TransformerCVN: Convolution Transformers for NOvA Event and Particle Classification

Alexander Shmakov, Alejandro Yankelevich, Jianming Bian, Pierre Baldi

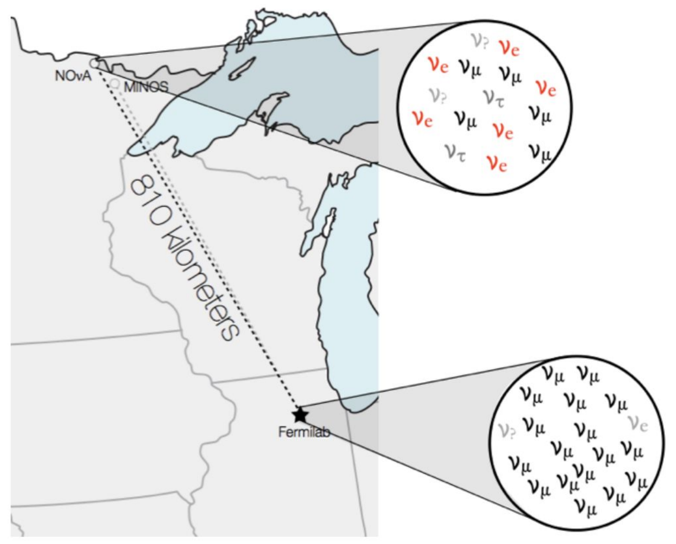
Background

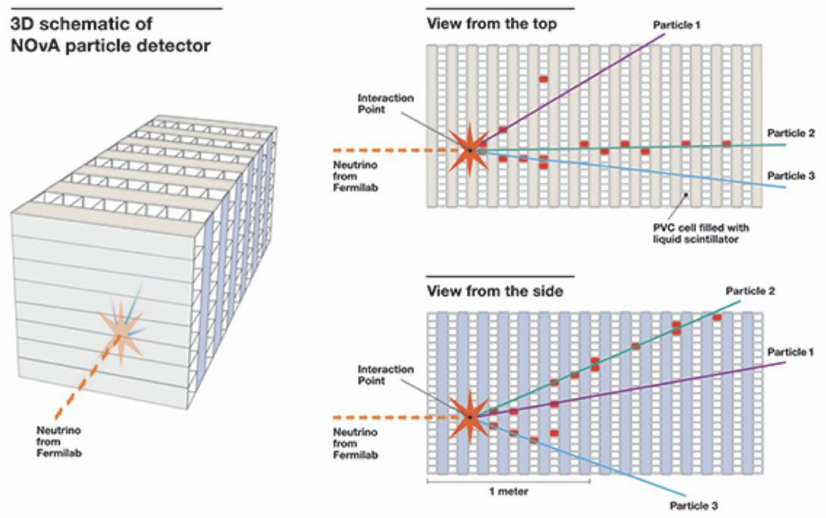
NOvA is a long-baseline neutrino oscillation experiment using Fermilab's NuMI beam [1]. The experiment consists of two functionally identical detectors 809km apart formed from plastic extrusions filled with a liquid scintillator.

The NuMI beam is predominantly ν_μ. The collaboration studies neutrino oscillations by detecting the rate of disappearance of ν_μ and appearance of ν_e at the NOvA far detector through charged current interactions. The goal of event reconstruction is therefore to identify each particle's cluster of scintillator cell hits known as a "prong" and to classify the overall event interaction type:

1. ν_μ charged current (produces μ)
2. ν_e charged current (produces e)
3. Neutral current (flavor cannot be determined)
4. Cosmic ray muons

