Transformer Network for Event/Particle Identification and Interpretability at NOvA

Jianming Bian

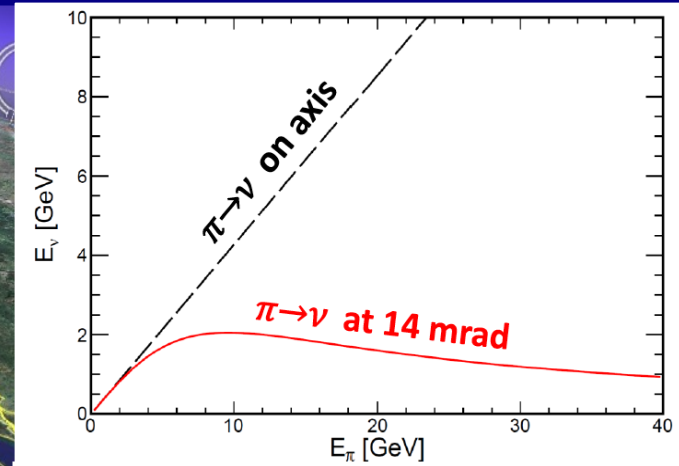
University of California, Irvine

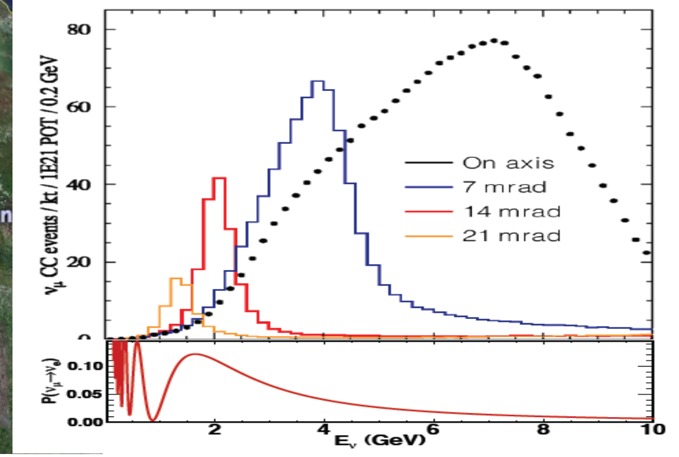
for the NOvA Collaboration

06-25-2024

NPML 2024, ETH Zurich, Switzerland

NuMI Off-Axis ν_e Appearance Experiment (NOvA)

