PROCEEDINGS OF SCIENCE

Convolutional Neural Networks for event classification

Adrián Rubio Jiménez,* José Enrique García Navarro and María Moreno Llácer

Instituto de Física Corpuscular (mixed centre CSIC-UV),
Catedrático José Beltrán 2, Valencia, Spain
E-mail: adrian.rubio.jimenez@cern, jose.enrique.garcia@cern.ch, maria.moreno.llacer@cern.ch

Cutting-edge Artificial Intelligence is being implemented in a wide range of tasks in High Energy Physics (HEP) in order to facilitate the analysis of large datasets. However, visual recognition has not been explored as much in HEP for event classification. This study shows how Convolutional Neural Networks could be applied for such an important task, for which a novel method to represent the event information in images is explored.
This technique is applied for a classification problem corresponding to a search for Dark Matter in proton-proton collisions. The results obtained with this technique are also compared with the performance of a Boosted Decision Tree.

The Ninth Annual Conference on Large Hadron Collider Physics - LHCP2021
7-12 June 2021
Online

*Speaker

PoS(LHCP2021)264

1. Introduction

This study presents an innovative way for classifying particle collisions as those taking place in the Large Hadron Collider (LHC). Since visual recognition belongs to the state-of-the-art in Artificial Intelligence algorithms, a new technique using Convolutional Neural Networks (CNNs) is explored for event classification [1]. The samples used were produced by the DarkMachines initiative [2], using a quick detector simulation for a modified version of the ATLAS detector card [3]. The signal events considered for this analysis are mono-top processes, in which a top quark is produced in association with Dark Matter (DM) particles. Due to the low interaction probability associated to DM particles, these are not expected to interact with the detector. In particular, in this study, the selected events are required to have large missing transverse momentum (MET > 150 GeV), together with exactly one charged lepton (i.e. electron or muon) and at least one b-tagged jet coming from the decay of the top quark. The main background processes for this search are $t\bar{t}$, *t-channel* single top and *W+jets*, which need to be distinguished from the DM signal.

2. Visual recognition for event classification

To deal with event classification, in which a very rare process needs to be identified over a large data volume, there are standard Machine Learning (ML) techniques such as Boosted Decision Trees (BDTs) and Neural Networks (NNs). These multivariate techniques provide good results. However, other Artificial Intelligence (AI) technologies based on visual recognition have shown big improvement over last years, which could be taken as an advantage for event classification.

CNNs consist in a class of Neural Networks that uses images as inputs, preserving the spatial symmetries and the local structures of the images. To do that, the first layers perform two-dimensional operations[1] on the pixels, learning from the spatial structure and acting as a feature extractor. Subsequently, the resulting pixels are flattened in an array, acting as the input for some dense layers added in order to perform the classification. In this way, it might be possible with an appropriate representation, that CNNs could learn physics features from an image.

2.1 Transfer learning

Training a NN as supervised learning consists in an iterative process in which a large amount of parameters are updated in each step in order to minimize a function, called *Loss function*, which is related to the proportion of events that are classified wrongly. This process can be speed up using *transfer learning* [4], which is based on the idea that most of the basic features for image classification (lines, edges, etc.) are common, so one would only need to train for the specifics of the problem. Thus, this study will use the power of the well-known architecture VGG16 [5], freezing part of its parameters during the training process.

However, to be able to use images, a collision event needs to be represented as an image and this is the most challenging and innovative aspect of this approach.

[1] There are also CNNs which make analogous operations in three dimensions, working with 3D images.

2.2 Image representation of the events

For each collision, each object will be represented by a circle and its colour will correspond to the type: green for charged leptons (electrons and muons), pink for jets, red for b-tagged jets and blue for the missing transverse momentum (MET or p_T^{miss}). The direction of the particles is completely determined by the coordinate system of ATLAS, with the x-axis corresponding to the pseudorapidity η in the range [-4.5, 4.5] (related to the angle with respect to the beam) and the y-axis corresponding to the azimuthal angle ϕ transverse to the beam axis in the range [-π, π]. This can be understood in Figure 1 for an image example[2]. Finally, the transverse momentum of each particle is proportional to the size of the circles, following a binned scaling and preventing larger p_T particles from exceeding the image limits. In this way, the four-momentum of the particles, their type and the most essential event characteristics are encoded in the image.

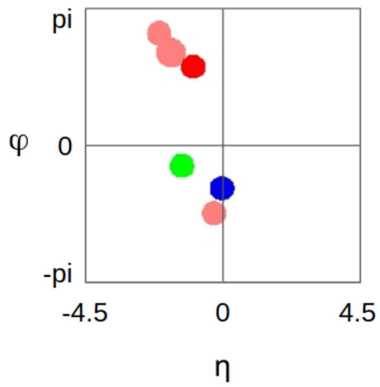


Figure 1: Image representation of the principal characteristics of a collision: pseudorapidity (x-axis), azimuthal angle (y-axis) and p_T (circle diameters) of every particle. Each colour corresponds to one type of particle: charged leptons (electron and muon) are green, jets are pink, b-tagged jets are red and the missing transverse momentum is given by a blue circle.

