Application of Machine Learning and a Convolutional Neural Network as a Neutrino Event Classifier (Jan 2022)

Author: Joshua Giblin-Burnham (SN:19052055)

Department of Physics and Astronomy, University College London

Contents

I. Introduction

Categorisation of particle interactions observed in detectors is a core problem in experimental high-energy particle physics (HEP); typically, categorisation is achieved using machine learning algorithms applied to reconstructed components such as tracks, showers, and jets associated with specific particle interactions and energies. While these techniques have successfully allowed HEP physicist to characterise much of prior particle interactions, they are prone to two flaws highlighted in the Journal of Instruments [1]: "mistakes in the reconstruction of high-level features can lead to incorrect categorization of the physics of the event, and the features used to characterize the events are limited to those which have already been imagined for the experiment".

Alternatively, we look to replace this methodology with Convolutional Neural Networks (CNNs) which have a wide breadth of application in feature learning and categorisation problems therefore playing a pivotal role in image recognition and analysis. What follows is an application of these technics to the problem of identifying particle interactions in high energy neutrino physics, specifically NOvA-like neutrino detector images. The discussion includes an overview of the core concepts of CNNs and deep learning, the techniques used to build and train these networks including using feature extraction and machine learning algorithm, then the application of those concepts to specific example of identifying NOvA-like neutrino interactions based on visual topology.

II. Background

A. Neutrino Background:

Neutrinos are massive, electrically neutral particles part of the lepton family, like electrons. There are three 'flavours' of neutrino (electron ν_e, mu ν_μ and tau ν_τ) which are each paired with their respective heavier charged lepton. Due to quantum mechanical phenomenon, the neutrinos can oscillate in flavour; these oscillations have been observed for muon neutrinos to tau neutrinos, however, they've not been observed for muon neutrinos oscillating into electron neutrinos. These oscillations are significant as the unknown factors that govern them could have important implications for our understanding of the particle physics and are seemingly the only process that'll been observed to break the Standard Model of particle physics.

Current Particle Physicists study these 'long-baseline' neutrino oscillation in the NuMI Off-axis ν_e Appearance (NOvA) experiment near Ash River, Minnesota which has been running since 2014. Its primary research goal is measuring three flavour neutrino oscillations. NOvA has two detectors, separated by 810 km; a Far Detector in Ash River, MN, which measures neutrinos after the possible oscillation and a Near Detector at Fermilab, IL, that measures the neutrino beam before oscillations occur. NOvA measures the 750-kW beam of mainly muon neutrinos produced by the NUMI beam line, and the Far Detector measures both electron neutrinos and muon neutrinos due to oscillations. However, the complex multi-step process involved in creating a neutrino beam means that the beam has a broad spectrum of energies and particle types (including some ν_e components and antineutrinos, like $\bar{\nu}_\mu$ and $\bar{\nu}_e$.)

Due to their electrical negativity, neutrinos are invisible to particle detectors and must be detected instead when they interact with the material in the detector; the interactions are extremely rare and occur by the weak processes. Therefore, to study neutrinos, they must be created in an intense beam and sent through the large detector continuously for a long period of time. The NOvA detector is filled with instrumented tubes filled with liquid scintillator. This is an organic molecule, mainly consisting of carbon atoms, thus, mostly the neutrinos interaction with carbon nuclei is studied. Neutrino are detected when they strike an atom in the liquid scintillator and subsequently releases a visual burst of charged particles. Wavelength-shifting fibres connected to photodetectors are then used to calculate the energy of the particles they come to rest. Using the pattern of light seen by the photodetectors, the energy and kind of neutrino that caused the interaction can be identified.

The interactions can then be categorised into two strata: charged-current (CC) processes, where the neutrino is turned into its charged partner (i.e., a electron-neutrino ν_e turns into an electron e −); or neutral-current (NC) processes, where the neutrino scatters but remains as a neutrino (not changing flavour). In a charged-current interaction, to balance charge, a neutron in the nucleus will also be converted to a proton. The data analysis in this report is simulated data similar to that produced by the NOvA detector, in which the images produced show the beam is travelling in the z direction and from both x and y view (i.e., view from above and the other the side), these images are show in Fig. 1.

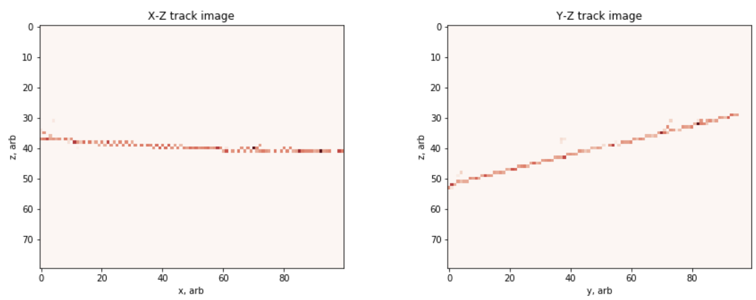


Figure 1: Sample image of simulated NOvA data used in the report

In addition, as nucleons (protons and neutrons) are not fundamental particles, instead composed of quarks and gluons, the interaction between nucleons and neutrinos can vary. Thus charged-current interaction can be further categorised into a set of different interaction types: quasi-elastic (the simplest interaction, at lowest energies: ν_μ + n → μ − + p), resonant (where the nucleon is excited to an unstable 'resonant' state, which then decays) or deep inelastic scattering (where the nucleon is broken up, and the neutrino scatters from a quark inside it).

