Towards a new Track PID method

Leigh Whitehead

DUNE FD Sim / Reco Meeting

16/03/20

UNIVERSITY OF CAMBRIDGE

DUNE
DEEP UNDERGROUND NEUTRINO EXPERIMENT

Introduction

- I have been investigating the use of 1D convolutions to perform particle ID on track-like objects
- Currently I am trying to classify particles from neutrino interactions as either muons, pions or protons.
- Use a sample of roughly 100,000 tracks from MCC11 files in the DUNE FD workspace geometry
  - I use 80k for training and 20k for testing
  - Those from the testing sample are used later for the distributions

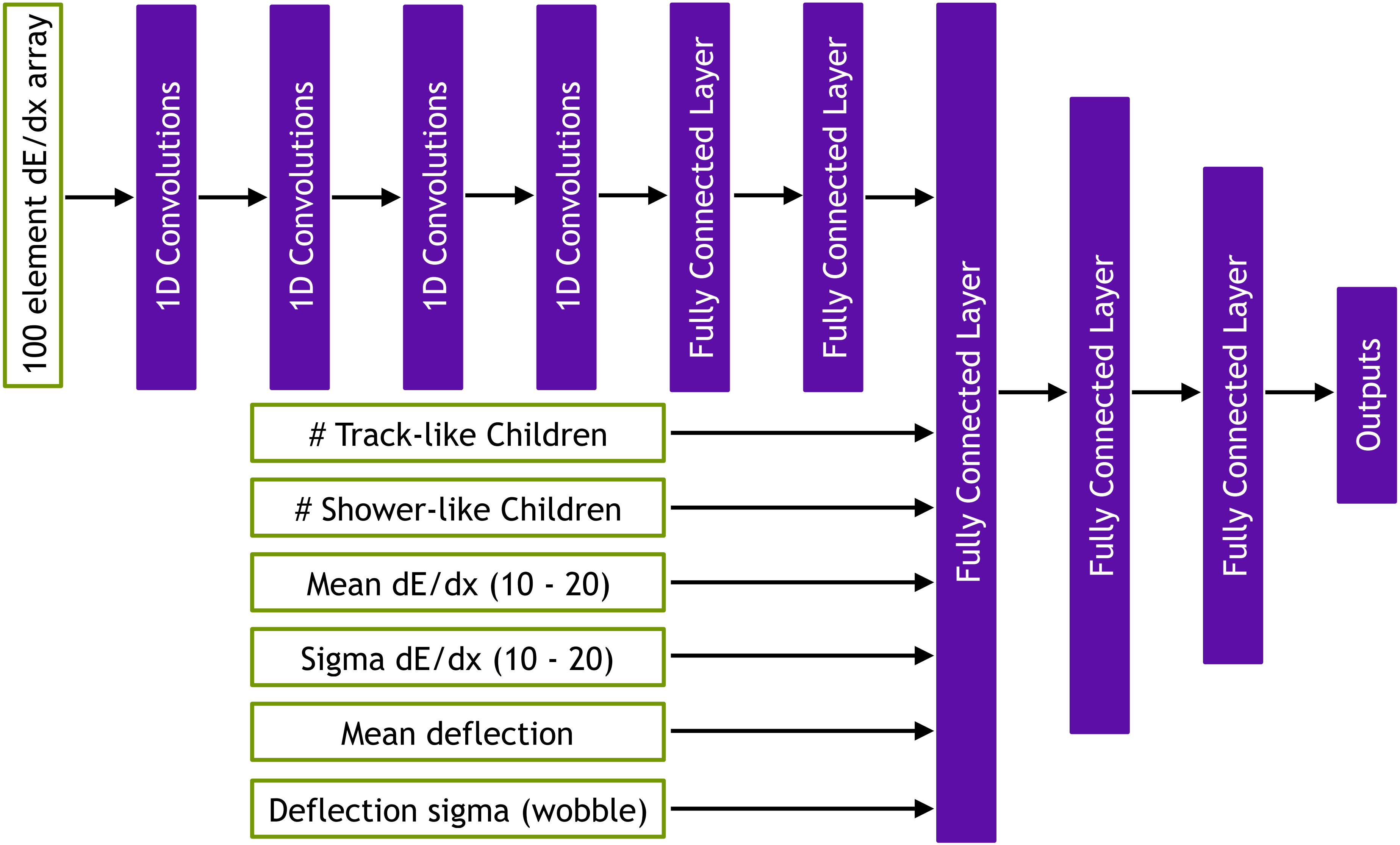
Introduction

- I take the final 100 points of the track as the 1D array

- If a track has 50 < points < 100 then I pad the start of the track as follows:
  - Calculate the mean and sigma of dE/dx for the central third of the points
  - Pad the array to length 100 using random values from a Gauss(mean,sigma) distribution

