PHILOSOPHICAL TRANSACTIONS A

rsta.royalsocietypublishing.org

Research

Article submitted to journal

Subject Areas:
Particle Physics, Machine Learning

Keywords:
Monopoles, Deep Learning, Convolutional Neural Networks, Nuclear Track Detectors

Author for correspondence:
A. J. Bevan
e-mail: a.j.bevan@qmul.ac.uk

Machine Learning Techniques for detecting topological avatars of new physics

A. J. Bevan[1,2]

[1]Particle Physics Research Center, Queen Mary University of London, London, E1 4NS, United Kingdom

[2]Alan Turing Institute, British Library, 96 Euston Road, London NW1 2DB

The search for highly ionising particles in nuclear track detectors (NTDs) traditionally requires experts to manually search through samples in order to identify regions of interest that could be a hint of physics beyond the Standard Model (SM) of particle physics. The advent of automated image acquisition and modern data science, including machine learning-based processing of data presents an opportunity to accelerate the process of searching for anomalies in NTDs that could be a hint of a new physics avatar.

The potential for modern data science applied to this topic in the context of the MoEDAL experiment at the Large Hadron Collider (LHC) at the European Centre for Nuclear Research, CERN, is discussed.

© The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.

THE ROYAL SOCIETY PUBLISHING

1. Introduction

The Higgs boson discovery at the CERN Large Hadron Collider (LHC) in 2012 was the culmination of the search for a particle that started in 1964. It took 48 years from the postulation of this particle to its discovery [1,2]. A monumental task that is unparalleled in particle physics. Magnetic monopoles were postulated well before the Higgs, yet as of the time of writing these proceedings no incontrovertible evidence of fundamental particle monopoles has been found. However there are some interesting hints, some of which are discussed below.

There are different types of magnetic monopole: Dirac [3], t'Hooft-Polyakov [4], and Cho-Maison [5,6] monopoles were postulated in 1931, 1974 and 1997, respectively. The theoretical motivation for these particles are varied. The Dirac monopole is a consequence of the properties of the wave function hence are predicted by quantum mechanics. The existence of isolated magnetic poles is linked directly to the quantisation of electric charge in the theory. The t'Hooft-Polyakov monopole is motivated by Grand Unified Theories (GUTs) and is expected to be massive. There could no monopoles within our observable Universe [7], the implication of that would be a futile search for a particle that would be effectively impossible to detect. From an experimental perspective the prospects of searching for that signature are extremely daunting. The Cho-Maison monopole is a hybrid between the Dirac and t'Hooft-Polyakov monopoles.

The theoretical motivation for such particles, including the link to the foundations of quantum mechanics and GUTs makes searching for these particles compelling. Reverting to the Higgs boson for a moment it is worth reflecting on the dual between application of quantum field theory to condensed matter and particle physics that led to the prediction of that particle in 1964. The fact that, in condensed matter, Dirac monopoles have been found to exist [8] is somewhat reassuring that there might also be a particle physics analogue.

Ionising particles passing through material will lose energy with a characteristic energy loss as discussed in [9]. Those with electric charge obey the Beth-Bloch energy loss relation:

$$\frac{dE}{dx} = 4\pi N_A r_e^2 m_e c^2 \frac{Z}{A}\frac{z^2}{\beta^2}\left[\frac{1}{2}\ln\frac{2m_e c^2\beta^2\gamma^2 T_{max}}{I_e^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2}\right], \tag{1.1}$$

where N_A is Avagadro's number, r_e is the classical electron radius, m_e is the electron mass, c is the speed of light in vacuum, $Z(A)$ is the atomic number (mass), z the charge of the incident particles in units of the electron charge, β the velocity of the incident particle as a fraction of c, $\gamma = \sqrt{1/(1-\beta^2)}$ and I_e is the mean ionisation potential of the medium. T_{max} is the maximum kinetic energy that can be transferred in a single collision. The δ term is an ionisation energy loss correction. However particles with magnetic charge obey a different energy loss characteristic that is not suppressed by a factor of β^2:

$$\frac{dE}{dx} = \frac{4\pi e^2(ng)^2}{m_e c^2} n_e \left[\frac{1}{2}\ln\frac{2m_e c^2\beta^2\gamma^2 T_{max}}{I_m^2}\frac{1}{2} - \frac{\delta}{2} + \frac{K(|g|)}{2} - B(|g|)\right]. \tag{1.2}$$

