Vertical Drift Studies

Nitish Nayak
14th Feb, 2022

Recap on Results

Current

```
INFO:root:flavour report:

              precision    recall  f1-score   support

     CC Numu       0.86      0.84      0.85     25609
      CC Nue       0.79      0.83      0.81     23491
    CC Nutau       0.48      0.27      0.34      5741
          NC       0.80      0.85      0.83     33863

    accuracy                           0.80     88704
   macro avg       0.73      0.70      0.71     88704
weighted avg       0.79      0.80      0.80     88704

INFO:root:flavour confusion matrix (rows = predicted classes, cols = actual classes):

[[21509   587  1176  1875]
 [  823 19435  1714  2486]
 [  299   698  1548   706]
 [ 2978  2771  1303 28796]]
```

Horizontal Drift

Reference

DUNE CVN (views 0, 1, and 2)

	precision	Recall	F1 score	#events
CC ν_μ	**0.93**	**0.96**	**0.95**	**26108**
CC ν_e	**0.93**	**0.97**	**0.95**	**25665**
CC ν_τ	0.66	0.37	0.47	5813
NC	0.94	0.95	0.94	42382

- Based on results given at collaboration meeting few weeks back, we were seeing a worrying drop in performance for the current VD training vs original CVN HD training.
- A number of suggestions were made about possible avenues to hunt it down

Fiducial Volume Cut

- Turns out, a change had been introduced in the CVN code a couple of years back where the requirement of neutrin vertex within the fiducial volume was removed for some reason.
- The original training relied on this to get a set of clean events, while our training didn't have this cut applied
- So reprocessed the datasets with this cut added back in and redid the training.
- Also processed the horizontal drift files in the original 1D simulation, similar to what was used in the training before
  - MCC11
  - I am training on similar amount of statistics as what I have for VD currently
- VD training/validation/test samples : ~550k, ~70k, ~70k
- My HD training/validation/test samples : ~530k, 66k, 66k
- Original training had ~3.2 million events so almost 5x the current statistics

Fiducial Volume Cut

- HD takes a fiducial volume as ~70 cm from the boundary walls
  - bool isFid = (fabs(vtx.X())<310 && fabs(vtx.Y())<550 && vtx.Z()>50 && vtx.Z()<1244)
- We want to do something similar for VD so the equivalent cut becomes
  - bool isFid = (fabs(vtx.X())<300 && fabs(vtx.Y())<680 && vtx.Z()>40 && vtx.Z()<850)
- This cut is also used in the energy estimator tuning, so is consistent with what Wenjie is doing as well
- Cut applied on true neutrino vertex

New Results

Before Fid Vol

```
INFO:root:flavour report:

              precision    recall  f1-score   support

     CC Numu       0.86      0.84      0.85     25609
      CC Nue       0.79      0.83      0.81     23491
    CC Nutau       0.48      0.27      0.34      5741
          NC       0.80      0.85      0.83     33863

    accuracy                           0.80     88704
   macro avg       0.73      0.70      0.71     88704
weighted avg       0.79      0.80      0.80     88704

INFO:root:flavour confusion matrix (rows = predicted classes, cols = actual classes):

[[21509   587  1176  1875]
 [  823 19435  1714  2486]
 [  299   698  1548   706]
 [ 2978  2771  1303 28796]]
```

After Fid Vol

```
              precision    recall  f1-score   support

     CC Numu       0.87      0.90      0.89     18715
      CC Nue       0.86      0.88      0.87     18625
    CC Nutau       0.50      0.32      0.39      4506
          NC       0.87      0.90      0.88     27786

    accuracy                           0.85     69632
   macro avg       0.78      0.75      0.76     69632
weighted avg       0.85      0.85      0.85     69632

INFO:root:flavour confusion matrix (rows = predicted classes, cols = actual classes):

[[16874   312   942  1227]
 [  274 16332  1121  1158]
 [  246   650  1441   532]
 [ 1321  1331  1002 24869]]
```

