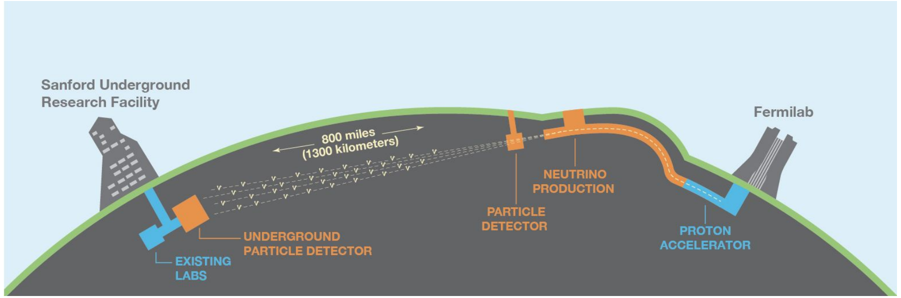
Sparse Convolution Transformers for DUNE FD Event and Particle Classification

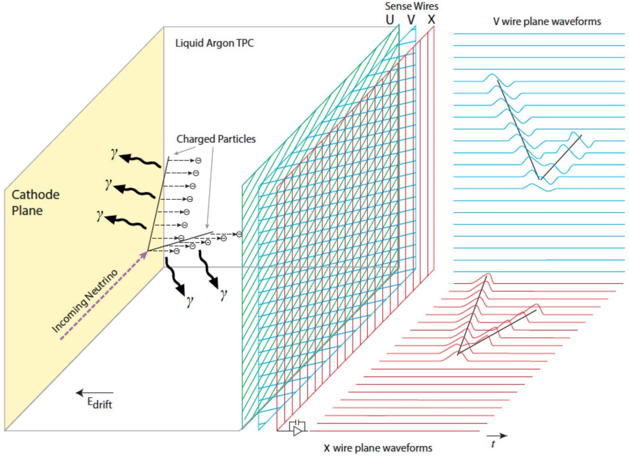
Alejandro Yankelevich and Alexander Shmakov
For the DUNE Collaboration

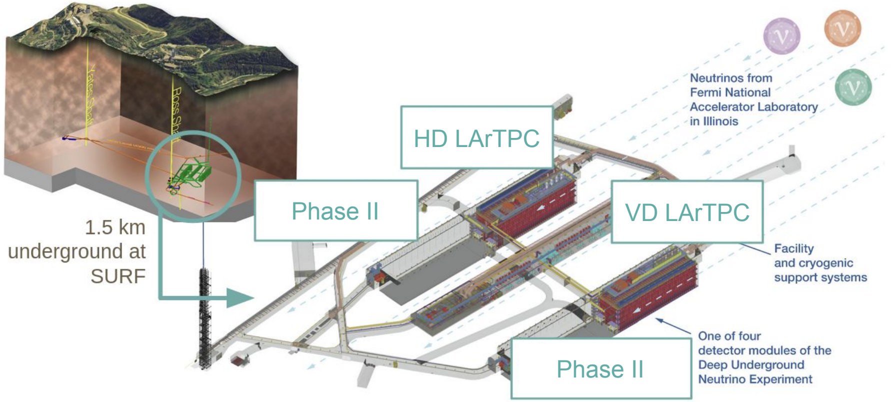
Wire-Cell Reconstruction Summit
Pattern recognition, AI/ML
April 12, 2024

DUNE

- 1300 km baseline neutrino oscillation experiment.
- 17 kt horizontal drift far detector module uses wire plane anodes with three readout planes.

