Neural network for He event reconstruction quality

AMS-Italy 30.11.2023

Francesco Rossi, Prof. P. Zuccon

Work strategy

Main goal:

Develop a fully connected neural network capable to identify well reconstructed events with charge 2.

How?

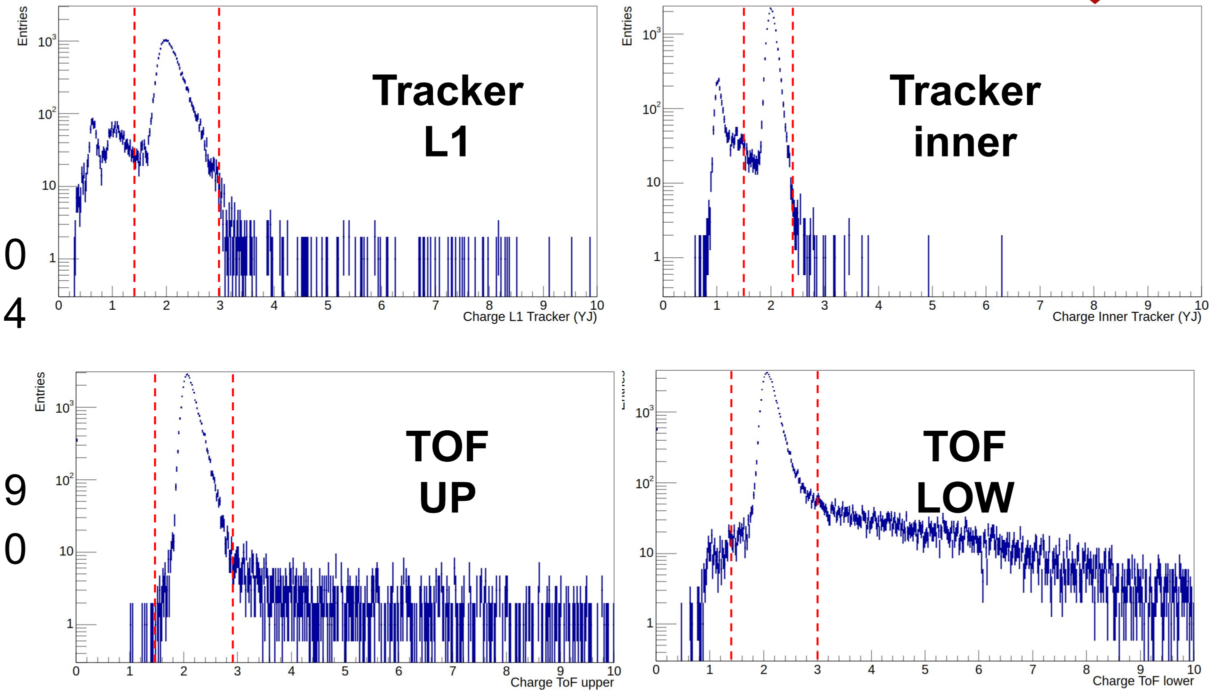
1. Obtain a charge 2 sample using ^{4}He Monte Carlo.
2. Define two event categories, the good (signal) and bad (background) reconstructed.
3. Identify useful variables to distinguish between those categories.
4. Create a new dataset containing only the only relevant variables (features).
5. Build, train and evaluate the performance of a classifier.

Monte Carlo samples and data pre-selection

L1-focused MC, **he4.pl1.l1.24000.6_02** with **NAIA v1.1.0 ntuples.**

Skimming selections:

Physics trigger:	true
L1 charge (YJ):	1.4 < |Q| < 3.0
Inner tracker (YJ):	1.5 < |Q| < 2.4
LayerGoodPathl	true
Upper TOF:	1.5 < |Q| < 2.9
Lower TOF:	1.4 < |Q| < 3.0

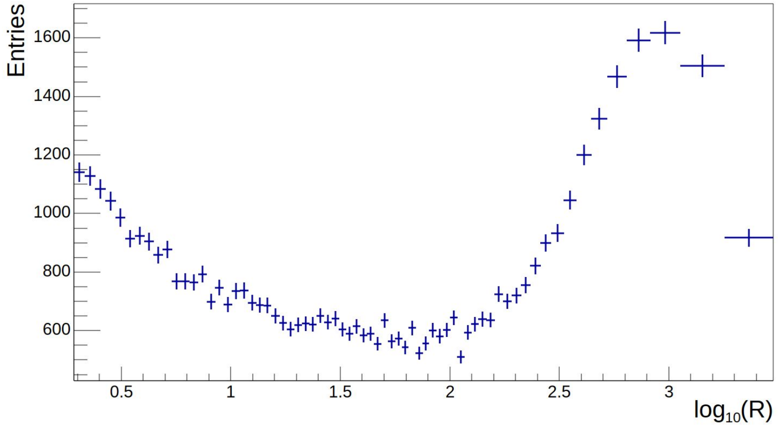


Signal and background definition

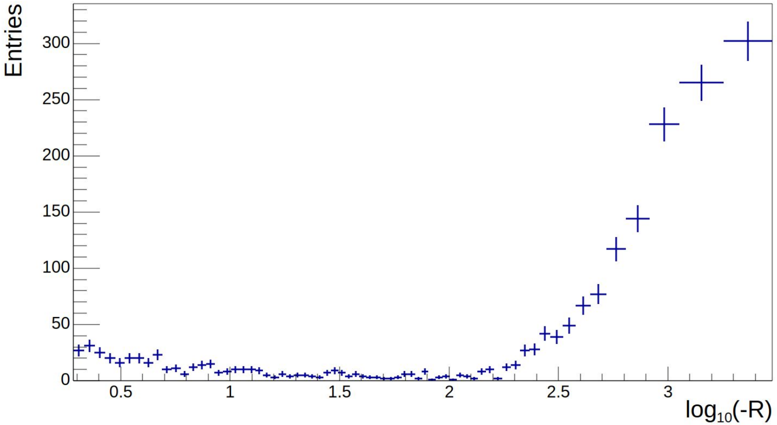
Sign of the reconstructed rigidity (R),inner-L1 using (GBL) to define signal and background.