B. Neural Network Background:

A neural network is an artificial computer network composed of 'neurons' (or often called nodes), which form a graph of connections that can be used for solving artificial intelligence problems. The connections of such networks are analogous to biological neuron but are modelled with numerical weights and biases between the nodes. These weightings reflect an excitatory and inhibitory connection for positive and negative values respectively. The given output value of a neuron, y, is calculated by applying the activation function, f, to the summation of all input, x_k, modified by the corresponding weight, w_k, and bias, b, the formula for this is given as,

$$y = f(\sum_{i=1}^{n} w_i x_i + b) \tag{1}$$

This activity is referred to as a linear combination and is calculated for all nodes/ neurons of the network. In HEP the traditional neural network used is the multilayer perceptron (MLP) which means a fully connected networks, in which each neuron in the input and intermediate layers / hidden layers are connected to all neurons in the subsequent layer as shown in Fig. 2. However, the "full connectivity" of these networks makes them prone to overfitting data and don't adapt well to variations in the data set.

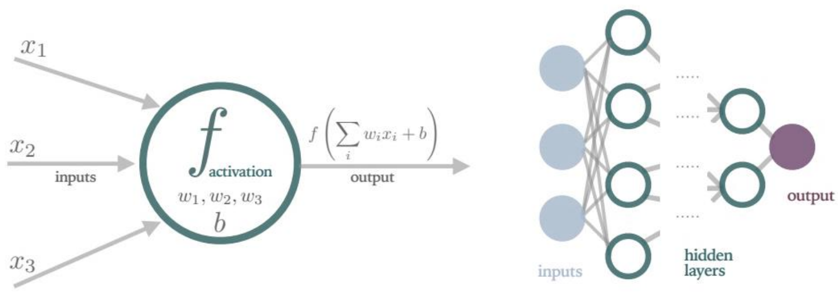


Figure 2: Multi-layered Perceptron Architecture, illustration from International Journal of Modern Physics [4]

In comparison, Convolutional Neural Networks (CNN) learn topological pattern in data rather than hard set patterns which is advantageous when applied to problems involving visual imagery; CNN are thus considered shift/ space invariant neural networks as they learn feature no mater locality in the image; this is achieved by computing the product of convolutional kernels with the image (sliding kernels along input) to produce invariant outputs known as a feature map.

In a CNN, convolutions are thus applied within the hidden layers. Typically, convolution works by sliding a kernel along the input image, calculating the dot product value for each translated convolution, such that a convolution operation generates a feature map which is the contributes of all these dot products. Thus for an the input image of shape [number of images x image height x image width x input channels] the output becomes a feature map with shape, [number of images x feature map height x feature map width x feature map channels].

Moreover, CNN also include pooling layers which reduce the dimensions of data by passing a 'frame' over the image and combining the 'pixel values' in the frame. Specifically, max pooling uses the maximum value of each pooling frame in the feature map. Following convolution and pooling a CNN then will use a fully connected layers to learn underlying trend in feature maps. It is the same as a standard multi-layer perceptron neural network.

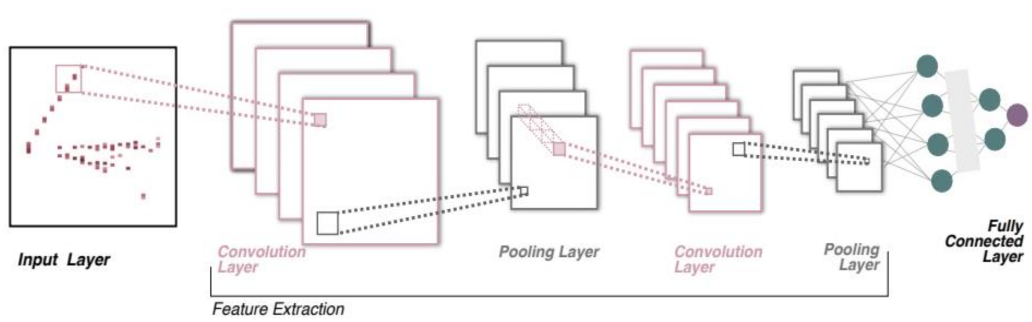


Figure 3: Multi-layered CNN Architecture, illustration from International Journal of Modern Physics [4]

C. *Application to NOvA event classification:*

We looked to construct and train our own CNN for the identification of neutrino events in NOvA-like experiment data. The NOvA experiment as outlined above take precision measurements of neutrino oscillation as the flavour of neutrinos changes spontaneously. We look to make a binary classify to identify specific event types, therefore, the CNN is used to extrapolate the features of each event type, the typical features are outlined in Journal of Instruments [1] which include:

- "ν_μ CC- One of the main topological features of these events is the long, low dE/dx track corresponding to the track of a minimally ionizing muon" [1]
- "ν_e CC- The electron topology is typically a wide shower, rather than a track, whose dimensions are related to the radiation length of the detector material" [1]
- "ν_τ CC- The tau is extremely short lived and not visible in the NOvA detector but decays immediately with varying final state probabilities" [1]

CC interactions are then further divided into quasi-elastic (QE), resonant (RES), and deep-inelastic scattering (DIS) categories which vary in topological/ visual complexity. QE events are two-bodied interaction in which the nucleon recoiling intact from the scattering lepton. In RES events the nucleon is knocked into a resonance and decay, and in higher energy DIS events the nucleon breaks up in the process of the interaction. Figure 4 shows example, simulated, events from these categories as they might be recorded by the NOvA detectors. A simplified categorisation of the interaction modes is:

• QE: Clean event, normally just two tracks
• RES: Something in the middle
• DIS: Messy event potentially many tracks and showers

