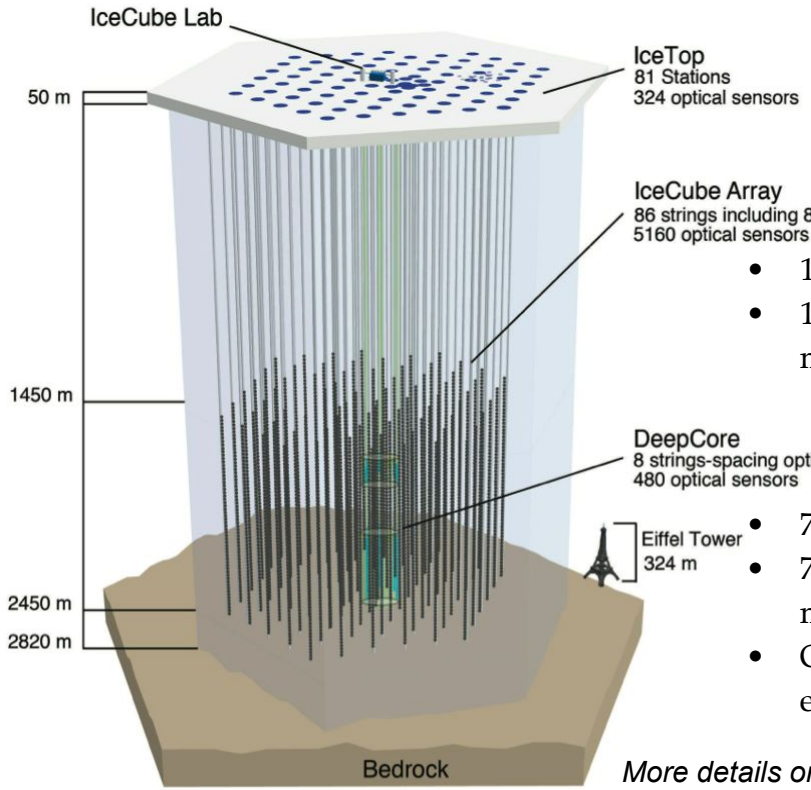
Optimizing a CNN to Reconstruct Low Energy IceCube Neutrino Events

Jessie Micallef
Michigan State University
micall12@msu.edu
NPML 2020

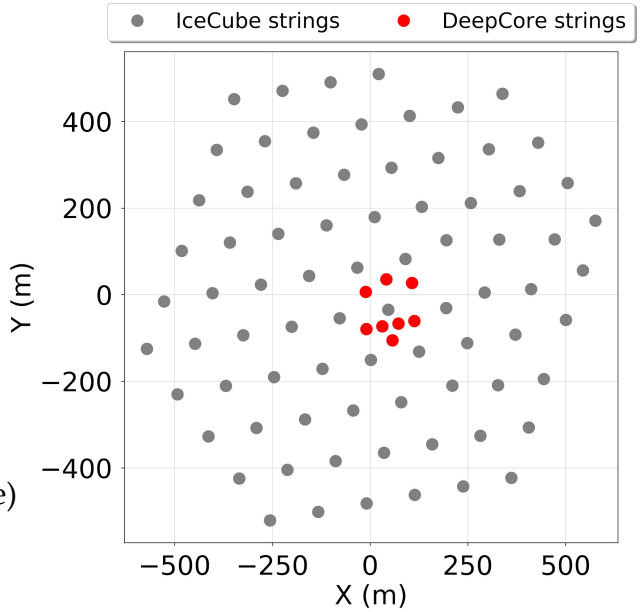
IceCube Neutrino Observatory



Goal of IceCube: Detect astrophysical and atmospheric neutrinos to determine their sources and measure neutrino characteristics, such as oscillation parameters.

- 125 meters between strings
- 17 meters between optical modules on string

- 72 meters between strings
- 7 meters between optical modules on string
- Optimized to detect lower energy events (10 GeV-scale)



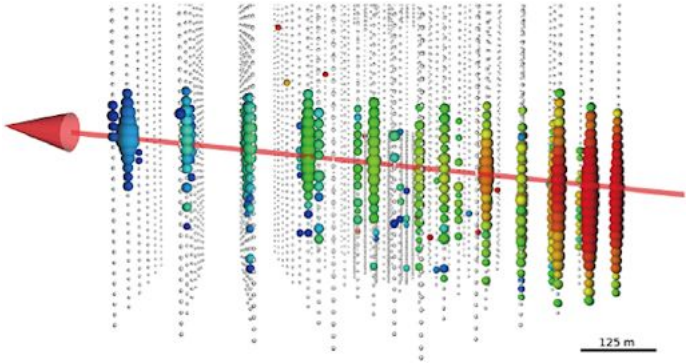
More details on DeepCore detector, see talk by Maria Prado Rodriguez

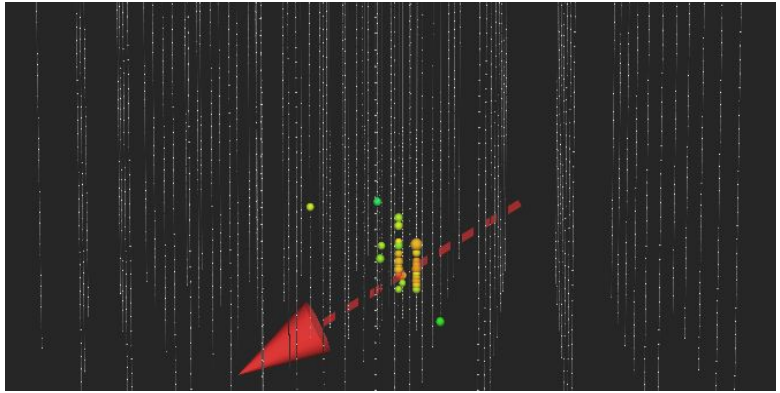
10 GeV-Scale Neutrino Events in IceCube

Why do we care about low energy neutrinos?

- Neutrino oscillation (flavor changes) observed at 10 GeV-scales in IceCube
- To measure oscillation parameters, need the neutrinos'...
  - Energies
  - Directions (as cosine zeniths)

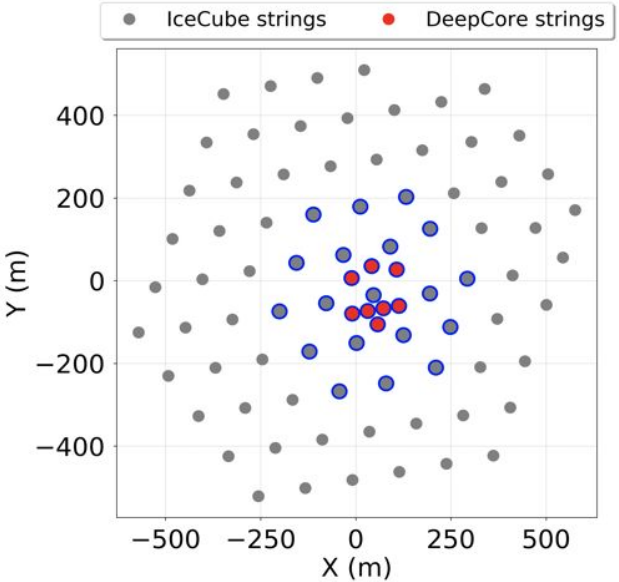
Goal: Use a CNN to reconstruct energy and cosine zenith for ν_μ CC (track-like) events



