Hyper-Kamiokande Event Reconstruction Using Machine Learning Technique

10 November 2021

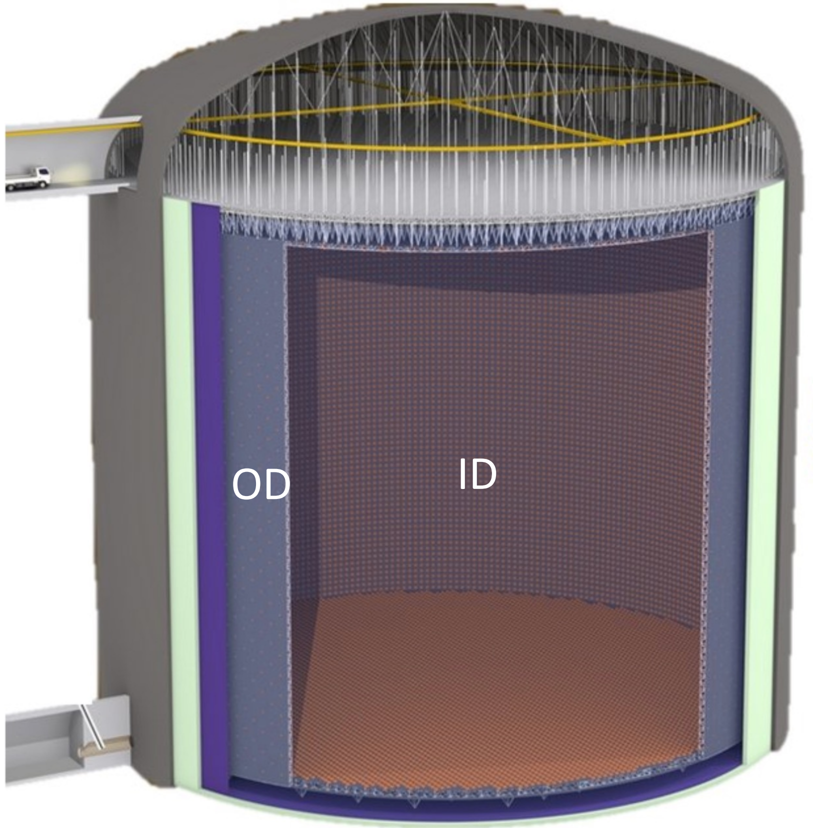
Joanna Gao

Content

- Introduction to Hyper-Kamiokande
- FiTQun, a traditional reconstruction tool
- PointNet, a machine learning technique
- PointNet for IWCD particle identification
- Data processing procedure
- Current result
- Future plans

Hyper-Kamiokande

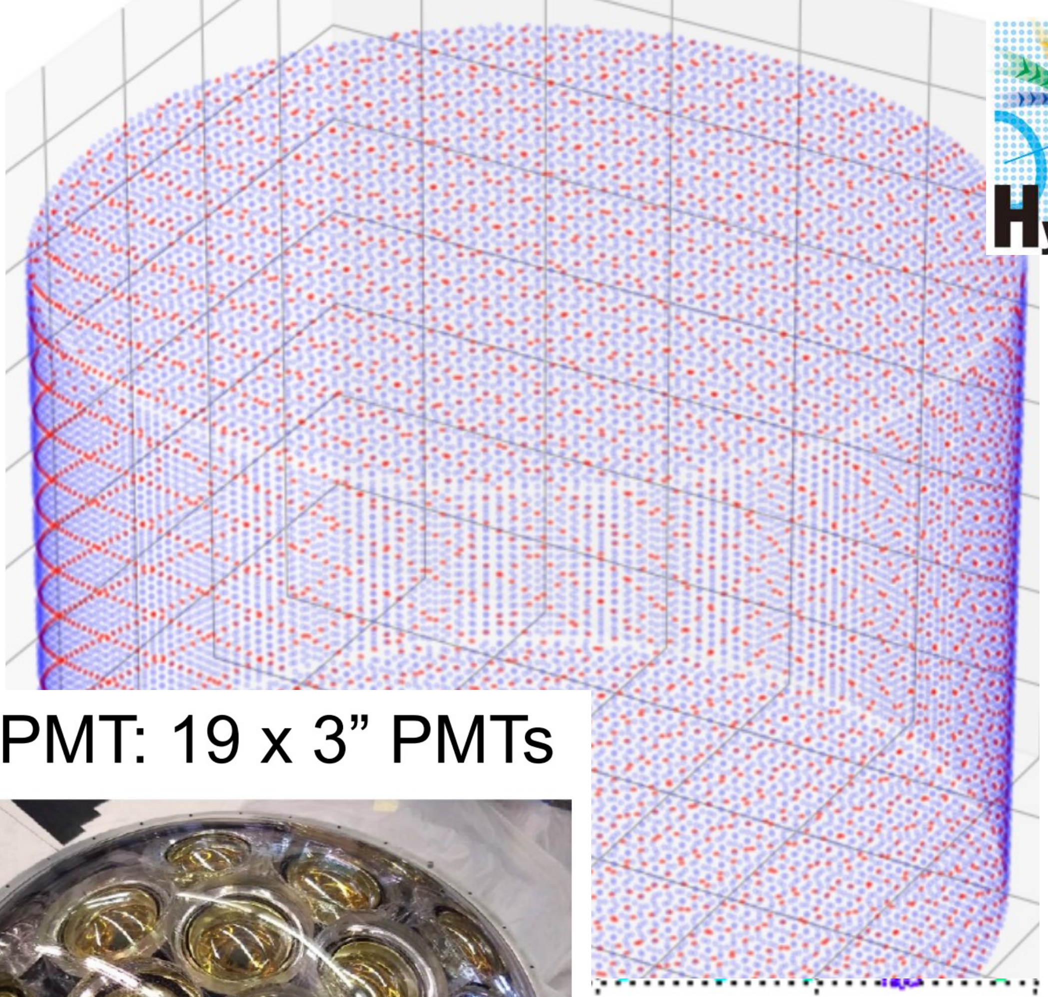
- Water Cherekov detector, the successor of Super-Kamiokande
- Located in the mountain in mid-west Japan, and it's 68 m in diameter and 71 m in height (3/4 of the Big Ben)
- 8 times the fiducial volume of Super-K