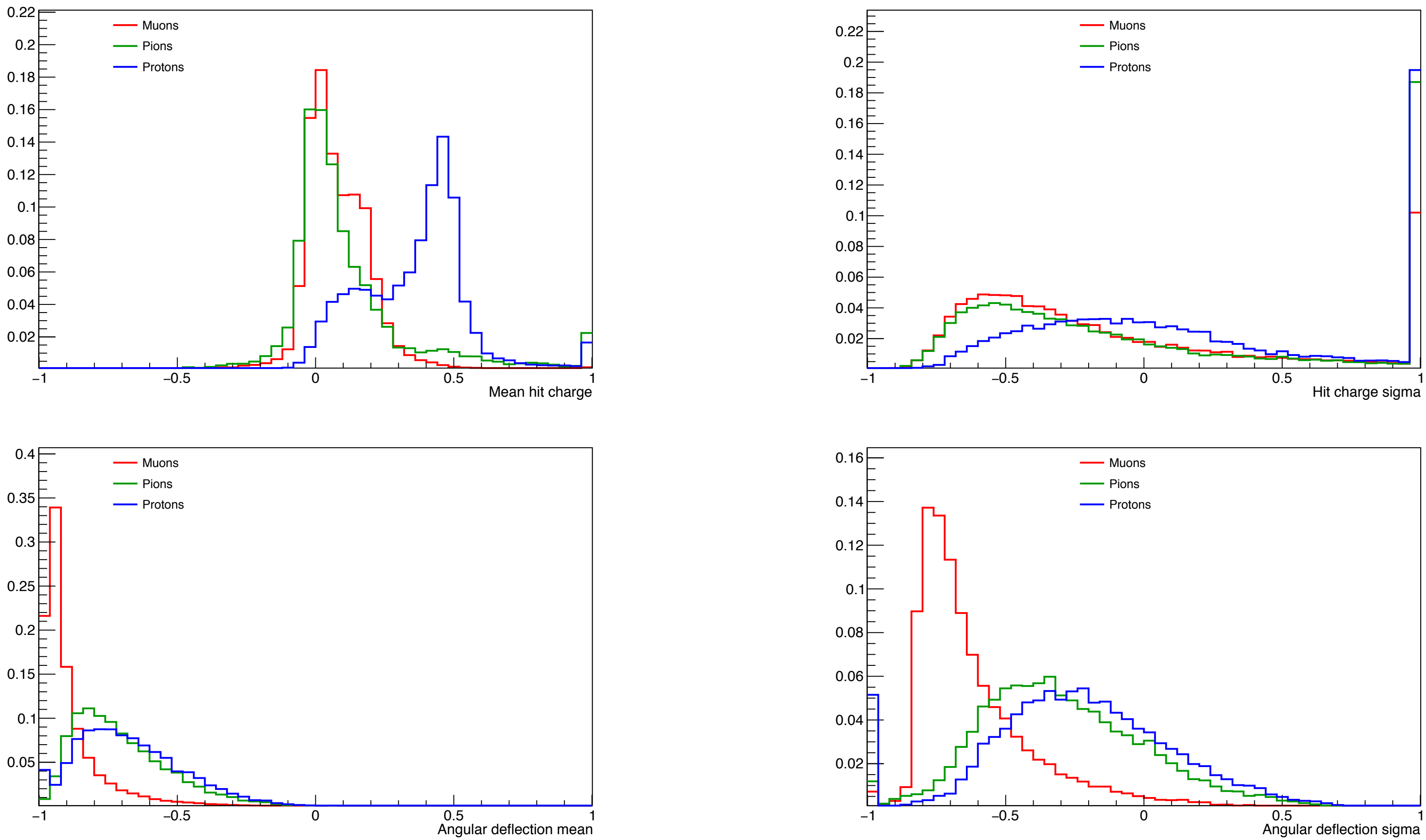
- Also store some additional variables
  - Mean and sigma of the dE/dx as calculated above
  - Mean and sigma of the distribution of angles between successive track points (how wobbly it is)
  - Number of track-like and shower-like child particles
  - Number of grand-children

Architecture



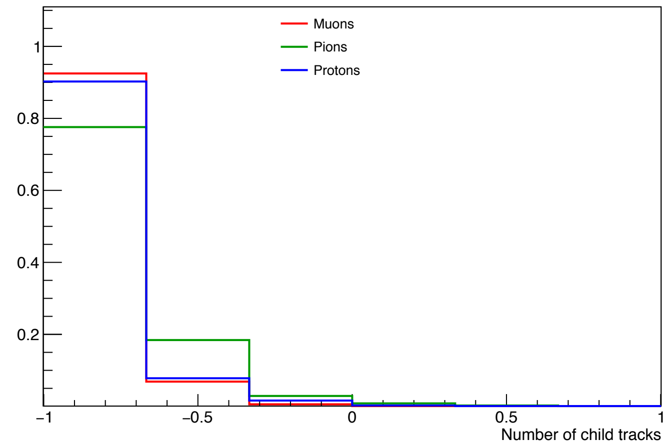
Input variable distributions

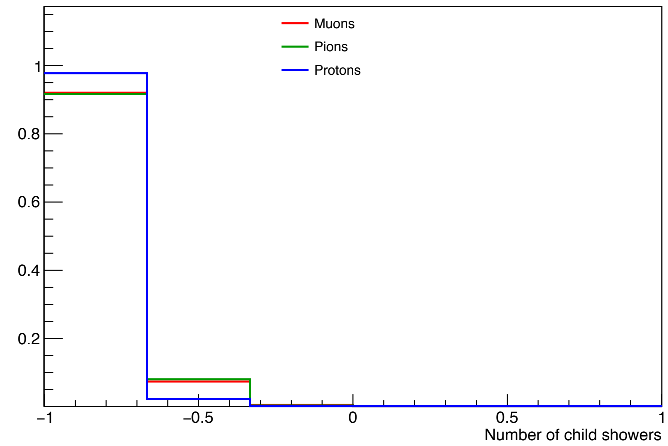
- Area normalised input distributions with normalised x-axes