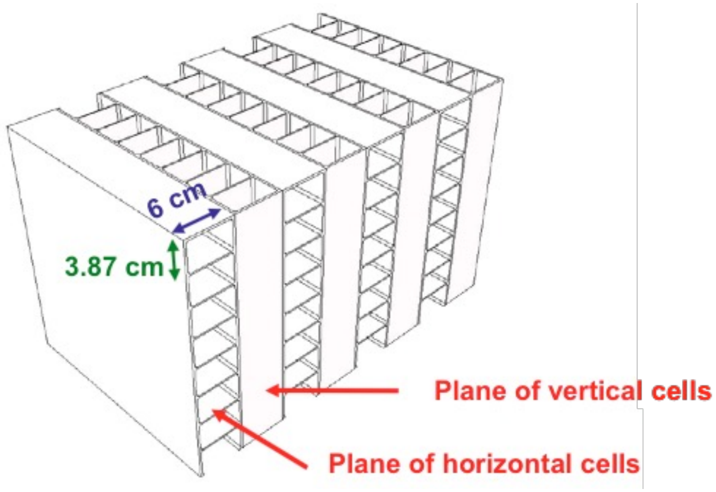


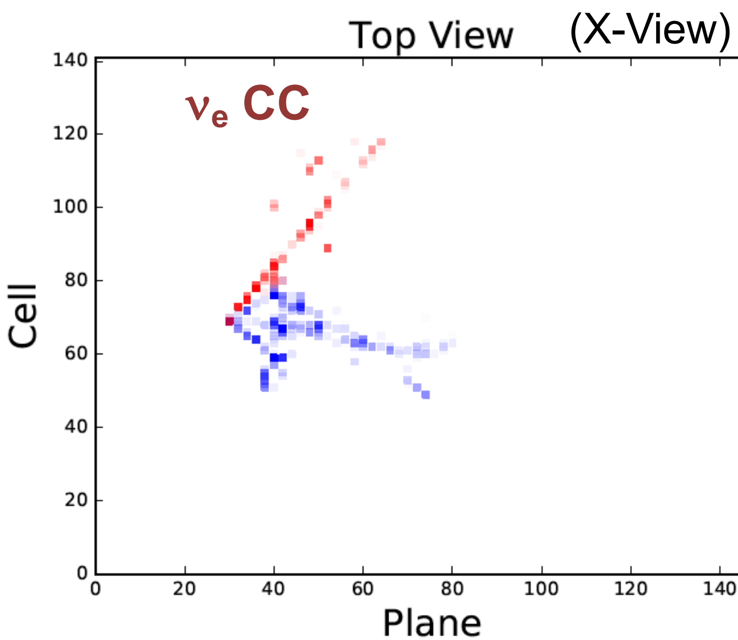


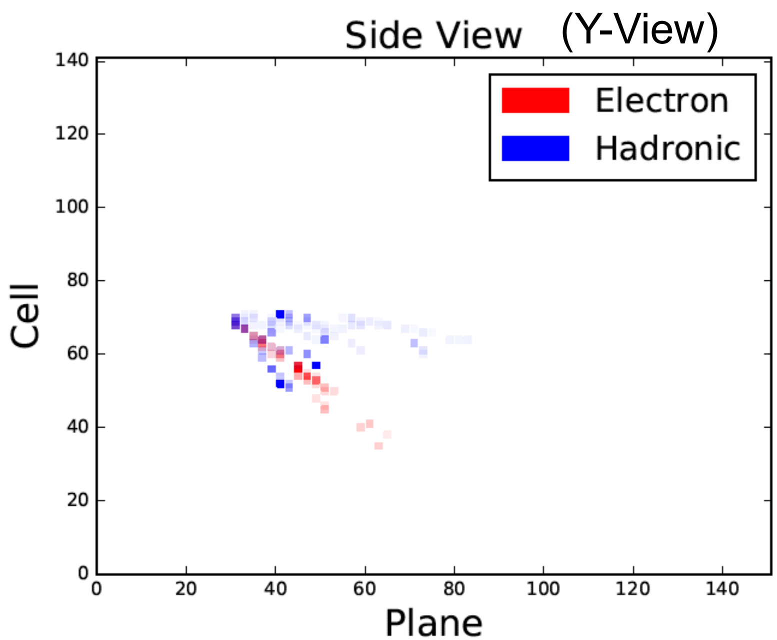
- Muon neutrino beam at Fermilab near Chicago
- Longest baseline in operation (810 km), large matter effect, sensitive to mass ordering
- Far/Near detector sited 14 mrad off-axis, narrow-band beam around oscillation maximum

NOvA Event Images

- NOvA detector cells arranged in planes, assembled in alternating X and Y directions
- Produce a pair of pixel maps (Cell Number vs. Plane Number) for the X and Y view of each interaction
- NOvA's 2-view pixel maps are ideal for image processing neural networks to reconstruct neutrino events

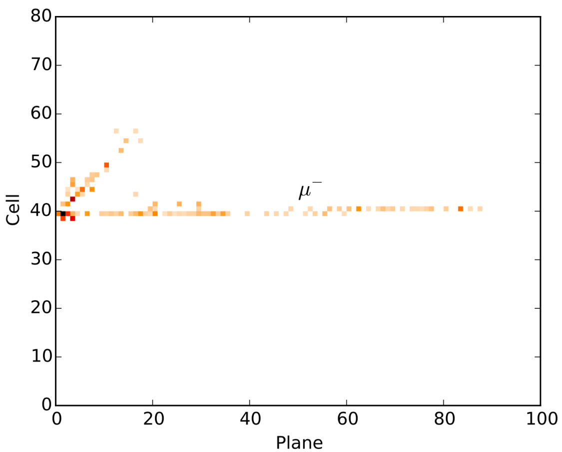




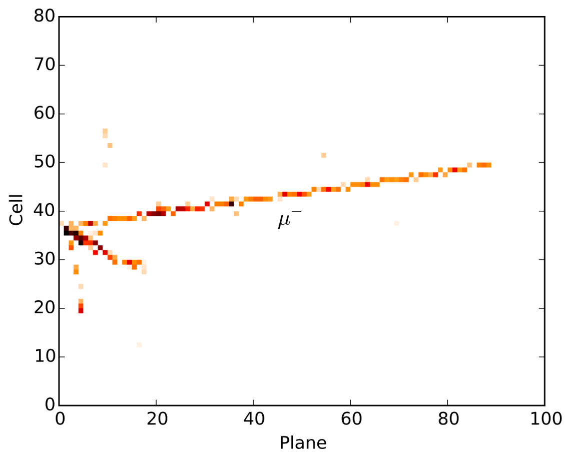


EventCVN and ProngCVN

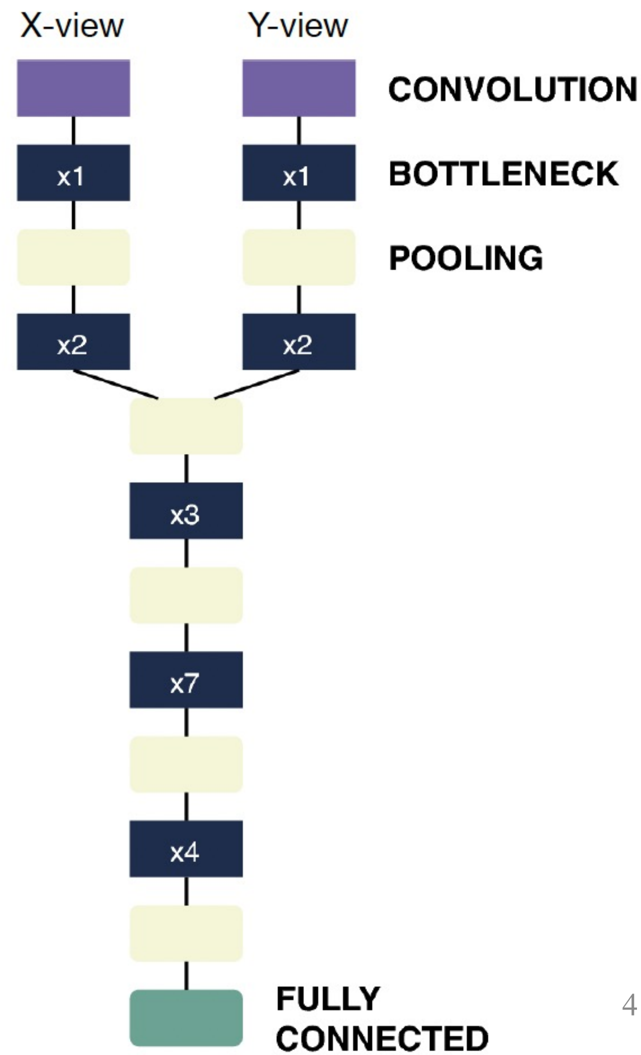
- EventCVN: neutrino event selection CNN trained on cropped XZ and YZ event images (pixel maps). Predicts [ν_μ CC, ν_e CC, NC, cosmic], used since 2017 [JINST 11, P09001 (2016), Phys.Rev.Lett. 118 (2017)]
- ProngCVN: predicts final state particle type from prong-only images and event images