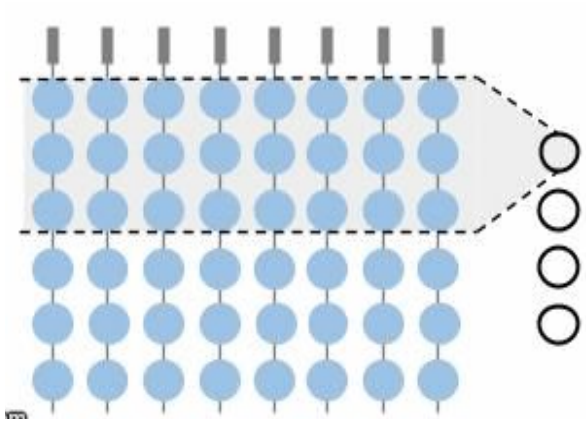


Applying CNN to Focus IceCube's DeepCore

- Focus on DeepCore using centermost strings for CNN (circled in blue)
- Per optical module approach: [string ID, depth ID, 5 summary variables]
- Inputs: 5 variables that summarize all pulses hitting optical module
  - sum of charge
  - time of first hit
  - time of last hit
  - charge weighted mean of times
  - charge weighted σ of times
- Split into two arrays
  - Different depths of modules for DeepCore vs IceCube

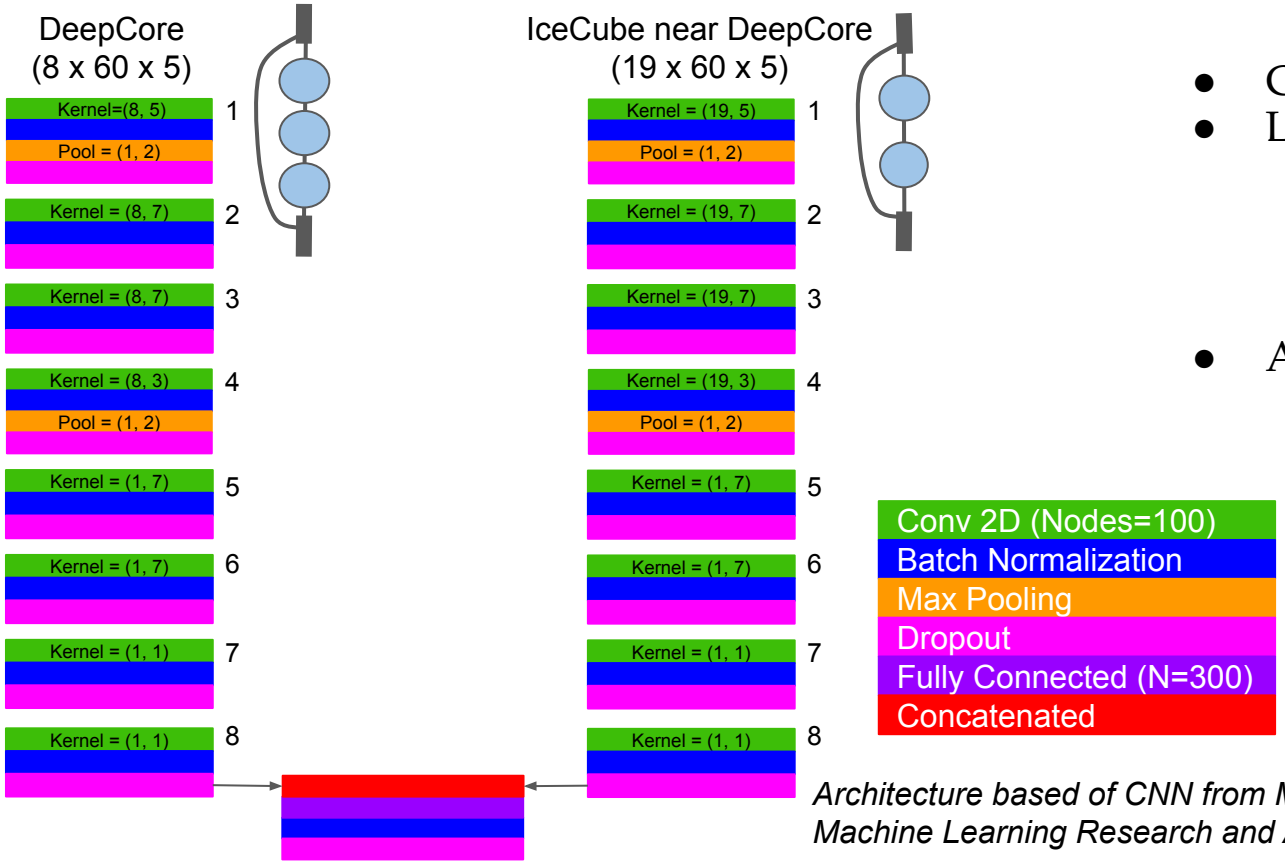


CNN on DeepCore Strings:



CNN kernel in depth going down optical modules

GeV-Scale CNN Architecture and Training Settings



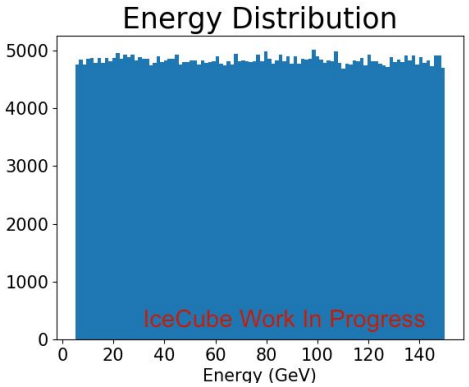
- CNN trained for **one** variable
- Loss optimized to variable
  - Energy: mean absolute percentage error
  - Cosine zenith: mean squared error
- Activation functions:
  - First layer uses tanh
  - Hidden layers use relu
  - Output layer uses linear

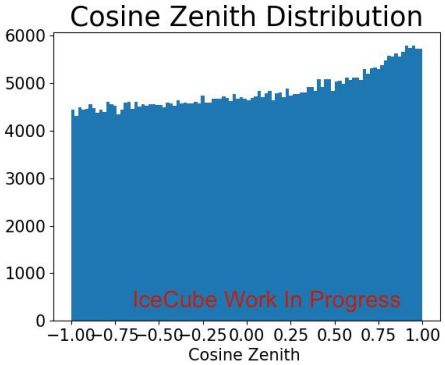
Architecture based of CNN from Mirco Huennefeld. See slides on Machine Learning Research and Applications in IceCube.

