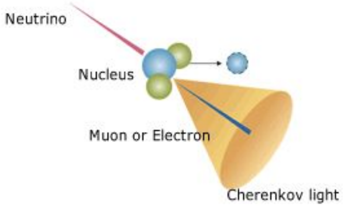
ResNet Particle Identification in Water Cherenkov Detectors

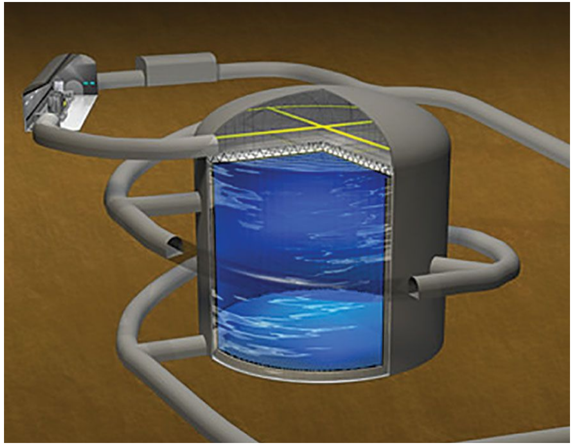
WNPPC

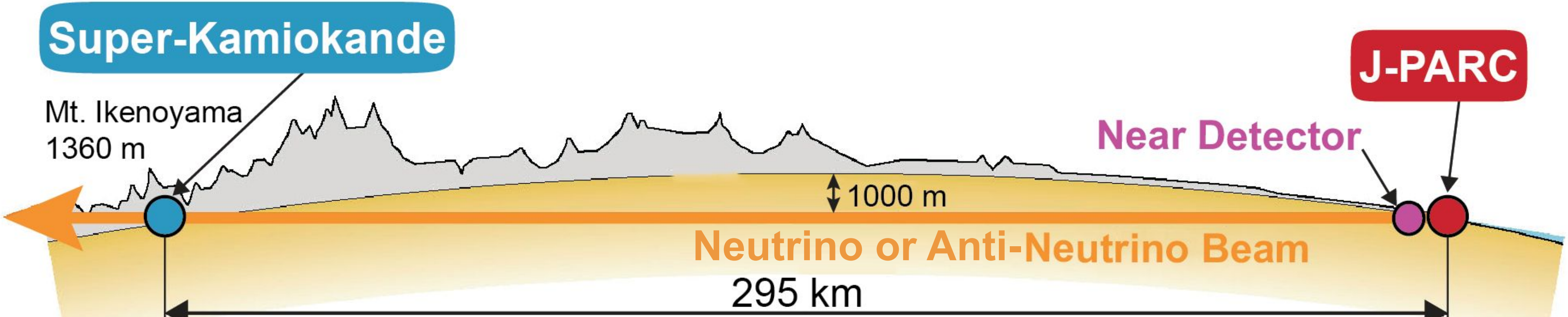
Joshua Tindall
TRIUMF
February 12, 2021

Hyper-K and Water Cherenkov Detectors

- Water Cherenkov Detector
  - Detect neutrino interactions from Cherenkov light produced by resulting charged particles in a medium
- Hyper-Kamiokande (Hyper-K)
  - Next generation successor to Super-K
  - Several ambitious physics goals
  - Also T2K

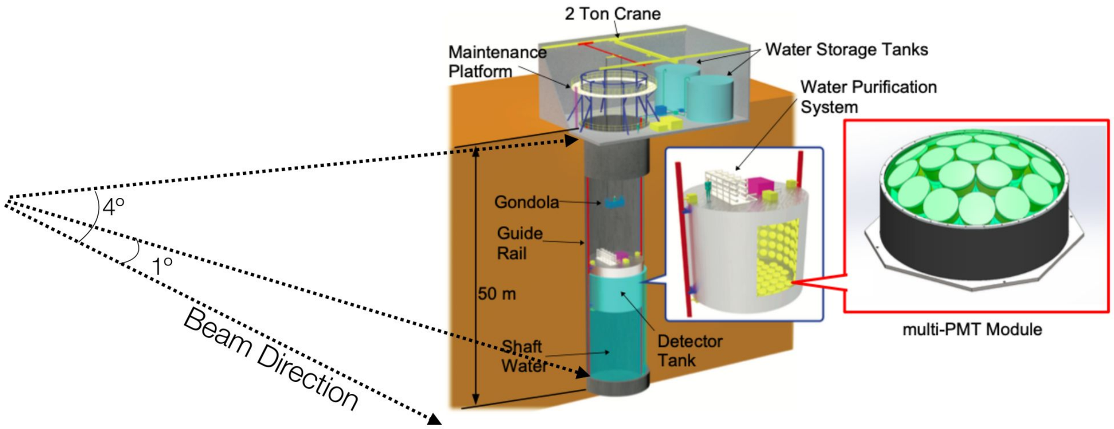


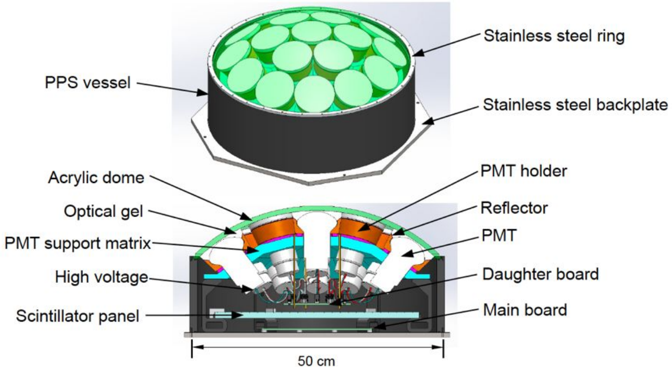




IWCD and mPMTs

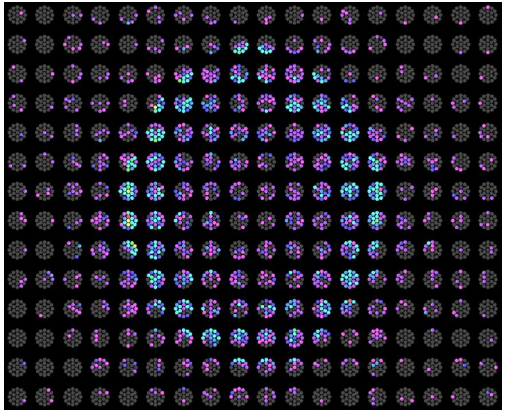
- Intermediate Water Cherenkov Detector (IWCD)
  - Located near (~1 km from) beam source
  - Characterize beam to control systematics
- Constructed using multi photomultiplier tubes (mPMTs)
  - Offers increased granularity and timing resolution



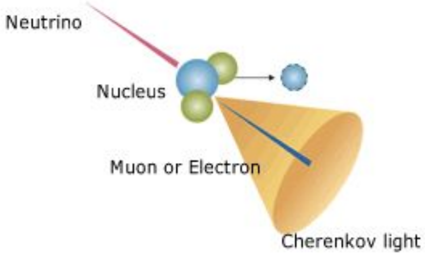


Event Reconstruction

- Inferring event parameters from Cherenkov detector data
  - Given charge and time data from PMT hits
  - Would like to reconstruct momentum/energy, particle type
- Regression and classification
  - Recover continuous parameters
  - Predict discrete classes