Since MC has He events, well reconstructed events must have R > 0 and bad reconstructed have R < 0.

Signal: He events with R>0.



Background: He events with R<0.

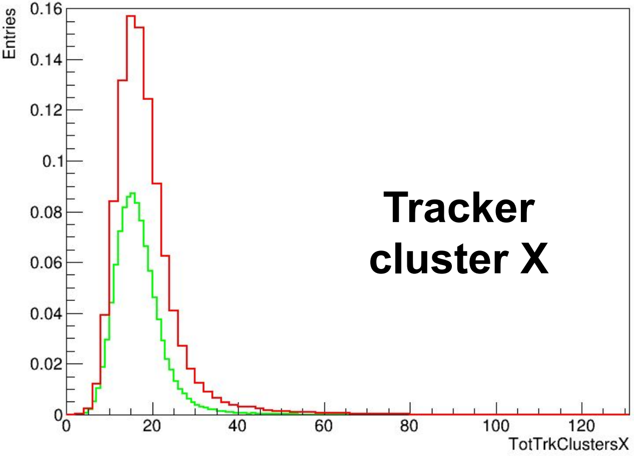


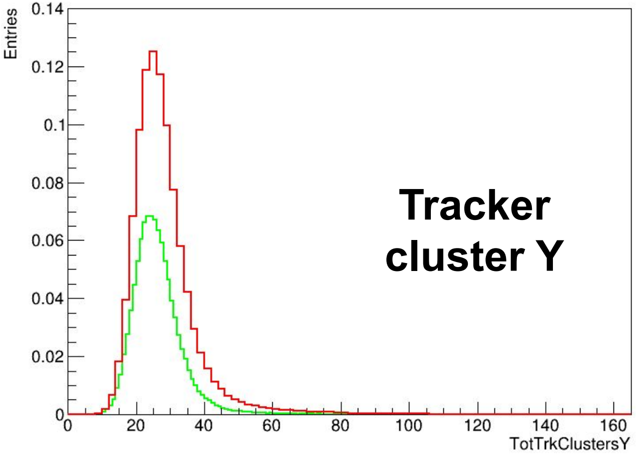
Interesting variables

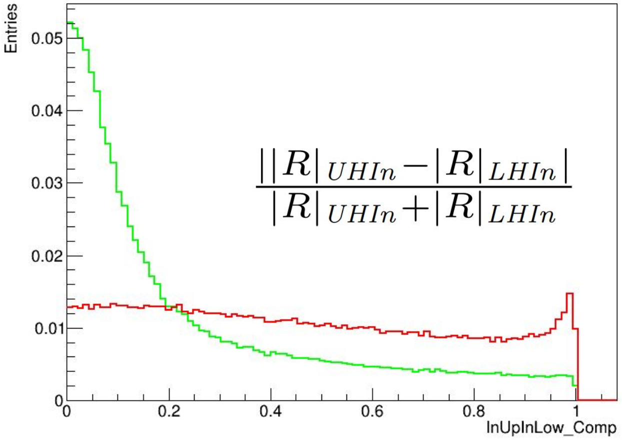
- Track pattern Y and XY sides (avoiding layer 9).
- Energy deposited in first eight layers of the tracker both Y and XY sides.
- Total number of tracker clusters, X and Y side.

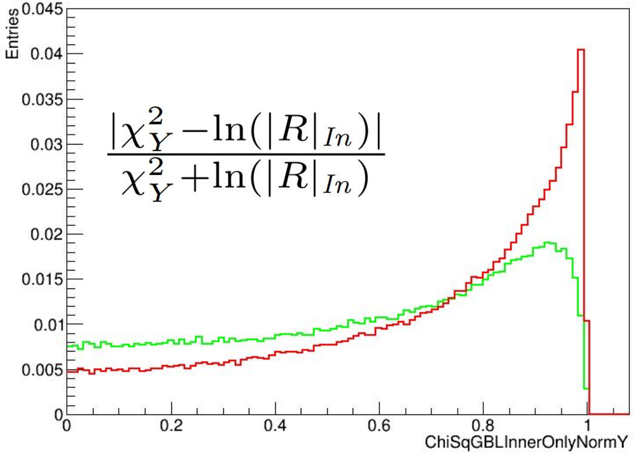
- Rigidity absolute values considering: inner-L1, inner only, upper and lower inner half
- Track $\chi^2_{X/Y}$ normalised by the rigidity $$\frac{|\chi^2_{X/Y} - \ln(|R|)|}{\chi^2_{X/Y} + \ln(|R|)}$$
- Compatibility between different spans: $$\frac{||R|_{UHIn} - |R|_{LHIn}|}{|R|_{UHIn} + |R|_{LHIn}}$$
- Number of ACC clusters.
- Number of ACC counters.

