Hit-reco: ProtoDUNE denoising with DL models

Marco Rossi[1,2]

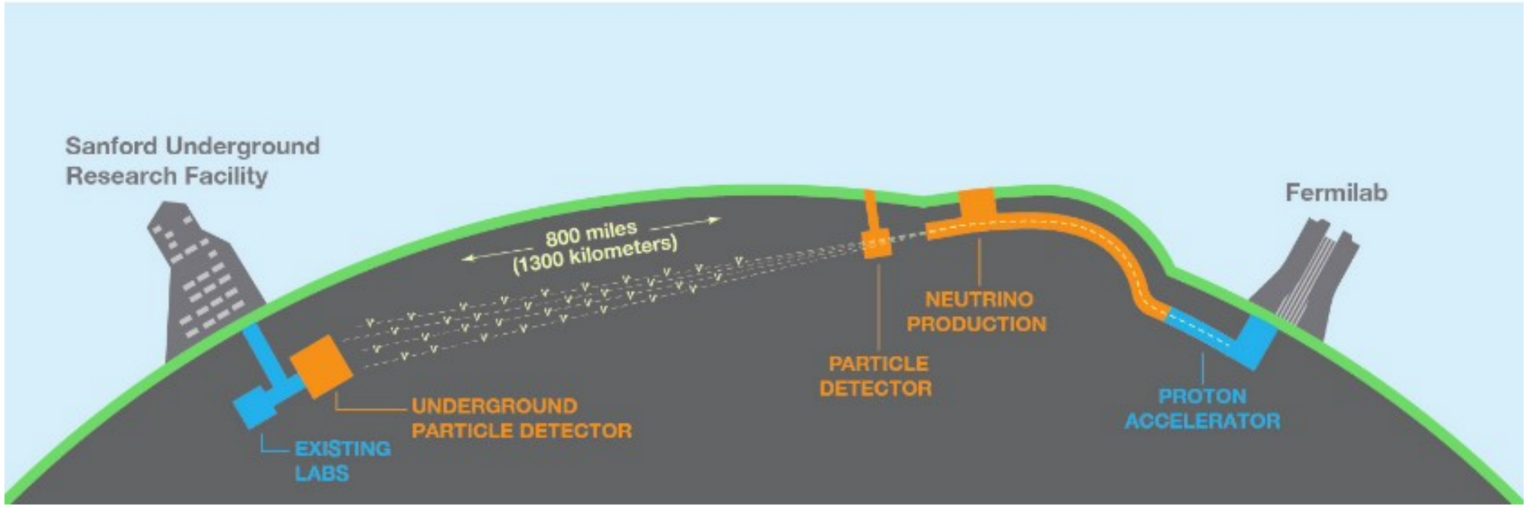
[1]University of Milan [2]CERN openlab

4th IML Machine Learning Workshop

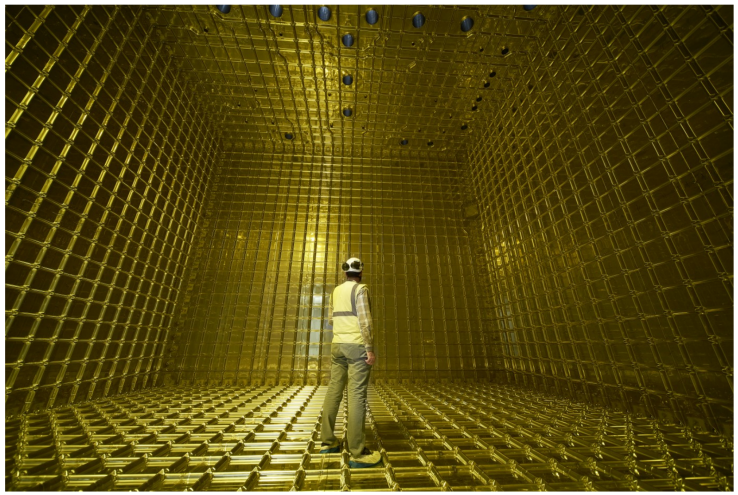
October 21st 2020

DUNE - Deep Underground Neutrino Experiment

DUNE (from 2026) $\longrightarrow$ Neutrino oscillations

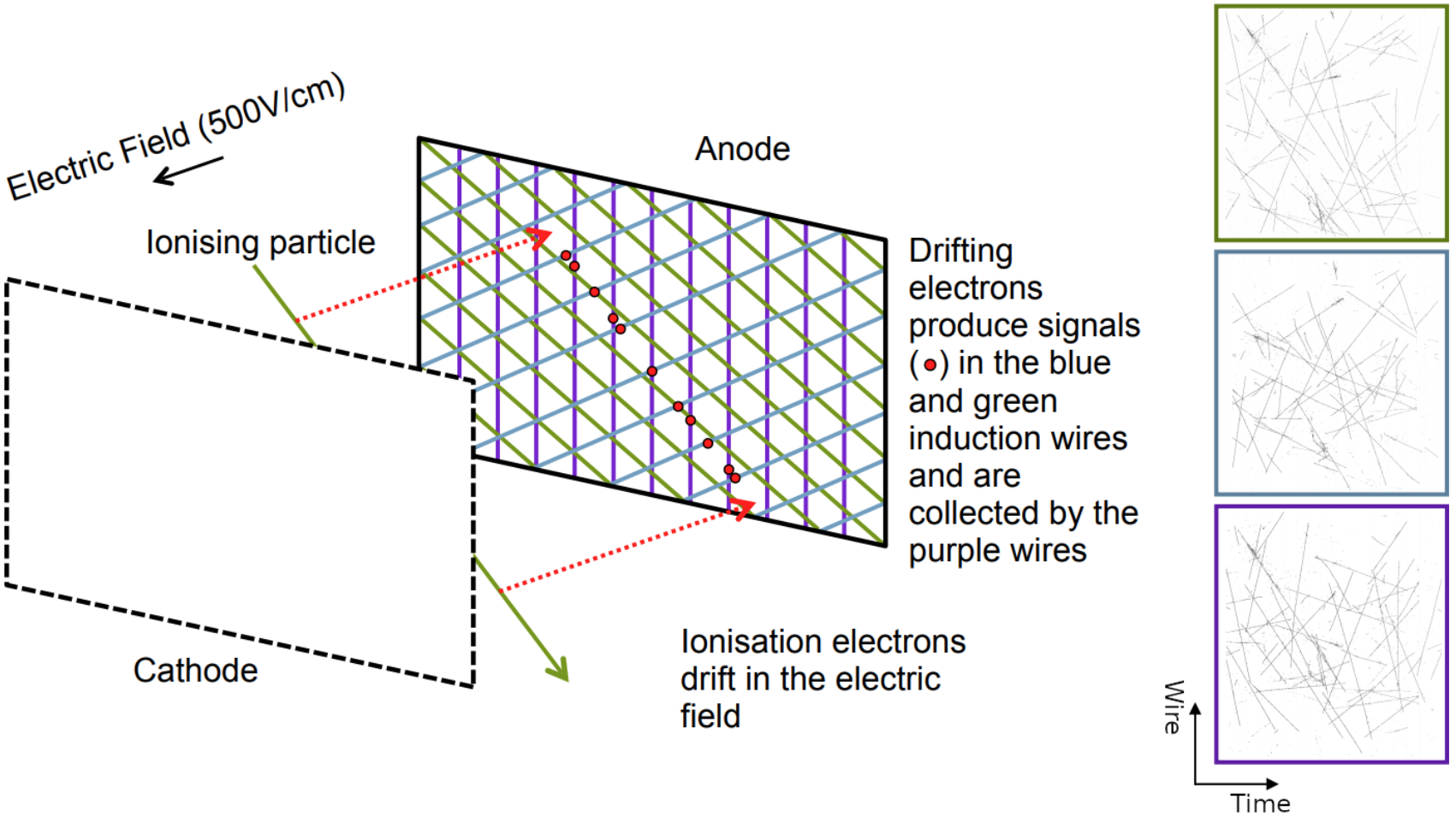


protoDUNE (from 2017) $\longrightarrow$ Test and validate technologies

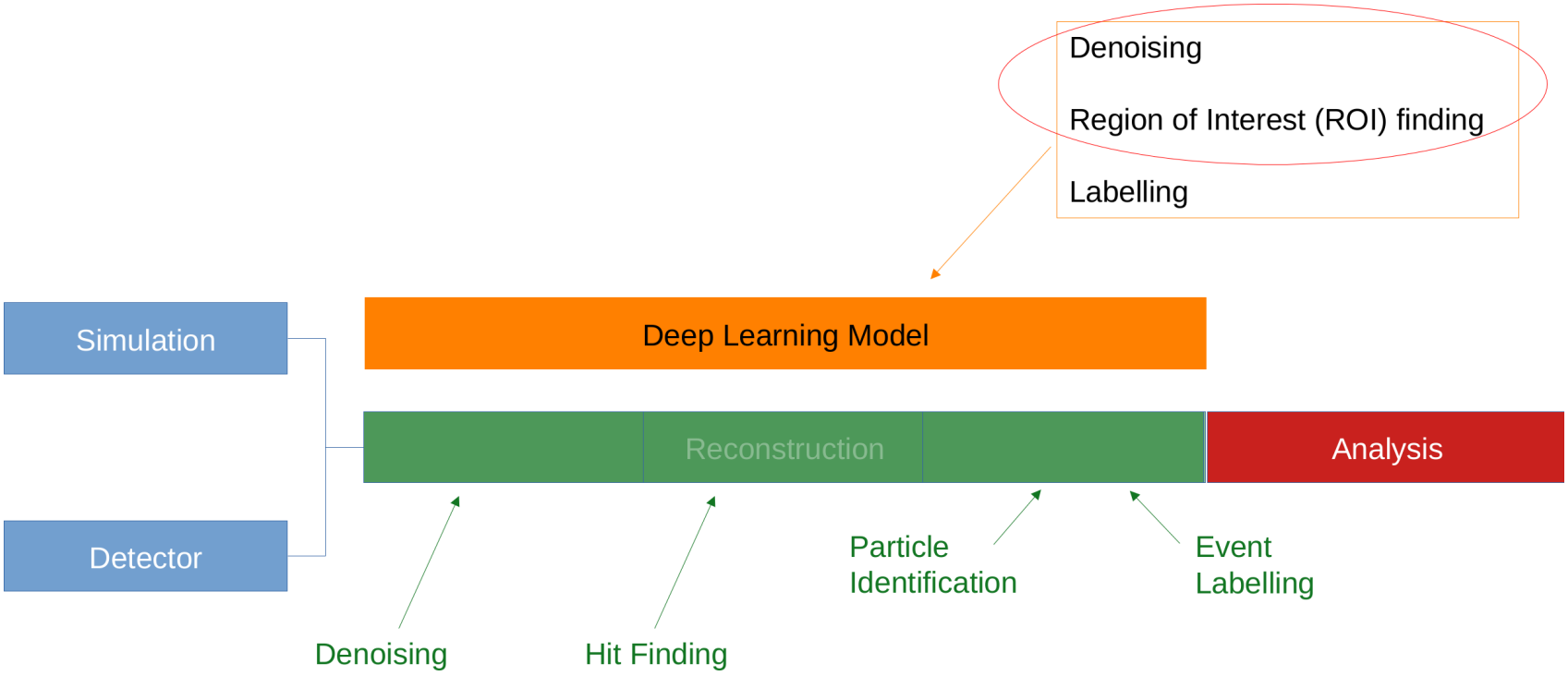


LArTPC - Liquid Argon Time Projecting Chamber

- Big box filled with liquid Ar