X-view



Y-view

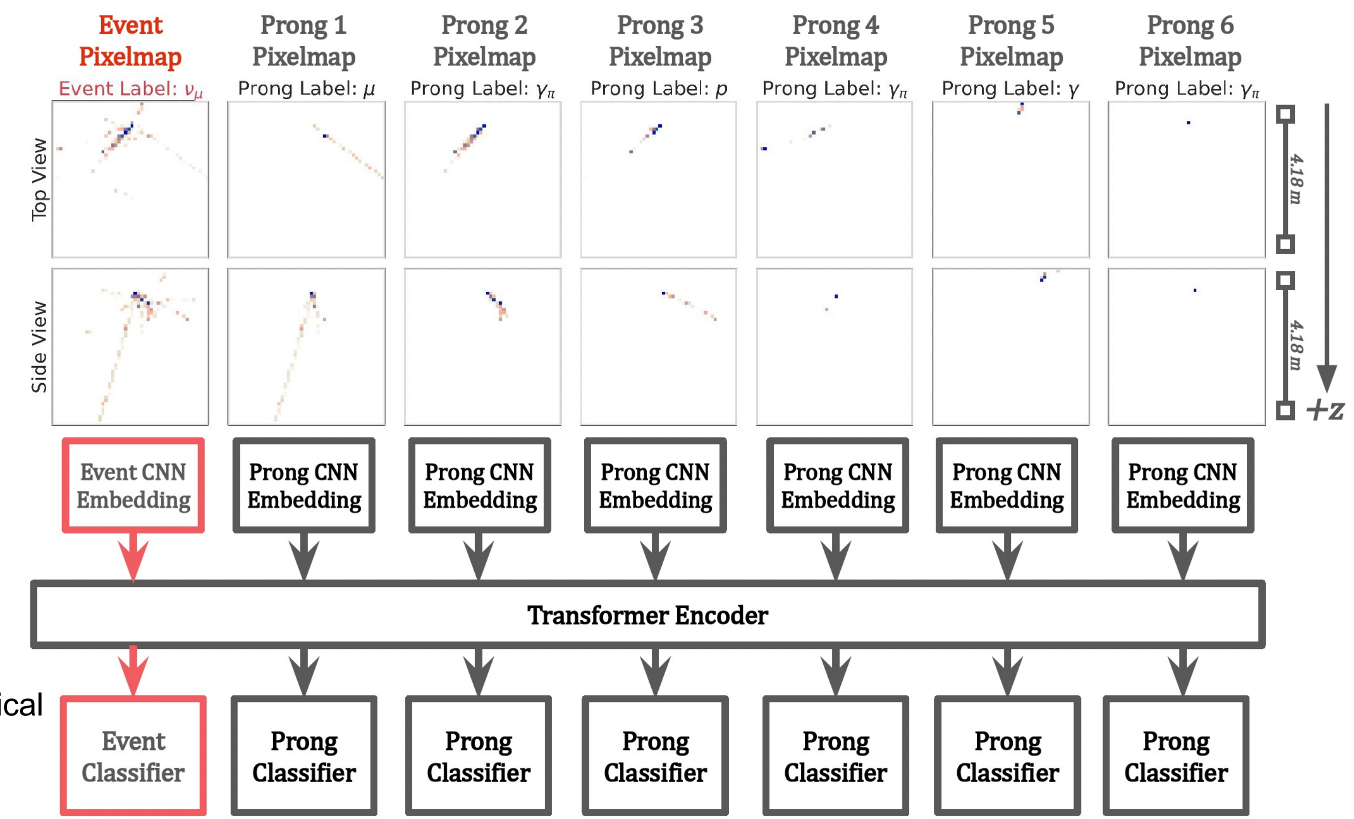


MobileNet-inspired architecture of 2-view EventCVN.

TransformerCVN for Event and Particle Identification

- Transformer: attention based network, foundation of ChatGPT, ideal for training on variable-length collection of object such as prongs
- Uses both event and prong images as inputs, identifies neutrino flavor and each particle simultaneously.
- Attention mechanisms automatically focus training and evaluation on image regions important to the final decision, significantly reducing the computing burden and enhancing training performance
- Attention mechanisms also provides interpretability, making deep learning more than just a "black box"

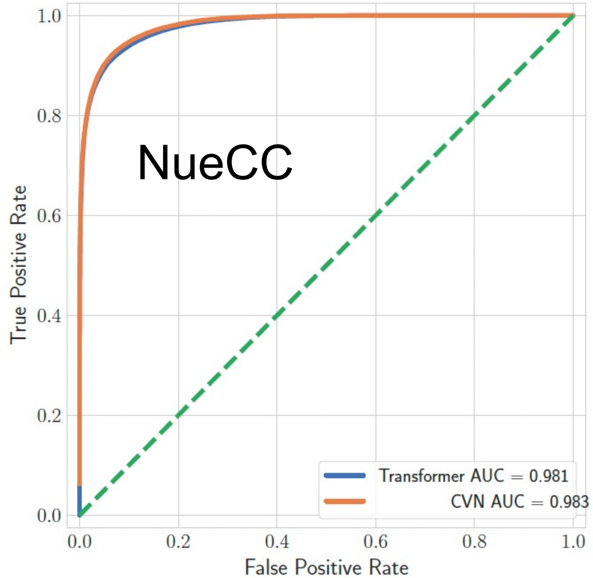
Transformer+CNN
="TransformCVN"