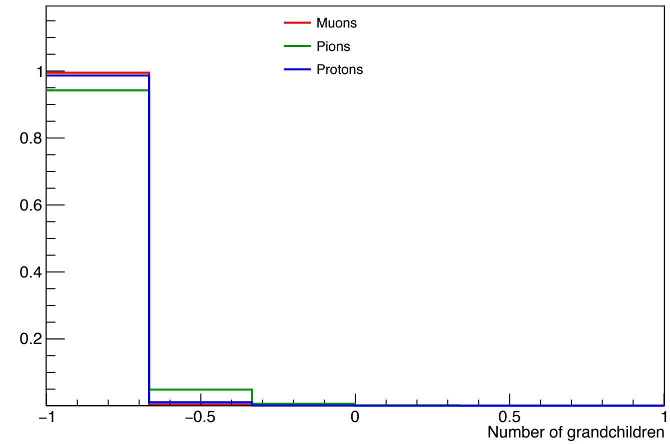


Input variable distributions

- Area normalised input distributions with normalised x-axes







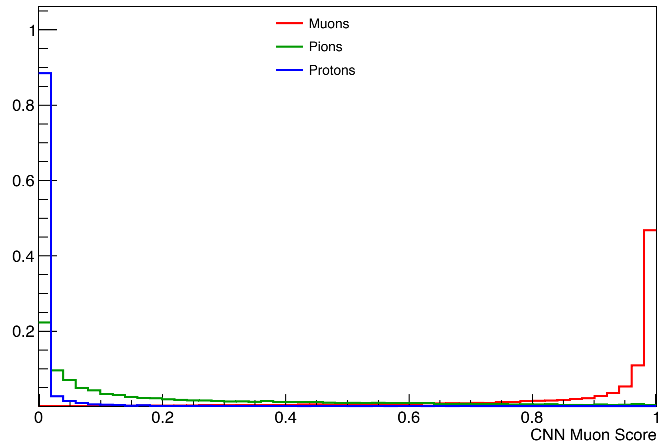
Results

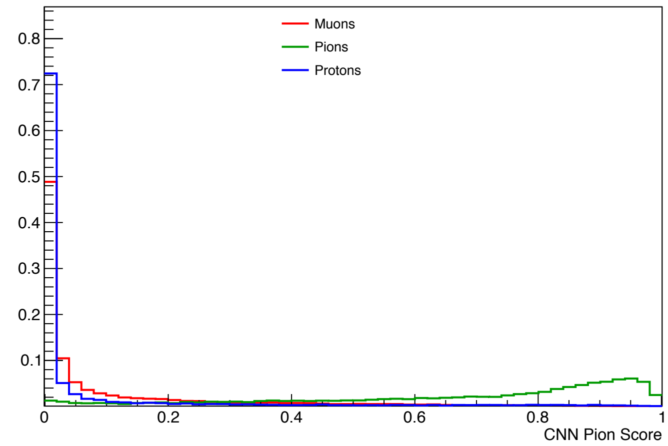
- I trained the network for 20 epochs
  - This only takes about 15 minutes on my laptop CPU
  - 1D convolutions aren't very CPU intensive compared to other approaches such as RNNs (typically used for signal processing tasks for speech recognition)

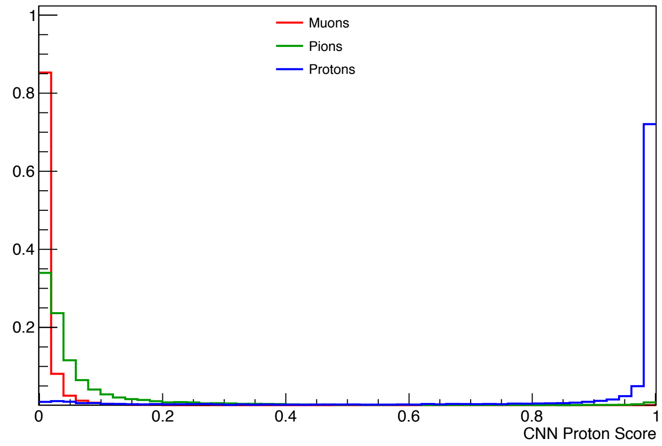
- The next slides show the output scores for each of the three categories (muon, pion and proton)