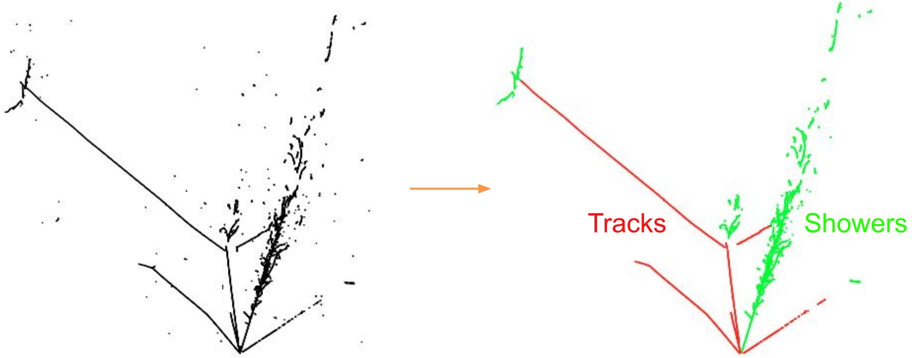


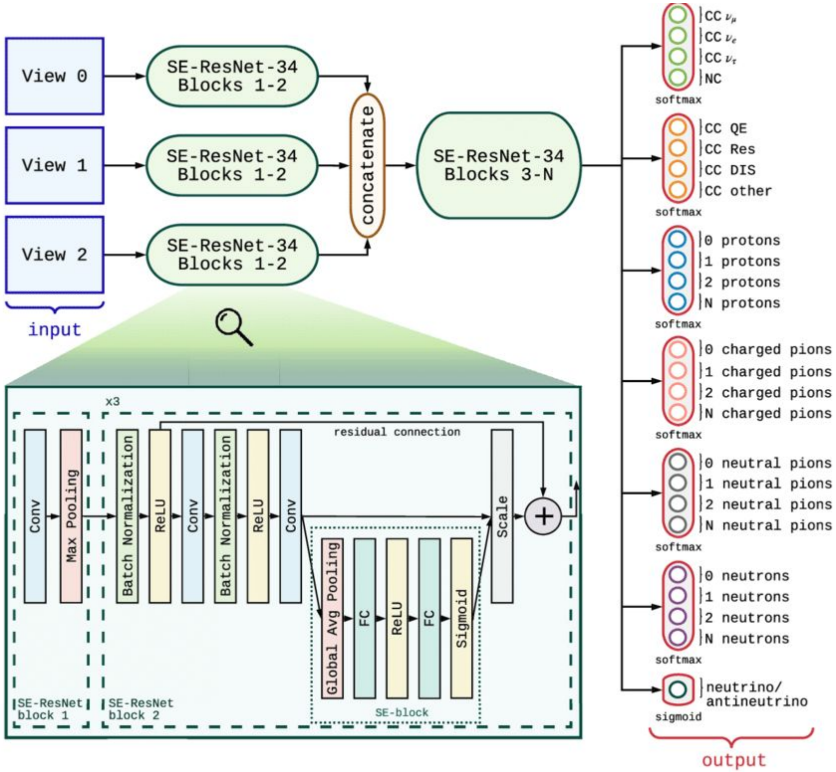




Reconstruction Techniques

- “Pandora” is a collection of reconstruction algorithms including vertexing, track and shower clustering.
- “CVN” network predicts event class and number of secondary particles (protons, pions) in event.

