Neutrino Event Classification with Deep Learning in NOvA

Grant Nikseresht

Department of Computer Science, Illinois Institute of Technology

June 11 - New Perspectives 2019
Neutrino Event Classification

Deep learning is used to classify $\nu_\mu$, $\nu_e$, neutral current, and cosmic events based on the topology of the interactions.

Figure 1: NOvA neutrino events.
Cell hits are captured along $XZ$ and $YZ$ planes and converted to pixel maps.

One charged current $\nu_\mu$ event (right) with two corresponding maps is a typical input to an event classifier.
Convolutional Neural Networks

- **Convolutional neural networks** (CNNs) learn features from images based only on pixels.
- Can be applied directly to images with minimal reconstruction, forgoing the need for manual feature extraction.
- CNNs have become the state-of-the-art technique for image detection problems across many fields.
Convolutional Visual Network

Figure 2: CVN Classic architecture. A pair of views for each event enter separate channels on the left and are mapped to an output class in the final fully connected layer.

- **Convolutional Visual Network** (CVN) was inspired by GoogLeNet [1] and characterized by a two tower structure and a sequence of Inception modules.
- CVN improved effective exposure by 30% for $\nu_e$ events and 10% for $\nu_\mu$ events compared to non-deep learning based reconstruction methods [2].
- In 2018, improvements in classifier efficiency were found by training RHC and FHC separately and by pruning layers from CVN Classic to create **CVN ShortSimple**.
Improving CVN

- There's a dedicated effort in NOvA to continue searching for new deep learning architectures and methods that can:
  1. Decrease the model complexity (number of trainable weights)
  2. Reduce training and inference time
  3. Maximize classifier performance in terms of efficiency and purity

- Two of these efforts are covered in the remainder of this talk:
  1. Development of CVN architectures based on residual learning
  2. Hyperparameter optimization using the DeepHyper framework at Argonne National Laboratory (ANL)

- All results here are preliminary and trainings are not final, but are shown to illustrate relative performance of different methods

- Models are compared using the harmonic mean of purity and efficiency,

\[ F_1 = 2 \times \frac{\text{purity} \times \text{efficiency}}{\text{purity} + \text{efficiency}} \]
Residual Learning

**Residual networks** (ResNet) are made up of chains of *residual blocks*

ResNet has been shown to outperform GoogLeNet on benchmark image classification problems [3]

**Residual blocks** are characterized by a *skip connection* that adds the input to a block back at the end

ResNets allow for training of much deeper networks
Parallel View ResNet

- ResNet50 was modified to accept XZ and YZ views as input to follow the two tower CVN methodology.
- Each view is passed through separate channels of 3 residual blocks before being combined and sent through another 13 residual blocks.
- Batch normalization, convolutional, dense, and pooling layers are contained in the residual block in between the skip connection.
Neutrino event classifiers were trained on a sample of 6.1 million Monte Carlo events using CVN ShortSimple (left) and ResNet50 (right) architectures.

ResNet50 continues to show performance improvement after CVN ShortSimple plateaus.
Model Performance - Efficiency Matrices

**Figure 3:** CVN ShortSimple, $F_1 = 0.8702$

Classification efficiency matrices for each classifier - ResNet50 achieves a higher efficiency for each class compared to CVN ShortSimple.

**Figure 4:** ResNet50, $F_1 = 0.8989$
Model Performance - ROC Curves

Figure 5: CVN ShortSimple

Figure 6: ResNet50

Receiver Operating Characteristic (ROC) curves for binary classifiers - ResNet50 achieves higher areas under the curve (AUC) for each class indicating better overall classifier performance.
Hyperparameter Optimization

- Training of neural networks is governed by hyperparameters that determine how the optimal model weights are found by gradient descent.
- For neural networks, heuristics are often employed for choosing hyperparameter values when computational resources are limited.
- Using ANL’s Cooley GPU cluster, we utilize a framework called DeepHyper [4] which allows for Bayesian hyperparameter optimization at scale.
DeepHyper

Figure 7: All trials over a 12 hour period

Figure 8: Subset of trials where validation loss was less than 1.5

Learning rate optimization using DeepHyper - a new learning rate is selected for each trial based on a maximum likelihood model.
Conclusion

- Improvement of CVN is one of the most critical ways to improve NOvA’s event selection
- Residual networks are a natural next step in the evolution of CVN and have shown promise for increasing event classification efficiency
- Scalable hyperparameter optimization can improve the performance of existing models that could be underoptimized
- Architectural and algorithmic improvements to CVN can improve a multitude of applications in NOvA including classification of prongs, cosmic rejection, and semantic segmentation
References I


Backup Slides
Model Performance - Purity Matrices

Figure 9: CVN ShortSimple, $F_1 = 0.8702$

Figure 10: ResNet50, $F_1 = 0.8989$
ResNet PIDs

Figure 11: CVN PIDs by class output from a trained ResNet50 model
CVN ShortSimple Architecture
Inception Modules

- **Inception modules** are a defining architectural feature of GoogLeNet.
- Convolutions are taken at different scales and the results are concatenated to extract features at different spatial resolutions.