- Electronics shapes electron induced current
- Plot Raw Digits as images (ADC counts on Time vs Wire)



Goals of the study

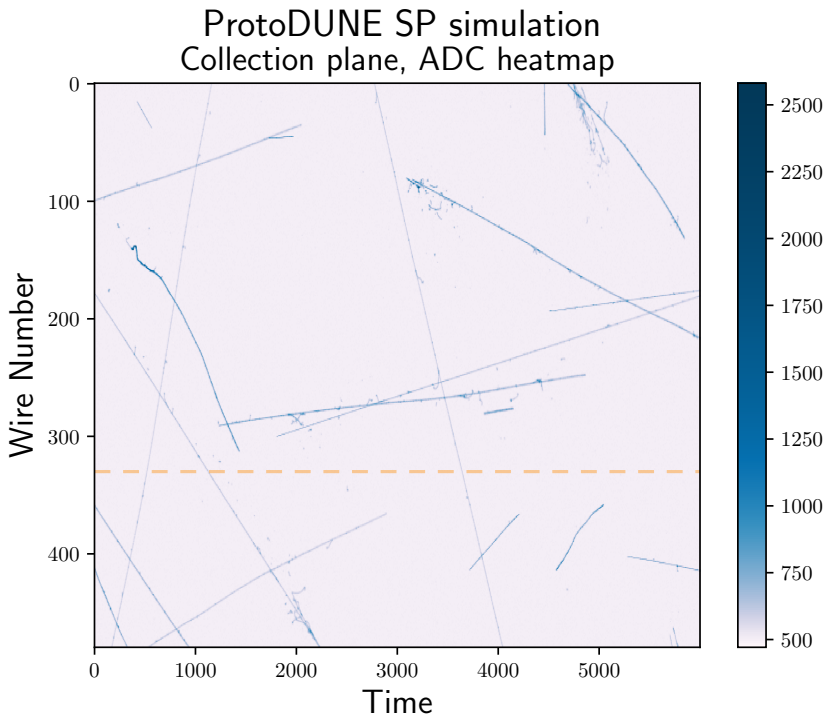


- Replace protoDUNE reconstruction with deep learning
- `Hit-reco` : region of interest (ROI) selection and denoising (DN) of raw data
- Assess capabilities of Graph Neural Networks

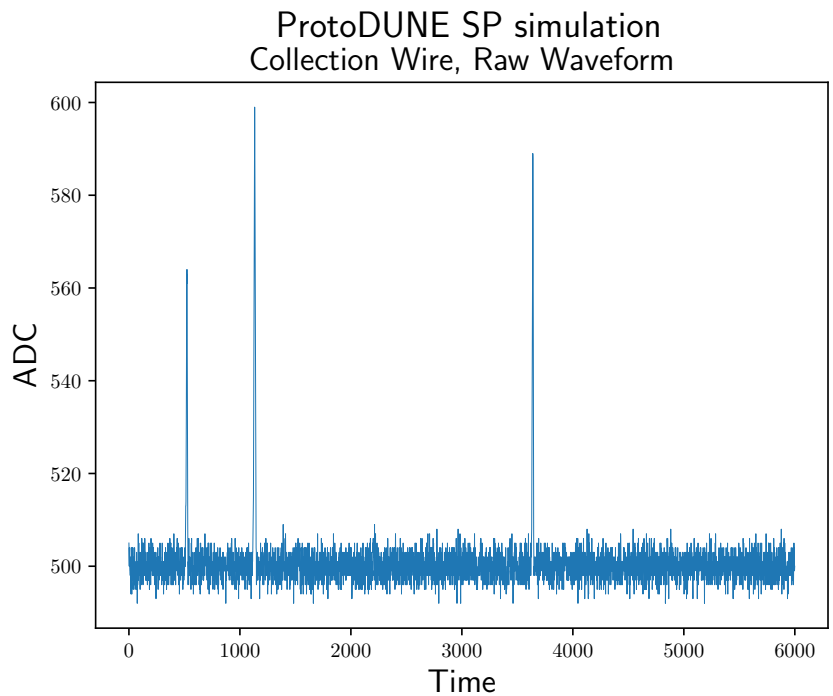
Raw Digits - an example

- Collection Plane view: 480 wire channels
- Detector outputs digitized form of the current: ADC counts
- Time window: 6000 detector timeticks @2 MHz $\rightarrow$ 3 ms