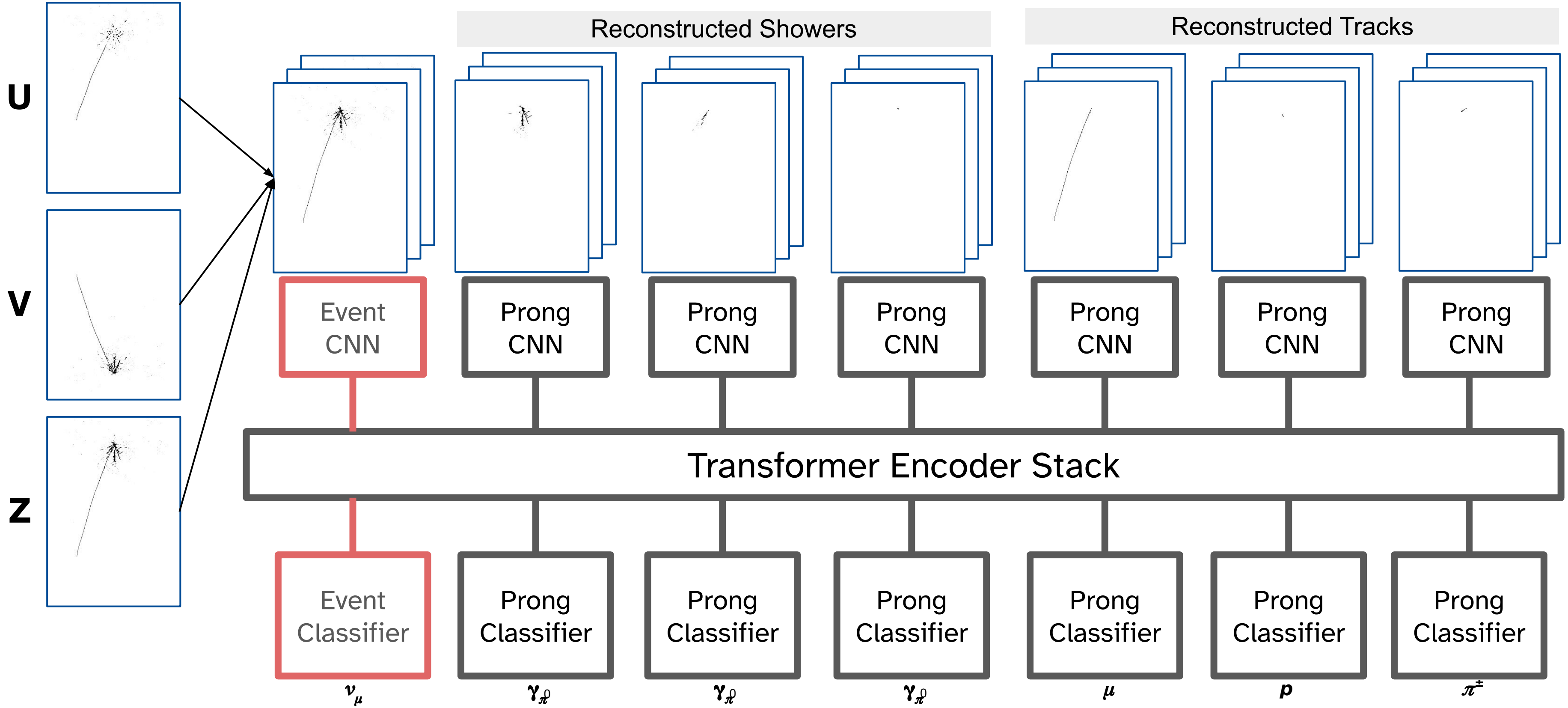




Motivation

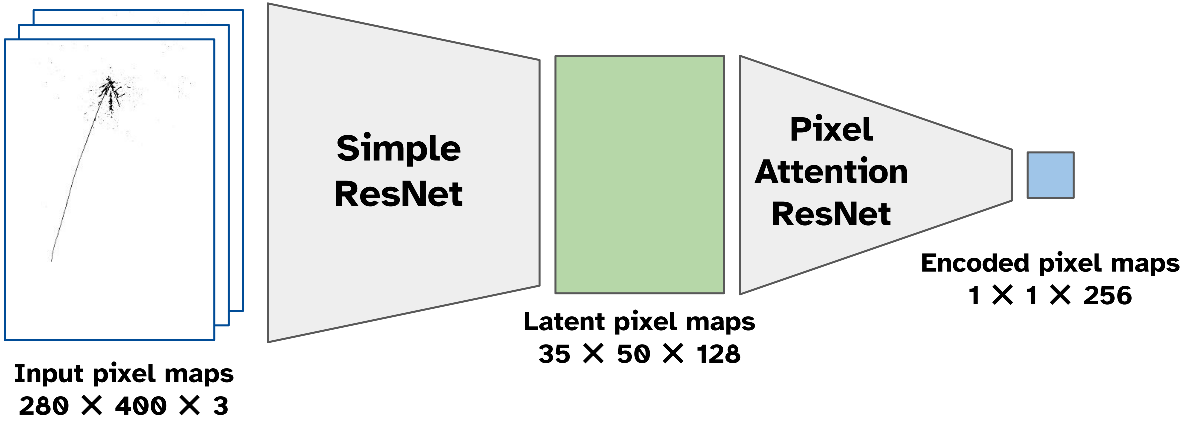
- Want a network that simultaneously uses information from both the overall event LArTPC images and images containing single prongs (reconstructed particle tracks or showers) to identify each prong.
  - Prongs can sometimes only be differentiated by their relationships to other prongs.
  - Provides more insightful interpretability than networks trained on single images.
- Permutation-invariant transformer architecture an ideal candidate for training on variable-length collection of object such as prongs.