Signal: 1–2
Background: 3–4

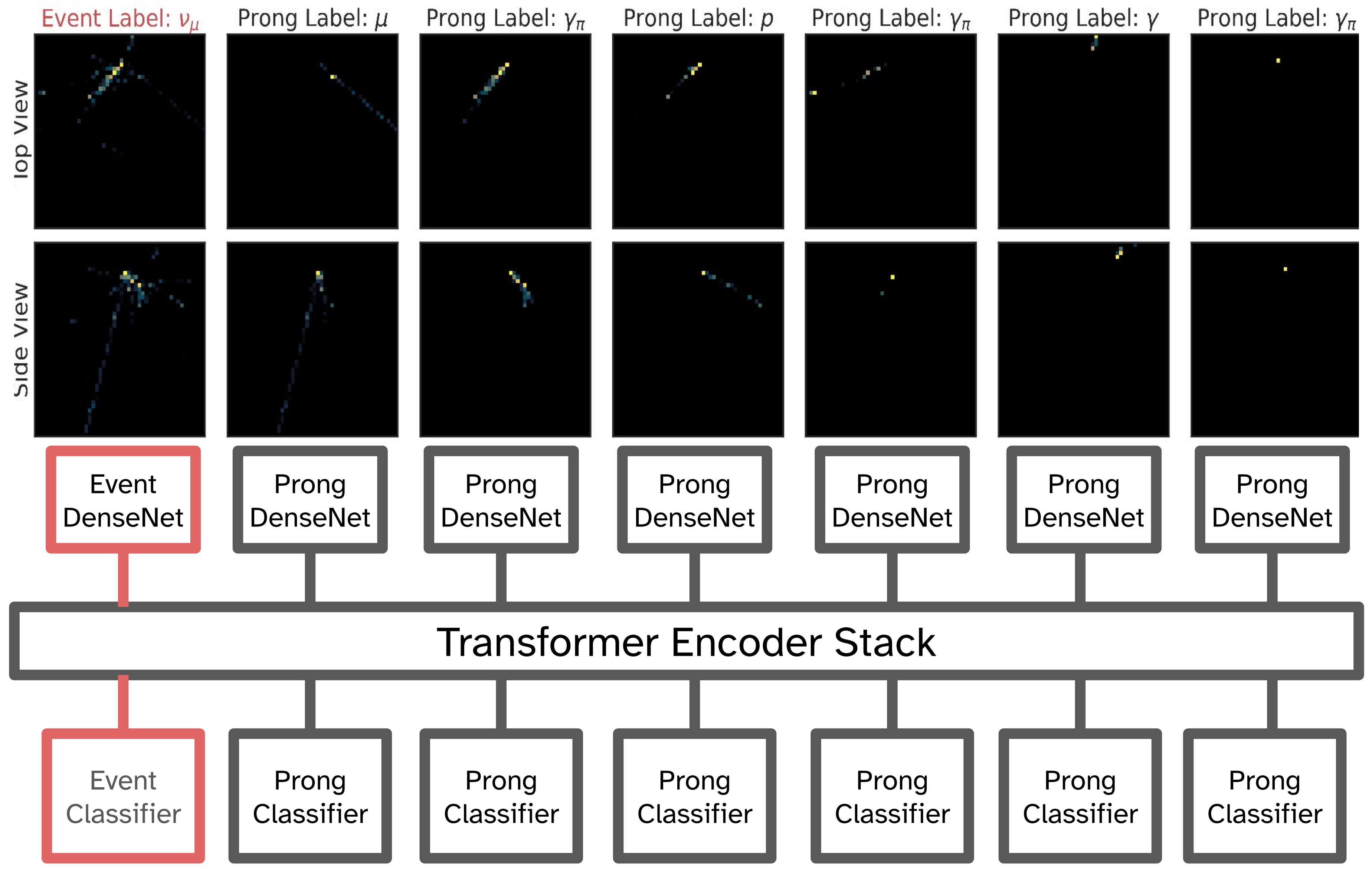




TransformerCVN

Two separate image-trained neural networks known as Convolutional Visual Networks (CVN) are currently used for the tasks of prong particle identification and overall event classification [2]. We introduce a combined end-to-end architecture to process and evaluate both the event and all associated prongs in a single network.