2D view

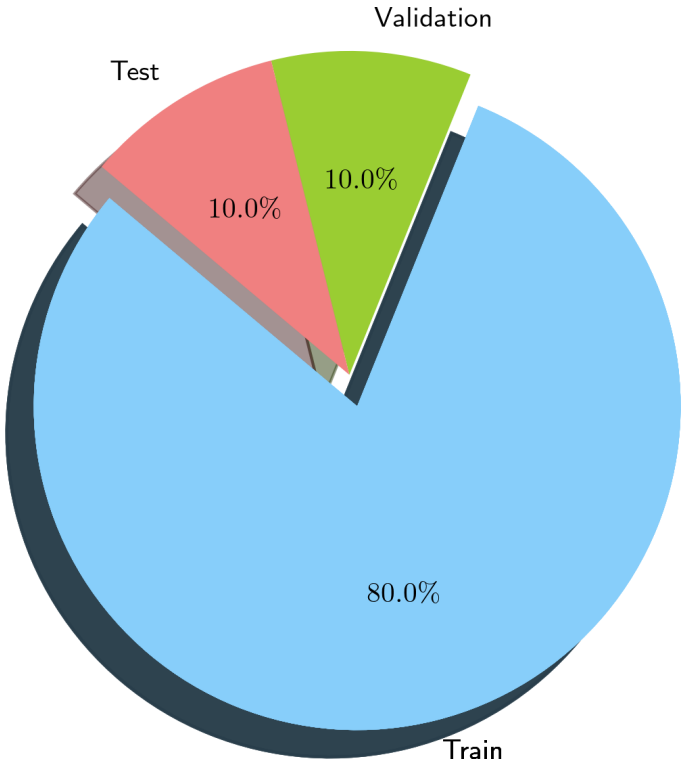


1D view



Dataset

- 10 simulated beam spill events with `duntepc v08_24_00`
- Model inputs are collection plane 2D views: arrays (960×6000)
- Current version focuses on collection plane only
- Planes are cropped in (32×32) images to fit memory
- Train set size $24k$ crops

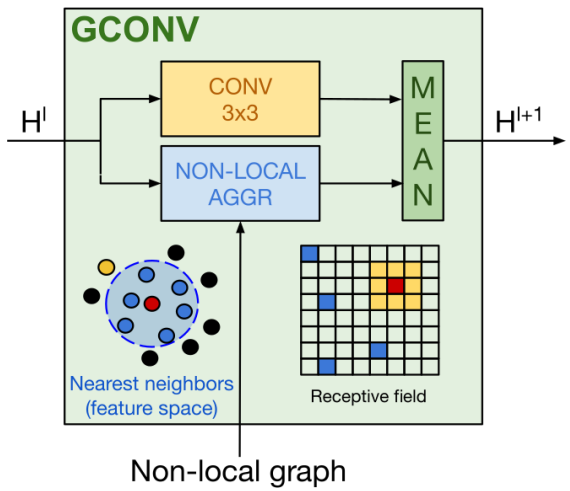


Proposed Model

Graph Convolutional Neural Network layer

- ✓ Non local graph
- ✓ KNN graph
- ✓ Long distance correlations
- ✓ $k = 8$
- ! Complexity $\mathcal{O}(n^2)$

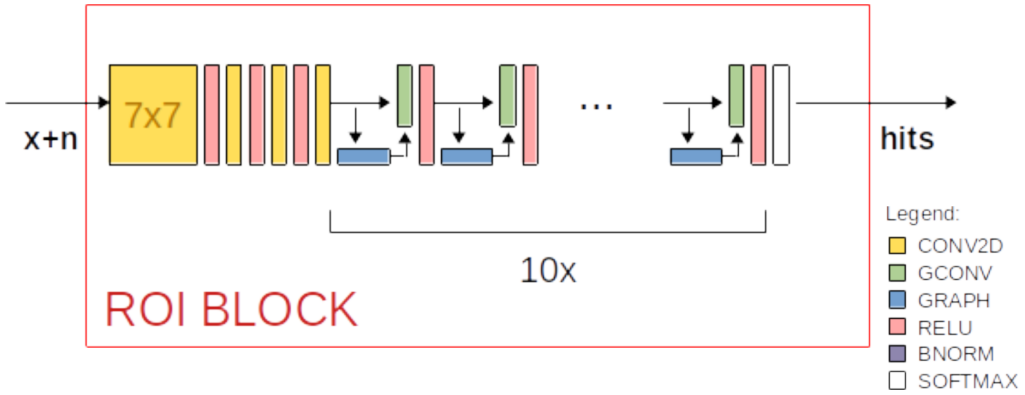
Reference: arXiv:1907.08448



- ✓ Convolutional filter
- ✓ Short distance correlations

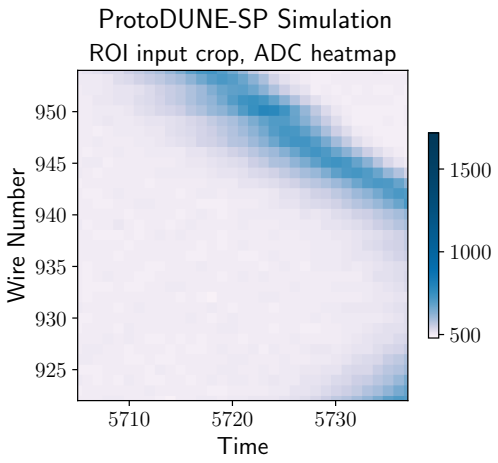
Model Overview - ROI

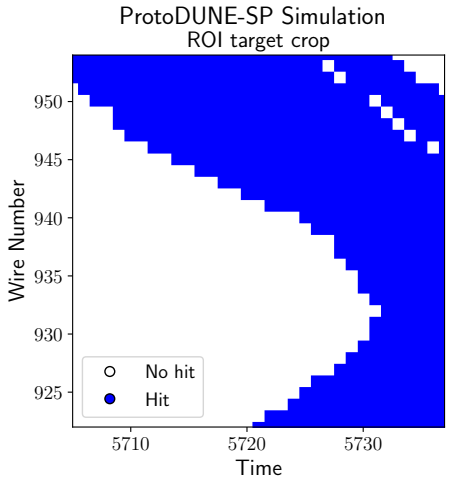
Region of interest finding (ROI)



Train a classifier:

1 $\longrightarrow$ pixel w charge
0 $\longrightarrow$ pixel w/o charge





What is signal?

