

# **Point Set Transformers for NOvA Multi-View Particle Segmentation**

**Edgar Robles, Dikshant Sagar, Alejandro Yankelevich,  
Jianming Bian, Pierre Baldi, for the NOvA Collaboration**



# Overview

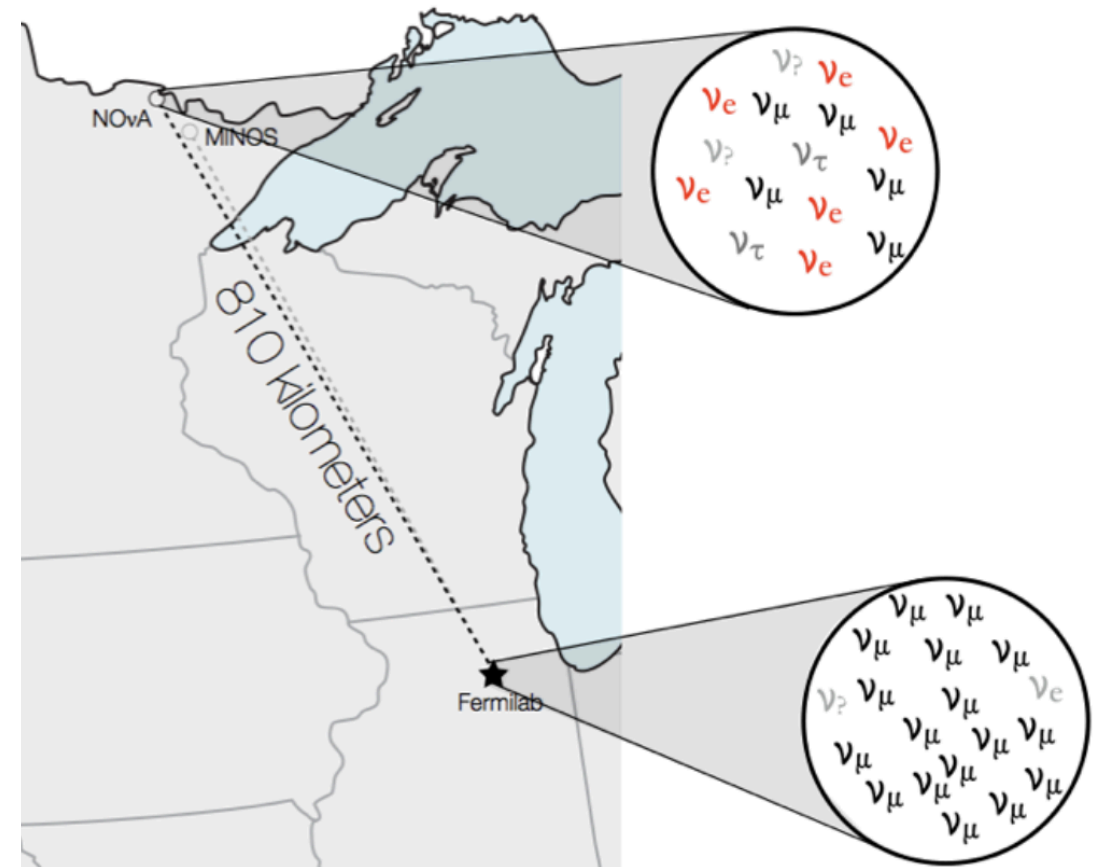
- NOvA Experiment
- Machine Learning in Reconstruction Chain
- Particle Segmentation
  - Current Methods
  - Motivation for Point Set Transformers
  - Architecture and Heterogeneous Attention
  - Performance



# NOvA

## Overview

- Long-baseline neutrino oscillation experiment.
- Fermilab's NuMI beam is predominately  $\nu_\mu$ .
- Two functionally equivalent detectors spaced 810 km apart.
- Observe  $\nu_\mu$  disappearance and  $\nu_e$  appearance to measure:
  - $\Delta m_{32}^2$  and mass ordering
  - $\sin^2(\theta_{23})$  and  $\theta_{23}$  octant
  - CP-violating phase  $\delta_{CP}$

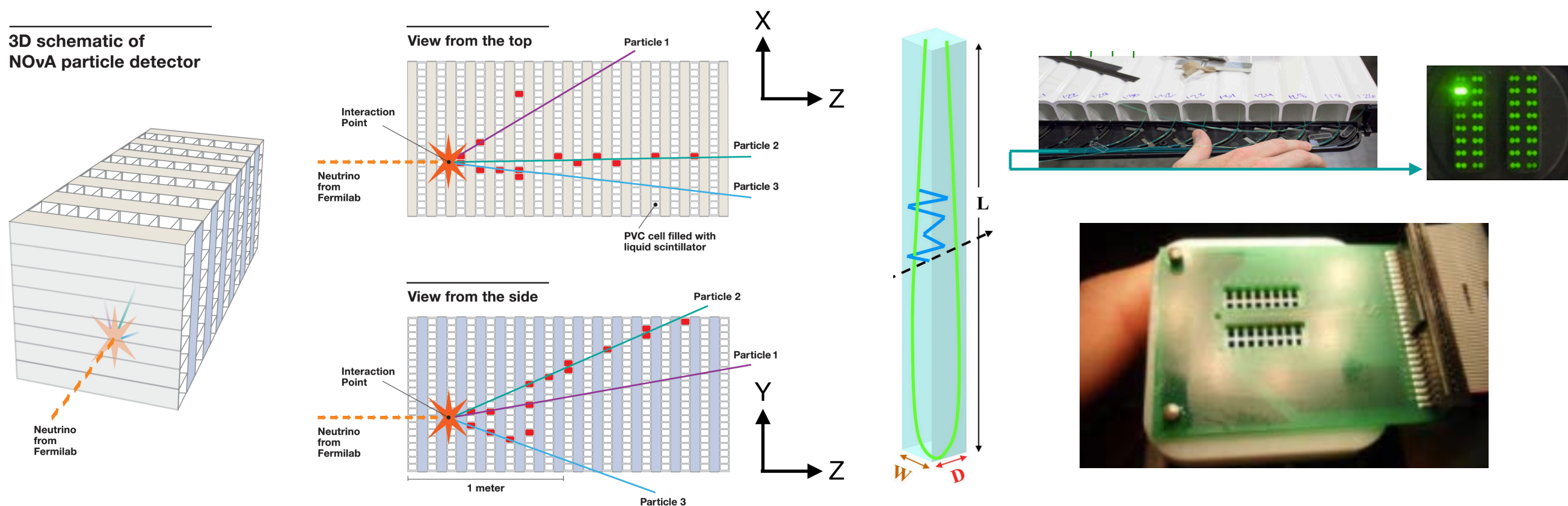




# NOvA

## Detector Technology

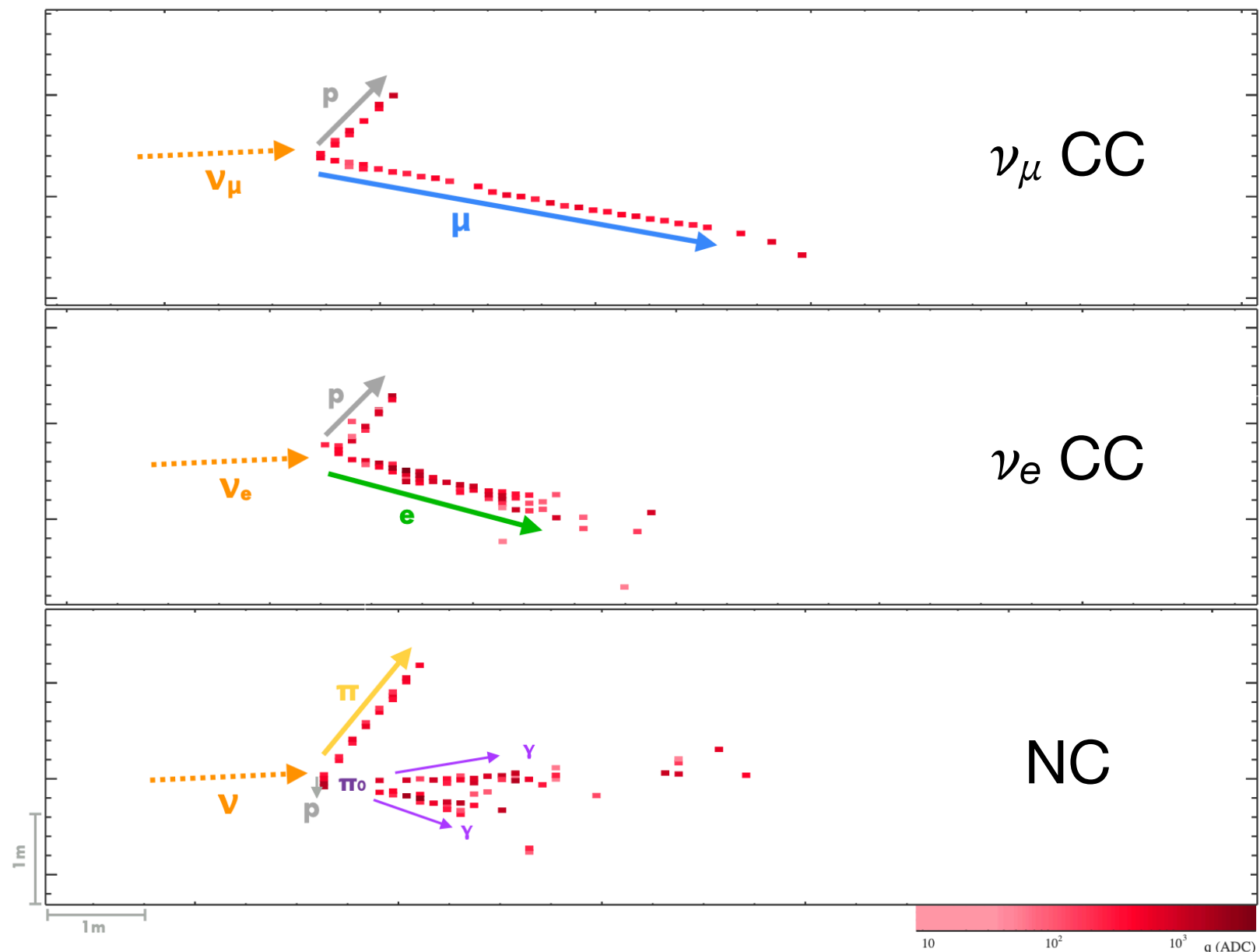
- Long 4cm x 6cm PVC extrusions filled with oil-based liquid scintillator. Wavelength shifter fiber through each cell leads to avalanche photodiode.
- Planes alternate in horizontal and vertical orientations to create two XZ and YZ views when viewed from the top or side.
- These views are fundamentally 2D. The Y position of a hit in a vertical cell cannot be known and has no correspondence to any hit in a horizontal cell.



