JOINT INSTITUTE FOR NUCLEAR RESEARCH
Dzhelepov Laboratory of Nuclear Problems

FINAL REPORT ON THE SUMMER STUDENT PROGRAM

Investigating Alternative Deep Learning Methods in the NOvA Experiment

Supervisor:
Oleg Samoylov and Chris Kullenberg

Student:
Bruna de Oliveira Stahlhoffer

Participation period:
July 02- August 25

Dubna, 2018

Investigating Alternative Deep Learning Methods in the NOvA Experiment

Bruna de Oliveira Stahlhofer

November 19, 2018

Abstract

In this paper, we will review the basic concepts of convolutional neural networks (CNN) and how CNNs are used to classify events in the NOvA experiment. We will review neutrinos in the standard model, and how the detection of such particles is necessary for the better understanding of the interaction between fundamental particles. We have investigated methods of improving NOvA's CNN classification. Our main focus has been the addition of event reconstruction variables to the network input. This has shown to improve categorization accuracy for small data samples, and a reduction in network over training in large data samples (with negligible improvement in accuracy).

Keywords: Convolutional neural Network; NOvA experiment; neutrino classification.

Contents

1 Introduction **4**

2 Standard Model **5**

2.1 Neutrinos 6

2.1.1 Leptonic Process 6

2.1.2 Semileptonic Process 7

2.1.3 Lepton Universality 8

2.1.4 CP Violation 9

3 The NOvA Experiment **9**

3.1 NuMI Beam (Neutrinos at the Main Injector) 10

3.2 Near and Far Detectors 11

4 Convolutional Neural Networks **12**
4.1 Caffe 13
4.2 Training on Neutrino Data 15
4.2.1 GoogLeNet 16
4.3 Input Improvements 19
4.3.1 Reconstruction Variables 19
4.3.2 Python layer 21

5 Results **22**

6 Conclusion **24**

7 Perspective **25**

1 Introduction

In the mid-1990s, Wilhelm Conrad Roentgen, working on cathode ray experiments, discovered that a plate with barium platinum-cyanide becomes fluorescent when near a Crookes tube where the cathode rays strike. Roentgen found that this new form of radiation appeared to propagate in a straight line, activated photographic plates, and had great penetration power. Becquerel later found that uranium crystals, even in the absence of sunlight, activated photographic plates. Thus, spontaneous radioactivity was discovered[1].

Marie (Sklodowska) Curie (1867-1934), together with her husband, proposed to investigate the nature of Becquerel's rays. She discovered the atoms of radium and polonium. Later, Rutherford classified three types of radiation: α rays, β rays, and γ rays. In 1914 Chadwick showed that the observed β-ray spectrum is continuous, and with this discovery the conservation of energy was called into question. This means, in the rest frame, the electron and the daughter nucleus do not carry the total initial energy. Many different hypotheses were made by the scientific community at the time.

> Gamow: "This would mean that the idea of energy and its conservation fails in dealing with processes involving the emission or capture of nuclear electrons. This does not sound impossible if we remember all that has been said about peculiar impossible properties of electrons in the nucleus".

To solve this problem, Pauli proposed a nucleus constituent called a "neutron" which interacts very weakly with matter and is a fermion with no electric charge. Later, because the neutron was discovered, Fermi called Pauli's particle "neutrino"[2].

Until 1950, neutrinos remained a hypothetical particle whose existence was proposed just to rescue the conservation laws, and subsequently model weak decays. Experimental evidence was necessary. For that, several experiments were developed at the Savannah River nuclear reactor in south Carolina. Finally, Cowan and Reines had success when, in a tank of water with cadmium chloride, they captured the positron produced in an inverse beta decay process, which they were able to interpret unambiguously as the presence of neutrino interactions.[3].

After the eletroweak theory was developed by Glashow, Weinberg, and Salam we know that neutrinos are fermions with no electric charge and carry three different flavours. However, massive neutrinos were incompatible with the Standard

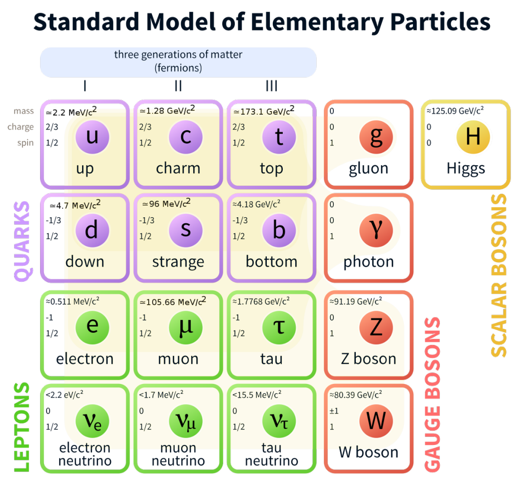
[1] Do átomo grego ao átomo de Bohr [13].
[2] A introduction to elementary particle physics [9].
[3] indroduction to elementary particles [12].

Model during much of the neutrino's history. Though we now know that they have mass, we still do not know which neutrino type is the heaviest. Furthermore, it is unknown if neutrinos are their own antiparticles or not.

To probe these issues we have modern neutrino detectors developed for different and specific proposes. For example, the NOvA experiment is a tracking style of detector, able track the path made by charged particles produced in neutrino interactions. Many studies can be performed using such detectors, like the mass hierarchy problem, leptonic number violation, among others. Therefore, for the sake of measuring neutrino oscillation parameters, which are tightly linked with some of the major puzzles related to neutrinos, we need very accurate measurements, and must be able to reliably classify events as having originated from one of the three neutrino flavors.