TransformerCVN Network Architecture



Embedding LArTPC Images

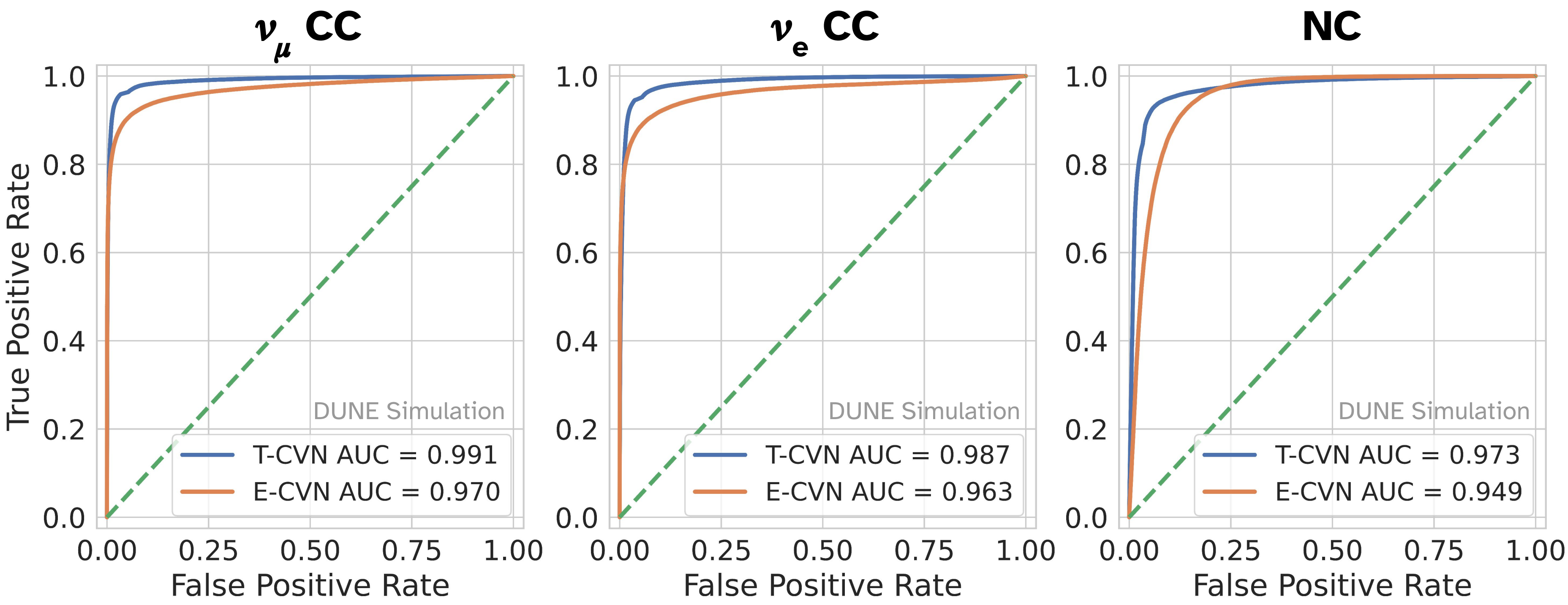
- DUNE LArTPC pixel map images are large, high resolution, and sparse.
- Use attention-based ResNet inspired by Stable Diffusion XL used in AI Art.
  - Combines Convolution and Attention layers to improve the regular ResNet.
  - Extracts pixel features and allows for long-distance information to be shared via attention.
  - Effective for utilizing all information in sparse pixel maps.
  - Smaller latent size for more efficient memory usage.



Training Details

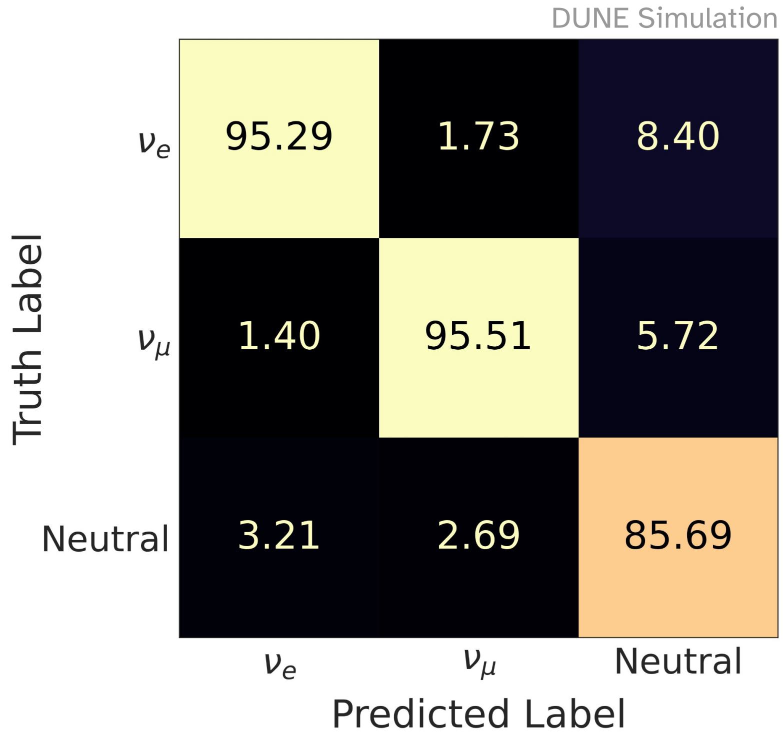
- Train on three 400 x 280 U,V,Z wire plane pixel maps for entire event and each reconstructed prong.
  - Prong pixel maps only contain hits assigned to prong by Pandora track and shower reconstruction.
  - Network size adapts to number of prongs in event, currently capped at 20.
- Predict event as (ν_μ CC, ν_e CC, NC, other).
  - CC: charged current, NC: neutral current
- Predict prongs as (e, μ, p, γ_n, $\pi^\pm$, γ_{π^0}, γ_{other}, other).
  - Subscripts denotes mother particle, but can combine all photons.

