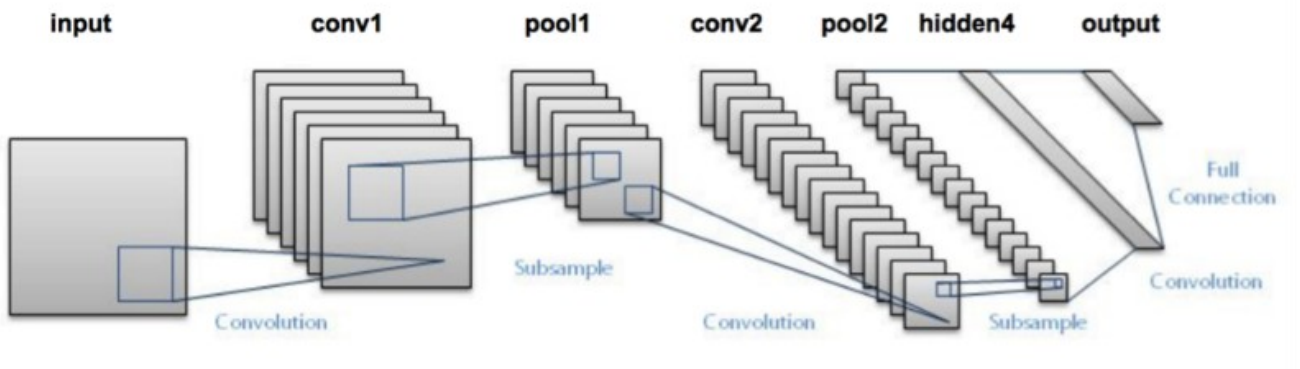
Update on convolutional neural network for ep event selection

- Development finished, framework take training data from ROOT files, and same with make predictions
- Training data: for now only train ep and non-ep events (cosmic, Moller ...), output layer has two nodes, easy to add ee events, cosmic events, need to collect reliable training data
