Hit Tagging Update

Aidan Reynolds

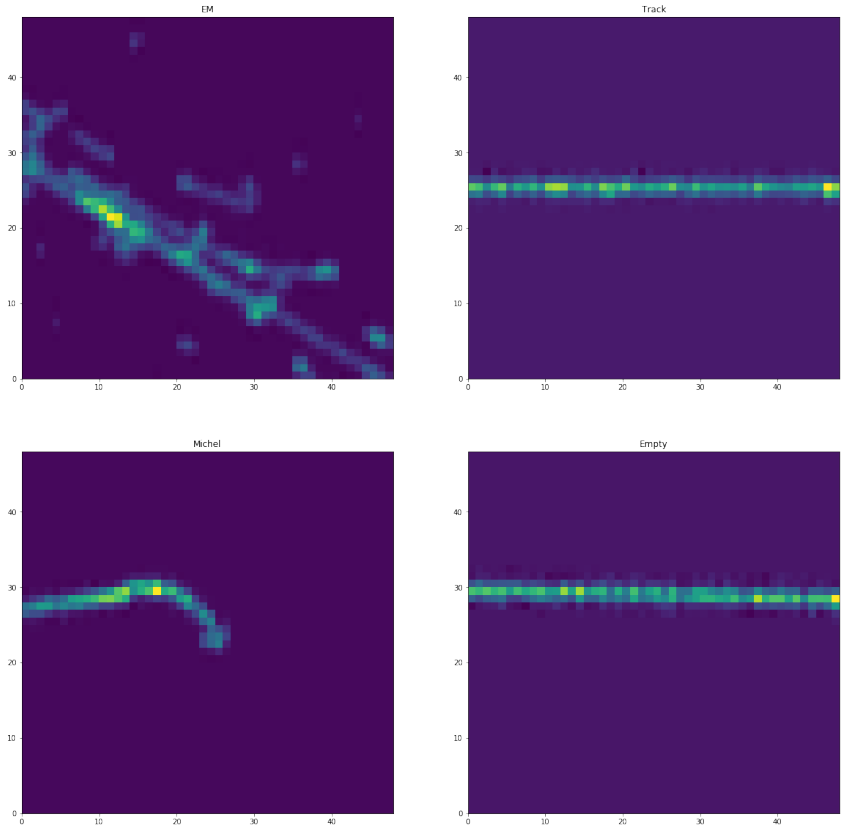
01.05.2019

Hit Tagging Goal and Motivation

- Want hit by hit level track shower separation
- Useful for generating calibration samples
  - Michel electrons
  - Delta ray removal for muon calibrations
- Potential to improve current algorithms with integration of alternative hit level PID

Implementation

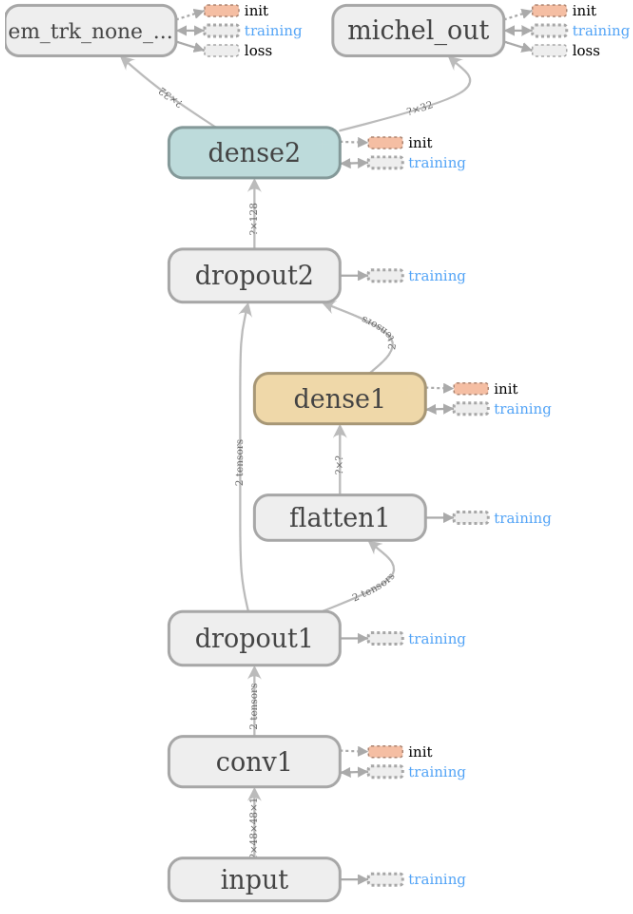
- Hit tagging is implemented with a simple CNN
- Input images are 48×48 pixels drawn from deconvoluted waveforms
- 4 types of images: EM, Track, Michel, and Empty
- Truth obtained by backtracking the true charge deposits
  - Michels are in both EM and Michel category
- Goal of the network is to identify the source of the charge at the centre of the image