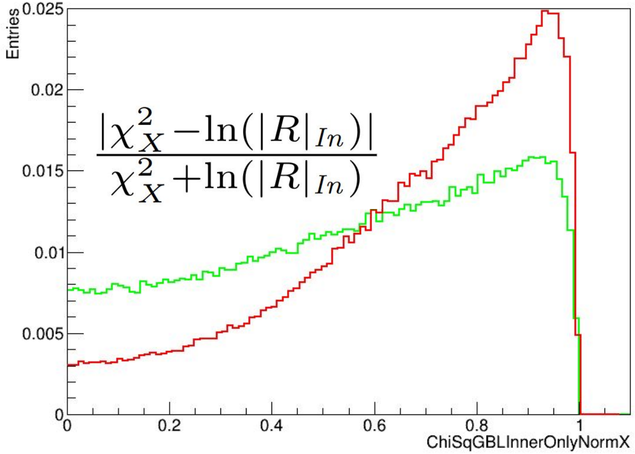
signal
background











Creation of the dataset

Further requirements:

- Track pattern requirement: one hit (Y or XY side) on 7/8 layers, L1 and L2 must be present.
- The rigidity fit (GBL) must exist for: inner-only, inner-L1, upper and lower inner half.
- $|R|_{In}$ and $|R|_{InL1}$ > 2 GV
- $|R|_{UHIn}$ and $|R|_{LHIn}$ > 1 GV

To avoid a bias, the dataset has the same amount of signal and background events. Only 1 out of 10 signal events is stored, until reaching the same amount of background events.

Total events number ~ 400 K

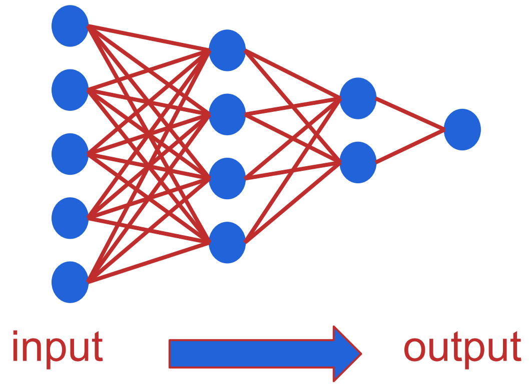
Fully connected neural network architecture

Fully connected neural network (FCNN) based on PyTorch.

The network has a customizable layers number and nodes per layer number.

The output of the network is a number ϵ [0,1].

The activation function is a Rectified Linear Unit (ReLU), and an *"Adam optimizer"* is used. The loss function is a Binary Cross Entropy.

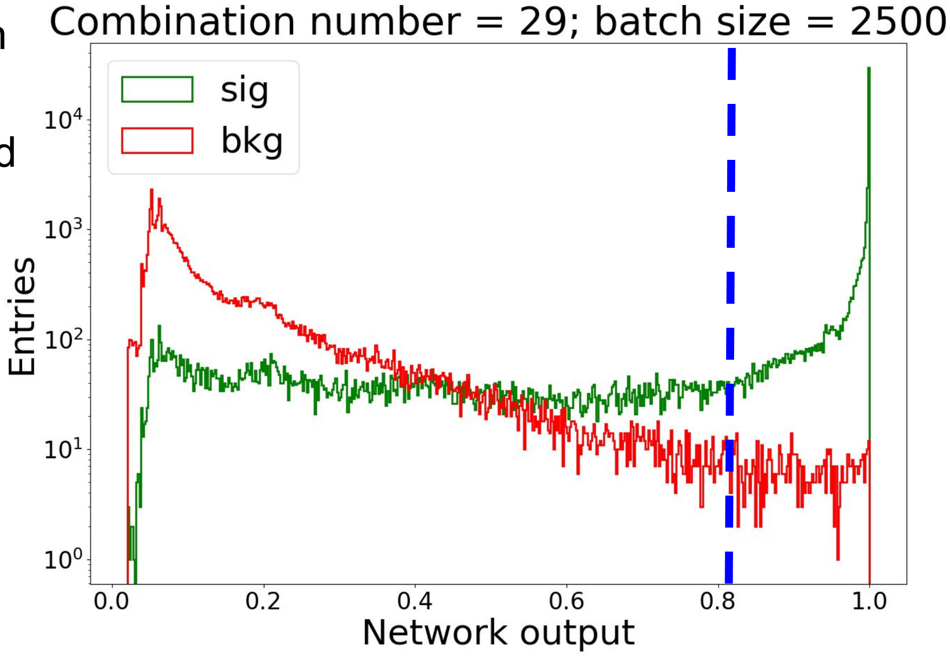


70% of the dataset is used for training, 30% for validation.
The area under the Receiver Operating Characteristic (ROC) curve is used to quantify the performance of the network.

ROC curve calculation strategy

Network output divided by true label indexes.

Signal efficiency: normalized signal integration above threshold.
Background efficiency: normalized background integration above threshold.
Background rejection: 1 - bkg efficiency.