Horizontal Drift

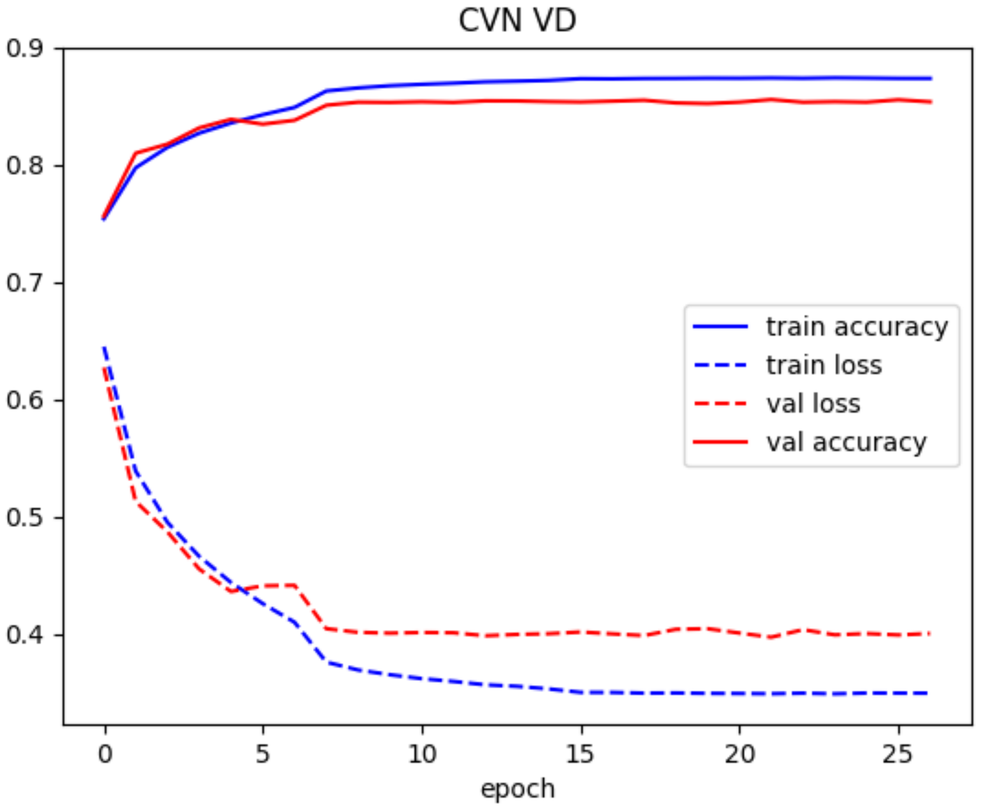
DUNE CVN (views 0, 1, and 2)

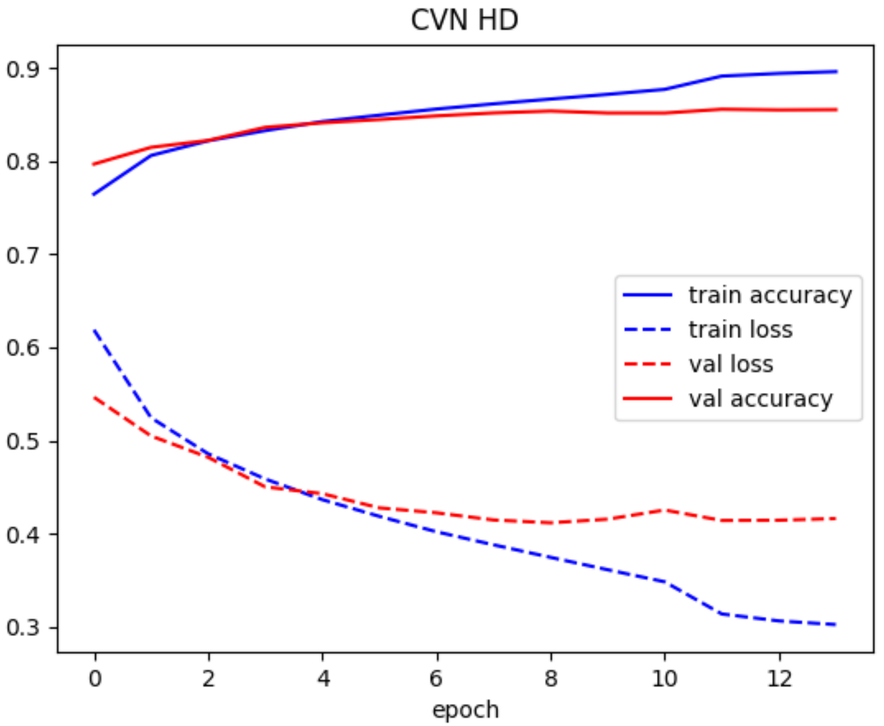
Reference

	precision	Recall	F1 score	#events
CC ν_μ	**0.93**	**0.96**	**0.95**	**26108**
CC ν_e	**0.93**	**0.97**	**0.95**	**25665**
CC ν_τ	0.66	0.37	0.47	5813
NC	0.94	0.95	0.94	42382

- Significant improvement in raw network performance!
- So as we'd thought, this was a pretty big contributor to the big drop in performance
- We're about halfway towards the original HD training result, which is encouraging

What about the rest of the way?