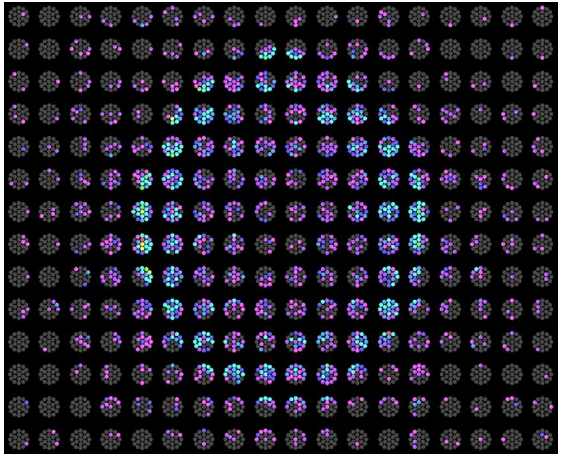
Reconstruction Algorithm

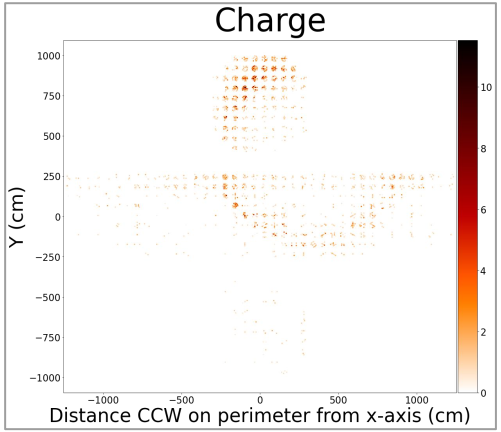


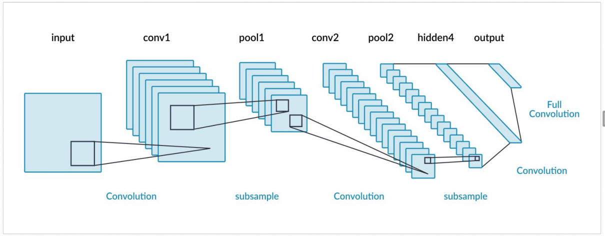
http://www-sk.icrr.u-tokyo.ac.jp/sk/detector/howtodetect-e.html

Particle Type Identification

- Classify events as one of e-, γ or μ-
  - Task objective is to identify the class of event associated with charge and time data from PMTs
- Important for discriminating between signal events and backgrounds
  - When inferring event statistics
- Good model problem for characterizing potential of machine learning techniques







γ

e-

μ-

Likelihood Based Event Reconstruction

- Conventional fitting algorithms such use maximum likelihood fitting
  - Produce likelihood of model assuming particle type
  - Predict based on relative likelihoods of different classes
- We have a likelihood based algorithm (fiTQun) that achieves good reconstruction performance with IWCD data
- Cost of fitting events high with increasing data resolution and model complexity
  - Reaching computational limits in performance on difficult tasks

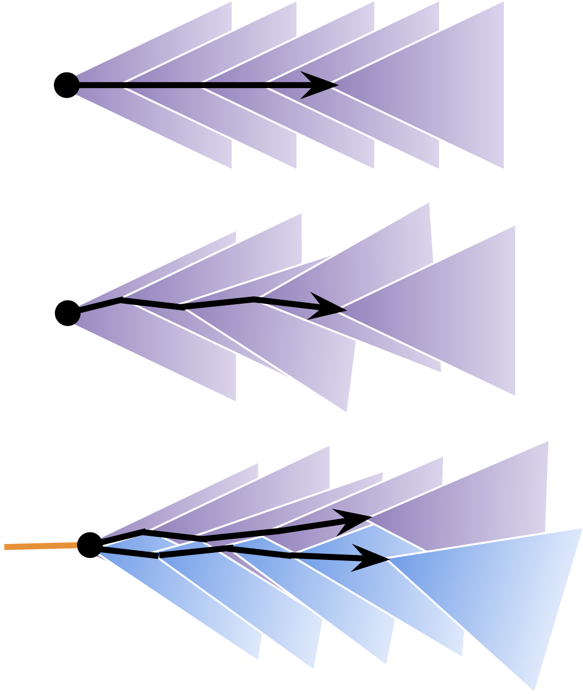
$$\underbrace{\boxed{L}}_{\text{Likelihood to maximise}}\left(\underbrace{\boxed{\mathbf{x}}}_{\text{Candidate event hypothesis}}\right) = \prod_{j}^{unhit} \underbrace{\boxed{P_j\left(unhit|\mathbf{x}\right)}}_{\text{Probability of no hit at PMT}} \prod_{i}^{hit} \underbrace{\boxed{P_i\left(hit|\mathbf{x}\right)}}_{\text{Probability of hit at PMT}} \underbrace{\boxed{f_q\left(q_i|\mathbf{x}\right)}}_{\text{Hit charge probability density}} \underbrace{\boxed{f_t\left(t_i|\mathbf{x}\right)}}_{\text{Hit time probability density}}$$

Particle Event Discrimination

muon
(1 track)

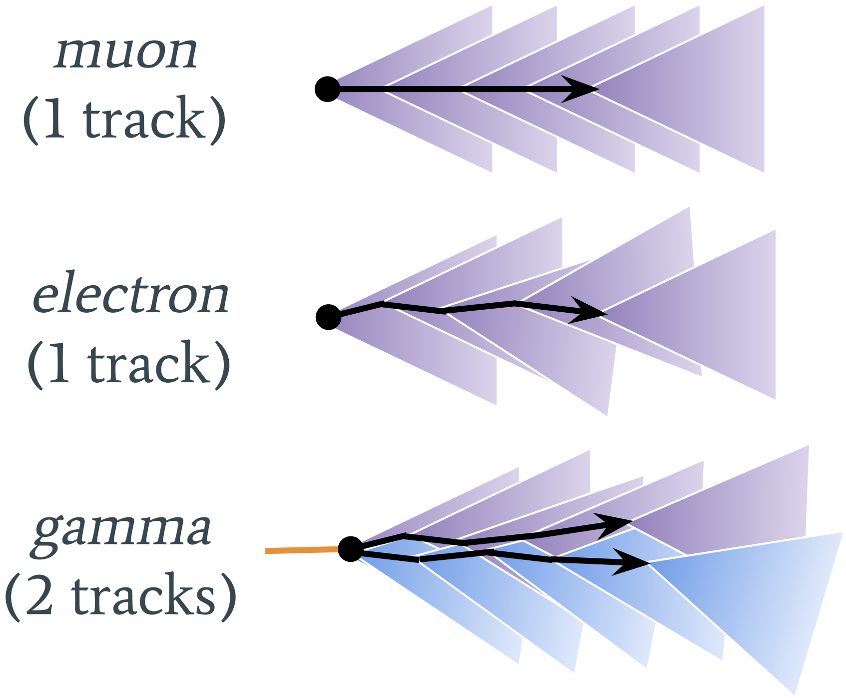
electron
(1 track)

gamma
(2 tracks)