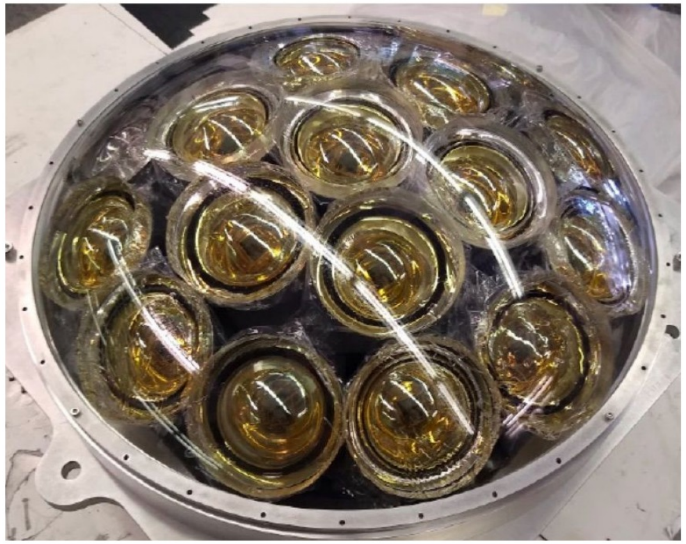
KING'S College LONDON

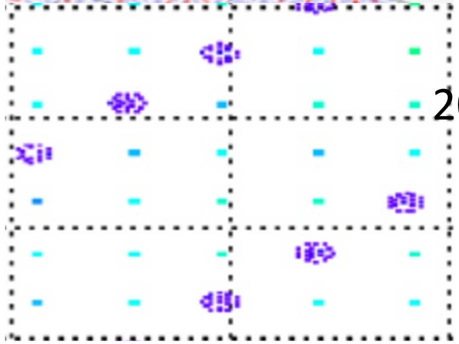
Hyper-Kamiokande

- It plans to have 20,000 20-inch PMTs and ~2,000 mPMTs (to improve granularity) in the inner detector (ID)
  and ~6,000 3-inch PMTs in the outer detector (OD) veto region
- Simulation on the right has ~19,000 20-inch PMTs and ~5,000 mPMTs