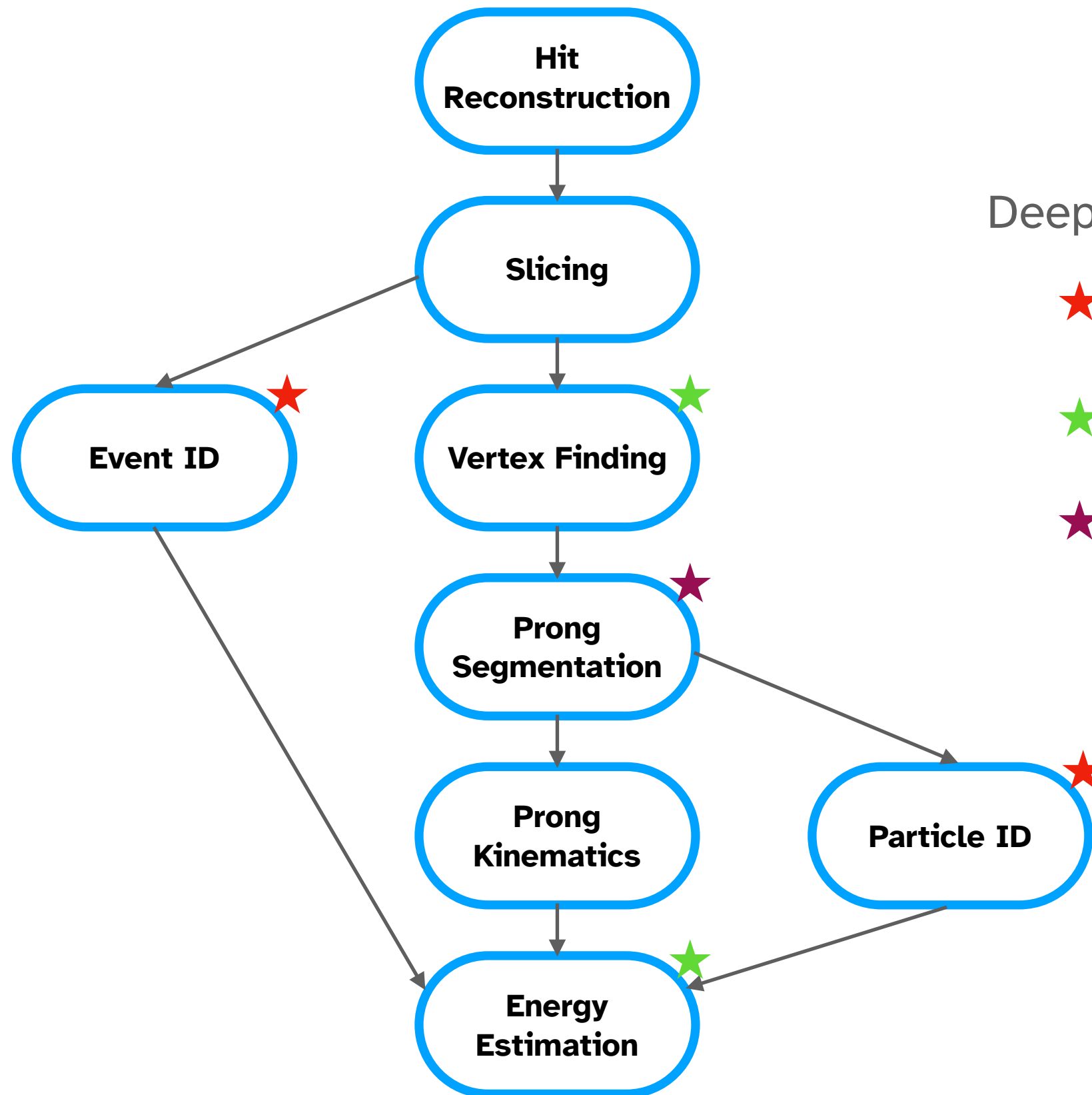
# NOvA

## Event Topologies

- Neutrino flavor can be determined with charged current (CC) interactions via primary lepton but not neutral current interactions (NC).
- Far detector on surface, so large cosmic ray background, but beam spill timing removes majority.



# Reconstruction Chain



Deep learning method...



Used in published oscillation analysis



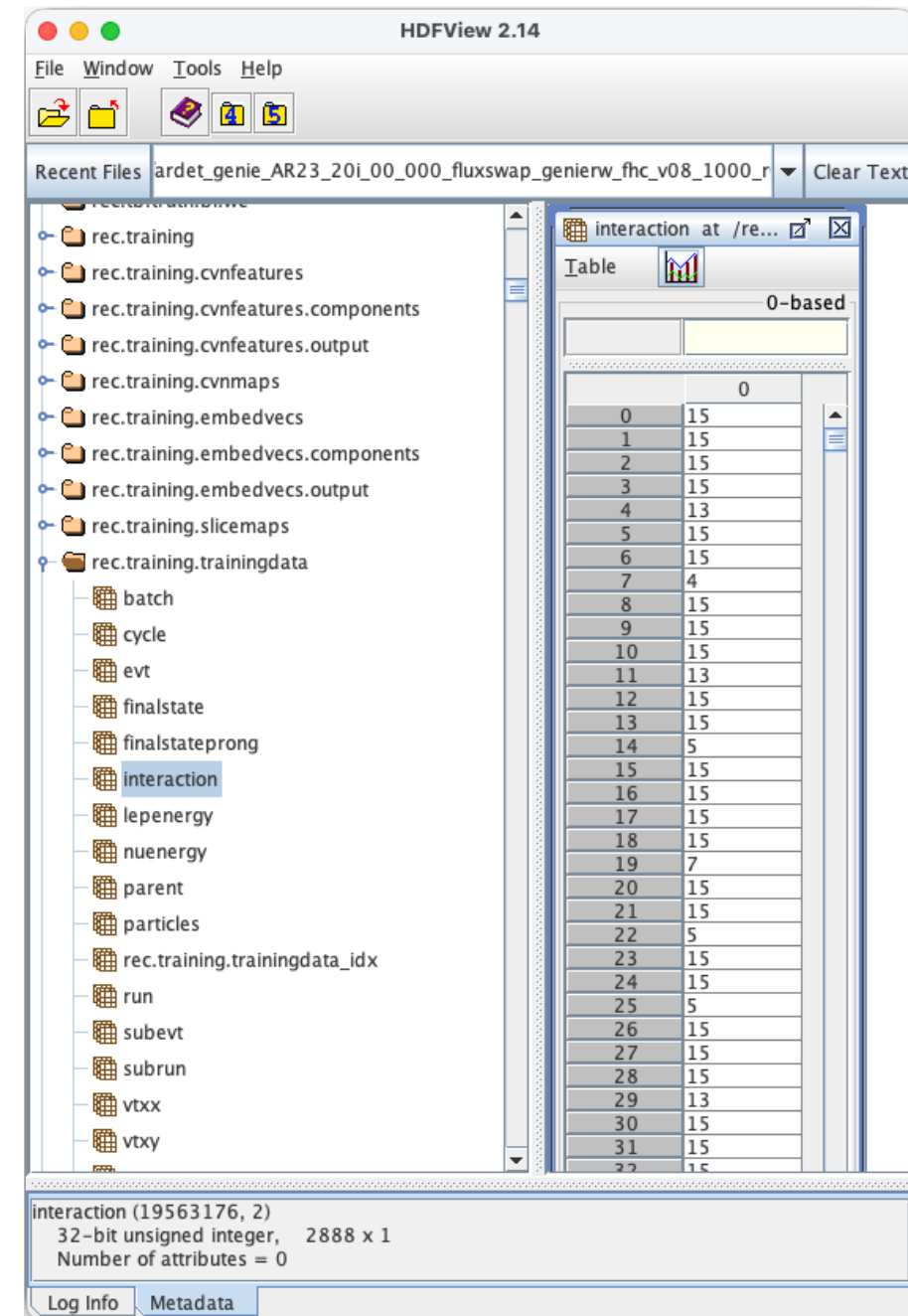
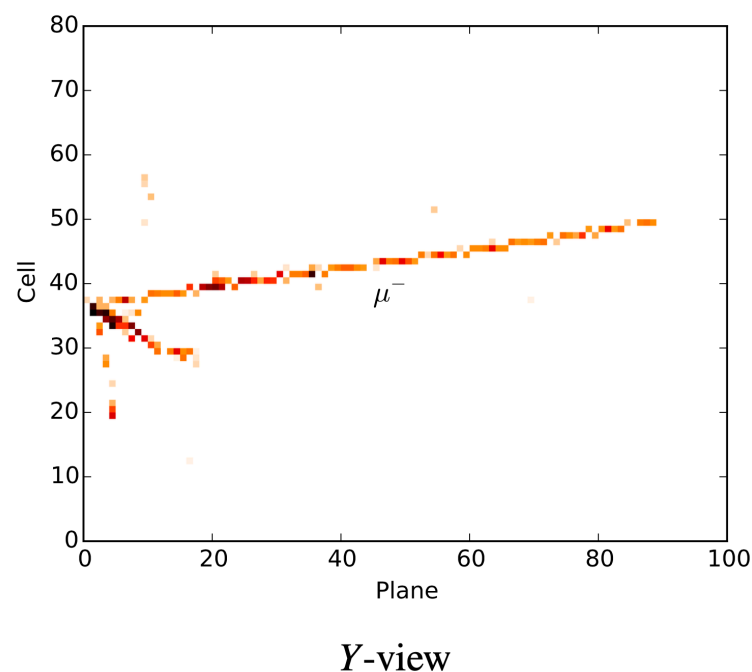
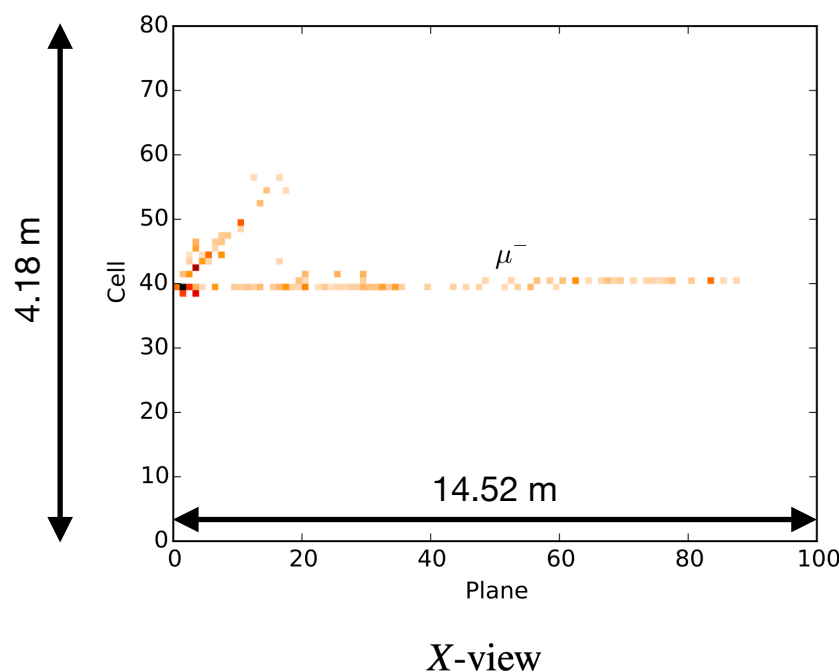
Ready to be used in next analysis



In development

# ML Framework

- Standard 2 x 100 x 80 images (“pixel maps”) cropped around neutrino interaction used for all image-based ML at NOvA.
  - One 8-bit channel per pixel: deposited charge 0-278 MeV
- H5CAFs: Common Analysis Framework (CAF) files in ROOT format converted to HDF5 for easy ML development.
  - Includes pixel maps and truth information in format suitable for training.

