

Machine learning techniques for event reconstruction in water Cherenkov detectors

Nick Prouse, TRIUMF
CAP Congress, Simon Fraser University
6th June 2019

Water Cherenkov detectors well suited to many applications

- Large target volumes possible
- Observe charged particles over wide energy range
 - Identify electron-like or muon-like rings
 - Reconstruct position, direction, energy

Reconstruction is essential to achieve physics goals

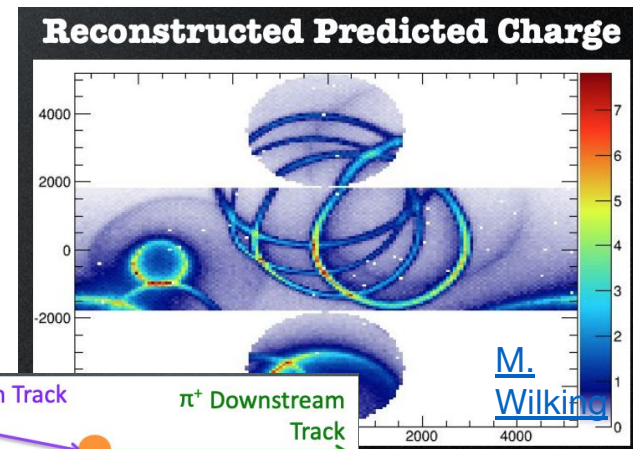
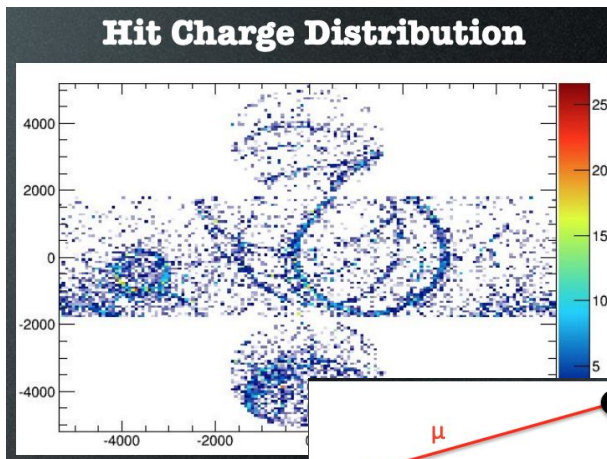
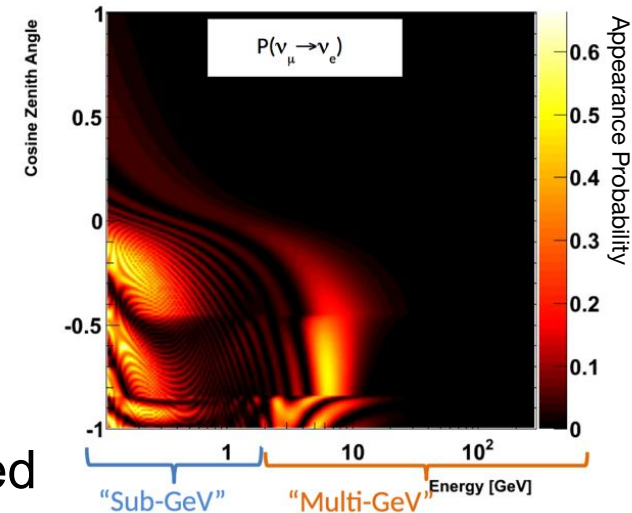
- Future measurements require improved precision
- Complex event topologies & enhanced background removal

Machine learning can improve on existing techniques

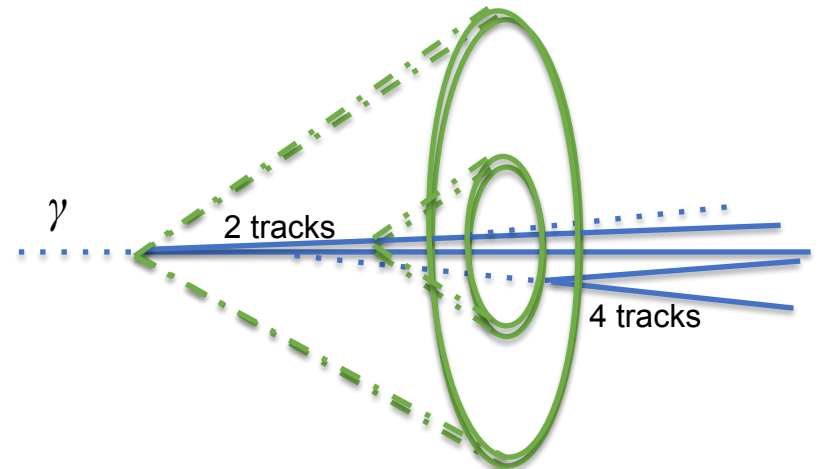
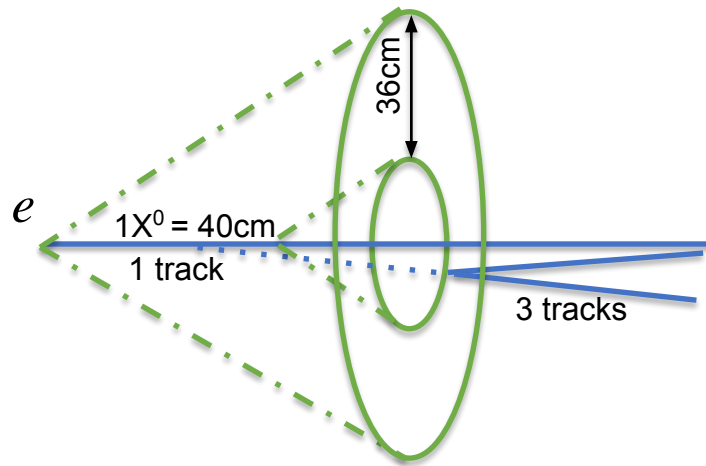
- Rapid advances in deep learning past few years
- Successfully used in other detector technologies