1. Events are preprocessed and cropped into a list two images (xz and yz view) each with size 100×80 pixels. We referred to these images as *pixel maps*.
2. Event pixel map is generated containing all hits in the event as well as pixel maps containing only hits assigned to a single prong using NOvA's prong reconstruction.
3. Pixel maps are embedded using a *DenseNet* [3].
   - DenseNet consists of a stack of convolution layers with skip connections between all layers, improving embedding of sparse prong pixel maps.
   - All prong pixel maps are fed through a single shared DenseNet.
   - A separately trained DenseNet is used for the event pixel maps.
4. Embedded Pixel maps are then processed by a stack of transformer encoders [6].
   - Transformers facilitate extraction of contextual information about prongs, processing each prong with respect to all other prongs in the event.
   - This allows information to be shared between prongs and allows the event side of the network to have access to prong-specific information.
5. Encoded pixel maps are processed by feed-forward classifiers to produce the event and prong reconstruction predictions.
   - A single shared classifier is used for all encoded prong pixelmaps.
   - A separate classifier is used for the encoded event pixelmap.

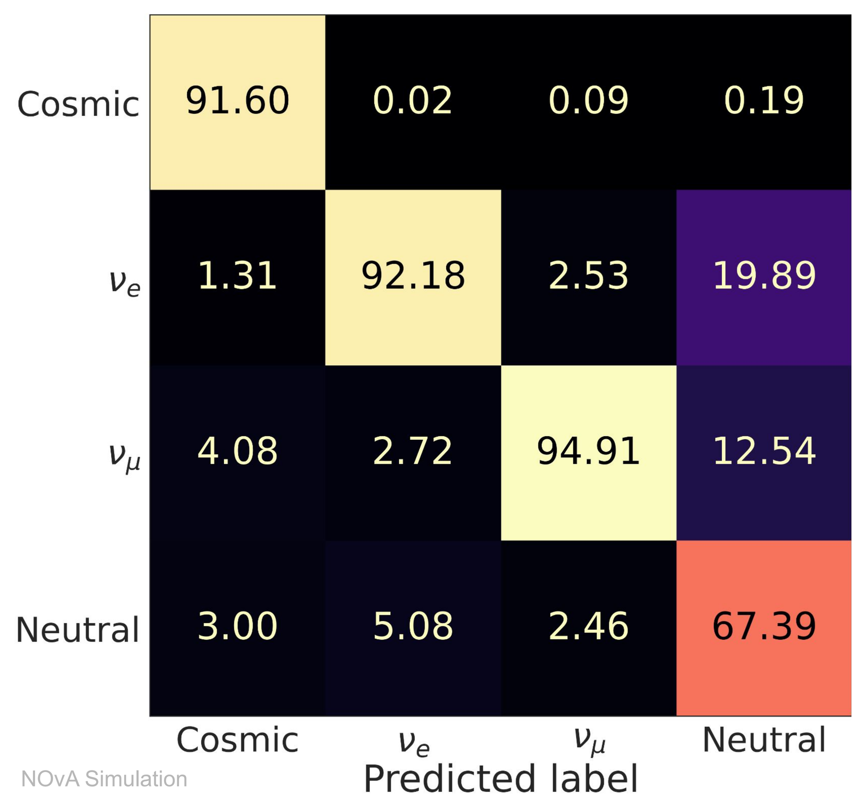


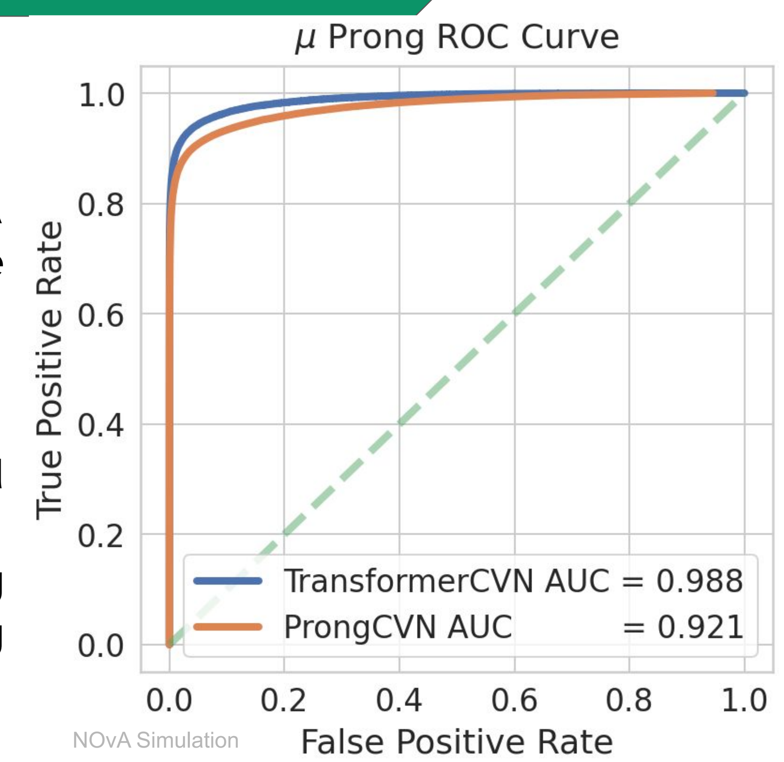
A complete diagram of *TransformerCVN*, including example pixel maps from a ν_μ current event. The separate *event path* is highlighted in red. The truth labels for each pixel map is provided at the top of the figure.