Training & Testing MC Samples

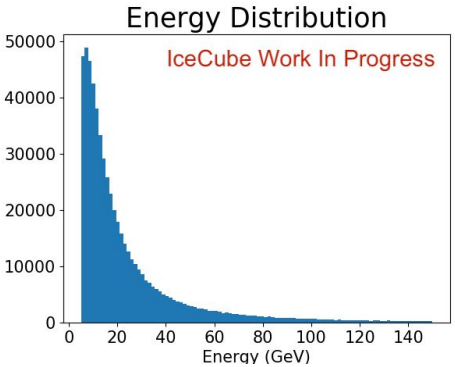
- Flat Training and Testing MC Sample
  - Energy and direction-independent for unbiased training
  - **Training from [5,150] GeV**
  - ~2.8 million events for training
  - ~400,000 events for testing
- Analysis-Like Testing MC Sample
  - Mostly low energy (< 20 GeV)
  - **Testing from [5,100] GeV**
  - ~600,000 events total for testing
- ν_μ charge current events only
- Starts in DeepCore and use hits within [-500,4000] ns of event start

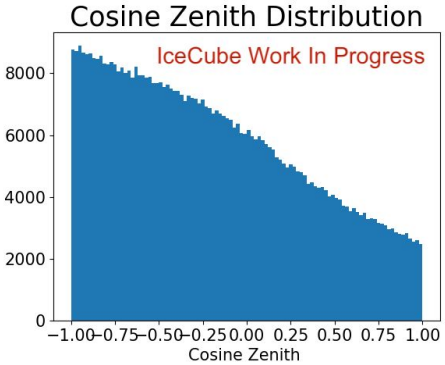
Flat Training & Testing Sample





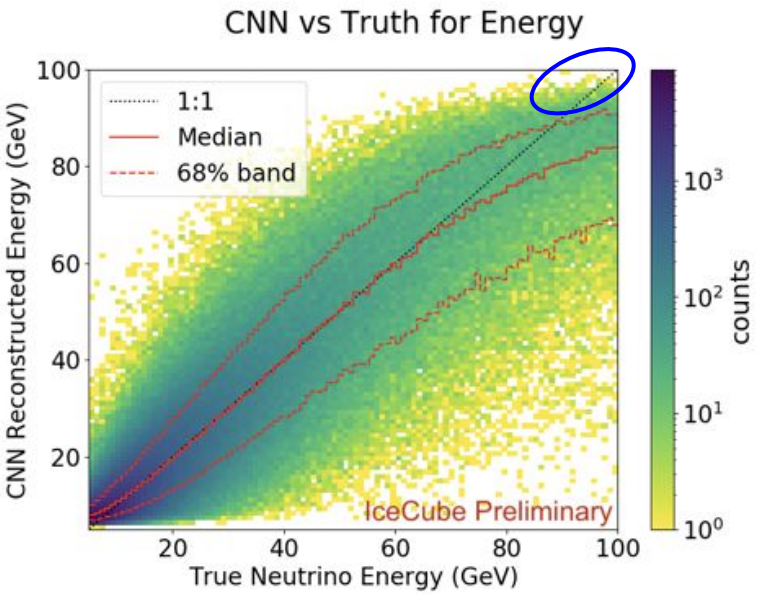
Analysis-Like Testing Sample





NPML Lightning Talk: Previous Performance

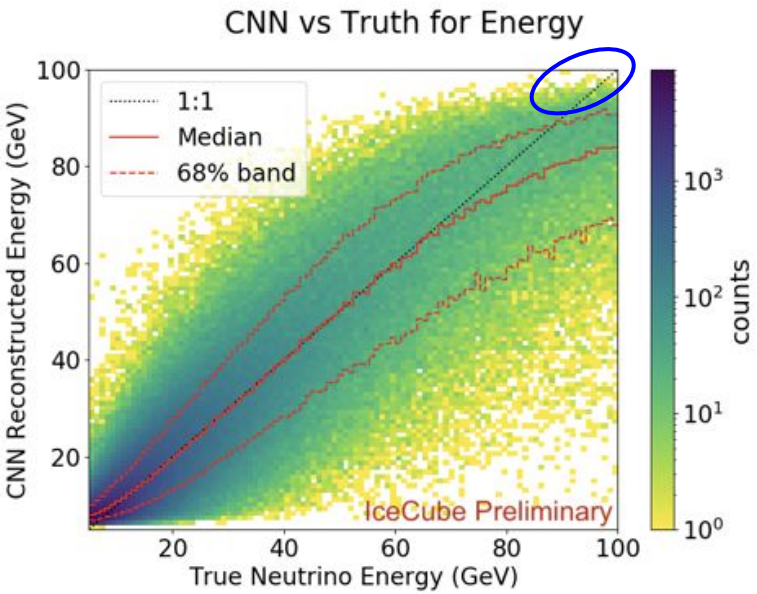
Old Model:
Energy Range 5-100 GeV



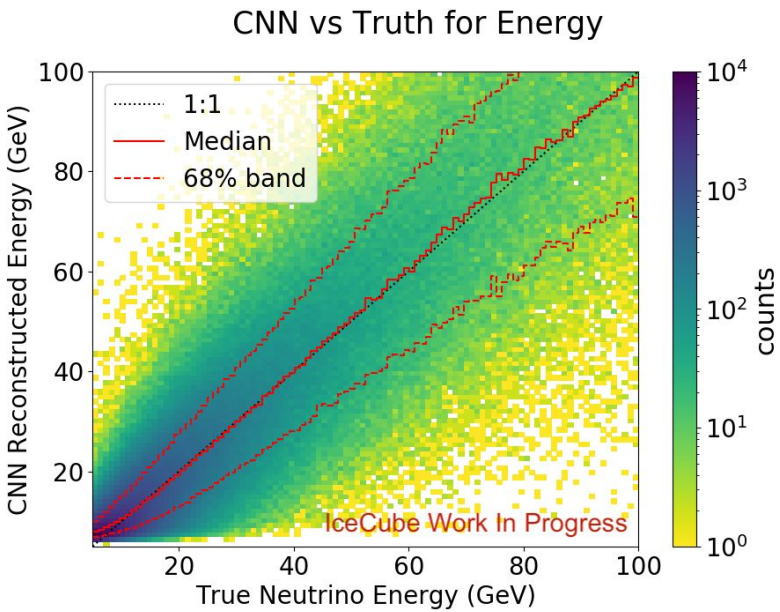
- [Presented poster at NPML Lightning]
- Energy CNN under reconstructing high energy events (> 80 GeV)
- Future work was to extend energy range when training → see next slide