Above Electronic Noise Charge

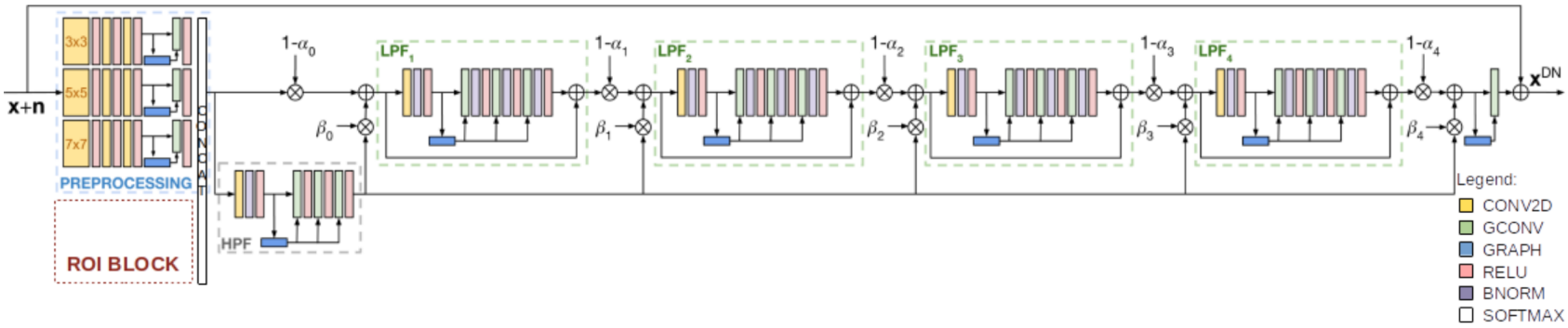
at protoDUNE-SP: $\sim 3.5\,ADC$
in collection plane [Reference, slide 18]

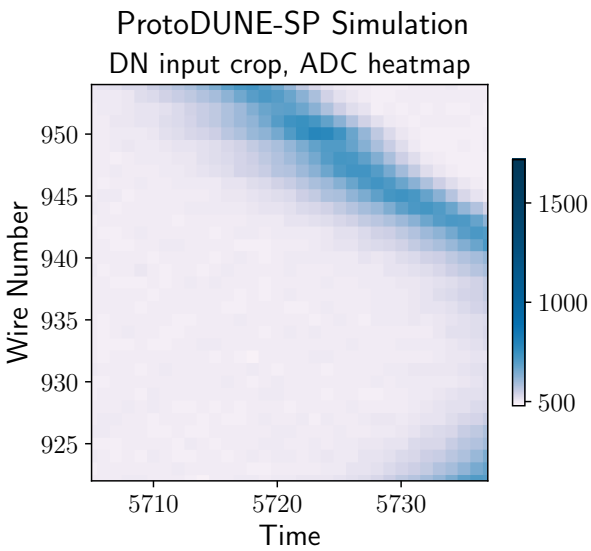
Technical features:

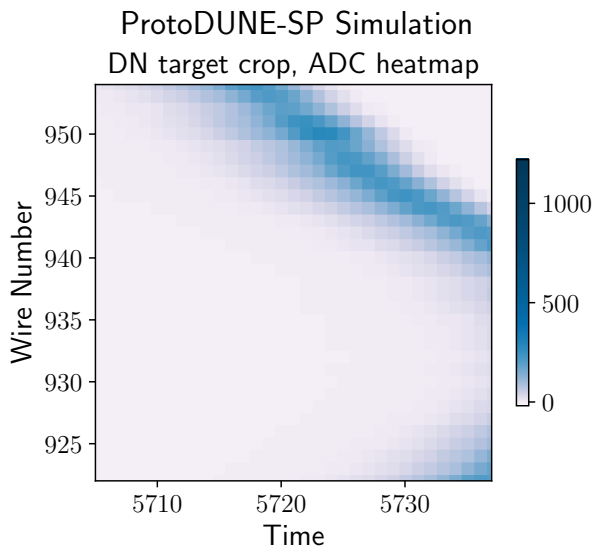
- Inputs are crops (32×32)
- $\mathcal{O}(10^5)$ trainable parameters
- Binary Cross Entropy loss function
- Adam Optimizer

Model Overview - DN

Reference: arXiv:1907.08448







Technical features:

- Inputs are crops (32×32)
- $\mathcal{O}(10^6)$ trainable parameters
- Loss function: $(1 - \text{SSIM}) + \text{MSE}$
- Adam Optimizer

Benchmark - ROI

Benchmark ROI model against traditional hit finding method

Larsoft product: `recob::Hit`

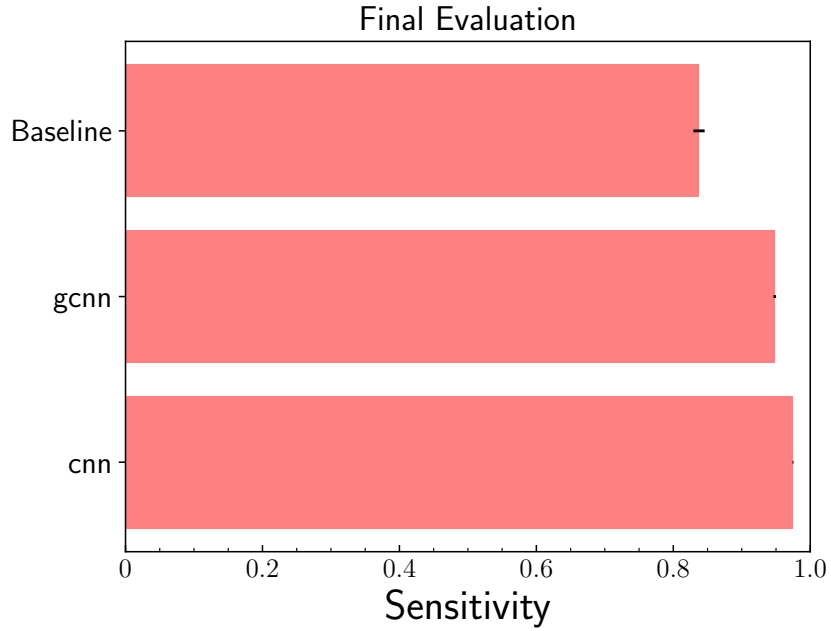
- 2D deconvolution
- Finds ROIs

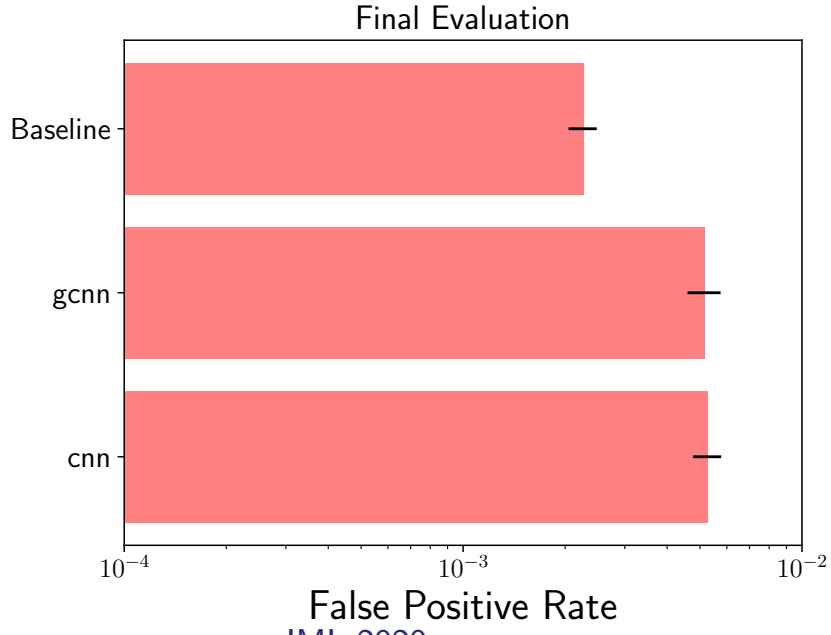
Total pixels in test set: $\mathcal{O}(34M)$

Balancing hits/non-hits: 1.9%

Figures of merit:

- sensitivity: $\frac{TP}{TP+FN}$
- false positive rate: $\frac{FP}{TN+FP}$

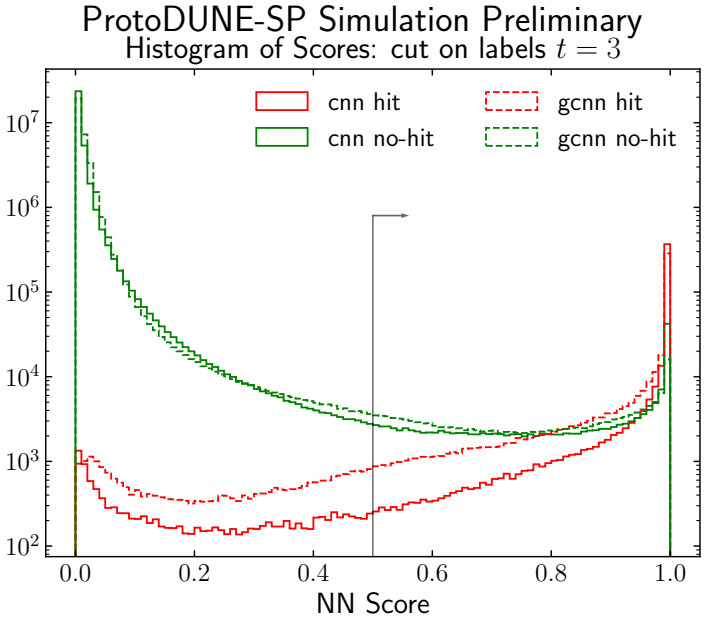




Benchmark - ROI

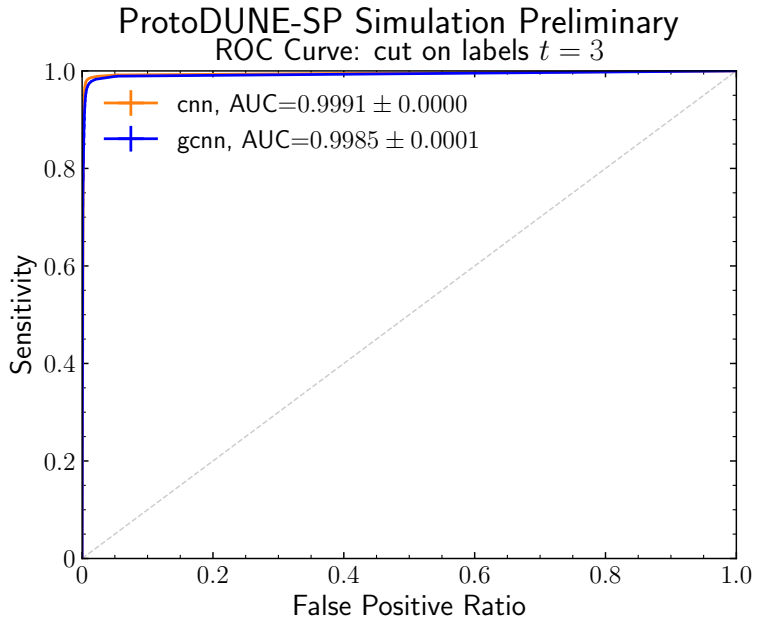
Scores histogram

The more separated the curve, the better



Receiving Operating Characteristic curve

Area under curve parameter



Tails are around three orders of magnitude lower than peaks.
AUC parameter is really close to unity.

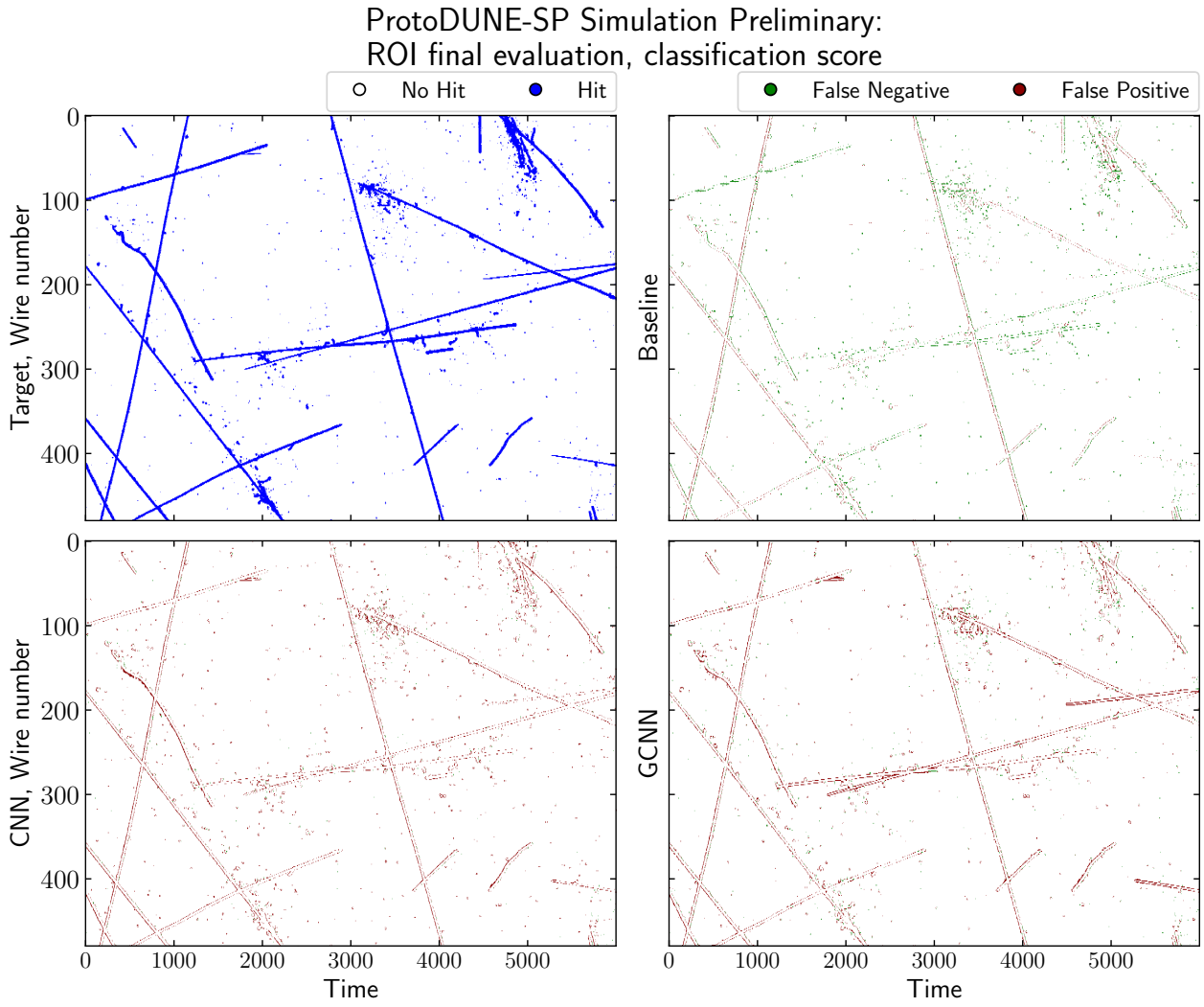
Results - ROI

Overall good agreement

NNs vs DUNE baseline tool:

- ✓ higher sensitivity
- ! higher FPR

Some tweak needed to clean around tracks



Benchmark - DN

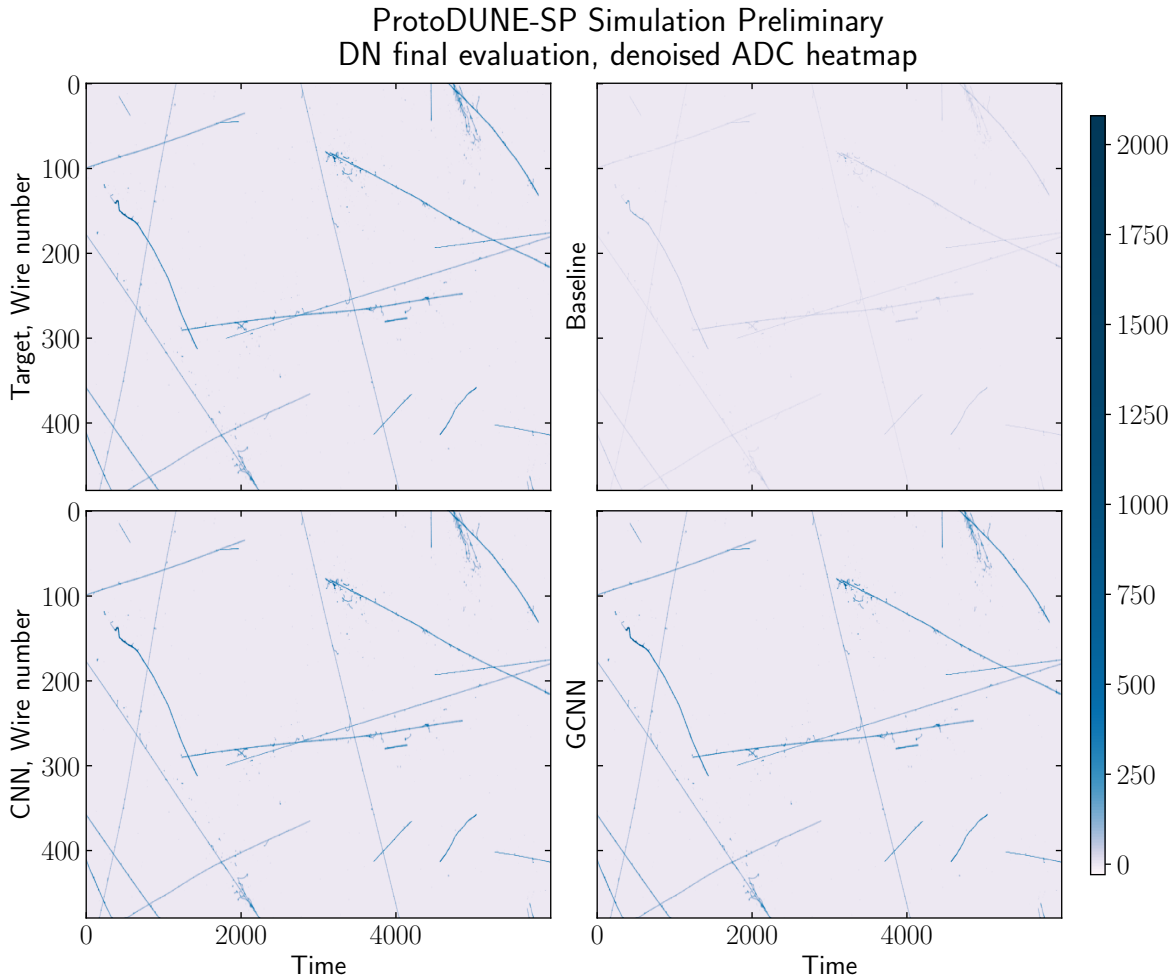
Benchmark DN model against traditional denoising method

Larsoft product: `recob::Wire`

- Takes ROI regions (deconvolution)
- Fits peaks with Gaussians

✓ Good match NNs-labels

! Need to change baseline's normalization

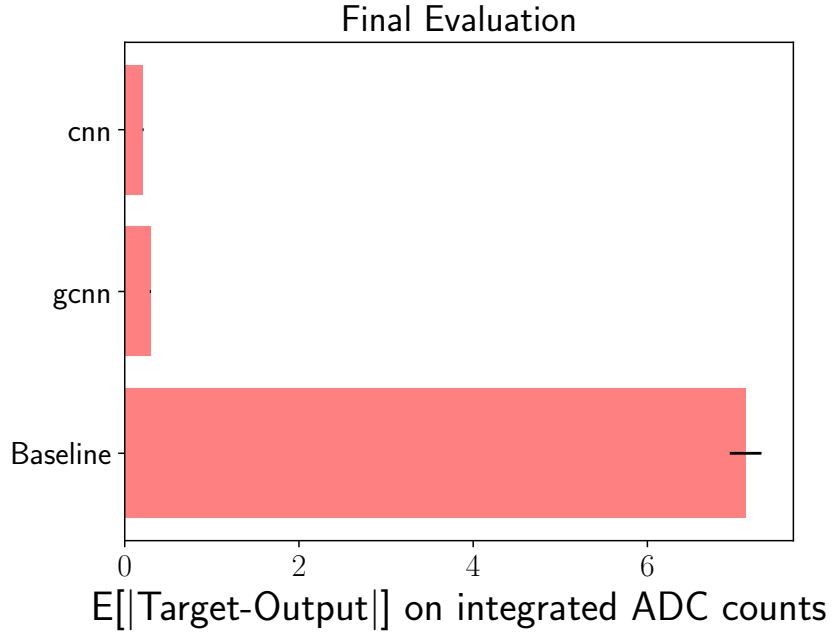


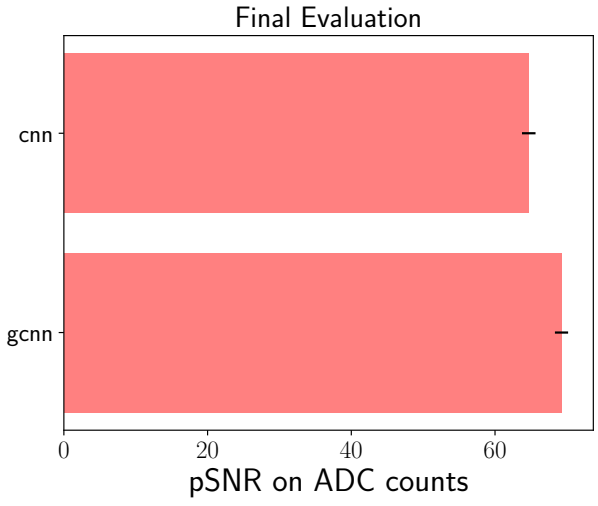
Results - DN

Deconvolution preserves areas, not amplitudes !

- Integrate ADC values over time
- Adjust the normalization
- Compute $\mathbb{E}[|\text{Target} - \text{Output}|]$

CNN slightly better GCNN on integrated ADC





Actually over the waveforms amplitudes (2D plane view):
GCNN performs better on pSNR

Summary

- Denoising and region of interest selection models for protoDUNE simulation events
- Benchmark CNN and Graph Networks against DUNE Baseline algorithms
- NNs succeded in outperforming traditional tools