Network Architecture

Final network architecture

- Single convolutional layer CNN
- Two dense layers
- In: 48×48 deconvoluted ADC
- Out: (EM, Track, Empty) + (Michel)
- Weighted loss: $L = 0.1 \cdot L_{ts} + L_m$
- Dropout for regularisation



Training Samples

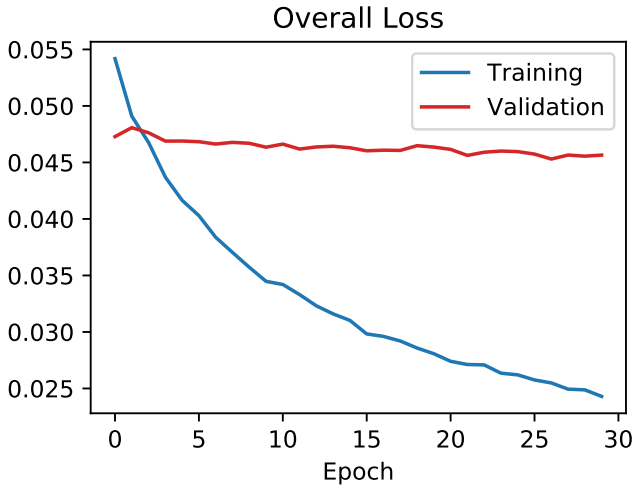
- Training data was built using MCC11 simulations
  - SCE on
  - Fluid flow on
  - All beam energies in sample
- Data samples split into training, test, and validation

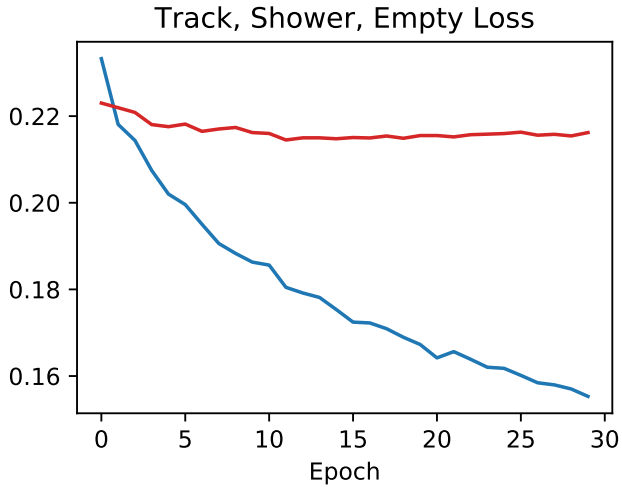
Patch Type	EM	Track	Empty	Michel
Training	13,493,982	9,727,604	2,517,882	731,456
Validation	734,673	562,038	141,388	42,727
Test	764,659	518,805	139,987	39,674