Event ROC Curves

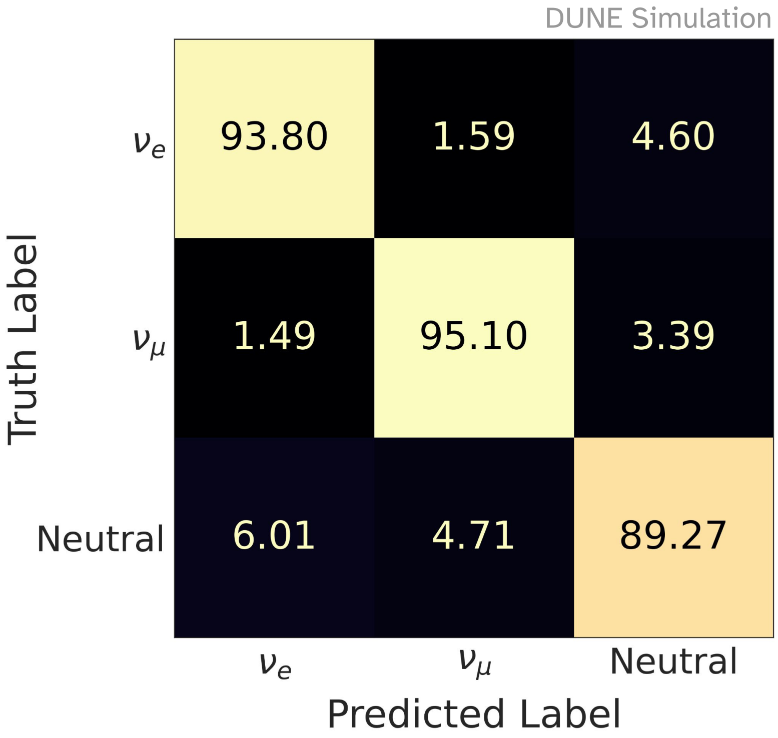


Event Confusion Matrices

Prediction Normalized



Truth Normalized



Prong Confusion Matrices

Prediction Normalized

DUNE Simulation

Truth Label \ Predicted Label	e	μ	p	$\pi^\pm$	γ
e	83.71	1.08	1.53	2.17	8.87
μ	0.67	91.73	0.88	5.80	1.50
p	2.42	1.69	85.59	17.68	5.39
$\pi^\pm$	2.78	3.60	8.72	69.98	6.30
γ	10.42	1.90	3.28	4.37	77.93

Truth Normalized

DUNE Simulation

Truth Label \ Predicted Label	e	μ	p	$\pi^\pm$	γ
e	82.19	0.62	2.24	1.55	13.40
μ	1.08	86.27	2.13	6.80	3.72
p	1.58	0.65	83.90	8.43	5.44
$\pi^\pm$	3.54	2.68	16.60	64.84	12.34
γ	7.47	0.80	3.51	2.28	85.95