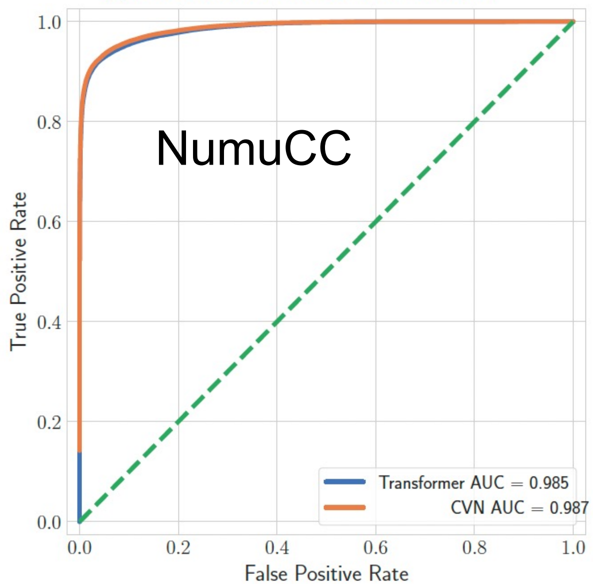
Machine Learning and the Physical Sciences, NeurIPS 2023
arXiv:2303.06201

Event and Prong ROC Curves

TransformerCVN/Event CVN

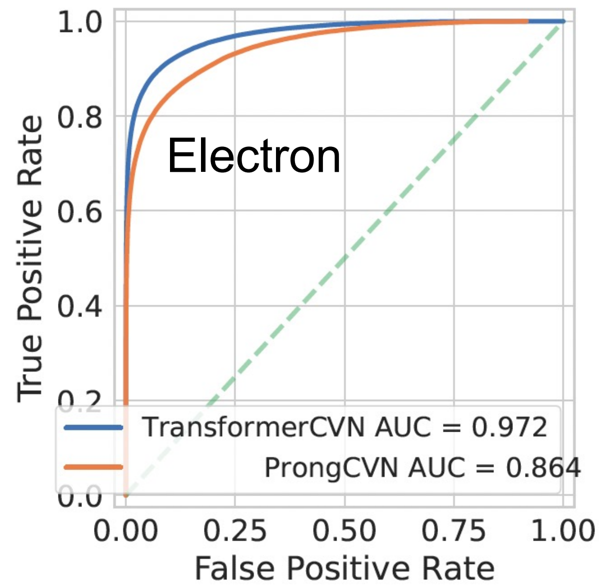


(a) ROC Curve for ν_e event reconstruction.

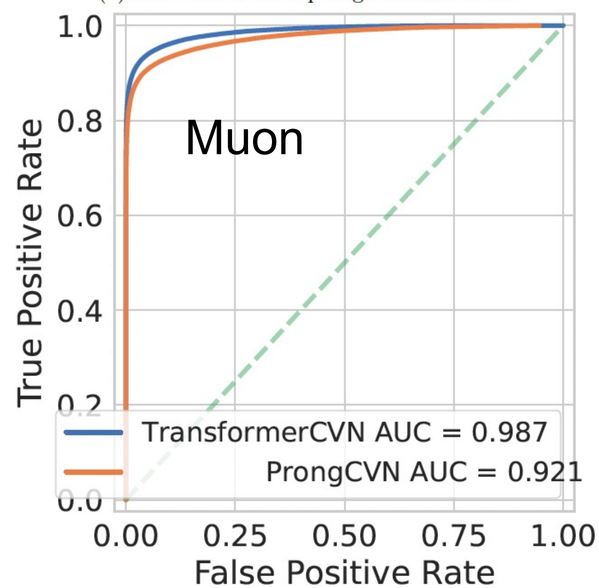


(b) ROC Curve for ν_μ event reconstruction.

TransformerCVN/Prong CVN

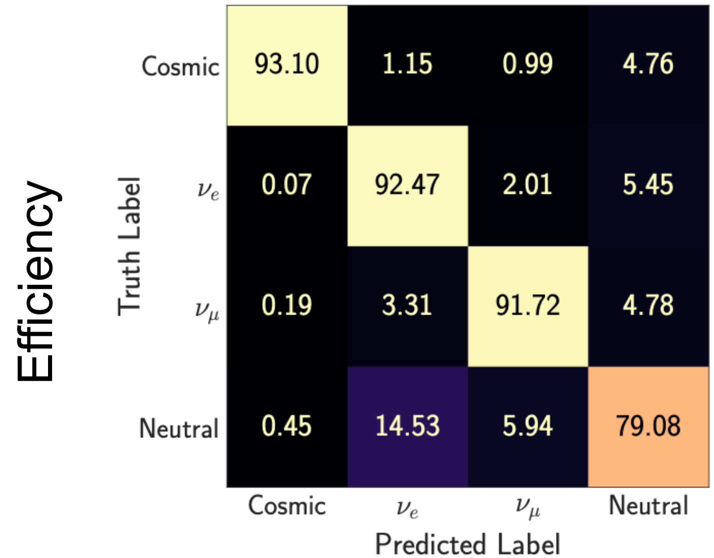


(a) ROC Curve for e prong reconstruction.

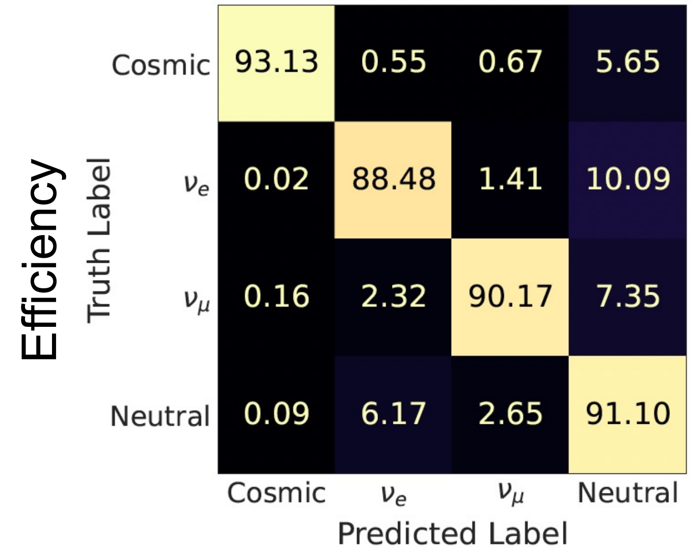


(b) ROC Curve for μ prong reconstruction.