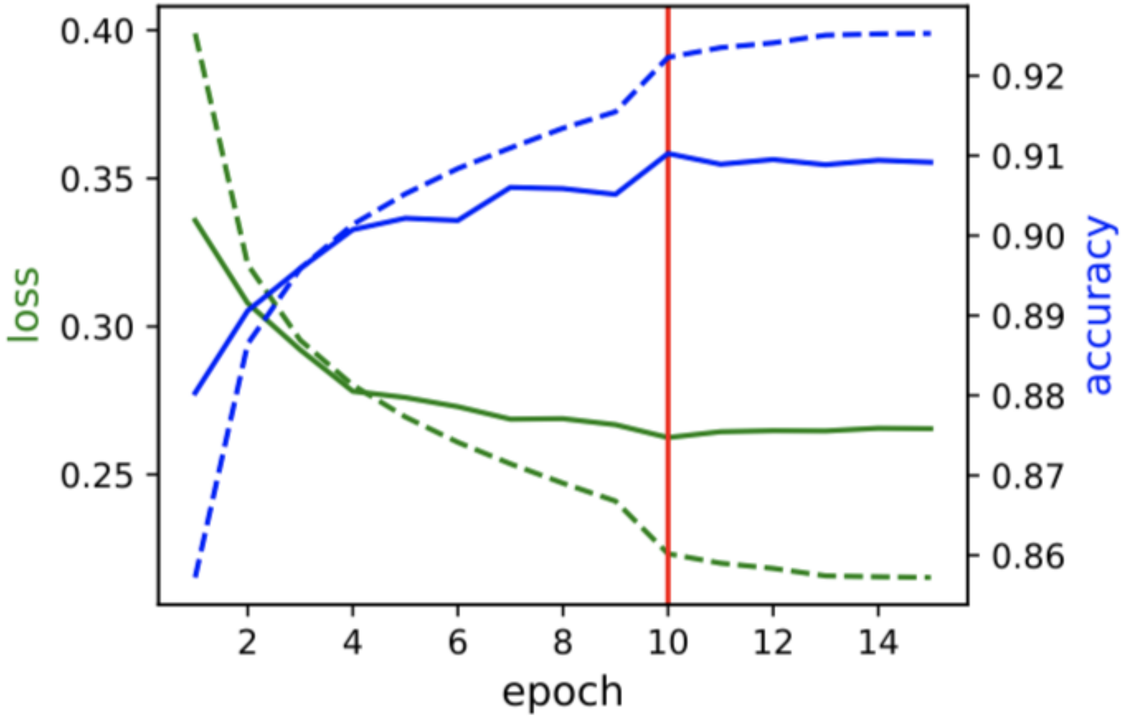




- CVN HD training w/ similar statistics as the VD training
  - Shows that the VD and HD training w/ similar statistics converge to ~ same overall accuracy in the validation sample (~0.86)
- Also implemented Leigh and Saul's suggestions on reducing the learning rate by a factor of 10 after the results are similar for a few epochs

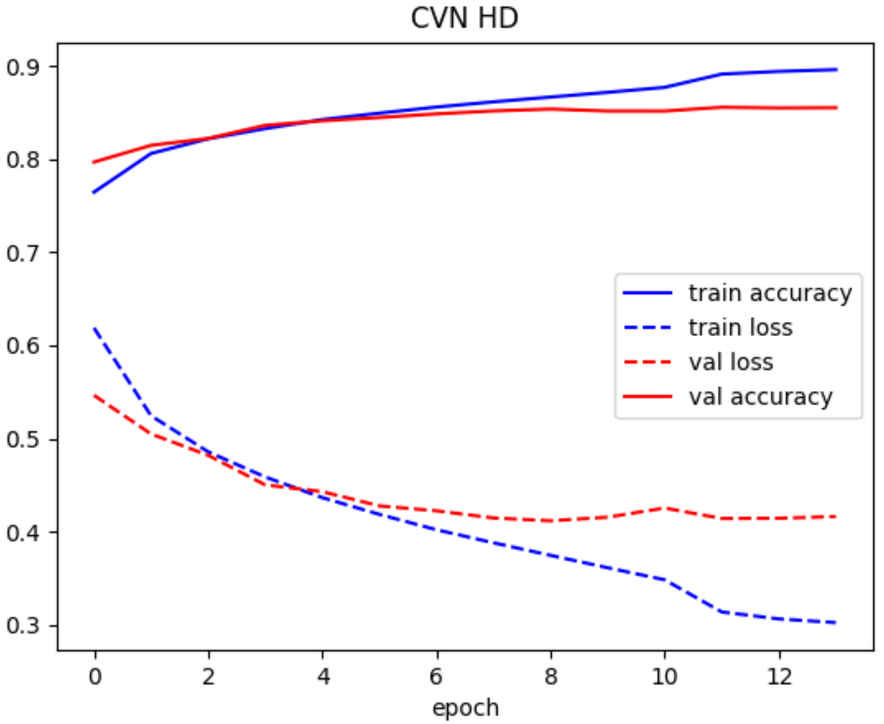
What about the rest of the way?

original CVN HD training



(a) Flavor.

my CVN training



- Overall training looks pretty similar in terms of convergence, but the final value it converges to are a bit different, 0.91 vs 0.86 in overall validation accuracy