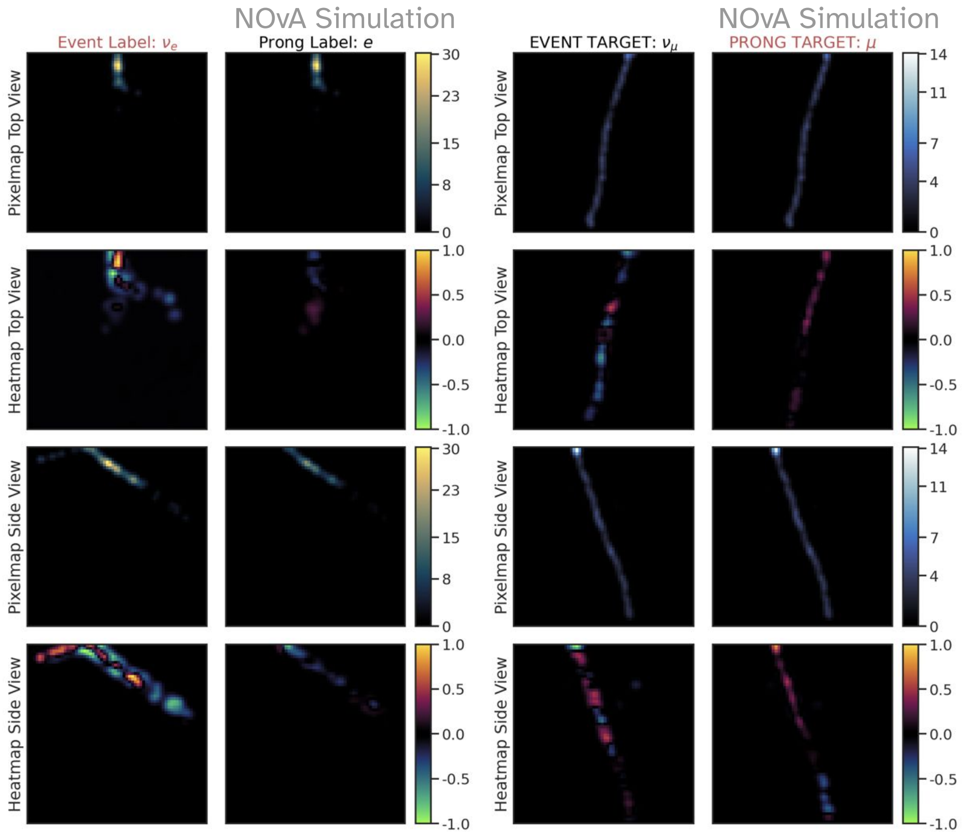
Interpretability

- Examine the internal activations and gradients within the network to determine which aspects of the input pixel maps cause one prediction over another.
- Pixel Gradients (Salience)
  - The gradient classification probability with respect to hits in a given location.
  - When aggregated, provides a template of a typical event for each prong type.
  - Red means more likely to predict the given prong type with more energy in that location.
  - Blue means less likely to predict if there is more activity (anti-correlation).
- Attention Scores
  - Indicate the importance of different elements to the events.
  - Allows us to find out which prongs are used for different types of events and prongs.

Note: These interpretability studies have not been conducted yet for DUNE simulation. The following slides show studies with a network trained on NOvA 2 x 100 x 80 pixel map images.

Individual Saliency

- Can calculate importance of different hits for each event.
- Useful for debugging wrong predictions.
  - Challenging to verify any learned physics.
  - When aggregated, provides a template of a typical event for each prong type.
- Need to aggregate multiple prongs of the same type to find patterns.

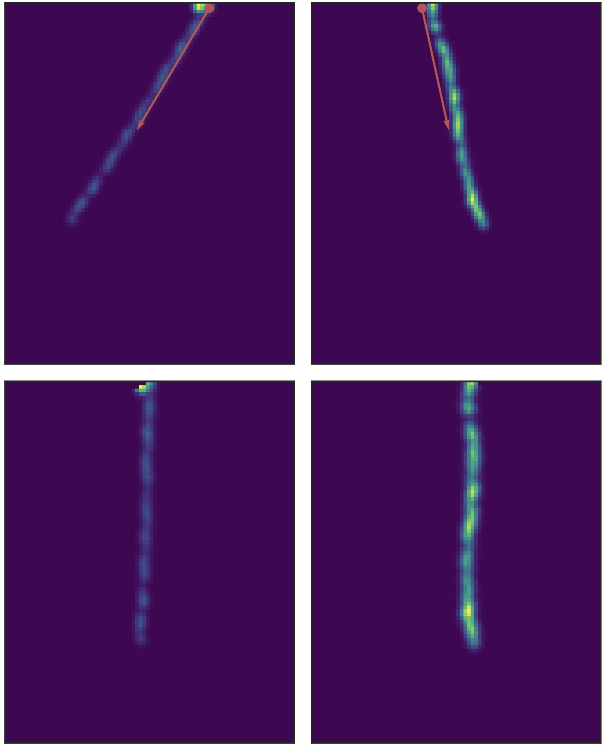


Example saliency maps for νe event prediction of ν_e event (left) and μ prong prediction of ν_μ CC event (right).