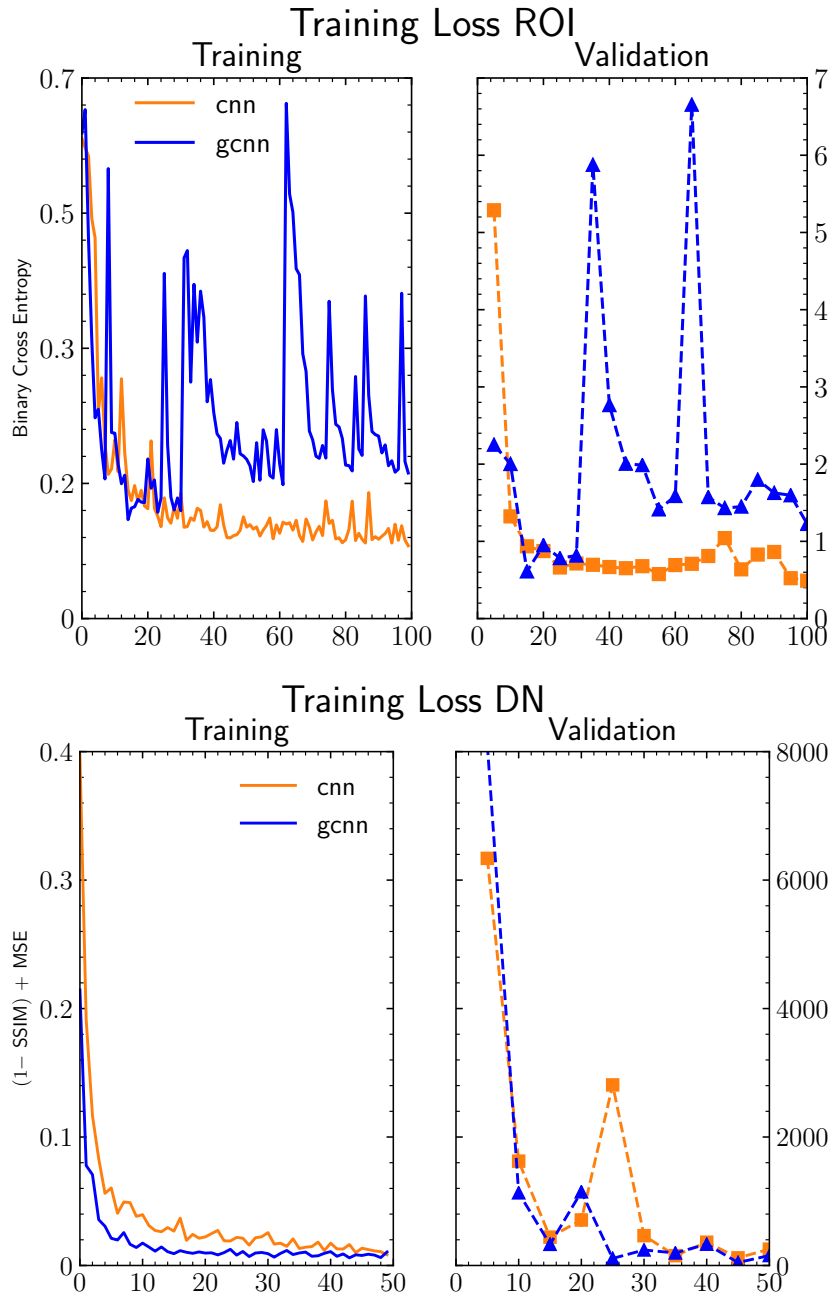
Future work:

- Enlarge training and test datasets
- Use latest simulation software (`dunetpc`) version
- Hyperparameter tuning to improve performances

Thanks!

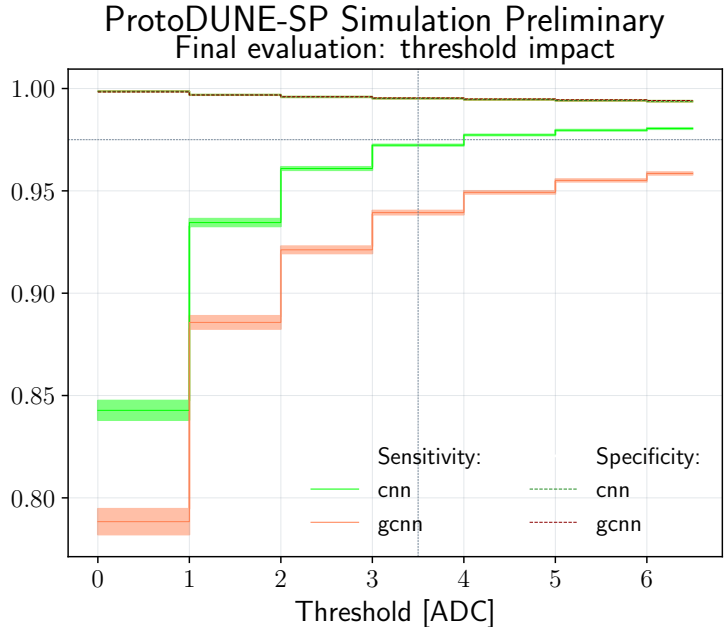
Backup - Training Strategy

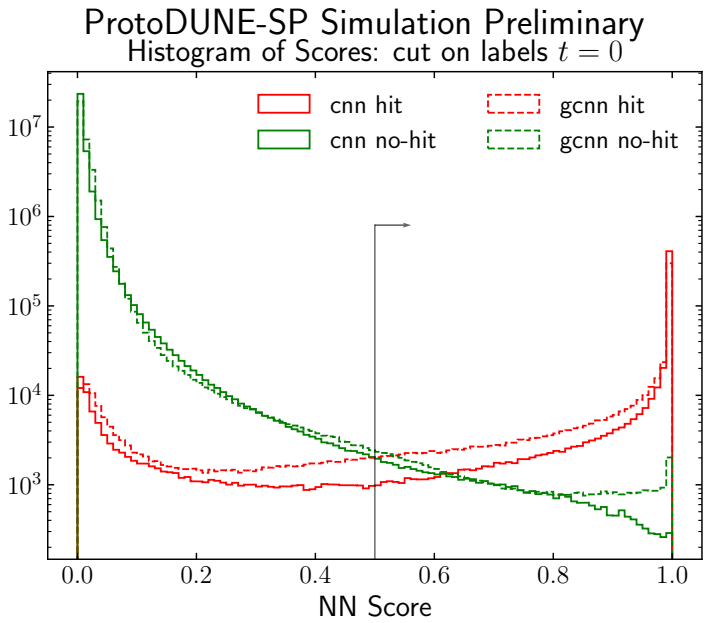
1. Train ROI for 100 epochs
2. Save best weights configuration
3. Load weights in ROI block for DN
4. Train DN for 50 epochs
5. Save best weights configuration

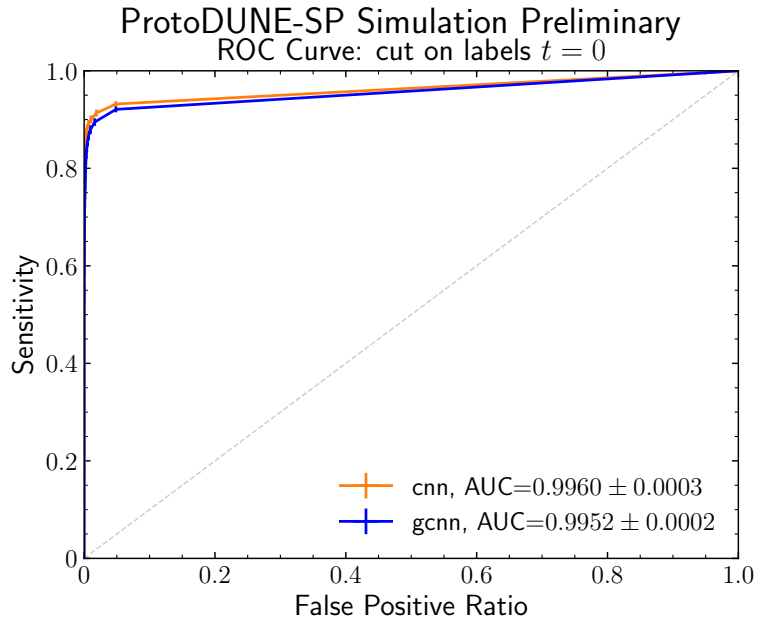


Backup - Label Cut Impact

- Cut on the labels has an impact on ROI performaces
- Sensitivity increases with the threshold
- Higher cut values mean only big peaks