- Classify events as one of e-, γ or μ-
- μ- produced by ν_{μ} interactions
  - Little deflection, produce crisp cones
- e- produced by ν_{e-} interactions
  - Undergoes scattering, producing a more diffuse ring
- γ particles are secondary products that undergo pair production
  - Produces two highly proximal electron-like rings

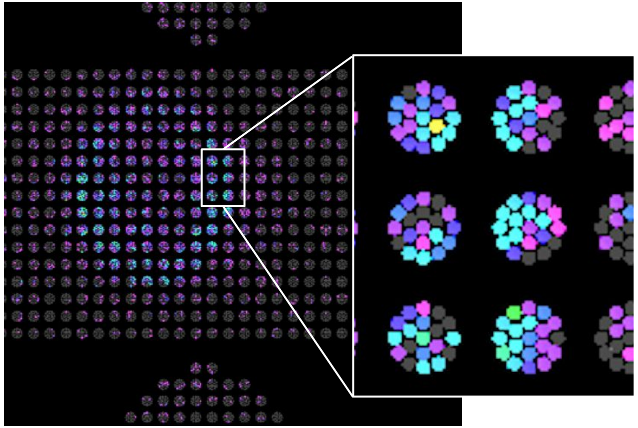
e-/γ Discrimination

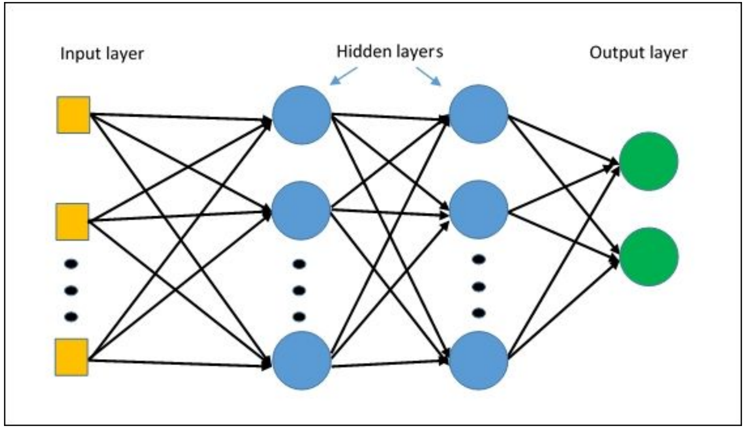


- Really two tasks, e-/γ vs μ- task and e- vs γ task
  - e-/γ both produce diffuse rings vs sharp μ- rings
- e-/γ vs μ- task is much easier than e- vs γ task, and treated well by existing methods
- e- vs γ task still a challenge for existing methods
- With the increased resolution of mPMTs, it may now be possible to distinguish gamma events
  - Hopefully achieve meaningful performance in discrimination in e- vs γ task

ML Based Event Reconstruction

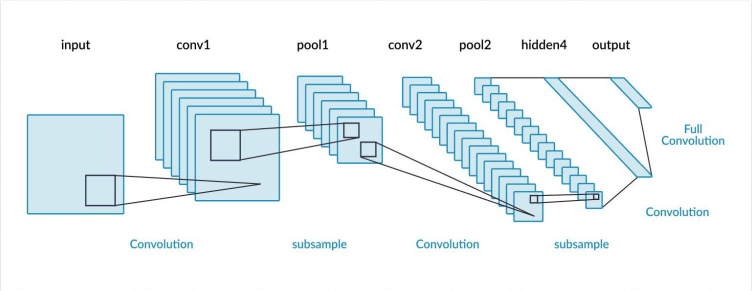
- Compared to likelihood models, neural networks
  - Much lower computational cost for prediction (once network has been trained)
  - Can avoid requiring simplifying assumptions about detector behaviour
- Hope is that neural network methods will allow us to make full use of increased resolution
  - Optimistic that this may help with difficult tasks such as e/gamma discrimination

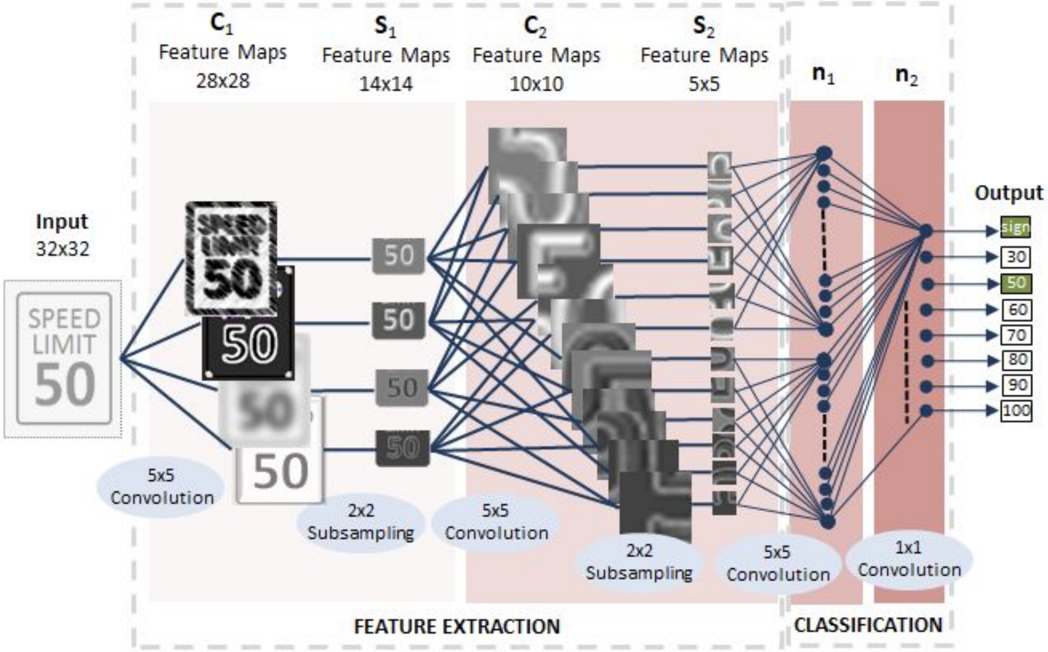




Convolutional Neural Networks

- Common for computer vision applications
- Learn convolutions to perform on the data
    - Kernels applied to input image data to extract image features into channels
- Effectively invariant to translations

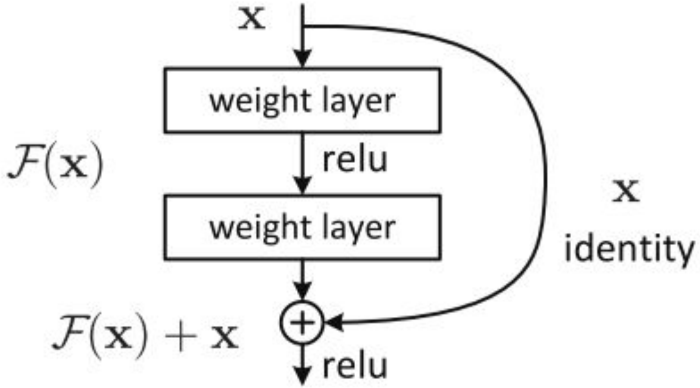


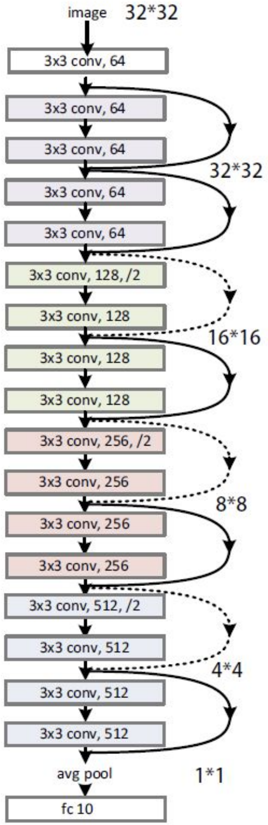


https://e2e.ti.com/blogs_/b/behind_the_wheel/archive/2018/02/08/ai-in-automotive-practical-deep-learning

