Context Enriched Prong CNN performance studies in NOvA

Akshay Chatla

New Perspectives 2021
16th-19th August 2021

August 19, 2021
The NOvA Experiment

- NOvA is a long baseline neutrino experiment with a near detector at 1 km and a far detector at 810 km.
- The Physics goals:
  - study of neutrino oscillations \((\theta_{23}, \delta_{13}, \Delta m^2_{32})\)
  - neutrino cross-section measurements
  - exotic physics
- These goals need good neutrino interaction classification and neutrino energy measurement. Machine learning plays an important role in achieving them.
Event Classification At NOvA

- NOvA was the first HEP experiment to use a Convolutional Neural Network (CNN) in a physics measurement to classify candidate neutrino interactions.
- NOvA uses Event CNN[1] which takes full event pixel map of both x and y views as input and the output is an event classifier.
- Event CNN can classify the events into 4 classes:
  1. $\nu_\mu$ CC interaction
  2. $\nu_e$ CC interaction
  3. NC interaction
  4. Cosmics

Particle classification

- While Event CNN can classify events, identification of final state particles of an event is needed to better our energy reconstruction and enable cross-section measurements of final states.
- Context enriched Prong CNN[2] is one of the networks being trained with a goal to identify all the final-state particles \((e^\pm, p^+, \mu, \pi^\pm, \gamma)\) of a given neutrino event.

Prong CNN[2] uses a four-tower siamese-type MobileNetV2 architecture for including context information i.e it takes both event (Context) views and prong (Independent) views.

Providing context to the network helps it to learn from physical quantities implying conservation laws as well as the relative topology of the particle with respect to the accompanying activity.

**Motivation**

- The previous iteration of Prong CNN used by NOvA trained on neutrino (FHC) and anti-neutrino (RHC) dataset separately. In this work, the network is trained on a combined (FRHC) neutrino and anti-neutrino dataset and compared with the separately trained networks.

- Following selection cuts are used on the training sample:

<table>
<thead>
<tr>
<th>Selection cuts</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Containment Cut</td>
<td>Selects the prong and the event contained within the detector boundaries</td>
</tr>
<tr>
<td>Cosmic Veto</td>
<td>Removes cosmic events</td>
</tr>
<tr>
<td>Purity Cut</td>
<td>Realistic looking cluster with prongs $\mu^\pm, \gamma : 50%$, $e^\pm : 40%$, $\pi^\pm, p^+ : 35%$</td>
</tr>
<tr>
<td>Prong length</td>
<td>Cut prongs with prong length more than 5 m</td>
</tr>
</tbody>
</table>
Before training the whole dataset, I took a part of FD simulated datasets for preliminary training. FHC had 1063406, RHC had 886712 prongs, Combined: 1950118
This training dataset is unbalanced, one needs to train the network on a balanced dataset to make sure that network does not learn bias towards the class with more samples.

Since, our training dataset is unbalanced, it is later balanced to contain an equal number of each type of particle.
Now, we have a choice of 4 kinds of datasets we can train.

1. A balanced neutrino dataset (FHC)
2. A balanced anti-neutrino dataset (RHC)
3. A half of $\nu + \bar{\nu}$ dataset (0.5 FHC + 0.5 RHC)
4. A double $\nu + \bar{\nu}$ dataset (FHC + RHC)

The balanced datasets look as follows:

For ease of comparison, I balanced the datasets to have the same number of total events, the 4th dataset will have double the data.
Only FHC Training/Evaluation

**FHC Network**

- **Loss**
- **Accuracy**
- **Validation Loss**
- **Validation Accuracy**

**NOvA Simulation**

<table>
<thead>
<tr>
<th>True Label</th>
<th>Electron</th>
<th>Muon</th>
<th>Proton</th>
<th>Pion</th>
<th>Gamma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electron</td>
<td>0.78</td>
<td>0.05</td>
<td>0.03</td>
<td>0.05</td>
<td>0.1</td>
</tr>
<tr>
<td>Muon</td>
<td>0.05</td>
<td>0.77</td>
<td>0.03</td>
<td>0.11</td>
<td>0.05</td>
</tr>
<tr>
<td>Proton</td>
<td>0.05</td>
<td>0.04</td>
<td>0.37</td>
<td>0.28</td>
<td>0.27</td>
</tr>
<tr>
<td>Pion</td>
<td>0.08</td>
<td>0.13</td>
<td>0.21</td>
<td>0.39</td>
<td>0.19</td>
</tr>
<tr>
<td>Gamma</td>
<td>0.21</td>
<td>0.02</td>
<td>0.17</td>
<td>0.13</td>
<td>0.47</td>
</tr>
</tbody>
</table>