Here ng is the magnetic monopole charge ($n = 1, 2, 3, \ldots$), n_e is the number of electrons per unit volume in the medium, $I_m = I_e e^{-D/2}$, where D is dependent on the element and the K and B factors are correction terms. This difference means that the usual Bragg peak characteristic of a highly ionising standard model (SM) particle such as a proton or alpha particle will be very different from the energy loss of a magnetically charged particle. The former are well known to range out as the the energy loss on entry to material is relatively low, and as the particle slows down there is an enhancement in energy loss, which is maximal as the particle comes to a stop. This is the well known underlying principle that underpins modern hadron therapy, such as proton therapy. Magnetically charged particles will not range out in the way that SM particles do, and this difference allows one to search for anomalous energy-loss signatures in detectors that would signpost avatars of new physics.

In 1975 the results of a balloon experiment using a stack of Lexan sheets were reported by Price et al. [10]. They observed a highly ionising particle traversing 33 Lexan sheets in the stack and concluded that this signature was compatible either with a nucleus with $125 \leq Z \leq 137$ and $\beta \leq$

0.92, or a magnetic monopole with $g = 137e$. The track structure was found to be consistent with a nucleus of $Z \simeq 80$ and $\beta = 0.5^{+0.10}_{-0.05}$ with a downward motion, or a monopole with that same value of g. These two interpretations of the data only had one consistent conclusion: that the particle was a monopole with $g = 137e$, and a limit on the mass was placed, it being greater than $200m_p$, where m_p is the mass of the proton. Following dialogue in the literature, subsequently the same team concluded other possible new particles could explain the event observed [11]. This mass range is within reach of experiments studying collisions of pairs of protons at the LHC, where ambiguous interpretation of high Z nuclei would not occlude interpretation; however other new physics scenarios, such as anomalously charged SUSY particles could do so.

Superconducting Quantum Interference Devices (SQUIDs) have been used to search for magnetic monopoles. Cabrera reported observing a single candidate monopole event on the 14th February 1982, consistent with a one Dirac unit of magnetic charge in a SQUID detector [12]. This candidate corresponds to a significant flux quanta jump in data at a given instant, which was uncharacteristic of stable operations or other interventions such as LN_2 or LHe transfers. A number of other potential causes were ruled out. A subsequent experiment by Caplin et al., similarly reported observing an unexplained event on the 11th August 1985 [13], while two other experiments by that time failed to observe any anomalies [14,15]. Similarly no subsequent experiment has observed anomalies that could be considered a monopole candidate. While it is not possible to rule out the monopole hypothesis for the Cabrera event, it is plausible that this could have been a spurious signal that subsequent experiments with redundant design have been able to reject as monopole candidates.

Detection mechanisms for magnetic monopoles pursued by experiments in recent years fall into three categories; one can either trap the particles in material, and subsequently search for magnetic anomalies in SQUID based magnetometer systems. The second follows the method of Price et al.: to use arrays of Nuclear Track Detectors (NTDs) that are chemically etched to enhance the damage caused by charged particles traversing the plastic (Section 2). The third method is to measure the electronic charge generated in some active sensor medium when a monopole passes through parts of its system.

The MoEDAL experiment at the LHC is dedicated to the search for magnetic monopoles and other highly ionising particles [16]. It consists mainly of two passive detector systems. One is a trapping detector: Aluminum rods that are used to trap monopoles passing through the material. MoEDAL has published a number of results on the search for monopoles using the trapping method, the most recent result of which can be found in Ref. [17]. The second passive detector system is the NTDs system - stacks of plastic that become damaged when charged particles traverse them. MoEDAL also has an array of active sensors that could be used to search for monopoles passing through them as a function of time. NTDs are discussed in more detail in Section 2. These proceedings introduce a new method of NTD analysis using modern Artificial Intelligence (AI) deep learning methods in Section 3, where opportunities and challenges for the method are discussed in Section 4.