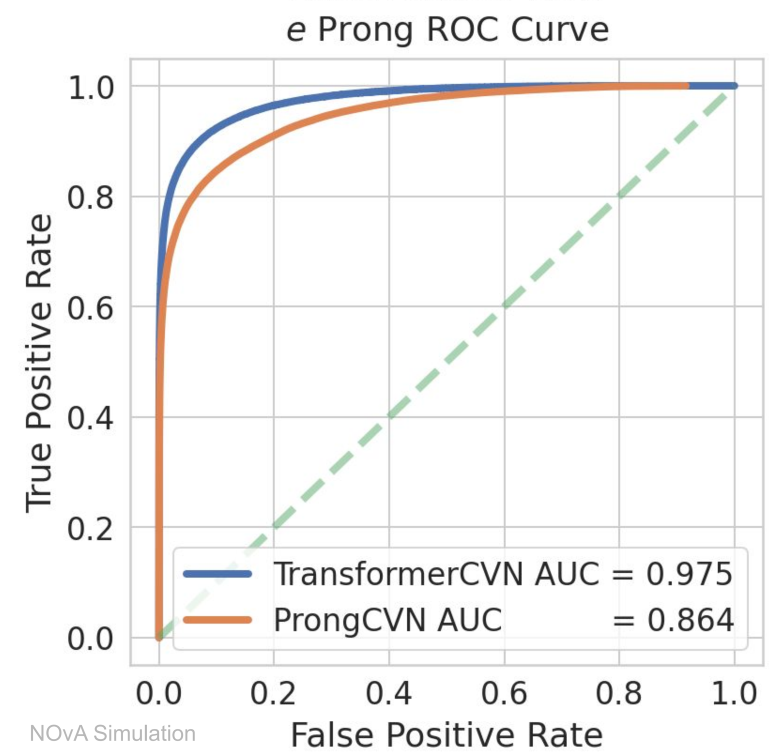
Results

Event Reconstruction TransformerCVN purity confusion matrix for event classification on NOvA Monte-Carlo data. Performance matches the specialized EventCVN.

Prong Reconstruction ROC curves for prong classification for electron and muon prongs and compare TransformerCVN to the ProngCVN. The end-to-end unified architecture improves prong reconstruction over the ProngCVN by exploiting contextual information.







Interpretability

TransformerCVN's end-to-end architecture allows us to generate *heatmaps* indicating correlations between different regions in the pixel maps and the network's predicted labels.

