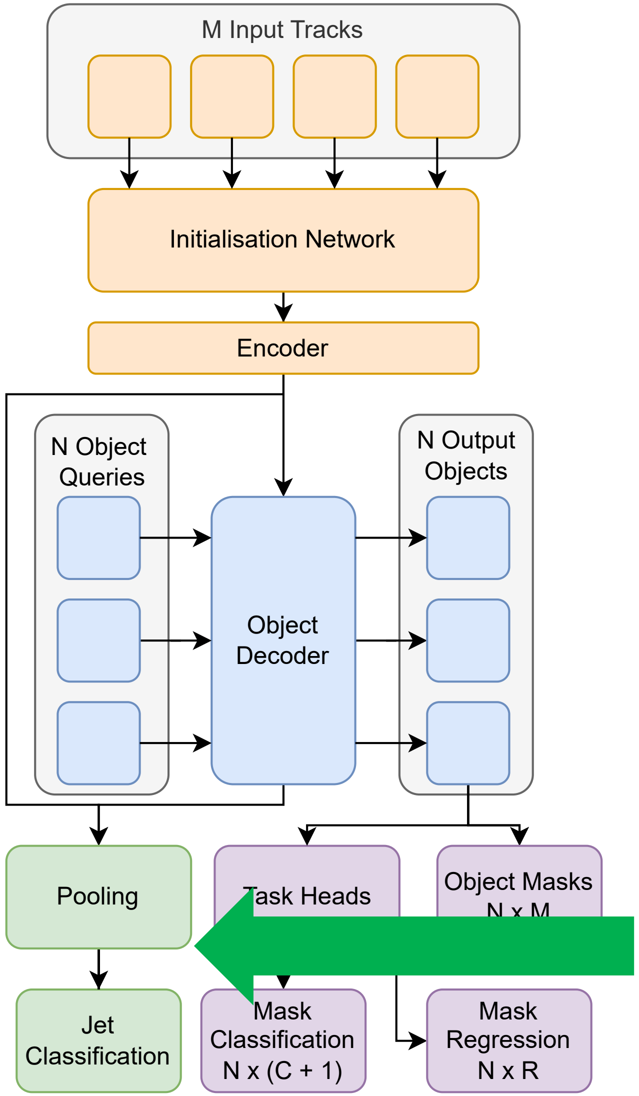
- Utilise M tracks per jet as inputs
- N object queries to generate up to N objects

MaskFormer in HEP



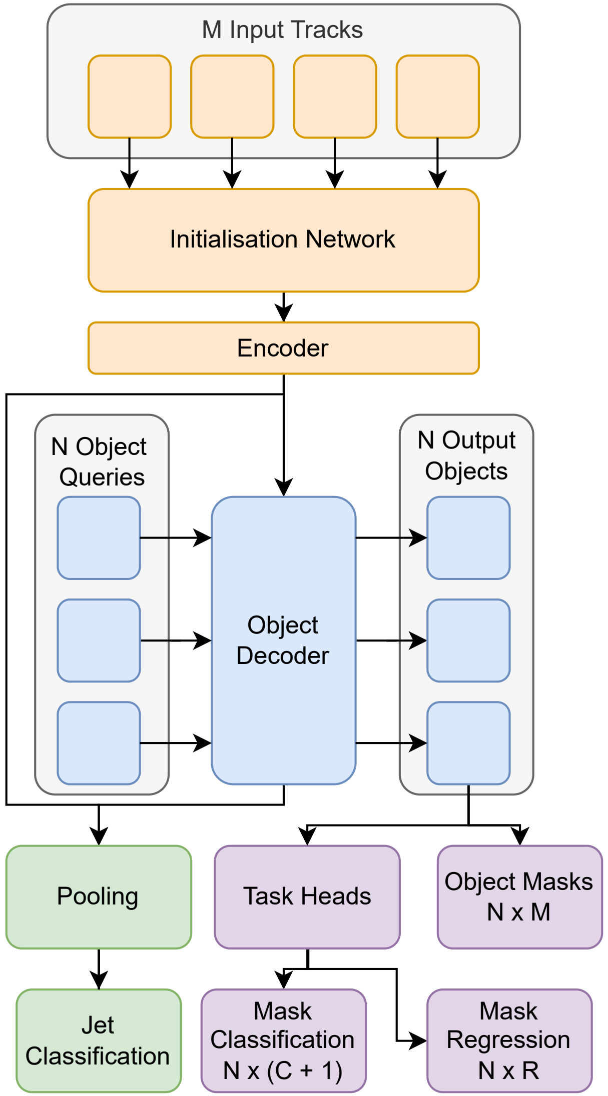
- Utilise M tracks per jet as inputs
- N object queries to generate up to N objects
- Generate masks associating tracks to vertex
- Object task heads
  - Classification: b, $c, null$
  - Regression: $p_T, \Delta R, L_{xy}$

MaskFormer in HEP



- Utilise M tracks per jet as inputs
- N object queries to generate up to N objects
- Generate masks associating tracks to vertex
- Object task heads
  - Classification: b, $c, null$
  - Regression: $p_T, \Delta R, L_{xy}$
- Pooling of decoder queries with encoder nodes
  - Predict jet classification: b, $c, light$

MaskFormer in HEP



Allows multiple-vertex finding, classification, and fitting of **all jet vertices**