CNN Improves with Extended Training Energy

Old Model:
Energy Range 5-100 GeV

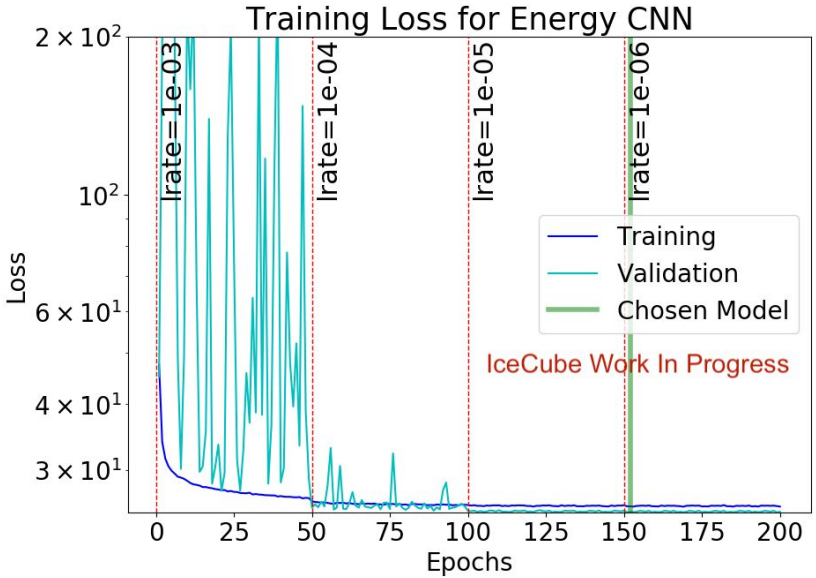


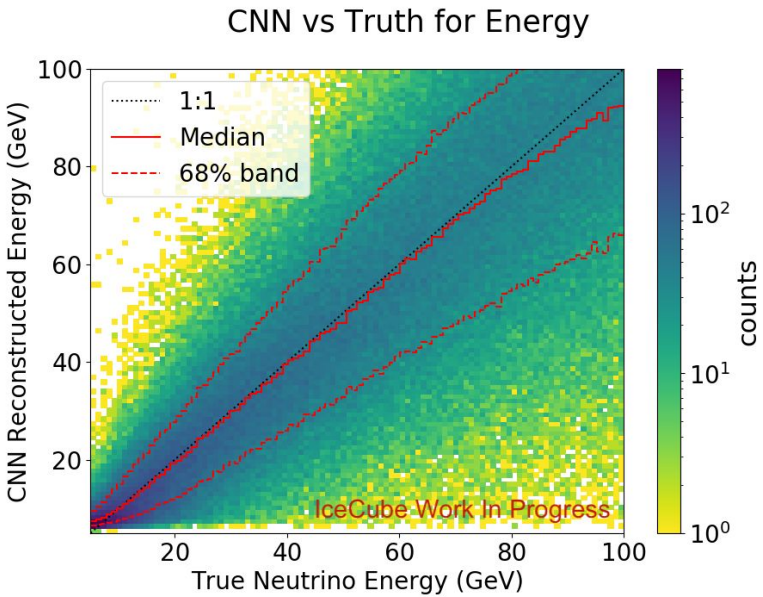
Extended Energy Model:
Energy Range 5-150 GeV



Training and Testing on Flat Sample: Energy CNN

→ Adjusting learning rate while training helps validation converge
→ CNN reconstructs well across entire flat energy distribution sample
→ Most difficulty at high energy, but median reconstruction near 1:1

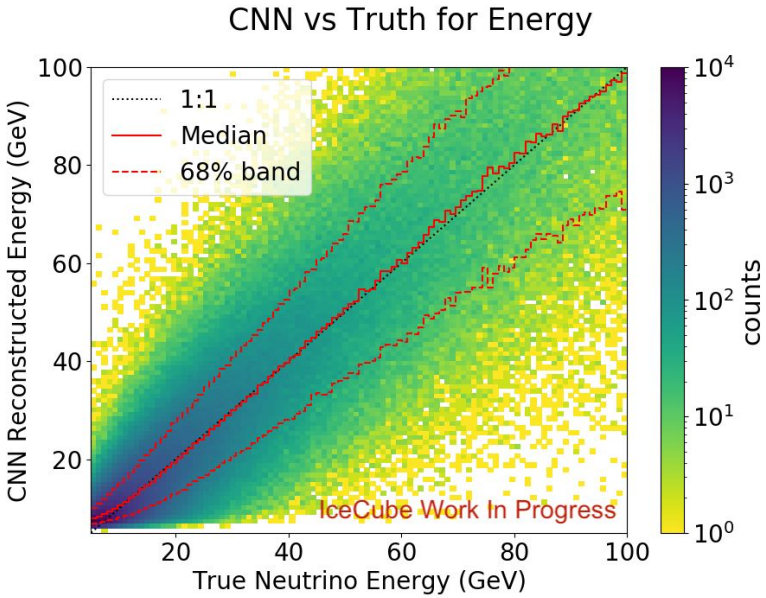


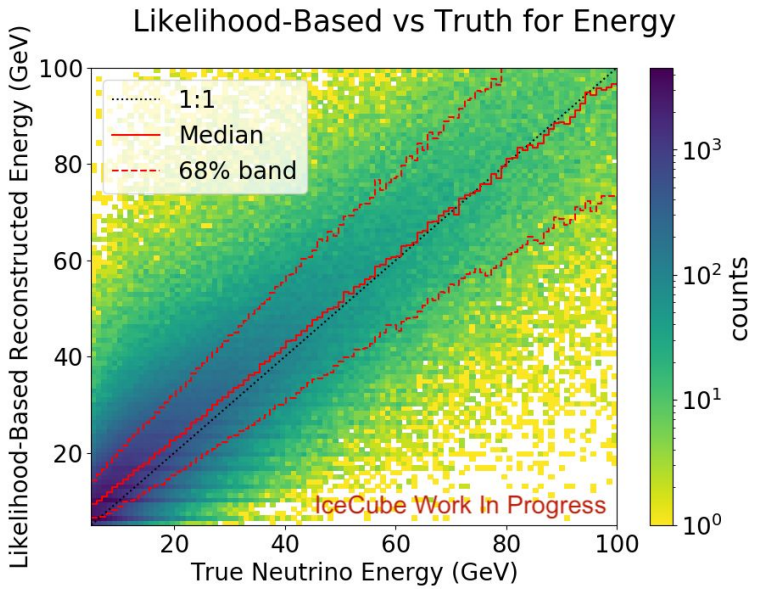


→ Loss calculated using mean absolute percentage error

Testing on Analysis-Like Sample: Energy CNN

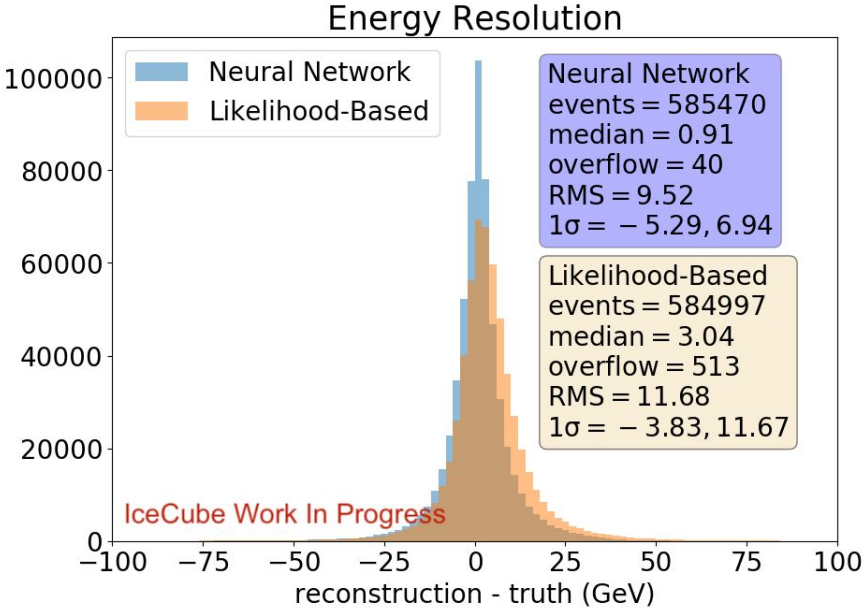
→ CNN resolution follows 1:1 across all energies
→ Some difficulty at low energy, but no worse than likelihood-based method
→ Recall sample has events with mostly < 20 GeV