2. Nuclear Track Detectors

The NTD (plastic) foils used by MoEDAL are CR39 and Macrofol, supplied by the UK company TASL. When an ionising particle traverses the foil it loses energy, some of which can damage the hydrocarbon bonds of the plastic. The plastic is subsequently exposed to an etching solution, where the nanoscopic damage to the polymer is amplified by removal of plastic in the vicinity of the damage. Stacks of NTDs exposed to Highly Ionising Particles (HIPs) will have a series of holes in them post-etching. The resulting etch pits in test-beam samples exposed to 13 GeV Xe ions are tens of microns across. Such holes are large enough to be imaged by a fairly standard scanning system, and dark field illumination is a convenient method of establishing if a hole candidate is through-going, or if it is just a surface pit. Details of the etching method are refined in order to ensure that the NTDs reveal the damage induced by charged particles, but not to the extent that they are eaten away by over etching. The etching process is illustrated in Figure 1. Figure 2 shows

CR39 and Macrofol NTD samples from the MoEDAL experiment that have been exposed to 13 GeV energy Xe ions in a test beam. The CR39 sample shows sharp dark circles, which have a faint 'shadow' above and to the right of those circles. The dark circles correspond to entry holes of the CR39 top surface and the shadow corresponds to the exit hole on the bottom surface. These examples indicate that the NTD foil was placed in the test beam at an angle of about $10 - 15^{\circ}$ relative to normal incidence. The Makrofol sample image from the same test beam is shown on the right hand side of Figure 2, and shows much more activity, including surface pits that are not associated with any through going holes. These pits can appear on the top surface or the bottom surface. As both images in the figure are focused on the top surface, the objects with sharp outlines are associated with the top of the NTD and the others are associated with the bottom surface.

The images in Figure 2 have been recorded using an Optical Gaging Products (OGP) Flash 200 "Smart Scope" coordinate measuring machine. Images corresponding to regions of interest from NTDs like this have been used to construct training and test data for the machine learning methods discussed in these proceedings. The method used is to acquire images with the Flash 200; these are manually inspected to identify regions of interest corresponding to through going holes and surface pits on the top and bottom of the NTD. This is a time consuming, but necessary precursor to applying a supervised learning technique to the data as described below. An advantage of using the Smart Scope for image acquisition is that we are able to make reliable measurements of any features observed. For example the through going holes for this sample were measured to have a diameter of 33μm in the CR39 sample.

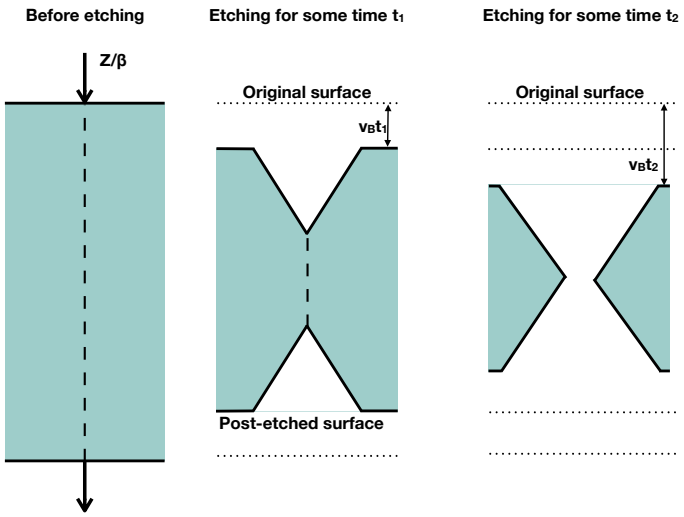

Figure 1. Cartoon of the effect of etching on the surface of NTD films. The etching rate is dependent on the damage.

3. Machine learning

Machine Learning and AI have long been used in particle physics for classification and regression problems. The recent development of deep learning has not only captured the imagination of scientists in many fields, but made a significant impact on results for certain problem areas. There is a broad set of models that can be applied to the task of identifying through going etch pits from images in NTD data. These proceedings consider the application of the deep learning Convolutional Neural Network (CNN) approach derived from the methodology developed by Fukushima [18] and inspired as an attempt to model the behaviour of the processing