This report is organized as follows: in the first two chapters we see the general ideas of the standard model and specifically about neutrinos, as well as how CNNs works. Next, we discuss the NOvA experiment and the tools developed to obtain accurate measurements. We will discuss some technical issues related to to training CNNs to classify neutrino event types, and in section 5 we show some results. Finally, in section 6 we summarize our work.

2 Standard Model



Currently, the four main forces are understood as resulting from the interactions between elementary particles. The most well-understood and successful theory is quantum eletrodynamics. It describes the interaction between electrons mediated by a gauge boson. In QED, this boson is a photon. The other force included in the Standard Model is the strong force describing the exchange of gluons by quarks. We call this theory quantum chromodynamics due to the color charge carried by the quarks. Finally, the last category is the weak theory that describes the exchange of the gauge bosons generating the weak force. The gravitational force is not included in the standard model so we are not going to consider it in this report.

We can see in figure 2 a table of elementary particles classified according to spin and charge. The fermions have fractional spin and the boson have whole-number spin. The fermions are classified according to the charge they carry, the leptons being those with electric charge and the quarks with flavor charge. Some bosons can carry charge, like the W^+ and the W^- that carry electric charge or the gluons which carry color charge.

2.1 Neutrinos

Neutrinos are leptons with spin 1/2 and they can be of three different flavors: electron neutrino, muon neutrino and tau neutrino, each one from their respective leptons.

The main method to probe weak interactions is with neutrinos. While weak interactions occur among standard charged particle interactions, the electromagnetic interactions occur so much more frequently that they drown out any chance of studying the weak force in detail. Because neutrinos have no electric or color charge, and only interact weakly, they are by far the best method to study the weak force. They are produced artificially in proton accelerators, and naturally in the fission reaction in stars, supernova explosions, and in interactions of charged particles with our atmosphere. The cosmos can produce neutrinos in other ways (like relic neutrinos from the big bang), and also contains other unknown sources of high energy neutrinos.

Neutrinos are involved in two process of weak interactions: leptonic process and semileptonic process.

2.1.1 Leptonic Process

When we say charged current (CC) this means that the charges of the fermions in the initial and final process differ by one unit. The boson exchanged in such a process is a $W^{\pm}$. On the other hand, if the boson exchanged is a Z boson, this means the charged in the initial and final process is the same. We call this process 'neutral current' (NC).

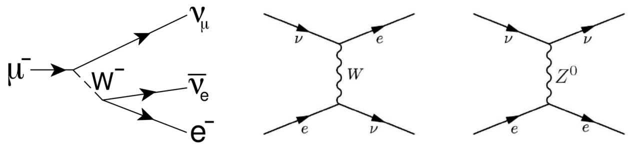


Figure 1: The diagrams are read from the left to the right. The first is a muon decay process and the other two show electron-neutrino scattering with charge and neutral current, respectively.

2.1.2 Semileptonic Process

Semileptonic processes are those where both leptons and hadrons are involved.

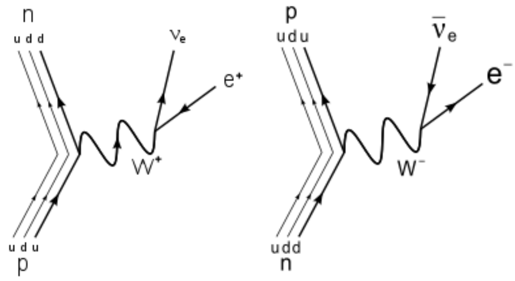


Figure 2: Beta decay and inverse beta decay, respectively. Read the diagrams from the bottom to the top.

Beyond the beta decay, neutrinos participate in quark-neutrino scattering, as we see in figure 2.1.2, where the first diagram is NC and the second CC. In weak interactions there is a third classification for non-leptonic processes, but the neutrinos don't participate so we will not discuss it in this report.

Another important classification necessary for our discuss is the range of energies involved in the scattering processes. These can be elastic, quasi-elastic, resonance, or deep inelastic scattering.

Under about 1 GeV we have elastic and quasi-inelastic scattering. For the former, no particles in the system change and a boson Z is exchanged carrying four-momentum. In the latter, the neutrino interacts with a nucleon and transforms into its charged counterpart through the exchange of a $W^{+/-}$ as we see in the process:

$$\nu + n, p \rightarrow e^{\pm} + p, n \tag{1}$$

As we move higher in energies, up to about 2 GeV, we begin to produce extra

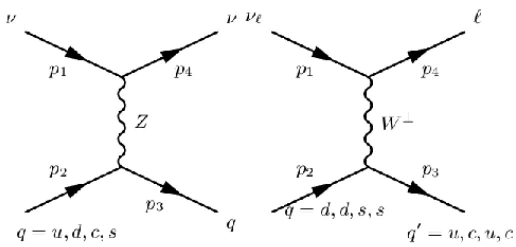


Figure 3: Quark-neutrino scattering.

resonant particles in the interaction that quickly decay, such as δ^{++}. At even higher energies, above 2 GeV, we begin to form any number of particles that are allowable for the given energy, in deep inelastic scattering (DIS) processes. An example of a DIS interaction is shown in figure 2.1.2.

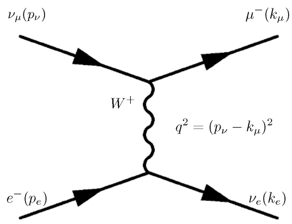


Figure 4: Deep inelastic scattering. Read from the left to the right.

2.1.3 Lepton Universality

There are many of unsolved problems involving neutrinos. They are a good way to probe some yet to be understood problems in the particle physics. One of then is the lepton universality.

Lepton universality means that one type of lepton should be produced as often as another. That is, the charged weak interaction is universal.

