Inferring Convolutional Neural Networks' accuracy from its architectural characterizations

Duc Hoang – Rhodes College

Supervisor: Dr. Gabriel N. Perdue

SIST – Final Presentation

1 August 2019

Outline

I. MINERvA experiment

a. Neutrinos and MINERvA detector
b. Vertex finding and hadron multiplicity problem

II. Deep Learning

a. Deep Neural Networks
b. Convolutional Neural Networks (CNNs)
c. CNN's design difficulties

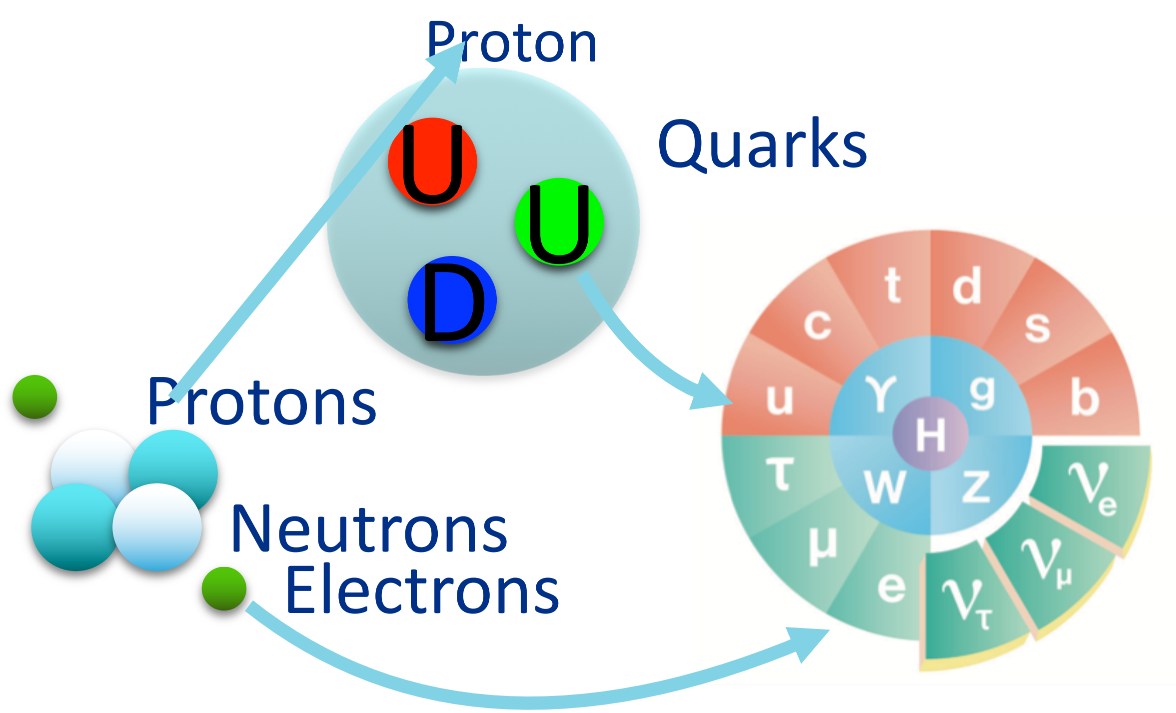
III. Inferring CNNs' accuracy before training time

a. Architectural characterizations
b. Predicting CNNs' accuracy based on characterizations – why is it useful?

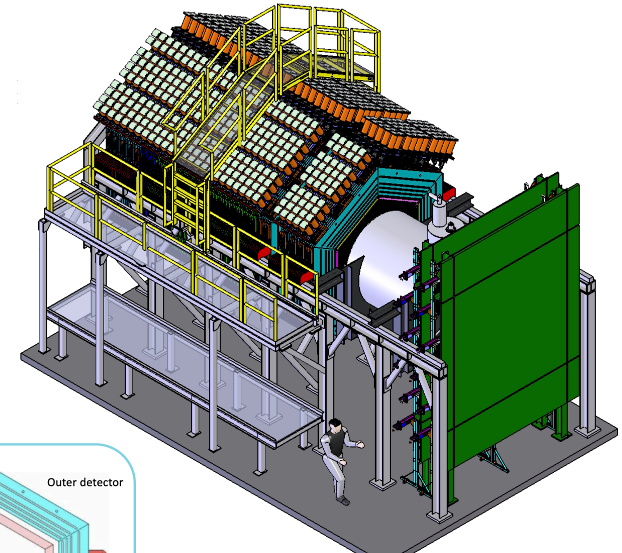
IV. Summary & Outlook

Neutrinos

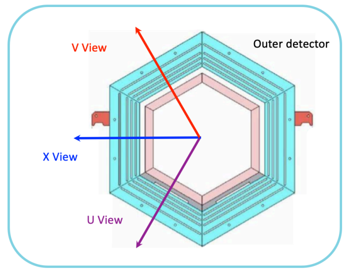
- Neutrinos are **fundamental**.
- They are **electrically neutral "partners"** of the familiar charged leptons (e.g., electrons).
- They are **very light,**
- **They very rarely interact** with other particles