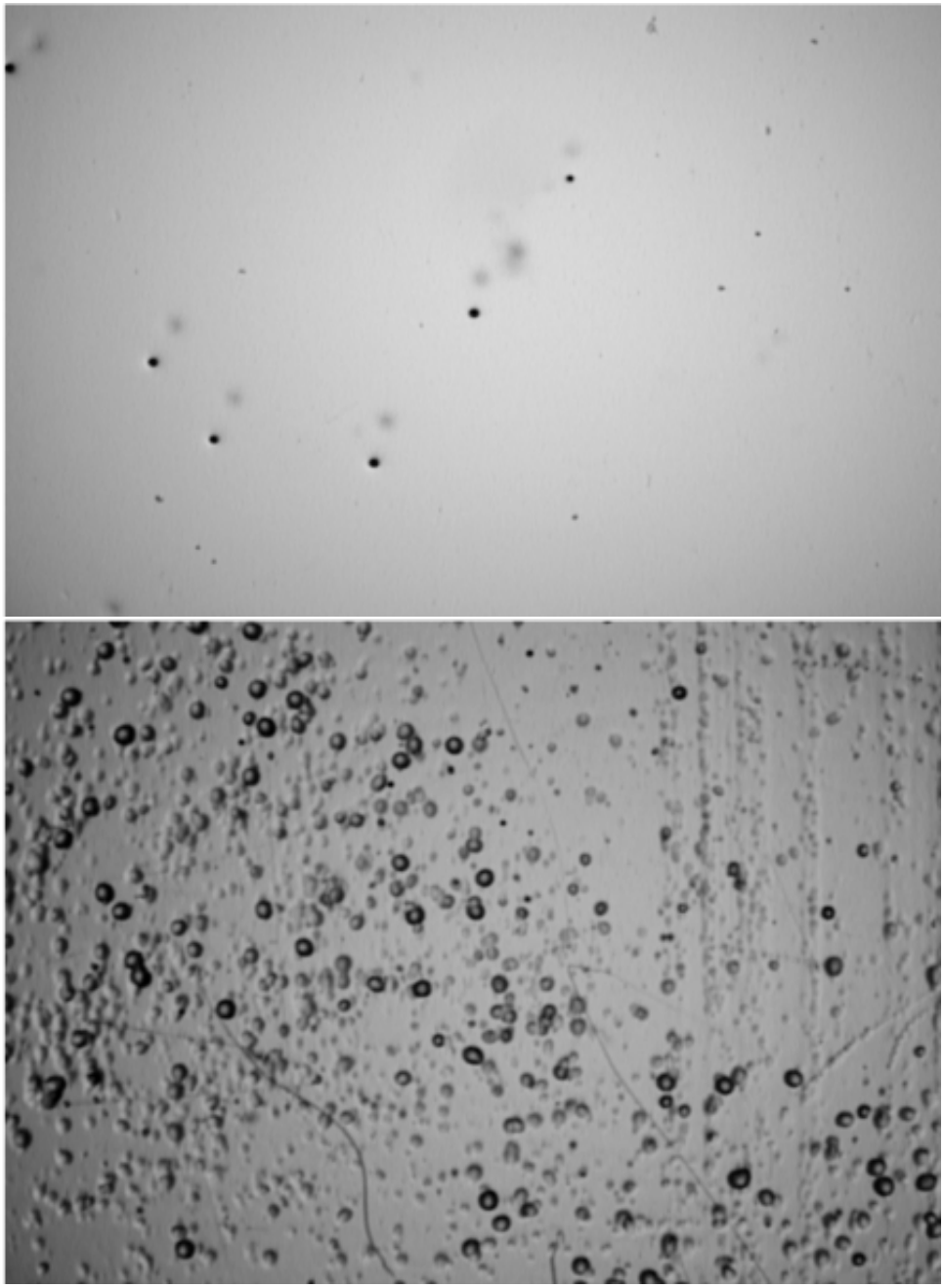

Figure 2. (top) CR39 and (bottom) Makrofol exposed to a Xe test beam and subsequently etched.

performed by the visual cortex of cats. The CNN approach was popularised by a series of high-profile computer science results where significant performance gains were made with image classification problems as a part of the Image Net Large-Scale Visual Recognition Challenge. The AlexNet and Inception CNN-based models being two important variants [19,20]. The models employed for this work are simpler architectures as a reflection of the simpler problem that is being addressed, and as initial test have shown high accuracy models are attainable, even with

small training samples. Ultimately more complex architectures may be beneficial, however we have insufficient training samples to explore such possibilities at this time.

The input feature space for a CNN is in the form of an image with a given (colour channel) depth. The MoEDAL data being processed use a depth of 1, namely a single number to assign a grey scale value to a given pixel. Thus the input feature space is just an $n \times n$ pixel array with each pixel corresponding to a single number [0, 255] renormalised to an appropriate range on the basis [0, 1] following standard data pre-processing conventions [21].