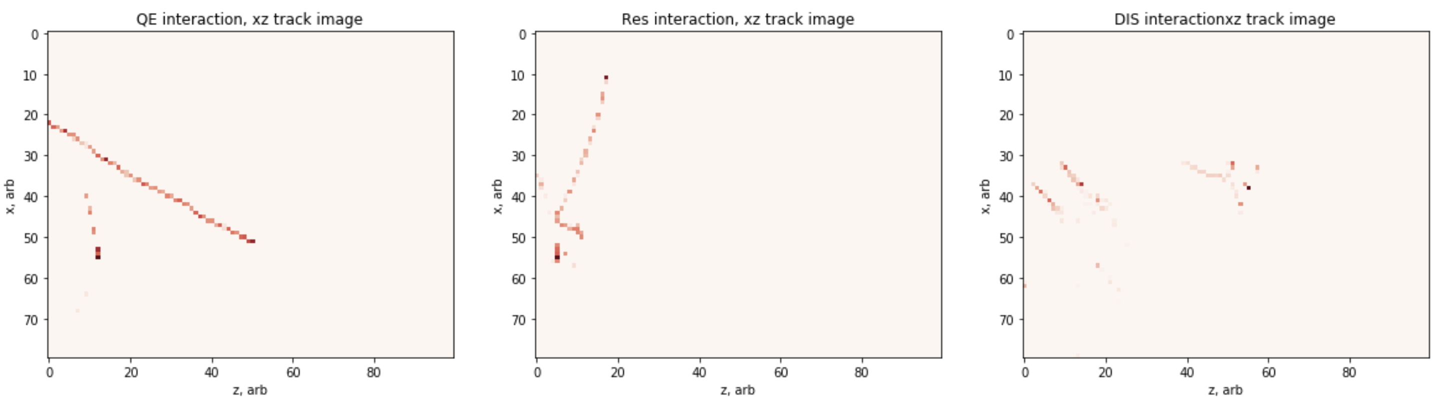


Figure 4: Sample image of different Interaction types

We also analysis the final state which refers to the final product of the neutrino interaction. When charged particles move through the NOvA detector, they leave distinctive patterns of energy deposits, and these can be used to identify the particles produced by the interaction. Different types of particles tend to produce different patterns; for example, muons often make long, straight tracks; protons and charged pions make shorter tracks, while electrons, photons, and neutral pions can generate electromagnetic showers, wider cone-like patterns of energy deposits. Each interaction has an associated neutrino energy which refers to the true energy of the simulated incoming neutrino and when that neutrino interacts, it will always produce a final-state lepton with set energy. These are all strata of data well suited for CNN categorisation. We will thus create a binary classifier in identifying muon neutrino charge carrier events.

III. Binary Classifier

A. Image Normalisation

A key consideration when constructing our CNN training is normalisation of our data set; broadly speaking, normalisation of the images allows faster convergence of the mode. As shared weights of the network in unnormalised data having varying calibrations for different features the cost function can converging very slowly and ineffectively thus normalising the data makes the cost function much easier to train. However, we must ensure our normalising preserves the same patterns in the batch, as although induvial images may appear the same, the correlation between pixel value and energy could be lost.

Shown in Fig. 5a is the distribution of energy against both maximum and minimum energy over batches of images with different normalisation; the total batch normalisation, such that the whole batch is divided by the total maximum value, preserves the distribution of unnormalized data, whereas individually normalised images, such that images are divide by the individual max value. This can lose the correlation in energy over the batch as the relative image pixel values contains information denoting energy- thus we use overall batch normalisation. Looking at histograms of energy and pixel values in Fig.5b, distribution between energy and pixel value show and highlights the batch normalisation preserves distribution.