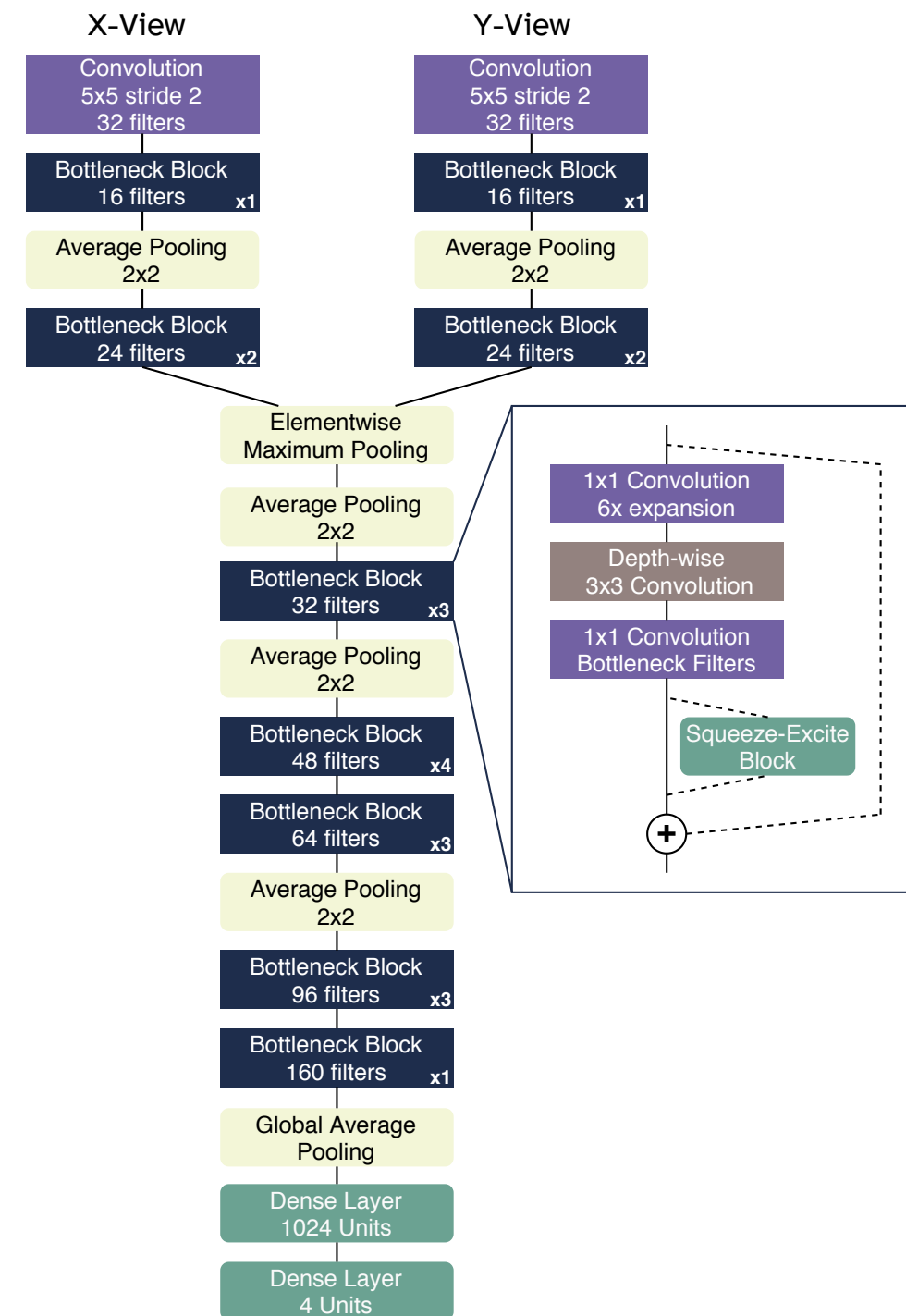
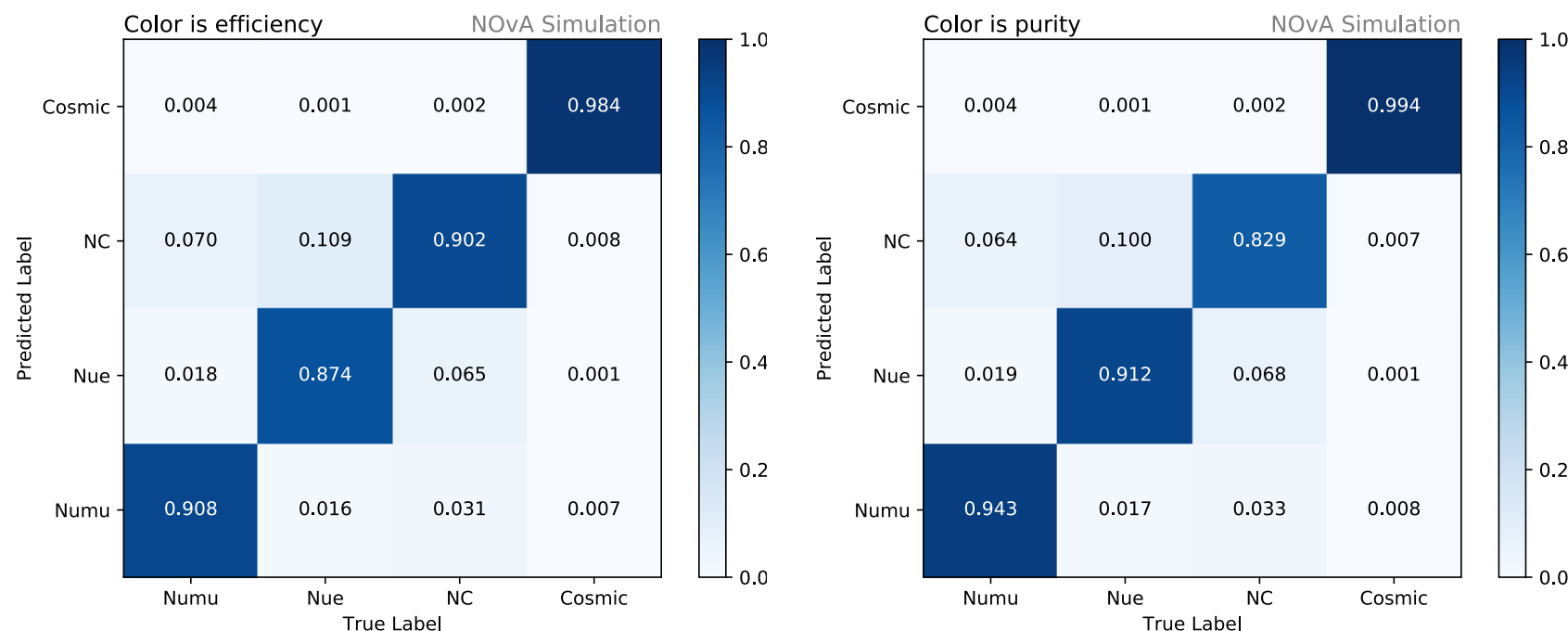
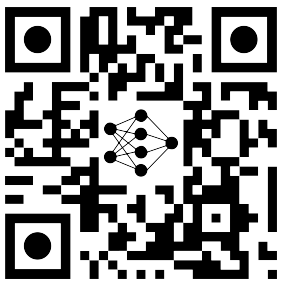


# EventCVN

## CNN for Event Classification

- First CNN used in major HEP analysis.
  - NOvA oscillation analysis.  
*Phys. Rev. Lett.* **118** 231801 (2017)
- 2-view Siamese CNN with MobileNetv2 architecture.

Can you beat  
the CVN? game

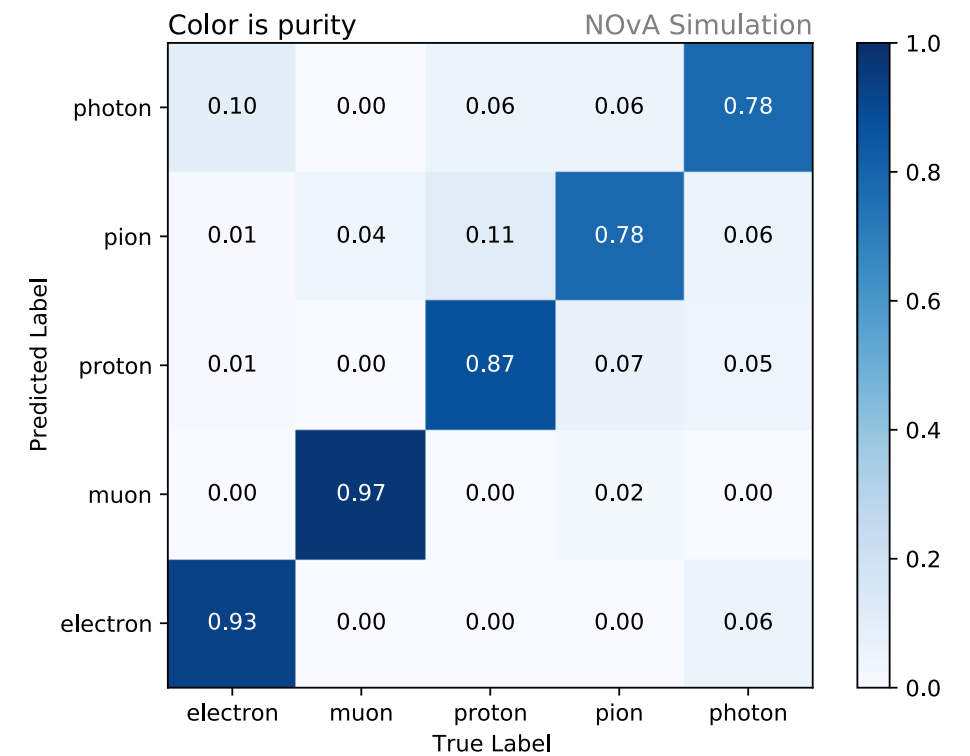
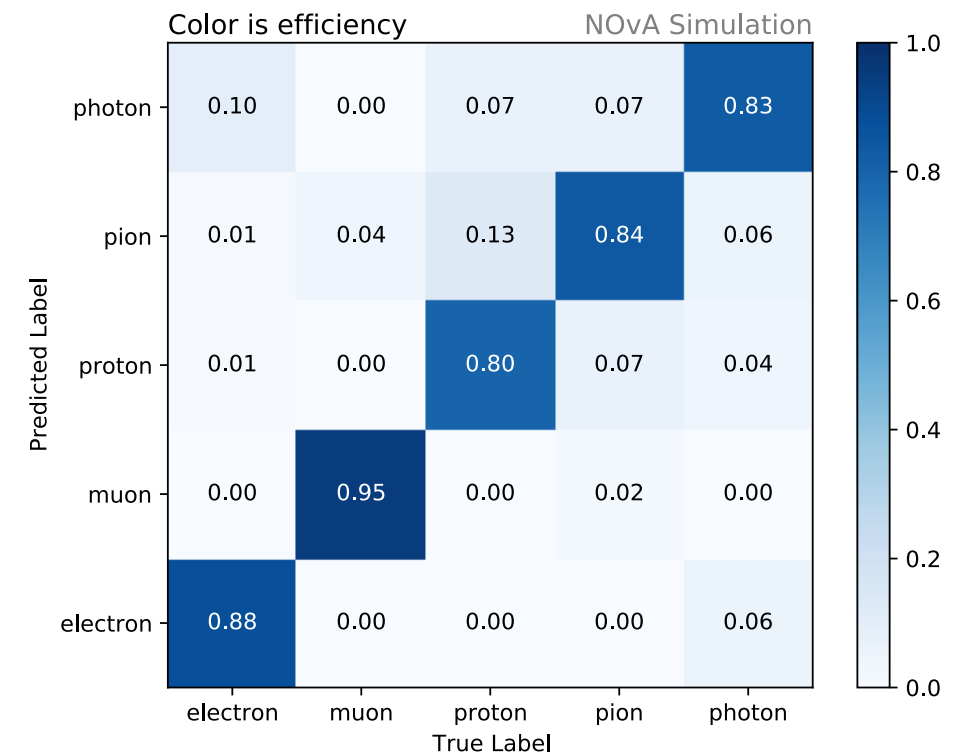
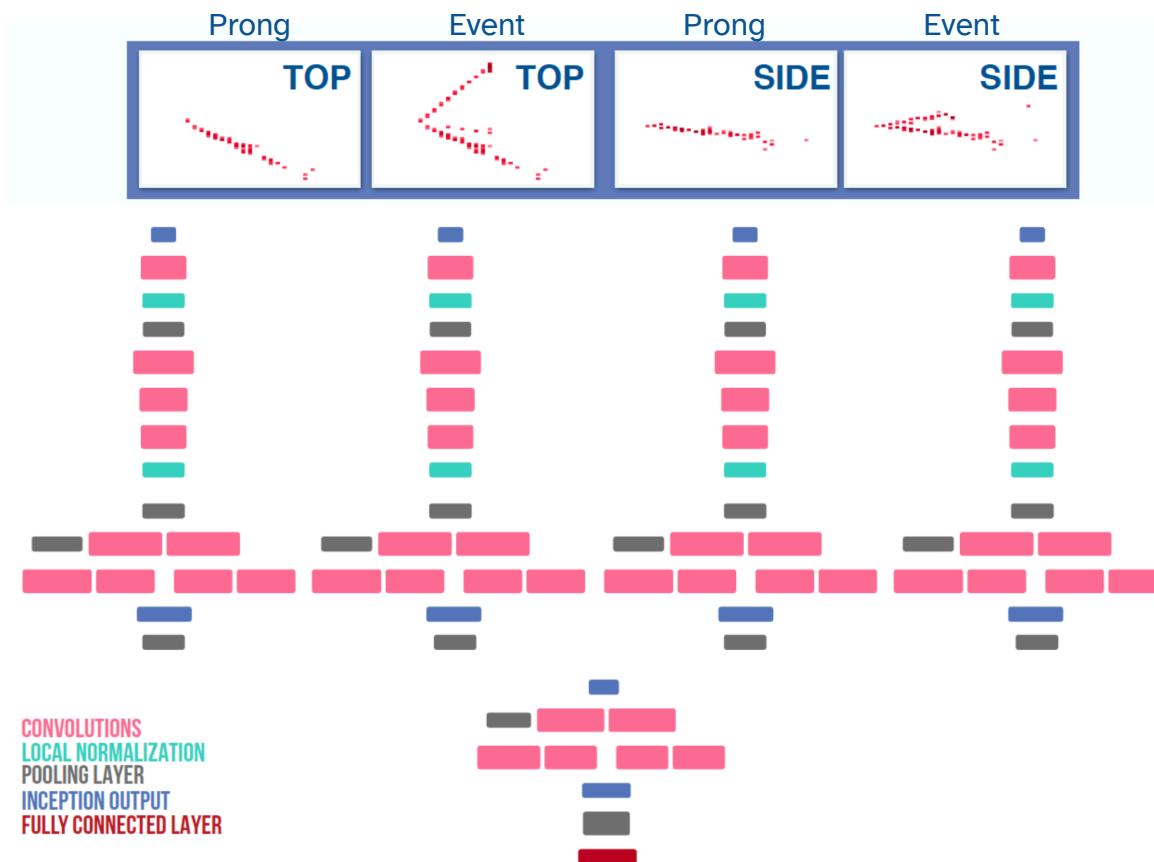


Methods paper: [JINST 11 P09001 \(2016\)](#)

# ProngCVN

## CNN for Prong Classification

- After prong segmentation, create pixel maps for each prong with only hits assigned to prong.
- 4-view Siamese CNN with both event and prong pixel maps.