Multi-GeV event reconstruction

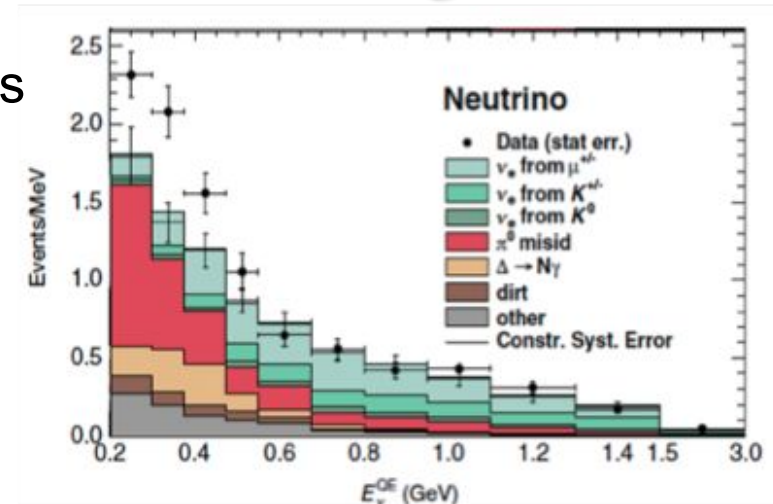
- Complex multi-ring event topologies
- Resolve more of atmospheric ν oscillogram
 - Measurement of mass hierarchy
- Observe ν_τ appearance
- Current reconstruction techniques are limited
 - Complex likelihoods, computationally expensive



e / γ discrimination




- Not achieved with existing techniques
- Reduce $\text{NC}\gamma$ background for ν_e appearance
 - CP measurement at Hyper-K



Dark noise removal

- Dark noise hits limit reconstruction
- Background from dark noise coincidence
- Key for low-E measurements
 - Neutron captures on H (2.2 MeV) or Gd (8 MeV)
 - Supernova relic ν 's, proton decay, backgrounds, $\nu/\bar{\nu}$ discrimination, etc.
 - Solar neutrino measurements
- Current reconstruction limited by assumptions
 - Difficult to build likelihoods for realistic dark noise
 - Correlation between PMTs ignored
 - Machine learning can train on real dark-noise data

Workshop on Machine Learning for Water Cherenkov



HyperK - Canada
MACHINE LEARNING WORKSHOP
APRIL 15-17, 2019

Machine learning for water Cherenkov detectors

The VISPA research centre at the University of Victoria hosted a workshop on the application of machine learning techniques for water Cherenkov detectors. The workshop was held on the campus of the University of Victoria from April 15-17, 2019.

The workshop included tutorials and working sessions using GPU servers to allow participants to gain experience in machine learning techniques. The focus was on developing techniques to analyze simulated photosensor data from the proposed intermediate and Hyper-Kamiokande water Cherenkov detectors. Participation was by invitation only.

The workshop was made possible with support from the University of Victoria Office of the Vice-President Research and Amazon Web Services.