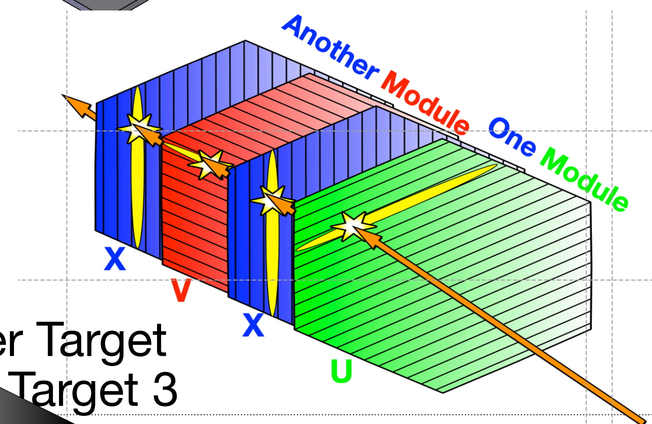
MINERvA

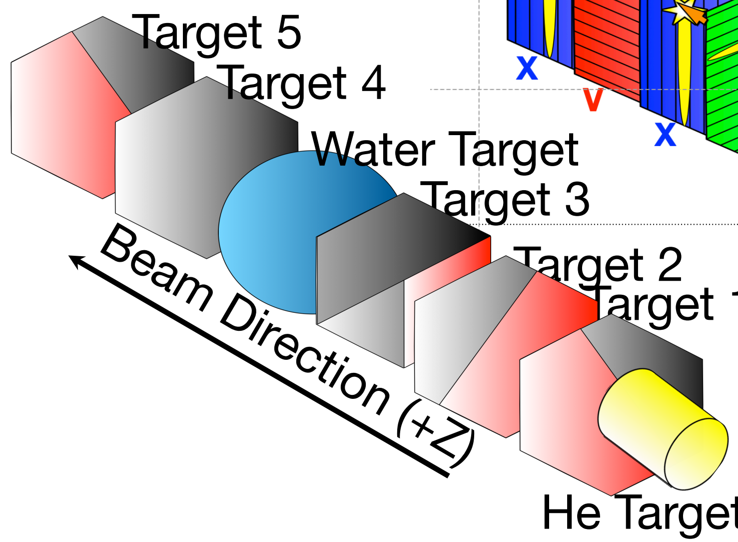


- **Nuclear effects with a variety of target materials ranging from Helium to Lead.**