- e-p event selection: Hycal cuts with GEM match



Convolutional neural network identify ep events

Event selection:

- HyCal cut: HyCal cluster size: 5-30, energy cut: 4 sigma
- With GEM match (either GEM1 or GEM2)
- If event pass HyCal cut, but no GEM match found, these event will not participate in training

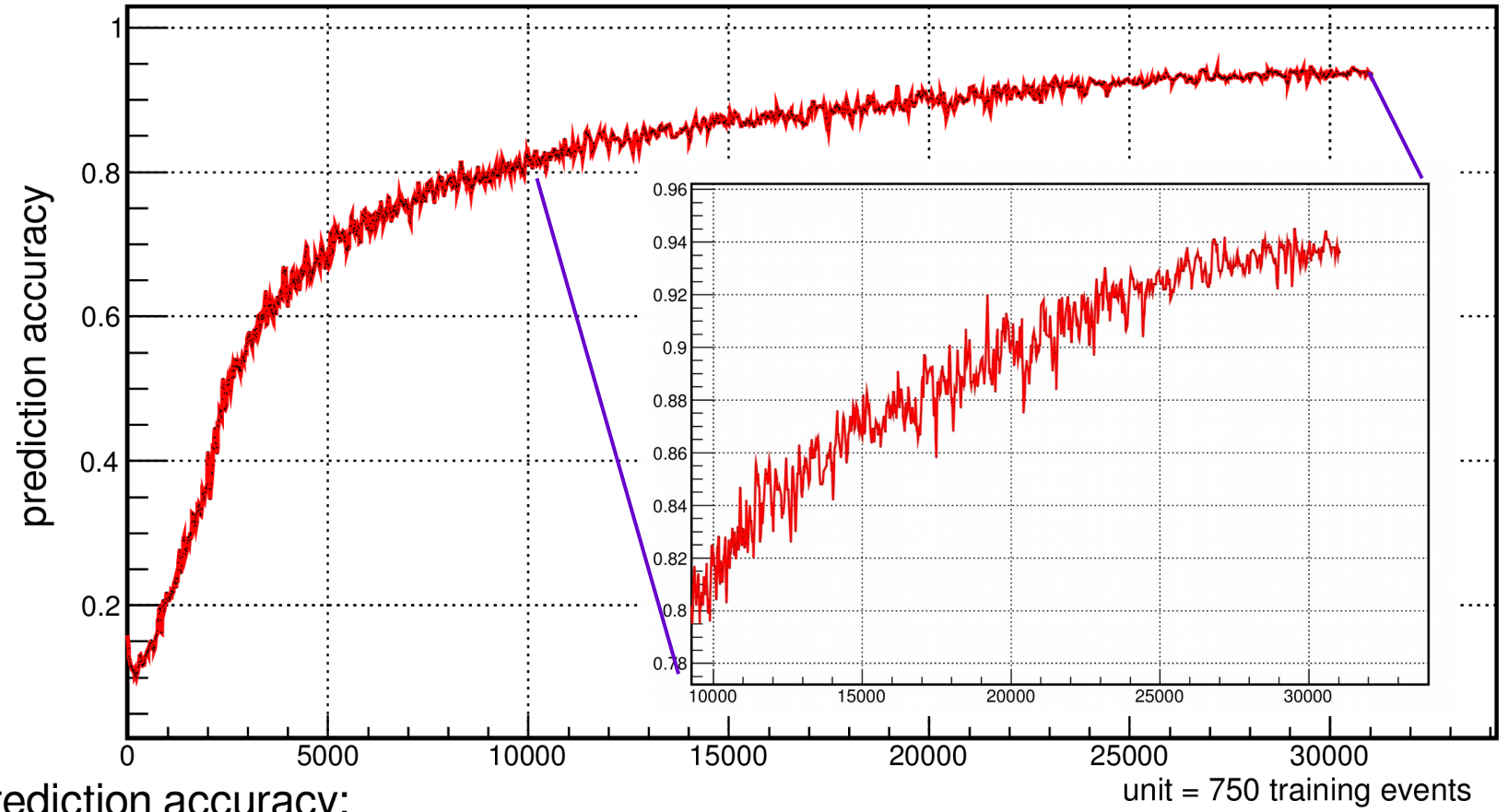
Model test:

- Used two datasets for training, simulation data and experiment data
- Prepare a dataset with pure known ep events (Cut on HyCal, require GEM match), both simulation and experiment
- After training, use trained NN to test the known dataset
- Prediction accuracy (using the prepared dataset):

$$accuracy = \frac{\text{number of events identified as } e-p \text{ by NN}}{\text{total number of } e-p \text{ events}}$$

Convolutional neural network identify ep events

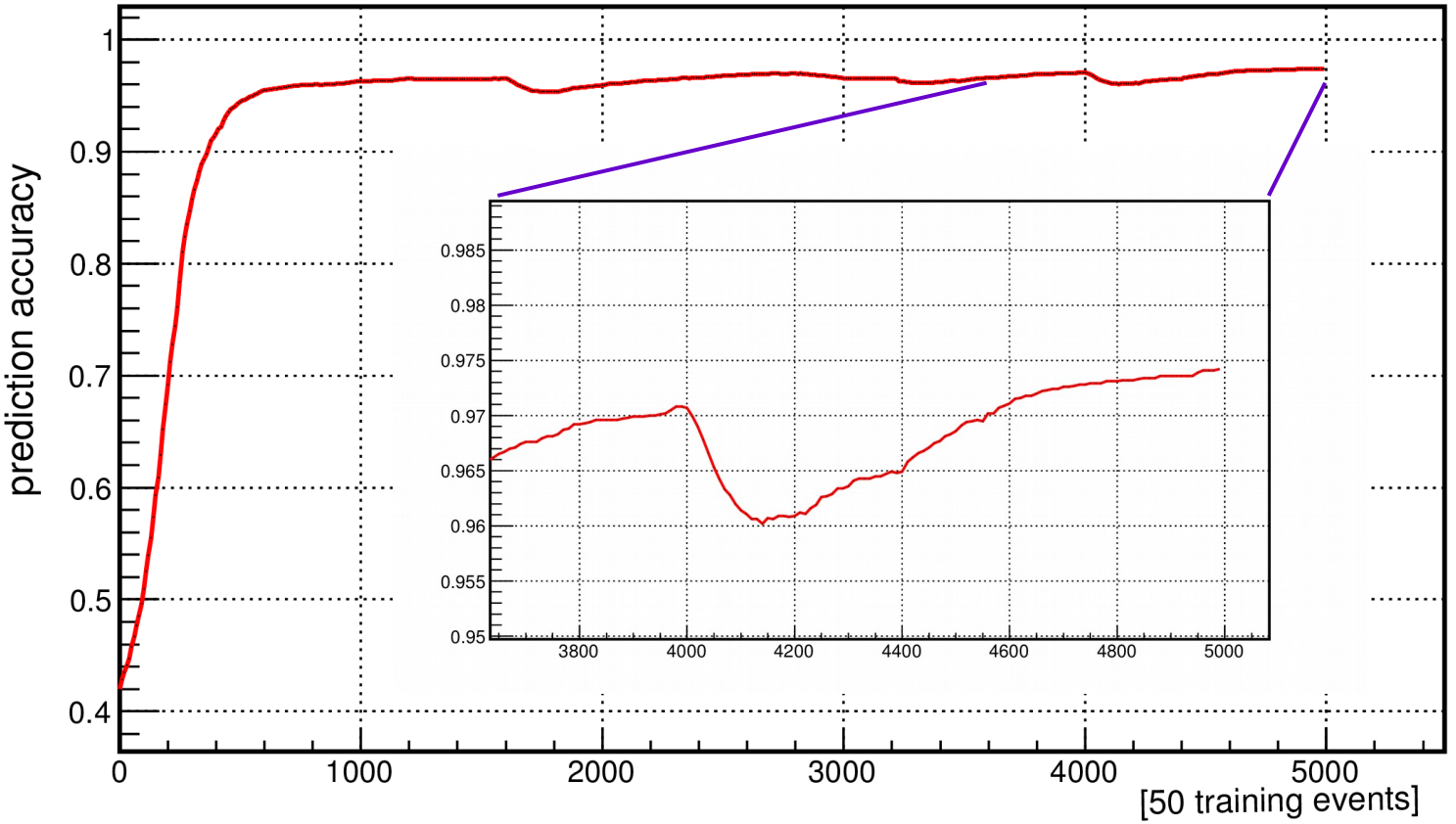
Use trained NN predicting experiment data



- Prediction accuracy:
  - 22M experiment events
  - Use nn see whether identify them correctly
- More training data, higher accuracy
- Seems that if feed more data to training, will reach higher accuracy
- Difficult to get training data clean

Convolutional neural network identify ep events

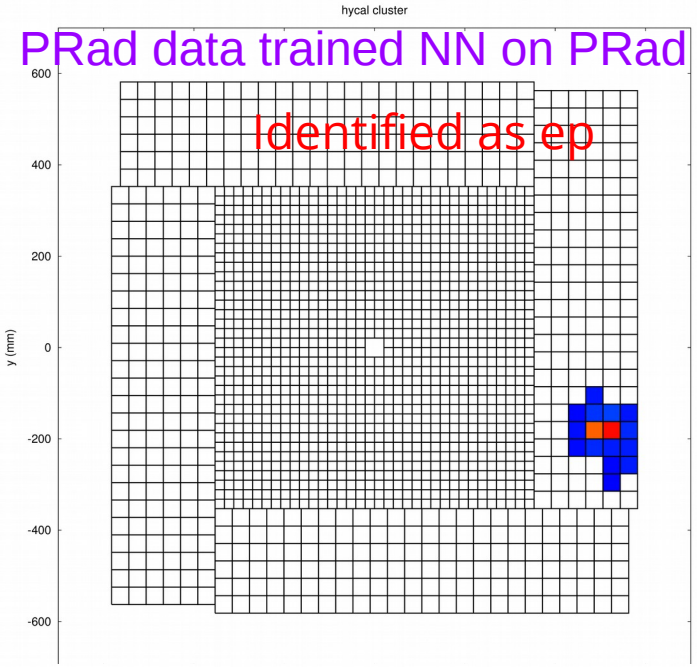
Use trained NN predicting simulation data