The MoEDAL NTD image data could be processed by a multilayer perceptron (MLP), an artificial neural network based on on an ensemble of Rosenblatt's perceptrons [22] each perceptron takes an input from all features in order to process that information to produce an output. The functional form of this transformation is based on the matrix multiplication $w^T x + \theta$ and the output of a node in the network is simply the mapping $y = f(w^T x + \theta)$. A 2-layer network can be represented as $y = g(f(w^T x + \theta))$, where weights w and θ in the second layer have been suppressed, and similarly for deeper networks. The functions f and g are one to one mappings known as activation functions. The Delta Rule, or back propagation can be used with a gradient descent based optimisation algorithm to determine the values of the weights. However one disadvantage of this method is that the MLP approach links all input features to all nodes in the first hidden layer of the network, and subsequently all nodes in that layer normally get linked to all notes in the next layer. The number of weights to be determined rapidly increases and the understandability of the model for MLPs becomes challenging.

CNNs are constructed from convolution and pooling layers. With a CNN only a local subset of input features are processed by a convolution kernel (typically and $M \times M$ filter, where M is an odd number). The convolution filter is then run across the image in order to reconstruct a new image (or feature map). This is the analogue of a node in an MLP. Multiple convolution kernels are used to create a set of feature maps. The convolution process again is described by $f(w^T x + \theta)$. However here instead of $n \times n + 1$ weight parameters we have $M \times M + 1$ weight parameters to determine for a convolution 'node' in the network. Each convolution kernel learns to identify some aspect of the input image (for example edges or shape information). We use padding around the input image to ensure that the feature maps obtained have the same dimensionality as the input images. This means we add a border of zeros around the input image with a size of $(M - 1)/2$ to avoid down-sampling the information content of the input image.

Pooling layers are used to intentionally down-sample the dimensionality of the information being processed in a CNN [23]. We use a 2×2 max-pooling algorithm that identifies the maximum value for each distinct 2×2 array of pixels in the image, and uses that value to represent the cluster of pixels in a down-sampled image. This means we can reduce the number of features from one set of feature maps to the next by a factor of 4. The use of the max-pooling algorithm is intended to suppress noise expected to dominate the data, for example in the case of the CR39 and Makrofol samples given in Figure 2, where the majority of the image area is devoid of both signal and spurious signal (one sided etch pits). The expectation is that max-pooling applied to feature maps for this problem will not lose information about the signal and spurious signal of interest.

After successive convolution and pooling layers one has to identify if a region of image contains signal, spurious signal or noise. For that final parse of the information normally one brings together all of the pixels in the feature maps of the final convolution or pooling layer and feeds that into a so-called fully connected layer. This fully connected layer is just like a hidden layer of an MLP described above. The CNN architecture used for this analysis is shown in Figure 3. A set of four repeating units of convolution kernels followed by max-pooling layers is used, each feeding into the next. The output of the final max-pool layer is then fed into the first of two fully connected layers, before a single node is used to combine the information into a single score. The CNN is implemented in TensorFlow and in Keras [24,25], and the ADAM optimiser [26] is used to learn the model hyper parameters by minimising an L_2 objective function. This can be

used for regression or classification of an image. We train three 'expert' networks to identify the following hypotheses:

- Through going holes (signal) vs noise;
- Top surface etch pits (spurious signal) vs noise;
- Bottom surface etch pits (spurious signal) vs noise.

The outputs of the three expert CNNs are then combined in order to identify which hypothesis is the most likely outcome for the trained models.

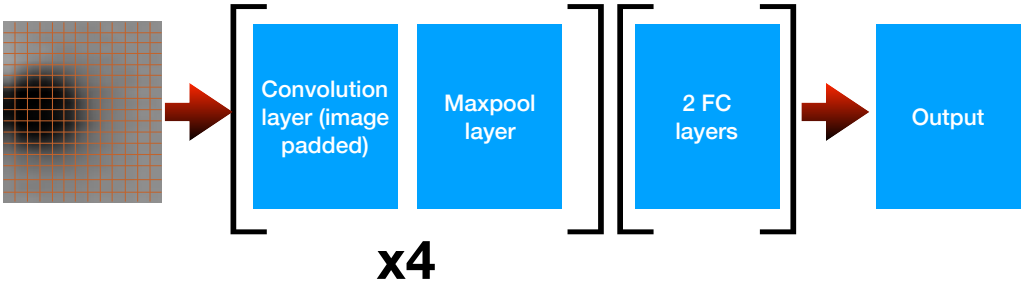

Figure 3. The CNN model used for this analysis: a set of 4 pairings of convolution and max-pool layers, followed by two fully connected (FC) layers that feed into an output node used either for inference or binary classification.