- Proof-of-concept results for ML-based **multi-vertex finding + fitting**

Improved scalability:

- Scales subquadratically, $N \times M$, compared to Edge Classifier M^2

Other Approaches To Vertexing

- Hand-crafted (HC) algorithms [1,2]
  - Can be manually modified for many outputs and situations
  - Time-consuming to re-optimize for different detector/environment
- ML Based
  - Edge Classification (EC)
    - Approach used in ATLAS GN1 [3]
    - No direct ability to fit vertices
  - Differentiable Vertex Fitting (NDIVE) [4]
    - Generate and fit a single inclusive vertex
  - Object Condensation (OC) [5]
    - Identify central points and clustering data towards them
  - ***MaskFormer***

	HC	EC	NDIVE	OC	MaskFormer
Learned Vertex Finding	✗	✓	✓	✓	✓
Learned Vertex Fitting	✗	✗	✓	✓	✓
Resolve Multiple Vertices	✓	✓	✗	✓	✓
Predict Truth Properties	✗	✗	✗	✓	✓

[1] *Secondary vertex finding for jet flavour identification with the ATLAS detector*, 2017, The ATLAS Collaboration, ATL-PHYS-PUB-2017-011

[2] *Topological b-hadron decay reconstruction and identification of b-jets with the JetFitter package in the ATLAS experiment at the LHC, 2018*, The ATLAS Collaboration, ATL-PHYS-PUB-2018-025

[3] *Graph Neural Network Jet Flavour Tagging with the ATLAS Detector,* 2022, The ATLAS Collaboration, ATL-PHYS-PUB-2022-027

[4] *Differentiable Vertex Fitting for Jet Flavour Tagging*, 2023, R. E. C. Smith et al., arXiv:2310.12804

[5] *Object condensation: one-stage grid-free multi-object reconstruction in physics detectors, graph, and image data*, 2020, Jan Kieseler, arXiv:2002.03605

Details

- Simulated $t\bar{t}, \sqrt{s} = 13$ TeV, events and parton shower in Pythia 8
- Detector response in Delphes, based on ATLAS detector
- Models trained on 13.5m jets, $20 < p_T < 250$ GeV, $|\eta| < 2.5$
  - Match kinematic distributions between jet flavours
- Input track features
  - 5 global track parameters + values relative to jet axis
- Vertexing performed only on heavy flavour (HF – containing b/c-quarks) hadrons:
  - Require truth hadron $p_T > 5$ GeV, $\Delta R(Jet, Hadron) < 0.4$
  - Herein described as *Hadron Finding/Fitting*
- All models utilise 6.3m trainable parameters

Definitions

- Define HF hadron finding efficiency for a loose recall/precision criteria

$$\epsilon = \frac{N_{HF-Tag,\geq 2\,Tracks}}{N_{HF-Total,\geq 2\,Tracks}} \qquad R = \frac{TP}{TP+FN} \geq 65\% \qquad P = \frac{TP}{TP+FP} \geq 50\%$$

- And for perfect reconstruction

$$R = 100\% \qquad P = 100\%$$

- Define Fake Rate as number of fake HF Hadrons predicted in light-jets

$$F = \frac{N_{HF-Fake\geq 2Tracks}}{N_{Light-Jets}}$$

- Compare to base-line: Transformer *EdgeClassifer*
    - Generates edge-relation score for each track pair
    - Group into vertices via Union Find as postprocessing
    - Based on ATLAS GN1
        - Shown to outperform hand crafted algorithms for vertex finding [1]

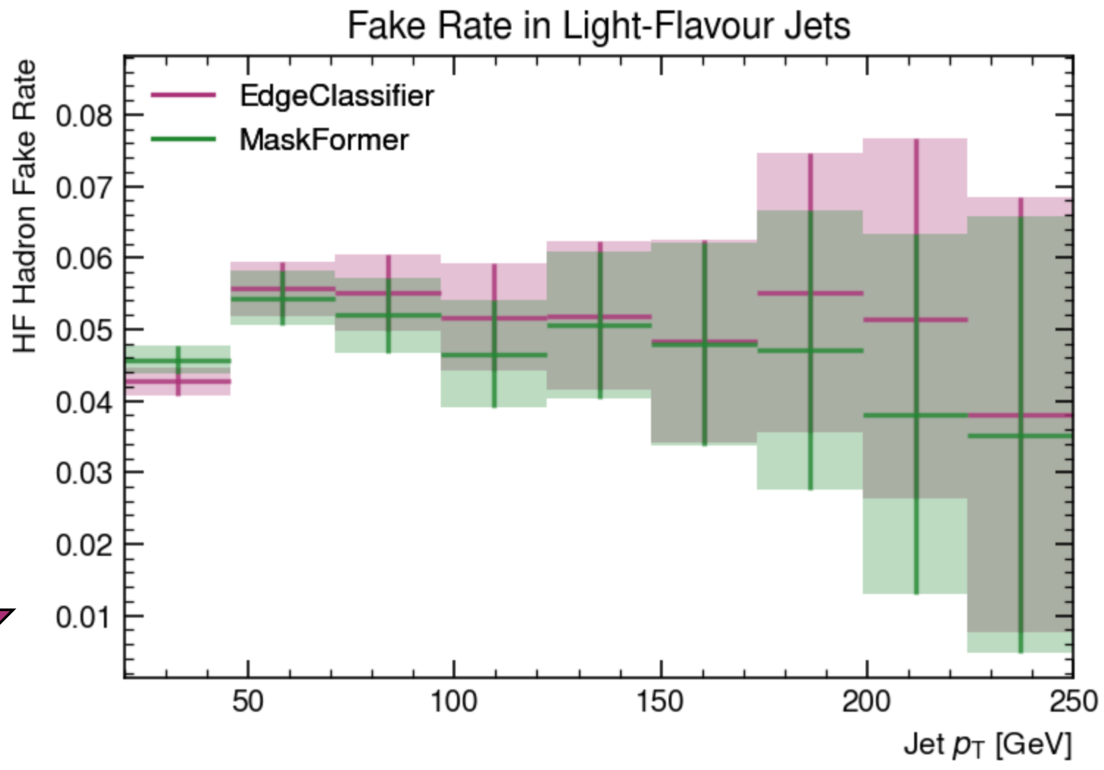
[1] Graph Neural Network Flavour Tagging and Boosted Higgs Measurements at the LHC, 2023, S. Van Stroud, CERN-THESIS-2023-142