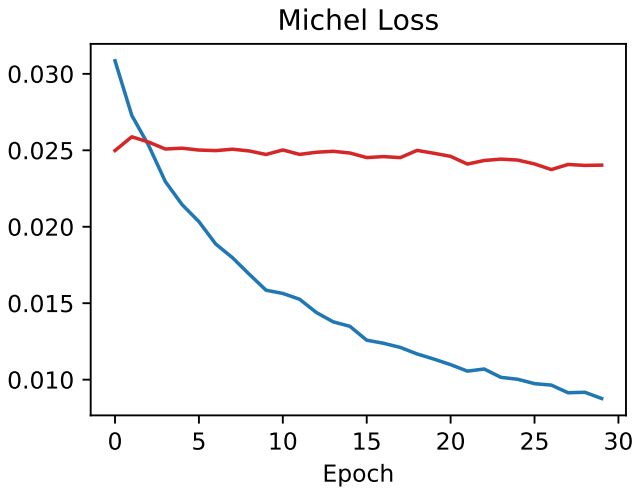
Monitoring of Training

Training monitored with Tensorboard

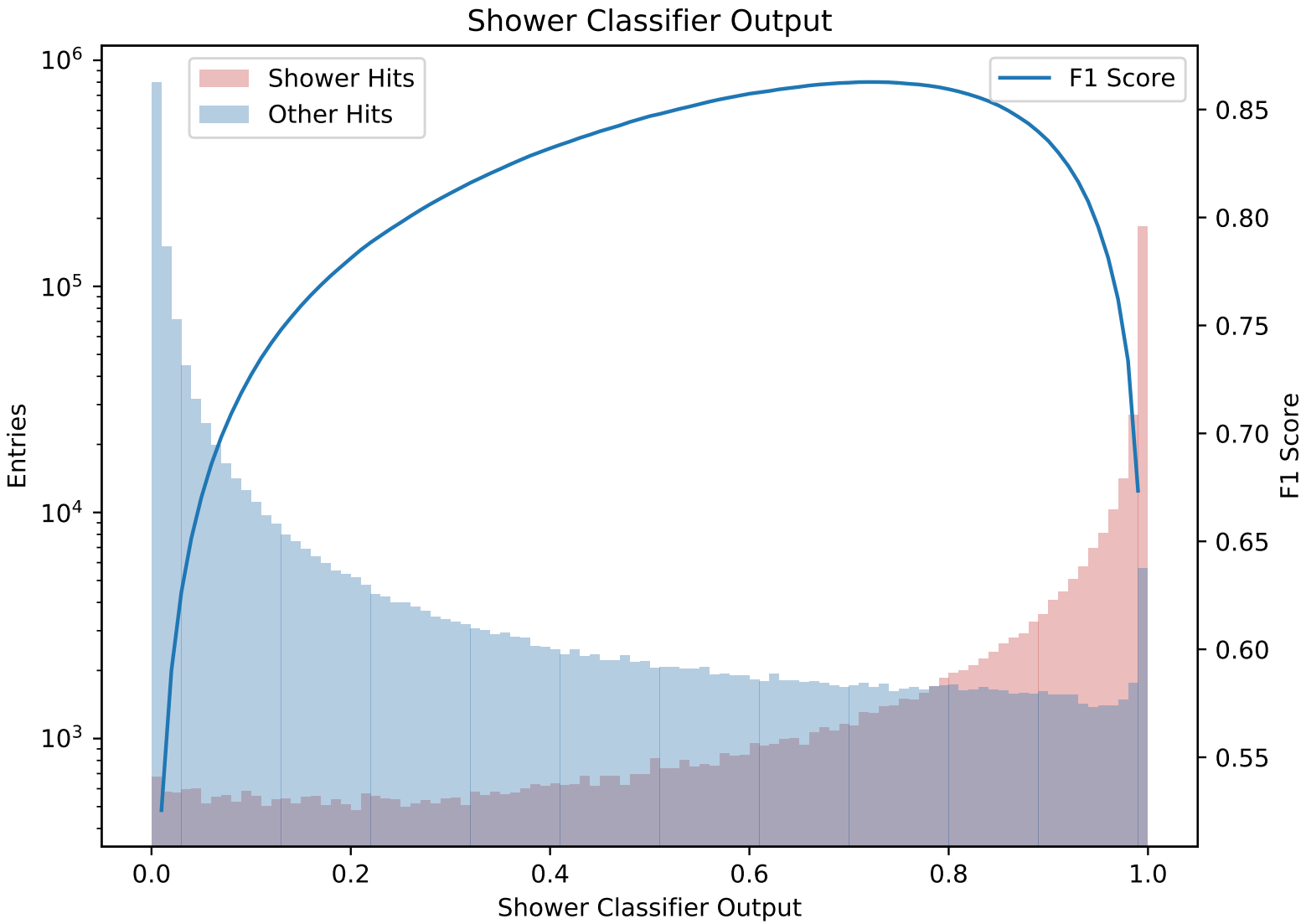
- Learned fast with negligible validation improvement after 1st epoch
- Dropout successful in preventing over–training