University of Victoria



Victoria Subatomic
Physics & Accelerator
Research Centre



powered by
aws



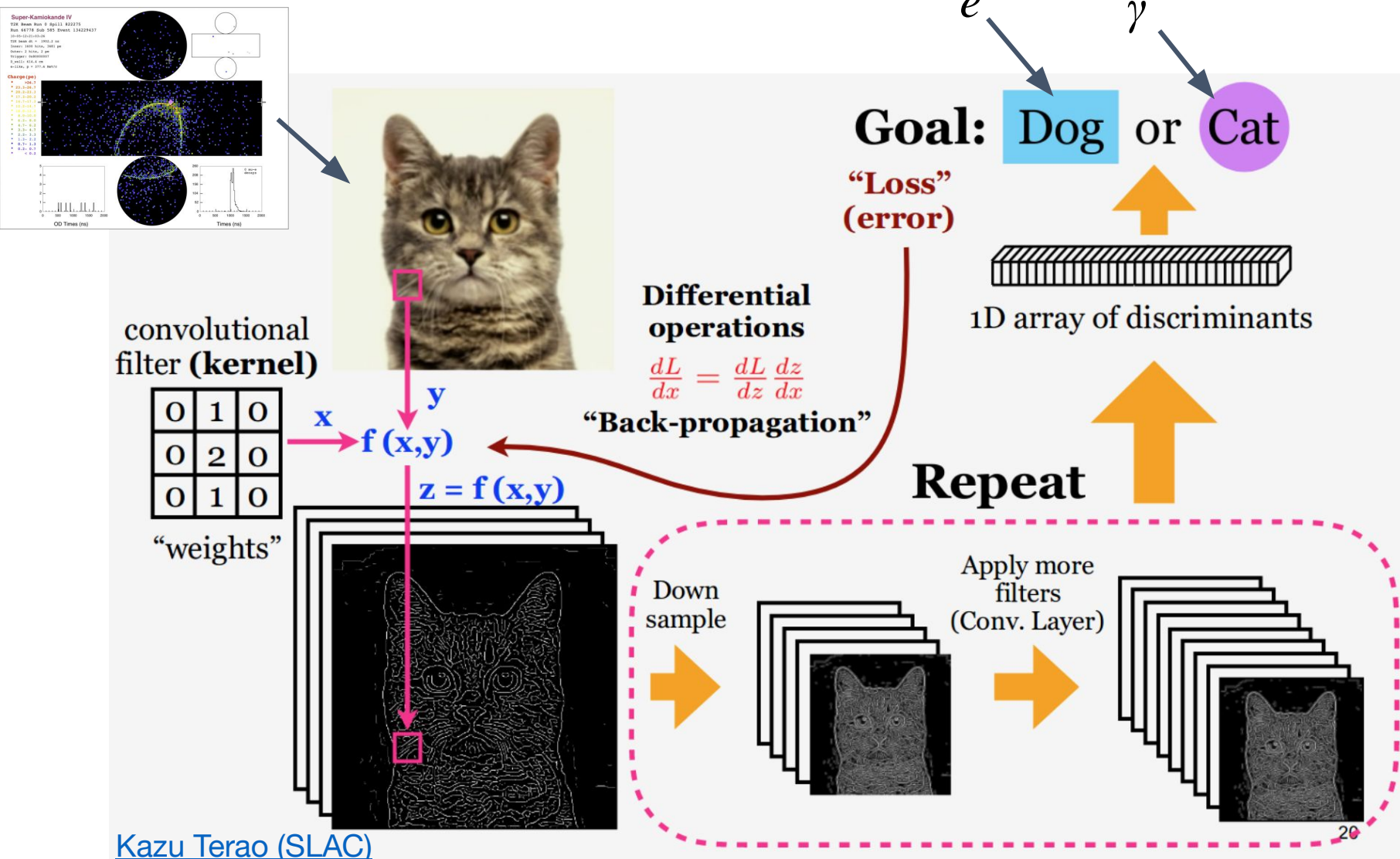
mlw.hyperk.ca/program/agenda
www.watchmal.org
github.com/watchmal



Successful workshop at UVic led by Kazu Terao (SLAC)

- Platform and data prepared in advance
 - Jupyter notebooks with PyTorch ML platform
 - Hyper-K Intermediate detector simulated data
- Working sessions
 - All 19 participants excitedly developed and ran code
 - Preliminary results show some promise
- WatchMaL group formed

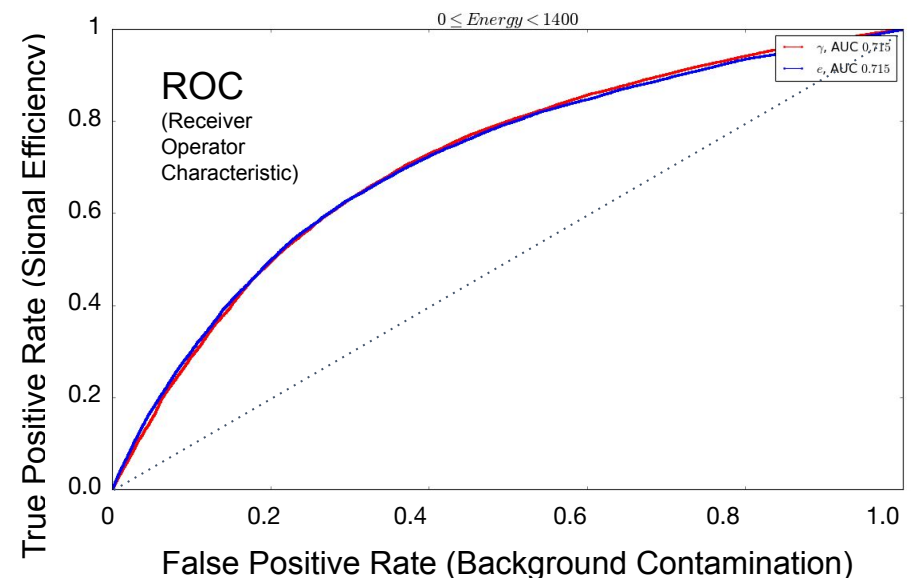
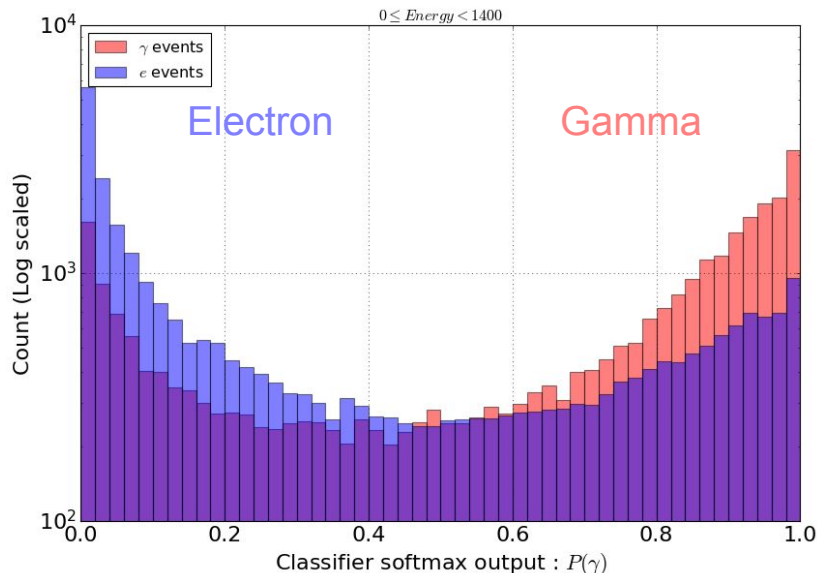
Convolutional neural networks