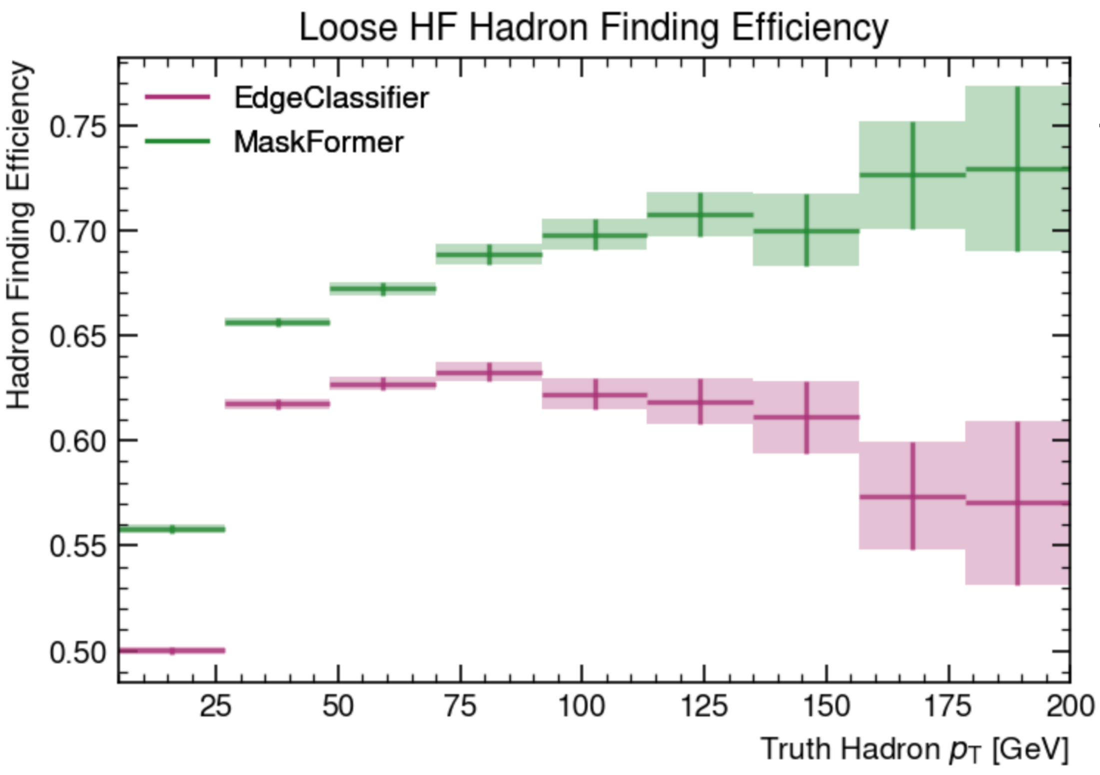
Hadron Finding Efficiency

MaskFormer is tuned to have equivalent fake rate to base-line **EdgeClassifier**

With **10% increase** for loose hadron finding efficiency in all HF-jets

Lower is better!