Let us look at an example of the tau decay:

$$\tau^- \to \mu^- + \overline{\nu}_\mu + \nu_\tau \tag{2}$$

$$\tau^- \to e^- + \overline{\nu}_e + \nu_\tau \tag{3}$$

These two interaction modes should have the same probability to occur. With very high accuracy we can measure these two cross sections:

$$\frac{\Gamma(\tau^- \to \mu^- \overline{\nu}_\mu \nu_\tau)}{\Gamma(\tau^- \to e^- \overline{\nu}_e \nu_\tau)} \propto \frac{g_\mu^2}{g_e^2} \tag{4}$$

Where g is the weak charge. The result is [9]:

$$\frac{g_\mu}{g_e} = 1.001 \pm 0.004 \tag{5}$$

2.1.4 CP Violation

Another big question involving neutrinos is their oscillation between different types, and how this can help the study about CP violation.

Charge-Parity violation has been observed with mesons K^0 and B^0. But this is not enough to understand why there is much more matter than antimatter in the universe, for example.

Once we have experimental evidence that the neutrinos have mass, we can describe the mass eigenstate as linear combination of three different flavours. This means the neutrinos can oscillate type, as Bruno Pontecorvo predicted. The unitary transform matrix is:

$$\begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{i\alpha} & 0 \\ 0 & 0 & e^{i\beta} \end{pmatrix}$$

Figure 5: The Pontecorvo–Maki–Nakagawa–Sakata matrix.

where c and s are sine and cosine of the mixing angle, respectively. δ is the CP phase violation.

CP violation continues to be a mystery, and there is much to discover. Neutrinos could be a good tool for this propose.

3 The NOvA Experiment

As we discussed, until 1950 the neutrinos were just a mathematical tool to preserve the conservation laws, though Fermi's development of Weak theory had thoroughly convinced theorists that the neutrino must exist. However, neutrinos interact very weakly with matter and are difficult to detect. The first success was obtained with a tank full of water and Cadmium chloride. They expected the inverse beta reaction:

$$\overline{\nu}_e + p \rightarrow n + e^+ \tag{6}$$

The positron signal was used as a trigger in coincidence with delayed signal from the neutrons absorbed by a cadmium nucleus.

Since the neutrino's first detection, many different detectors have been constructed for different proposes. As we saw, one of the actual challenges is to understand neutrino oscillations. The NOvA detector was developed for this main propose: to measure the electron-neutrino appearance and the muon-neutrino disappearance from Fermilab's NuMI neutrino beam.

The method used for the detection of neutrino interactions is a liquid scintillator. The photos emitted are captured along a wavelength-shifting fiber and directed to a photodetector, which transform the light into an electrical signal.

For accomplishing the main goal of the NOvA experiment is important not just detect the occurrence of neutrino events, but to measure tracks within the detector. This is possible through a sophisticated structure made of hundreds of rectangular PVC cells filled with liquid scintillator. The figure 3 shows a schematic of a single cell [7].

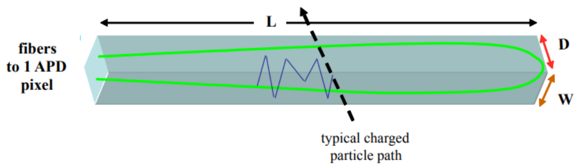


Figure 6: Schematic illustration of a PVC cell of dimensions WxDxL containing liquid scintillator and a wavelength-shifting fiber (green). A charged particle incident on the cell produces scintillation photons (blue line) that bounce off the cell walls until absorbed by the fiber or lost.

The neutrino beam is formed in Fermilab, USA. The experiment consist of a 300-ton near detector and a 14000-ton far detector located in the Ash River, 810 km to the North of Fermilab, in Minnesota. The initial neutrino content of the beam is measured at the near detector, and interactions at the far detector allow us to measure the proportion of neutrinos that have oscillated into another flavor.

3.1 NuMI Beam (Neutrinos at the Main Injector)

The Fermilab accelerator is capable of delivering 700 Kilowatts of power to the NuMI beam. Protons are accelerated and hit a graphite target. A strong magnet focuses the beam of charged particles, like pions, which decay into muons and muon neutrinos. The beam starts out 150 feet below ground at Fermilab and is directed at a 3,3^{o} downward angle. The muons are stopped with thick walls of rock, and the beam of neutrinos continues its journey unhindered through the earth until passing through the detectors.

3.2 Near and Far Detectors

The Near Detector is located approximately 500 m downstream of the neutrino production target. They are made of blocks of 16-cell PVC extrusions. We can see schematic these structures in the figure 3.2. The Far Detector has a total of

Figure 7: (a) Close-up photos of one 16-cell PVC extrusion, 15 cm long. (b) Two full-size 16-cell extrusions 15.5 m long placed side-by side form the basis for an extrusion module.

344,064 PVC cells, each cell measures 3.9 cm wide, 6.0 cm deep and 15.5 meters long. Both detectors are located 14 milliradians to the west of the central axis of the NuMI beam. In this way, it is expected that the neutrino energy spectrum will be around 2 GeV. In the figure 3.2 we show a picture of the detectors' size [6].

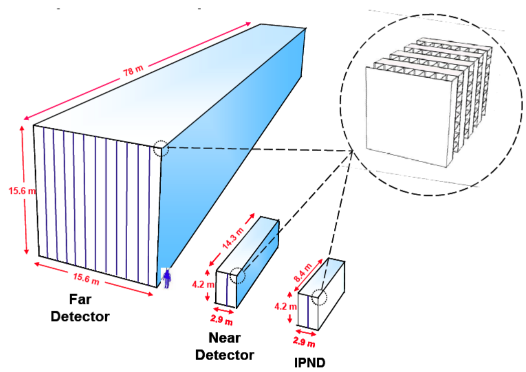


Figure 8: The far and near detector. The far detector is the largest free-standing man-made plastic structure in the world.