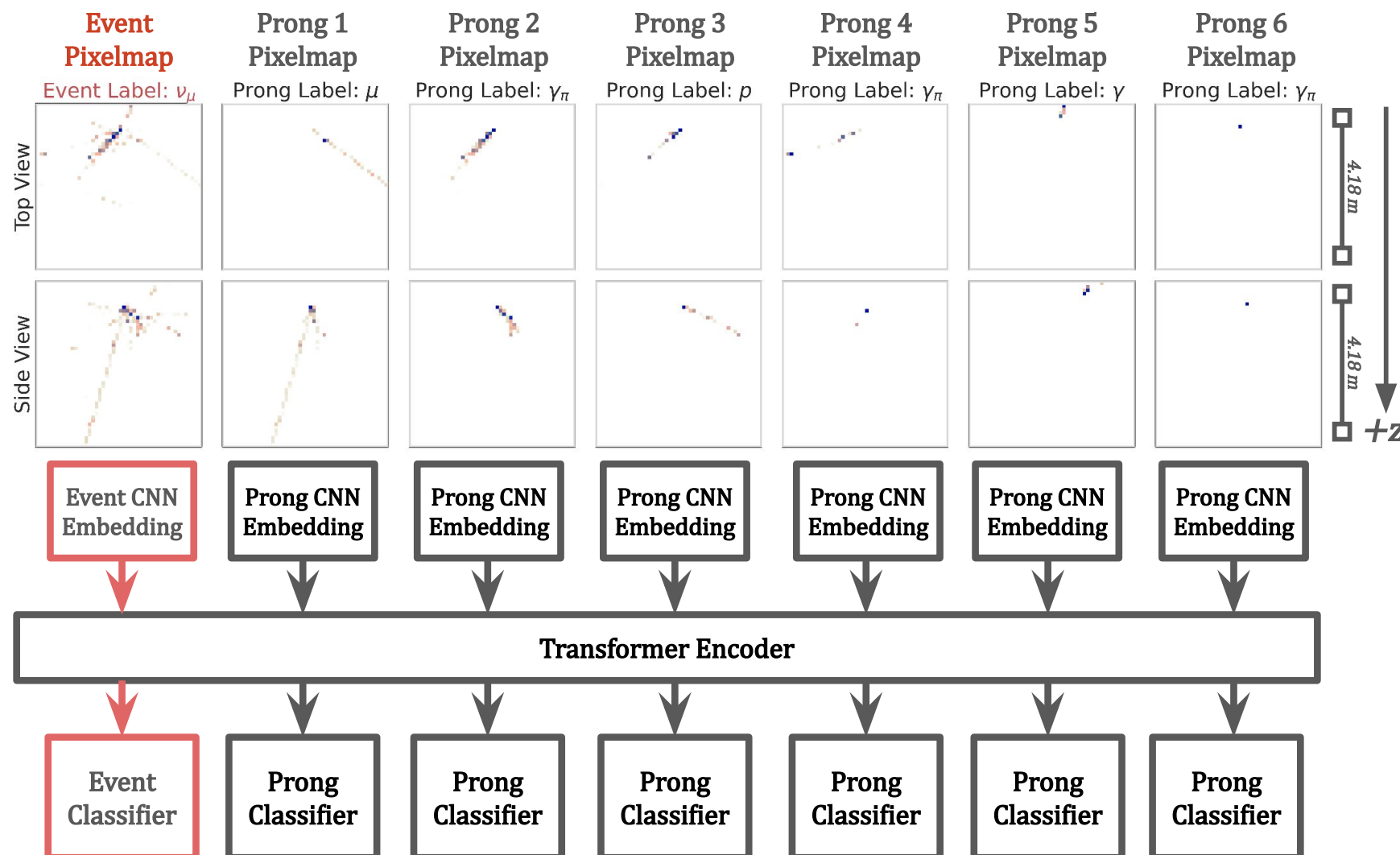
Methods paper: [Phys. Rev. D 100 073005](#)



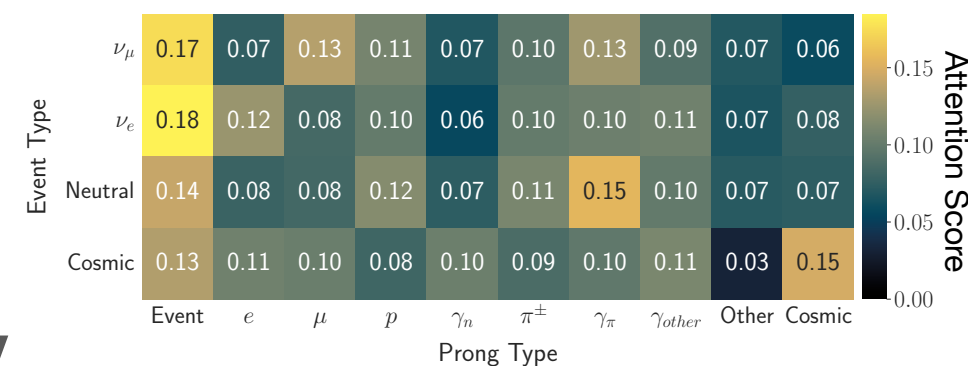
# TransformerCVN

## Transformer for Joint Event+Particle Classification

- Simultaneous predictions with weight sharing for increased interpretability.
- 2 views are treated as channels (100 x 80 x 2).



Prong Influence on Event Classification



| Metric    | Event |       | Prong |       |
|-----------|-------|-------|-------|-------|
|           | T-CVN | E-CVN | T-CVN | P-CVN |
| Accuracy  | 0.894 | 0.897 | 0.783 | 0.726 |
| Precision | 0.894 | 0.908 | 0.783 | 0.760 |
| ROC AUC   | 0.982 | 0.984 | 0.951 | 0.932 |

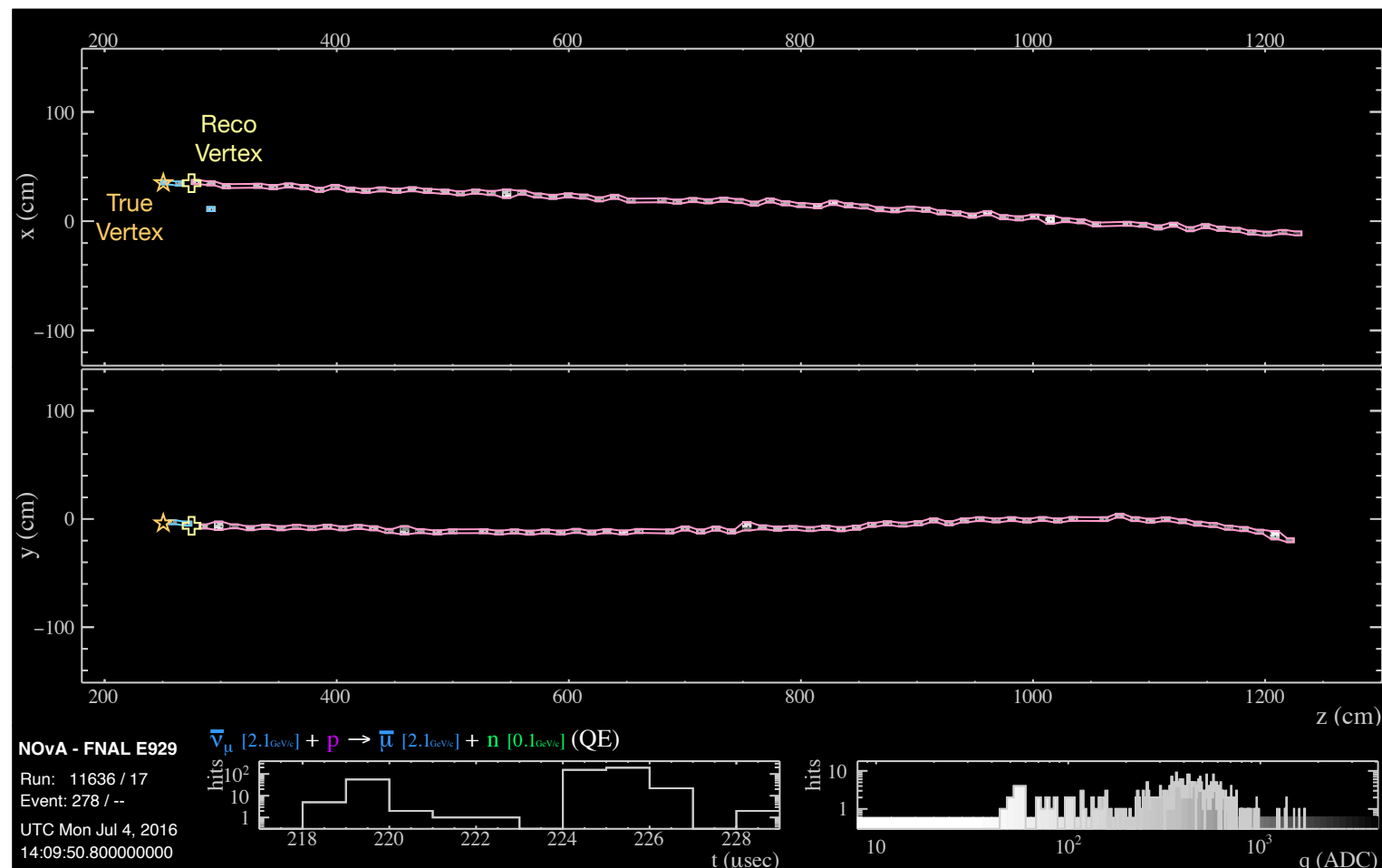
Methods paper: [ML4PS 2023 168](#)



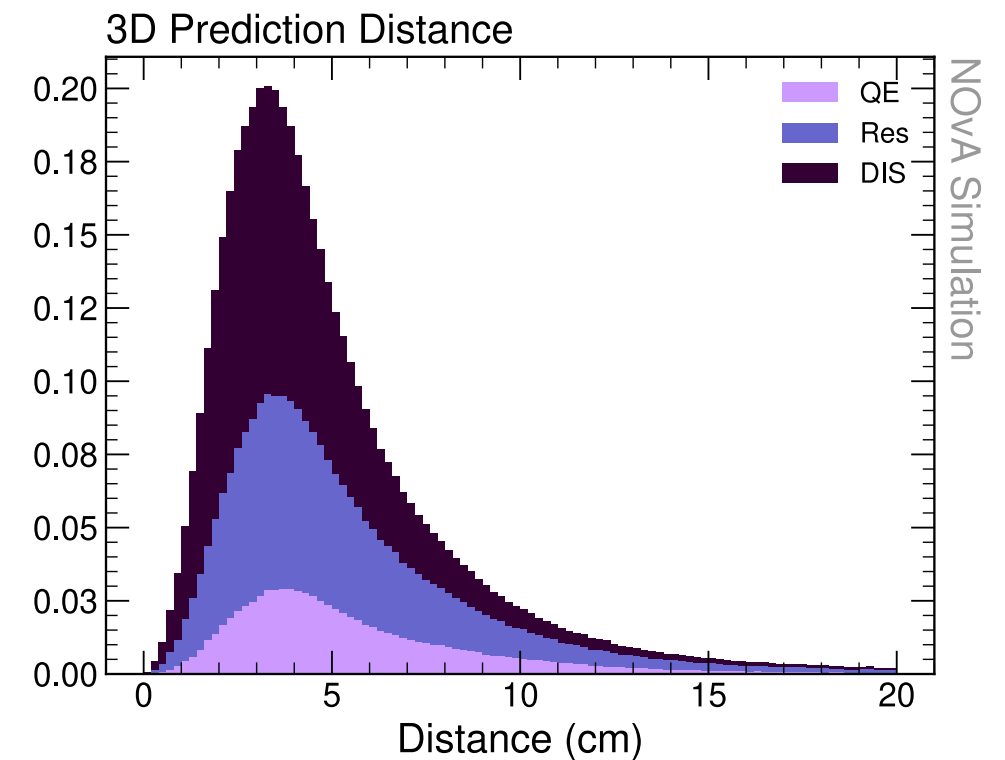
# VertexCVN

## CNN for Vertex Regression

- EventCVN architecture modified for vertex regression.
- Improves vertex resolution to within cell width.