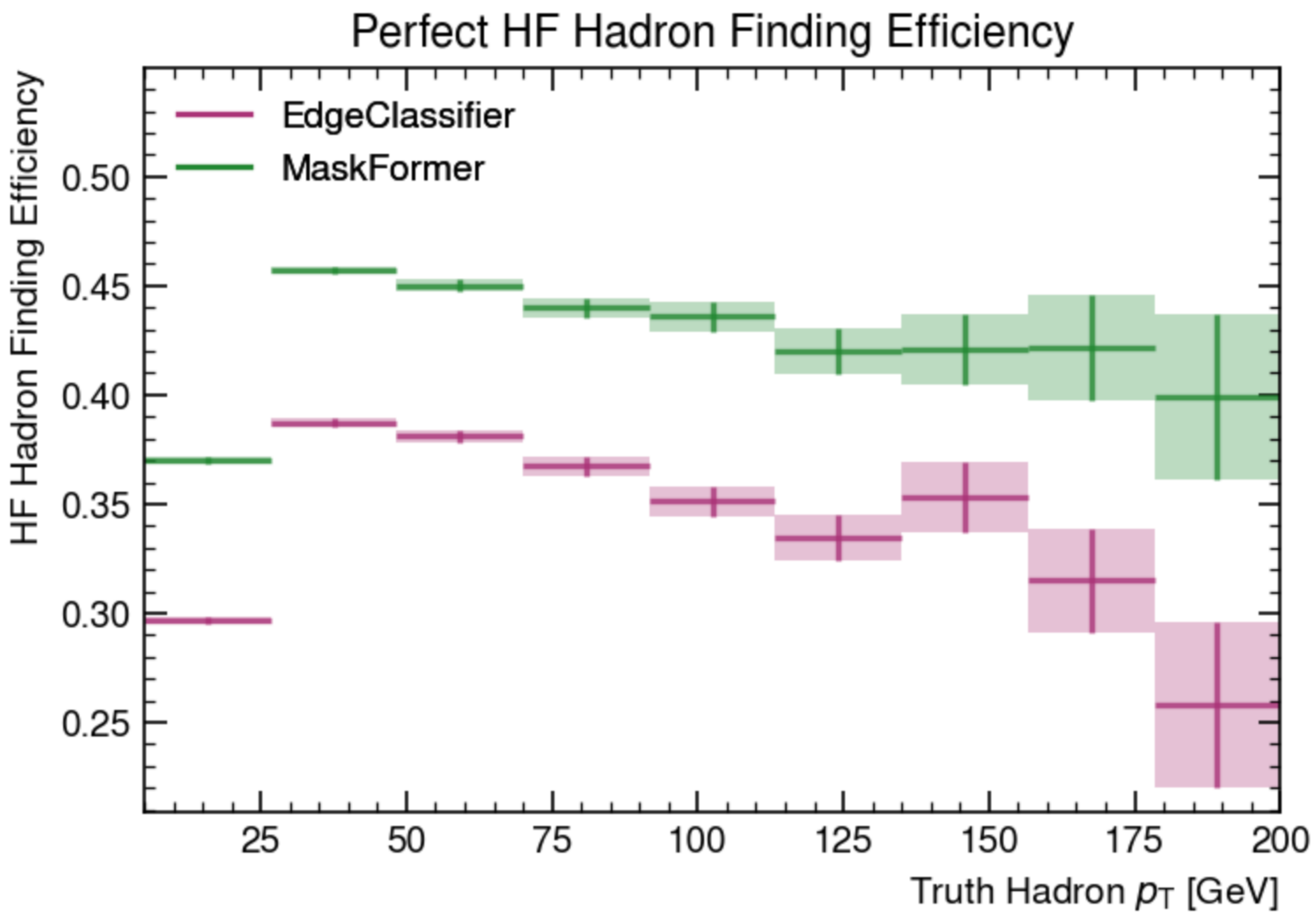


Higher is better!

Perfect Hadron Finding Efficiency

- Absolute increase in efficiency of $\mathbf{5-10\%}$
  - Relative increase of $\mathbf{15-20\%}$

- **MaskFormer** consistently achieves improved performance compared to EdgeClassifer for an equivalent fake rate