(a) Image resolution

The study performed here uses data acquired using a high resolution optical coordinate measuring machine. The precision of the measured surface obtained optically (or using a laser scan of the plastic surface) results in a sub-micron lateral measurement accuracy and spatial resolution of the imaging system exceeding that specification. Understanding the resolution required to solve the problem of distinguishing between through going holes, surface etch pits (top or bottom surfaces) or noise is essential to ensure that computing resources are used appropriately, and this is also linked to the training sample size required to avoid over training the model. Figure 4 shows typical benchmarking results for ensembles of pseudo images generated from stitching together regions of sample images in as an attempt at a fast simulation of NTD data. The high resolution images were found to have an issue with over-training evident as the accuracy of model predictions varied by a few percent for the test and train samples. Down-sampling the images results in a reduction in the number of hyperparameters required for the CNN. Correspondingly well trained models are obtained as a result. For samples with 128×128 pixels or smaller the accuracy of model prediction is typically in excess of 99%. There is a factor of 7 reduction in training time required between a 384×384 and a 128×128 set of image data; with further reduction in time required to learn a robust model for even smaller image sizes.

(b) Feature maps and model performance

Figure 5 shows a set of feature maps derived from input images. The input images are generated by oversampling an original image in order to make a unique independent image to process using a model. Example feature maps obtained from three convolution layers are shown. As the convolution layer number increases, the level of detail in a feature map decreases as the

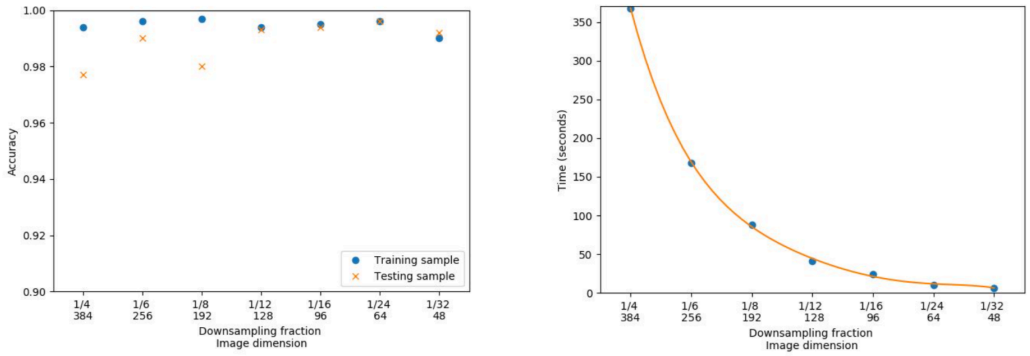

Figure 4. (left) Model accuracy for training (dots) and test (crosses) samples of data as a function of the image size in pixels (right) the corresponding change in time required for model training on a Tesla K40 GPU.

number of convolution and pool layers increases. i.e. Successive feature maps are blurred by their predecessor convolution and pooling steps. The feature maps reveal aspects of the images of interest, including the hole, and edge features. The fact that edge features have been learned by the feature maps in this example raises a concern as to what the corresponding convolution filters are learning. While the assumption is that labeled example images are presented to the CNN in order for the model to learn the difference between signal and noise, there is a potential disconnect as to what in the images is being used to learn that difference. It is important to check that the model has learned to address the intended problem.

While inspection of the feature maps of a CNN can provide information about what the model is learning, especially when there are complicated inputs, that is not sufficient to ensure that one can verify the model is addressing the particular question being posed through the training process. Much like the phrasing of the null and alternate hypothesis for a hypothesis test, it is all too easy to think you know what question is being addressed by a ML model when training or applying it to data. A closer analogy can be found in the problem of parameter estimation using the likelihood or least squares methods. Some problems have challenging convergence issues, or may converge to unphysical solutions. For example a polynomial probability density function written as $\sum_i a_i x^i$, where a_i is the coefficient of some feature x, and the sum is over n terms in the model, will readily yield negative probabilities to accommodate regions of low statistics.

Biases in the data set, such as the edge effects obtained in the oversampled images found in Figure 5 may influence the model. In retrospect this should not be a surprise at all as after all the CNN is just a function approximator used to map an input image into a single number. This issue of explainability and interpretability is something that is currently a popular topic in computer science, in part because of this kind of issue. There are methods designed to provide both local and global measures of the explainability of models. Three of these: Guided backpropagation, Grad CAM and the hybrid Guided Grad CAM [27] have been applied to this problem, and examples where signal has been correctly identified by the model have indeed been found to have been classified due to edge effects that bare no relation to signal in origin. While this work has addressed a number of issues regarding developing a new method of NTD analysis, there are still challenges to be overcome before a physics-ready system can be deployed for the LHC.