- Seems like more statistics might bridge this gap further

CVN VD

```
              precision    recall  f1-score   support

    CC Numu       0.87      0.90      0.89     18715
     CC Nue       0.86      0.88      0.87     18625
    CC Nutau      0.50      0.32      0.39      4506
          NC      0.87      0.90      0.88     27786

    accuracy                          0.85     69632
   macro avg      0.78      0.75      0.76     69632
weighted avg      0.85      0.85      0.85     69632

INFO:root:flavour confusion matrix (rows = predicted classes, cols = actual classes):

[[16874   312   942  1227]
 [  274 16332  1121  1158]
 [  246   650  1441   532]
 [ 1321  1331  1002 24869]]
```

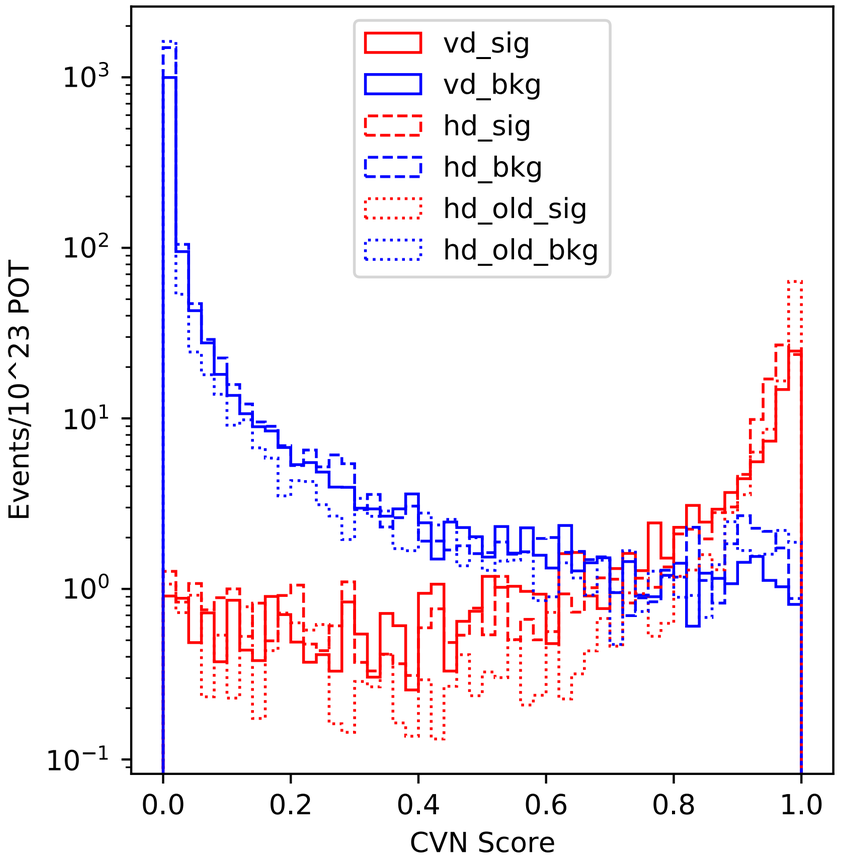
CVN HD

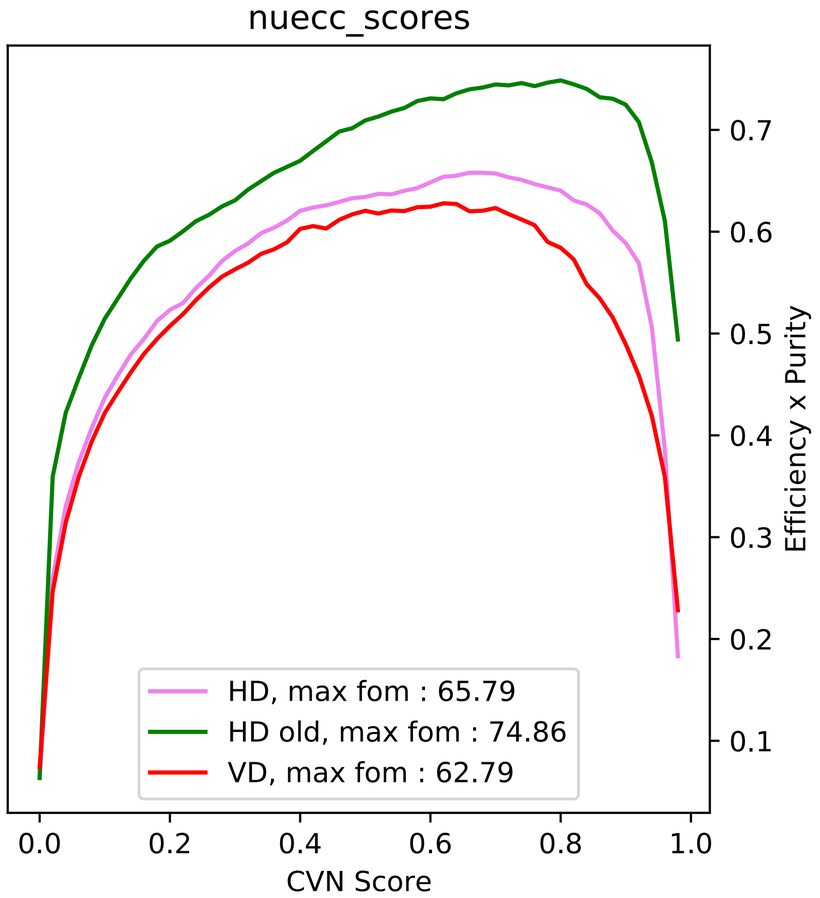
```
              precision    recall  f1-score   support

    CC Numu       0.87      0.92      0.90     18541
     CC Nue       0.88      0.88      0.88     18060
    CC Nutau      0.46      0.27      0.34      4503
          NC      0.86      0.89      0.87     25456

    accuracy                          0.85     66560
   macro avg      0.77      0.74      0.75     66560
weighted avg      0.84      0.85      0.85     66560

INFO:root:flavour confusion matrix (rows = predicted classes, cols = actual classes):

[[17089   306  1077  1063]
 [  157 15816   964  1030]
 [  294   462  1229   679]
 [ 1001  1476  1233 22684]]
```