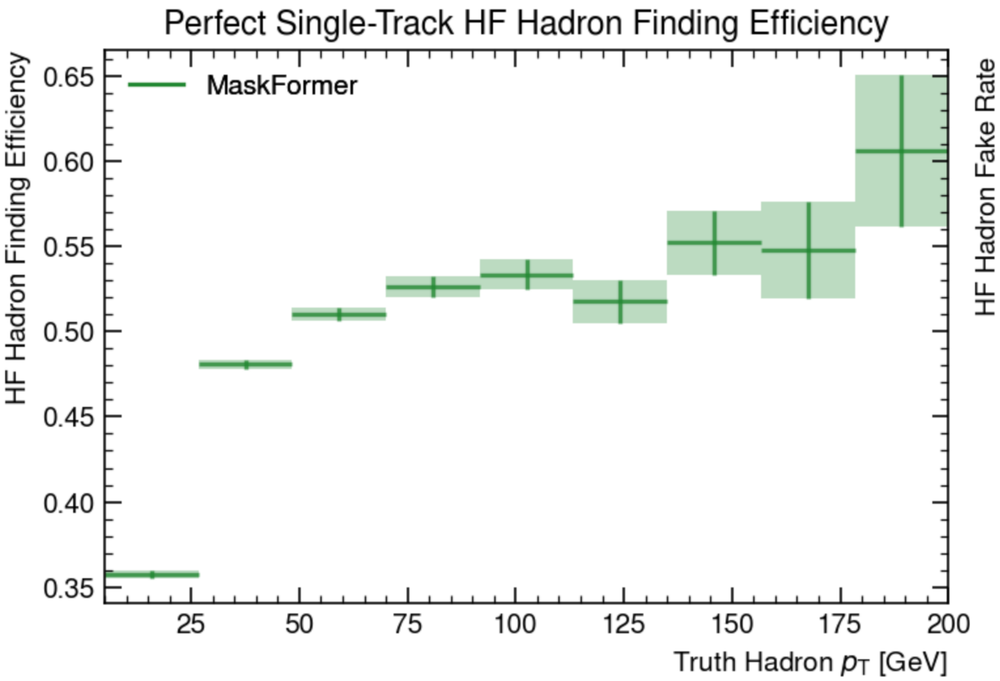


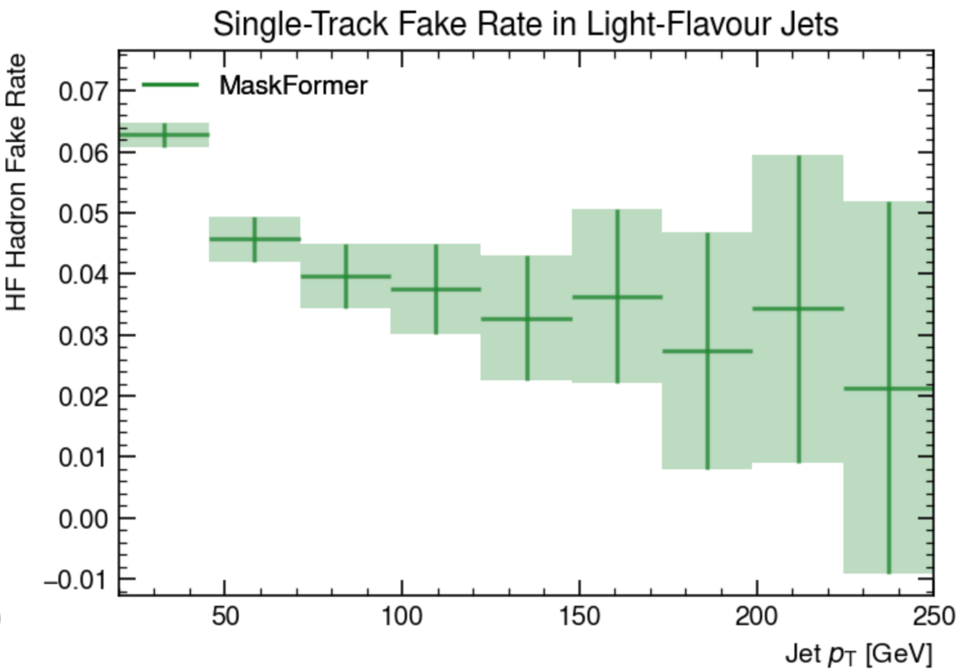
MaskFormer Single-Track Performance

- MaskFormer can successfully attribute individual tracks to single-track vertices
  - In this configuration, not possible for Edge Classifier
- Correctly assigns single-track HF Hadrons perfectly $\approx 50\%$ of the time
  - Only 5% fake-rate