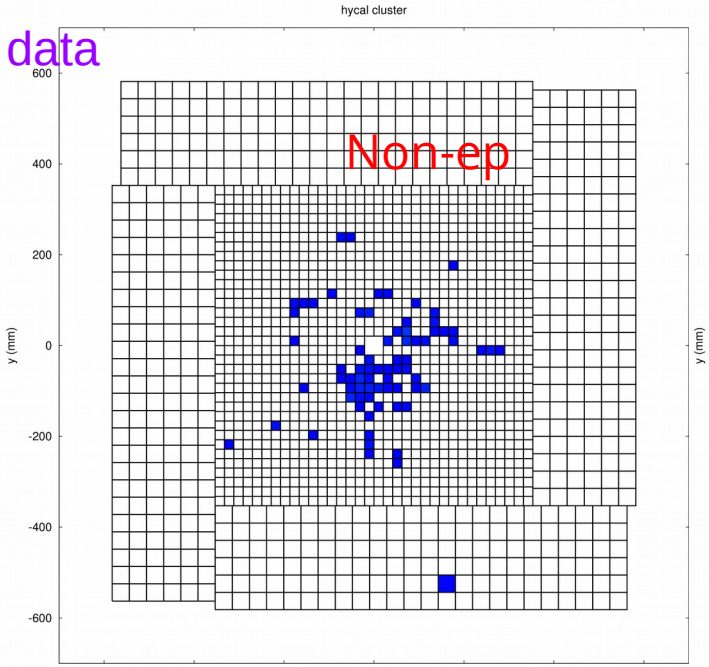


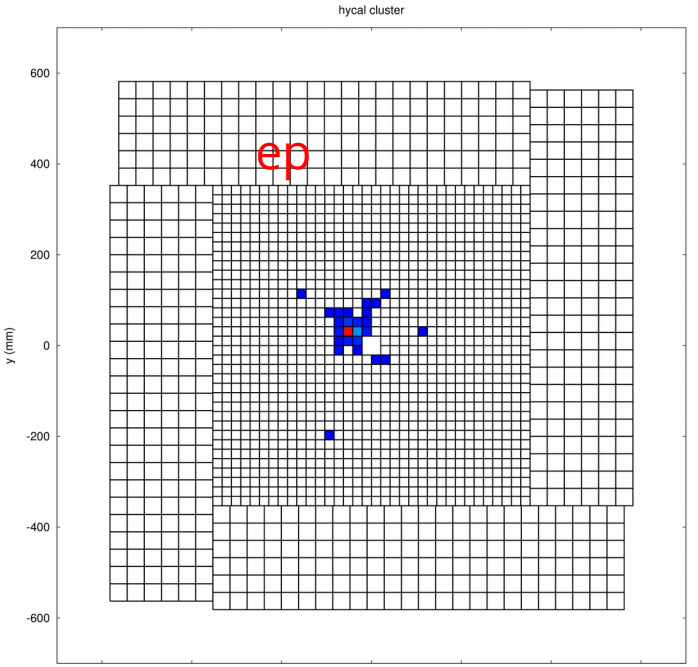
- Prediction accuracy:
  - 10k simulated ep events
  - Use nn see whether identify them correctly
- More training data, higher accuracy
- Can go higher accuracy with more training data

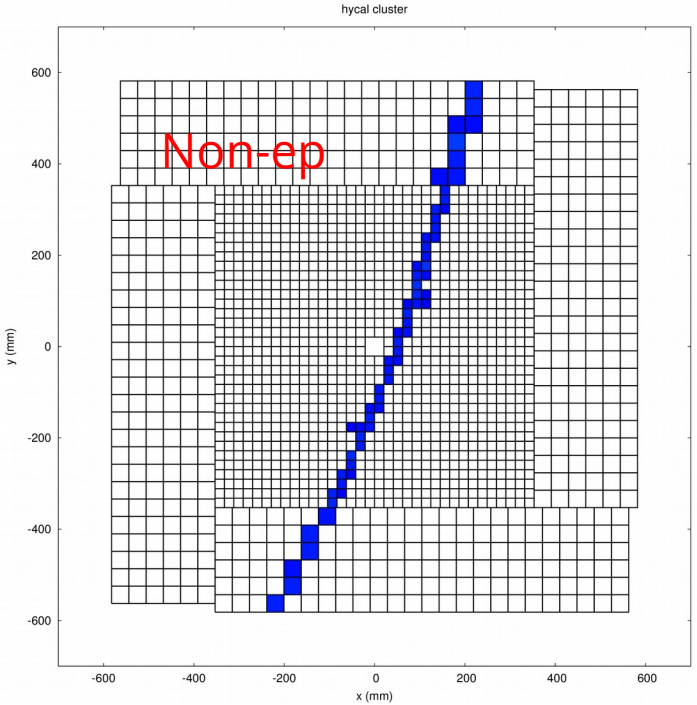
Convolutional neural network identify ep events