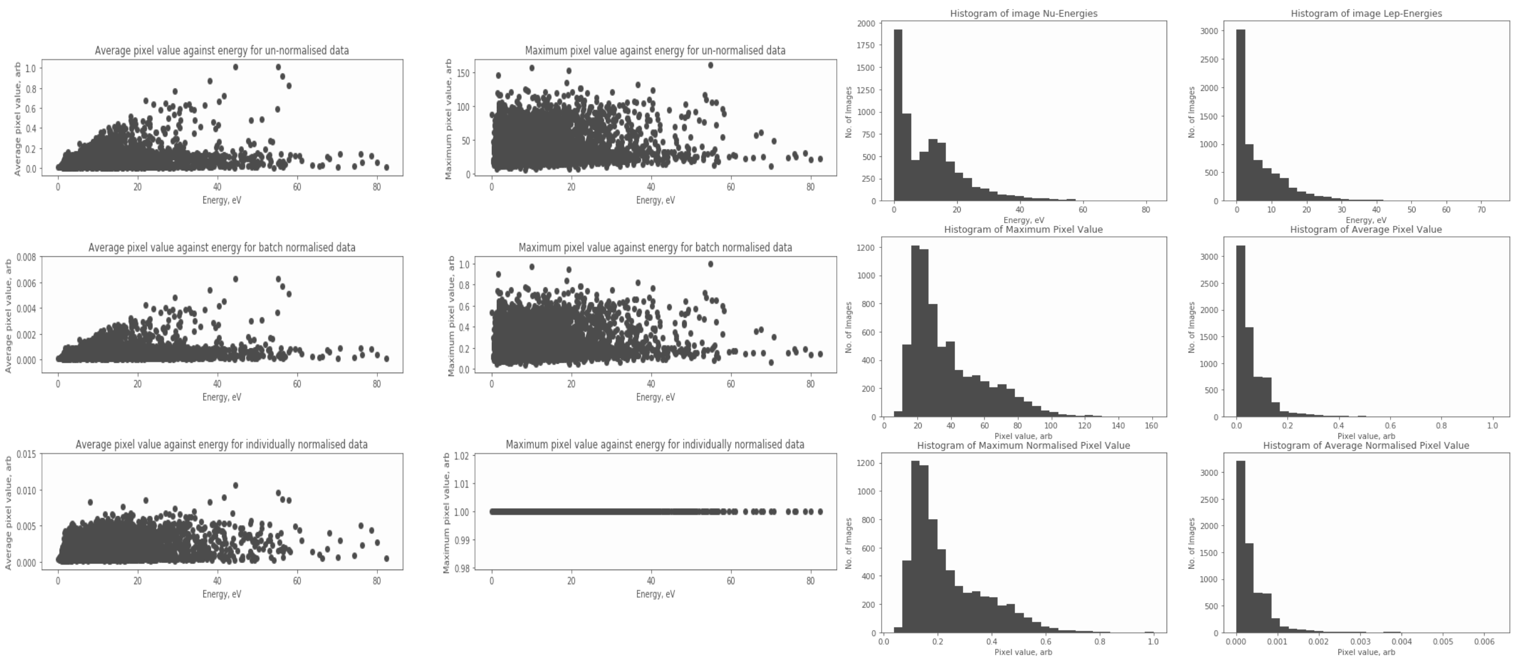
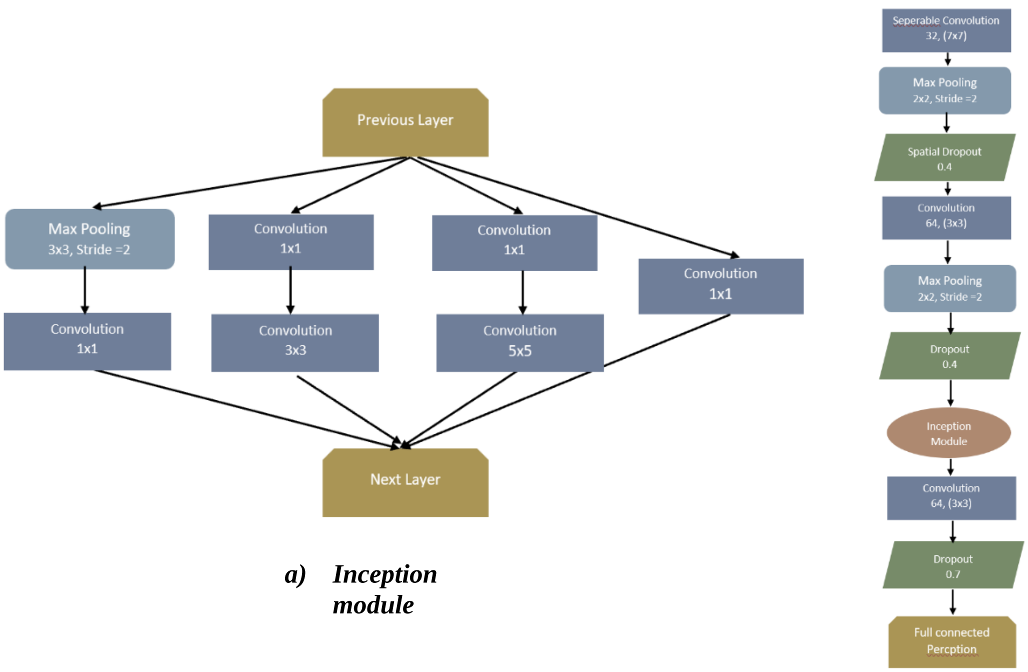


Figure 5: Various plots for trends in normalisation data

B. Model Architecture:

Our model is a binary classify to identify between muon neutrino charge carrier events against all other events. Our implementation of CNN use architecture like that used in paper by Aurisano, A. et al [1] and developed by Caffe [3] framework. We input the images together with each special dimension as input channels into a Separable Convolution of 32, 7x7 kernels followed by 3x3 Max Pooling and a 0.3 Spatial Dropout. The separable convolutions first performing a separate spatial convolution on individual channels followed by a pointwise convolution mixing the resulting output channels. The pooling acts to reduce dimensionality and we use spatial dropouts to avoid overfitting the data; it performs the same function as a normal dropout; however, instead of individual elements and neuron connections entire 2D feature maps are dropped and thus will help promote independence between feature maps and prevent overfitting.

Our model then passes through another convolution and pooling then we pass the data through an Inception layer. The Inception layer is a combination of a 1×1 convolutional channel, 3×3 convolutional channel, 5×5 convolutional channel and a 3x3 pooling by concatenation into a single output. Inception layers allow for more efficient feature learning through a dimensionality reduction with stacked convolutions and the module further helps with overfitting. Conceptually the convolution filters of different sizes will handle objects at multiple scales better and thus provide better feature learning. Further convolutions are applied before feeding the inputs into a fully connected perceptron layer with dropout to further avoid overfitting. The model then has a single output as it is a binary classifier, and our loss function is Binary Cross- Entropy and we use Binary accuracy as our metric. Visualisation of the model is given in Fig.6.



a) Inception module

b) Model architecture

Figure 6: Model Architecture Visualised

C. Training:

To train our model we concatenate multiple data sets of NOvA like images to give a batch of around 14,0000 images, of which 33% are used as validation data. Before training we ensure the binary classes are equalised, this is important as the data set given has biased, as discussed in the paper by Saito and Rehmsmeier [5], which means "the visual interpretability of [data] in the context of imbalanced datasets can be deceptive with respect to conclusions about the reliability of classification performance, owing to an intuitive but wrong interpretation of specificity". In addition, as the simulated data given has imbalance of about 88:12 for ones to zeros, this affects training as the classifier is biased to guessing ones and therefore gets in a fault minimum as it can get 88% accuracy only guessing 1; learning the plateaus and thus the model will not work.