Example event with split muon track due to traditional method misplaced vertex.

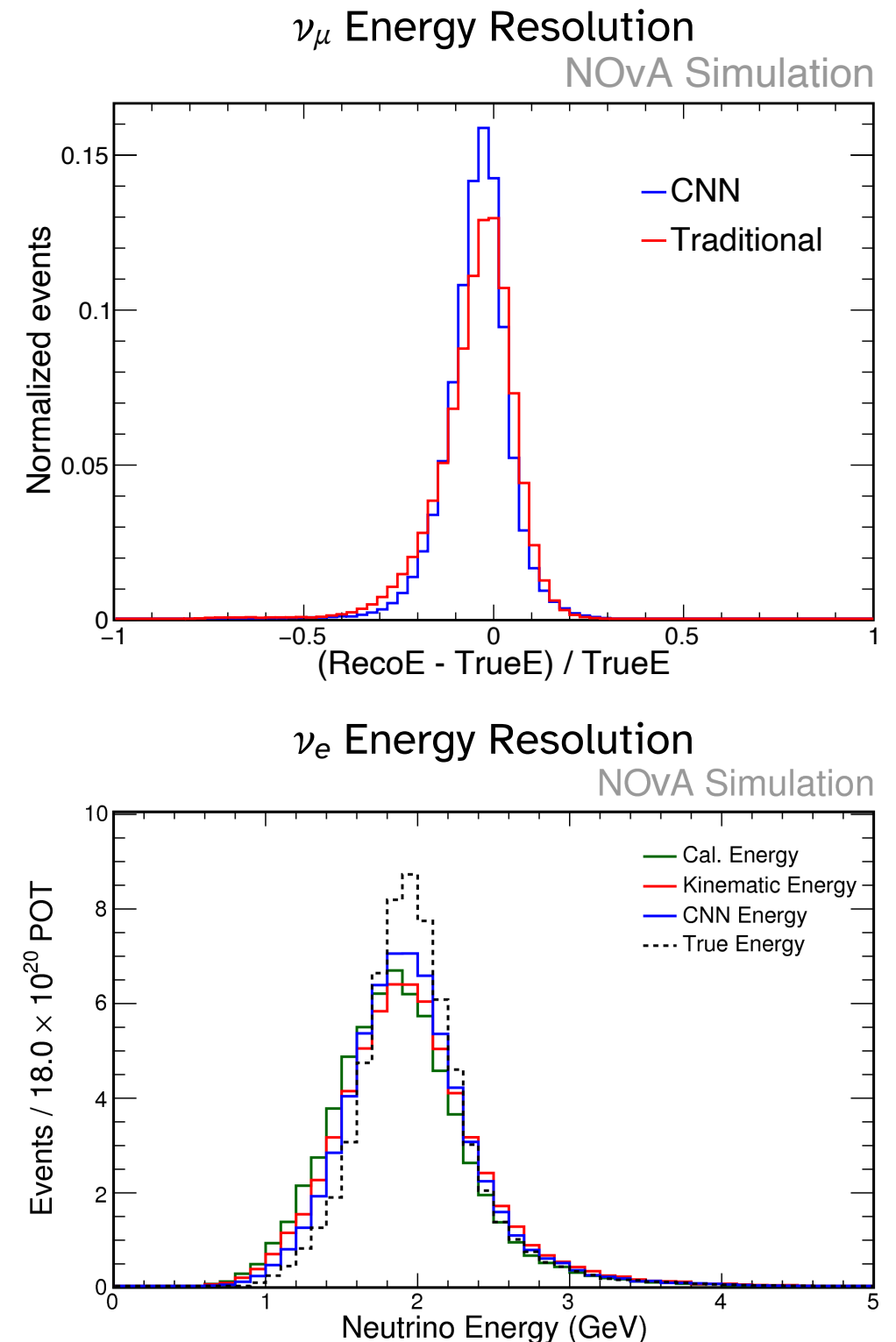


# RegCVN

## CNN for Energy Estimation

- Various ResNet CNN models for  $\nu_e$ ,  $e$ , and simultaneous  $\nu_\mu/\mu$  energy.
- $\nu_\mu/\mu$  model takes reco. track length of longest prong as additional input for muons that leave cropped pixel map region.
- Gaussian noise added to pixel values for robustness against calibration systematic error.

Methods paper: [Phys. Rev. D 99 012011](#)



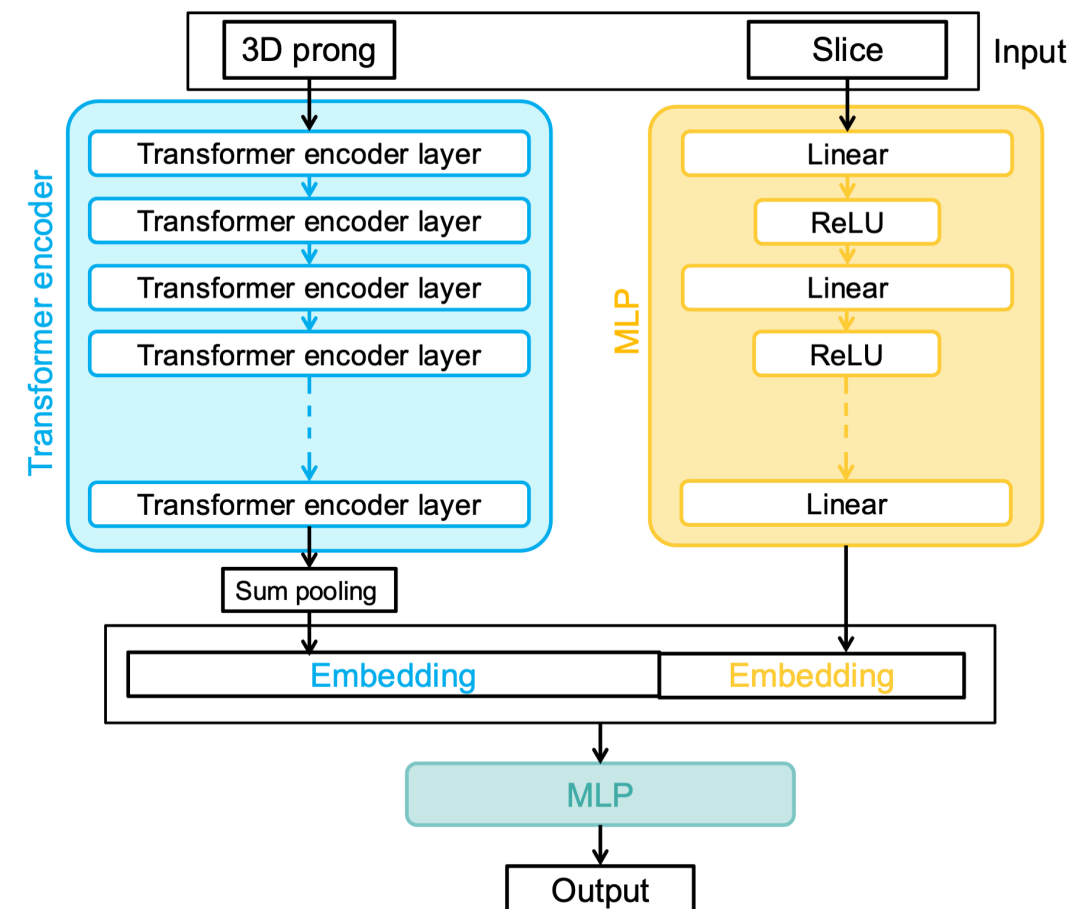
# TransformerEE

## Transformer for Energy Estimation

- Takes event-level and prong-level reconstructed variables as input.
- Transformer network replaces previous LSTM architecture.
- Gaussian noise added to energy-related variables for robustness against calibration systematic error.
- Simultaneous  $\nu_\mu/\mu$  energy predictions.

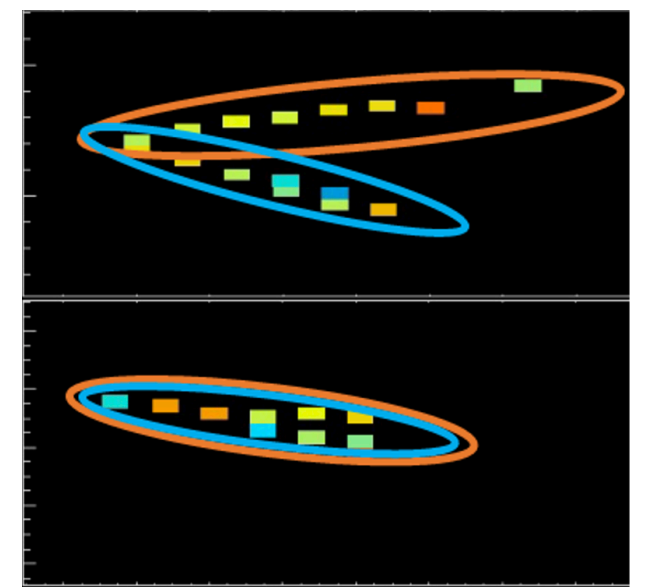
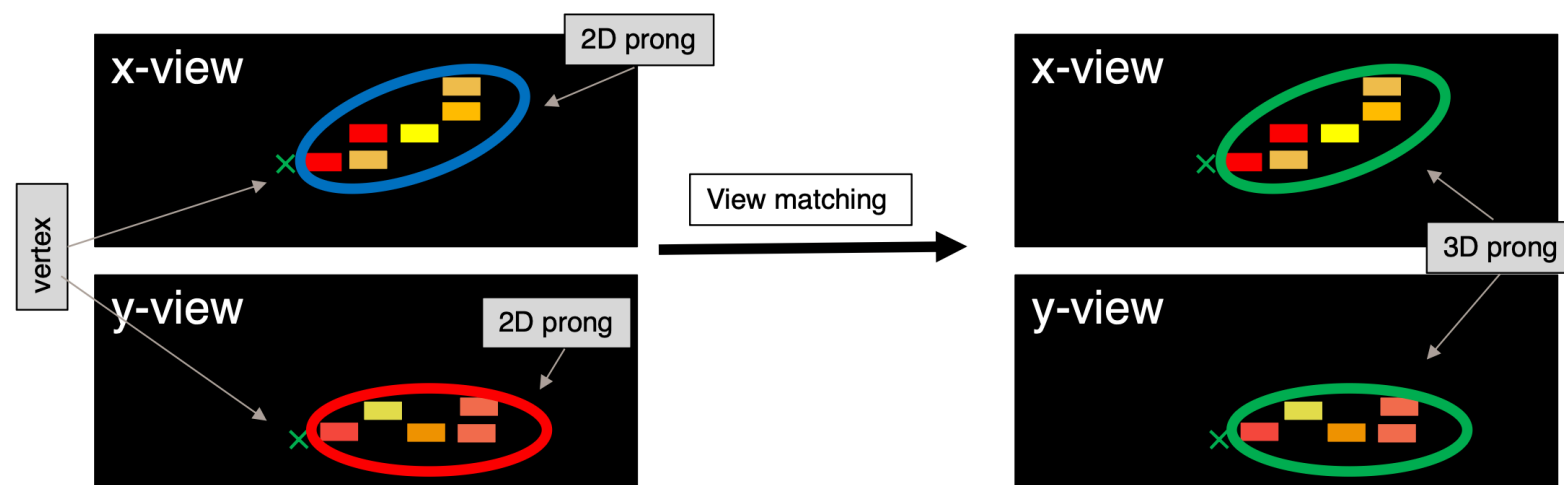
|                                                                                                                            |
|----------------------------------------------------------------------------------------------------------------------------|
| x,y,z components of the direction vector for 3D prong                                                                      |
| x,y,z components of the position vector for 3D prong                                                                       |
| transformerCVN output scores representing the likelihood that the prong was produced by a muon,electron,pion,proton,photon |
| Estimated energy by Break Point Fitter under different particle assumptions (0: proton, 1: pion, 2: muon)                  |
| BPF energy of the 3D prong shared with other prongs                                                                        |
| x,y,z components of the BPF momentum for 3D prong                                                                          |
| Prong length                                                                                                               |
| Number of hits in the 3D prong                                                                                             |
| Number of hits of the 3D prong in XZ, YZ view                                                                              |
| Number of planes crossed by the 3D prong track                                                                             |
| Sum of the calibrated energy of all hits associated with the 3D prong.                                                     |

|                                                                    |
|--------------------------------------------------------------------|
| Sum of the calibrated energy of all hits in the slice              |
| calE - sum(png.calE)                                               |
| Number of hits in the slice                                        |
| Calorimetric energy of hits that don't appear in any FuzzyK prongs |