mPMT: 19 x 3” PMTs



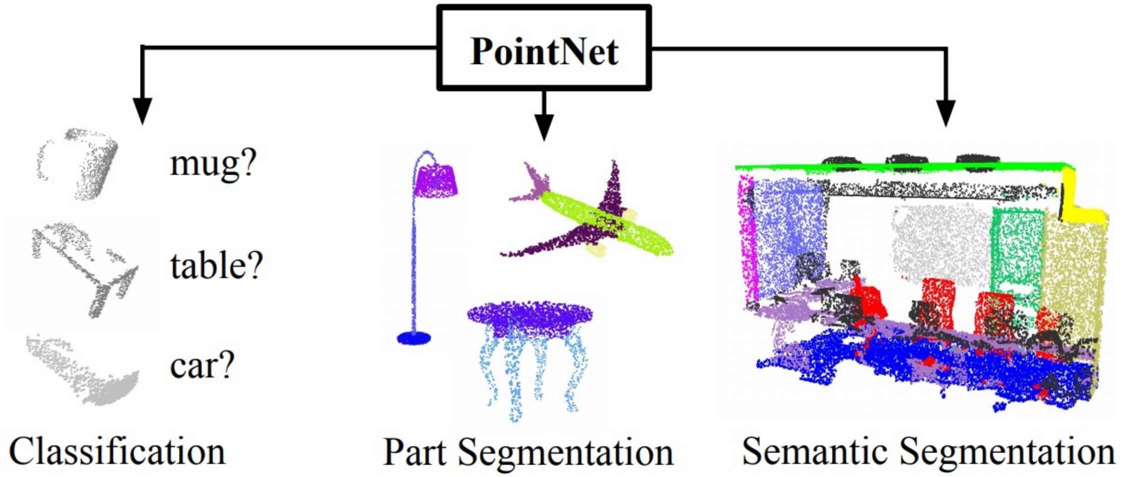


FiTQun, a Traditional Reconstruction Tool

- Likelihood based, developed by Super-K and migrated to Hyper-K
- Good at energy, direction, vertex location reconstruction, also performs good e- and pi0 separation
- Not so good at e- and gamma separation as the signals from the two are quite similar
- Very slow, ~ min per event

PointNet, a Machine Learning Technique

- Unlike traditional Convolutional Neural Network (CNN), which unwraps a 3D data into a 2D image (e.g. ResNet), PointNet is a 3D classification and segmentation tool that accept 'point cloud'
- Advantage:
  - retaining location, timing and charge relation between hit PMTs;
  - can apply to any detector size and geometry
  - Quick to use after training

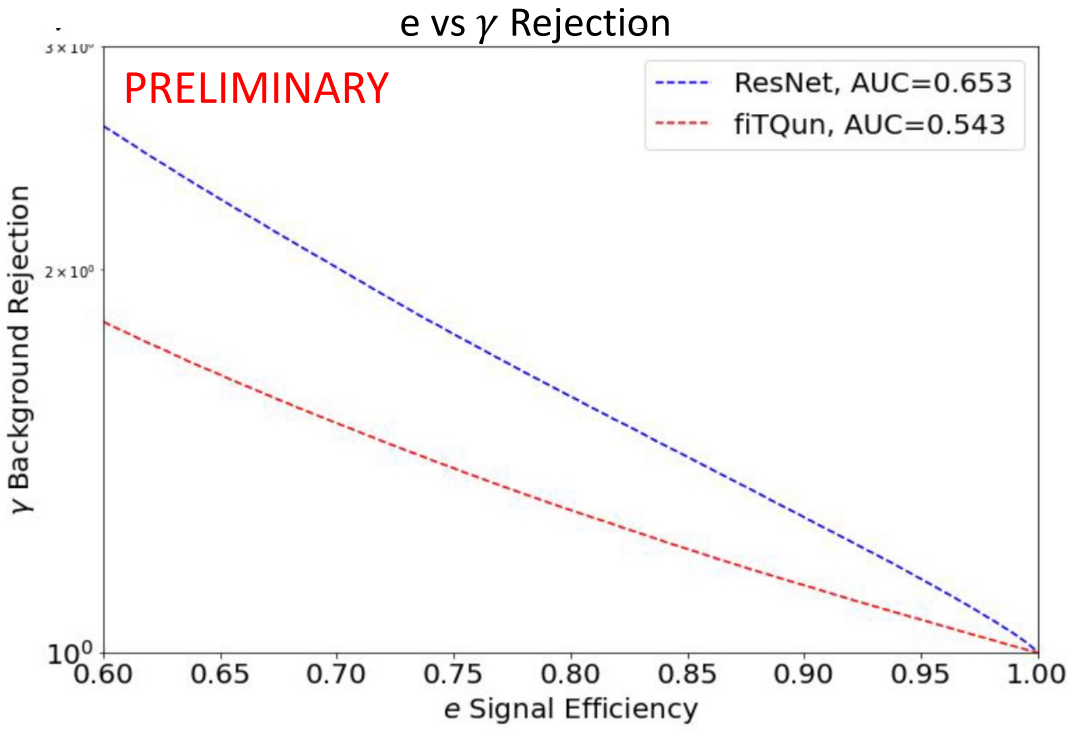


[1]