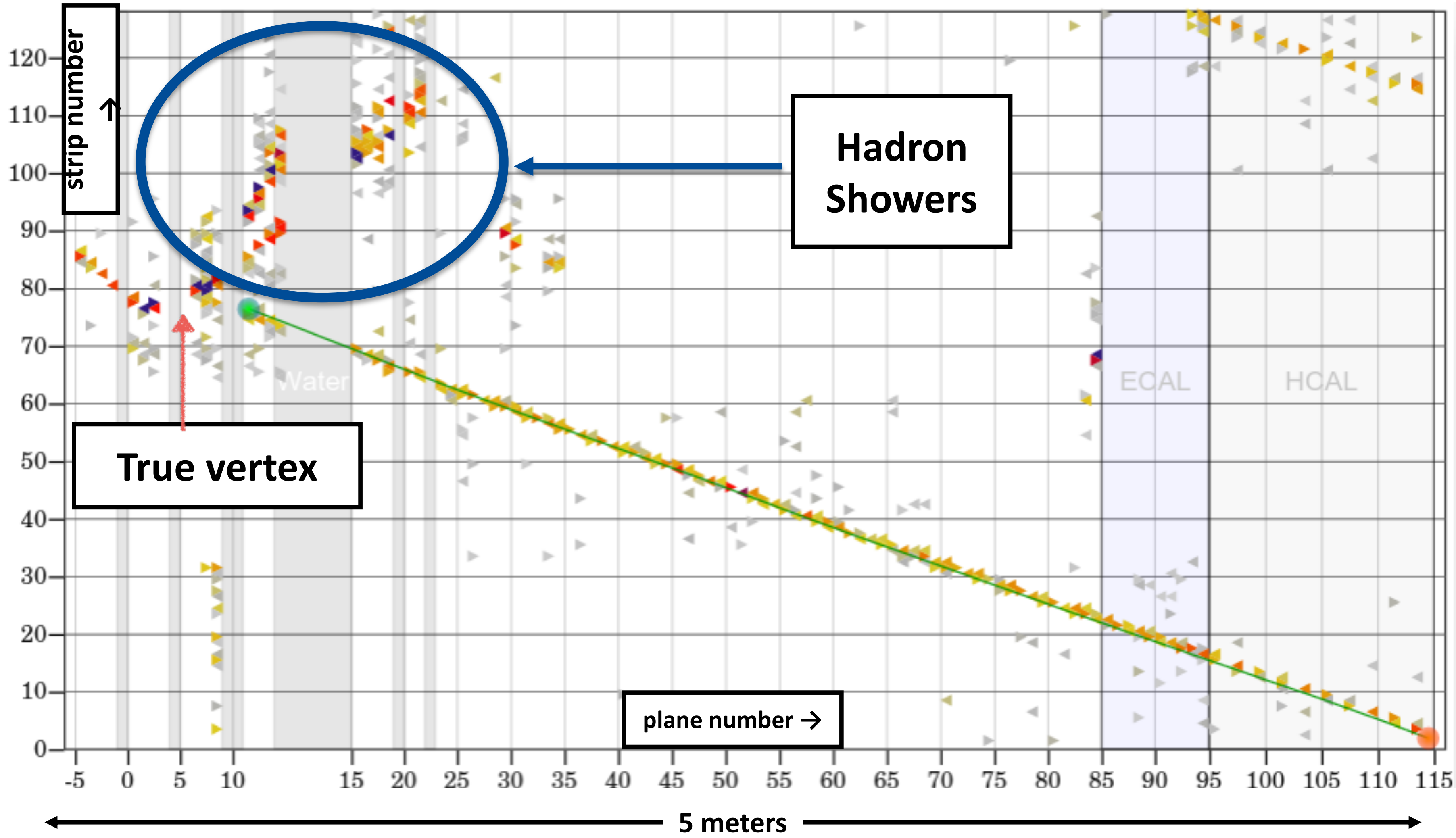


- **Fine-grained resolution for excellent kinematic measurements.**



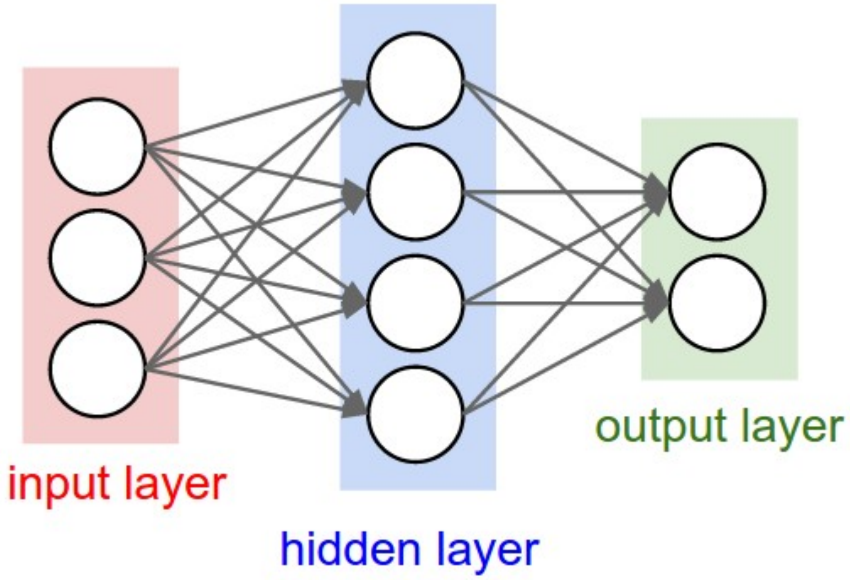


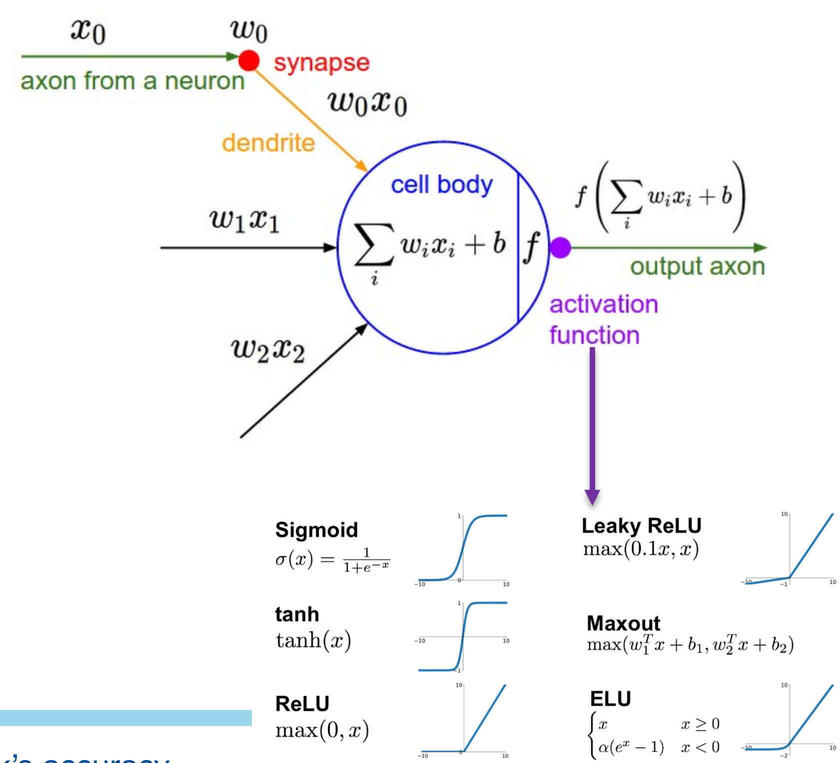
Vertex Finding and Hadron Multiplicity problem