$$\epsilon = \frac{N_{HF-Tag,=1\,Track}}{N_{HF-Total,=1\,Track}}$$

$$F = \frac{N_{HF-Fake,=1\,Track}}{N_{Light-Jets}}$$



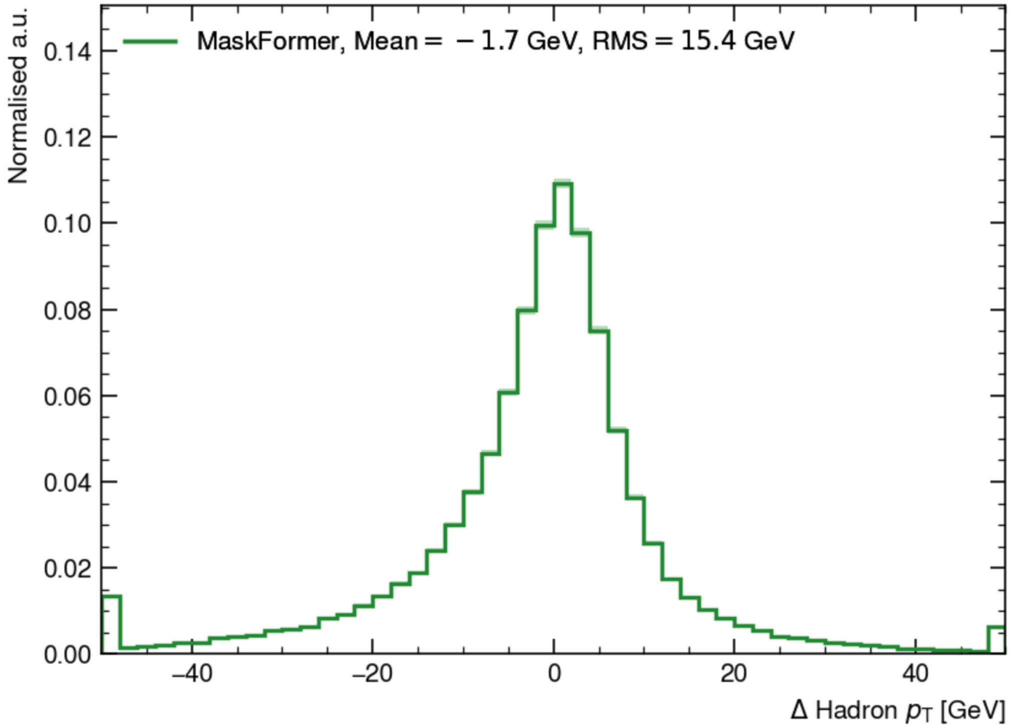


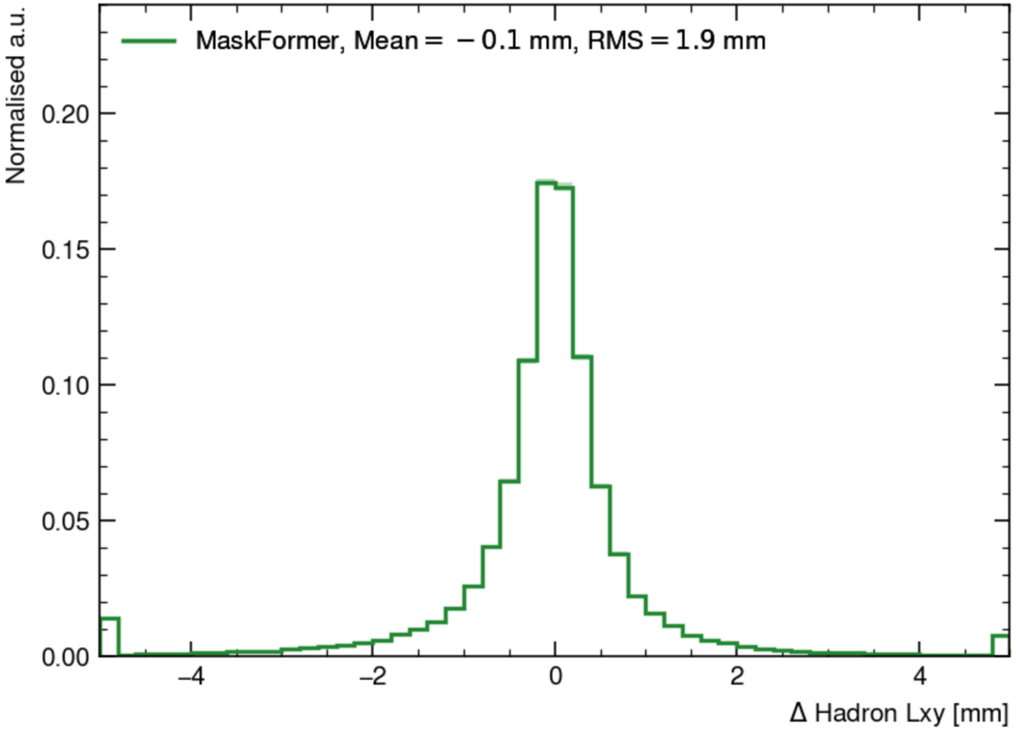
Hadron Fitting

- Reconstruct hadron properties directly as regression targets
- Require MaskFormer to predict mask as b/c-hadron
- Regress hadron properties to truth level