4. Challenges and opportunities

A significant challenge for NTD analysis is obtaining sufficient training and validation examples to learn a robust model. While the performance of the models obtained here is good, with a high

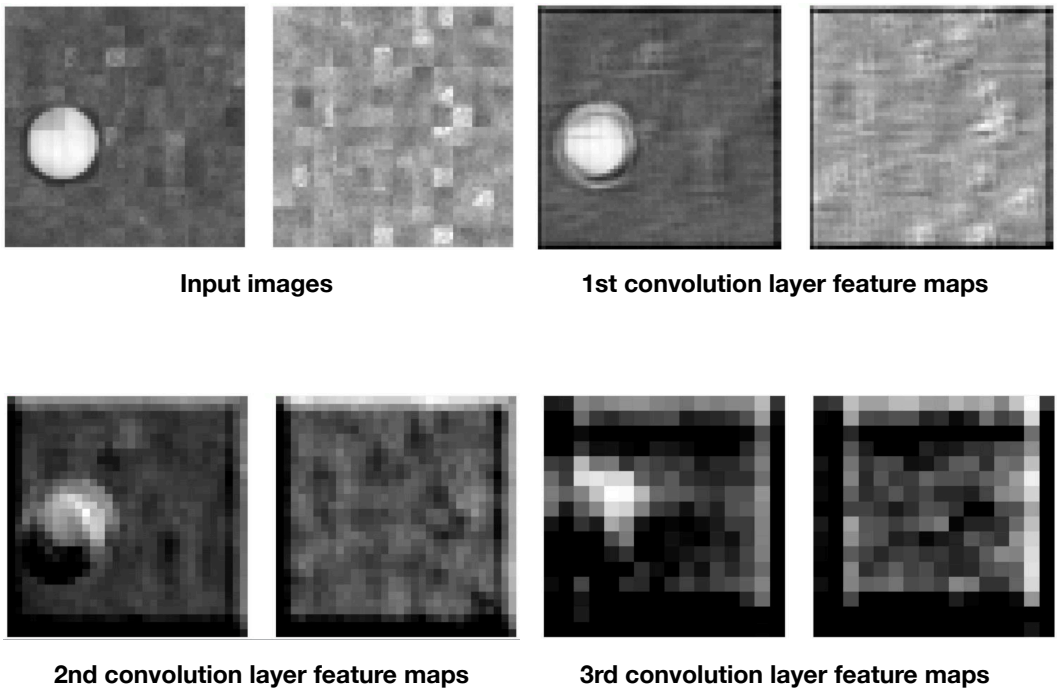

Figure 5. (top-left) Input image examples for evaluation by the CNN model and (other) feature maps created by the 1^{st}, 2^{nd} and 3^{rd} convolution layers.

classification accuracy, it is a concern that examples have been classified using features in images that are not associated with that classification type. Classifying examples correctly, but for the wrong reasons can lead to an unreasonable expectation of performance as has been reported in a number of other use cases, and is referred to as Horses in some circles [28]. Developing a means of understanding what makes a horse, and how to avoid that kind of misclassification will hopefully lead to insights that allow for more robust models to be developed for solving these problems.

The limited training samples available leads us to consider the possibility of using pseudo labeling to develop more robust trainings based on using an initial training on a subset of data to automatically label a larger set of data to use for a subsequent training; allowing for a semi-supervised learning. While this does not remove the issues of bias related to the smaller training, it can help develop a better generalisation performance [29].

An alternative to a semi-supervised learning approach could be to develop an detailed NTD simulation; modeling the fundamental interactions between ionising particles and the plastic NTD foils; simulating the damage induced by chemical etching and finally simulating the ability to image data. This task is non-trivial, however the understanding for experimentalists to simulate the response for both SM and non-SM HIP particles exists; as does the understanding of chemical etching of plastic. The final step requires translating the surface into a simulated image. This would have an advantage in that one could simulate images resulting from particles with different angles of incidence on NTDs. The change in trajectories from normal (or as in the case of the available test beam data, an angle of incidence of $10 - 15^{\circ}$ from the normal) will affect a given model's ability to learn how to discriminate between etch pits and through going holes. The ability for a given model to adapt to different particle trajectories should be studied to further explore the robustness of a given model, which in turn may affect the ability to track HIPs.