ResNet and Residual CNNs

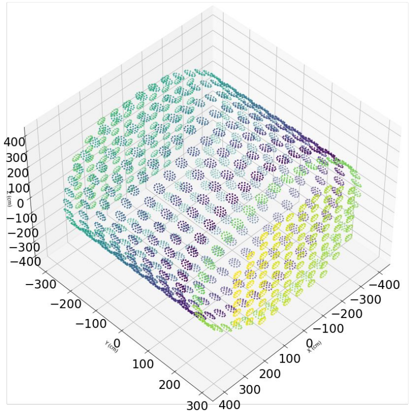
- Residual CNNs contain skip connections which pass identity forward several layers
  - Allows smoother flow of gradient
- Residual CNNs have seen widespread success on computer vision tasks
- Architecture used is based on ResNet-18

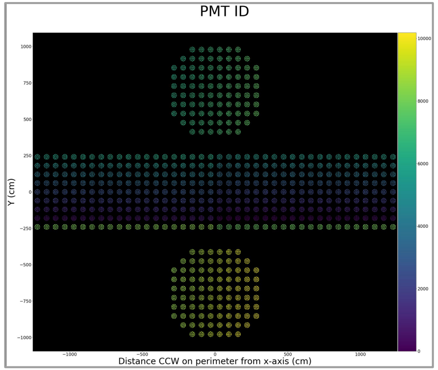


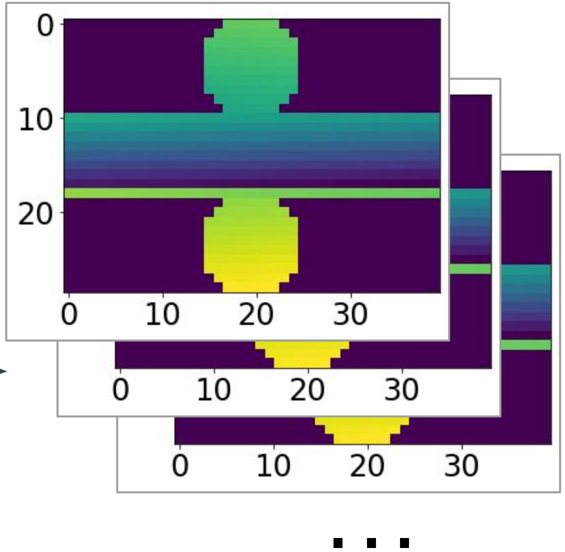


Data Format

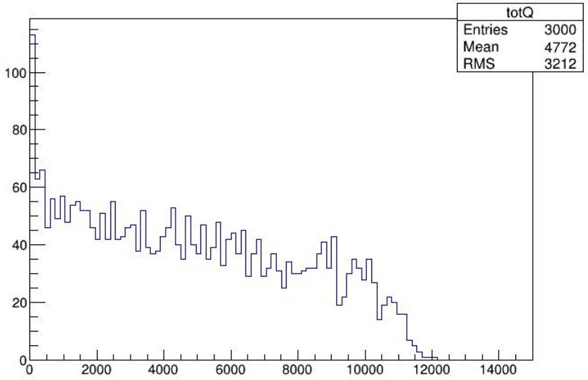
- Collapsed ResNet data format
  - Unfold cylindrical data into flat map
  - Unpack PMTs into 19 channels
- Currently using "digitized" hit data
- More sophisticated approaches possible
  - Topological mapping used for regression tasks

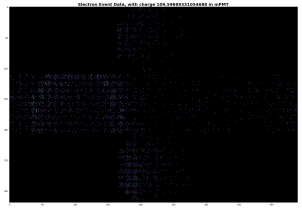


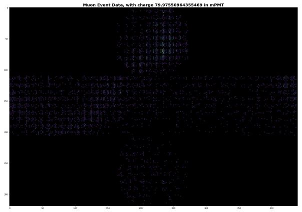


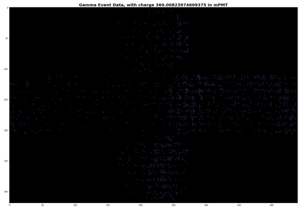


Dataset





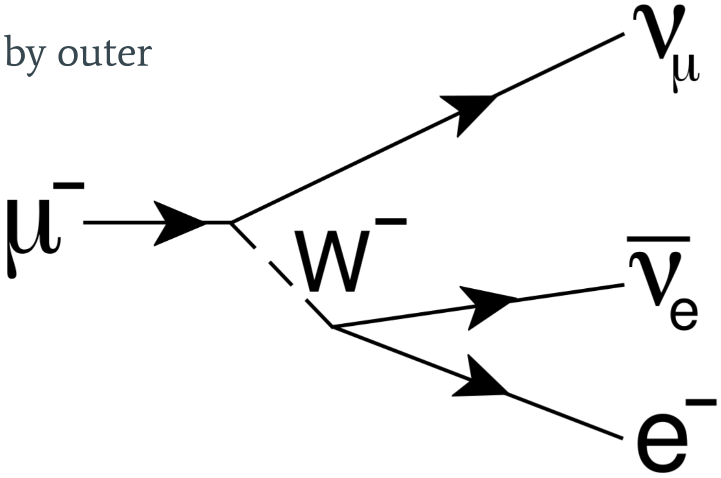




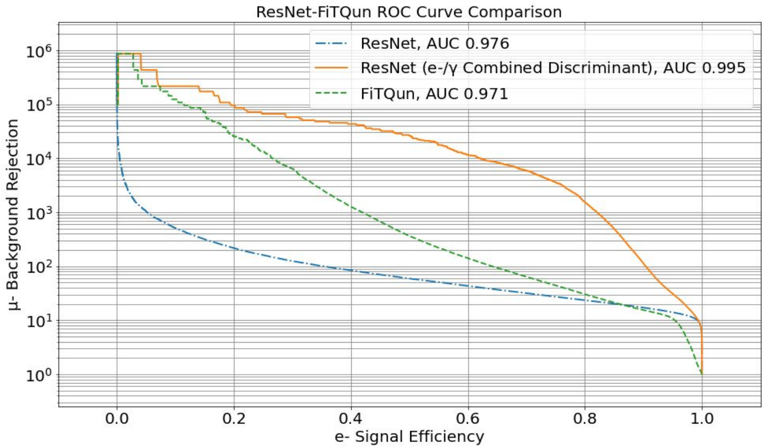
- Particle identification dataset consists of events from three kinds of particle: e-, γ and μ-
- 3M events were generated for each particle type using WCSim
  - Uniform distribution of kinetic energy over the range 0-1000 MeV above Cherenkov threshold
  - Isotropic distribution of angles
  - Event origin distributed uniformly over IWCD volume
  - Events with no hits were removed
- Simulations yield time and charge of all PMT hits
  - Used only charge data and 'digitized' hits
- Events were split 50/10/40 among training/validation/test datasets