# Prong Segmentation

- Current method uses a fuzzy k-means clustering algorithm to reconstruct prongs: J. Phys. Conf. Ser. 664 072035 (2015).
- Clustering first applied independently to each view to create 2D prongs.
- OPALProng view-matching algorithm then creates 3D prongs.
- Motivation: Want to study whether deep learning method can improve prong segmentation without exceeding NOvA computational resource limitations.



Two **O**verlapping **P**Artic**L**es  
successfully reconstructed

# ML Segmentation Methods

- Convolutional Neural Networks (CNNs)

NERDProng CNN for NOvA: [NPML 2020](#)

- Image-based networks, so naturally dense
- Sparse CNNs exist, but must approximate convolutions

- Graph Neural Networks (GNNs)

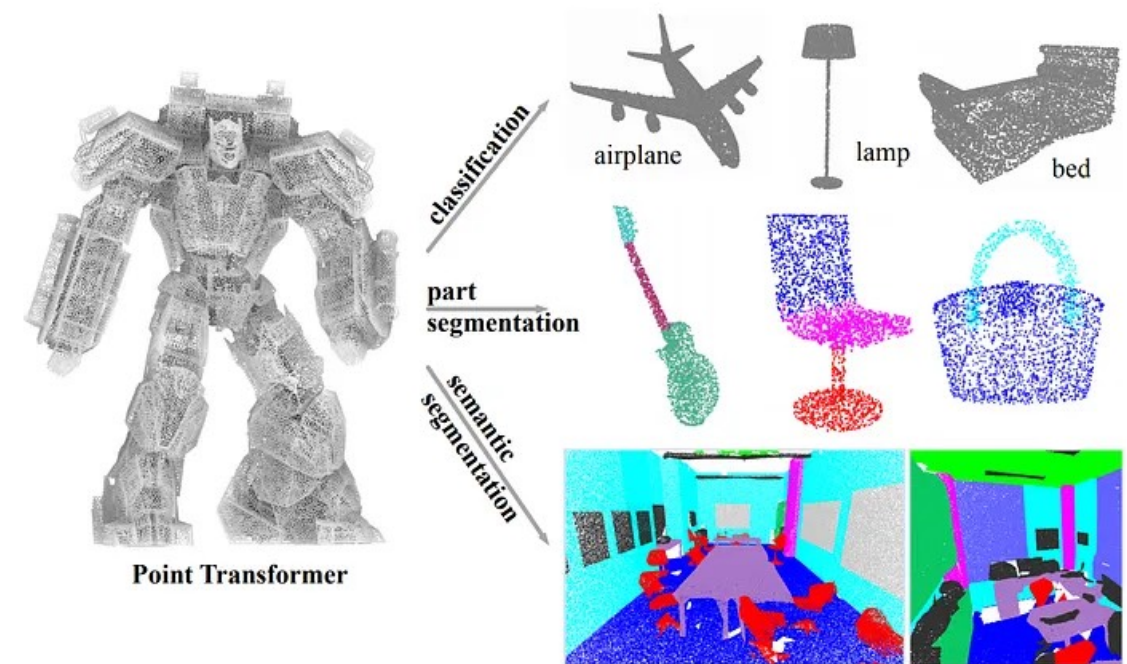
- Uses point clouds with edge connections, so naturally sparse
- Typically need to establish rigid graph structure, usually by k-nearest neighbors (kNN)
- Graph Attention Network (GATs) use attention mechanism for neighborhood aggregation of points

- Point Set Transformers (PSTs)

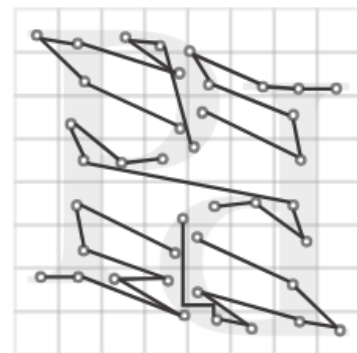
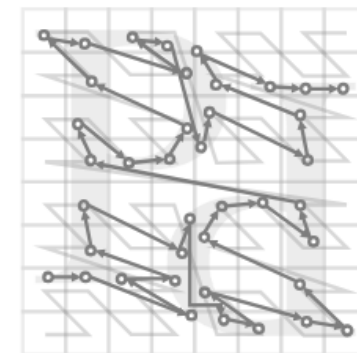
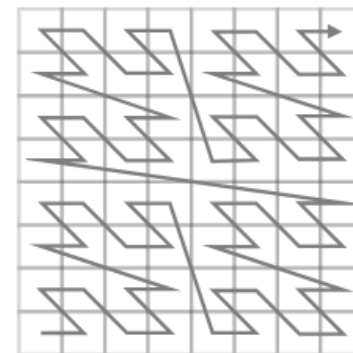


# Point Set Transformers

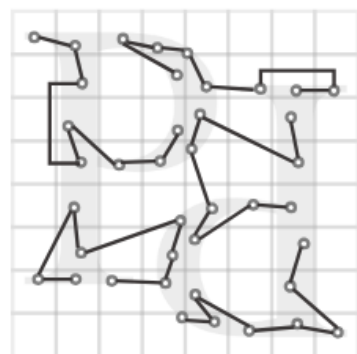
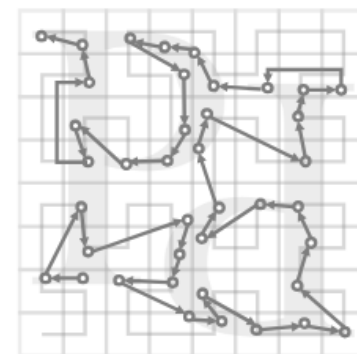
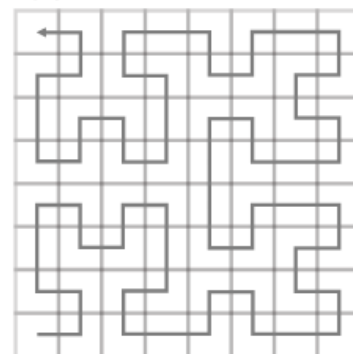
- PST applies self-attention between points in point cloud.
- Permutation-invariant attention is appropriate for unordered point sets.
- Naturally sparse.
- Complexity grows as  $N^2$ , so attention operation usually limited to kNN.
- Point Transformer v3 replaced kNN with serialization of points to improve efficiency.



(a) Z-order



(b) Hilbert



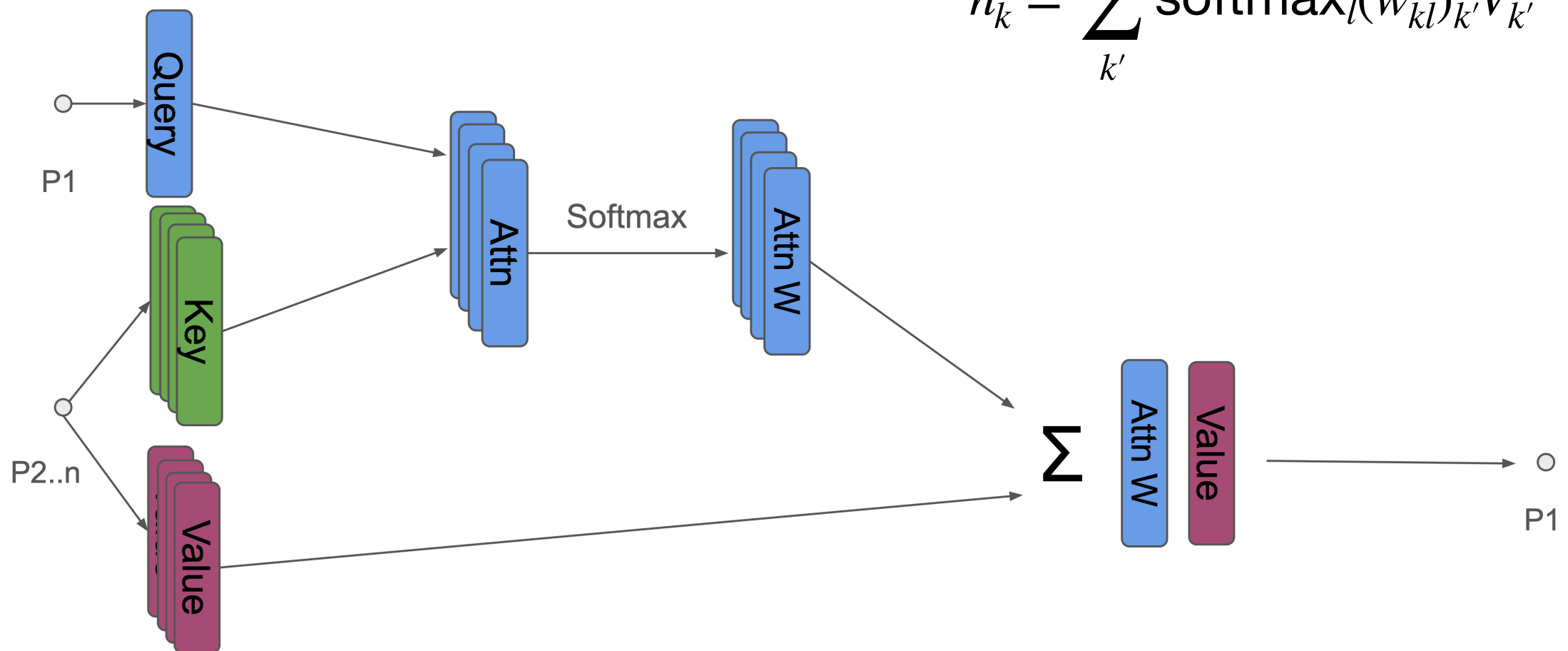
"Point Transformer" IEEE / CVF ICCV 2021