$$\Delta X = X_{pred} - X_{truth}$$

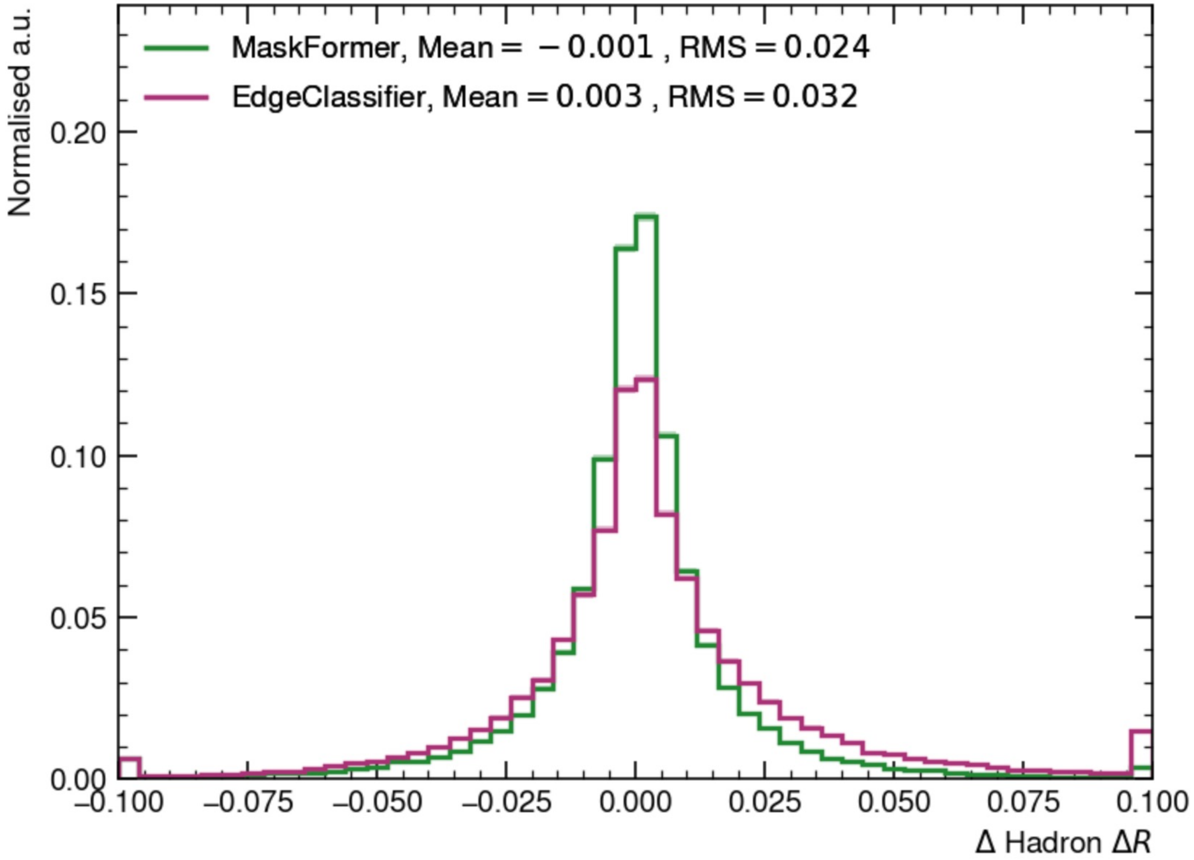
For property X





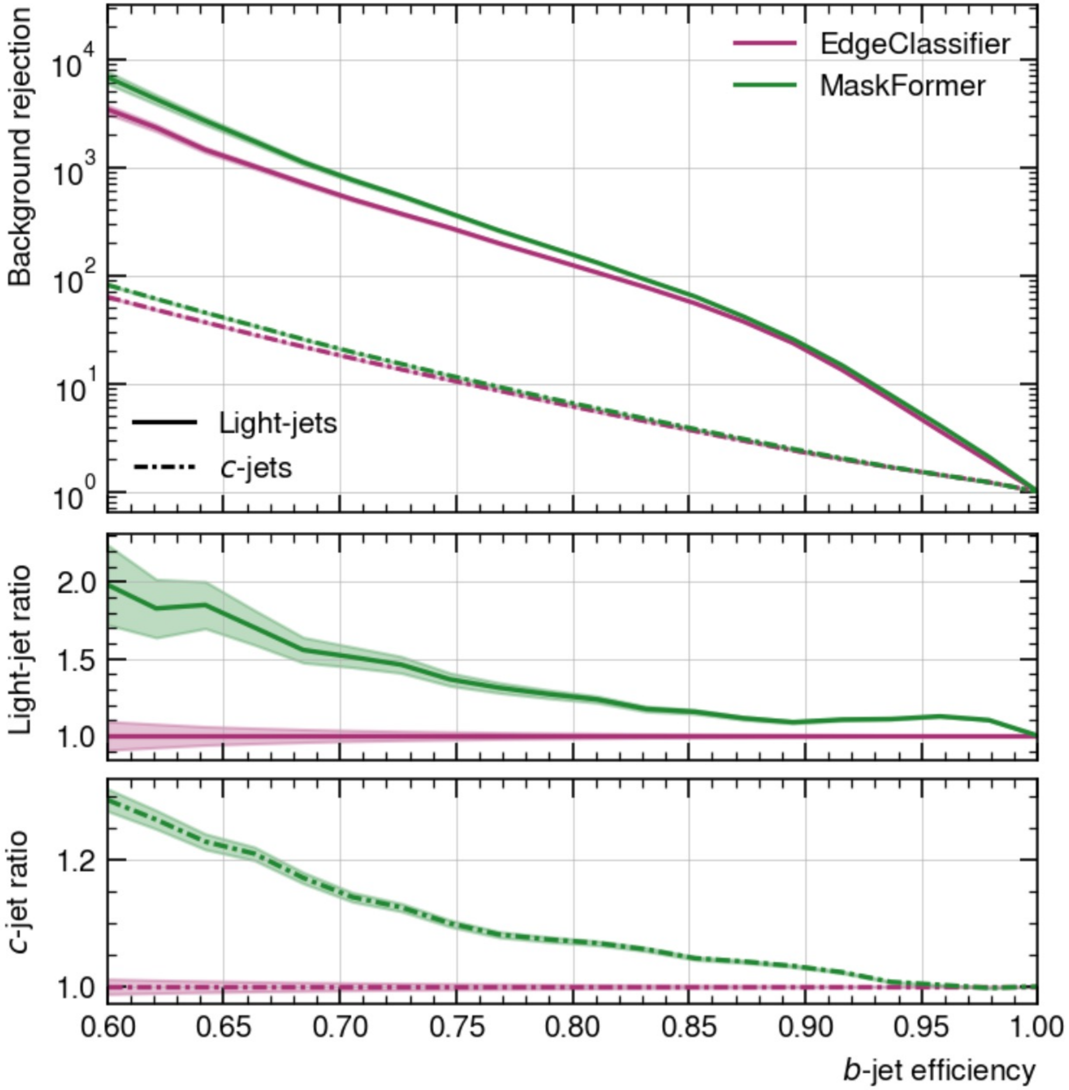
Hadron Fitting

- MaskFormer directly regresses hadron properties
- Can reconstruct some properties with Edge Classifier out of the box
  - ΔR reconstructed from sum 4-vector of tracks associated to hadron
  - No explicit latent representation of hadrons
    - Can't directly reconstruct other properties at a per-hadron level
  - **Less accurate** predictions than **MaskFormer**
- **MaskFormer** easily extendable to other classification or regression targets



Improvements to b-tagging

- Decay topology vital component for b-tagging
- Track-hadron association and hadron regression/classification aids b-tagging
- Note
  - Performance not directly comparable to standard taggers (*GN1* or *ParticleNet* [1])
  - We utilise simplified track input features
  - Reduced set of auxiliary training tasks



[1] *ParticleNet: Jet Tagging via Particle Clouds*, 2012, H. Qu, L. Gouskos, arXiv:1902.08570

Conclusion and Further Work

- We present proof-of-concept preliminary results for
  MaskFormer for HF Vertex Reconstruction
- Allows explicit multi-object reconstruction for multiple-inputs
  - Transformer based encoder/decoder architecture
- Allows multiple HF hadronic vertices to be reconstructed
  - Competitive HF-hadron finding performance over range of HF targets
  - Robust predictions of HF-hadron properties
- Improved b-tagging performance
- Concept of *'M-to-N'* reconstruction may have wide use case in HEP
  - Easily extendable to other tasks

Thank you for listening!