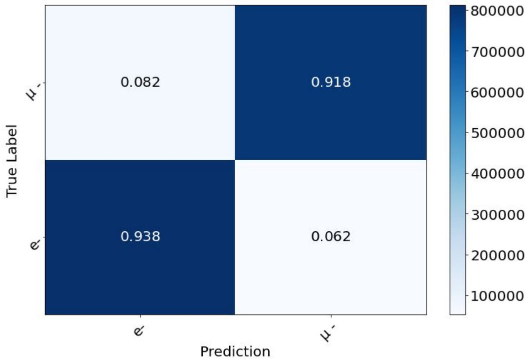
Event Flags

- Examined impact of possibly irregular events
  - Flagged events with Michel electrons
    - Detector triggered after a muon had produced an electron
    - Will not be present in future datasets (e.g. short tank geometry)
  - Added an outer detector (OD) event flag
    - Events which would not have been detected by outer volume elements
    - Included from the start for future datasets
- No noticeable effect on performance

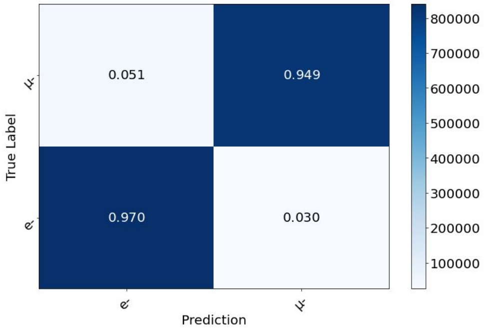


Performance and Previous Results





FiTQun confusion matrix



ResNet (Combined discriminant) confusion matrix

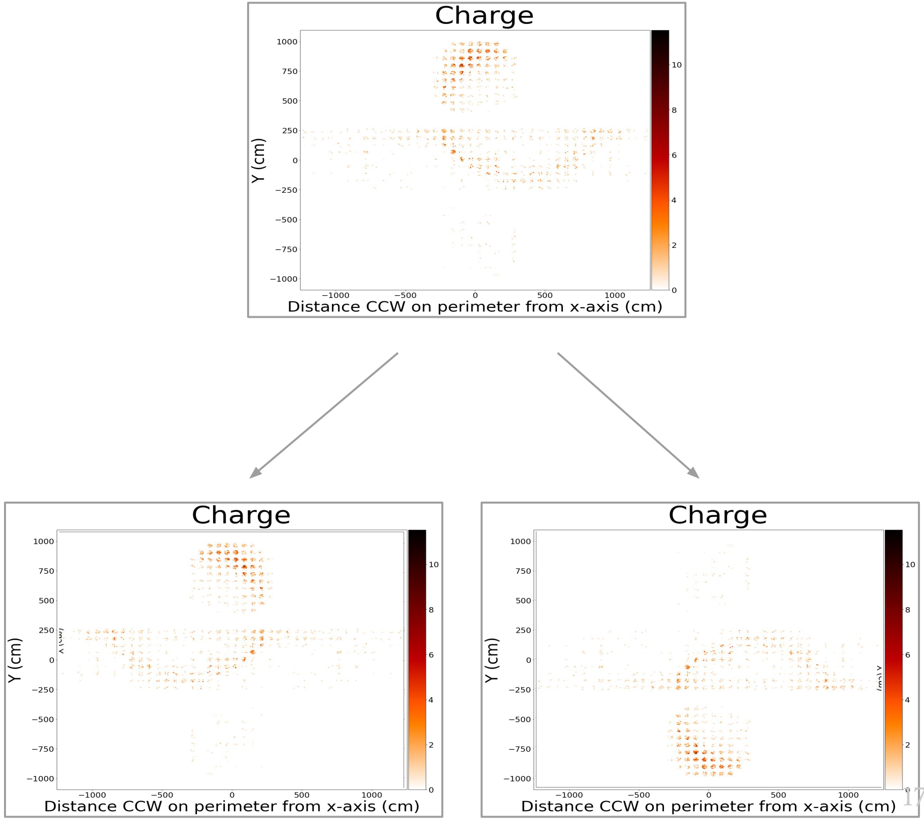
- Previously results had been obtained with a 3-class ResNet-18 model
- Performance appears to be competitive when compared with FiTQun
- Also investigated dependence of performance on some key features such as event energy and particle trajectory

Performance Improvements

- Two main lines of improvement
  - Data Augmentation (addition of transforms)
  - 2-Class Models

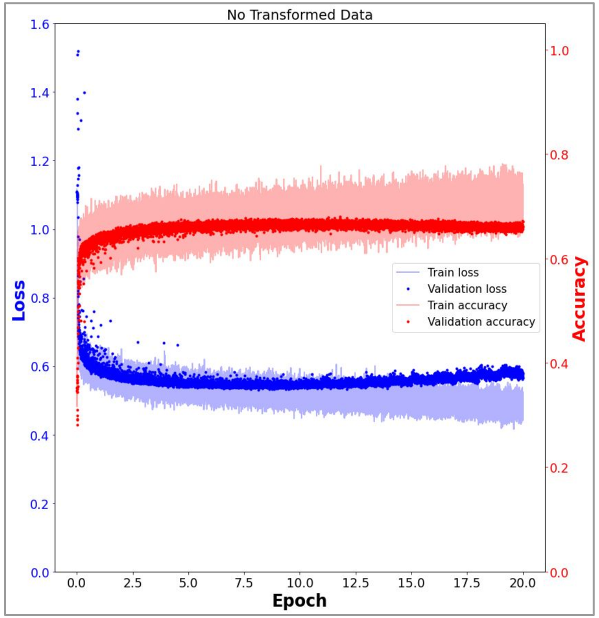
Data Augmentation

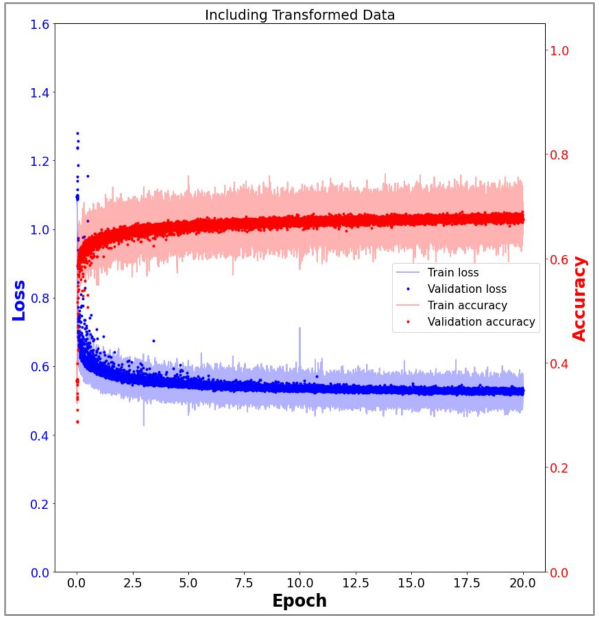
- Introduced random horizontal and vertical flipping
    - Uses dataset data to generate related data
- Effectively increases the size of the training dataset
    - Larger representative sample to train on