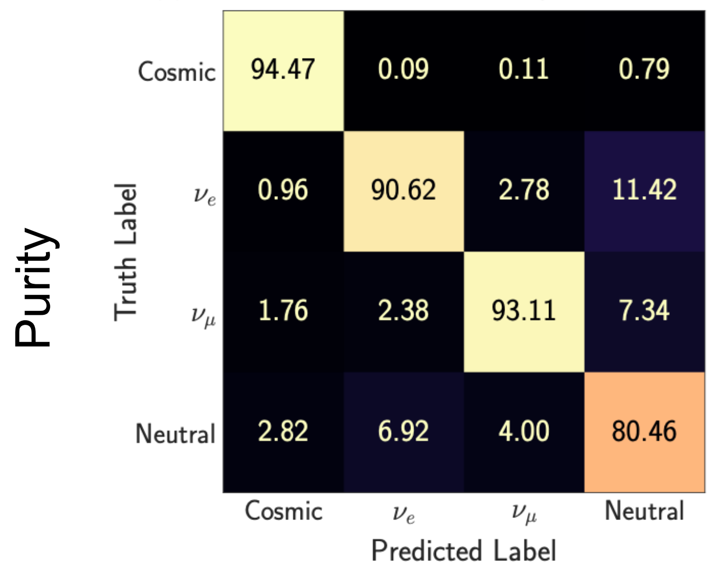
Event Confusion Matrices



(a) **Efficiency** matrix, normalized along truth labels.

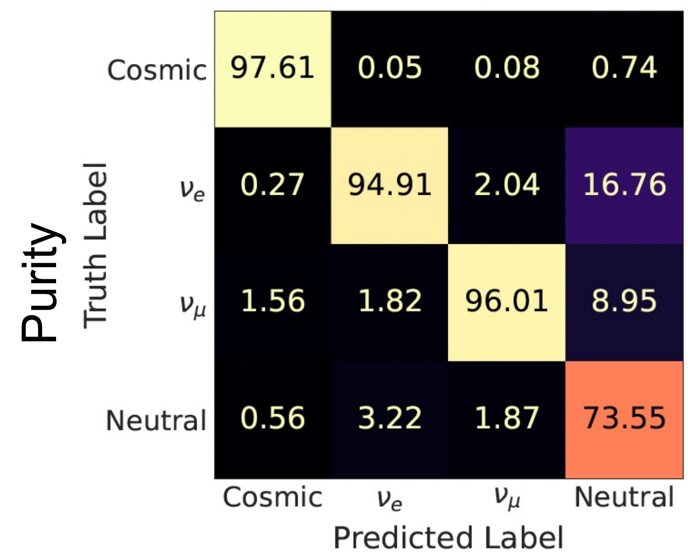


(a) **Efficiency** matrix, normalized along truth labels.



(b) **Purity** matrix, normalized along predictions.

TransformerCVN



(b) **Purity** matrix, normalized along predictions.

Event CVN

Prong Confusion Matrices

Efficiency

Truth Label \ Predicted Label	e	μ	p	$\pi^{\pm}$	γ
e	83.00	0.56	3.32	1.33	11.79
μ	0.56	89.78	2.50	4.30	2.87
p	1.69	1.00	77.91	6.57	12.83
$\pi^{\pm}$	1.73	4.64	22.48	51.08	20.06
γ	4.70	0.50	8.94	3.81	82.04

(a) **Efficiency** matrix, normalized along truth labels.

Purity

Truth Label \ Predicted Label	e	μ	p	$\pi^{\pm}$	γ
e	87.52	0.85	2.00	2.44	6.59
μ	0.39	90.83	1.01	5.27	1.07
p	3.03	2.56	79.65	20.44	12.18
$\pi^{\pm}$	1.18	4.56	8.78	60.75	7.27
γ	7.88	1.20	8.56	11.11	72.89

(b) **Purity** matrix, normalized along predictions.

TransformerCVN

Efficiency

Truth Label \ Predicted Label	e	μ	p	$\pi^{\pm}$	γ
e	77.94	0.29	3.02	4.39	14.35
μ	0.65	82.36	2.03	12.54	2.42
p	1.87	0.25	64.73	17.72	15.43
$\pi^{\pm}$	2.13	1.73	12.64	68.22	15.27
γ	7.35	0.19	6.32	10.68	75.46

(a) **Efficiency** matrix, normalized along truth labels.

Purity

Truth Label \ Predicted Label	e	μ	p	$\pi^{\pm}$	γ
e	82.37	0.51	2.28	4.22	8.34
μ	0.46	96.25	1.02	8.06	0.94
p	3.36	0.74	82.93	28.90	15.23
$\pi^{\pm}$	1.46	1.97	6.19	42.51	5.76
γ	12.35	0.54	7.58	16.32	69.74

(b) **Purity** matrix, normalized along predictions.