Any questions?

Backup – Regression Distributions and Relative Differences

Define relative residual as

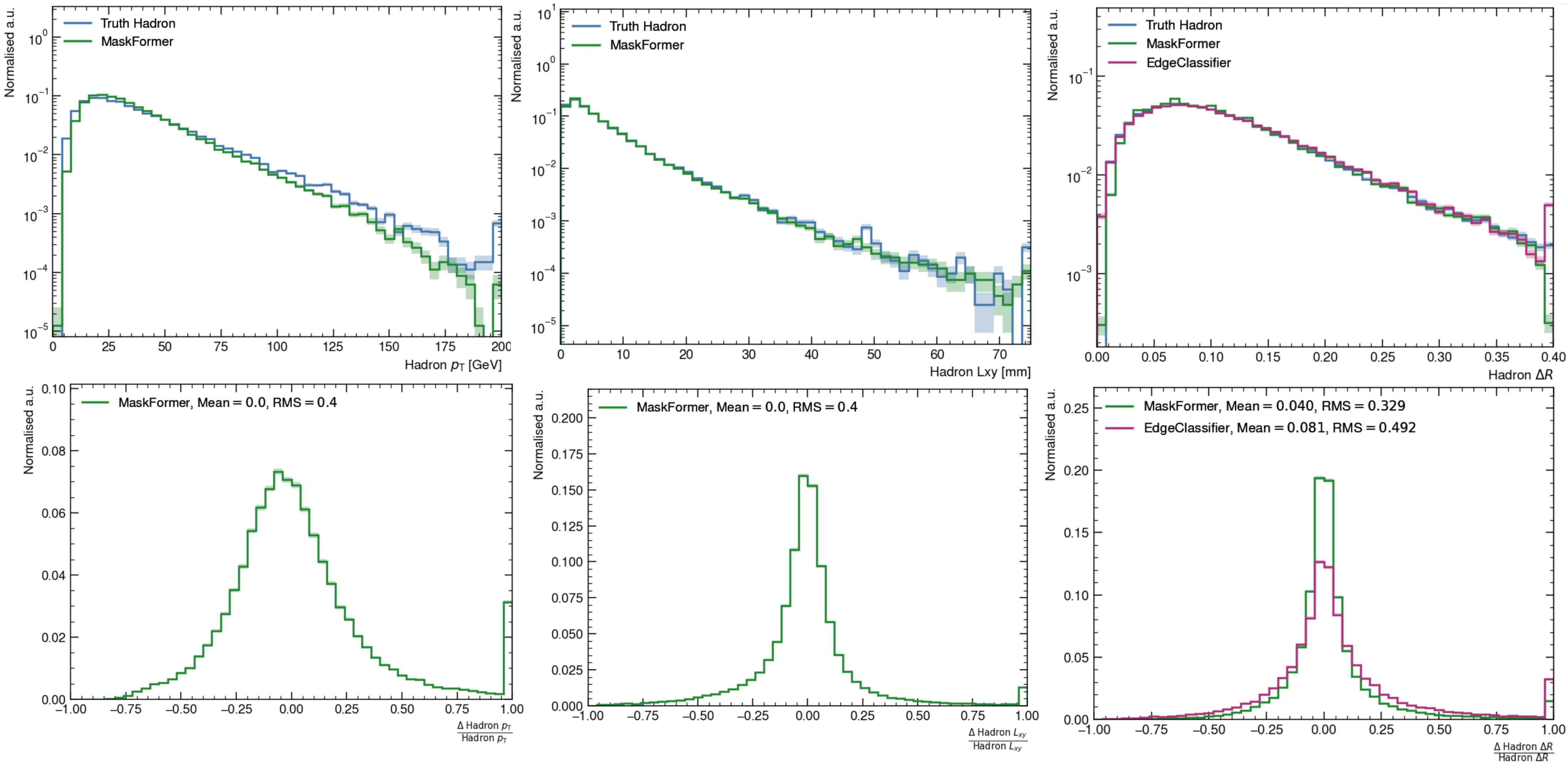
$$\frac{X_{pred} - X_{truth}}{X_{truth}}$$

For relative Plots

$$L_{xy} \geq 1mm$$

$$\Delta R \geq 0.01$$

Bulk +-25%
Outliers cause
Larger RMS



Backup- Edge-Classifier Model

- Edge Classifier consists of encoder only architecture
  - Concatenate encoder embedding for $[Node_i, Node_j]$
  - Predict edge relationship E_{ij}
- In postprocessing
  - Symmetrise $\boldsymbol{E}$
  - Apply threshold cut
  - Run Union Find to group edges into vertices
    - Associate vertices to truth hadron for performance comparison