e/γ discrimination

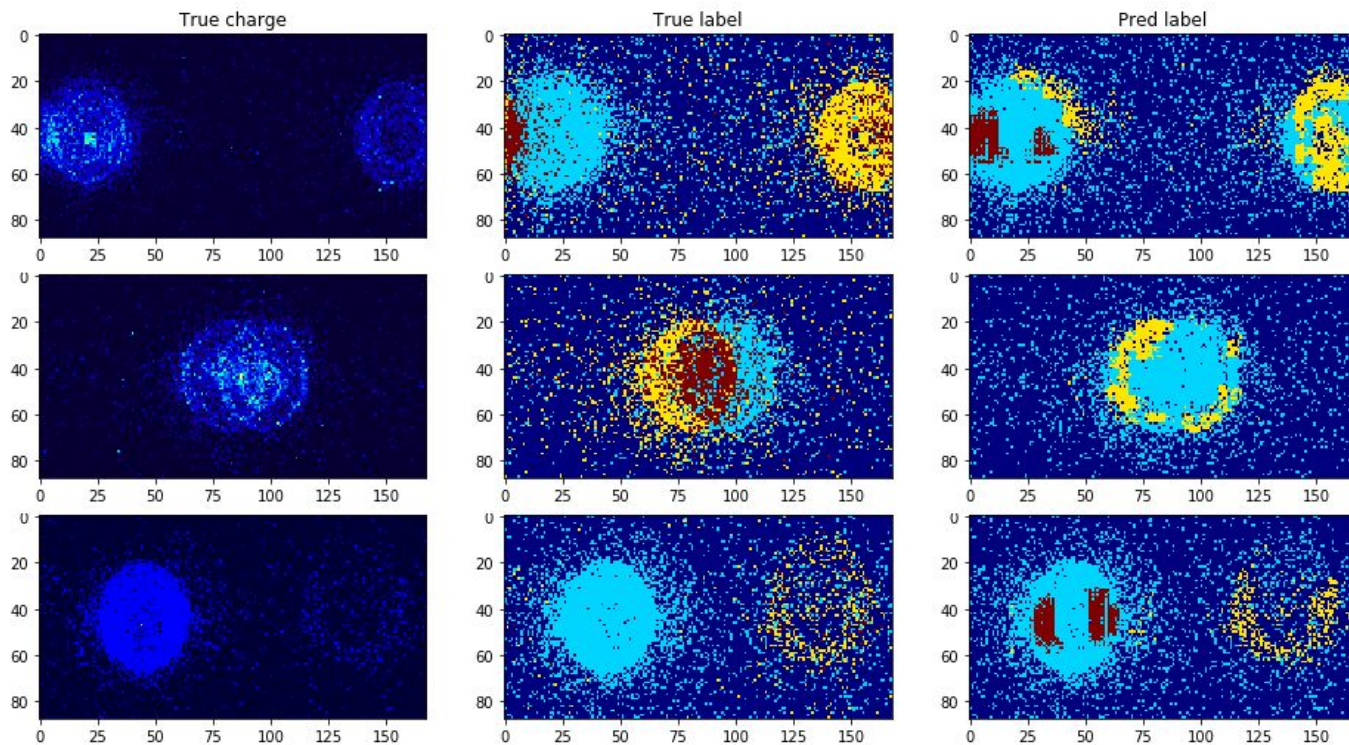
Wojciech Fedorko (TRIUMF, Data Scientist)
Julian Ding, Abhishek Kajal (TRIUMF, Co-op Students)

- Tested on Hyper-K Intermediate detector
 - 10.4 m tall, 7.4 m diameter, 12,160 x 3" PMTs
 - Trained on 300,000 events / particle type
 - Simulated at tank centre, varying direction & energy
- Out-of-the-box implementation already looks promising



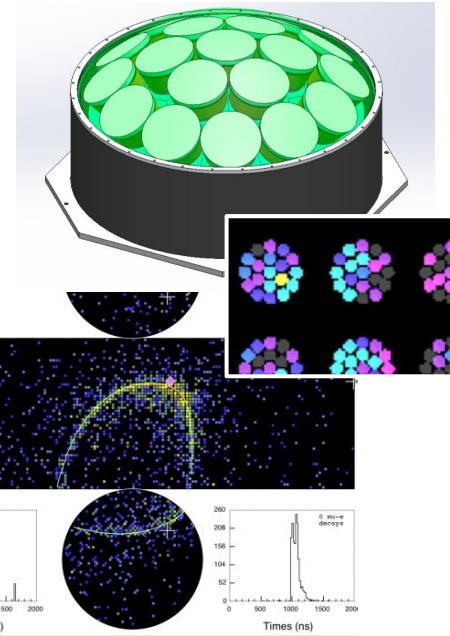
Multi-ring separation

- Combined simulated e and μ events into multi-ring sample
- Trained network to perform ‘semantic segmentation’ on pixels
 - Label PMT hits from e ring, μ ring, or overlap of both
- First attempt okay, but a lot of room for improvement



Non-trivial detector geometry

- Combining barrel and end-caps of cylindrical tank
- Multi-PMT modules further complicate geometry
- Investigate alternative network architectures
 - Sparse submanifold NN or graph NN
- Use physics knowledge to preprocess data