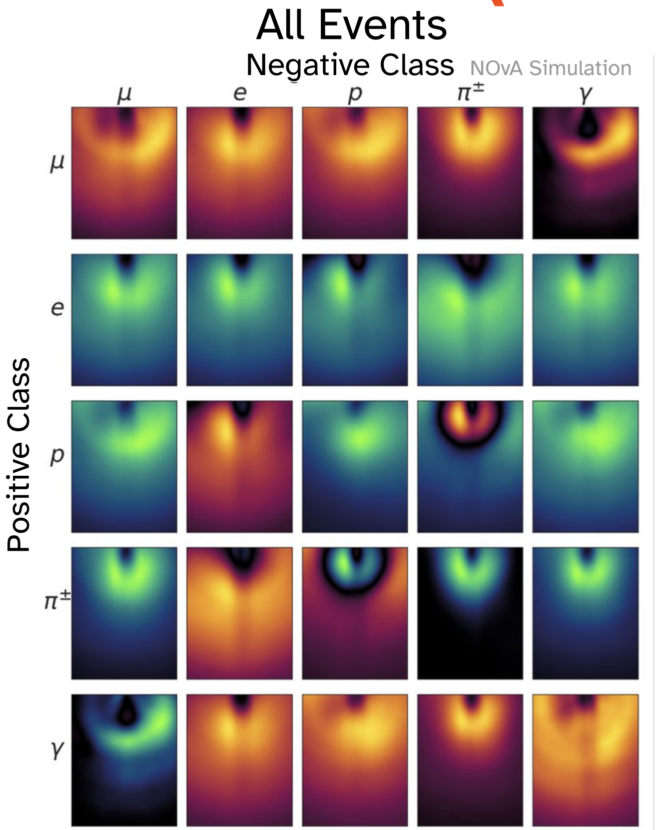
Salience Aggregation

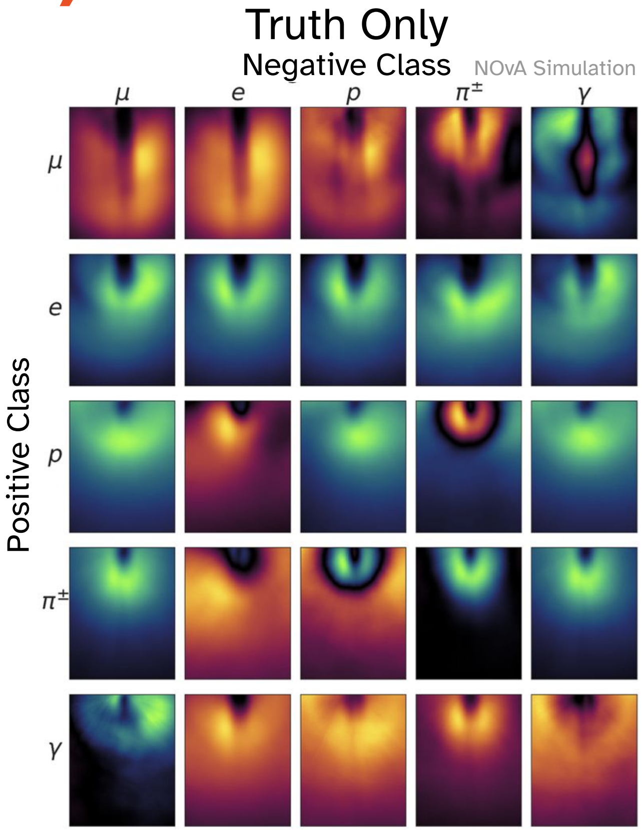
- Calculated these values for every prong in ~10,000 events and every possible output of the network.
- Rotated, translated, and averaged each image using the vertex information associated with each prong.
  - Every prong forced to have vertex at (40, 0) and facing toward +z.
  - Possibly limit event by track length to compare similar lengths.
- Plot grid of results.
  - Diagonal displays the gradients for each class.
  - Off-Diagonal elements display difference between the two classes (“Positive Class” - “Negative Class”).
  - Truth-Only plots contain only prongs whose truth label matches the Positive Class.