Deep Neural Networks

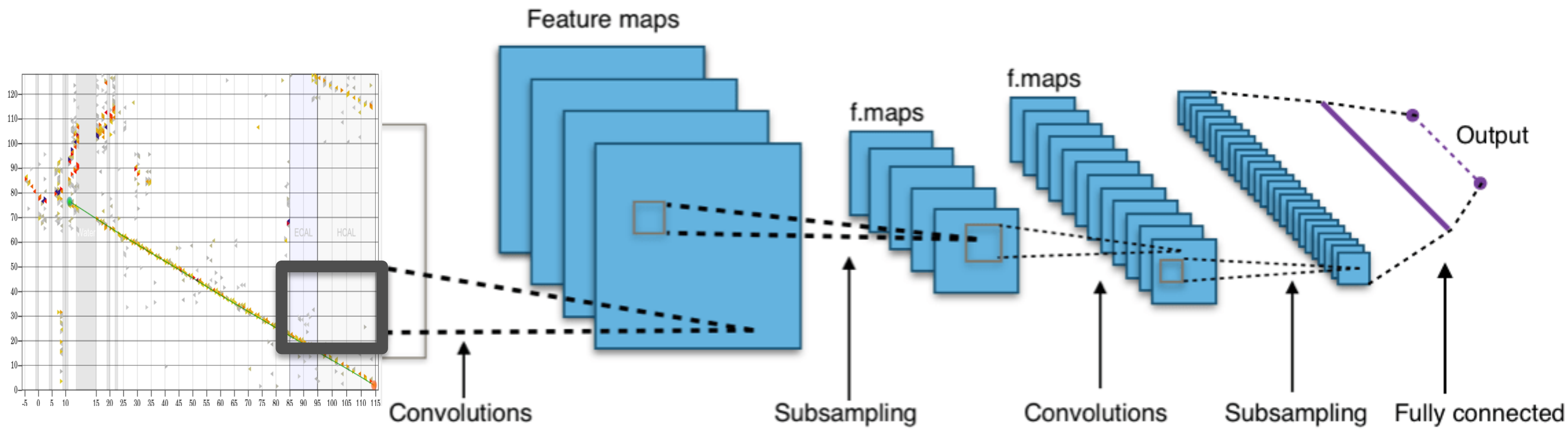
- Fully-connected architecture
- Each **input** multiplied by a **weight**.
- **Weighted values** are summed, **Bias** is added.
- Non-linear **activation function** is applied
- Trained by varying the **parameters** to minimize a loss function (quantifies how many mistakes the network makes)

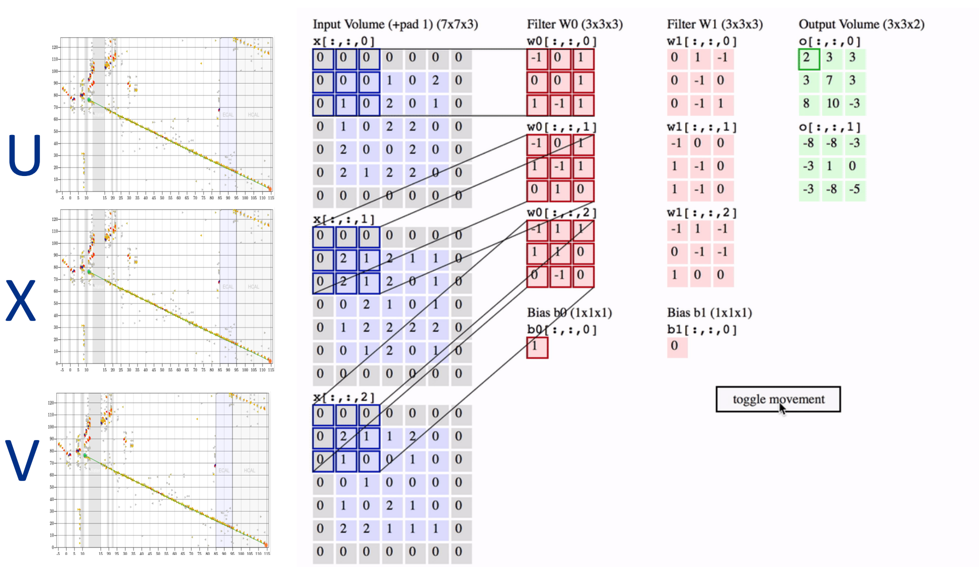




Convolutional Neural Networks (CNNs)

- Similar concept to Deep Neural Networks, but highly effective for **image inputs,** and modern neutrino detectors are **imaging detectors.**