PRad data trained NN on PRad data

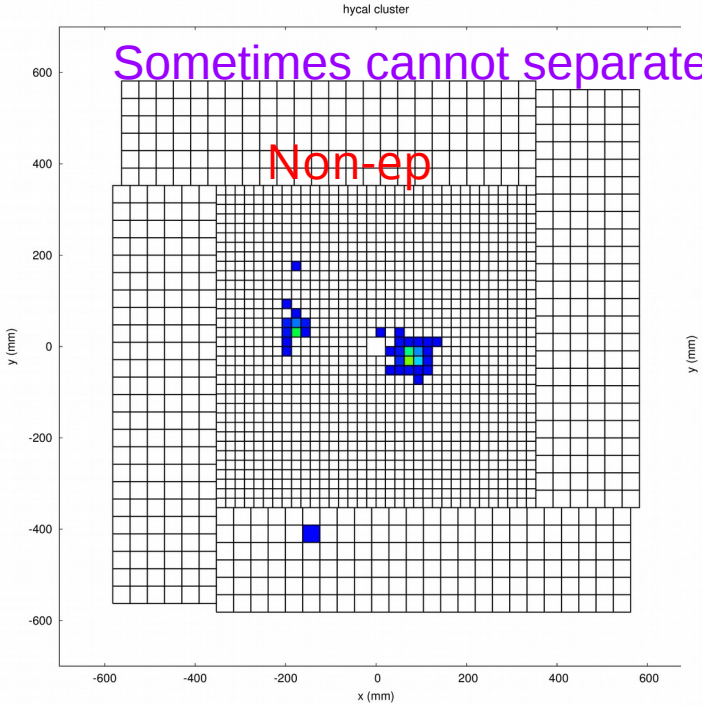


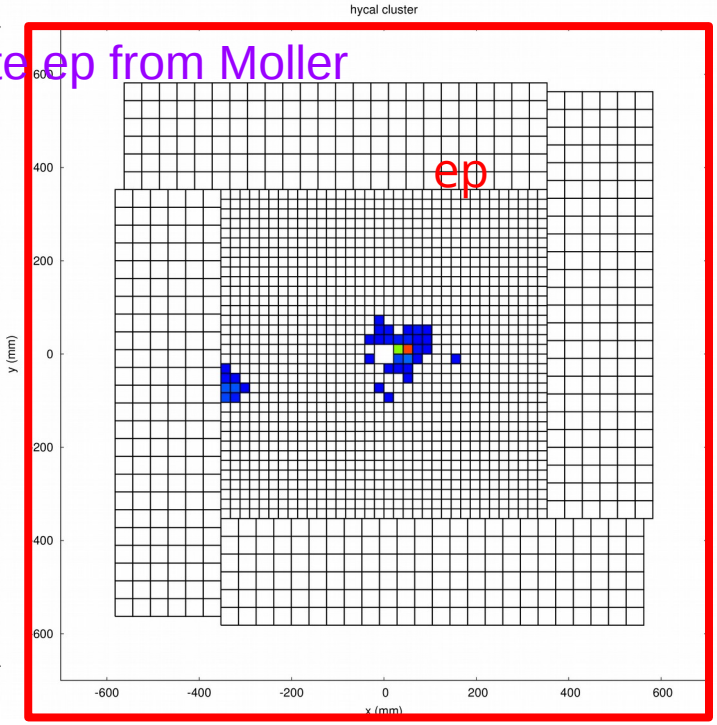






Sometimes cannot separate ep from Moller





Convolutional neural network identify ep events

Results:

	Simulated ep events (20k)	Simulated cosmic events (20k)	Experiment ep events (10k)	Experiment cosmic events (10k)
Experimental data trained NN	89.06%	99.98%	93.94%	99.67%
Simulation data trained NN	97.96%	97.24%	99.28%	95.7%

- Table shows how many events were correctly identified (in percentage)
- For cosmic events from experiment, require # of fired hycal modules > 4
- # of experimental data used in training: 22M
- # of simulation data used in training: 20k
- Experiment training data not clean, contaminated by Moller events, the low number probably due to that some ep events were identified as Moller events
- Angle independent