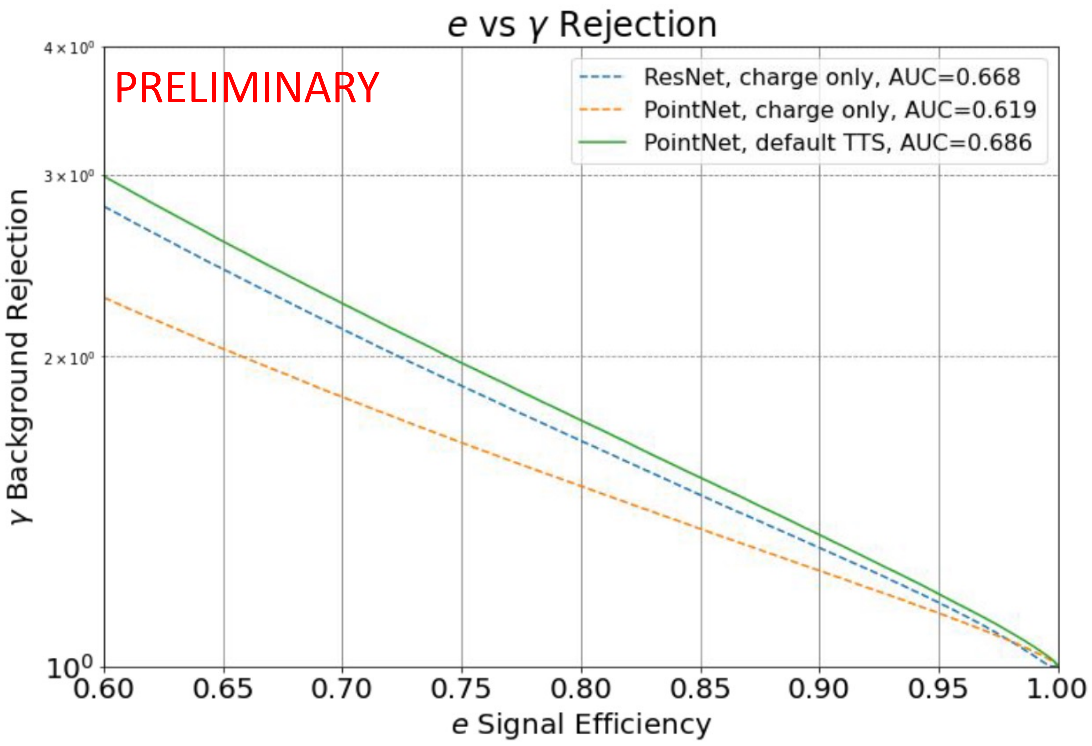
Fig. 1 Diagrams illustrate the input dataset for PointNet

PointNet for IWCD PID

- IWCD explored the efficiency of performing e/γ discrimination with fiTQun, ResNet and PointNet
- Figure of merit: AUC, [0, 1], the larger the better
- ResNet better than fiTQun
- PointNet better than ResNet when using both <u>charge</u> and <u>timing</u> information (default TTS curve in bottom right)



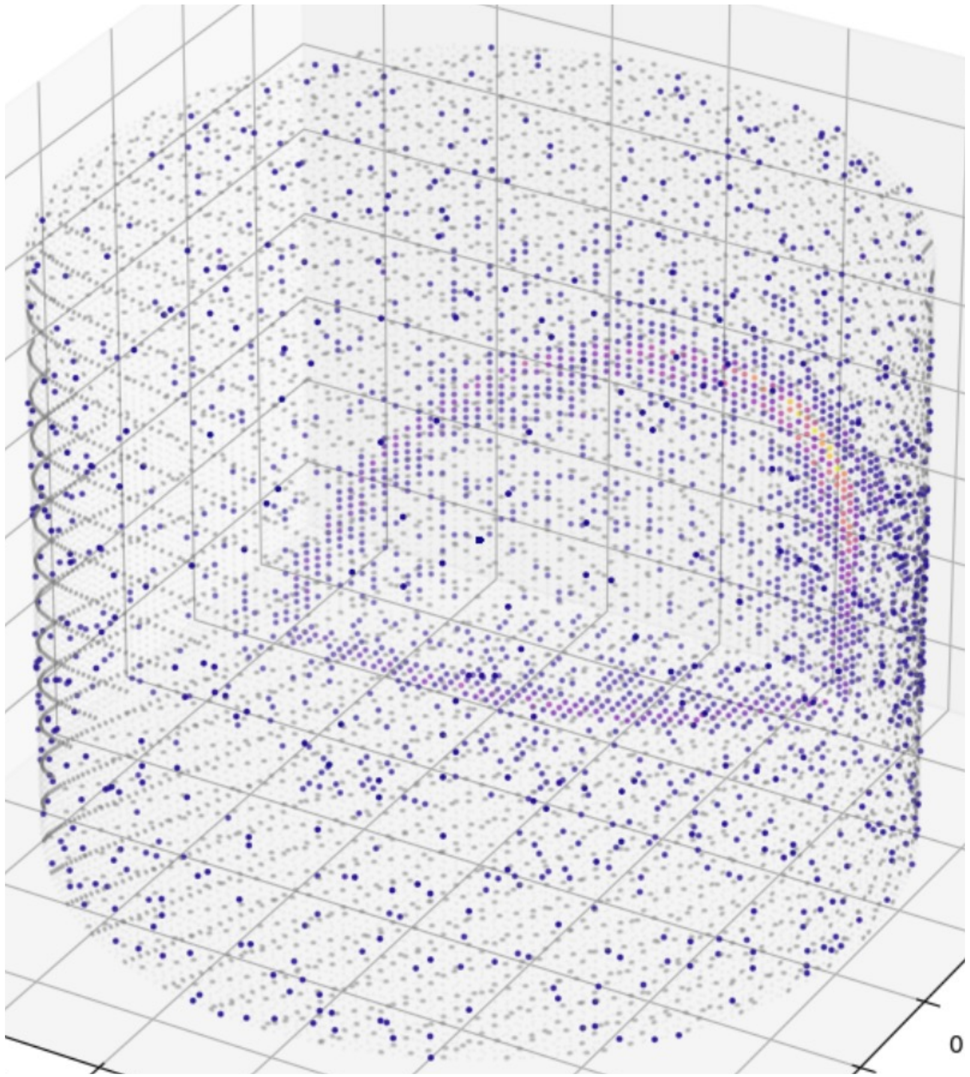
[2]