Output distributions

- The CNN scores for different particle types



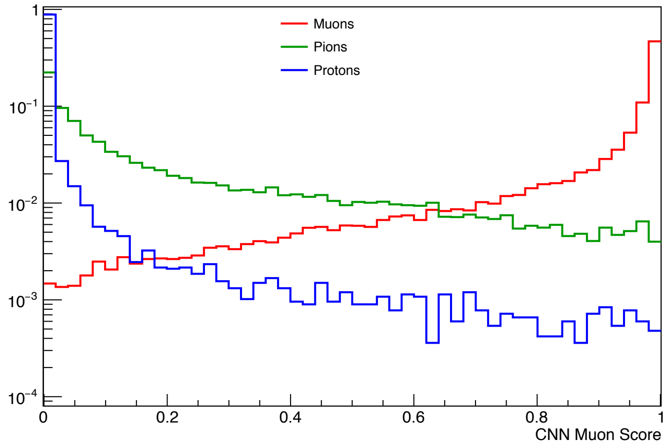


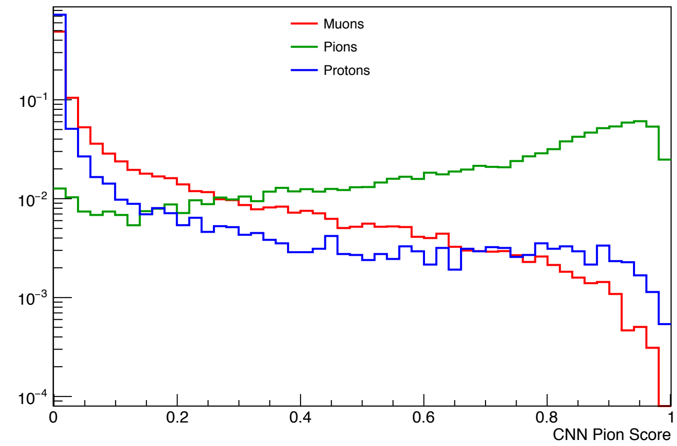


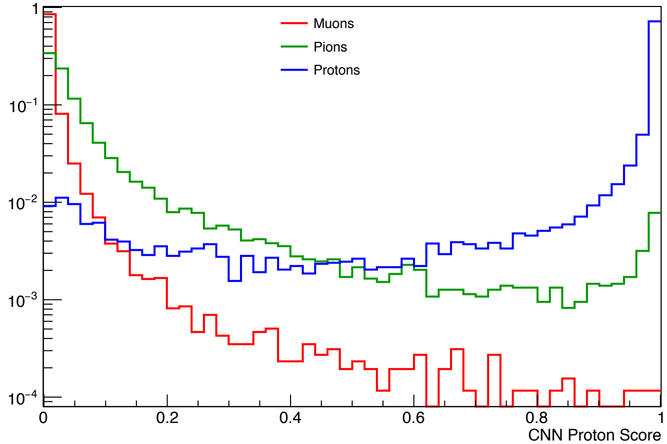
Muon and proton scores well peaked at 1 (signal) and 0 (background)

Output distributions

- The CNN scores for different particle types







Muon and proton scores well peaked at 1 (signal) and 0 (background)

Simple selections

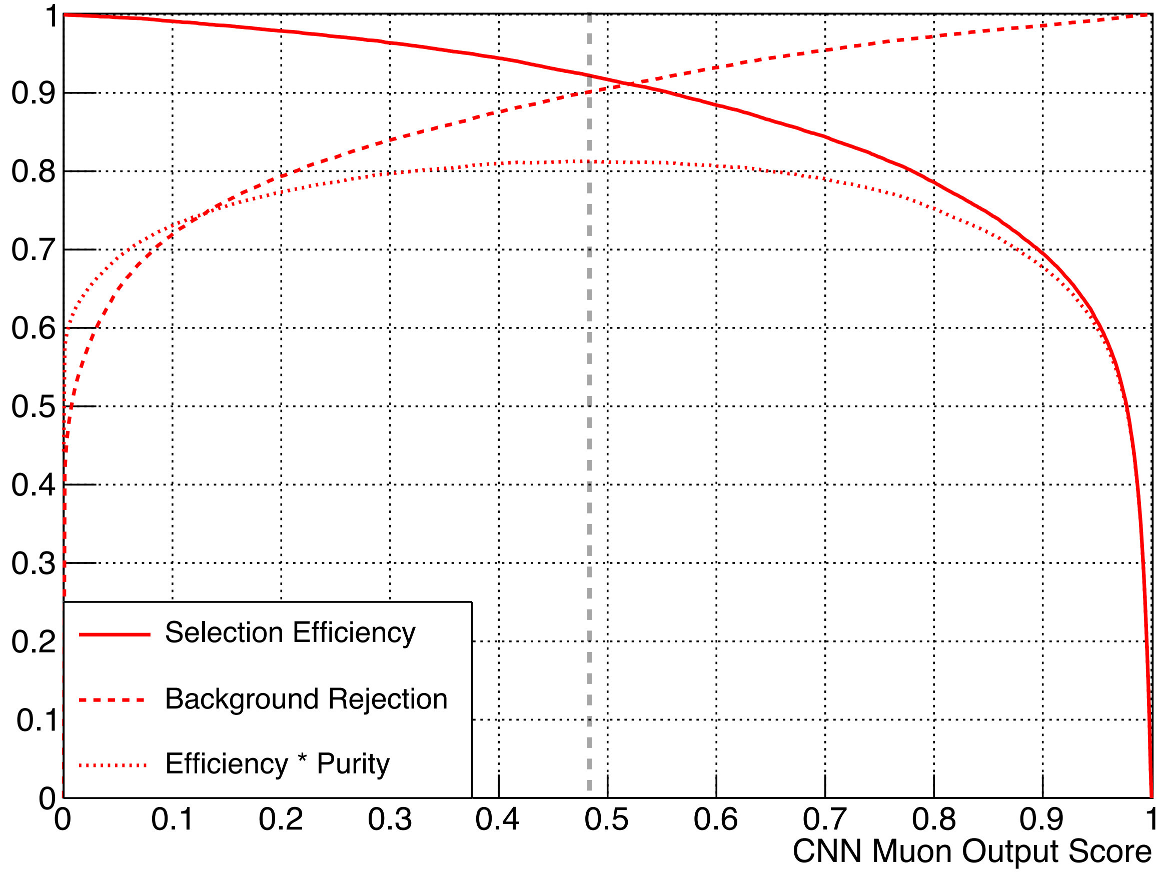
- Try to make selections of muons / pions / protons