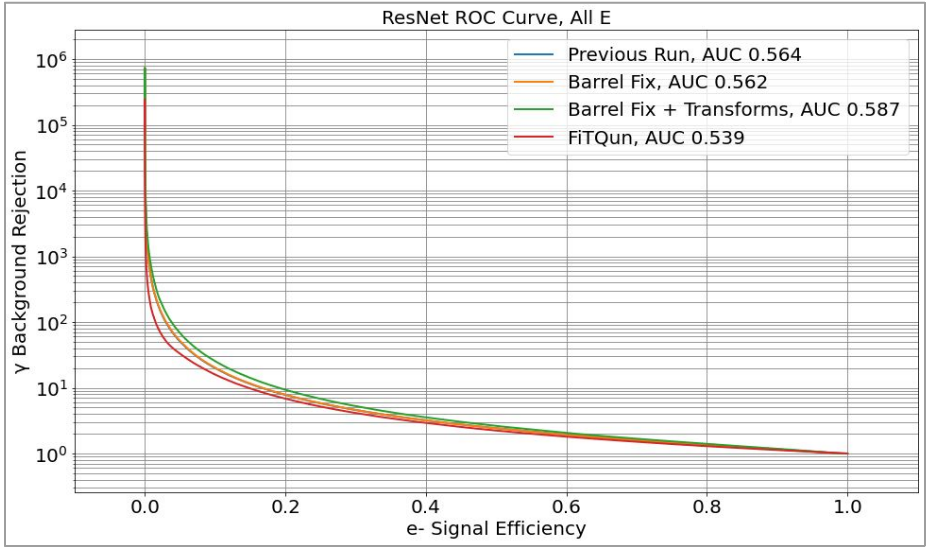
Data Augmentation Effect on Overfitting

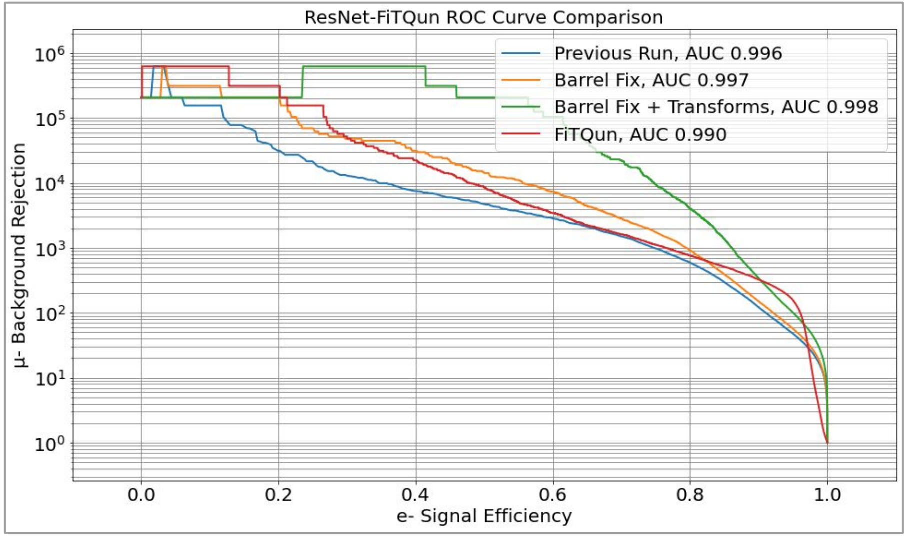
- Introduction of transforms appears to help reduce overfitting
- Models trained with augmentation avoid overfitting otherwise observed by around 20 epochs





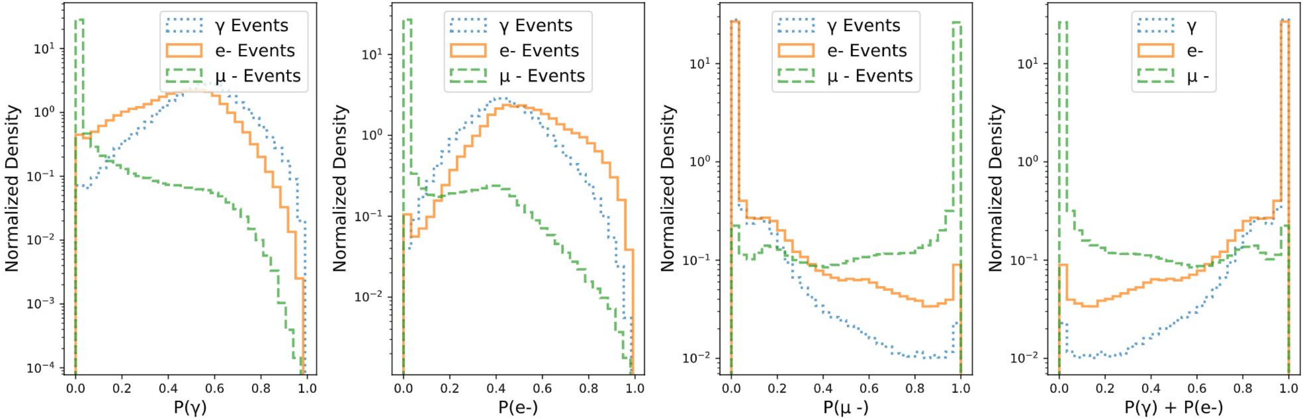
Data Augmentation Performance





- Reduced overfitting does in fact lead to improved performance on the test set
- In combination with other improvements, yields substantial jump in performance over previous models