[3]

Reconstructing Events

- Currently looking at e- and mu- from neutrino CC interactions (simulated events)
- Performing classification using PointNet and compare the performance to FiTQun



Input Data Processing Procedure for PointNet

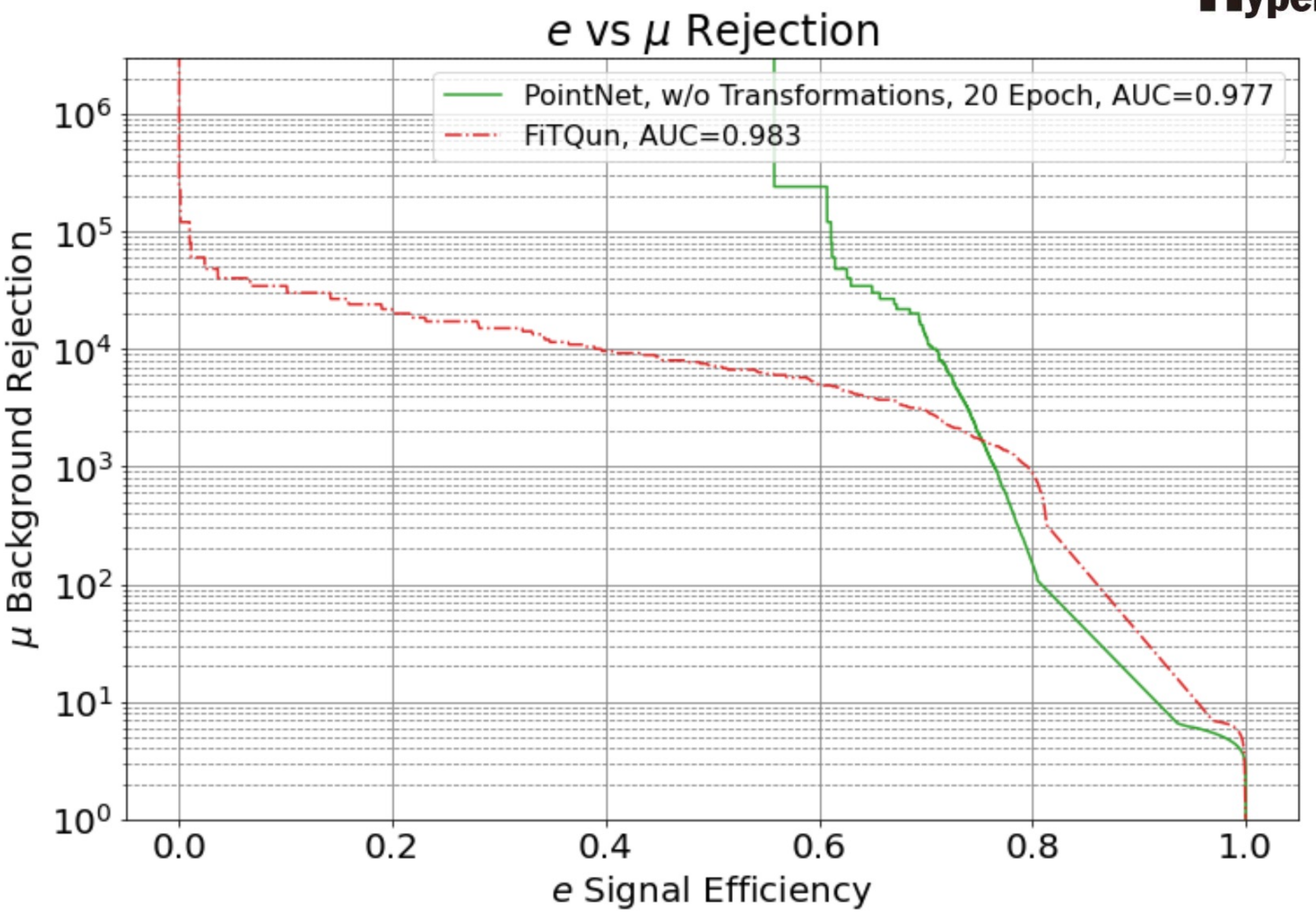
Simulation Output Data → npz files* → 1 hdf5 file → PointNet