Highly sensitive to data / MC discrepancies

- Techniques to allow training on real data
- Adversarial training prevents training on features in MC but not data

Understanding ‘black-box’ algorithm

- In principle, calculate systematic errors same as traditional reconstruction
- Visualisation of convolutions & layers, investigating marginal events, etc. can help understand what the network has learnt

Improved reconstruction required for future WC measurements

- Complex multi-ring topologies of high-E events
- Improved PID including e / γ discrimination
- Removal of dark noise for low-E events

Machine learning can provide improvements

- CNNs very successful in image processing
- Early results look promising
 - e / γ discrimination in simplified simulations
 - First attempts at separating overlapping rings

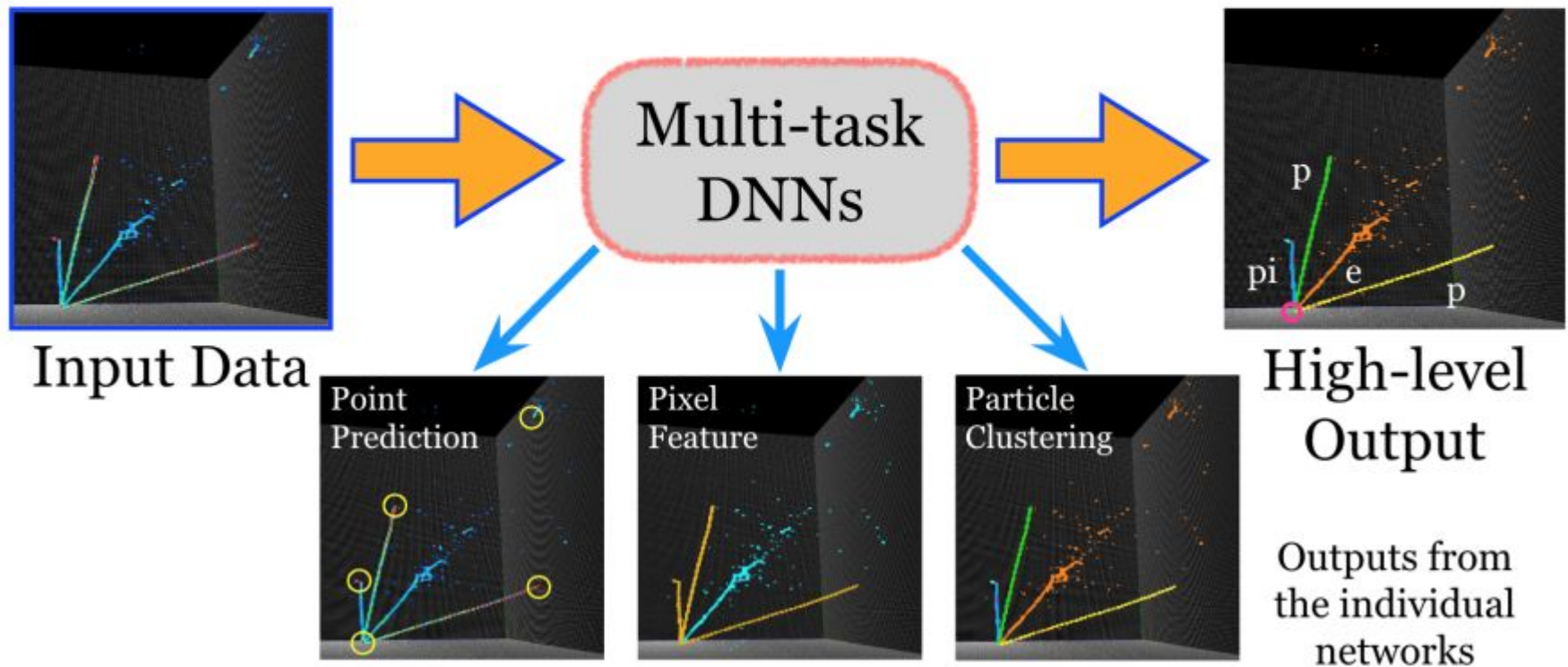
Many remaining challenges

- Work with full cylindrical (and multi-PMT module) tank geometry
- Investigate other network architectures & other reconstruction goals
- Ensure network is not overtraining or learning data / MC differences
- Understand the trained networks for validation and diagnosing issues

Backup

Kazu Terao (SLAC)

Introduce physical feature extraction tasks (auxiliary targets) to bias the data transformation path to support producing a logical conclusion. Optimize the whole reconstruction chain.