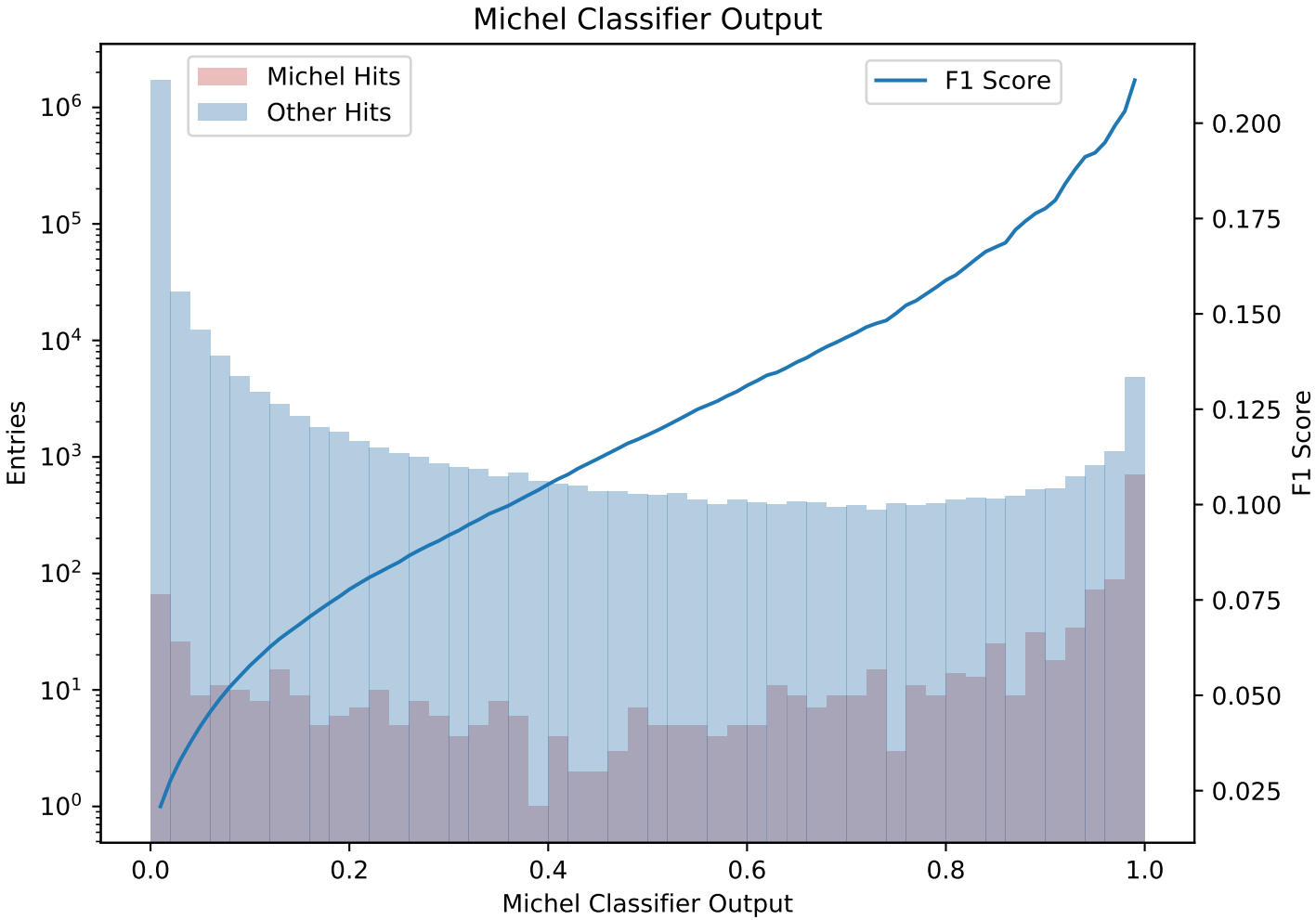




Performance Tests MC

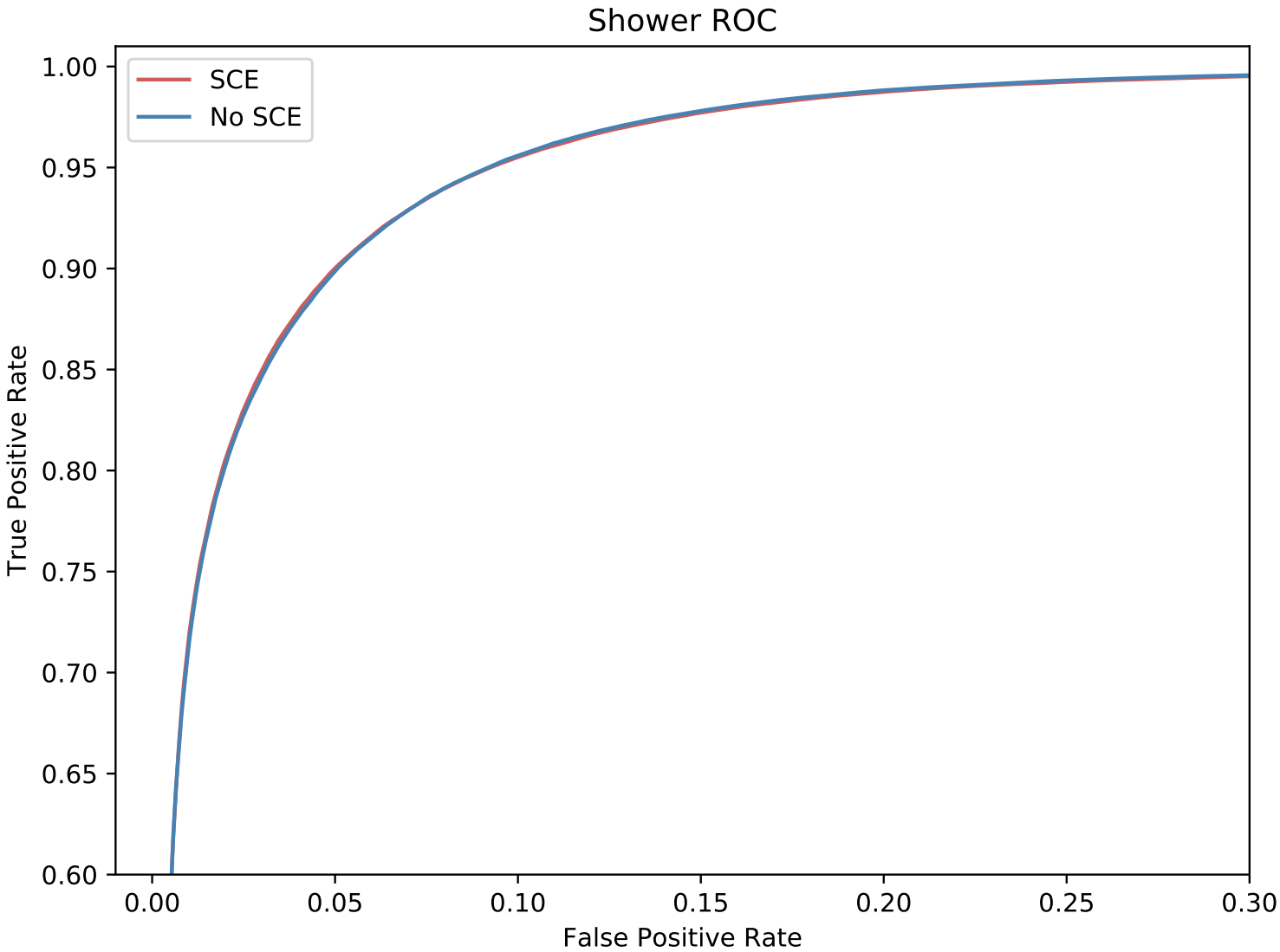


Performance Tests MC

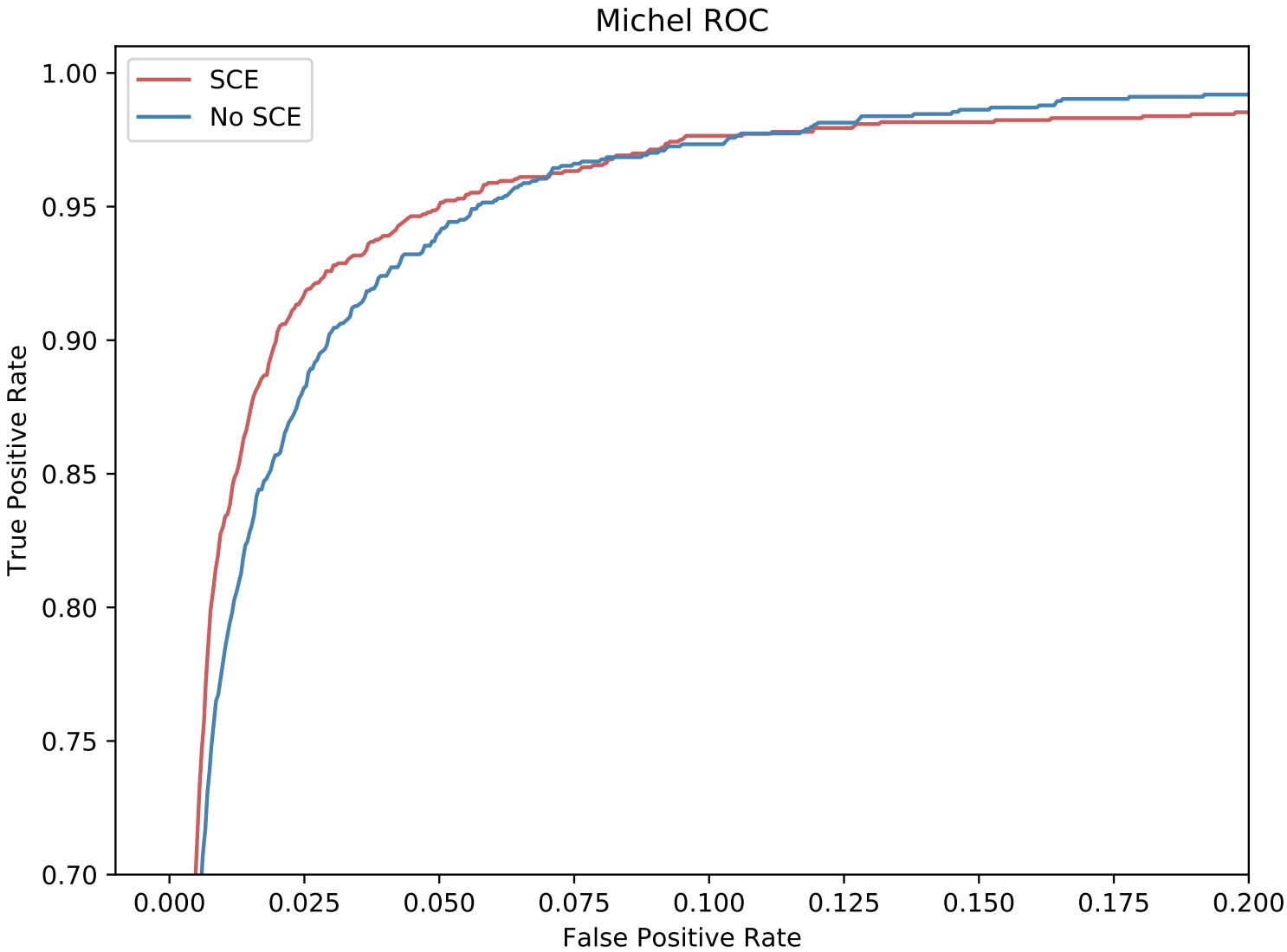