Speed of processing:

- For 6M events, takes ~6 days from simulation output to PointNet results, ~0.01 minute per event
- FitQun processes straight from simulation output, few minutes per event

Analysis of the ML and FiTQun Results

- Currently the classification using FiTQun negative log likelihood value performs better than PointNet
- Not ideal but the difference is small
- Could be due to a simulation bug, currently processing new data without bug



Next Step

- Compare the performance of e-/gamma separation using PointNet and FiTQun
- Incorporating particle energy/direction and vertex location reconstruction into the PointNet analysis
- Ultimately move on to reconstructing high energy (TeV scale) events

Backup Slides

WatChMaL

WatChMaL: cross-collaboration group formed to explore ML for WC

Common challenges for ML with WC detectors

- Cylindrical geometry
- High-resolution, sparse data

Many physics goals

- Maximise precision of new detectors
- Reconstruct complex event topologies
- Discriminate electron and gamma rings
- Improving detector calibration & systematics

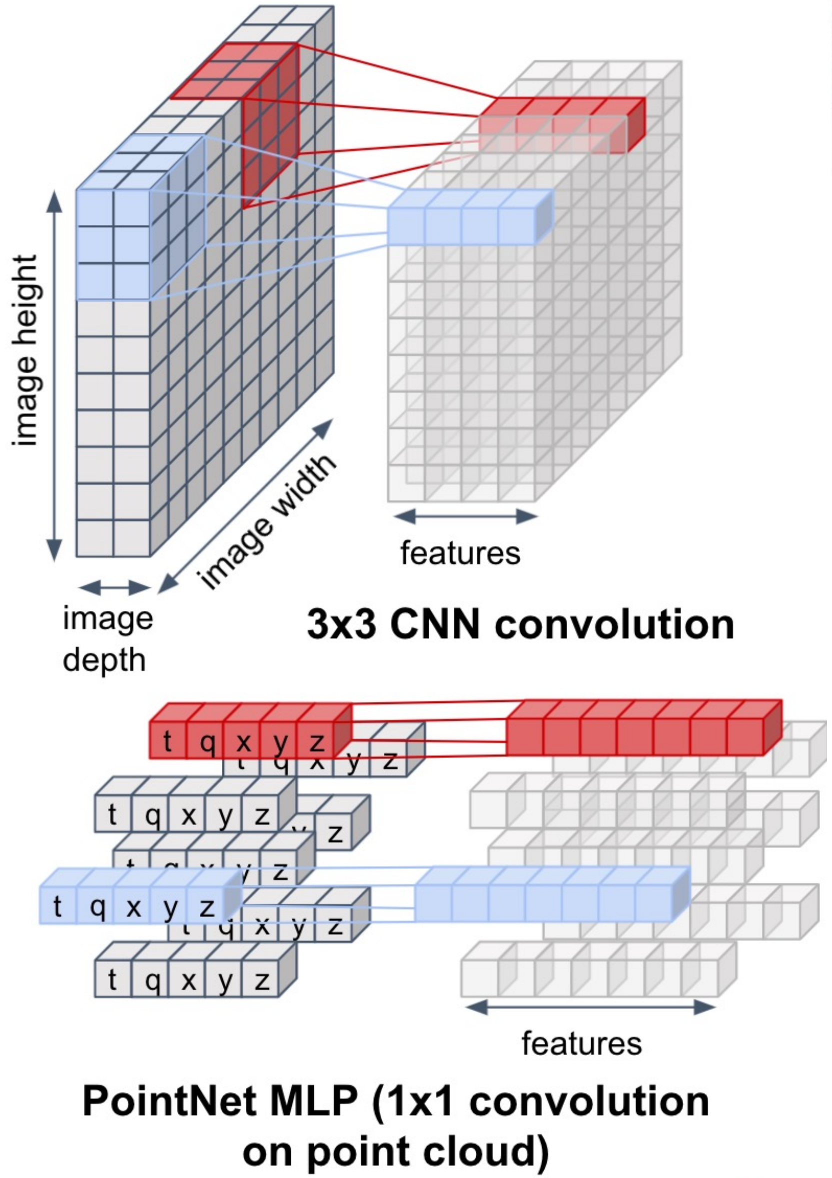
WatChMaL.org

[2]