- Simplest approach:
  - Optimise the three selections individually based on the muon, pion and proton scores from the network
  - This doesn’t guarantee mutual exclusivity

- Vary the cut value and plot the efficiency, background rejection and FOM (efficiency * purity in this case)

Muon Selection

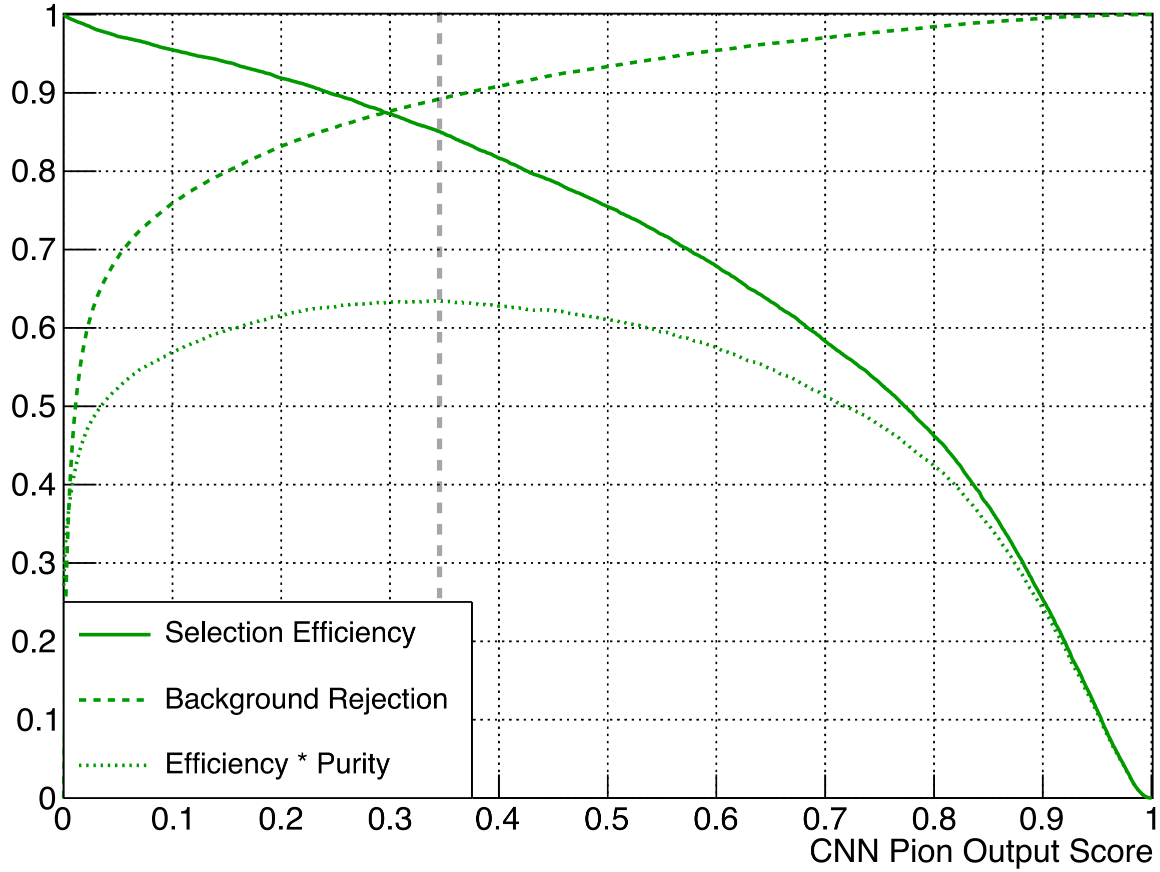
- Distributions for the muon selection



Optimal cut at 0.48

Pion Selection

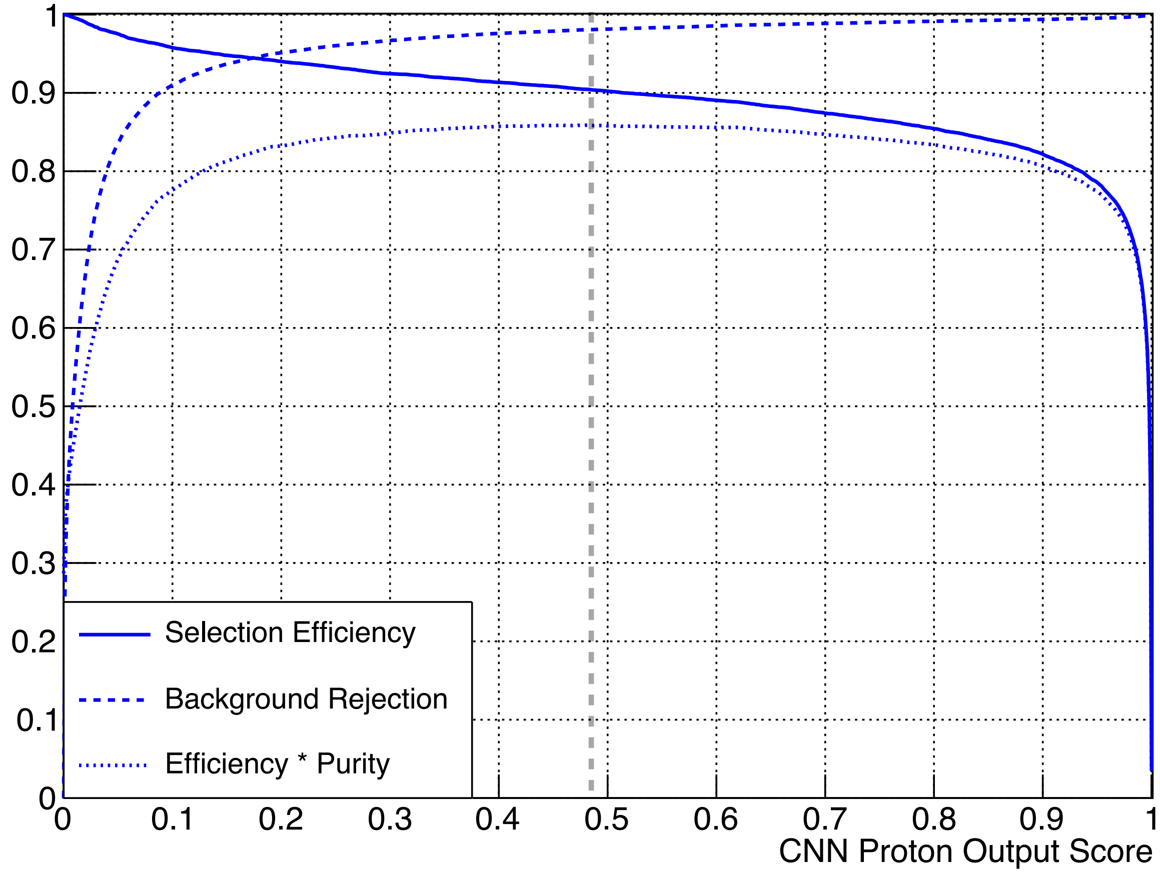
- Distributions for the pion selection



Optimal cut at 0.35

Proton Selection

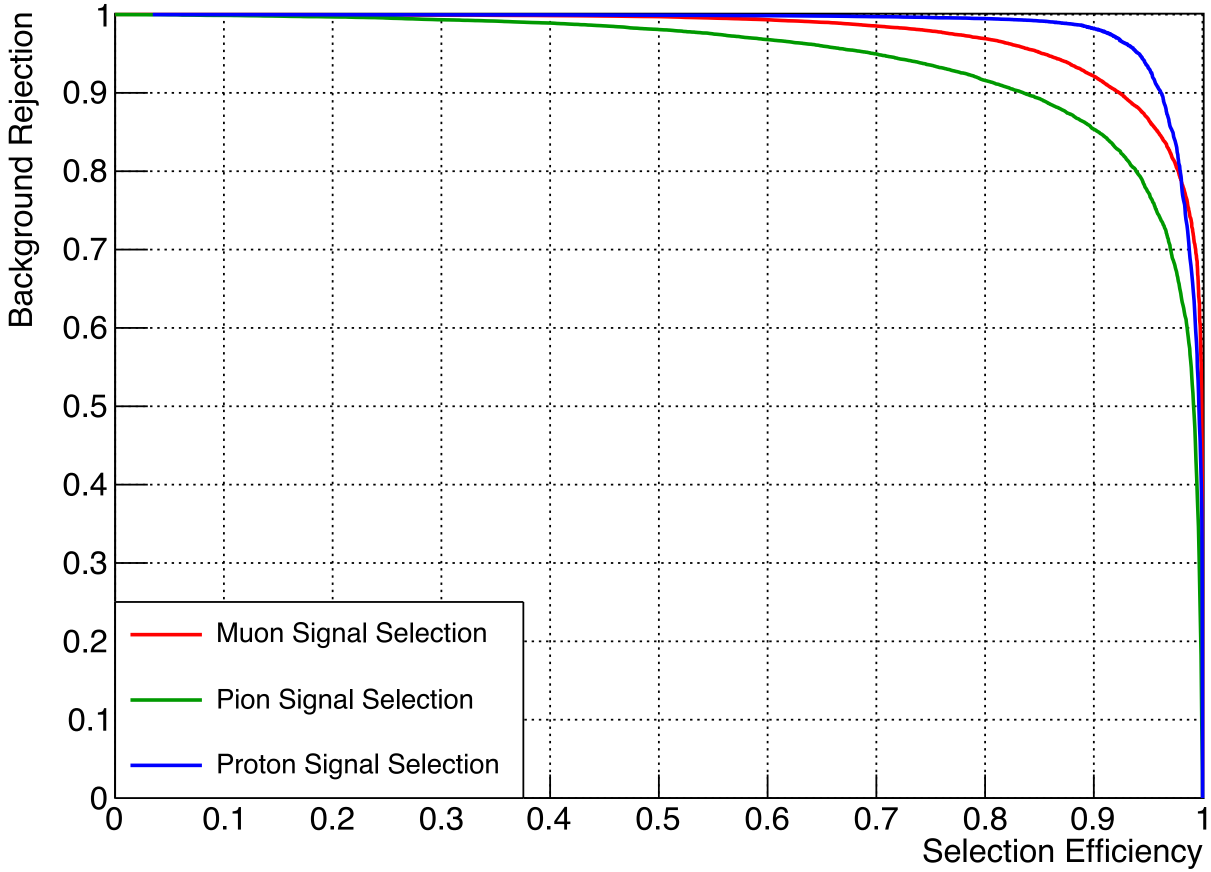
- Distributions for the proton selection