- CVN HD is my training on MCC11 files w/ similar statistics as VD
- CVN HD Old is applying the old trained model in the CVN paper from Leigh and Saul on the MCC11 HD dataset
  - Can easily imagine ~2-3 points in precision/recall coming from more training statistics

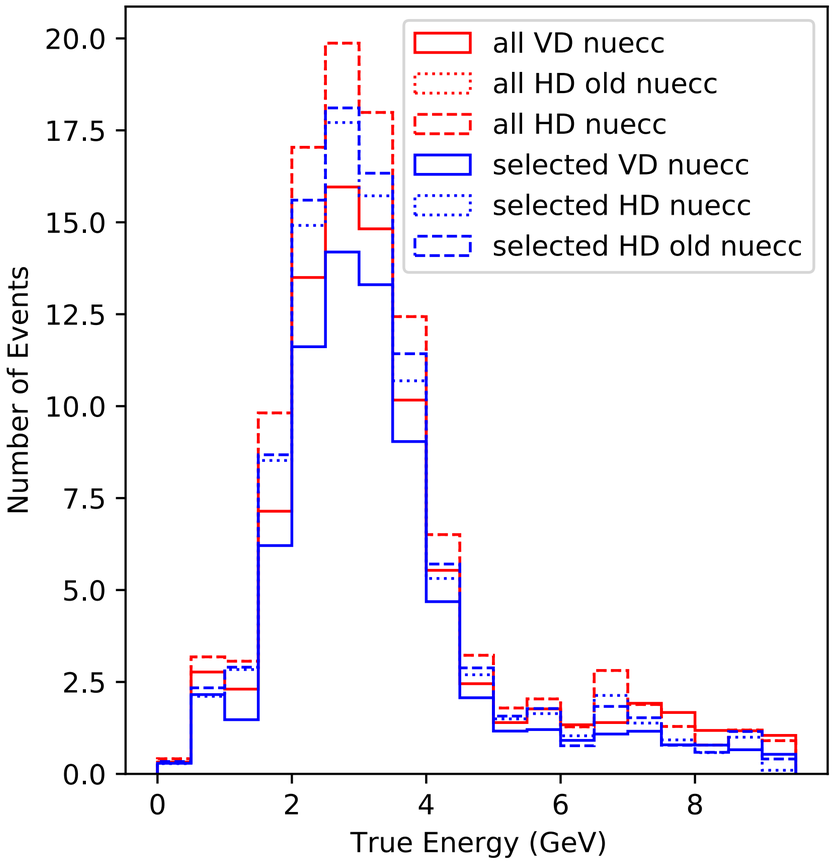
CVN HD Old

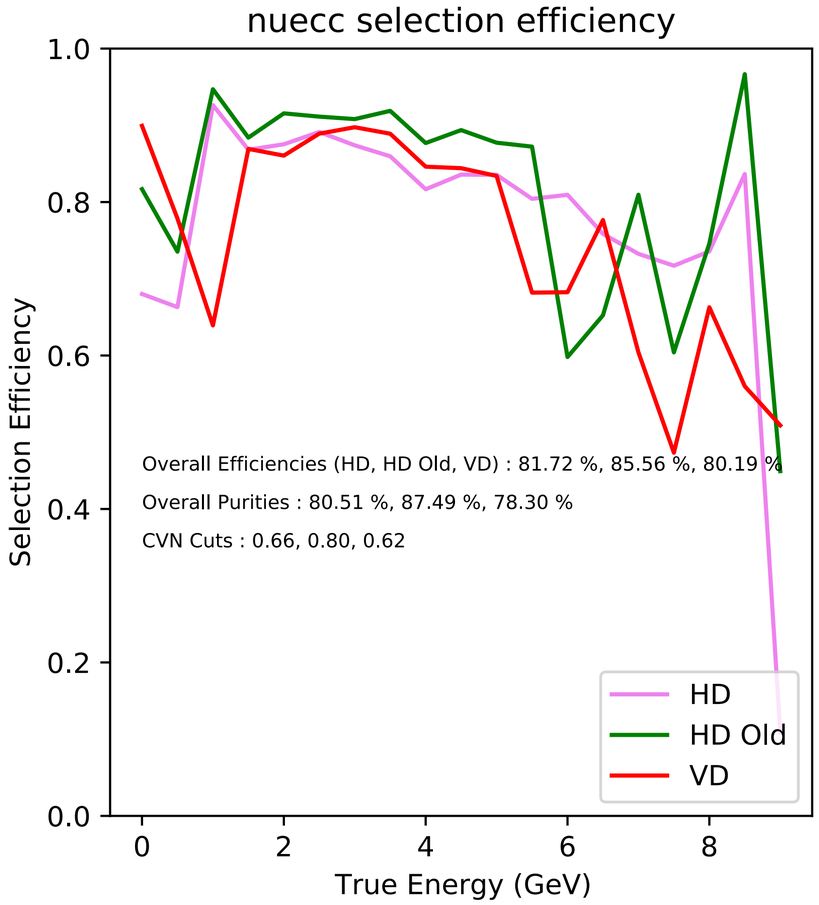
```
              precision    recall  f1-score   support

    CC Numu       0.91      0.93      0.92     18541
     CC Nue       0.90      0.92      0.91     18060
    CC Nutau      0.54      0.33      0.41      4503
          NC      0.88      0.91      0.90     25456