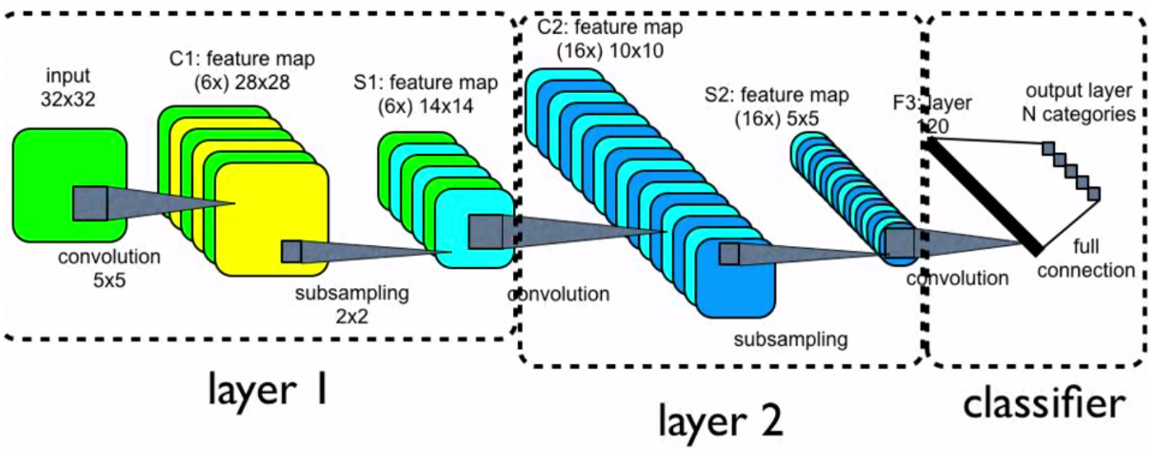




Fermilab

Difficulty

- There is no universal CNN design for every tasks.
- And designing an appropriate structure/architecture of CNN takes a lot of time and effort even for the experts.
- There is no systematic way to design CNNs: mainly rely on human intuition and random/grid search.
- Computationally expensive to train a CNN model.



Fermilab

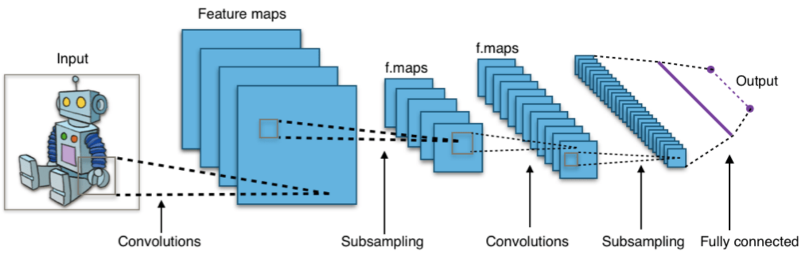
Objectives

I. Propose a **systematic language** to characterize CNN's architecture, and demonstrate that they can be **predictive** of a CNN's accuracy.

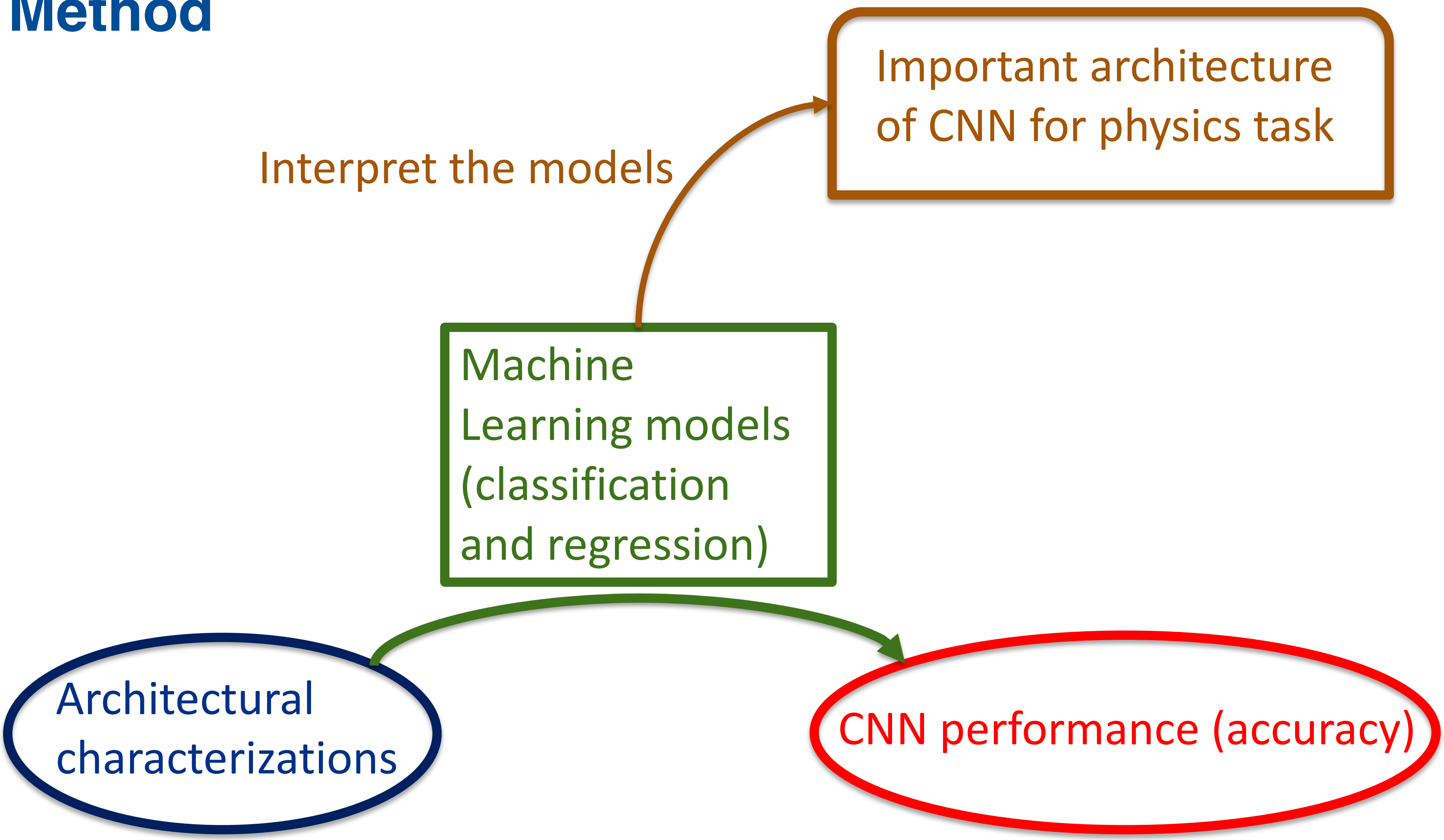
II. Suggest **architectural changes** to CNNs for different physics tasks (vertex finding and hadron multiplicity)