With this construction is possible reconstruct the tracks of the charged particles passing through the detectors. For this goal, it is necessary to develop sophisticated computational tools for reconstruction and classification, as we will see next.

4 Convolutional Neural Networks

Convolutional neural networks (CNN) have been used for a long time in image recognition and analysis. With the track images as input, it's possible develop a specific CNN capable of classifying neutrino event types. Let us take a very simple example to see how this works.

Suppose we want an algorithm to recognize handwritten numbers [3]. For humans is an easy task, but we need to find a way which the machine is capable to learn. We can build filters to look for patterns and calibrate the machine learning with a known result.

In practice, the input images are pixel arrays. The filters, sometimes called kernels, are used to search for patterns within parts of the image. This kernel can learn to recognize patterns such as lines, curves, or generally whatever is necessary to recognize our number. The network contains weights combined in various ways, that together approximate a function that should give us the desired output. The network will 'learn' by modifying these weights until the function's output approaches the correct solution.

Figure 9: An example of inputs to training the network. Each image will pass through the network, along with its correct category label.

If the network is capable of learning well, then at the end of our training we should obtain, as we see in example of the figure 4, the first number classified as a zero.

Mathematically speaking, the network evaluates the weights with a loss function, which estimates the error in the answers produced by the network. The first weights are randomly chosen. After evaluating the loss function we use a method known as Gradient Descent, which basically attempts to minimized the loss function by modifying the values of the weights. In the figure 4, we can see the general structure of a CNN.

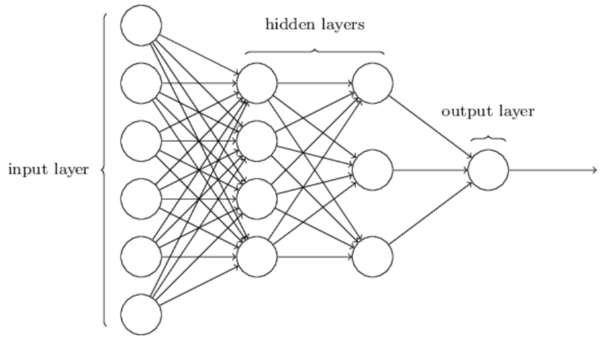


Figure 10: The general structure of a CNN with four layers.

Let us present some standard layers of a CNN.

- **Convolution:** The main goal of the convolution operator is to extract features from the input image. A feature map is produced by multiplying sections of the input image by the kernels. We see a pictogram in figure 4 representing this operation. The particular kernels are learned by the network, rather than programmed explicitly.

- **Pooling:** The pooling layer is used to reduce the size of the array, as we see in the figure 4. Passing too much data through the network can produce unreasonable training times.

- **ReLU:** The ReLU operator introduces non-linearity to the CNN. The negative pixel values in the feature map are replaced by zero [8].

- **Fully Connected:** This is a traditional neural network. The previous convolution steps are used to extract useful patterns from the images, keeping important information, while reducing the data flowing into the final neural network.

4.1 Caffe

Caffe is a deep learning framework made with expression, speed, and modularity in mind [4]. Let's look in the main structure of caffe.

- **Nets, Layers and Blobs:** the blob is the standard array and unified memory interface for the framework [4]. It is a basic a N-dimensional array stored

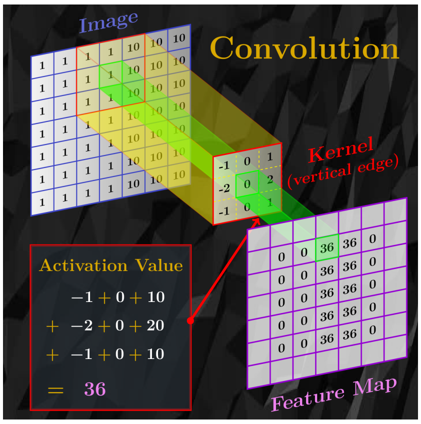


Figure 11: A diagram showing the convolution of a portion of an image with a kernel that searches for vertical lines.

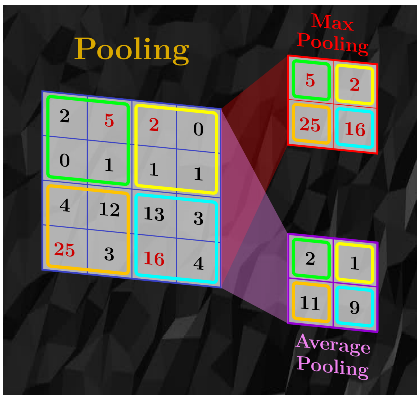


Figure 12: A schematic of a pooling layer.

with usually four standard dimensions [N, channel k, height h, width w] where N is the batch size of the data. The channel can be the RGB images, or for us will be the X and Y views of the detector.

- **Forward / Backward:** Given an input from the bottom the forward operation the forward step computes the output and sends it to the top layer. The backward step calculates the gradient to the input and sends it to the bottom layer.

- **Loss:** The goal is to find a value for the weights that minimizes the loss function. A standard choice of loss function is SoftmaxWithLoss, which scales the outputs so that they represent a kind of probability.

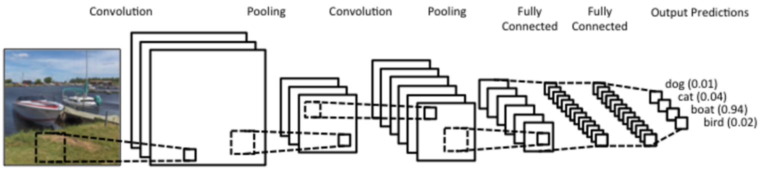


Figure 13: Basic structure of a CNN.

```
layer {
  name: "loss"
  type: "SoftmaxWithLoss"
  bottom: "pred"
  bottom: "label"
  top: "loss"
}
```

Figure 14: Network definition of the loss function.