PointNet

PointNet is designed to work on 'point clouds' rather than images

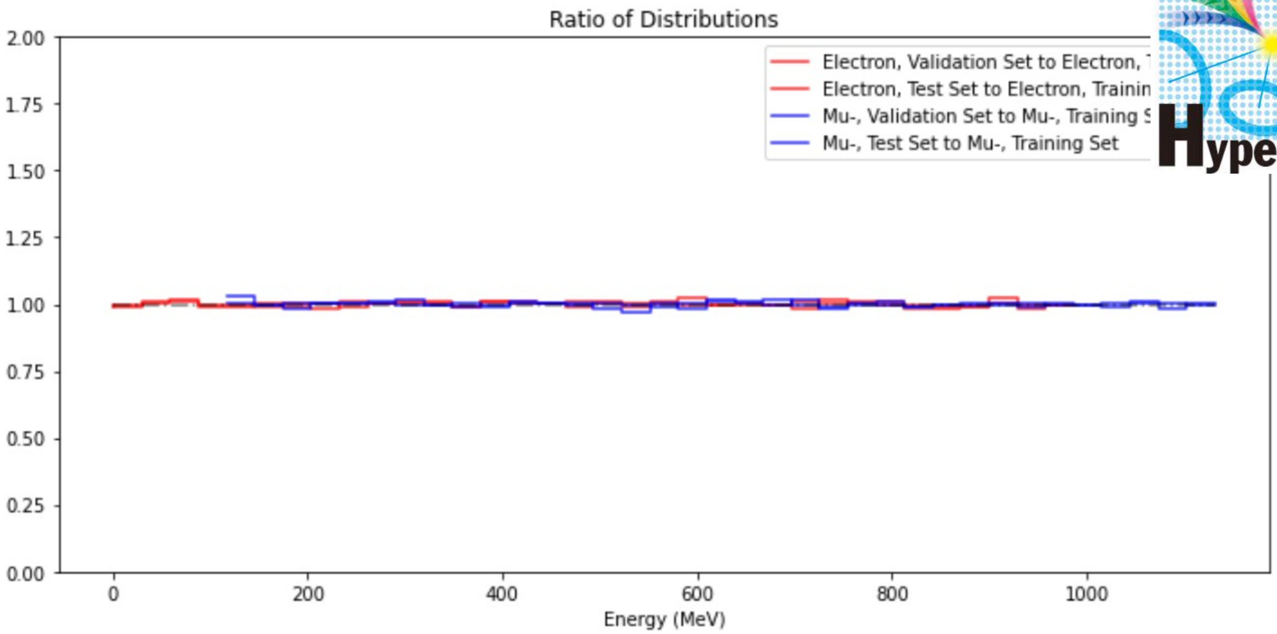
- Each hit PMT is a 'point' with time, charge & position, not fixed to grid
  - CNN learns translation-invariant functions on image
  - PointNet learns symmetric functions on point clouds
    - Symmetric: ordering of points cannot affect outcome
- Convolution-like operations act on each point's charge, time and position
- Information flows between points by learning global transformations applied to all points
- Can apply to any detector geometry