Examples of architectural attributes we extracted (32 in total):

- Number of convolutional layers.
- Number of rectified linear unit (ReLU) activated convolutional layers.
- Average depth

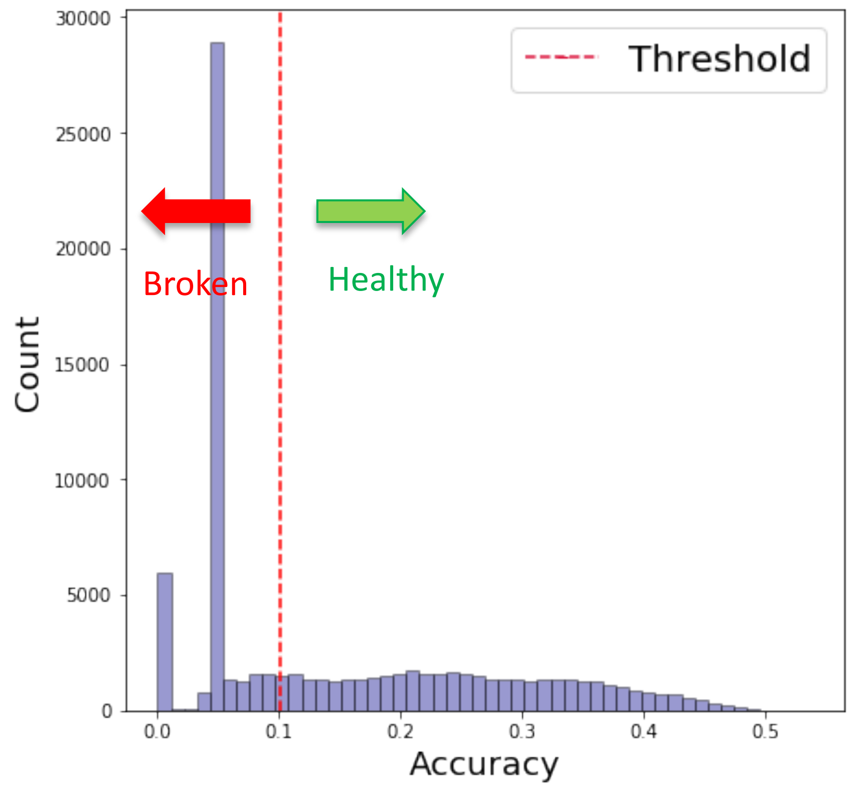


Method


Important architecture of CNN for physics task
Interpret the models
Machine Learning models (classification and regression)
Architectural characterizations
CNN performance (accuracy)

Classification

- Divide data set into ***broken*** and ***healthy*** networks.
- Use **Random Forest** and **Extremely Randomized Tree** to predict each CNN's class.

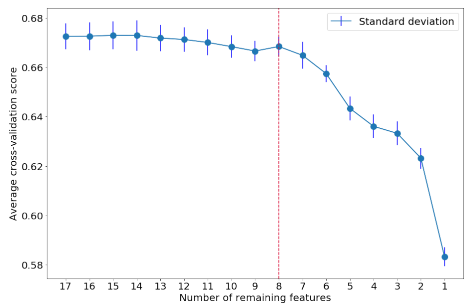


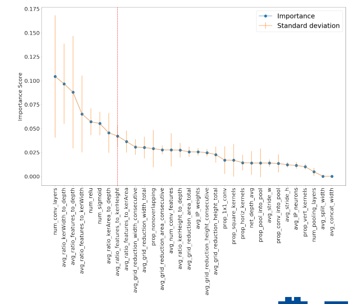
Classification results

- Machine Learning models perform **significantly better** than random guessing (50% when there is no class overflow):

Model	Average 5-fold cross-validation scores	Accuracy on test set
Random Forest	70.3 ± 0.006 %	70.6 %
Extremely Randomized Tree	70.2 ± 0.003 %	70.5 %