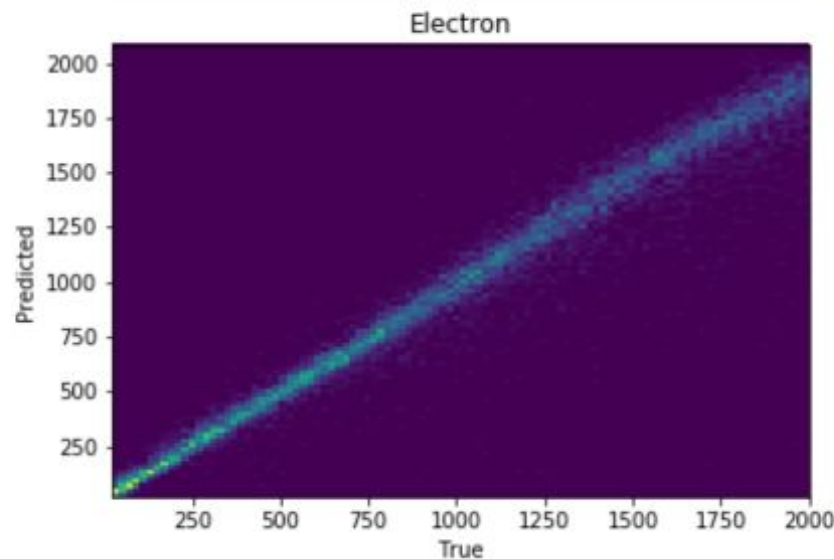
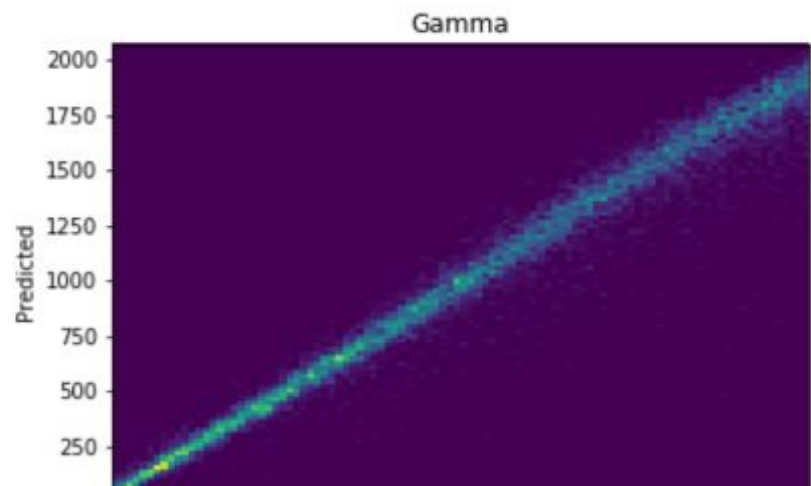
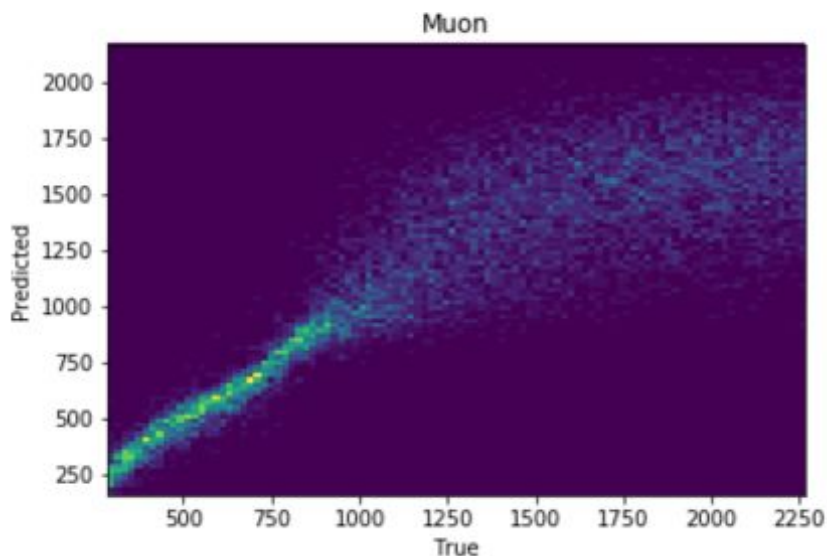
Can we do the same with water Cherenkov events?

- Based on residual learning framework introduced in arXiv:1512.03385
- The framework provides a solution to the degradation problem exposed in deep neural networks in which as the neural network depth increases, the model accuracy gets saturated and then degrades rapidly
- 900000 simulated events with 80-10-10 (training-validation-test) split for 10 epochs
- This model can accept images of size 16 x 40 x 38 [Height x Width x (2 x #PMTs in mPMT)] and generate softmax output for 3 classes i.e. P(gamma), P(electron) and P(muon).