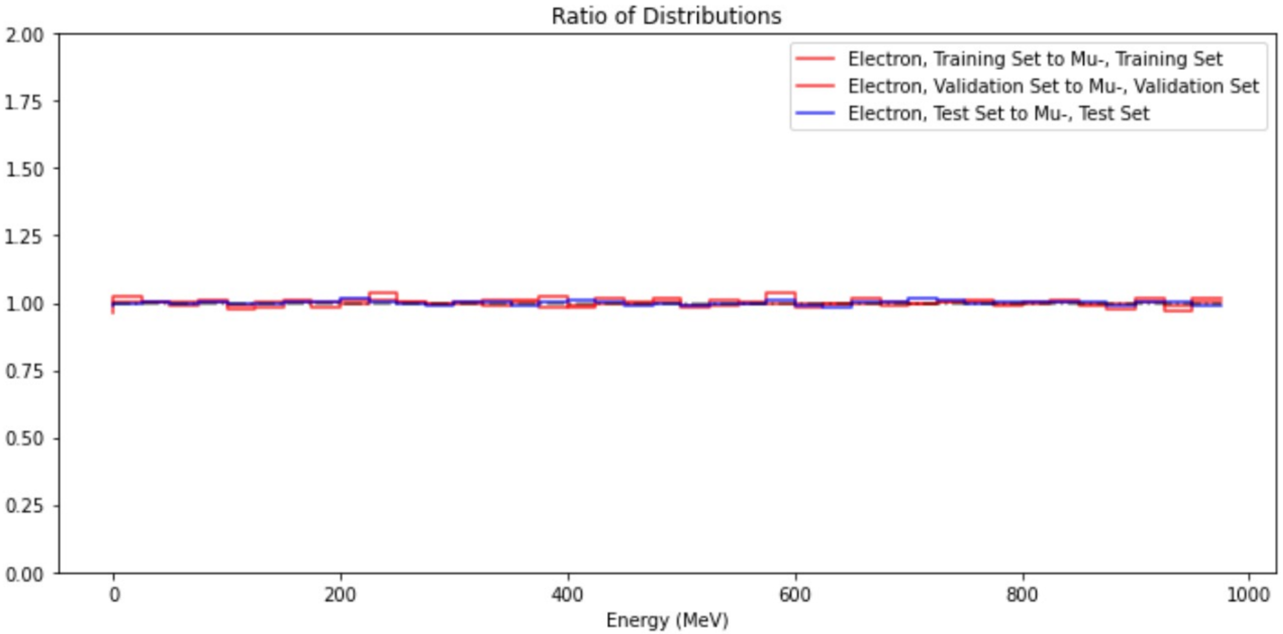


[2]

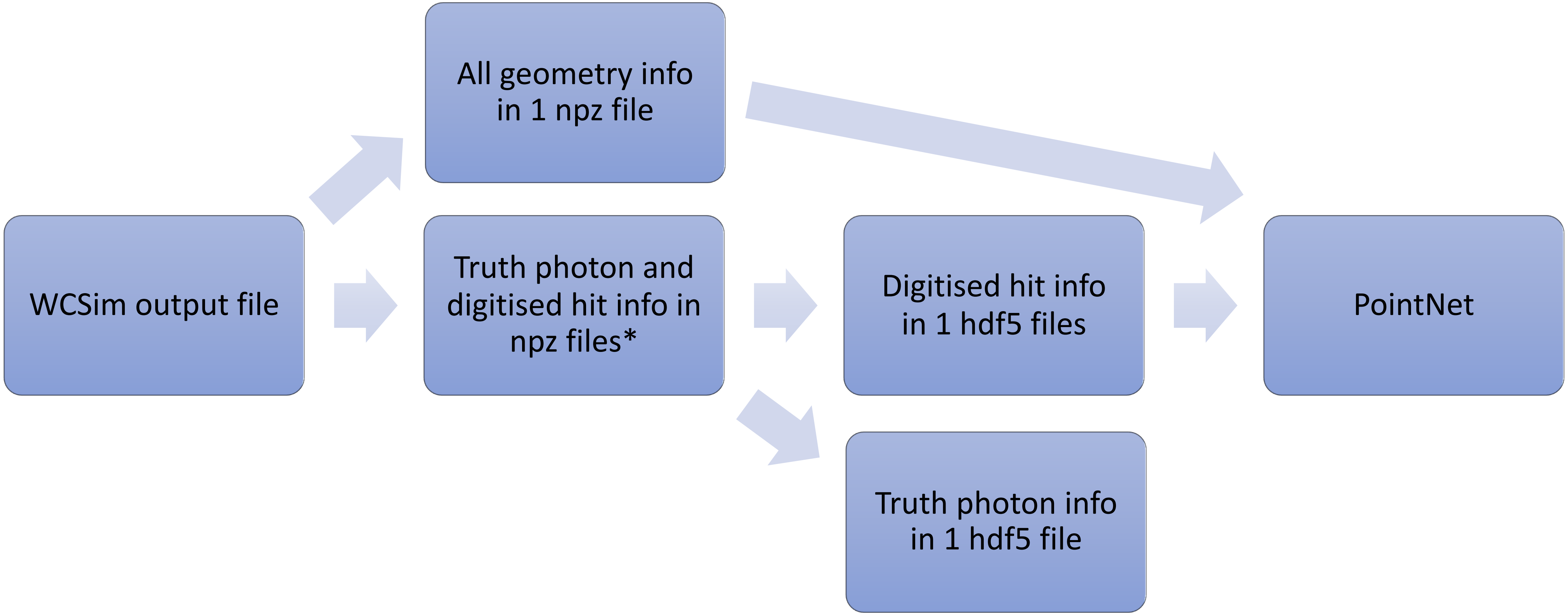
Machine Learning Dataset Split

- The whole dataset is divided into three:
  50% are training sets;
  10% are validation sets;
  40% are test sets.





Input Data Processing Procedure for PointNet



* 1 ROOT file to 1 npz file

References

[1] Qi, C. et al. *PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation.* 2017. Available from: https://arxiv.org/pdf/1612.00593.pdf [Date of Access: 21st July 2021]

[2] Prouse, N. *Machine Learning Techniques for Water Cherenkov Event Reconstruction*. [Presentation] CAP Congress Meeting. TRIUMF. Available from: https://indico.cern.ch/event/985448/contributions/4295792/attachments/2259596/3834940/CAP_%20WatChMaL.pdf [Date of Access: 7th June 2021]

[3] Prouse, N. *PointNet e_γ performance with varying PMT timing resolution.* [Presentation] WatChMaL weekly meeting