Once data is balanced the data is given to the model in batch sizes of 100 and we train over 35 epochs with adam optimiser with learning rate of 0.001; achieving a final training binary accuracy of 0.8355 and loss of 0.3927; the validation loss is then 0.4107 and validation binary accuracy 0.8338 and as we can see the model has not overfitted as accuracy and loss have stayed below validation value, although as learning plateaus there is some erratic behaviour in the validation trend line due to the dropout layers(Fig.7).

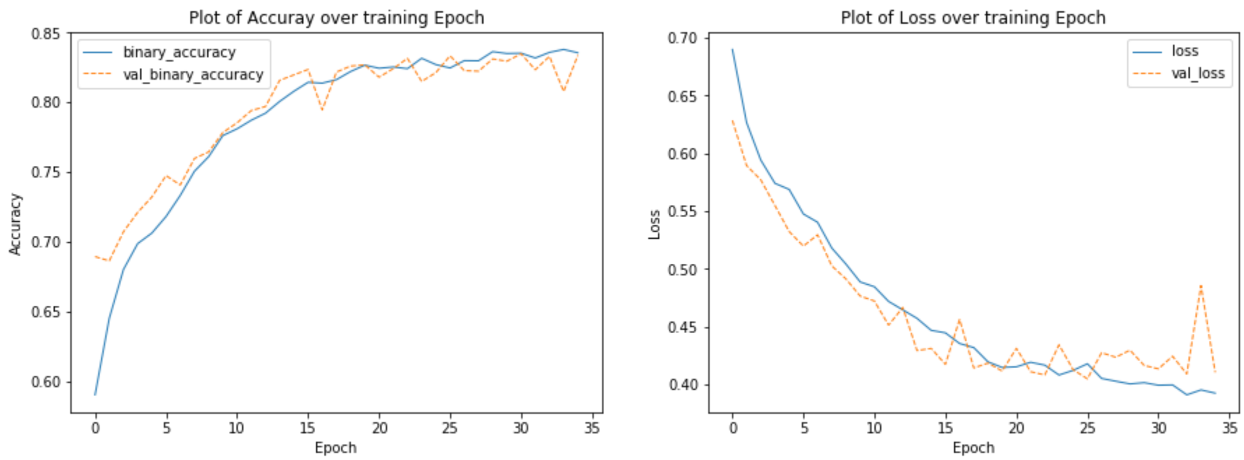


Figure 7: Training plots

D. Performance:

To test our model, we concatenate multiple data sets of NOvA like images to give a batch of around 60,0000 images. For comparison we test on both this unbalanced data set as well as balanced data set; the simulated data given has imbalance of about 88:12 ones to zeros, this gives test accuracy of 77% compared to the 82% on the balanced data which implies lower accuracy in predicting ones as the accuracy drops. To further this analysis the test images are subsequently used to produce an array of prediction; histogram of these predictions for both balanced and unbalance data sets are given in Fig. 8.

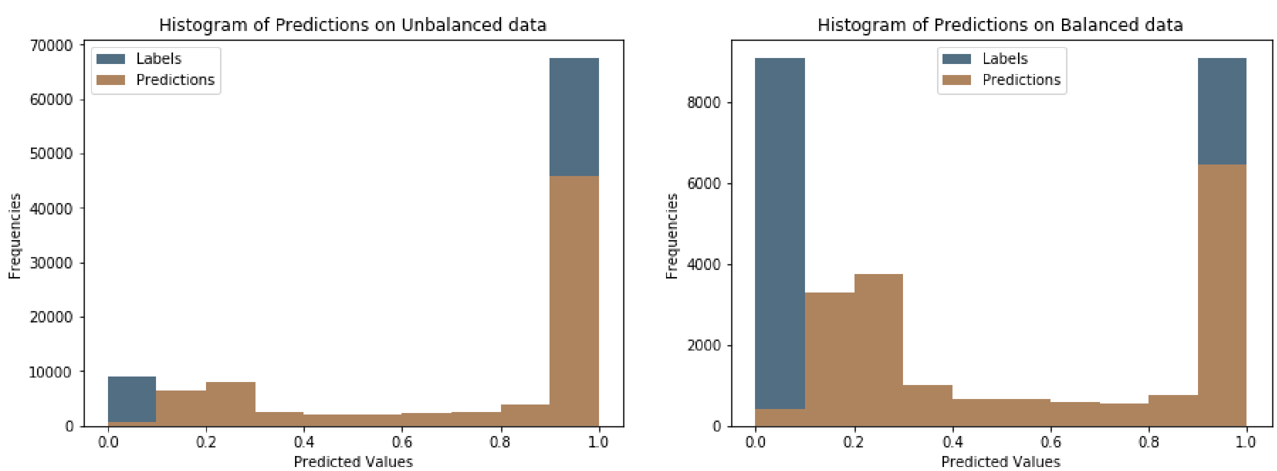


Figure 8: Histogram plots with prediction overlayed over labels

To further investigate this, we look at the variation of false positive and true positive rates of the predictions over different probability thresholds and the ROC curve produced. A receiver operating characteristic curve, ROC curve, is a plot of False Positive rate against True Positive rate and is used to illustrates the diagnostic ability of a binary classifier system as threshold is varied; an ROC curve thus highlights the trade-off between two rates for a predictive model. The ROC curve for this model characterises the performance well as the curve is close to the top-left corner indicate a better performance (random classifiers are expected to give a diagonal FPR = TPR). Our curve in Fig. 9 shows shape close to the top-left corner which again shows good classification; however, the non-symmetric shape shows the false positive rate increases faster than the true positive rate, showing overall bias in classifier to predict false positive.