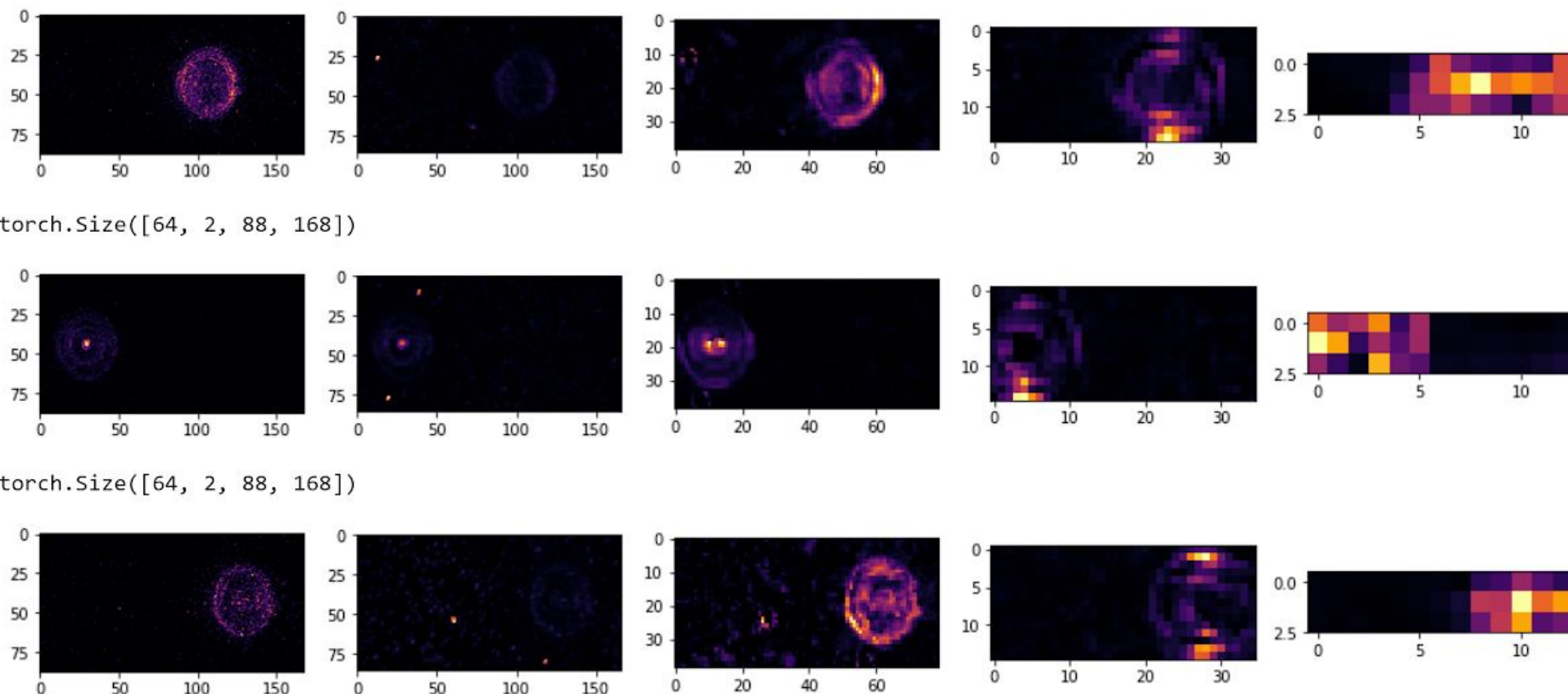
Layer Name	Output Size	ResNet-18
conv1	$112 \times 112 \times 64$	$7 \times 7, 64$, stride 2
conv2_x	$56 \times 56 \times 64$	3×3 max pool, stride 2 $\left[\begin{array}{c} 3 \times 3, 64 \\ 3 \times 3, 64 \end{array} \right] \times 2$
conv3_x	$28 \times 28 \times 128$	$\left[\begin{array}{c} 3 \times 3, 128 \\ 3 \times 3, 128 \end{array} \right] \times 2$
conv4_x	$14 \times 14 \times 256$	$\left[\begin{array}{c} 3 \times 3, 256 \\ 3 \times 3, 256 \end{array} \right] \times 2$
conv5_x	$7 \times 7 \times 512$	$\left[\begin{array}{c} 3 \times 3, 512 \\ 3 \times 3, 512 \end{array} \right] \times 2$
average pool	$1 \times 1 \times 512$	7×7 average pool
fully connected	1000	512×1000 fully connections
softmax	1000	

Reconstruct continuous quantities (e.g. energy) in addition to discrete classifications

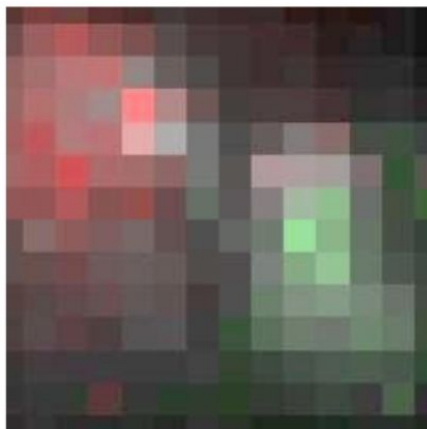
[M. Hartz \(TRIUMF\)](#), [C. Vilela \(Stony Brook\)](#), [B. Richards \(QMUL\)](#)



By separating the convolutional layers, the location of filter activation can be plotted showing which features the neural net is interested in between layers



Looking inside CNNs



<https://distill.pub/2017/feature-visualization/>

Attribution¹ studies what part of an example is responsible for the network activating a particular way.