Network architecture

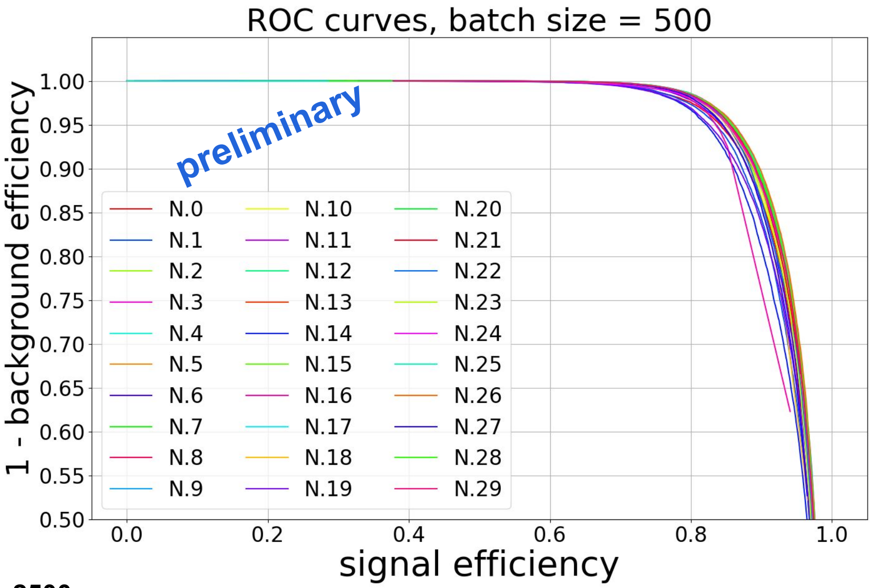
batch size = 2500
learning rate = 5e-3
Layers number = 8
Nodes per layer = 128, 128, 128, 128, 128, 128, 128

Grid search

Free parameters:

- 2 batch size: [500, 2'500]
- 2 learning rate: [5e-4, 5e-3]
- 5 layer multiplicity: [4, 5, 6, 7, 8]
- 7 nodes per layer: [128, 64, 32, 16, 8, 4, 2]

60 different architectures have been evaluated.

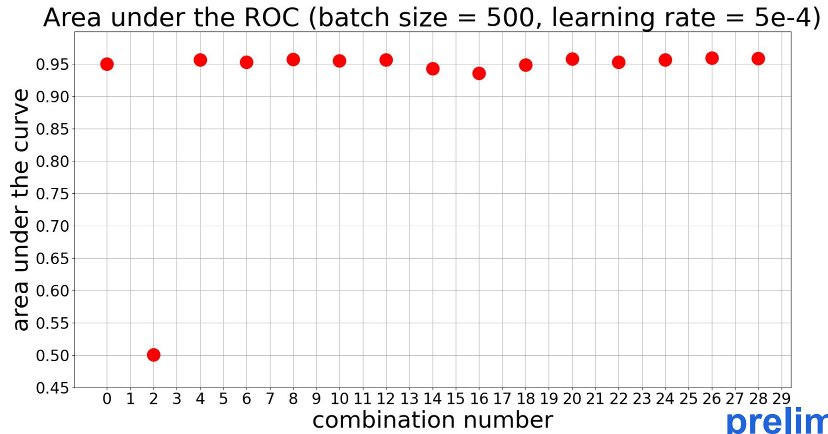


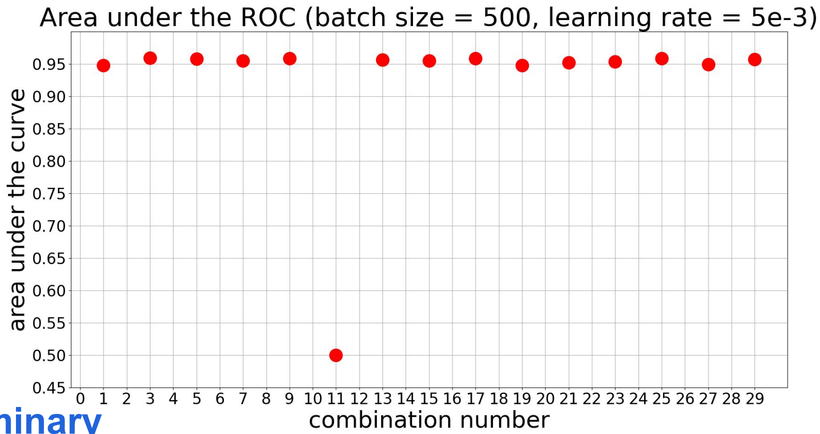
Batch size = 500, 2500

Combination number	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Layers numbers	4	4	4	4	4	4	5	5	5	5	5	5	6	6	6
Nodes per layer	64, 32, 16	64, 32, 16	16, 32, 64	16, 32, 64	64, 64, 64	64, 64, 64	64, 32, 16, 8	64, 32, 16, 8	8, 16, 32, 64	8, 16, 32, 64	64, 64, 64, 64	64, 64, 64, 64	64, 32, 16, 8, 4	64, 32, 16, 8, 4	4, 8, 16, 32, 64
learning rate	5e-4	5e-3	5e-4	5e-3	5e-4	5e-3	5e-4	5e-3	5e-4	5e-3	5e-4	5e-3	5e-4	5e-3	5e-4