ProngCVN

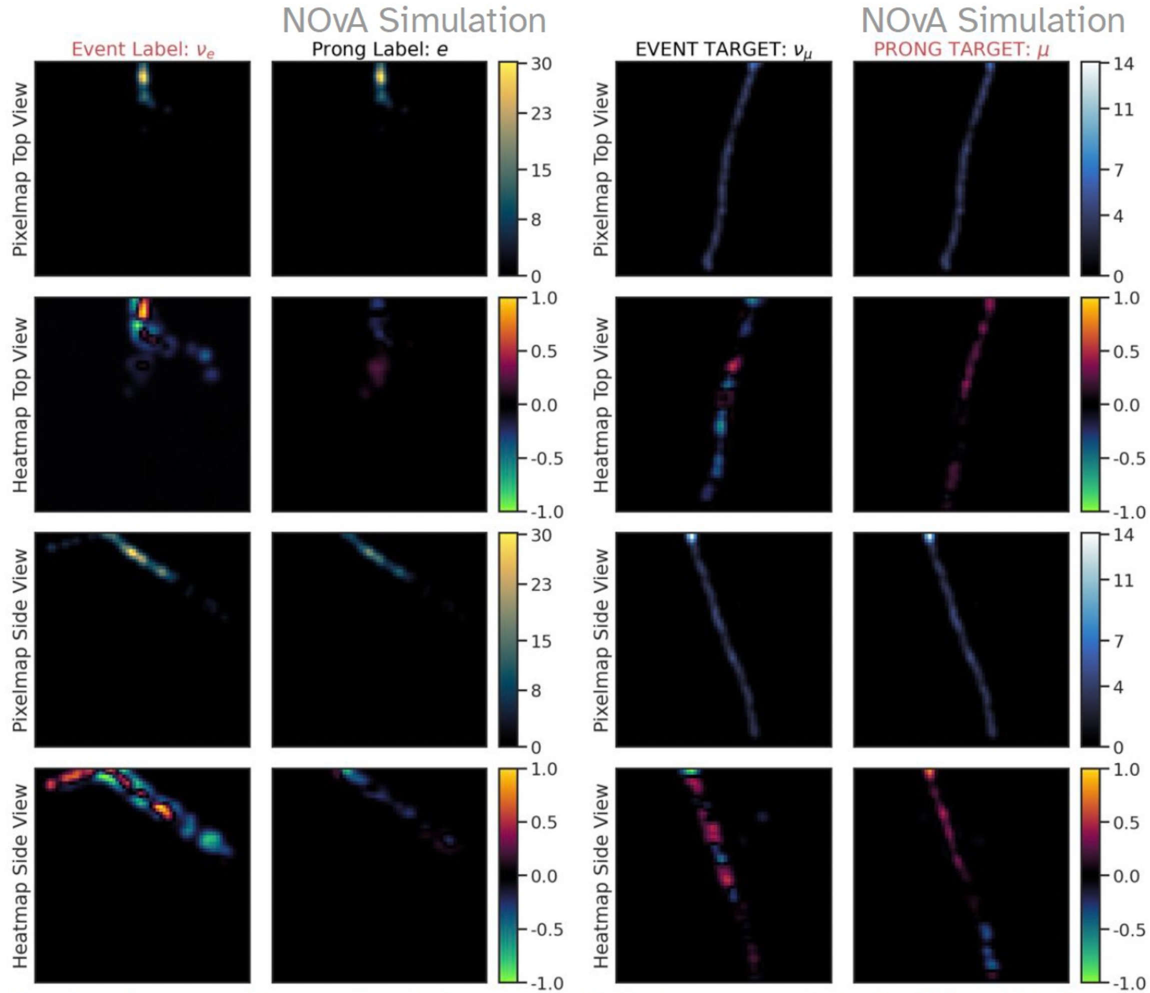
Interpretability

- Pixel Gradients (Saliency)
  - Saliency: gradient of output classification probability with respect to charge in each hit
  - Attention mechanism enhances the saliency of key regions in the input
  - Study saliency to understand which regions the Transformer focuses on to identify a particle
  - When aggregated, provides a template of a typical pattern for each prong type

- Attention Scores
  - Indicate the importance of different elements to the output
  - Allows us to find out which prongs are most useful to identify different types of events
  - Diagnose neural network and explain decision

Individual Saliency

- Calculate saliency of different hits for each event
- Useful for debugging wrong predictions
- Need to aggregate multiple prongs of the same type to find patterns



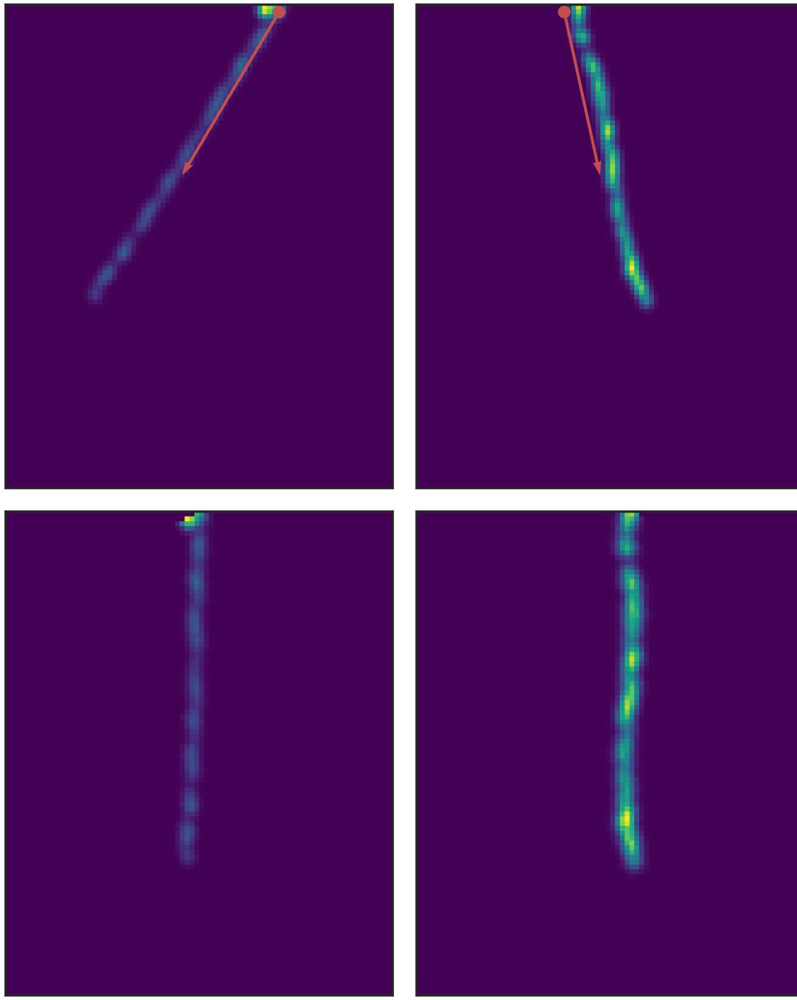
Example saliency maps for νe event prediction of ν_e event (left) and μ prong prediction of ν_μ CC event (right).