2-Class Model Motivation

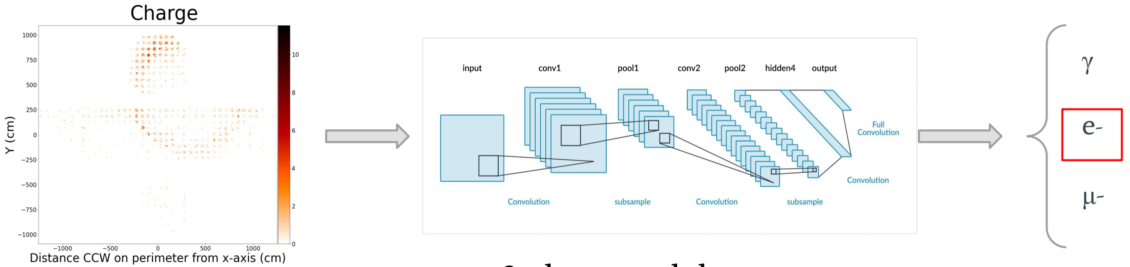


Likelihoods for true events of different types for a 3-class model

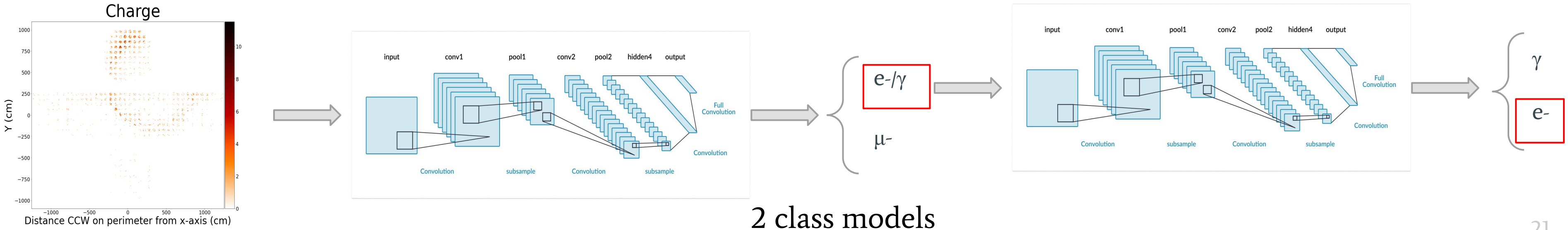
- Previous 3-class results indicated that γ rejection is more difficult than μ- rejection
  - Stronger discrimination between e-/γ class events and μ- than between e- class and γ class events
- Solutions?
  - Split the problem into binary tasks to investigate whether this would improve γ rejection

2-Class Models

- First tackle separating e-/γ class from μ- class with one classifier
- Have second network tackle more difficult discrimination of e- class from γ class
    - This classifier can train only on e- and γ data

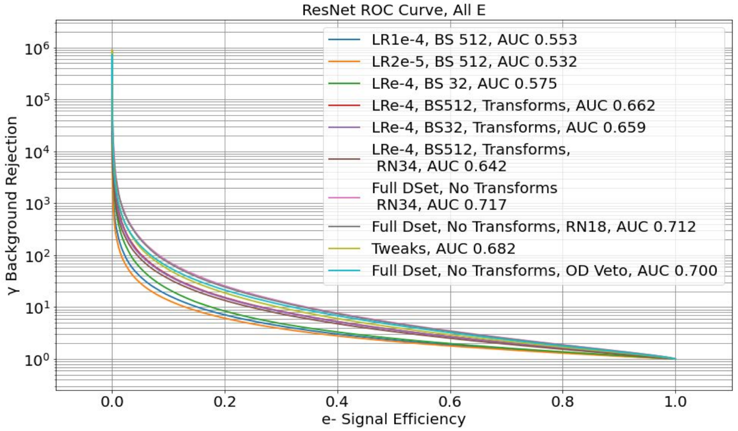


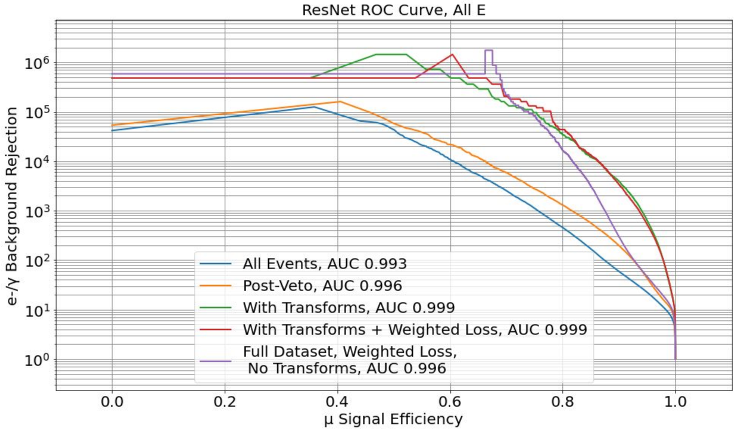
3 class model



2 class models

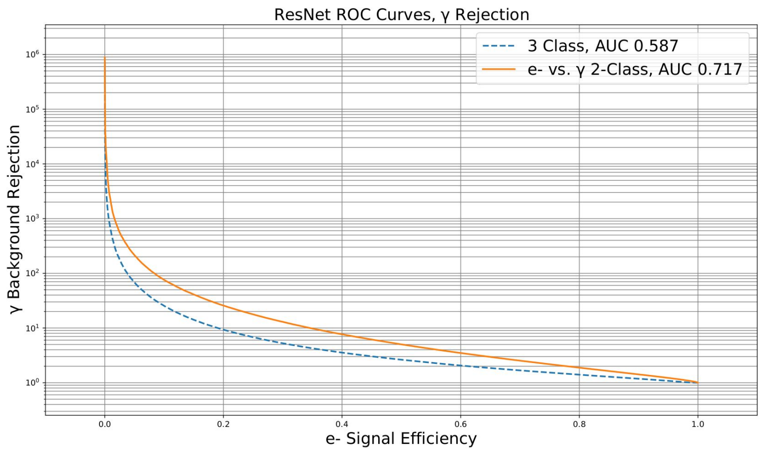
2-Class Tweaking

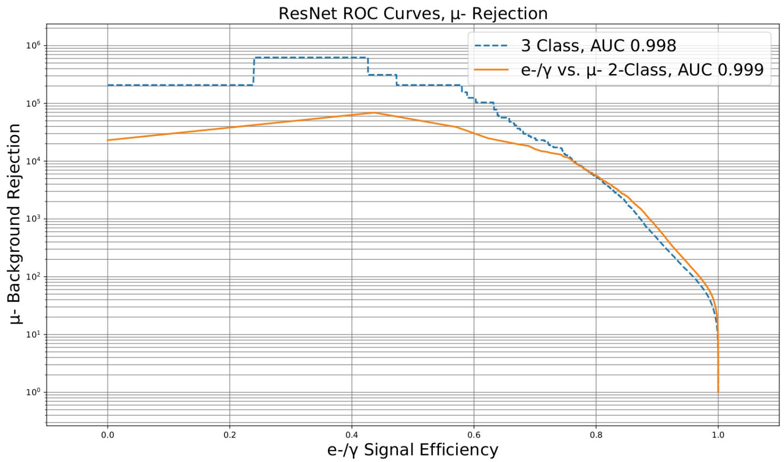




- 2-class models were trained with a variety of architecture and hyperparameter tweaks for optimization
  - Varied learning rate, batch size, Resnet-34
- Combined with other improvements such as transforms
- To fully train e-/γ discriminator, additional e- and γ events were generated and added to the dataset