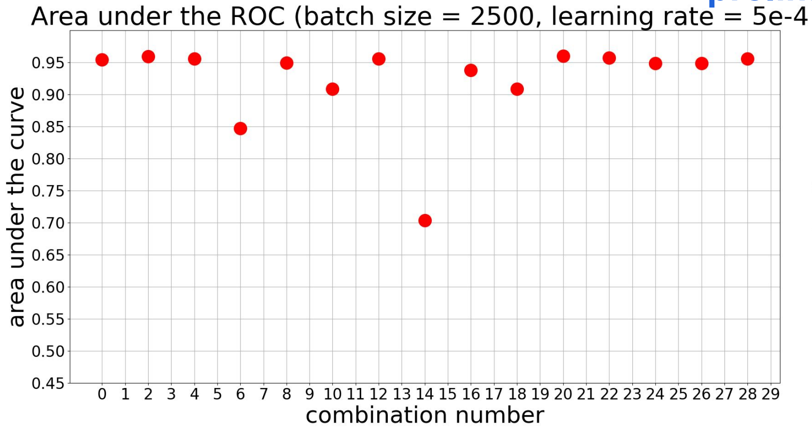
Combination number	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
Layers numbers	6	6	6	7	7	7	7	7	7	8	8	8	8	8	8
Nodes per layer	4, 8, 16, 32, 64	64, 64, 64, 64, 64	64, 64, 64, 64, 64	64, 32, 16, 8, 4, 2	64, 32, 16, 8, 4, 2	2, 4, 8, 16, 32, 64	2, 4, 8, 16, 32, 64	64, 64, 64, 64, 64, 64	64, 64, 64, 64, 64, 64	128, 64, 32, 16, 8, 4, 2	128, 64, 32, 16, 8, 4, 2	2, 4, 8, 16, 32, 64, 128	2, 4, 8, 16, 32, 64, 128	128, 128, 128, 128, 128, 128, 128	128, 128, 128, 128, 128, 128, 128
learning rate	5e-3	5e-4	5e-3	5e-4	5e-3	5e-4	5e-3	5e-4	5e-3	5e-4	5e-3	5e-4	5e-3	5e-4	5e-3

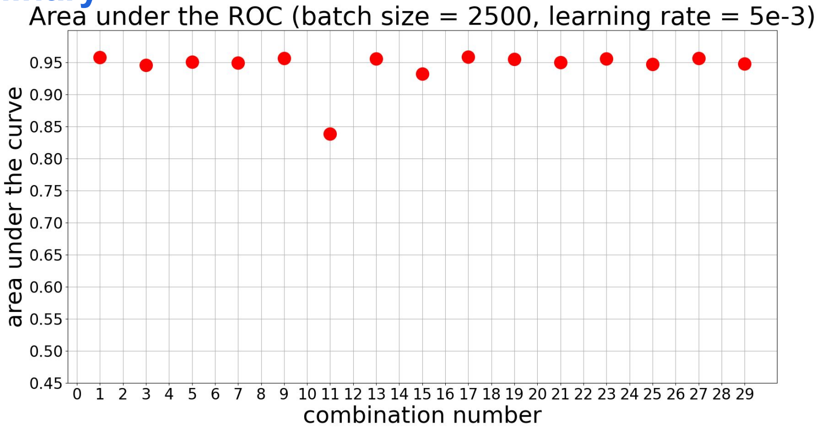
Area under the ROC curve





preliminary





Conclusions

The data preprocessing and the preliminary performance of a FCNN for He event reconstruction quality have been presented.

Next steps

- Further tests are required to understand if it is optimal to use the absolute value of rigidity as an input feature.
- Data preprocessing could be improved scaling all the features in the same range.
- MC is generated flat in logarithm, it might be necessary to reproduce the true spectra (upsampling?).
- Study the few background events with a network output ~ 1.

Conclusions

The data preprocessing and the preliminary performance of a FCNN for He event reconstruction quality have been presented.

A special thanks to
Greta Brianti
&
Francesco Follega
for the help

Next steps

- Further tests are required to understand if it is optimal to use the absolute value of rigidity as an input feature.
- Data preprocessing could be improved scaling all the features in the same range.
- MC is generated flat in logarithm, it might be necessary to reproduce the true spectra (upsampling?).
- Study the few background events with a network output ~ 1.

Conclusions

The data preprocessing and the preliminary performance of a FCNN for He event reconstruction quality have been presented.

THANK YOU FOR YOUR ATTENTION!

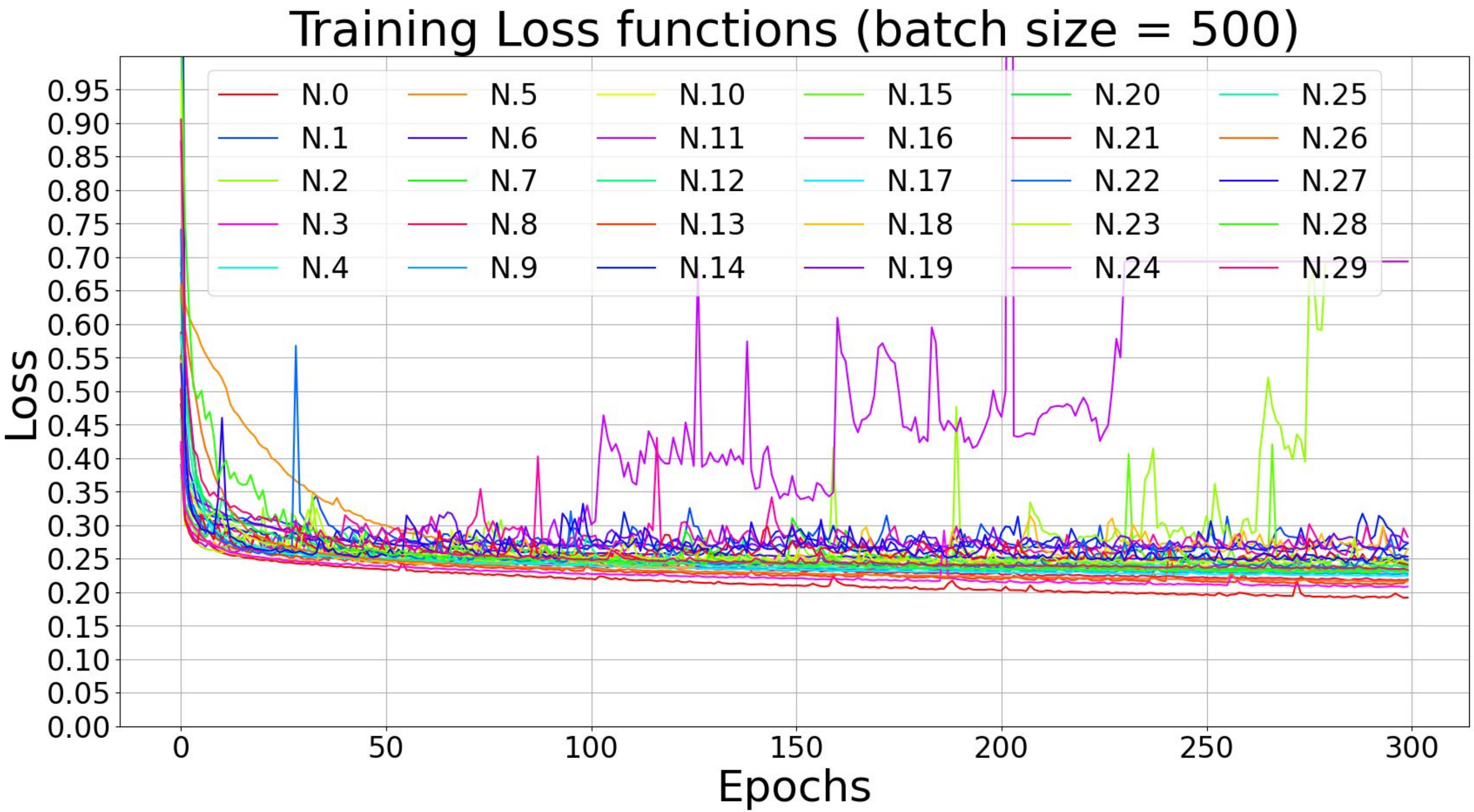
A special thanks to
Greta Brianti
&
Francesco Follega
for the help