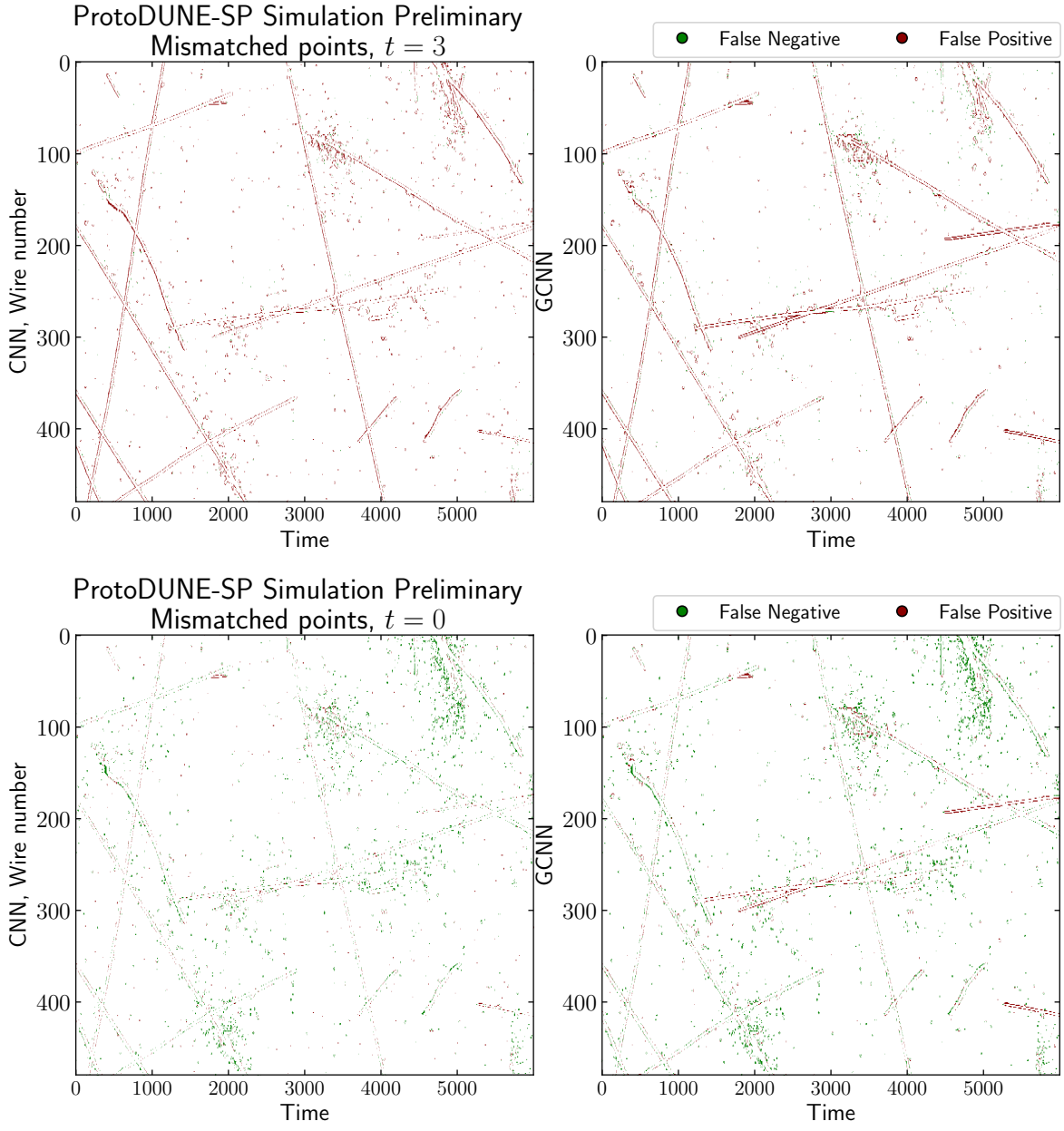




Backup - Label Cut Impact, Mismatched Points

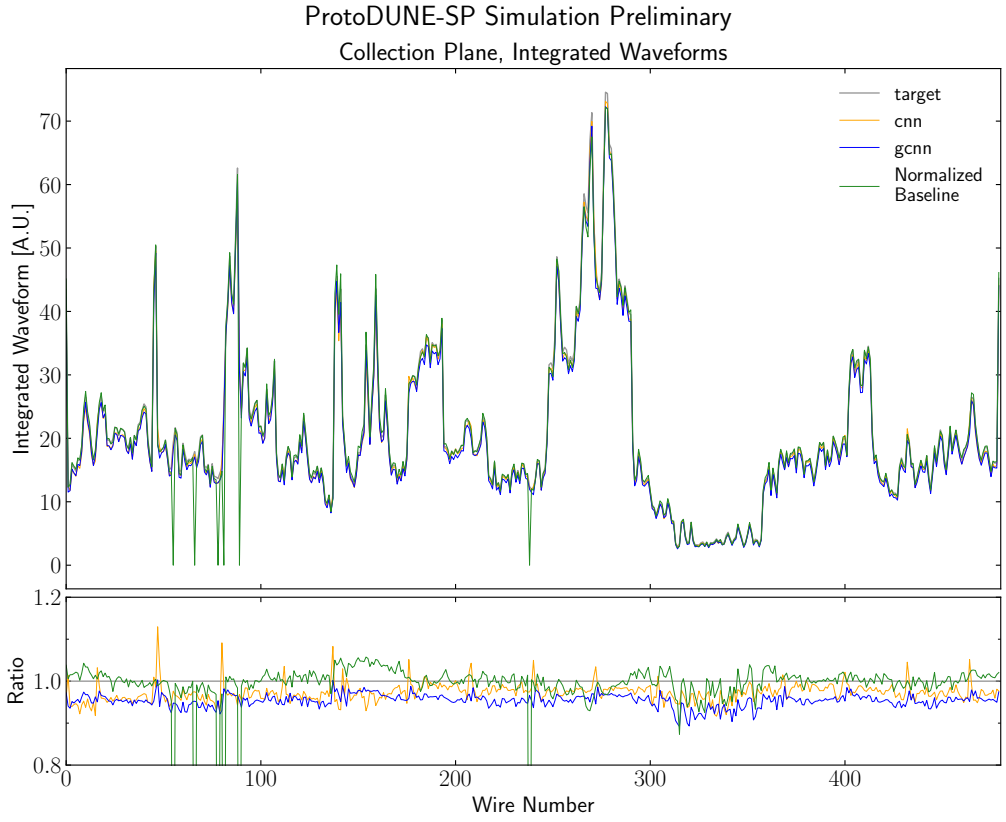
Increasing t values:

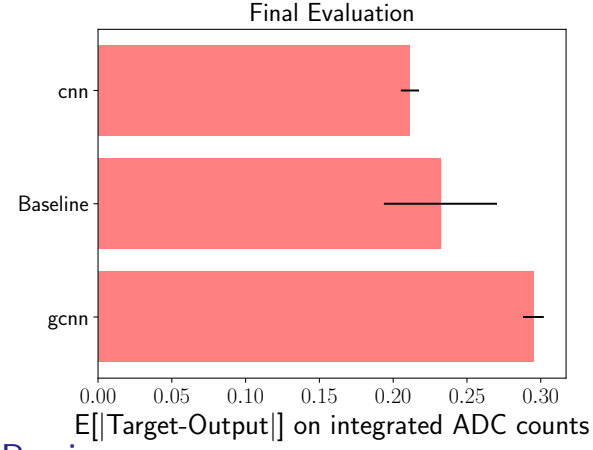
- FNs lower
- FPs increase
- ✓ FP clusters around tracks
- ! How to make tracks sharper?



Backup - Filter out bad baseline denoised channels

- ✓ NNs good profiling
- ! Baseline may fail (zero integrated ADC)
- ▶ Good ratios overall





- ▶ Remove channels with no integrated charge from baseline
- ✓ Algorithms show comparable results
- ! Baseline has large uncertainty