- **Solver:** Coordenates the network's forward inference and backward gradients to form parameter updates that attempt to improve the loss [4].

  Each iteration does the following:

  1 Calls network forward to compute the output and loss.

  2 Calls network backward to compute the gradients.

  3 Incorporates the gradients into parameter updates according to the solver method.

  4 Updates the solver state according to learning rate, history and method to take the weights all the way from initialization to learn model.

4.2 Training on Neutrino Data

Our goal is to obtain correct classifications for the different types of neutrino interactions. We need a CNN capable of distinguishing between electron-neutrino (ν_e), muon-neutrino (ν_μ), tau-neutrino (ν_τ), and the background (cosmic neutrinos, for example).

For training we use Monte Carlo simulation events from a data set generated by the NOvA group at Fermilab. As we saw, we use the Caffe framework to develop our CNN. This packaged is originally configured with GoogLeNet, which we can use to learn the features of Caffe.

4.2.1 GoogLeNet

GoogLeNet is an improvement over the pioneer LeNet-5 (1998). It was capable of recognizing digits and hand-written numbers, but now the numbers of parameters was drastically reduced with very small convolutions.

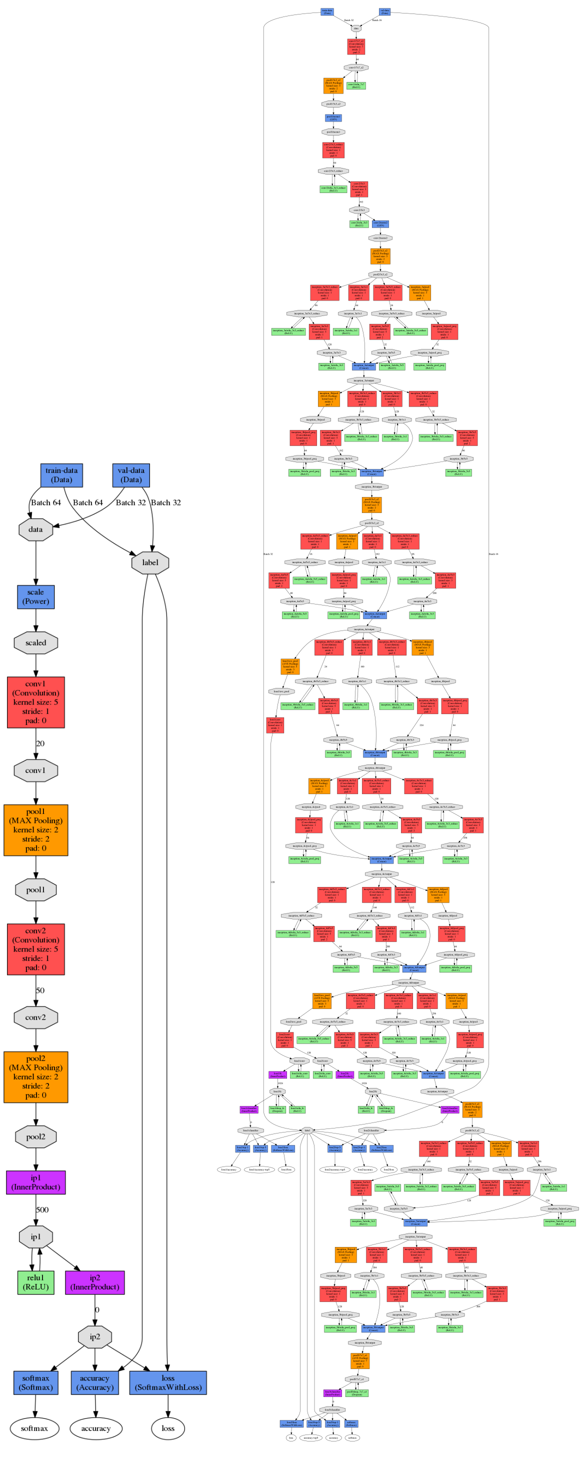


Figure 15: Lenet and GoogLenet CNN.

For our research, we used a customized network. We can see a result of a training using an LMDB dataset in the figure 16