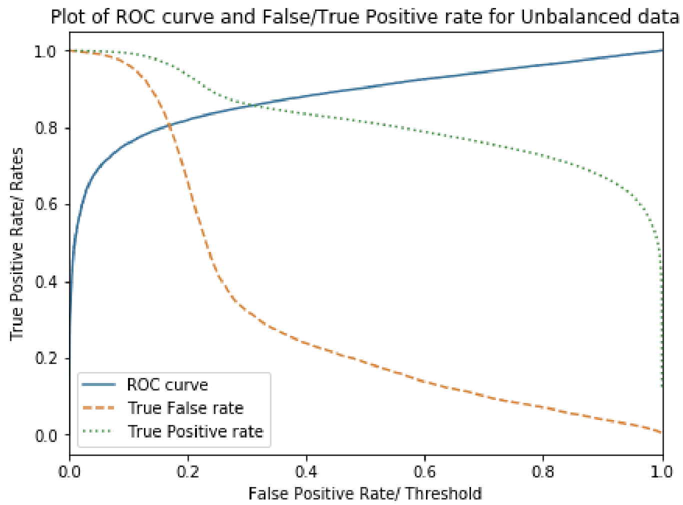


Figure 9: ROC and rates plot

Subsequently, we evaluate the accuracy of the predicted values over changing threshold on both balanced and unbalanced data to identify an unbiased threshold; for the unbalanced data this bias is evident in Fig. 10 as reducing the threshold, such that predictions all become false, increases accuracy as the overall unbalanced distribution has approximately 88% false. This is regardless of classifier performance and the reverse is seen when increasing threshold to one. Comparatively, for balanced data the accuracy for a threshold of both one and zero gives 50% giving a more accurate baseline for analysis. We can use the intersection of the two graphs to set an unbiases threshold for analysis of the meta data as increased accuracy beyond the intersection is only due to any bias in data set. Thus, the threshold used in further analysis is 0.5 which indicates good separation between classification; this is further supported by the shape which shows high accuracy over a wide threshold range implying correct classifications and separation over range.

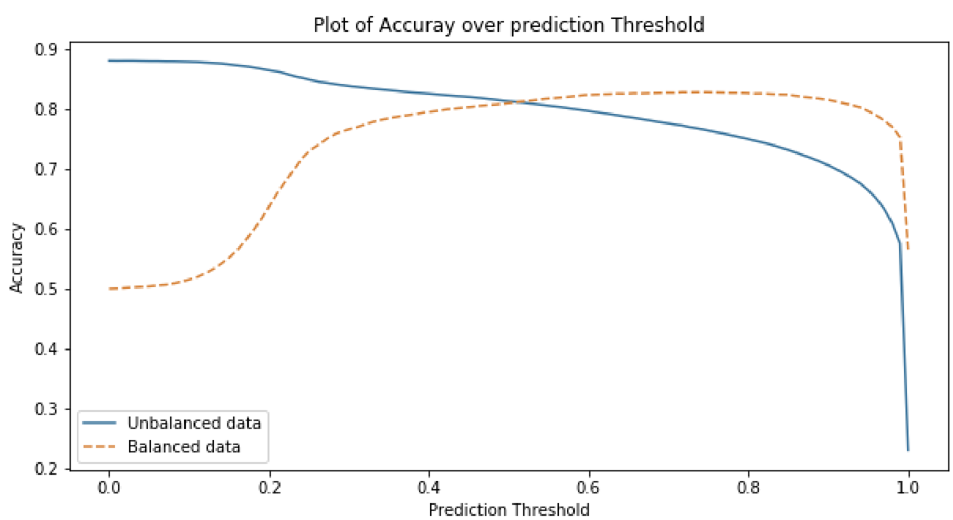


Figure 10: Accuracy of model over thresholds

IV. Testing Meta Data

A. Interaction Type:

Next, we analysis the performance of our model on the interaction type and find the comparative accuracy for QE, Res and DIS interactions. Predicting on these interaction image strata and using a threshold of 0.5 to calculate accuracy, we then find the corresponding accuracies, shown in Fig.11, are 93.8% for QE, 85.9% for Res and 76.5% DIS, comparatively testing other interactions gives 93.1%. This is the expected result as the QE are clearer two track interaction thus are more easily distinguished compared to DIS which are messier potentially with many tracks and showers. RES events have an accuracy in between as their tracks are less degradation then DIS but still less clear than QE events.

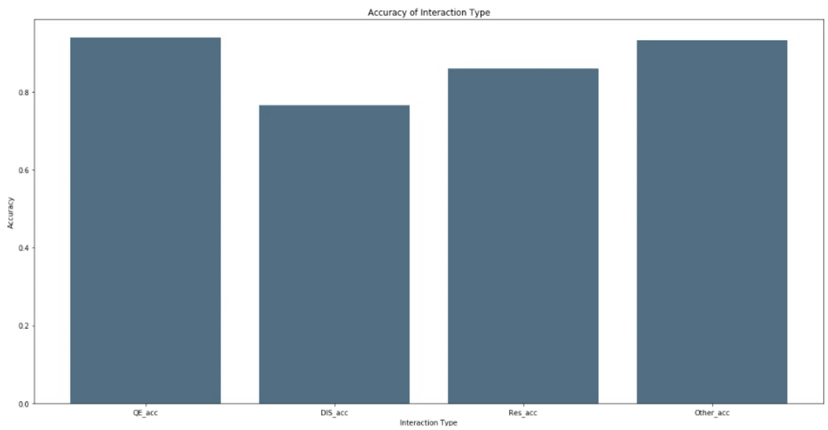


Figure 11: Accuracy for interaction type

To ensure this behaviour is not a feature of the threshold chosen, the accuracy for the interactions is then plotted over a range of frequencies. Fig.12 shows same characteristic shape of the unbalanced data in Fig. 11, however, here the interaction type is split and show same ordering as above over the whole range, thus the correlation holds over the whole range.