Pixel Map Alignment



Pixel Gradients (Salience)

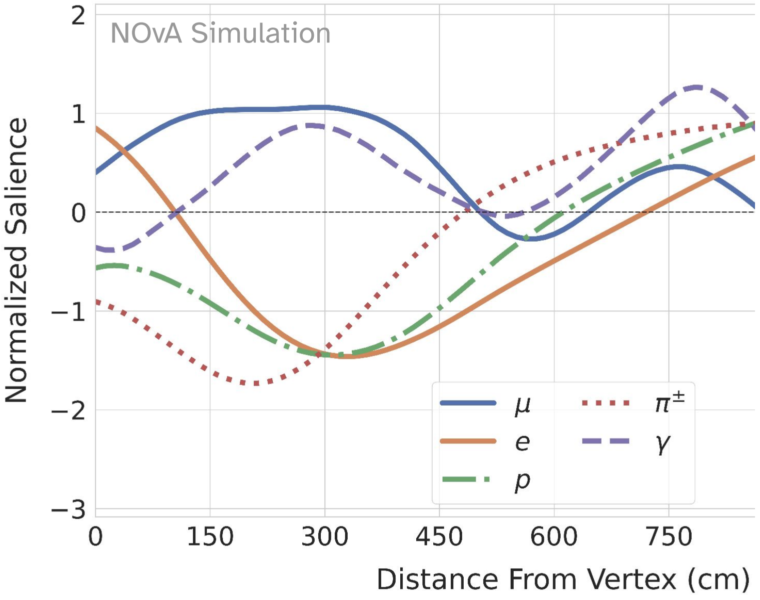




Integrated Importance Maps

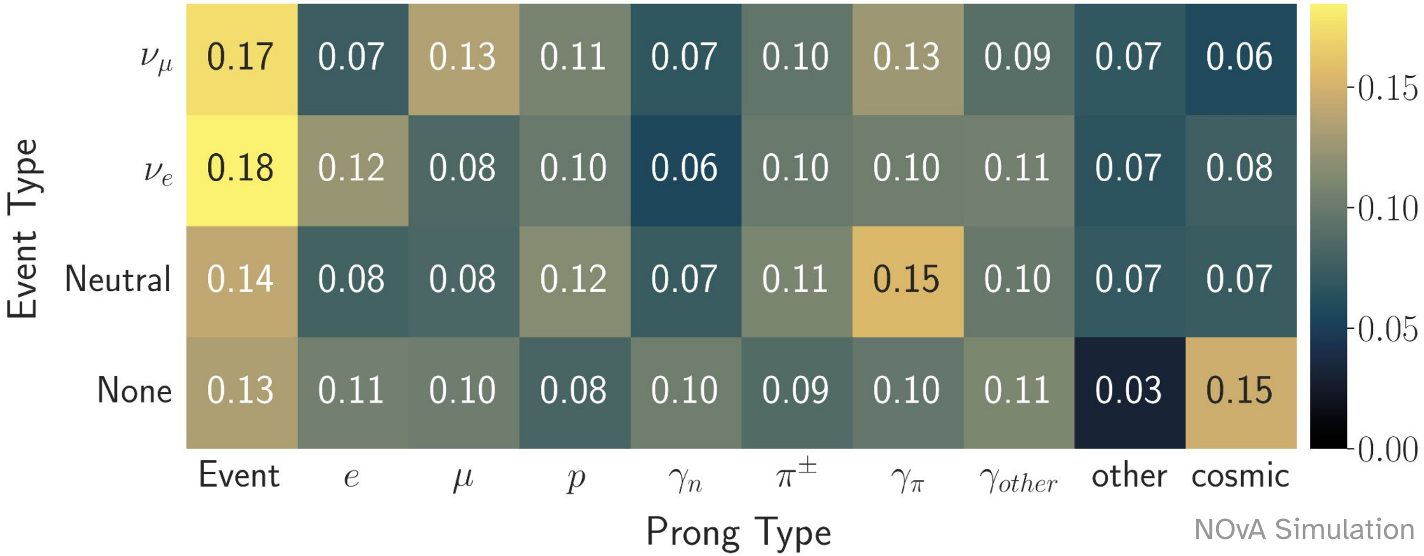
- Comparing 2D maps is challenging, so integrate importance along the width of the detector and plot them all on the same axis w.r.t distance from the vertex.

- 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 π0 for important for NC.



Summary and Future Work

- "TransfomerCVN" architecture uses contextual learning capability of transformer networks for improved joint event/particle classification.
- Attention-based CNN makes approach feasible for large LArTPC images without large memory impact.
- Will quantify potential benefits to DUNE oscillation analysis.
- Will use increased interpretability potential to conduct studies on relationships between event/particle types and importances of inputs.