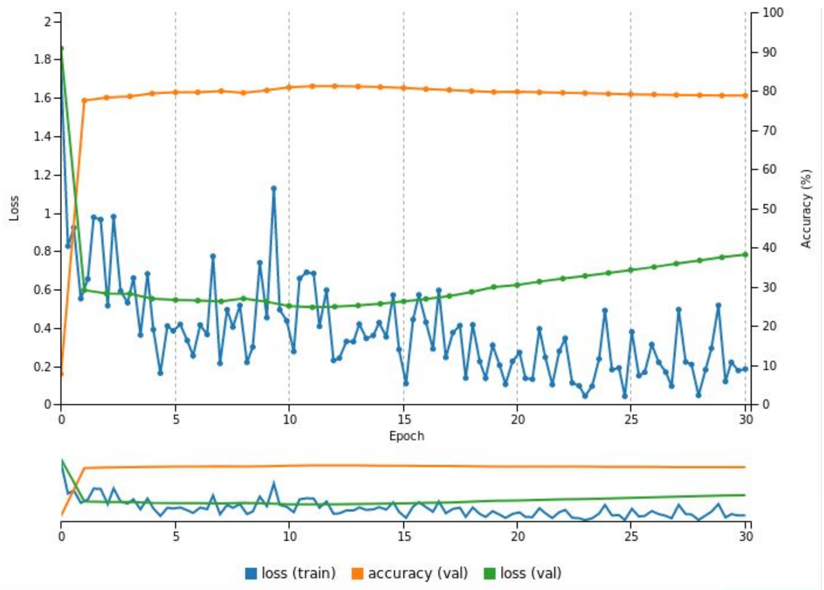


Figure 16: loss and accuracy

The data used for network training comes from Monte Carlo simulations of neutrino interactions in the NOvA detector. A separate sample is used for validation, to test the network's generality when applying it to data that it has not seen before. While the loss decreases, the accuracy increases accordingly. This can tell us how well our network is training. The loss function is calculated for both the training set, and the testing/validation set, though accuracy is calculated only from the test dataset. If the training loss continues to go down, while the testing loss goes up, it means that the network is losing its generality, and is being 'over trained', or is learning the specific features of the training sample too well. In figure 16 one can see an example of over training when comparing the blue (train) and green (test) loss functions.

We can also choose how the learning rate changes over time. In the beginning, we need more variance in the rate to search for a region containing in the weight space that contains a minimum loss value. As learning continues we need to decrease the rate at which we modify the weights, so that we do not jump back-and-forth around the actual minimum. Figure 17 is schematically showing a step decrease rate, though one can use any number of decreasing functions, such as an exponential.

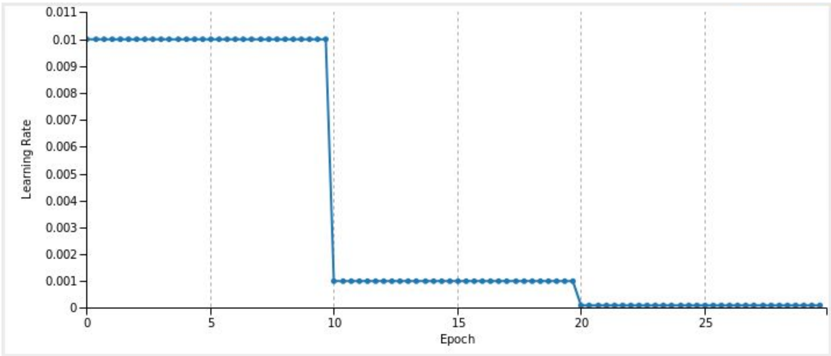


Figure 17: A step-down learning rate decrease

4.3 Input Improvements

In order to improve the network, one of our possibility is to append the input with extra information in the hopes of training faster and more efficiently. The way we choose to do that is by adding a third layer in addition the standard two, containing the X and Y views.

This allows us pass the 3-layer.png files through the network as a test. Then, we can add simple reconstruction information for each event, for example, energy, angles, and vertex position.

4.3.1 Reconstruction Variables

The hypothesis is that adding some reconstruction information directly into the input of the final fully-connected network can help the training. The idea being that this additional preprocessing information might aid the network in event identification. We can reconstruct the vertex position from the event image, for example. Additionally, we can fit the tracks with lines, and associate each point with a track. We can see some examples of our simple reconstruction figure 18.

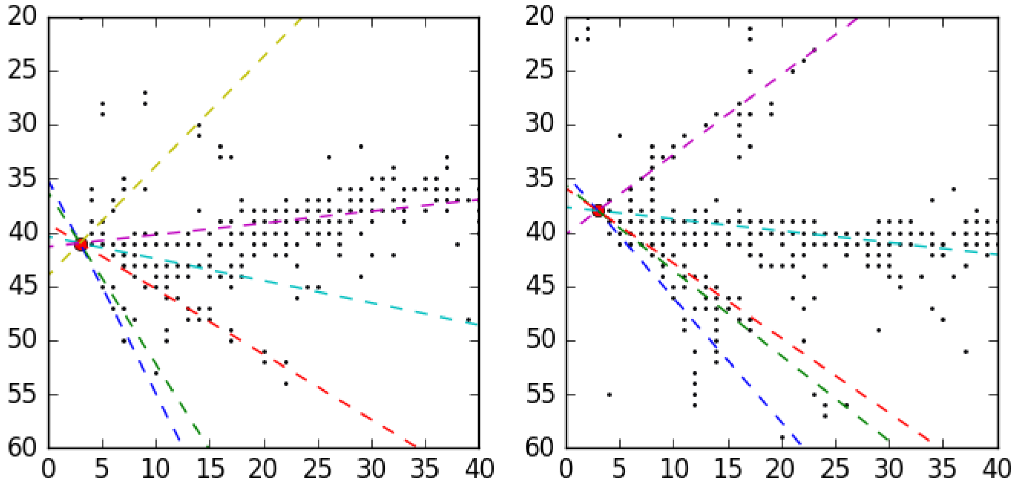


Figure 18: Some examples of simple track and vertex position reconstruction.

The third layer needs to have the same format of the two standard layers, containing the X and Y view images. In order to save disk space, NOvA has chosen to store these training images as 8-bit unsigned integers, with a range of 0-255. Therefore, when adding the values of the new variables to the third layer we must rescale into this range as well.

We can see in the figure 19 and 20, some other examples of reconstruction variables we can put in the third layer. We can see with these plots how useful the variables may be to the network. Each color shows a different interaction type, all with plot area scaled to unity to compare shapes. The better the separation between colors then the better the variable will be for classifying the events. For example, in figure 20 the maximum track is a good variable for identifying background and muon-neutrinos events.