    accuracy                          0.88     66560
   macro avg      0.81      0.77      0.78     66560
weighted avg      0.87      0.88      0.88     66560

INFO:root:flavour confusion matrix (rows = predicted classes, cols = actual classes):

[[17291   151   858   719]
 [  181 16667   958   700]
 [  220   272  1477   765]
 [  849   970  1210 23272]]
```





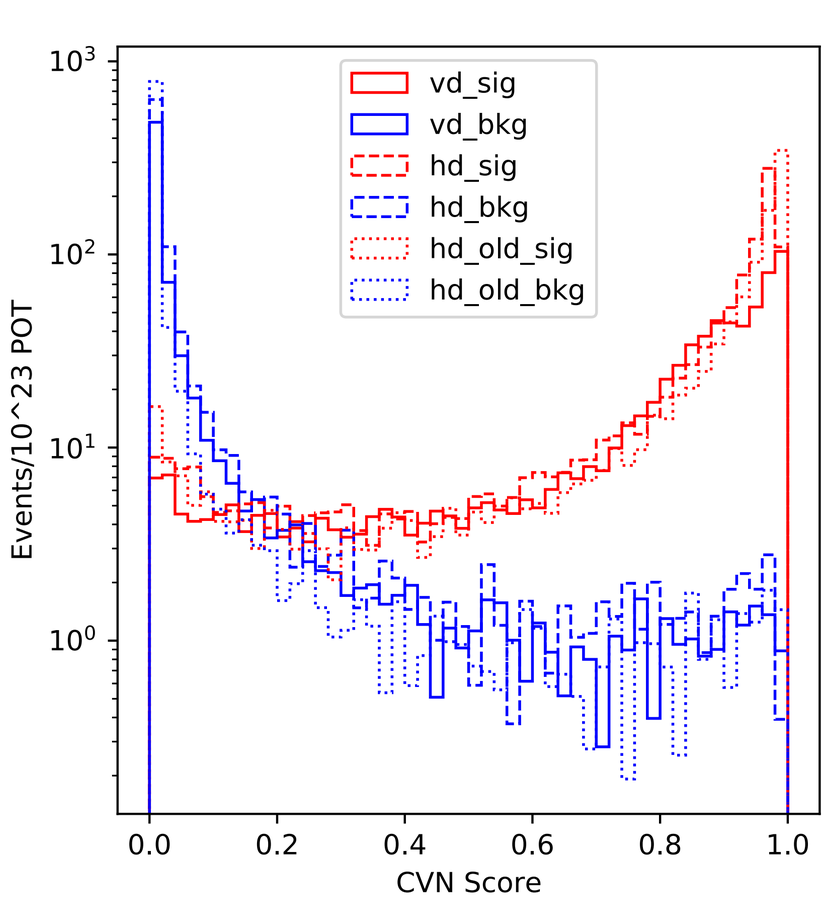
- Comparing the 3, it seems VD and HD are close — whereas “HD Old” outperforms them both
- POT Normalization approximately, ran over 2000 events for each swap (numuCC, nueCC, nutauCC) and extrapolated total POT from there — nuTauCC background approximately halved from before