An improvement over the approach described here would be to replace the final output node with a softmax activation function, and use a softmax cross-entropy loss function. This would allow the model to simultaneously learn how to classify, or predict, a vector of outputs; making for a more complicated single training model, but removing the need for the three expert approach.

5. Summary

The advent of modern computing and AI, including deep learning methods presents a significant opportunity for NTD-based experiments to automate pattern recognition tasks and accelerate the ability to perform physics analyses. These proceedings have presented the problem of searching for highly ionising avatars of new physics at the MoEDAL collaboration, and explored modern deep learning algorithms implemented using TensorFlow and Keras to the problem. The opportunity that this technology brings is scalable fast reconstruction of data so that physicists can focus on the problems of reconstructing any events of interest, and subsequent interpretation. The challenges faced include the preparation of sufficiently large samples of data with which one can train robust models, with acceptable bias and variance, and with which one can benefit from the full potential of modern deep learning. A NTD simulation would help overcome a number of the current limitations encountered with this method. To truly rely on deep learning for physics applications one has to be able to explain the outcomes, and here I have presented some methods that have been explored in this emerging area of computer science.

Funding. This work has been supported by the Science and Technology Facilities Council, part of the UKRI.

Acknowledgements. The author acknowledges the MoEDAL collaboration and in particular the following collaborators in this area of work: Tom Charman, Krzysztof Furman, Jon Hays, Muhammad Karim, and Lewis Milward. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Tesla K40 GPU used for this research.

References

1. Aad G. et al. [ATLAS Collaboration], 2012, *Phys. Lett. B* **716** 1-29.
2. Chatrchyan S. et al. [CMS Collaboration], 2012, *Phys. Lett. B* **716** 30-61.
3. Dirac P, 1931, *Proc. R. Soc. Lond. A* **133** no.821, 60-72.
4. t'Hooft G, *Nucl. Phys. B* **79** 276-284 (1974); Polyakov A, 1974 *JETP Lett.* **20** 194-195, 1974 *Pisma Zh.Eksp.Teor.Fiz.* **20** 430-433.
5. Cho Y. M. and Maison D., 1997, *Phys. Lett. B* **391** 360-365.
6. Bae W. S. and Cho Y. M., 2005, *JKPS* **46** 791-804.
7. Lazarides, G. et al., 2918 *Phys. Rev. Lett.* **49** 1756.
8. Ray M. W. et al., 2014, *Nature* **505** no.7485, 657-660.
9. B. Acharya et al., 2014, *Int. J. Mod. Phys. A* **29** 1430050.
10. Price P. B. et al., 1975, *Phys. Rev. Lett.* **35** 487-491.
11. Price P. B. et al., 1978, *Phys. Rev. D* **18** 1382-1421.
12. Cabrera B., 1982, *Phys. Lett. B* **48** 1378-1381.
13. Caplin A. D. et al., 1986, *Nature* **321** 402-406.
14. Incandela J. et al., 1984, *Phys. Rev. Lett* **53** 2067-2070.
15. Bermon S., et al., 1985, *Phys. Rev. Lett* **55** 1850-1853.
16. Pinfold J. et al. [MoEDAL Collaboration], CERN-LHCC-2009-006, MoEDAL-TDR-00.
17. B. Acharya *et al.* [MoEDAL Collaboration], 2019, *Phys. Rev. Lett.* **123** 021802.
18. Fukushima K., 1980, *Bio. Cybernetics* **36** 193-202.
19. Krizhevsky A et al., 2012, *Proc. Advances in Neural Information Processing Systems 2012*.
20. Szegedy C., et al., arXiv:1409.4842.
21. Ed. Montavon G. et al., 2012, Neural Networks Tricks of the Trade, Chapter 1, Springer.
22. Rosenblatt F, 1958, *Psychological Review* **65** 386-408.
23. Boureau Y. L., 2010, *International Conference on Machine Learning*, Haifa, Israel.
24. Abadi M et al., 2015, https://www.tensorflow.org/.
25. Chollet F et al., 2015, https://keras.io.
26. Kingma D. P. and Ba J., 2014, arXiv:1412.6980.
27. Selvaraju R. R. et al., 2016, arXiv:1610.02391.

28. Sturm B. L., 2014, *IEEE Trans. Multimedia* **16** 1636âĂŞ1644.
29. Lee D. H., 2013, *ICML 2013 Workshop*.