Slightly positive
activation examples



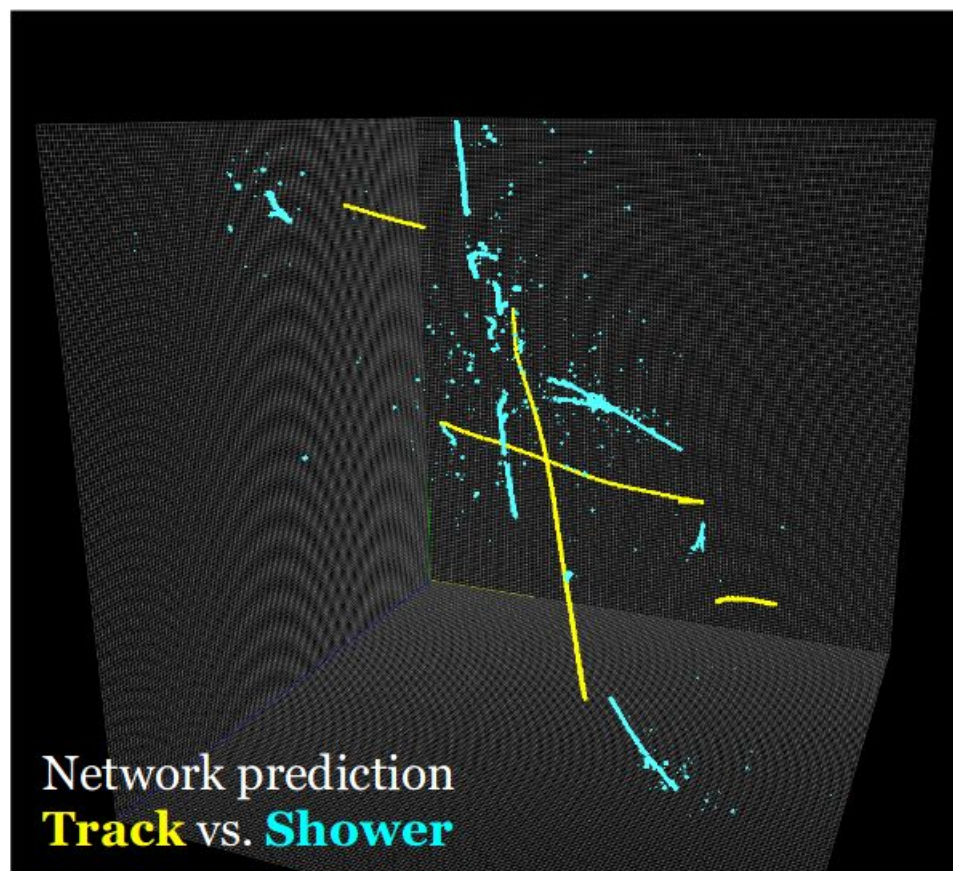
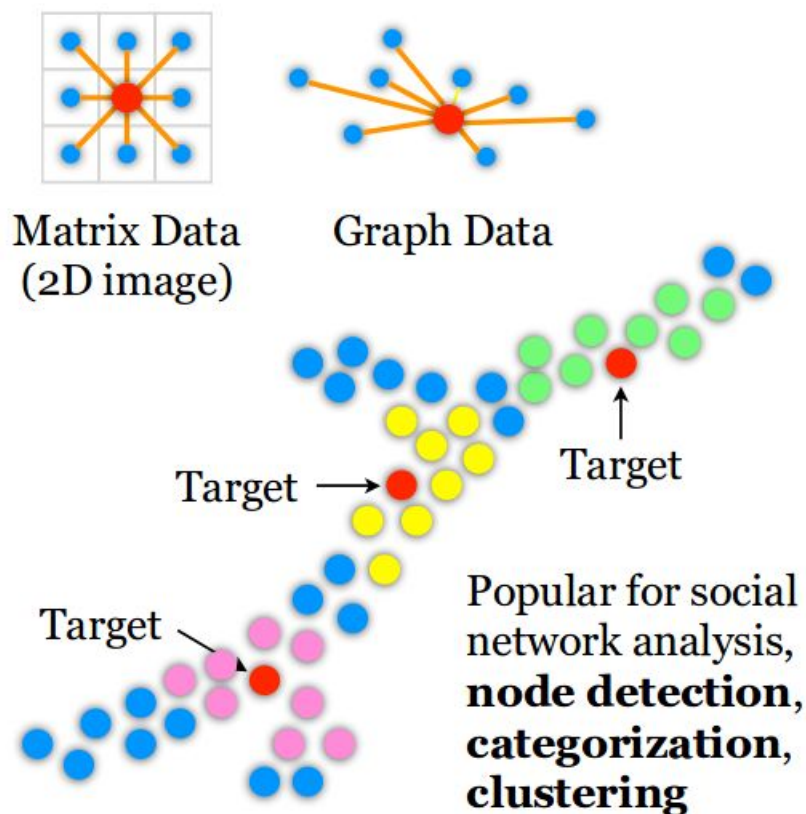
Maximum activation
examples



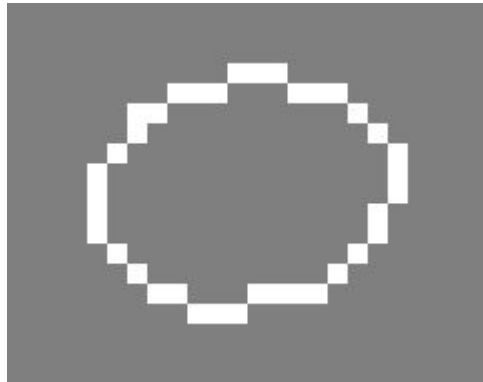
Positive optimized

Kazu Terao (SLAC)

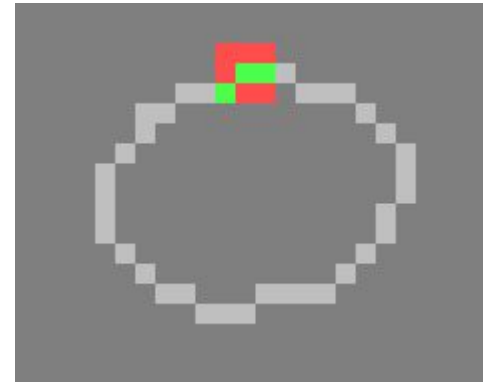
- Sparse data representation = **Can address “dense CNN” issues**
- But **not ideal for locally dense geometrical feature** extraction



Submanifold Sparse CNNs



CNN



SparseConvNet



[arXiv:1706.01307](https://arxiv.org/abs/1706.01307)

Train network on real data, use to generate more samples



Randomly generated images of non-existent people
thispersondoesnotexist.com

Variational autoencoder: Use real data, modify with generative network to different event type, etc. More accurate (data-based) MC?