2-Class Performance Comparisons

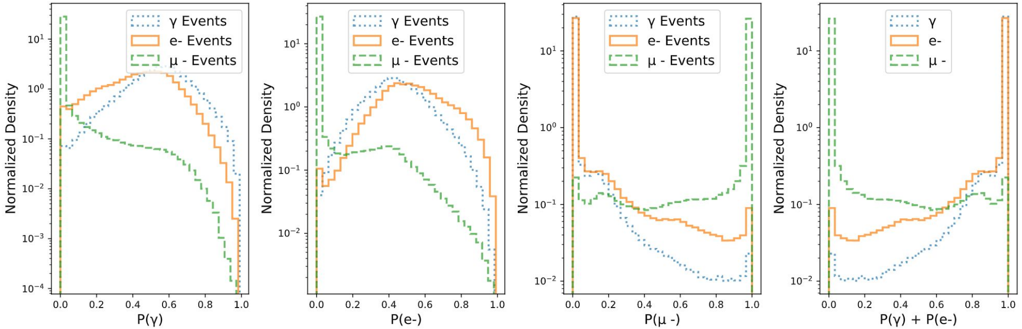




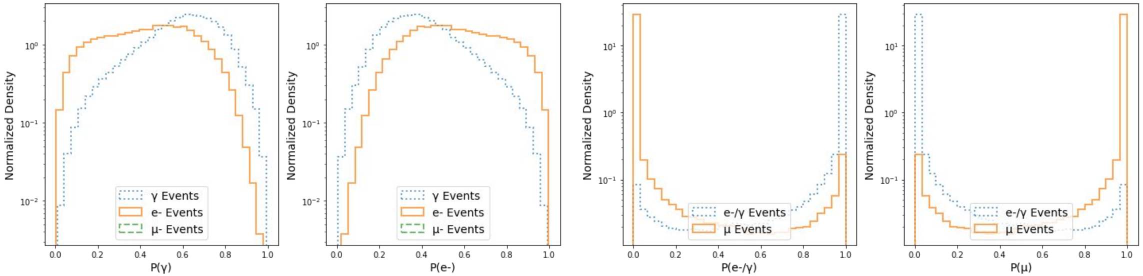
- To assess improvements obtained with 2-class models, the performance was compared with 3-class models
- Switching to a 2-class model did appear to yield performance improvements
    - Particularly for the e-/γ task

2-Class Performance Comparisons

- Some improved separation in e-/γ task



3 class model likelihood histograms

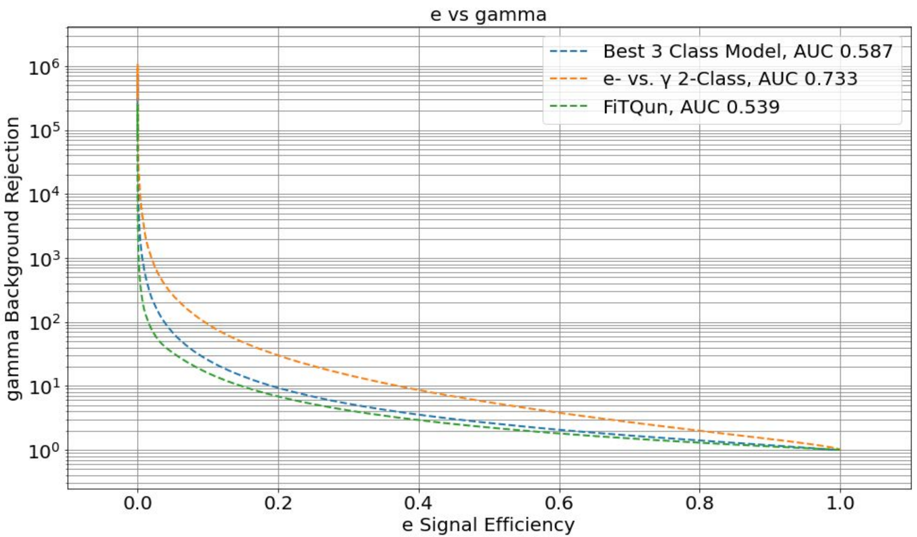


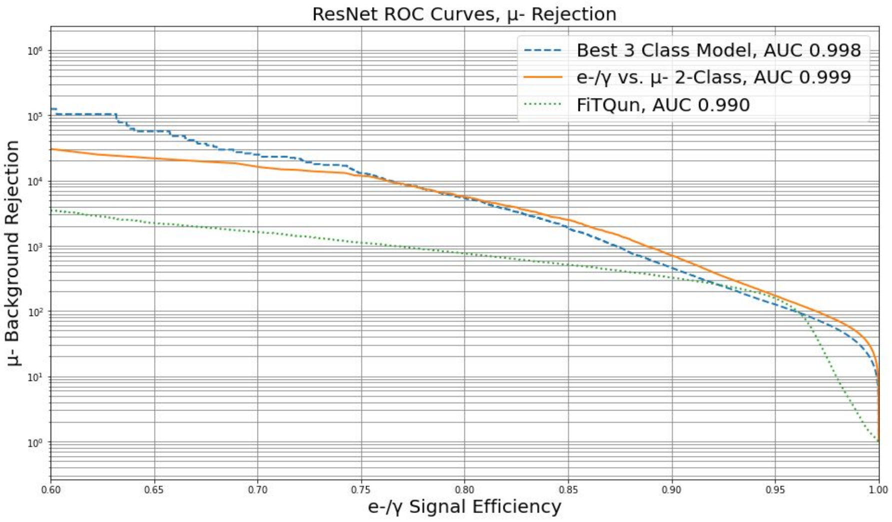
2 class model likelihood histograms

Additional Modifications...

- Training with weighted loss function
- Weight decay
- Some bugs found

Best Results



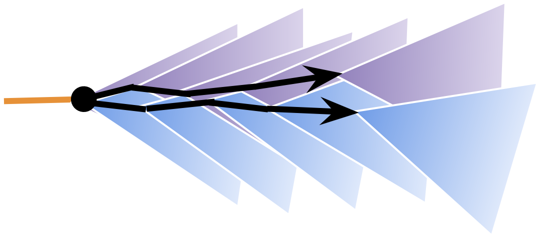


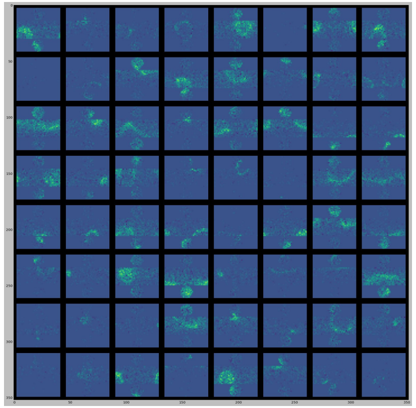
- Reduced overfitting does in fact lead to improved performance on the test set
- In combination with other improvements, yields substantial increase in performance over previous models

Next Steps

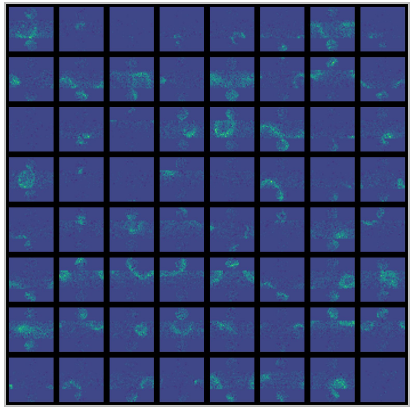
- Extending model to additional classes
  - Addition of 2 track π0s
- Mock statistical analysis
  - Mock analysis of MC data using model to characterize background
  - Understand effects of errors and systematics on future analysis
- Resnet based GANs
  - Some very preliminary work done (qualitative only)
  - Facilitate performance evaluating and understanding of systematics

π0
(2 tracks)





WCSim Simulated Data



GAN output

Summary

- Resnets show promise for event reconstruction tasks in water Cherenkov detectors
- 2 and 3 class ResNet models were trained on MC generated IWCD data to discriminate e-, γ and μ- events
- Performance on particle event classification seems comparable to existing likelihood based fitting algorithms
- In particular seems to show promise on the problem of e-/ γ separation

Thank you!
Questions?