Basic clustering helps a lot here

Performance Tests MC



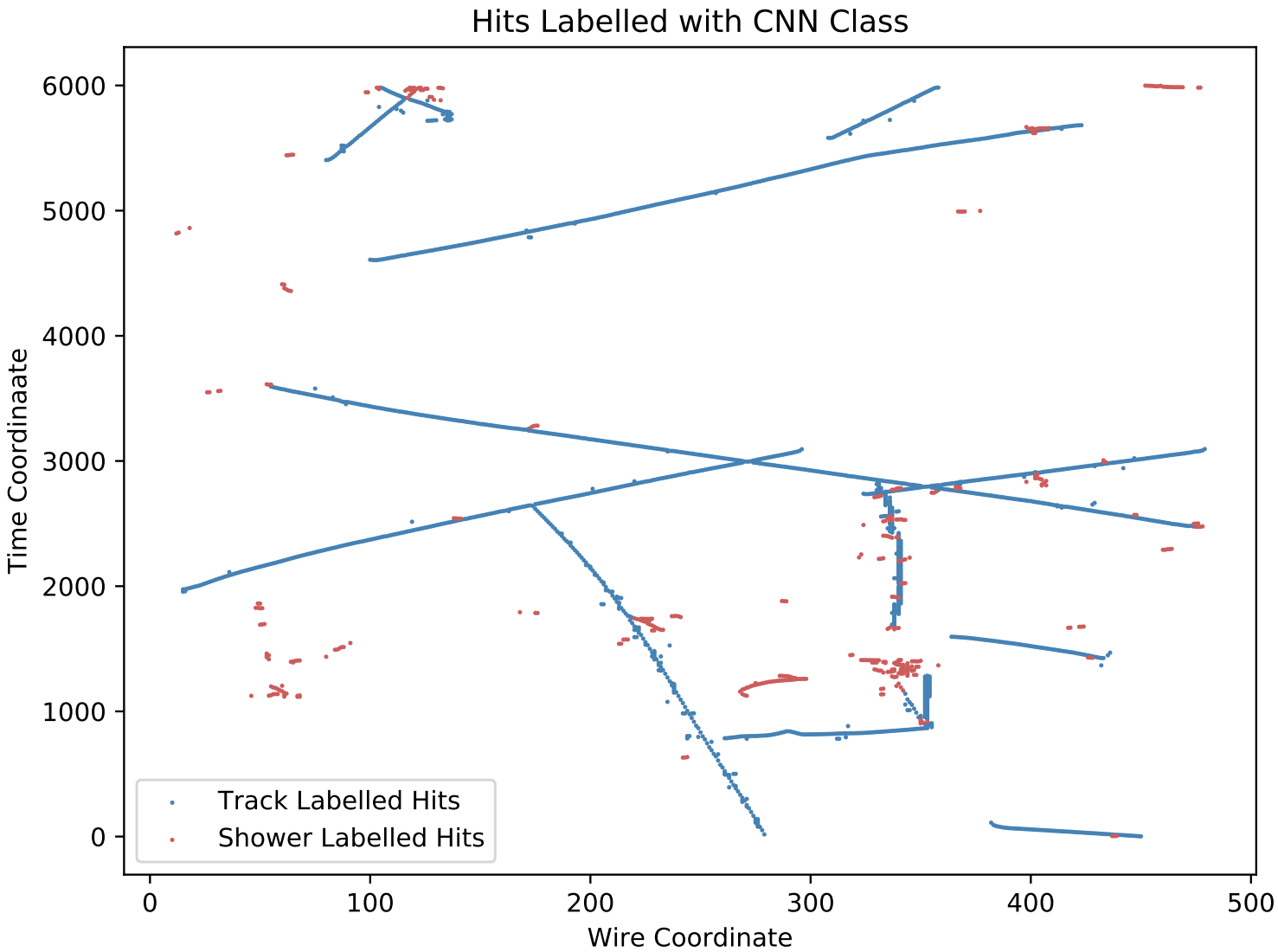
Performance Tests MC



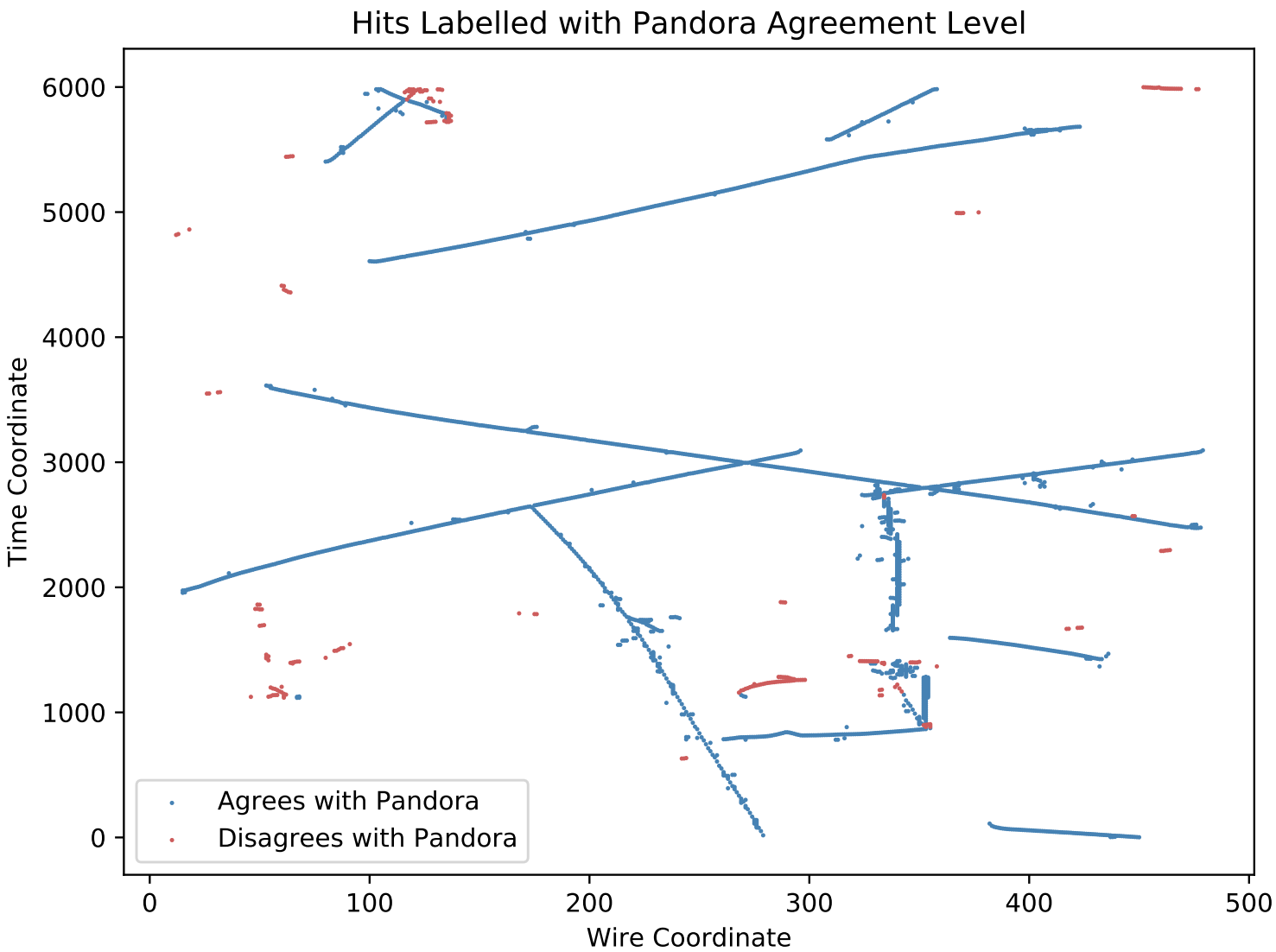
Performance Tests Data

- 2D deconvoluted data used for tests
  - Run number 5387
- Tests were done by comparing with the output of Pandora
  - Cross validation of algorithms
  - Makes interpretation a little more difficult due to effects for both algorithms
- Looked at CNN score for hits in Pandora Tracks and Pandora Showers