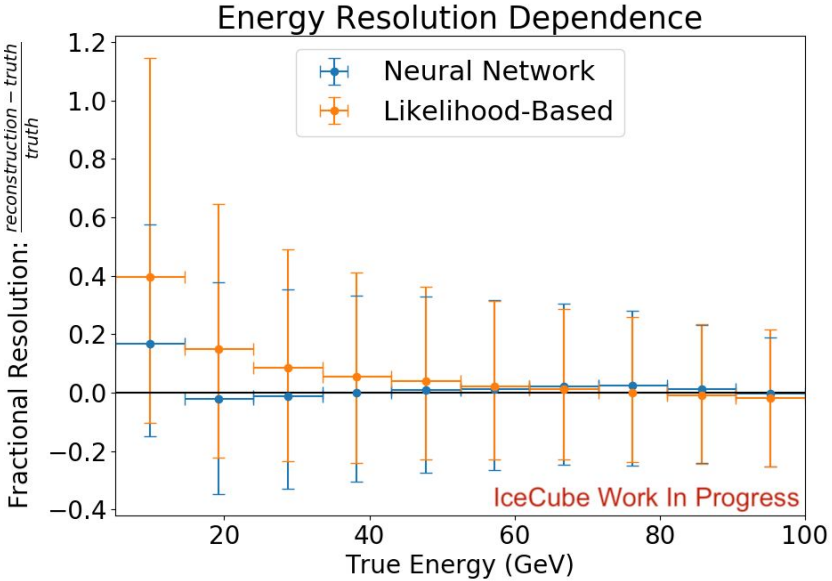




Comparing to Analysis-Like Sample: Energy CNN

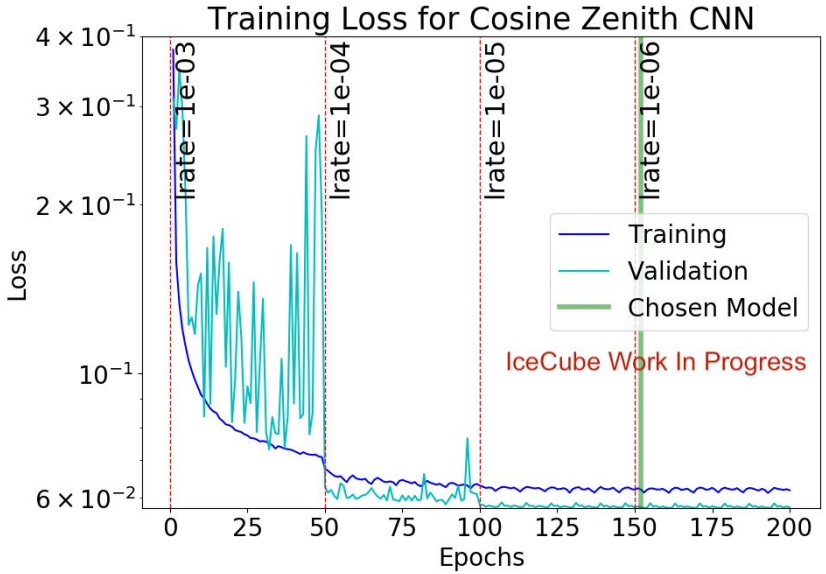
→ CNN has minimal bias at lowest energies, compared to previous method
→ Recall sample has events with mostly < 20 GeV
→ CNN takes O(milliseconds) vs. Likelihood-Base takes O(minutes)



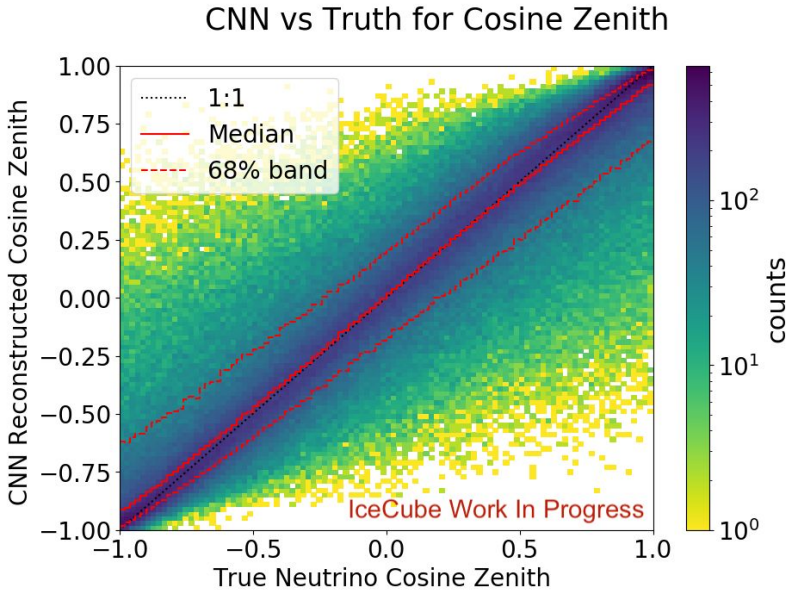


Training and Testing on Flat Sample: Direction CNN

→ Adjusting learning rate while training helps validation converge
→ CNN reconstructs well across entire flat cosine zenith distribution sample
→ Most difficulty reconstructing at boundaries