"Point Transformer V3: Simpler, Faster, Stronger" IEEE / CVF CVPR 2024

# Point Set Attention

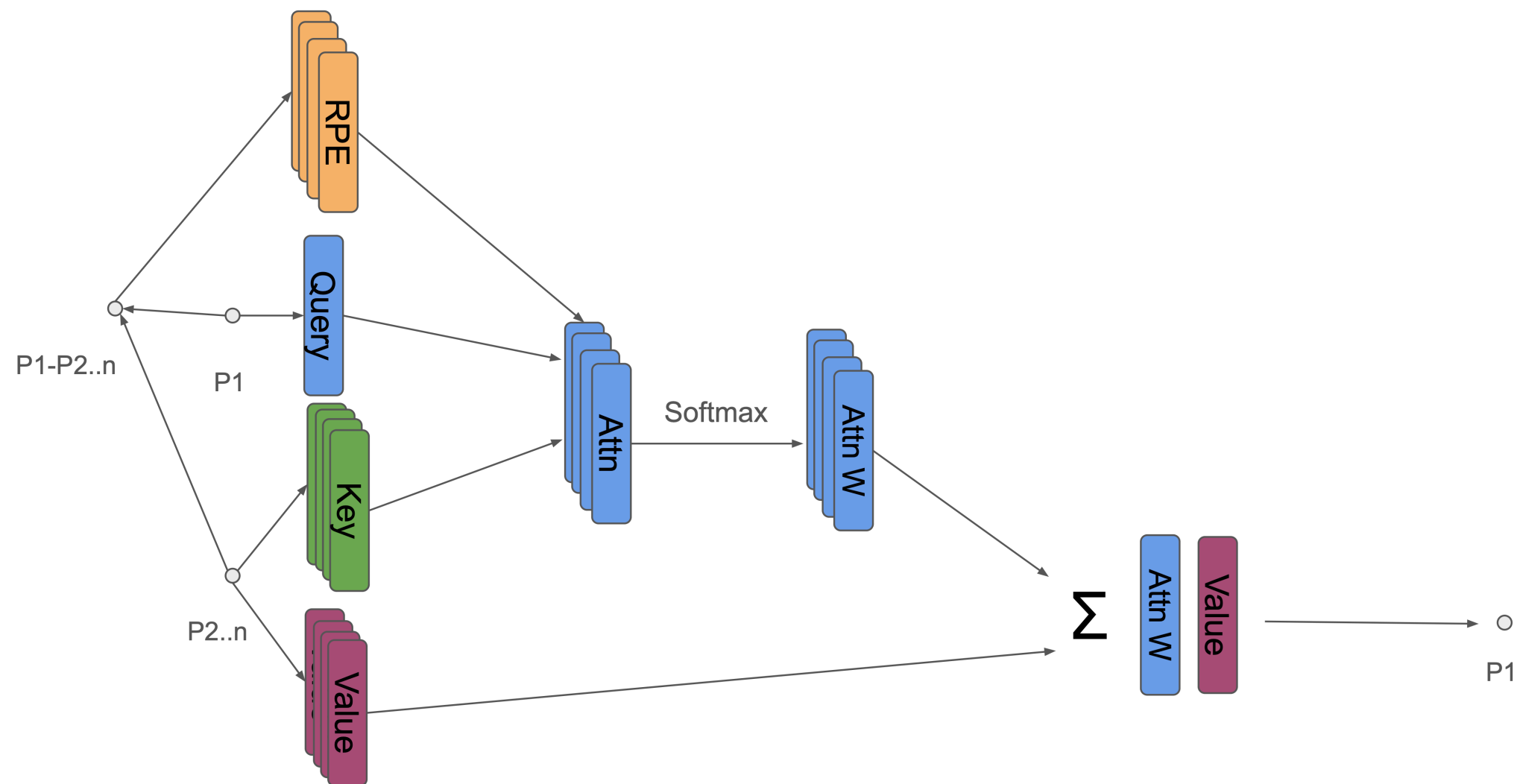
- Consider dataset represented by **event**  $i$ , **view**  $j$ , **point**  $k$
- Edge score between two points  $x_k$  and  $x_{k'}$ :  $w_{kk'} = Q_k K_{k'}$
- Scores for all points connected to  $x_k$  normalized with softmax:

$$h_k = \sum_{k'} \text{softmax}_l(w_{kl})_{k'} V_{k'}$$



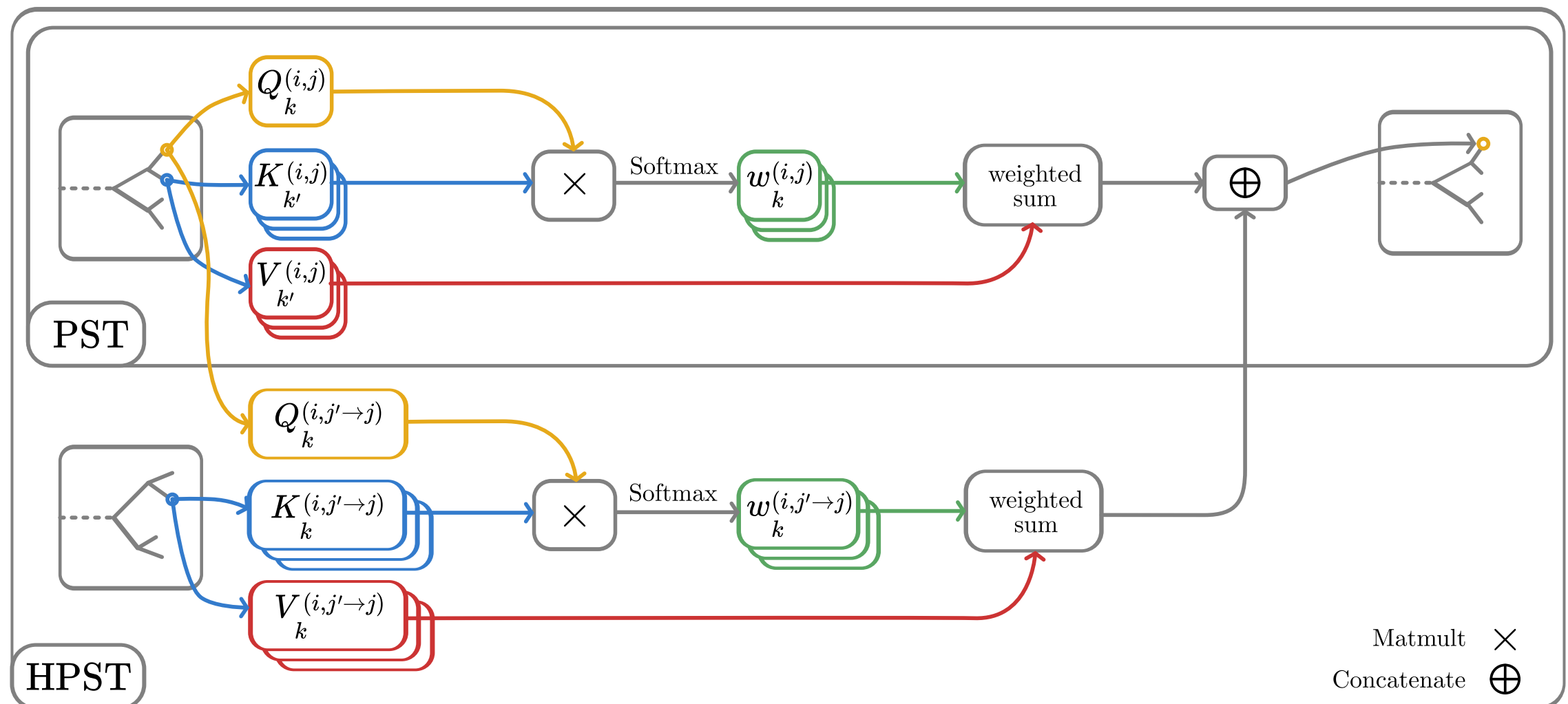
# Relative Positional Encoding

- Add relative positional encoding (RPE), a Linear layer based on difference of pair's features:  $\text{RPE}(x_k, x_{k'}) = \text{Linear}(x_k - x_{k'})$
- Edge score is then  $w_{kk'} = Q_k K_{k'} + \text{RPE}(x_k, x_{k'})$



# Heterogenous Attention

- Want to share information between views from the beginning (early fusion).
- Define some inter-view distance in addition to intra-view distance, then create connections between points across views as well.
- Inter-view edge score:  $w_{kk'}^{(i,j' \rightarrow j)} = Q_k^{(i,j' \rightarrow j)} K_{k'}^{(i,j' \rightarrow j)}$  where view  $j' \neq j$

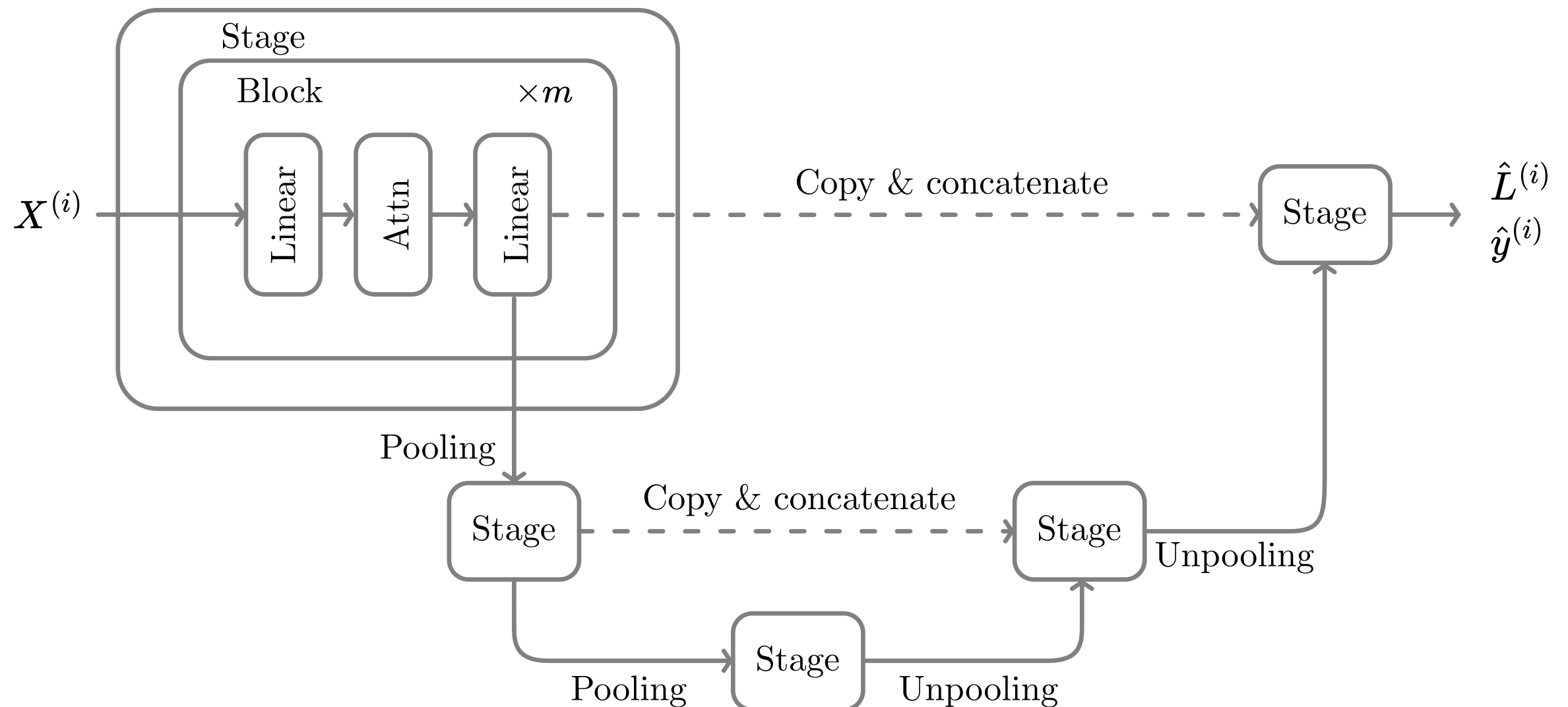


# Pooling

- Need pooling layer to aggregate attention weights across regions separated in space.
- Pooling nearest neighbors is computationally expensive so approximate it with a grid  $G$ .
- For each square of grid  $G_l$ , simply average features of all hits within square,  $v'_l = \frac{1}{|G_l|} \sum_{v_{i,j} \in G_l} v_{i,j}$
- Pooling applied separately between views.

# Architecture

- Finally, apply unpooling to create U-net-like architecture, with skip connections.