Next steps

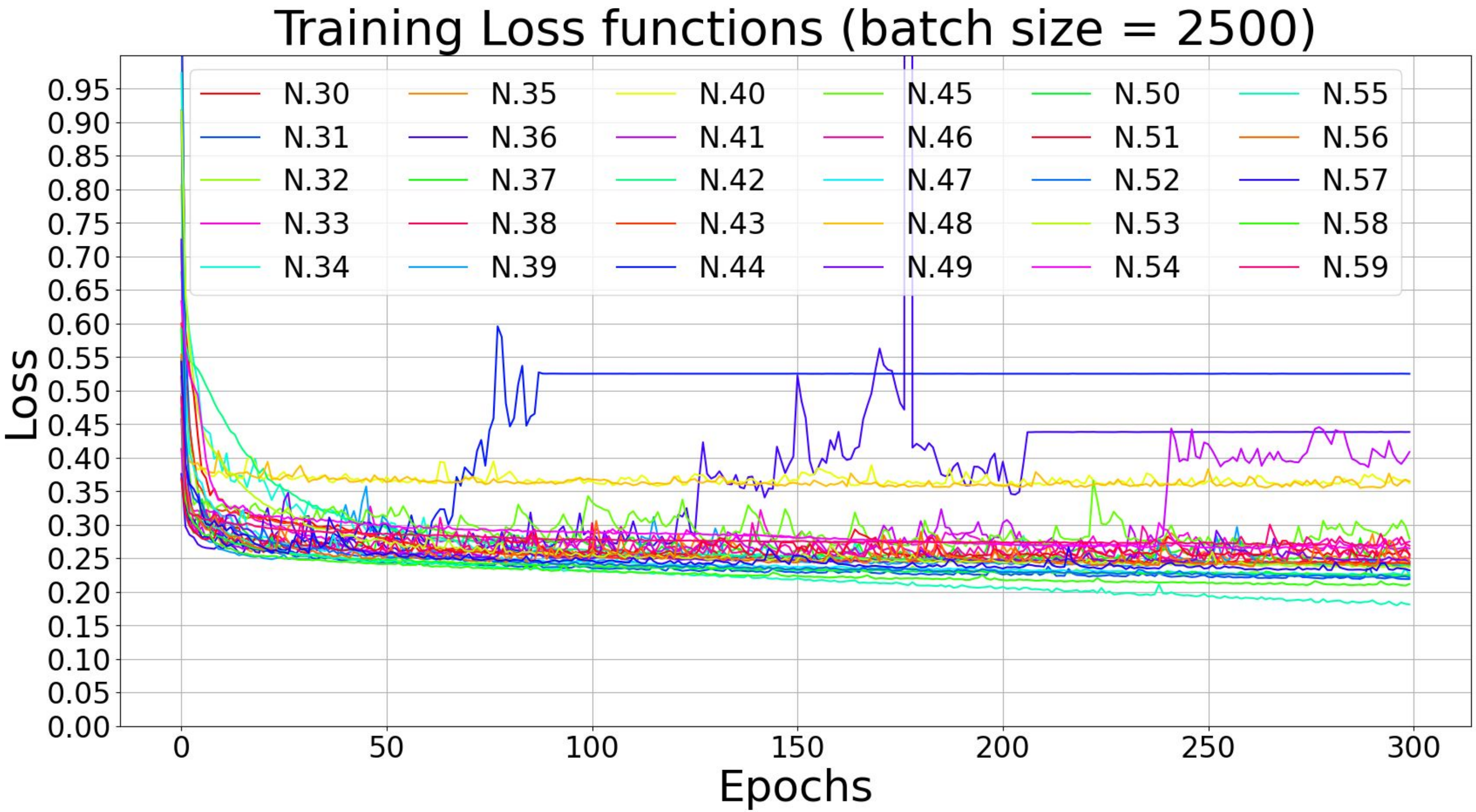
- Further tests are required to understand if it is optimal to use the absolute value of rigidity as an input feature.
- Data preprocessing could be improved scaling all the features in the same range.
- MC is generated flat in logarithm, it might be necessary to reproduce the true spectra (upsampling?).
- Study the few background events with a network output ~ 1.