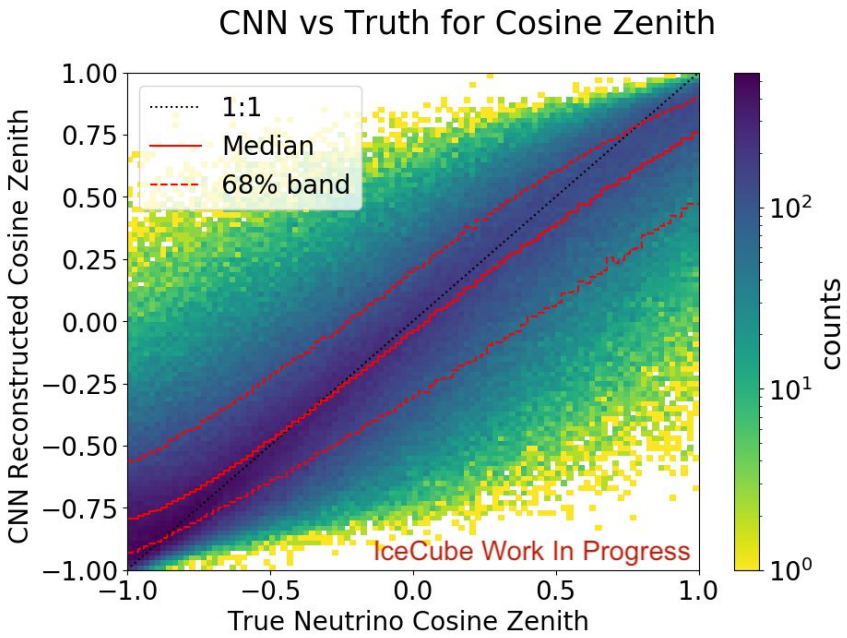


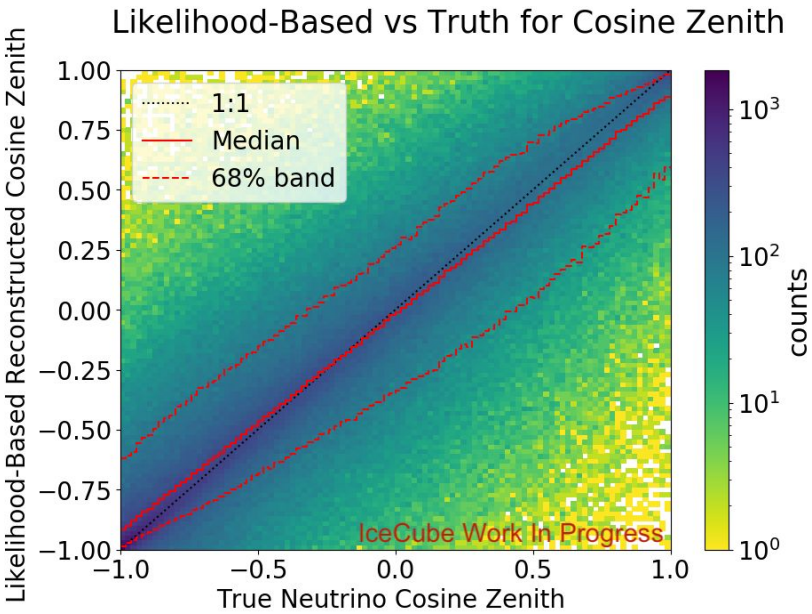
→ Loss calculated using mean squared error



Testing on Analysis-Like Sample: Direction CNN

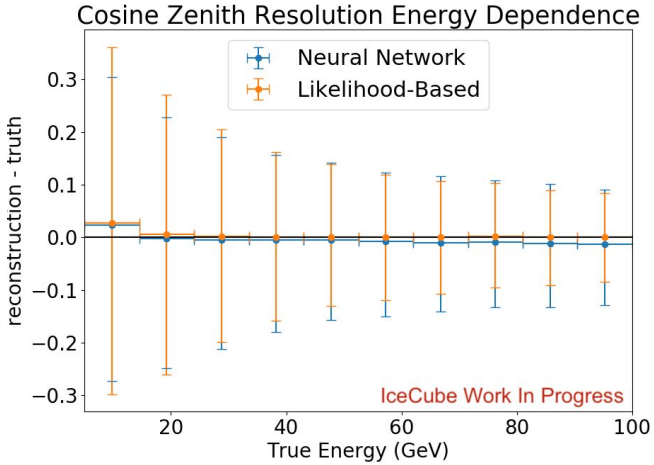
→ CNN resolution spread slightly is more narrow
→ CNN has most difficulty at boundaries

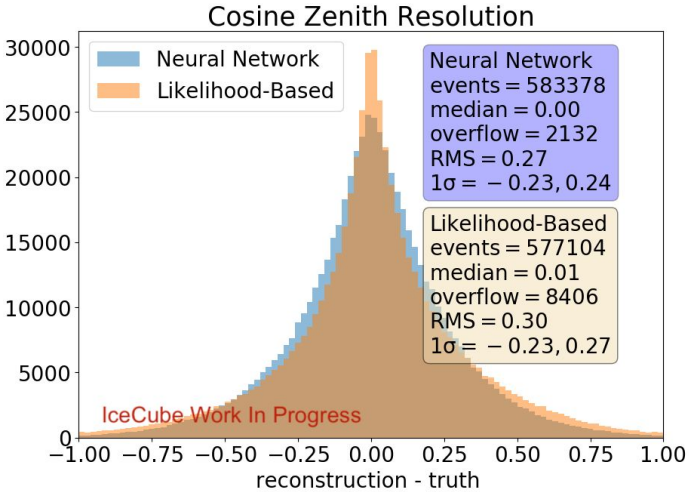


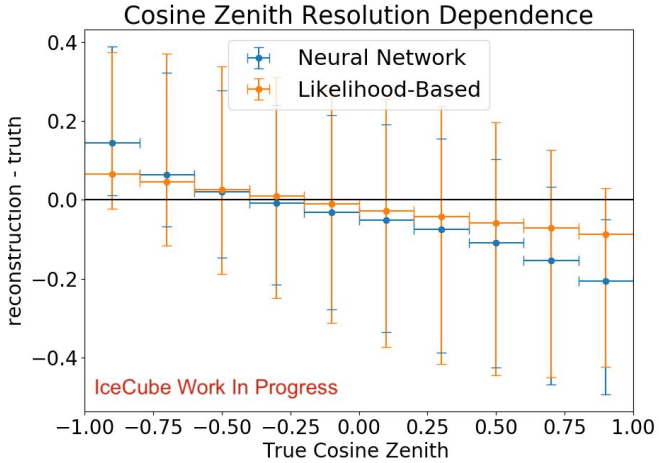


Comparing to Analysis-Like Sample: Direction CNN

→ CNN resolution very comparable
→ Move to reconstructing dx, dy, dz since can't extend cosine boundary?
→ CNN is 10^6 times faster!







Conclusions

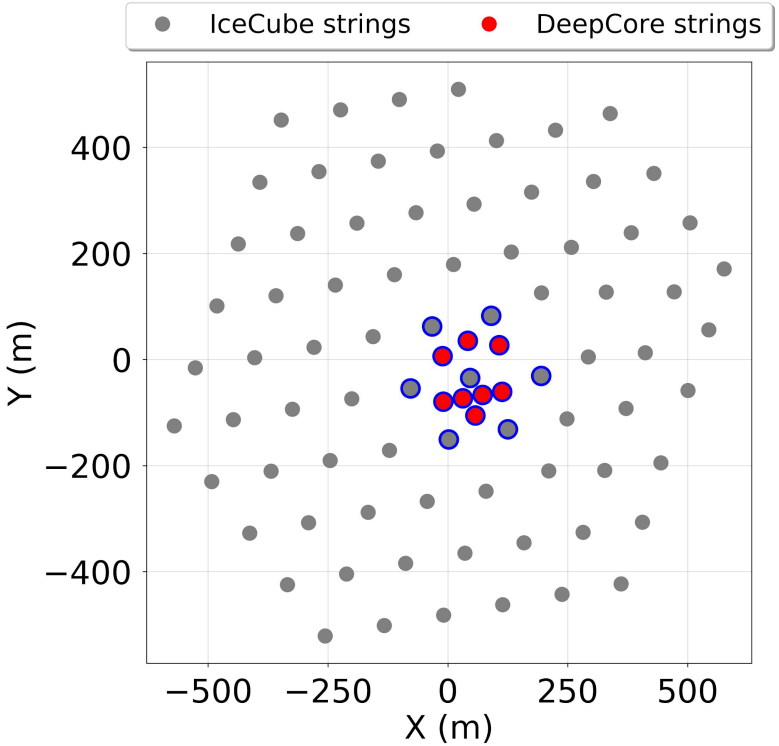
- The GeV-scale CNN shows comparable resolution to previous methods for track-like ν_μ events
  - Extending energy range for training sample resolved high energy boundary condition effects shown previously
  - Adaptive learning rate helps network converge
- Reconstruction speed 10^6 times faster than previous methods!
- Future work:
  - Improve resolution with CNN
    - Add more MC training data
    - Train on vector components for angle to improve boundary effect
  - Apply training and reconstruction to other neutrino flavors

Acknowledgements:
"This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE1848739. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation."