# Objectives

## Loss Function

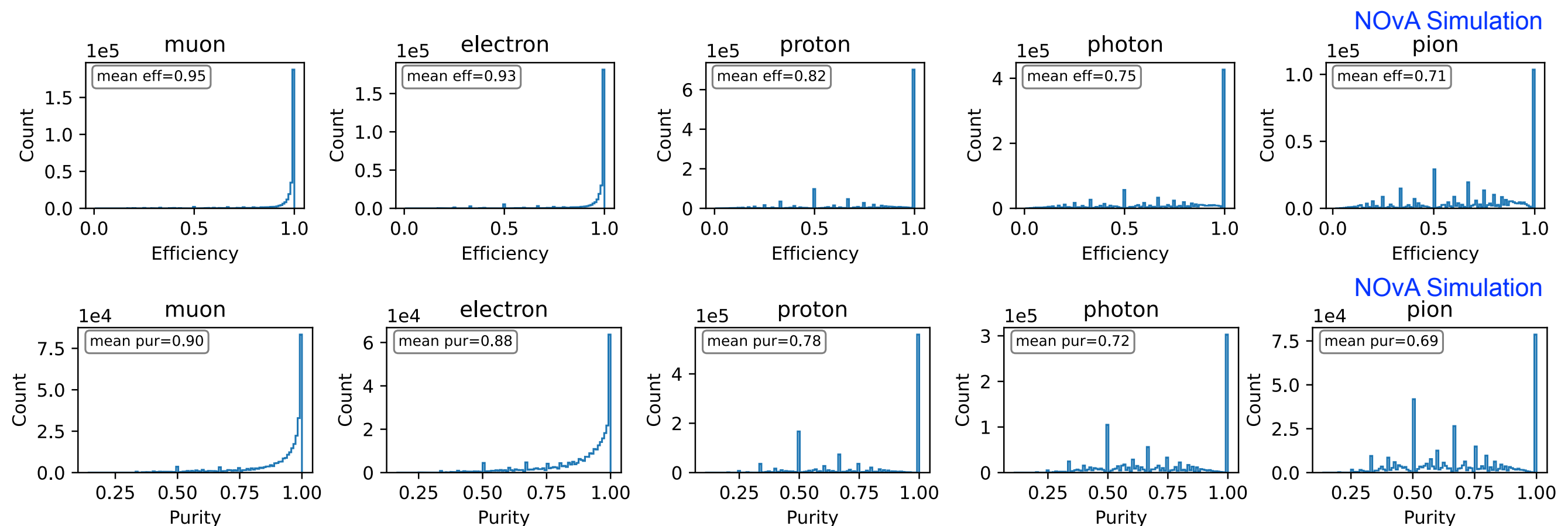
- Networks performs panoptic segmentation, both instance segmentation (clustering) and semantic segmentation (classification):  $\mathcal{L} = \lambda \mathcal{L}_{\text{sem}} + (1 - \lambda) \mathcal{L}_{\text{ins}}$
- Semantic segmentation loss is multi-class cross-entropy.
- Instance segmentation compares assignments to best possible assignments of points to segments.
  - Optimal assignment calculated with the Hungarian algorithm.

Kuhn, H. W. "The Hungarian method for the assignment problem" *Naval Research Logistics* **2** 83-97 (1955)

# HPST Performance

## Clustering (Instance Segmentation)

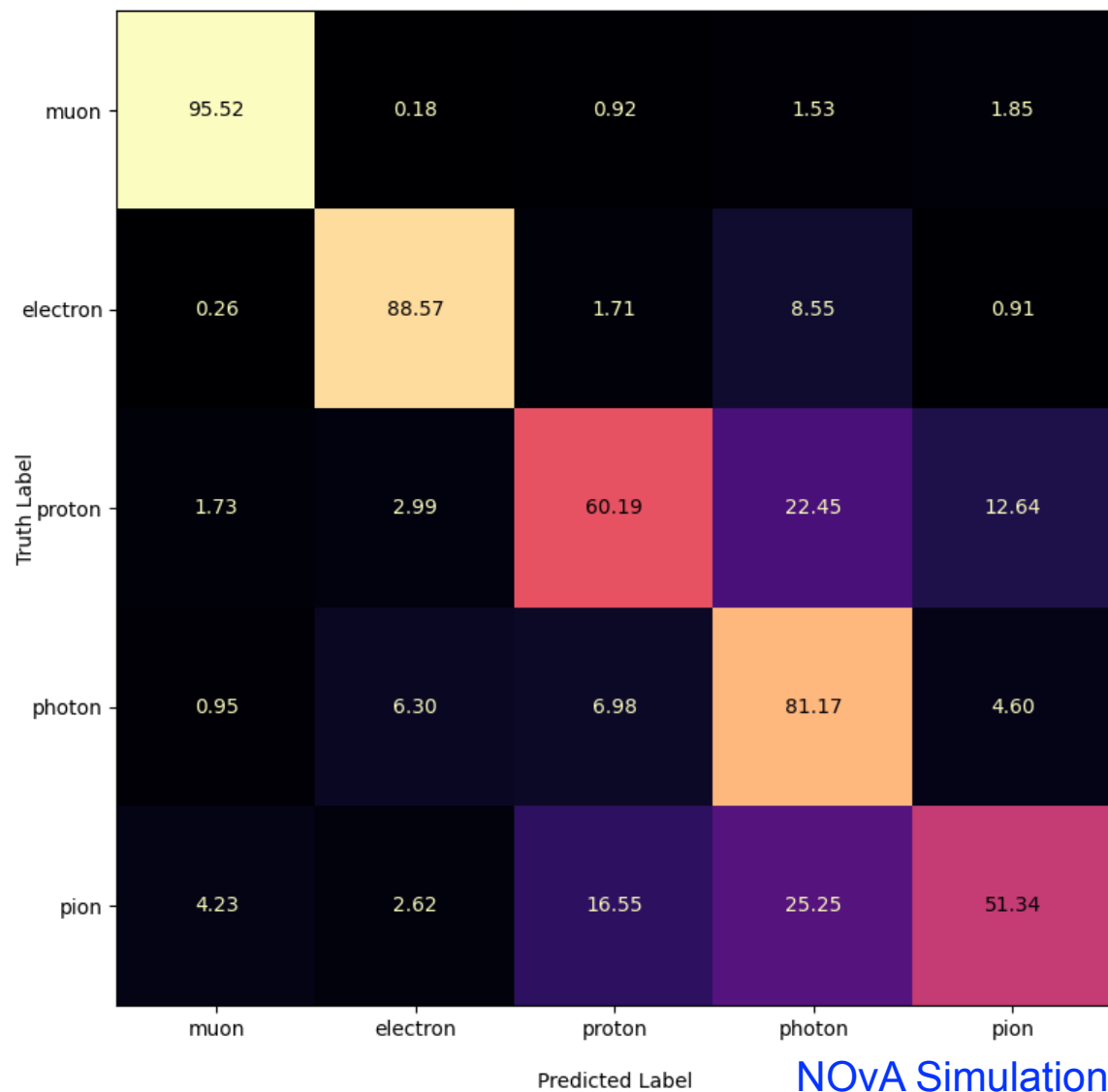
- Efficiency = % of predicted prong assigned to correct prong.
- Purity = % of true prong predicted correctly.
- Good performance, particularly in majority classes.



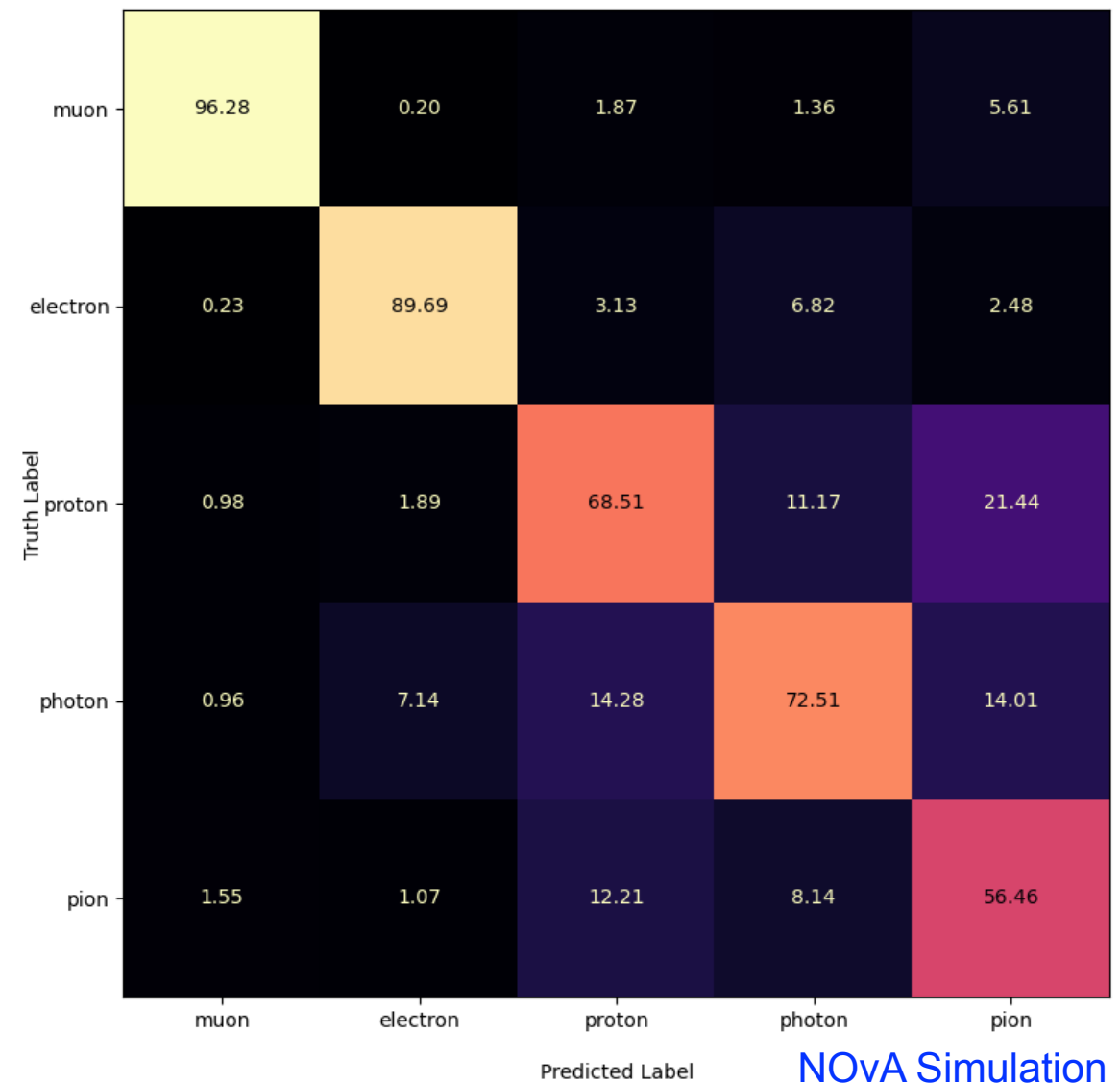
# HPST Performance

## Classification (Semantic Segmentation)

Efficiency (Normalized by rows)



Purity (Normalized by columns)



# Architecture Comparisons

- Considerable performance and resource utilization improvement compared to region-based convolution network (R-CNN), and some improvement over graph attention network (GAT).

| Model       | Memory usage (MiB) | Time per sample (s) | OVR AUC | Segmentation accuracy |
|-------------|--------------------|---------------------|---------|-----------------------|
| R-CNN       | $151.9 \pm 11.950$ | $3.1449 \pm 0.0660$ | 0.732   | 0.343                 |
| GAT         | $29.10 \pm 0.001$  | $0.3911 \pm 0.0010$ | 0.854   | 0.659                 |
| HPST (ours) | $29.35 \pm 0.559$  | $0.4636 \pm 0.0032$ | 0.968   | 0.835                 |

“Mask R-CNN” IEEE / CVF ICCV 2017

“Graph Attention Networks” ICLR 2018

# Summary

- Widespread use of machine learning in NOvA reconstruction with various approaches for use of two 2D views.
- Developed heterogeneous point set transformer model for NOvA prong segmentation, a task for which computational resource limitations previously precluded deep learning methods.
- Self-attention between hits in different views allows for information sharing between views at the hit level.
- Improvement over R-CNN and GAT models also tested.
- Network creates 2D prongs, but want to extend to 3D prongs with interpolation between views.
- More details:
  - ML4PS paper: [arXiv:2510.09659](https://arxiv.org/abs/2510.09659)
  - HPST for LArTPCs: [JINST 20 P07030 \(2025\)](#)