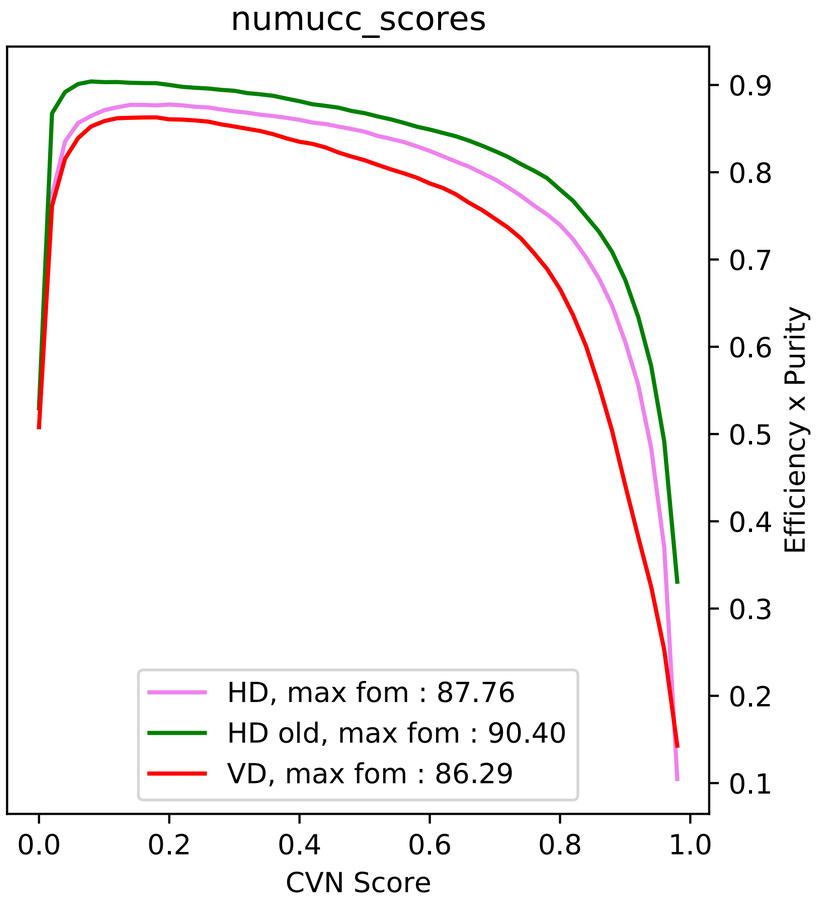




- For efficiency and purity numbers, after approximate POT normalisation, I get ~80% efficiency and purity for both VD and HD

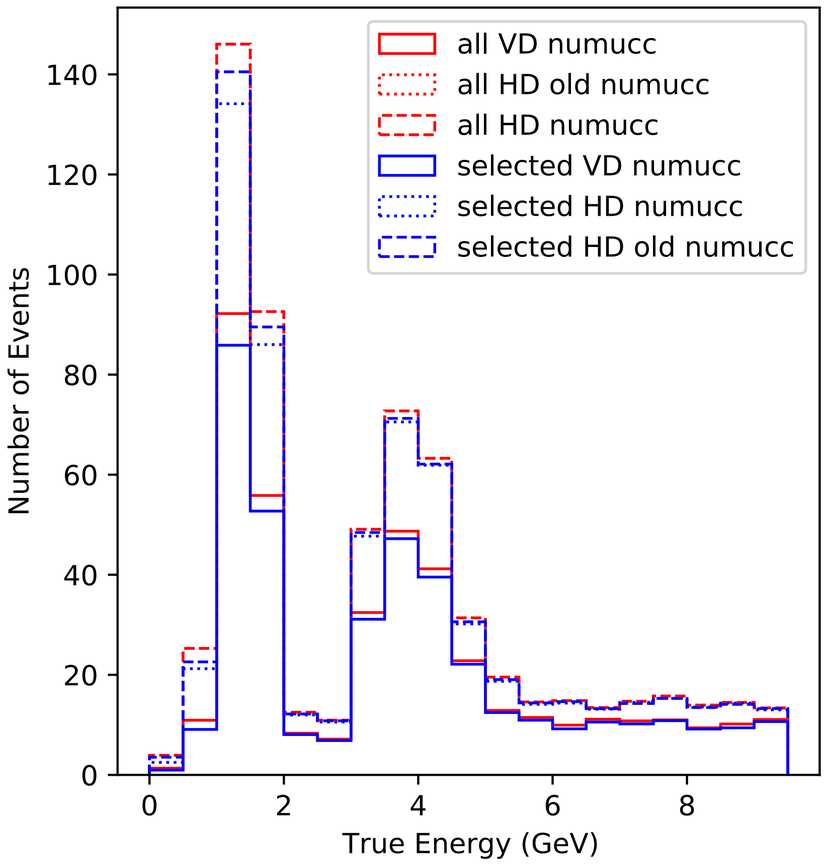
- For HD old, its 86% and 88% (eff, purity) respectively — which is pretty close to what they report in the paper (NB : they report it wrt reco energy since at nueCC peak, high energy NCs dominate background, not 2-4 GeV NCs like here)