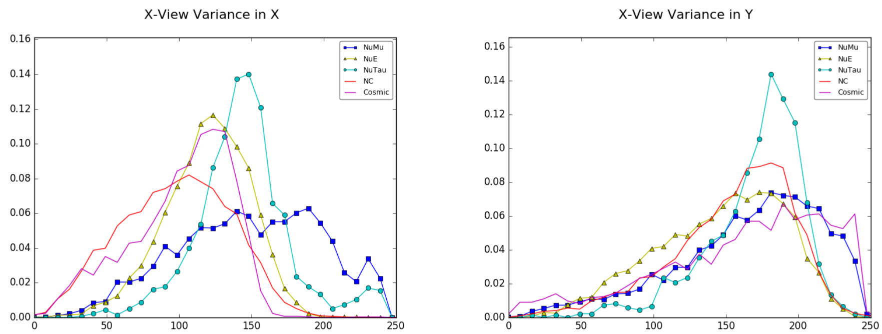


Figure 19: Standard reconstruction variables about X and Y view.

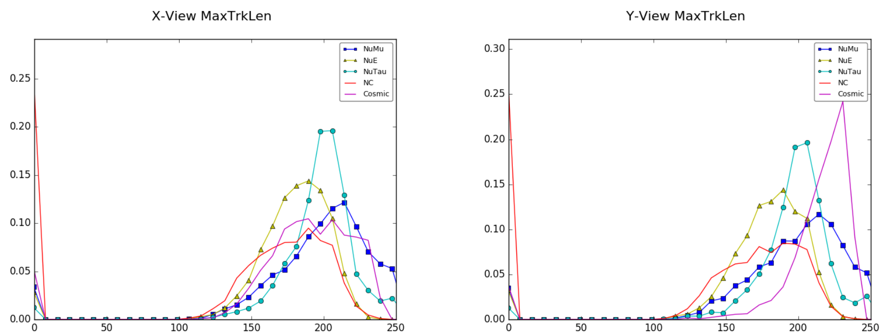


Figure 20: Here we show the maximum track length of an event.

4.3.2 Python layer

We need a method to implement the third layer in the network. The Caffe framework allows for Python code to be called in the network. In this way, the python layer extracts the reconstruction information from the third layer, which is then concatenated with the convolutional network output.

It is possible to calculate some reconstruction variables with a *python layer* at same time as the training. However, by default Python will use the CPU for calculations, which causes a bottleneck in the network. To use on-the-fly calculations Python must be implemented in a way that takes advantage of the GPU's parallel computation. What we have chosen to do, instead, is add the reconstruction variables to the dataset, and simply use a Python layer to pass the information further down the network, which is a very computationally light task and does not affect the speed of training.

5 Results

There are a large number of possible reconstruction variables, but not all of them improve the network's performance. We found that the addition of some variables may even reduce the total accuracy of the network, perhaps because they simply do not distinguish between event types well enough to provide useful information.

We have chosen a set of variables that maximized the improvement in training (both in accuracy and in the reduction of over training).

Our results were obtained with the follow variables:

- X-view variance in hit positions in X/Y
- Y-view variance in hit positions in X/Y
- X/Y view maximum track length (reconstruction variable)
- X/Y view average hits per track (reconstruction variable)
- X/Y view X-coordinate of average hit position

The results are:

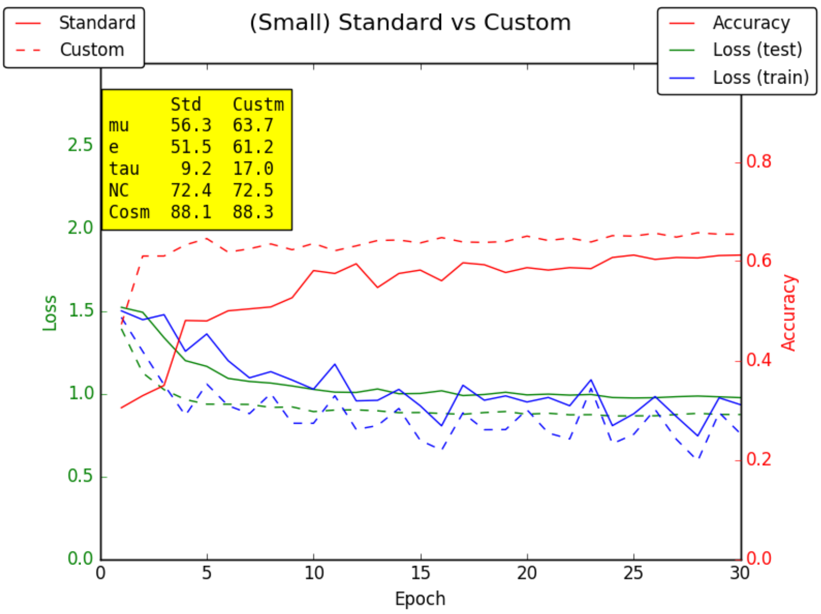


Figure 21: Custom accuracy and loss for the small data sample. The plot is showing comparation between standard network and the custom one with some reconstruction variables.