Red: more likely to predict the given flavor/particle type with more energy in that location.
Green: less likely to predict if there is more activity (anti-correlation).

Saliency Aggregation

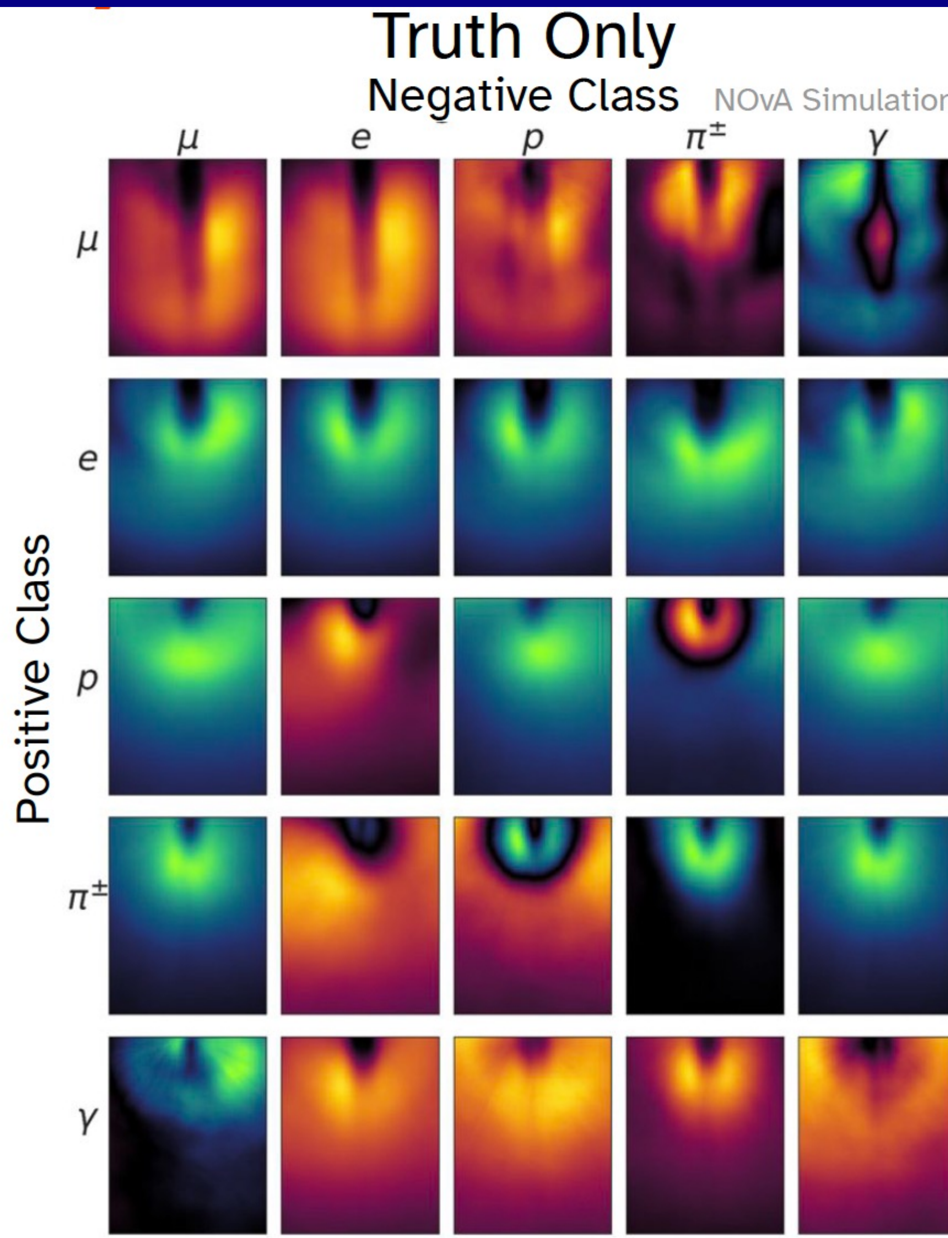
- Calculated average saliency values for each type of particle in ~10,000 events
- Rotated and translated each prong image using the vertex and direction information associated with each prong.
    - Every prong forced to have vertex at (40, 0) and facing toward +z.
    - Possibly limit event by track length to compare similar lengths.
- New tool to analyze saliency pattern