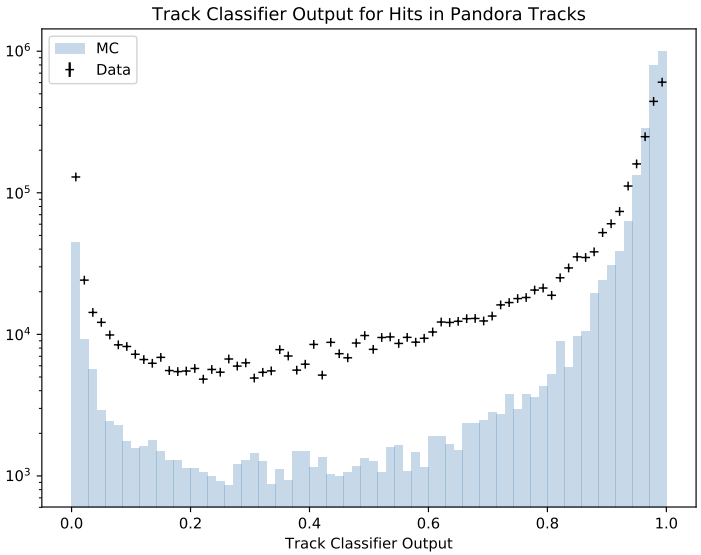
Performance Tests Data

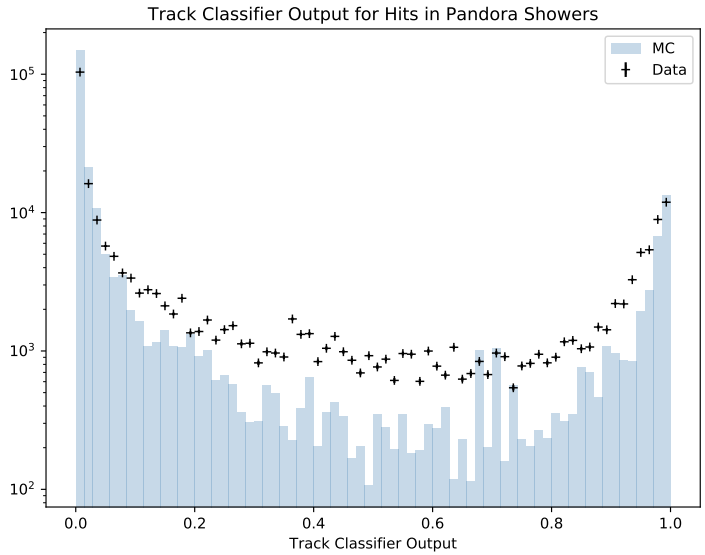


Performance Tests Data



Performance Tests Data

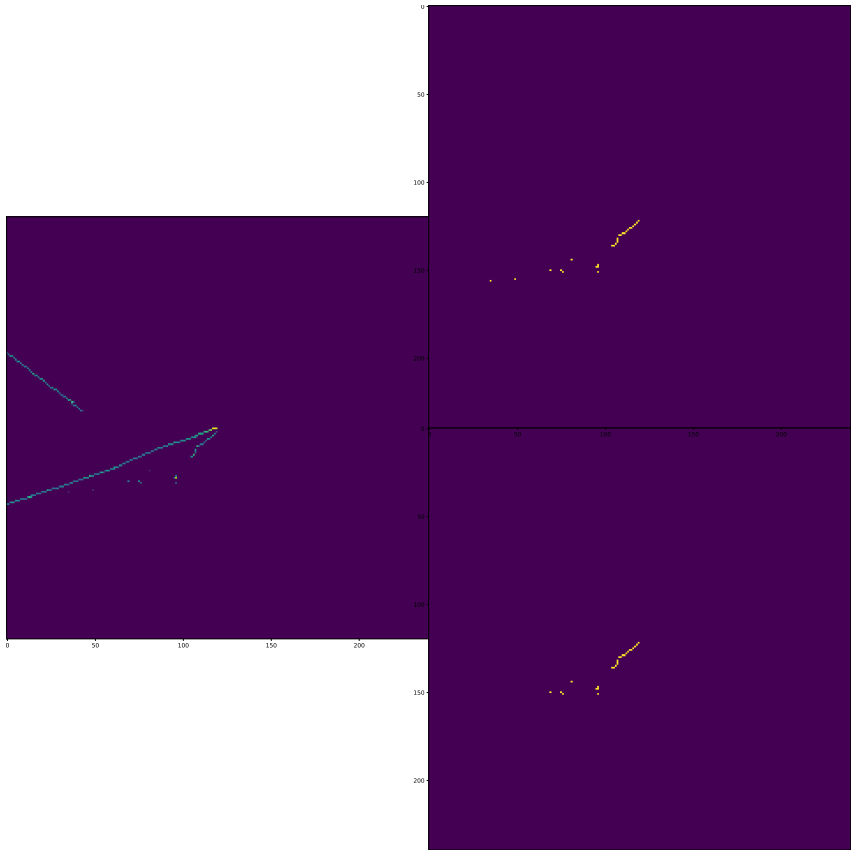




Ideas for Improvement

CNN slow on CPU's, $\sim$ 40s / event

- Would like to use deeper networks but time is prohibitive
- Often redundancy with neighbouring network evaluations
- Slow running due to many image creations and network evaluations
  - Semantic segmentation could be faster
    - Evaluate many hits at once
    - Fewer images and evaluations



Hit Tagging Use Cases

Some potential uses for labelled hits

- Michel electron reconstruction (In progress)
- Defining clean calibration samples, e.g. removing delta hits for muon samples
- Integration into current algorithms to aid reconstruction