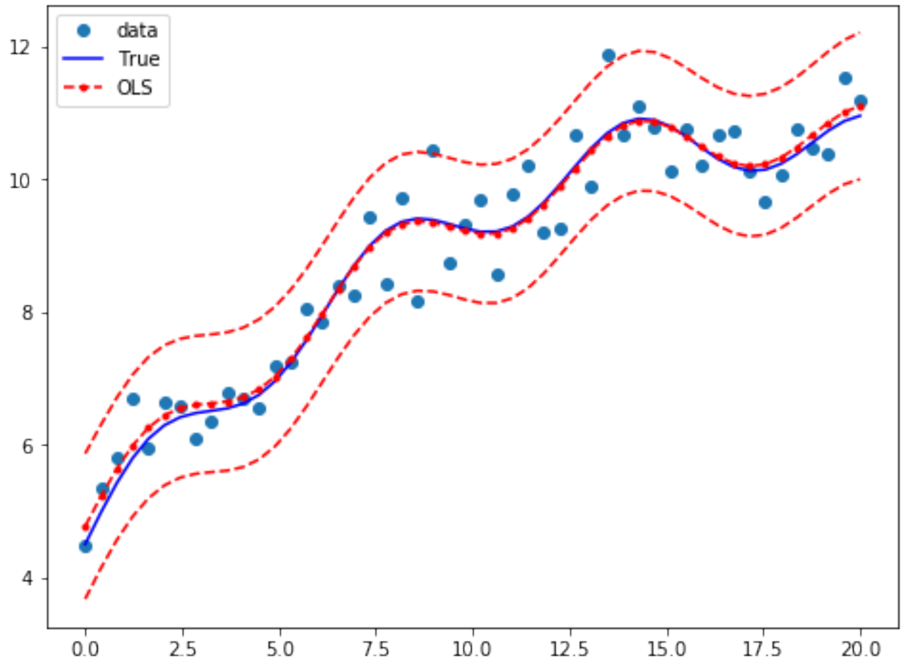
- Were also able to extract important features to suggest architectural changes to CNNs. More details in paper.





Regression on just healthy networks

- Fitted an non-linear Ordinary Least Square model on just the ***healthy*** networks.



Regression results

- Models fitted on two populations of vertex finding CNNs:

Population	R^2 of OLS	Number of *Healthy CNNs*
First	0.426	49276
Second	0.295	21415
Combined	0.961	70691

- **Limitations:** Still not enough parameters to characterize the detailed relationship between CNN's architecture and its performance. ▶ Planning to extend attribute set in the future.

Outlook & Summary

- Proposed a systematic language to characterize convolutional neural networks architecture.
- Successfully demonstrated that we can use those parameters to predict whether a network is “good”.
- There are limitations to predict the exact accuracy, but initial results are promising. Extension of the attributes set might help in the future.
- One of the early studies about relationship between CNN’s architecture and its performance.
- More details in our up-coming paper.

Acknowledgements

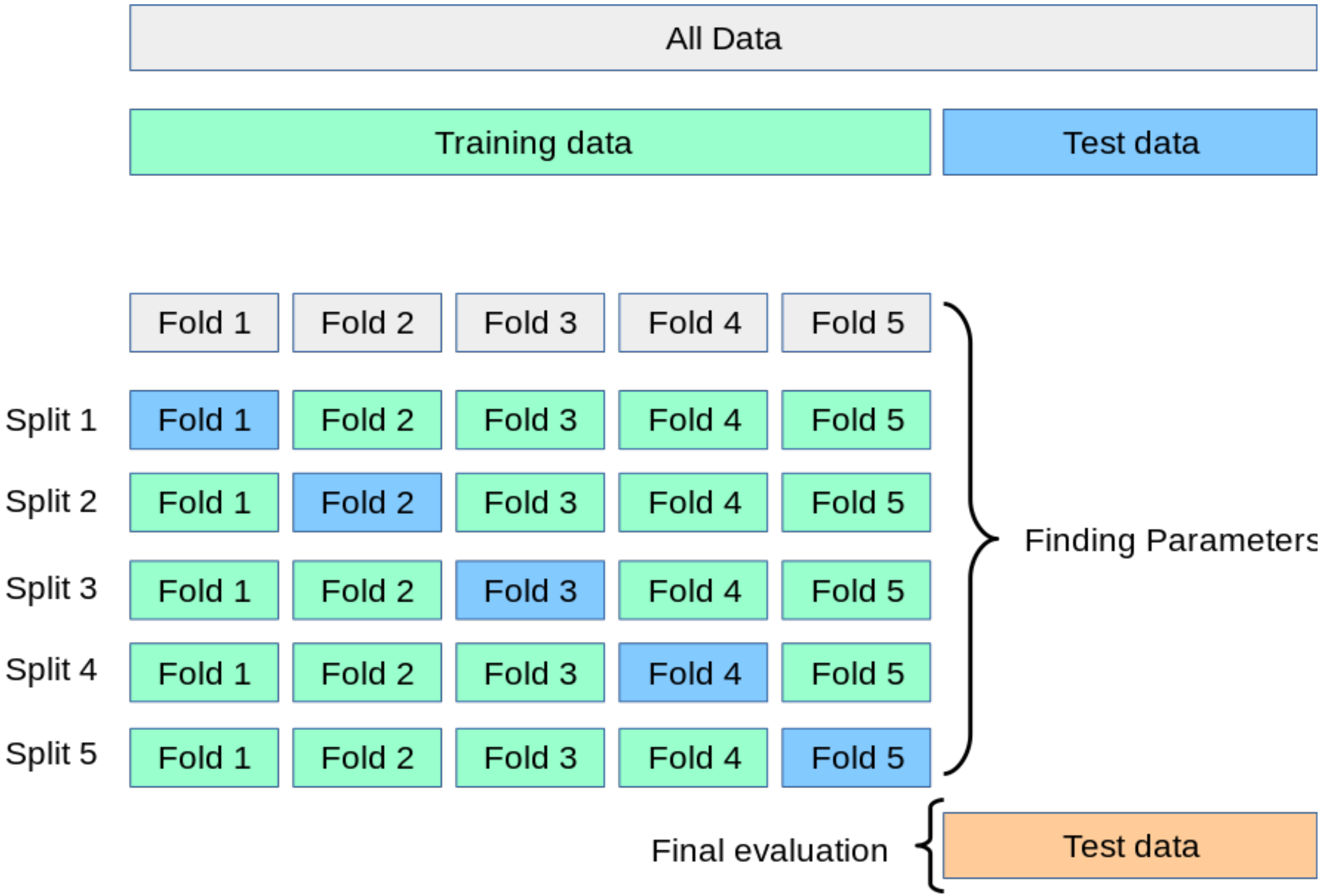
- Special thanks to my supervisor, Dr. Gabriel N. Perdue.
- To members of MINERvA's machine learning group for their mentorship and useful conversations:
  - Anushree
  - Steven (Oak Ridge)
  - Nafis
  - Luis
- To members of SIST committee and my mentor group and other awesome interns:
  - Laura Fields
  - Sandra Charles
  - Judy Nunez
  - Alexander Martinez
  - Raul Campos
  - Matthew Alvarez