Pixel Map Alignment



Saliency Aggregation

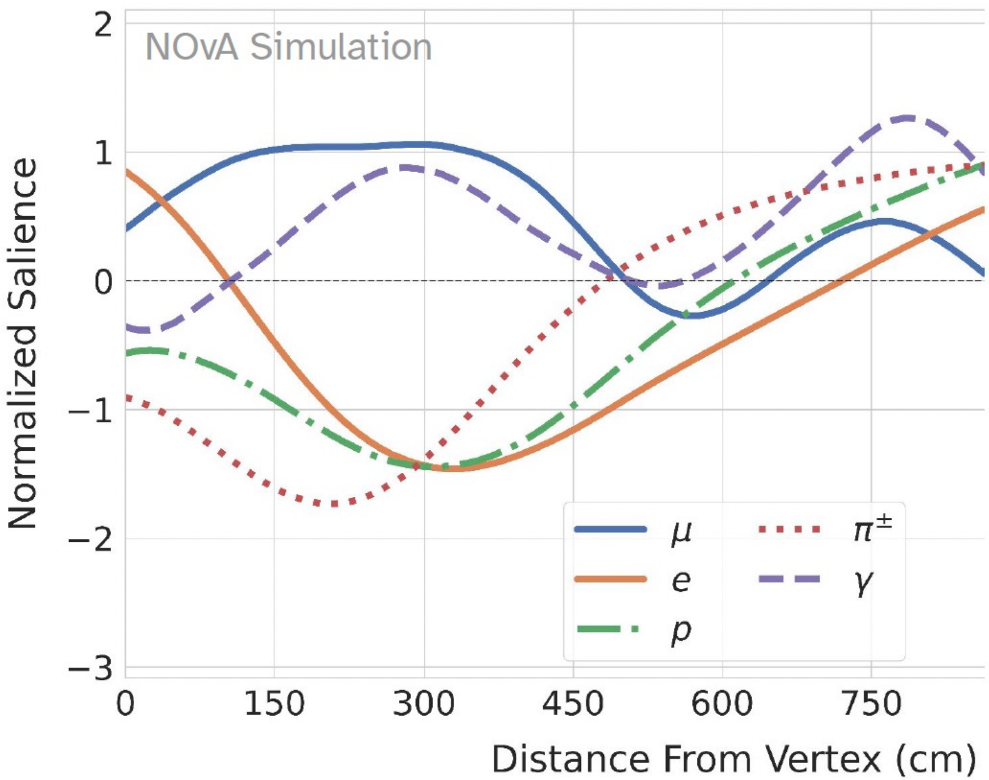
- Truth-Only plots contain only prongs whose truth label matches the Positive Class.
- Diagonal displays the gradients for each class.
- Off-Diagonal elements display Positive Class Saliency - Negative Class Saliency



Red: more likely to predict the given flavor/particle type with more energy in that location.
Green: less likely to predict if there is more activity (anti-correlation).

Integrated Saliency Maps

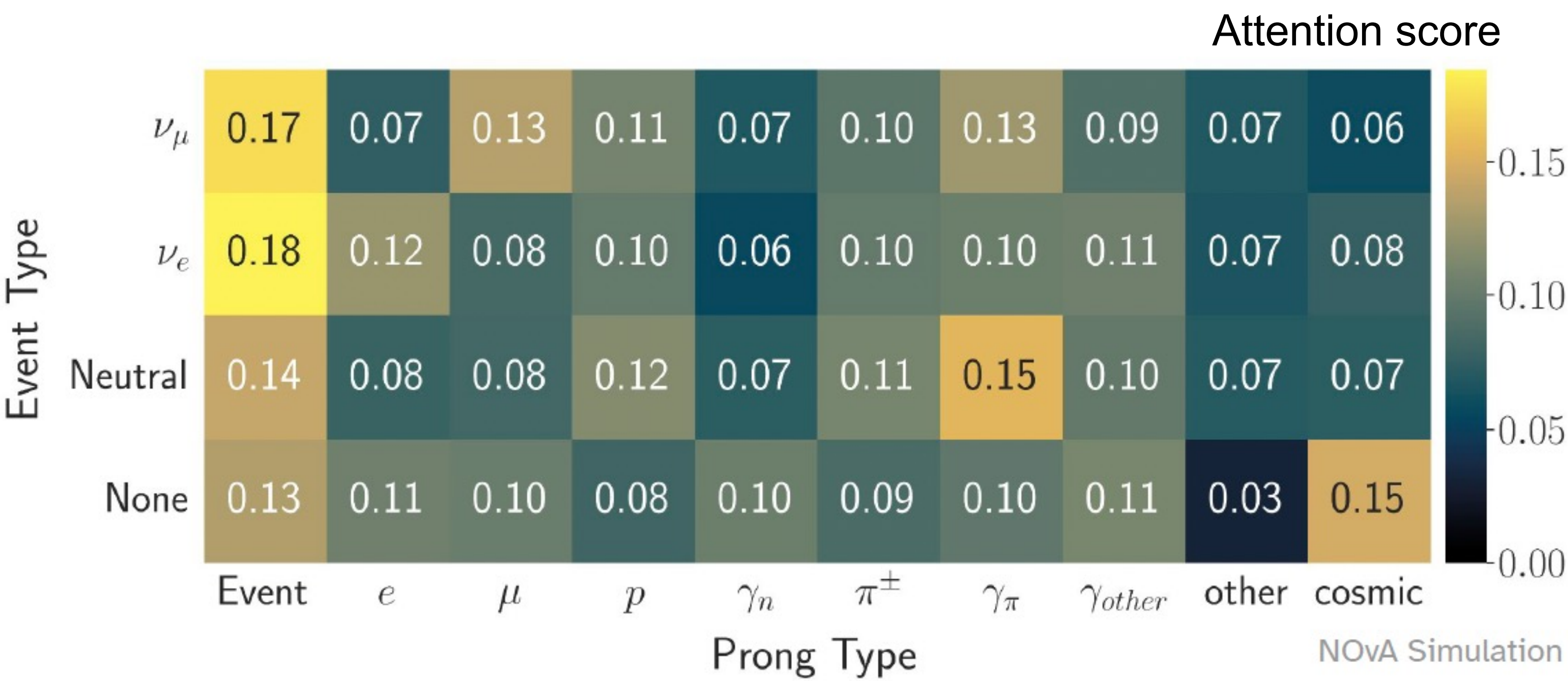
- Comparing 2D maps is challenging, so integrate importance along the width of the detector and plot them all on the same axis w.r.t distance from the vertex
  - Electrons peak early, fall off.
  - Muons have a long, flat profile along track.
  - Photons feature delayed peak.
  - Tail values (>500 cm) tend to go wild due to sparse data in that region



Attention Matrix

Importance of different prong types for classifying the event type

- - e, μ important for corresponding CC events.
- - p and π^0 important for NC.



Summary

- "TransfomerCVN" architecture has been developed for joint event/particle classification
- Performance comparable or better than traditional CNN
- Use Saliency to identify shower/track regions that are important to the final decision
- Use Attention scores to interpret relationships between output event and particle classifications

Thank you!