Backup

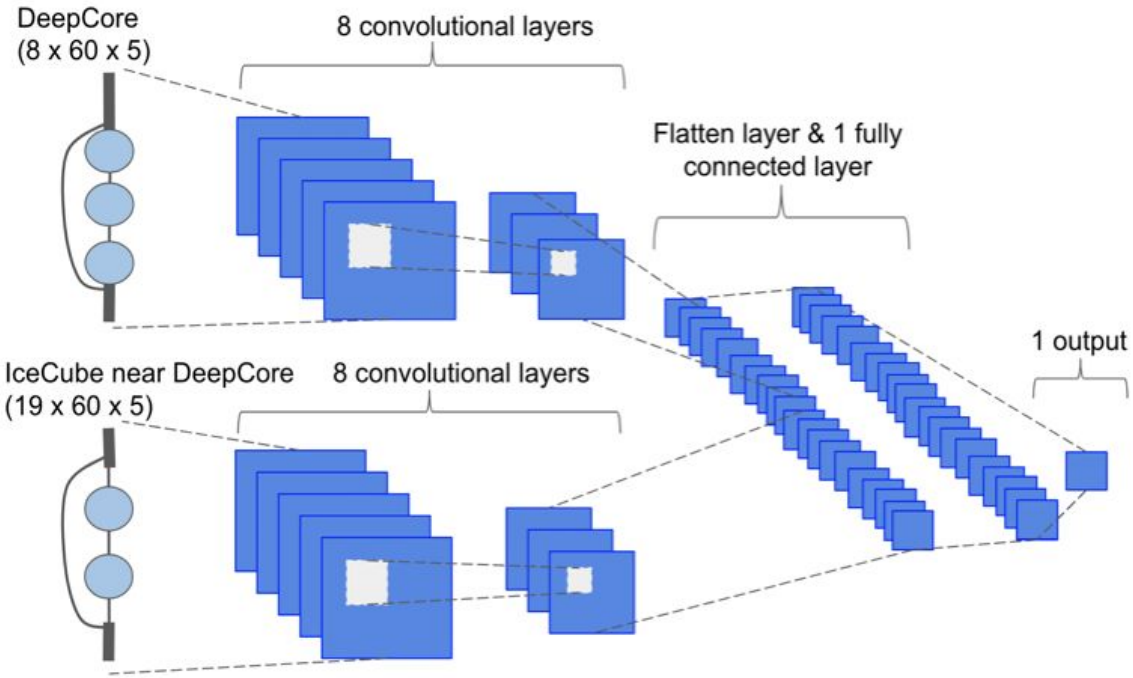
GeV-Scale CNN Setup

- CNN uses per-optical module approach
  - Summarize all pulses that hit each module
  - Total of 5 charge and time based variables
- Two subarrays: one for DeepCore (DC) strings and one for IceCube strings near DeepCore (IC near DC)
  - Different depths of optical modules
  - Structure = [String x DOM x variables]
  - DC = [8 x 60 x 5]
  - IC near DC = [7 x 60 x 5]



Based off Mirco Huennefeld's DNN for high energy IceCube neutrino events:
https://icecube.wisc.edu/~mhuennefeld/DNNreco/wikipage/material/MasterThesis.pdf

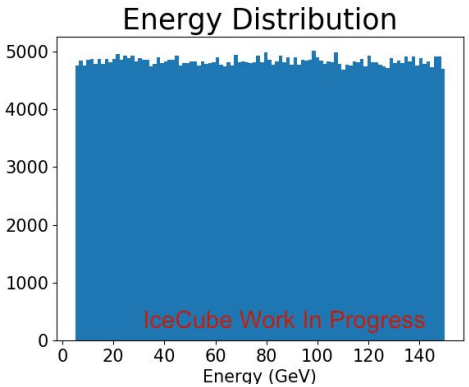
GeV-Scale CNN Architecture and Training Settings

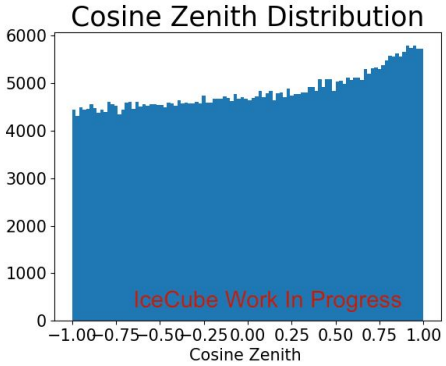


- CNN trained for **one** variable
- Loss optimized to variable
  - Energy: mean absolute percentage error
  - Cosine zenith: mean squared error
- Learning rate changes
  - Starting at 0.001
  - By 0.1
  - Every 50 epochs
- Applied every layer…
  - Dropout: 0.20
  - Batch normalization

Training Sample: Energy-Independent MC

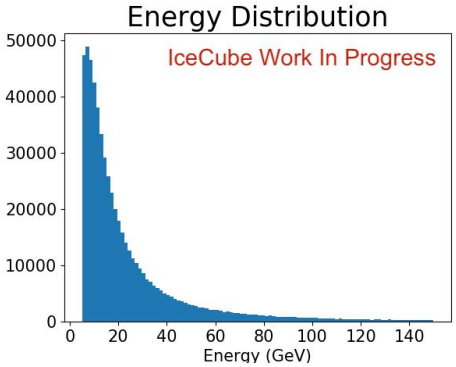
- Use energy-independent MC so that CNN trains without bias in energy
  - Flat energy distribution from [5, 100] GeV
  - Goal: [5, 60] GeV for oscillation analysis
- Use cosine zenith (flat distribution) for unbiased training
- ν_μ charge current events only (track-like signature)
- Starts in DeepCore and use hits within 4000 ns
- Train network with ~2.8 million events
  - Validate with ~700,000
  - Test with ~400,000 events