References

- Neural networks and Convolutional Networks visualization: Source
- Random Forest: Source

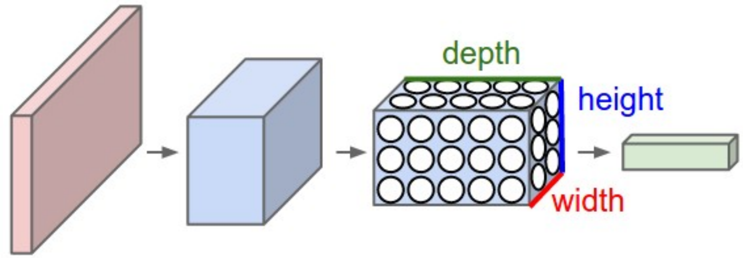
Back-ups

K-folds cross validation



Architectural characterizations of CNNs

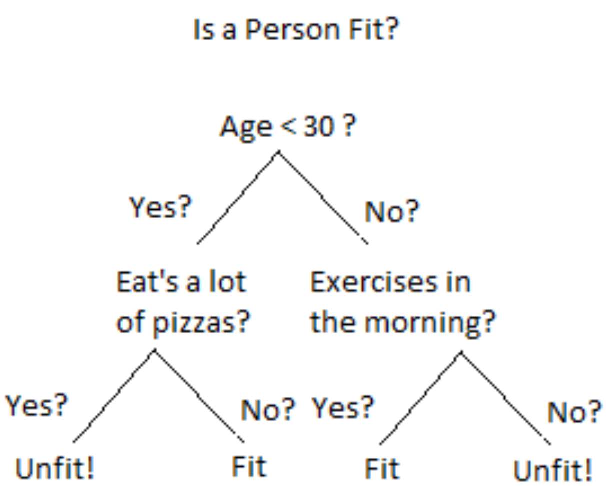
- Different deep learning problems require different network architecture.
- However, selecting an appropriate architecture for CNNs is usually done by human intuition or random search
- If we have a way to uniformly characterize a network architecture, then it would be particularly useful.

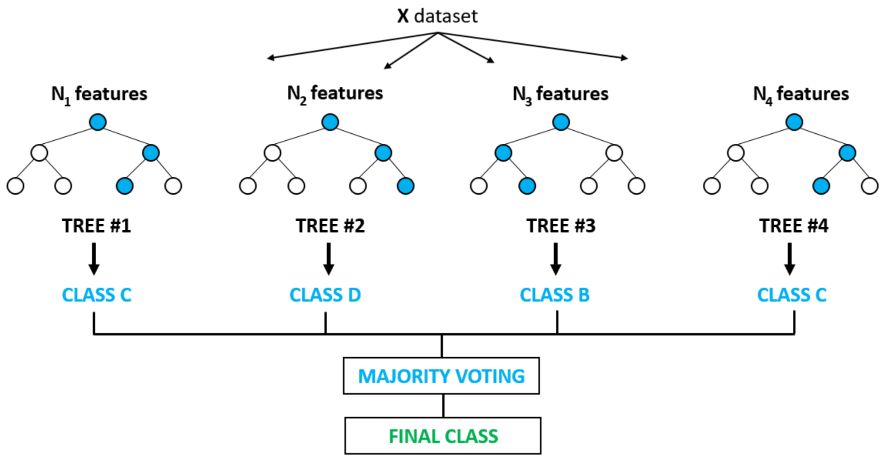


Good for the task?

?

Random Forest and Extremely Randomized Tree





Fermilab