Optimal cut at 0.48

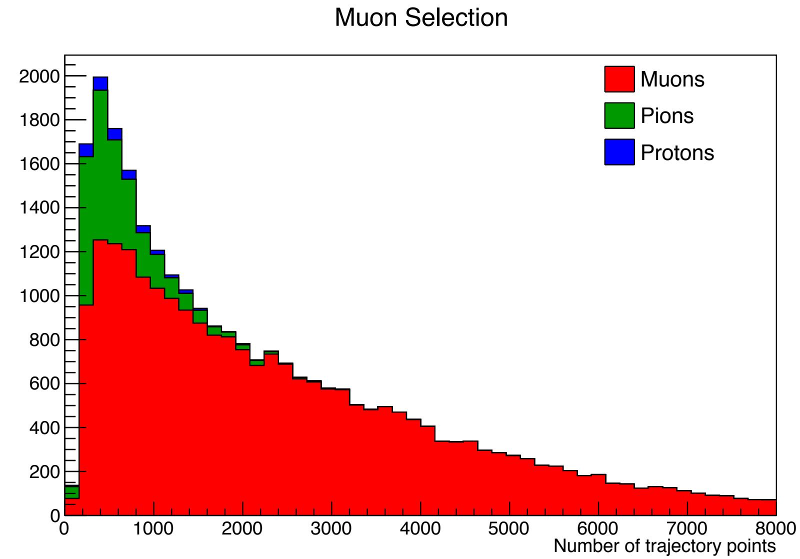
Efficiency vs Rejection

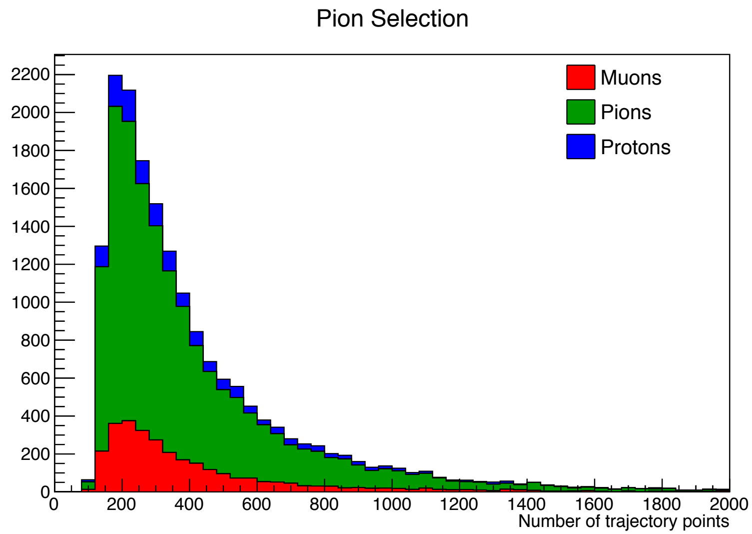
- Shown for all three selections

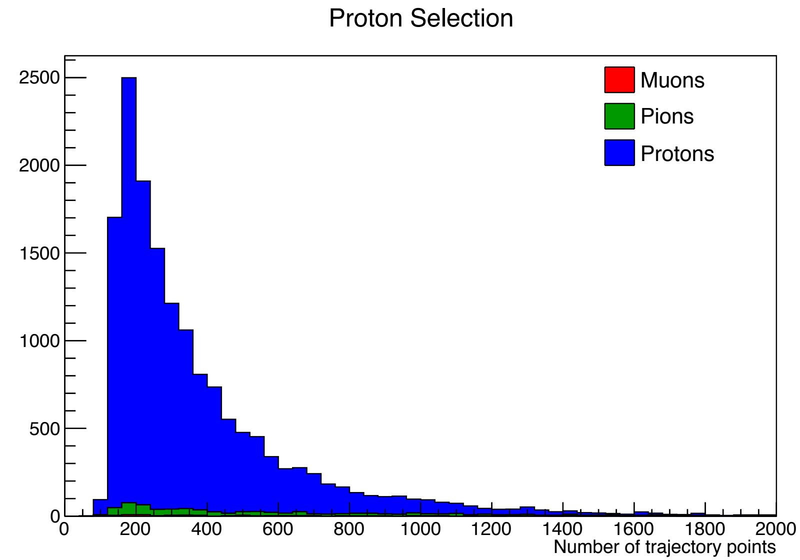


Event selections

- Number of hits for each selection using the optimal cut values listed previously
  - Some tracks could be counted twice (could pass two cuts)