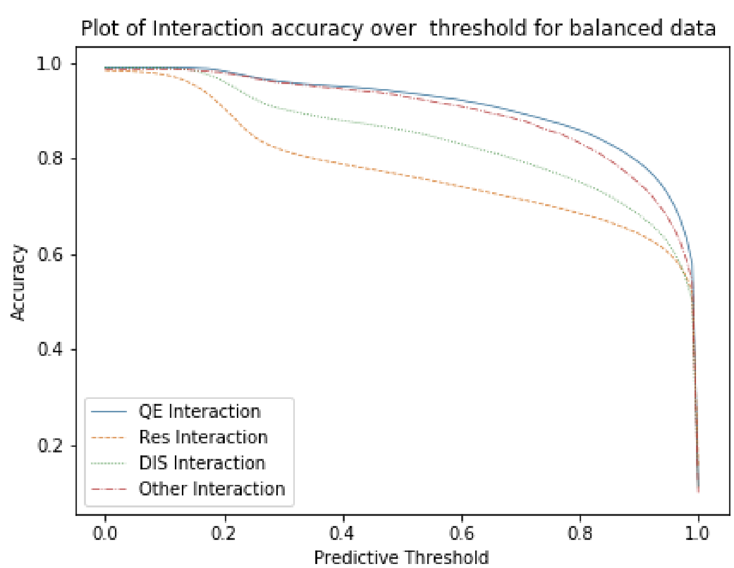


Figure 12: Accuracy of interaction type over thresholds

B. Energy:

Next, we look at how accuracy varies with both energy of the neutrinos and with the lepton energy; images are grouped into 5eV bins and the model is used to get predicted values and a 0.5 threshold is used to calculate the corresponding accuracy. Fig. 12 shows a plot of the accuracy over these respective energy ranges and the correlation between energy and accuracy. It can be seen from the plots there is a positive correlation between the neutrino/lepton energy value and the accuracy of the model. This corelation could be due to a variety of factors in the images; the higher the energy the clearer the tract as greater 'exposure' in the image and high energy particles are more penetrative therefore will have long and more distinct tracks, in addition, there is less overlap visual in differences interaction type at high energy. This allows for better feature extraction and correlation of underlying image topology to given interaction.

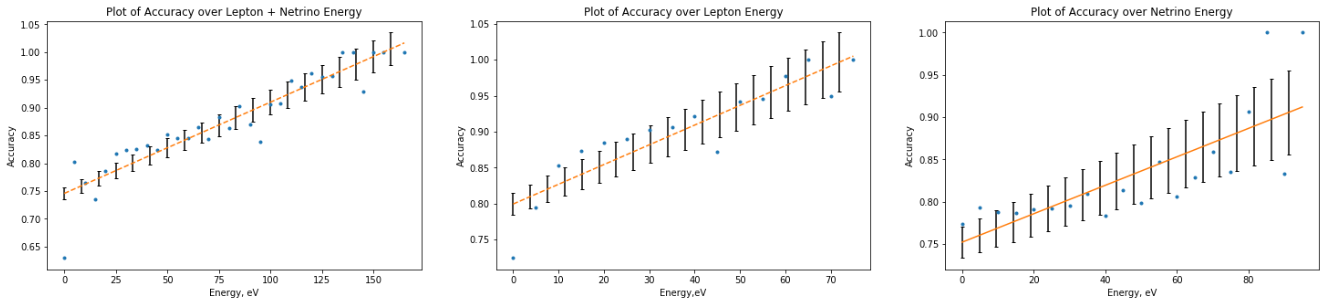


Figure 13: Accuracy against energy

C. Other Meta Data

Finally, we look to find if the model is dependent on any other meta data types; we test on images in strata of different meta data. Fig.14 show the accuracies for the final states of the images, for muon final states have lower accuracies than that of the other final state; when averaged we find muon final states have an average accuracy of 74.2% compared to the other final states which have average accuracy of 88.6%. This could be due to the ubiquity of muon there is a larger number and thus wide variety of interactions recorded, therefore, greater overlap with other interaction types which leads to incorrect categorisation. In addition, muons often make long, straight tracks, these simile tracks may also get miss categorised with other states.