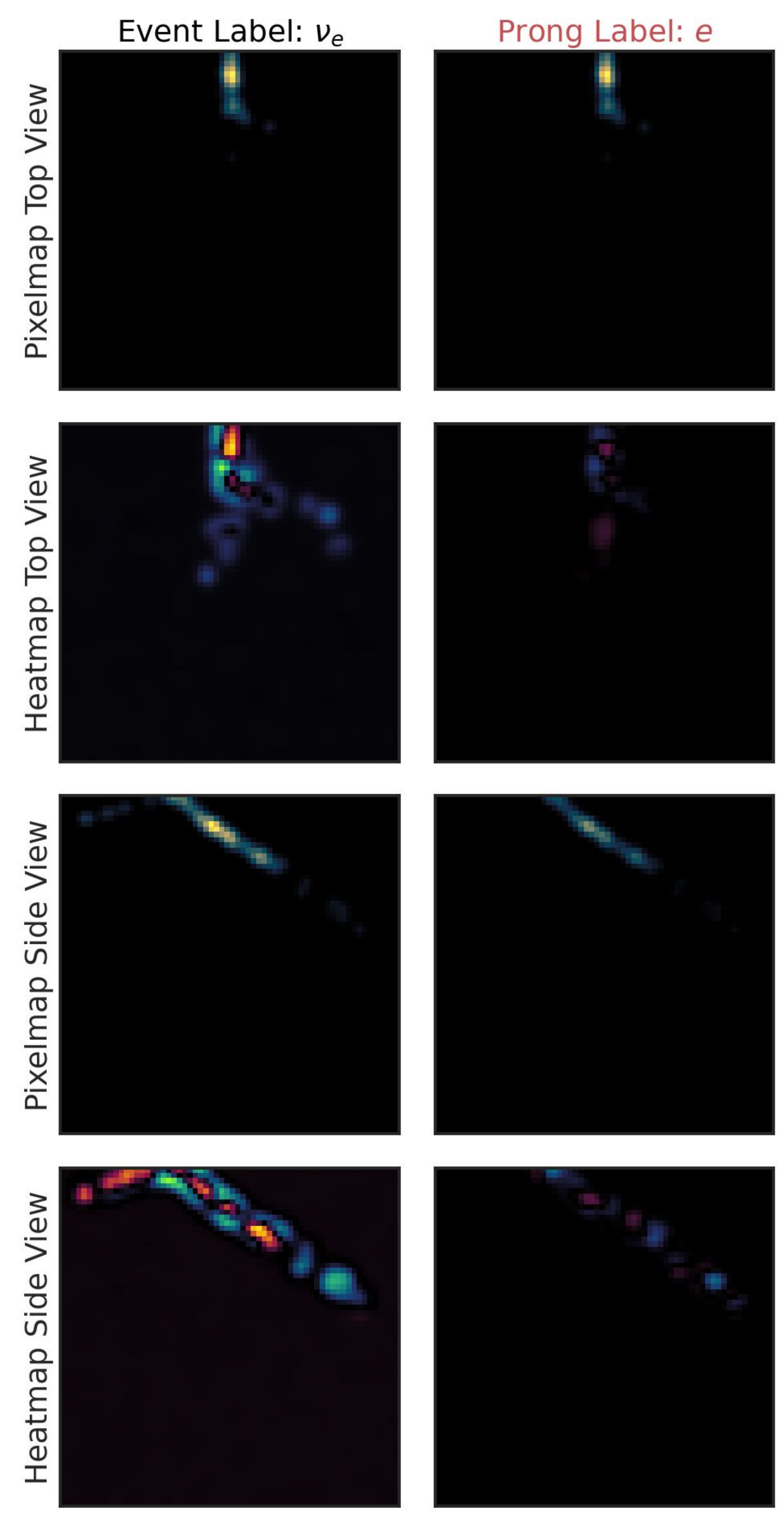
- Allows us to *peek* inside the network's reasoning behind the classifications assigned to particular events.
- Allows us to interpret which aspects of the input result in the prediction.
- Provides a method for analyzing incorrect reconstructions.

Presented to the right is an example heatmap for a ν_e current event.

- Regions of the heatmap which are red are **positively correlated** with the electron prong prediction
- Blue regions are **negatively correlated** with the electron label.

Heatmaps are generated by

1. Computing the gradient of the electron prong classification output *w.r.t* the input pixel map [4],
2. Weighting the gradient by the Class Activation Mapping (CAM) [5] from the DenseNet convolution layers
3. Weighting the gradient by the attention weights from the transformer encoder layers [6] to create full heatmap list.



References

[1] D. S. Ayres et al. The NOvA Technical Design Report. 10 2007.
[2] A. Aurisano, *et al.* A convolutional neural network neutrino event classifier. JINST, volume 11, P09001, 2016
[3] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.
[4] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In 2017 IEEE International Conference on Computer Vision (ICCV), pages 618–626, 2017.
[5] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. In Workshop at International Conference on Learning Representations, 2014.
[6] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, L ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, Advances in Neural Information Processing Systems, volume 30. Curran Associates, Inc., 2017.