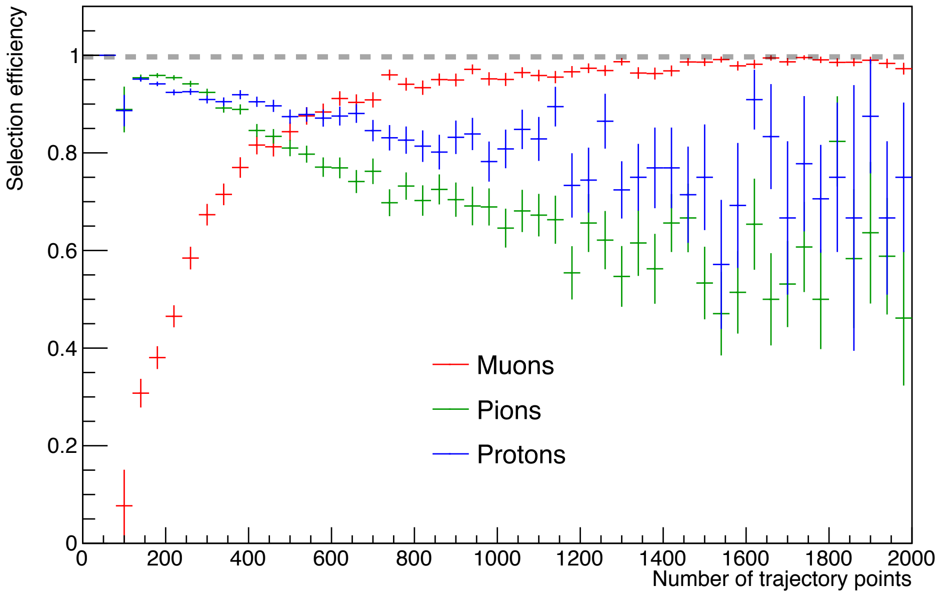


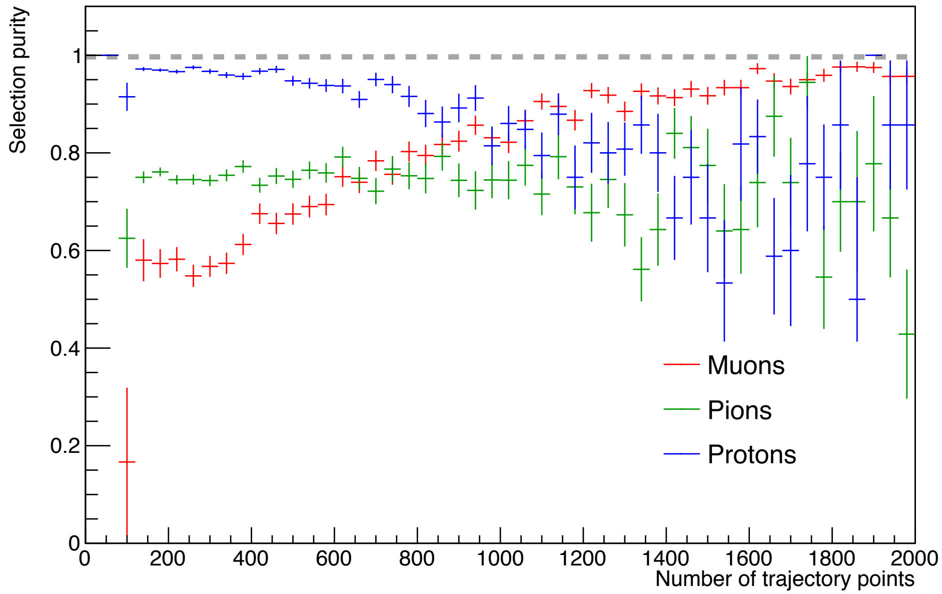




Event selections

- Efficiency and purity of the event selections
  - Muon selection is bad at low numbers of hits… mostly due to the sample make-up
  - This was worse before I removed the number of hits from the training variables, though





Results

- Optimum cut values come out to give the following:

	Selected Signal	Selected Background	Efficiency	Purity
Muon Score > 0.48	23769 (25754)	3234 (32478)	92.3%	88.0%
Pion Score > 0.35	13383 (15788)	4514 (42444)	84.8%	74.8%
Proton Score > 0.48	15098 (16690)	808 (41542)	90.5%	94.9%

- Again, note that since the cut values are less than 0.5, these samples are not guaranteed to be mutually exclusive

Summary

- I have developed a track PID method using 1D convolutions to extract features from the dE/dx distribution

- Looks to be working well
  - See high purity and efficiency for selecting muons and protons
  - Some muon / pion confusion for short tracks

- Need to work on the interface to LArSoft now to get it integrated and tested