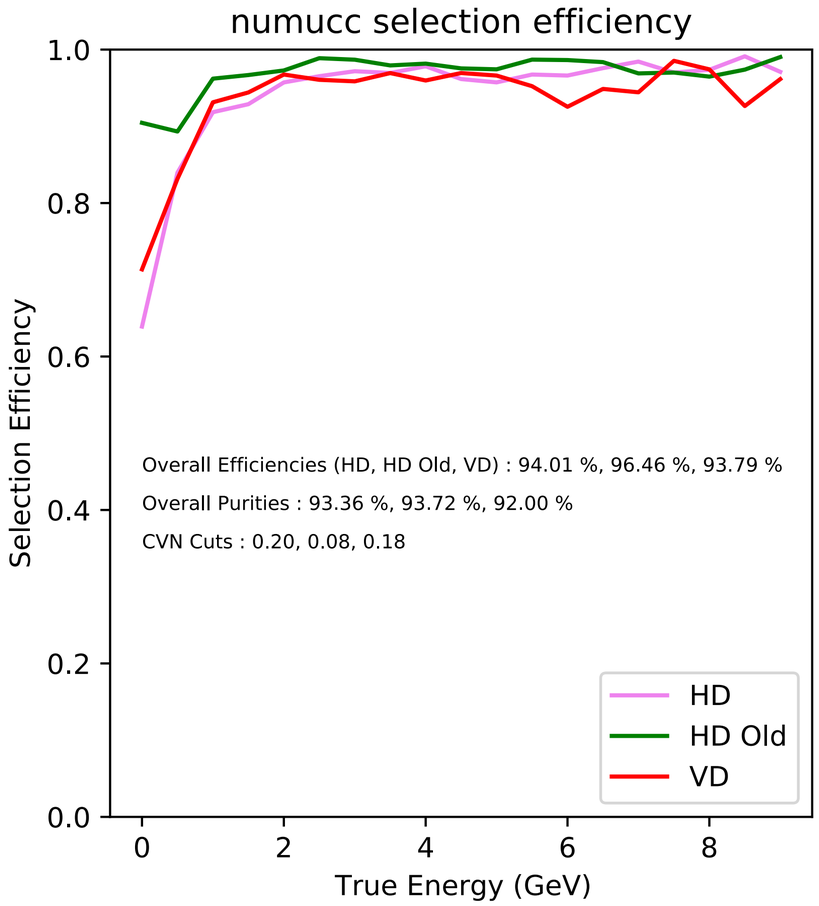




- Comparing the 3, it seems VD and HD are close — whereas “HD Old” outperforms them both

- POT Normalization approximately, ran over 2000 events for each swap (numuCC, nueCC, nutauCC) and extrapolated total POT from there — nuTauCC background approximately halved from before





- For efficiency and purity numbers, after approximate POT normalisation, I get ~94% efficiency and 92-93% purity for both VD and HD

- For HD old, its 96% and 93% (eff, purity) respectively — which is pretty close to what they report in the paper

Summary

- Seems like we're close to the original CVN performance :
  - Running old model on my processed dataset gives similar results to paper
  - Retraining just on my smaller statistics gives similar results to VD training
  - Remaining gap hopefully just a matter of enlarging the training sample to similar levels as before
- If it turns out that VD is slightly worse off than before, I don't think that's a big problem? A lot has changed since then and not sure we expect similar levels of performance anyway

Next Steps

- Want to retrain CVN on full VD reco dataset to see the improvement
  - Planning to do it on Wilson Cluster
  - Pre-processing step on FermiGrid seems a bit inefficient as it copies the gz/info output files for every event from the grid nodes? Is there a better way to do this? This might be a problem on the full dataset given the huge number of events
- Also try out trainings with wirecell/simchannels pixel maps to see the effect of signal processing and the new 2D simulation
- Not a big difference in accuracy b/w training and validation sets. Is this acceptable? Since we might be forced to train and study sensitivities on the same sample?

Backup

Abstract

The Deep Underground Neutrino Experiment is a next-generation neutrino oscillation experiment that aims to measure CP-violation in the neutrino sector as part of a wider physics program. A deep learning approach based on a convolutional neural network has been developed to provide highly efficient and pure selections of electron neutrino and muon neutrino charged-current interactions. The electron neutrino (antineutrino) selection efficiency peaks at 90% (94%) and exceeds 85% (90%) for reconstructed neutrino energies between 2-5 GeV. The muon neutrino (antineutrino) event selection is found to have a maximum efficiency of 96% (97%) and exceeds 90% (95%) efficiency for reconstructed neutrino energies above 2 GeV. When considering all electron neutrino and antineutrino interactions as signal, a selection purity of 90% is achieved. These event selections are critical to maximize the sensitivity of the experiment to CP-violating effects.