Predicted Label
Only RHC Training/Evaluation

RHC Network

NOvA Simulation

Efficiency of RHC Network on RHC Dataset

<table>
<thead>
<tr>
<th>True Label</th>
<th>Electron</th>
<th>Muon</th>
<th>Proton</th>
<th>Pion</th>
<th>Gamma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electron</td>
<td>0.81</td>
<td>0.02</td>
<td>0.06</td>
<td>0.04</td>
<td>0.07</td>
</tr>
<tr>
<td>Muon</td>
<td>0.02</td>
<td>0.77</td>
<td>0.04</td>
<td>0.14</td>
<td>0.02</td>
</tr>
<tr>
<td>Proton</td>
<td>0.02</td>
<td>0.03</td>
<td>0.57</td>
<td>0.22</td>
<td>0.16</td>
</tr>
<tr>
<td>Pion</td>
<td>0.02</td>
<td>0.1</td>
<td>0.27</td>
<td>0.47</td>
<td>0.14</td>
</tr>
<tr>
<td>Gamma</td>
<td>0.14</td>
<td>0.02</td>
<td>0.33</td>
<td>0.2</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Epochs

loss
acc
val_loss
val_acc
0.5 FHC + 0.5 RHC Training/Evaluation

0.5 FHC + 0.5 RHC Network

Efficiency of 0.5 FHC+0.5 RHC Network on FHC+RHC Dataset

- Electron
  - True Label: 0.75
  - Predicted Label: 0.05
- Muon
  - True Label: 0.01
  - Predicted Label: 0.79
- Proton
  - True Label: 0.02
  - Predicted Label: 0.06
- Pion
  - True Label: 0.02
  - Predicted Label: 0.34
- Gamma
  - True Label: 0.14
  - Predicted Label: 0.04

NOvA Simulation
Double FHC + RHC Training/Evaluation

FHC + RHC Network

NOvA Simulation

Efficiency of FHC+RHC Network on FHC+RHC Dataset

<table>
<thead>
<tr>
<th>True Label</th>
<th>Electron</th>
<th>Muon</th>
<th>Proton</th>
<th>Pion</th>
<th>Gamma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electron</td>
<td>0.8</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
<td>0.1</td>
</tr>
<tr>
<td>Muon</td>
<td>0.01</td>
<td>0.82</td>
<td>0.01</td>
<td>0.13</td>
<td>0.03</td>
</tr>
<tr>
<td>Proton</td>
<td>0.01</td>
<td>0.01</td>
<td>0.7</td>
<td>0.16</td>
<td>0.12</td>
</tr>
<tr>
<td>Pion</td>
<td>0.01</td>
<td>0.14</td>
<td>0.16</td>
<td>0.52</td>
<td>0.18</td>
</tr>
<tr>
<td>Gamma</td>
<td>0.17</td>
<td>0.03</td>
<td>0.08</td>
<td>0.16</td>
<td>0.56</td>
</tr>
</tbody>
</table>

Epochs

loss
acc
val_loss
val_acc
Network Comparison

NOvA Simulation

Efficiency of RHC Network on RHC Dataset

<table>
<thead>
<tr>
<th></th>
<th>Electron</th>
<th>Muon</th>
<th>Proton</th>
<th>Pion</th>
<th>Gamma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electron</td>
<td>0.81</td>
<td>0.02</td>
<td>0.06</td>
<td>0.04</td>
<td>0.07</td>
</tr>
<tr>
<td>Muon</td>
<td>0.02</td>
<td>0.77</td>
<td>0.04</td>
<td>0.14</td>
<td>0.02</td>
</tr>
<tr>
<td>Proton</td>
<td>0.02</td>
<td>0.03</td>
<td>0.57</td>
<td>0.22</td>
<td>0.16</td>
</tr>
<tr>
<td>Pion</td>
<td>0.02</td>
<td>0.1</td>
<td>0.27</td>
<td>0.47</td>
<td>0.14</td>
</tr>
<tr>
<td>Gamma</td>
<td>0.14</td>
<td>0.02</td>
<td>0.33</td>
<td>0.2</td>
<td>0.3</td>
</tr>
</tbody>
</table>