Thus, an intuitive representation is proposed in this study, in which more specific observables (beyond the kinematic variables) could be added to provide more information to the CNN. In addition, the CNN is seeing not only the four-momentum of the individual objects, but also the spatial relations between them, from which some important information can be learnt. In this sense, it is important to note that by using this technique a lot of information is provided, but the CNN will decide by its own "where to look" during the training process. However, this is not an obscure algorithm in which we do not have any access but there are ways to look into the intermediate layers to check what features of the images is the CNN looking to, but this is left for future work.

3. Results

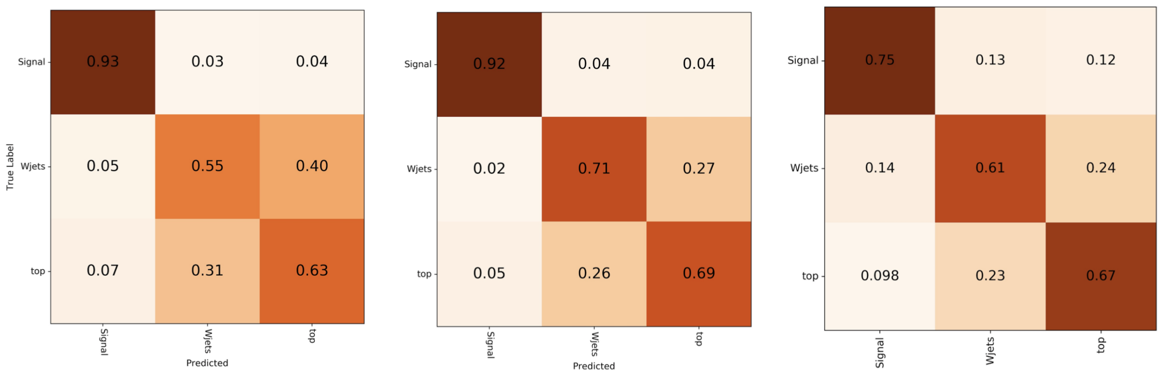
Balanced datasets are used (around 11900 for each class), which have been split into training (70%), validation (10%) and test (20%) samples. Physics processes including top quarks, like $t\bar{t}$,

[2]The MET, for its part, always has pseudorapidity equals to zero because this quantity is defined as the imbalance of momentum in the transverse plane to the beam axis. By the definition, the blue circle will appear in the centre (η = 0) of all the images.

PoS(LHCP2021)264

t-channel and Wt single top processes, have been merged in the same "*top*" class. Therefore, the classifications performed consist of three classes: *W+jets*, "*top*" and *signal* (a mono-top process), and the output of the CNN is the label of the class.

A comparison of confusion matrices is shown in Figure 2 between a deep CNN with 13 layers trained from scratch (Figure 2a) and the VGG16 architecture using transfer learning (Figure 2b). A similar result is observed, in which the signal is classified much better than the background processes, but transfer learning achieved a better performance in general. The deep CNN has been completely trained, consisting of $\sim$ 2.5 millions of parameters, whereas the VGG16 parameters have been partially tuned, freezing 7.6 millions of parameters from a total of 16.3 millions. Both of these approaches have room for optimisation, but it can already be seen that transfer learning is in fact a very suitable option.



(a) **(b)** **(c)**

Figure 2: Confusion matrices for: (a) Deep CNN with 13 layers trained from scratch. (b) Transfer learning using the well-known architecture VGG16. (c) Boosted Decision Tree using the XGBoost library[6].

Comparing with the performance achieved using a more standard technique is another important check (Figure 2c). In this case, we have chosen a very extended ML technique such as a Boosted Decision Tree (BDT), using XGBoost [6]. Here, the procedure is different since the observables (the four-momentum of the particles and the event characteristics) are given explicitly to the BDT as numbers, so very different type of input preprocessing is required. A similar behaviour is also observed, although the BDT seems to identify the signal worse than CNNs for this particular problem (see Figures 2a and 2b). Obviously, this does not have to be the case for any other classification. In addition, both the CNNs and the BDT can be further optimised.

4. Conclusions

This study shows a feasible technique for event classification, for which a visual recognition algorithm together with an original way to encode the event information in images are used. In this way, the transfer learning approach looks to be very promising because the limits of the most powerful architectures has not been explored in depth for classifying physical processes. And the most important result comes from the similar performances provided by these CNN approaches and the BDT, which is widely used in many HEP analyses. This implies that the CNNs are actually learning physics from the images, instead of irrelevant details.

References

[1] C. F. Madrazo, I. Heredia, L. Lloret, and J. Marco de Lucas. Application of a Convolutional Neural Network for image classification for the analysis of collisions in High Energy Physics. *EPJ Web of Conferences*, 214:06017, 2019.

[2] Darkmachines initiative. https://www.phenoMLdata.org.

[3] Model-independent signal detection: A challenge using benchmark monte carlo data and machine learning. *Contribution 23 from Les Houches 2019 Physics at TeV Colliders: New Physics Working Group Report*, 2020.

[4] C. Tan et al. A survey on deep transfer learning: 27th international conference on artificial neural networks, rhodes, greece, october 4–7, 2018, proceedings, part iii. pages 270–279, 10 2018.

[5] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014.

[6] XGBoost documentation. https://xgboost.readthedocs.io/en/latest/.