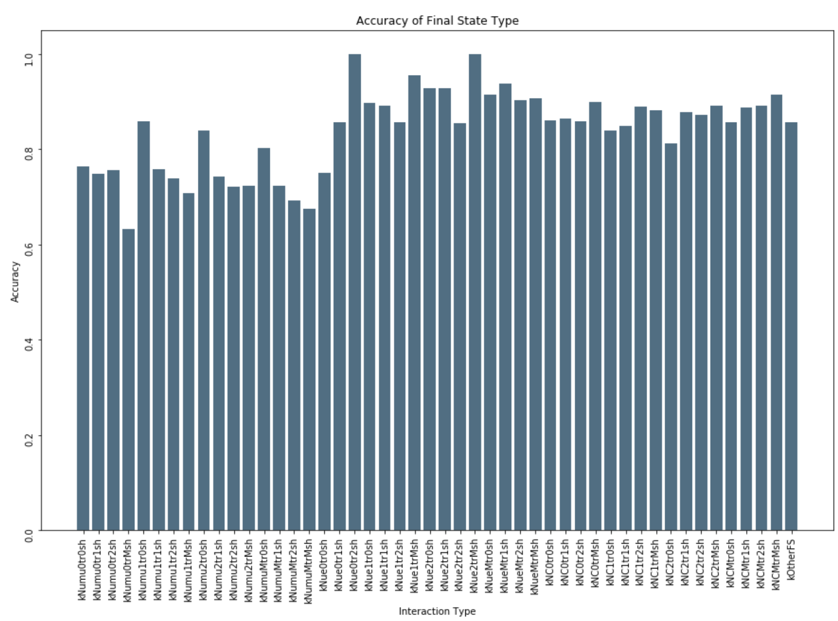


Figure 14: Accuracy Final State type

Analysing the other meta data shows no other dependency on accuracy as can be seen in the data in Fig.15. There is not trend in the data, however, this may require processing to highlight any trend, which could be done with further study.

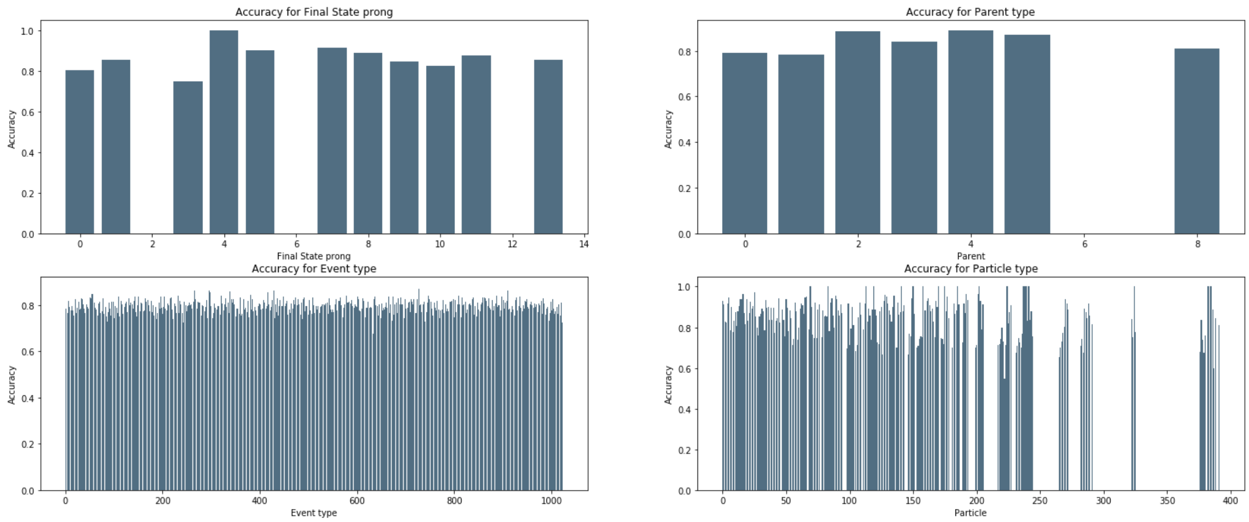


Figure 15: Accuracy against other meta data

Conclusion

Discussion and Extensions:

To conclude we have successfully shown the application of convolutional neural networks to a neutrino oscillation experiment, the Convolutional Neural Network (CNN) successfully acts as a binary classifier in identifying muon neutrino charge carrier events. We have shown the potency of inception modules and separable convolutions in creating a model to tackle this categorisation problem and produce a model with 83% accuracy in predictions; considering the large variation and subtlety in track types, this is a highly successful implementation.

Therefore, with no event reconstruction, we can build and train a single model which achieved good separation of desired signal (muon neutrino CC events) and background (other events) for NOvA-like data. CNN are subsequently a powerful tool for HEP and the problem of event classification. This approach should be expected to be transferable to a wide range of HEP images and analyses.

It has been showing the accuracy of such models depend on; the interaction type with quasi-elastic (QE) having highest accuracy and deep-inelastic scattering lowest accuracy; the neutrino and lepton energy with higher energies having greater accuracies; and the final state such that muons have lower accuracies.

This could extend by moving from a binary classifier to a sparse categorical classifier for all interaction types, or similar models could be applied to recover energy of images. This application can also be applied to other image analysis such as that of supernova and astrophysics data as well as other high energy physics.

References

[1] Aurisano, A., Radovic, A., Rocco, D., Himmel, A., Messier, M.D., Niner, E., Pawloski, G., Psihas, F., Sousa, A. and Vahle, P. (2016). A convolutional neural network neutrino event classifier. Journal of Instrumentation, 11(09), pp. P09001–P09001.

[2] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A. (2014). Going deeper with convolutions. [online] Available at: https://arxiv.org/pdf/1409.4842.pdf.

[3] E. Cengil, A. Çınar and E. Özbay. (2017). "Image classification with caffe deep learning framework," 2017 International Conference on Computer Science and Engineering (UBMK), pp. 440-444, doi: 10.1109/UBMK.2017.8093433.

[4] Psihas, F., Groh, M., Tunnell, C. and Warburton, K. (2020). A review on machine learning for neutrino experiments. International Journal of Modern Physics A, [online] 35(33), p.2043005. Available at: https://arxiv.org/pdf/2008.01242.pdf.

[5] Saito, T. and Rehmsmeier, M. (2015). The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. PLOS ONE, 10(3), p.e0118432.