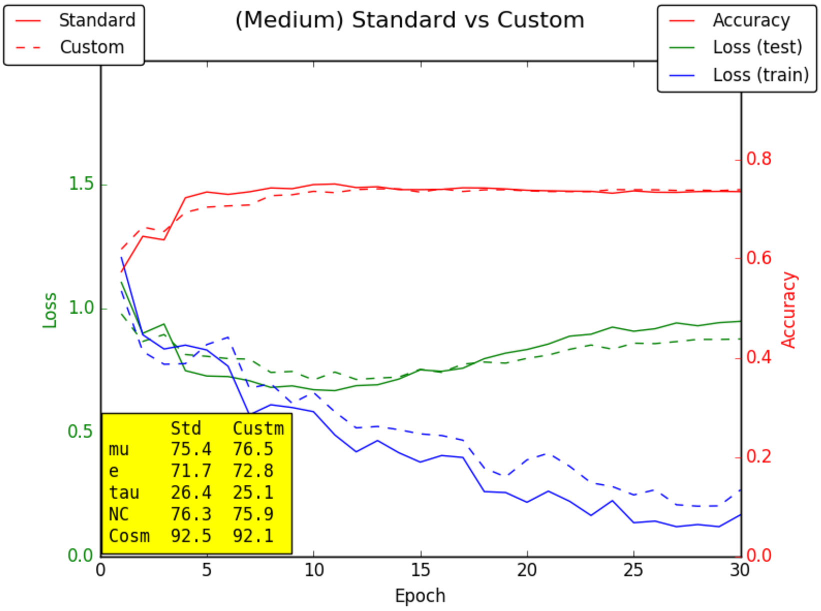


Figure 22: Custom accuracy and loss for the medium data sample. Unlike the small sample, there is no much improvement with the custom network in comparation with the standard one.

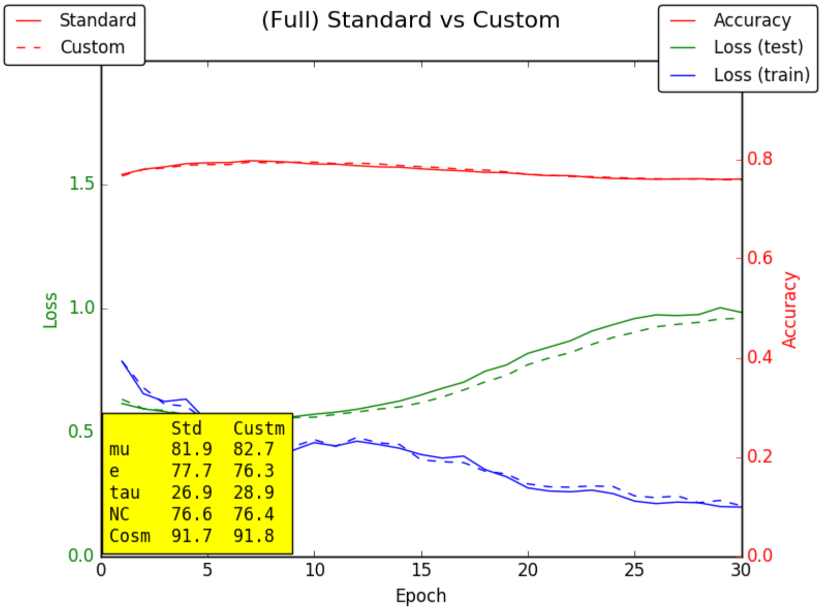


Figure 23: Custom accuracy and loss for the full data sample. We can see that there is almost no difference between the standard and custom network.

As we see in the figure 21 the custom loss shows some improvements, mostly for the muon-neutrino and electron-neutrino. This is a good result as these are generally the more difficult categories to classify. Additionally we see an increase in the total accuracy of the network with this small data sample, and we achieved this increased accuracy much more quickly.

For the medium and large data sets we did not see the same improvement in accuracy, as can be seen in figures 22 and 23 for the medium and full samples, respectively. There is slight improvement in the total accuracy, but it is marginal. The main improvement in the larger data sets is in the reduction of the over training of the network, where we can see the blue and green lines do not diverge quite as much.

6 Conclusion

We attempted to improve the network by adding a third layer as input into the network, in witch it is possible add simple reconstruction variables for each event. The results are preliminary. We showed that this method generally reduced the over training of the CNN, meaning that it learned more general features of the events. However, we were surprised to see that the accuracy was not improved by much. For the small data set we did see an impressive jump in accuracy. But it seems that for large data sets the network seems to learn to classify very well based only on the images, without much improvement by adding extra information. But this is only a first attempt, and there is still room for improvement in the methods.

7 Perspective

We used a very rudimentary method of calculating reconstruction variables. The quality of the variables input into the network could greatly affect the results. So our method may already see improvement by incorporating NOvA's official reconstruction calculations. Additionally, the choice of variables used and the structure of the network should be studied more. We made a careful study of each individual variable's effect on the network's results, but this is time consuming and does not take into account possible useful correlations between variables. There may be ways to improve such studies, both in quality and speed.

References

[1] A convolutional neural network. Aurisano et al. 2016. 10.1088/1748-0221/11/09/P09001.

[2] https://adeshpande3.github.io/adeshpande3.github.io/A-Beginner's-Guide-To-Understanding-Convolutional-Neural-Networks

[3] Neural network and deep learning. Virtual book. http://neuralnetworksanddeeplearning.com

[4] articlejia2014caffe, Author = Jia, Yangqing and Shelhamer, Evan and Donahue, Jeff and Karayev, Sergey and Long, Jonathan and Girshick, Ross and Guadarrama, Sergio and Darrell, Trevor, Journal = arXiv preprint arXiv:1408.5093, Title = Caffe: Convolutional Architecture for Fast Feature Embedding, Year = 2014

[5] http://novaexperiment.fnal.gov

[6] "PVC Extrusion Development and Production for the NOvA Neutrino Experiment". R.L.Talaga et al. 2015. 10.1016/j.nima.2017.03.004.

[7] Liquid scintillator production for the NOvA experiment. S. Mufson et al. 2015. 10.1016/j.nima.2015.07.026.

[8] https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/

[9] Bettini, A. Introduction to elementary particle physics. 2008.

[10] Becchi, C. M., Ridolfi G. 2006. An introduction to relativistic processes and standard model of electroweak interactions.

[11] Schwartz, M. D. 2014. Quantum Field Theory and the Standard Model.

[12] Griffiths, D. 2004. Introduction to elementary particles.

[13] Peduzzi, L. O. Q. 2015. Do átomo grego ao átomo de Bohr.