Backup

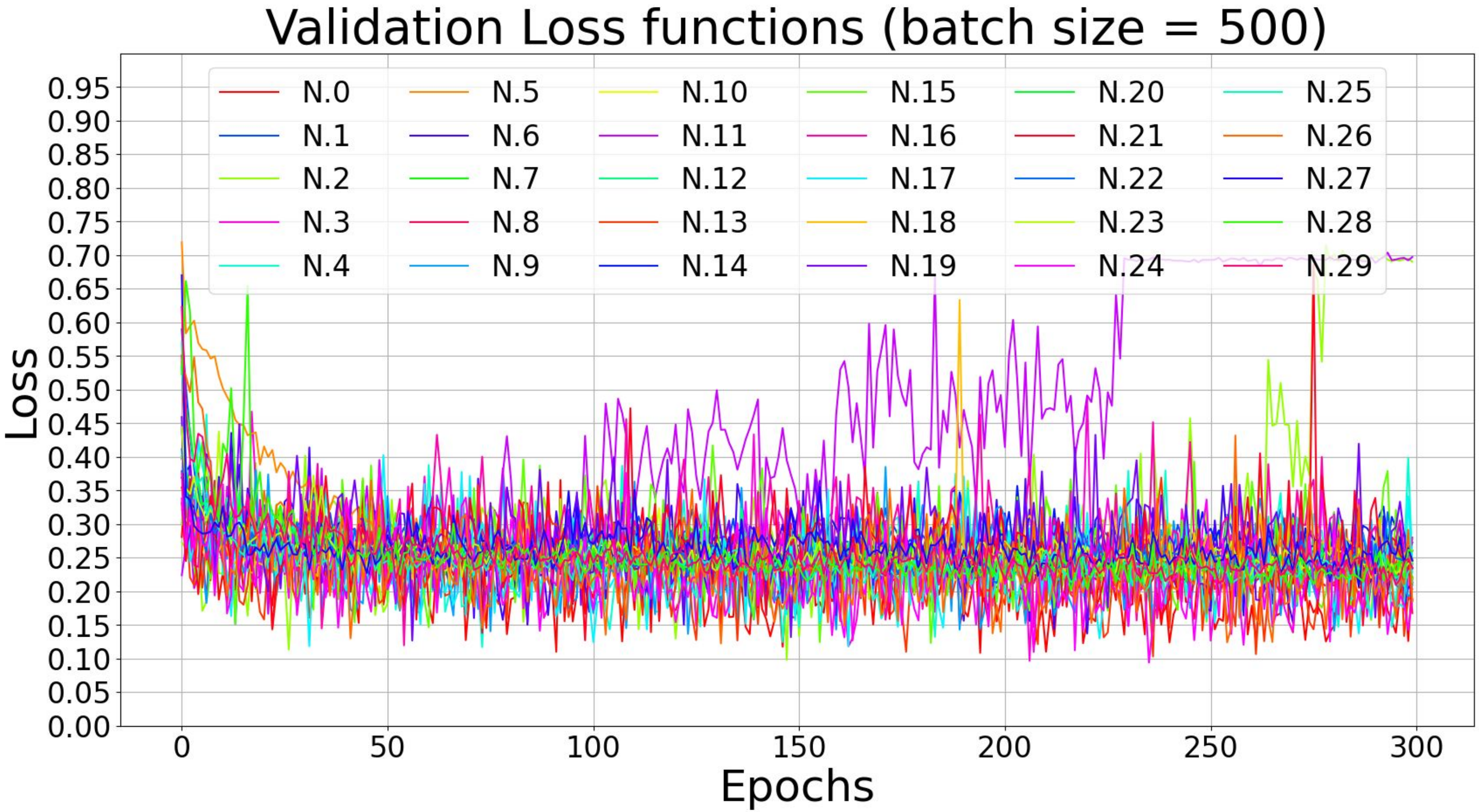
Training loss functions



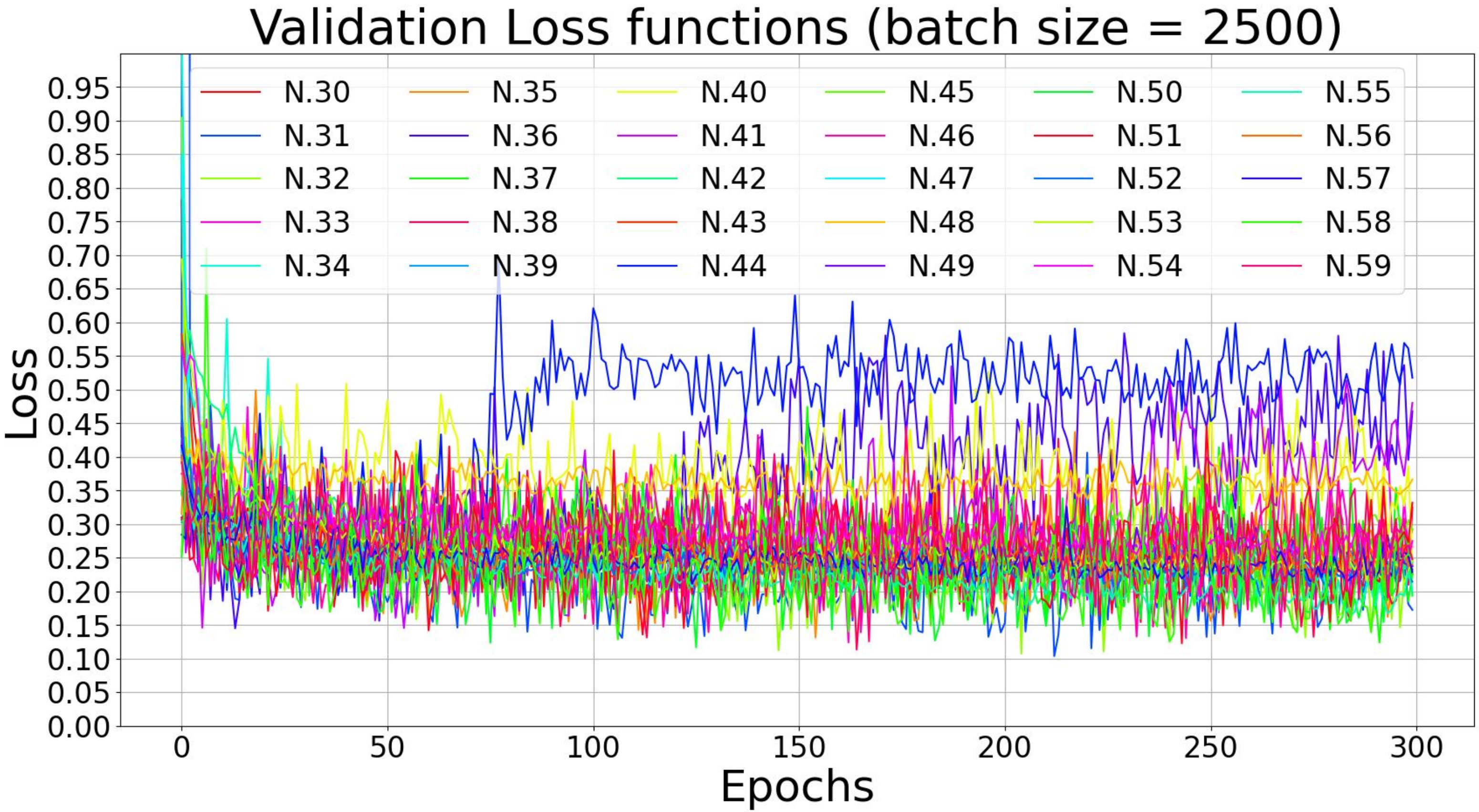
Training loss functions



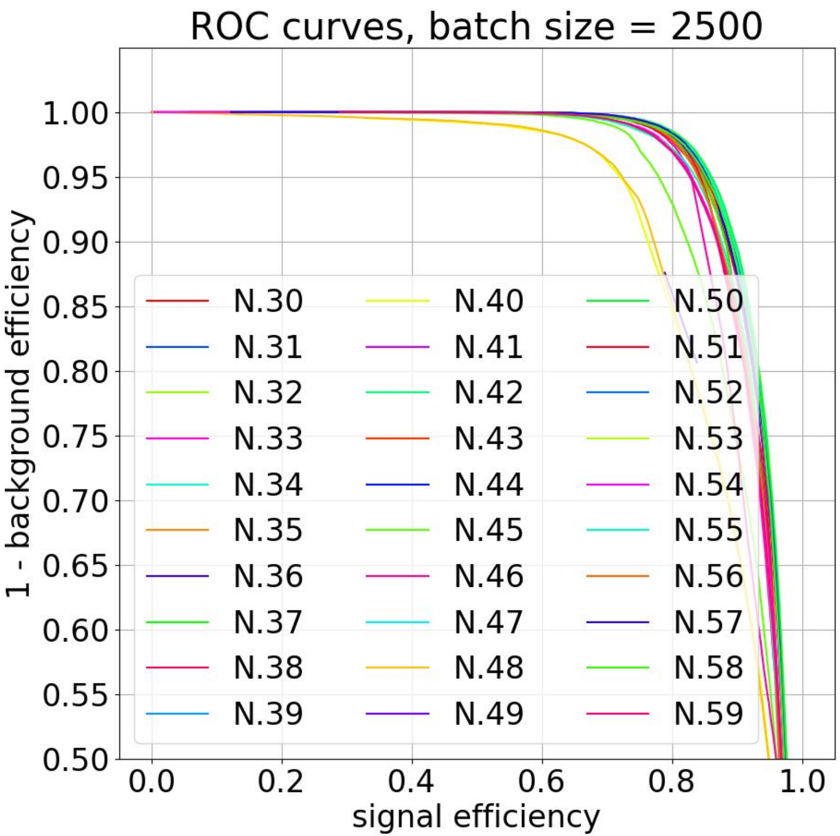
Validation loss functions



Validation loss functions



Roc curves


ROC curves, batch size = 2500
1 - background efficiency
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.0
0.2
0.4
0.6
0.8
1.0
signal efficiency
N.30
N.31
N.32
N.33
N.34
N.35
N.36
N.37
N.38
N.39
N.40
N.41
N.42
N.43
N.44
N.45
N.46
N.47
N.48
N.49
N.50
N.51
N.52
N.53
N.54
N.55
N.56
N.57
N.58
N.59