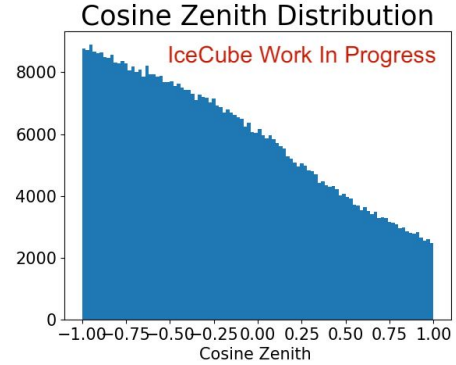




Testing Sample: From Previous Study

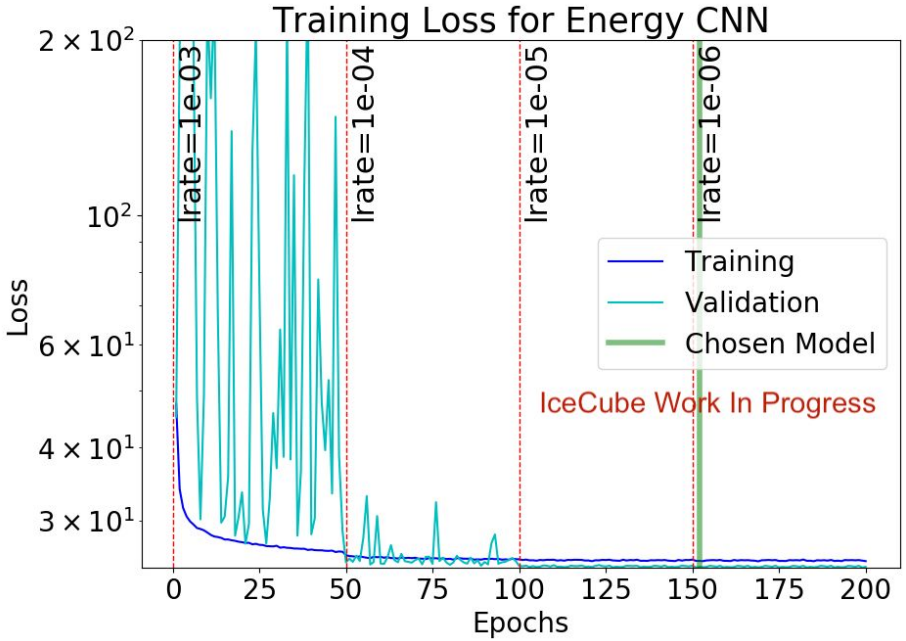
- MC sample generated for 6-56 GeV Atmospheric Neutrino Oscillation paper[1]
- Mostly low energy (< 20 GeV)
  - Focus on energy range [5, 60] GeV
  - Goal: [5, 60] GeV for oscillation analysis
- Cosine zenith has more upgoing (negative values) due to background cuts
- ν_μ charge current events only (track-like signature)
- Starts in DeepCore
- ~585,000 events matching criteria





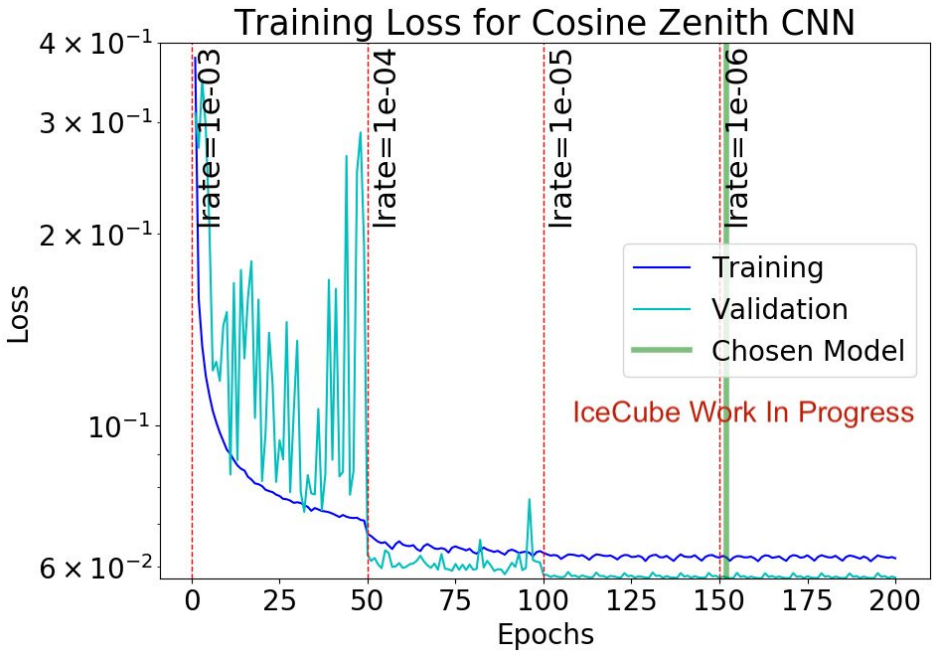
[1]Mark Aartsen *et al.* 2018. Measurement of Atmospheric Neutrino Oscillations at 6-56 GeV with IceCube DeepCore. *PHYS REV LETT*. Vol. 120, Iss. 7 (Feb. 2018). DOI: hptts://doi.org/10.1103/PhysRevLett.120.071801

Energy CNN Performance



5 epochs is 1 full pass through training data. Chosen model at 152 epochs = 19 full passes through data. Training set: 2,783,200 events. Validation set: 695,800 events. Testing set: 386,000 events.

Direction CNN Performance



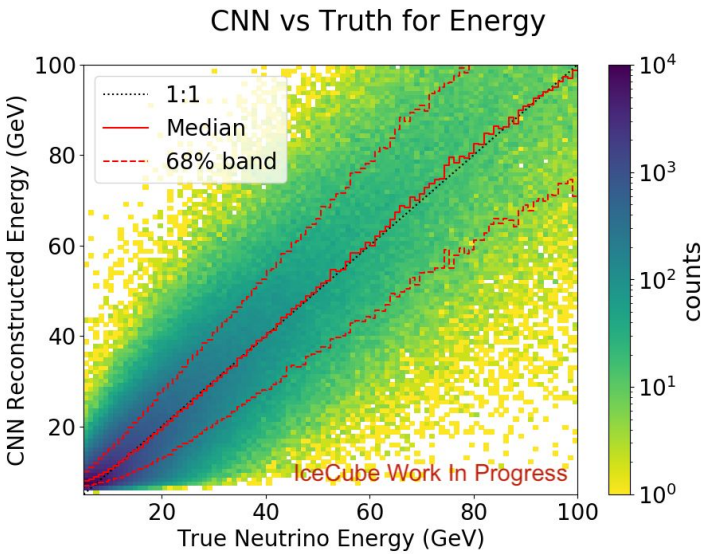
5 epochs is 1 full pass through training data. Chosen model at 152 epochs = 19 full passes through data. Training set: 2,783,200 events. Validation set: 695,800 events. Testing set: 386,000 events.

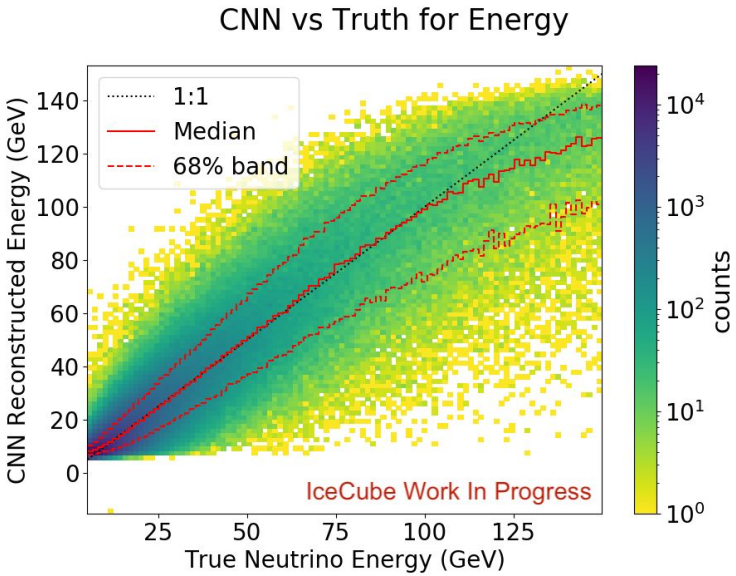
Unbounded plot: Testing up to 150 GeV

Sample Trained on

NuMu CC - **Energy**

Plot on left and in the middle are tested with the same network model, except the plot is bounded on the left to only show up to 100 GeV. Boundary condition still clear near edge of testing energy (near 150 GeV)





Unbounded plot

Sample Trained on

NuMu CC - **Cosine Zenith**

Plot on left and in the middle are tested with the same network model, except the plot is bounded on the left to only show up to [-1,1].