NOvA Simulation

Efficiency of FHC Network on FHC Dataset

<table>
<thead>
<tr>
<th></th>
<th>Electron</th>
<th>Muon</th>
<th>Proton</th>
<th>Pion</th>
<th>Gamma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electron</td>
<td>0.78</td>
<td>0.05</td>
<td>0.03</td>
<td>0.05</td>
<td>0.1</td>
</tr>
<tr>
<td>Muon</td>
<td>0.05</td>
<td>0.77</td>
<td>0.03</td>
<td>0.11</td>
<td>0.05</td>
</tr>
<tr>
<td>Proton</td>
<td>0.05</td>
<td>0.04</td>
<td>0.37</td>
<td>0.28</td>
<td>0.27</td>
</tr>
<tr>
<td>Pion</td>
<td>0.08</td>
<td>0.13</td>
<td>0.21</td>
<td>0.39</td>
<td>0.19</td>
</tr>
<tr>
<td>Gamma</td>
<td>0.21</td>
<td>0.02</td>
<td>0.17</td>
<td>0.13</td>
<td>0.47</td>
</tr>
</tbody>
</table>

NOvA Simulation

Efficiency of FHC+RHC Network on FHC+RHC Dataset

<table>
<thead>
<tr>
<th></th>
<th>Electron</th>
<th>Muon</th>
<th>Proton</th>
<th>Pion</th>
<th>Gamma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electron</td>
<td>0.8</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
<td>0.1</td>
</tr>
<tr>
<td>Muon</td>
<td>0.01</td>
<td>0.82</td>
<td>0.01</td>
<td>0.13</td>
<td>0.03</td>
</tr>
<tr>
<td>Proton</td>
<td>0.01</td>
<td>0.01</td>
<td>0.7</td>
<td>0.16</td>
<td>0.12</td>
</tr>
<tr>
<td>Pion</td>
<td>0.01</td>
<td>0.14</td>
<td>0.16</td>
<td>0.52</td>
<td>0.18</td>
</tr>
<tr>
<td>Gamma</td>
<td>0.17</td>
<td>0.03</td>
<td>0.08</td>
<td>0.16</td>
<td>0.56</td>
</tr>
</tbody>
</table>
Electron ROC Curve

- **FHC on FHC (AUC = 0.914)**
- **FRHC on FRHC (AUC = 0.919)**
- **FRHCd on FRHC (AUC = 0.937)**
- **RHC on RHC (AUC = 0.937)**
Muon ROC Curve

![Muon ROC Curve](image)

- FHC on FHC (AUC = 0.946)
- FRHC on FRHC (AUC = 0.948)
- FRHCd on FRHC (AUC = 0.971)
- RHC on RHC (AUC = 0.958)
Proton ROC Curve

Proton ROC Curve

- FHC on FHC (AUC = 0.829)
- FRHC on FRHC (AUC = 0.827)
- FRHCd on FRHC (AUC = 0.936)
- RHC on RHC (AUC = 0.833)
Pion ROC Curve

True Positive rate

False Positive Rate

FHC on FHC (AUC = 0.767)
FRHC on FRHC (AUC = 0.782)
FRHCd on FRHC (AUC = 0.856)
RHC on RHC (AUC = 0.826)
Preliminary Conclusions

- Preliminary training networks show that double FHC + RHC network is comparatively better than RHC and FHC networks trained separately.

- **Future**: Will try to find if the network performance can be improved by hyperparameter optimization.

- **Future**: Train with complete dataset and compare the three networks.

- **Future**: Check the effect of purity cuts on network performance.

THANK YOU
Backup Slides
Context enriched Prong CNN Architecture

Convolution
5x5 stride 2
32 filters

Bottleneck Block
16 filters

Average Pooling
2x2

Bottleneck Block
24 filters

Elementwise
Maximum Pooling

Average Pooling
2x2

Bottleneck Block
32 filters

Bottleneck Block
48 filters

Bottleneck Block
64 filters

Average Pooling
2x2

1x1 Convolution
6x expansion

Depth-wise
3x3 Convolution

1x1 Convolution
Bottleneck Filters

Squeeze-Excite
Block

Continued

Bottleneck Block
96 filters

Bottleneck Block
160 filters

Global Average Pooling

Dense Layer
1024 Units

Dense Layer
5 Units

Continued