Fast convolutional neural networks for identifying long-lived particles in a high-granularity calorimeter

J. Alimena\textsuperscript{a,1}, Y. Iiyama\textsuperscript{b} and J. Kieseler\textsuperscript{a}

\textsuperscript{a} EP Department, CERN, Espl. des Particules 1, 1211 Meyrin, Switzerland
\textsuperscript{b} International Center for Elementary Particle Physics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

E-mail: juliette.alimena@cern.ch

ABSTRACT: We present a first proof of concept to directly use neural network based pattern recognition to trigger on distinct calorimeter signatures from displaced particles, such as those that arise from the decays of exotic long-lived particles. The study is performed for a high granularity forward calorimeter similar to the planned high granularity calorimeter for the high luminosity upgrade of the CMS detector at the CERN Large Hadron Collider. Without assuming a particular model that predicts long-lived particles, we show that a simple convolutional neural network, that could in principle be deployed on dedicated fast hardware, can efficiently identify showers from displaced particles down to low energies while providing a low trigger rate.

KEYWORDS: Pattern recognition, cluster finding, calibration and fitting methods; Trigger algorithms; Trigger concepts and systems (hardware and software); Analysis and statistical methods

ArXiv ePrint: 2004.10744

\textsuperscript{1}Corresponding author.

\cite{ArXiv:2004.10744}
1 Introduction

Particles with long lifetimes are an important possibility in the search for new phenomena, and often appear in beyond the standard model theories, notably in models that describe the elementary particle nature of dark matter. When produced at the CERN Large Hadron Collider (LHC), these long-lived particles (LLPs) have a distinct experimental signature: they can decay far from the primary proton-proton (pp) interaction but within a detector such as ATLAS or CMS, or even completely pass through the detector before decaying. For example, neutral LLPs could travel a significant distance through the detector before decaying into displaced leptons, photons, or jets [1–5].

The data at the ATLAS and CMS experiments are collected using triggers, which select events in real time, reducing the event rate from the 40 MHz bunch crossing rate down to about 1 kHz that can be written to disk. Most triggers assume that the particles originate from the pp interaction vertex and are not displaced. Thus, dedicated triggers for displaced particles are necessary to maximize the chances of catching new phenomena at the LHC, in particular for its future data-taking runs.

The trigger system of the LHC experiments is usually organized in stages. In the CMS experiment, events of interest are selected using a two-level trigger system [6]. The first level (L1), composed of custom hardware processors, uses information from the subdetectors and will reduce the data rate to 750 kHz in CMS at the High-Luminosity LHC (HL-LHC) [7], which is planned to start taking data in 2027. In this phase, the upgraded L1 trigger will also feature inputs from the silicon tracker, allowing for real-time track fitting and highly efficient particle-flow reconstruction [8] of objects at the trigger level. The logic will be implemented in field-programmable gate arrays (FPGAs).

Deep neural networks (DNNs) of limited size can be deployed on FPGAs using dedicated tools such as HLS4ML [9], and can therefore now be included directly in the L1 trigger. Given the recent success of DNNs in high energy physics, in particular for complex pattern recognition problems such as b jet identification or heavy flavour jet identification, anomaly detection, as well as shower reconstruction in highly granular calorimeters and particle flow [10–23], this opens up new possibilities for triggers with simultaneously high computing and physics performance.
Due to the higher occupancy with up to 200 pp interactions per bunch crossing, in particular in the forward region, a new endcap calorimeter will be installed in CMS for the HL-LHC [24]. The interleaved High-Granularity Calorimeter (HGCal) detector layers within the absorber structure will feature a high-granularity electromagnetic section using 28 layers of silicon sensors with pad segmentation, and a hadronic section of 22 layers using the same technology in its innermost layers, and a less segmented scintillator tile section at higher radii. The high granularity of this system will allow for the measurement of particle showers in five parameters: three space dimensions, time, and energy. The HGCal will be the first imaging calorimeter in a running experiment at a high-energy collider, which generates many new opportunities, such as using it for a pattern-recognition-based trigger for displaced particles.

An example of a LLP signature that produces such a displaced, forward signature in form of jets are so-called “emerging jets” [25, 26]. Emerging jets contain electrically charged standard model (SM) particles that are consistent with having been created in the decays of new neutral LLPs produced in a parton-shower process by dark quantum chromodynamics (QCD). Dark QCD is a new strong dynamics, similar to SM QCD, but in a separate dark sector. Dark QCD is proposed in order to explain the origin of dark matter [25]. Other possible LLP signatures that could be identified in the HGCal are discussed for example in ref. [27].

This paper presents the first proof of concept of using CNN based pattern recognition to trigger on calorimeter signatures, as opposed to an energy over threshold. It has been shown in ref. [28] that two dimensional calorimeter images can be used to detect a variety of displaced signatures using convolutional neural networks (CNNs) [29]. The study presented here is made model independent by investigating the identification of electromagnetic showers that, in general, do not point to the primary pp interaction vertex. The angle between the projection direction and the particle momentum (angle to projection axis) is in the following referred to as $\alpha$. This trigger improvement will allow us to extend LLP searches with the HL-LHC to low mass and large displacements.

The study is performed using a toy calorimeter, similar to the HGCal, described in section 2 together with the generated data set. The architecture and training of the DNN is presented in section 3 and the results are presented in section 4.

2 Detector and data sample description

The endcap calorimeter is built using Geant4 [30] and is placed at $z = 3$ m distance to the interaction point. It covers a pseudorapidity ($\eta$) between 1.5 and 3.0, has a depth of 34 cm, and consists of 14 equidistant layers. Each layer comprises a 10.4 mm lead absorber and 300 $\mu$m silicon sensors. The sensors are placed in 30 rings in $\eta$, each containing 120 segments in $\phi$, leading to 50400 sensors in total, each with a size of approximately 0.05 in $\eta$ and $\phi$. This configuration corresponds to approximately 60 radiations lengths, and therefore covers electromagnetic showers only. The number of layers and the cell size approximate the granularity of the planned HGCal at first trigger level. Charged particles are subject to a magnetic field of 1 T in $z$ direction.

The signal data set is produced by generating photons at $z = 299$ cm with a flat energy spectrum between 10 and 200 GeV. The angle with respect to the projection axis is uniformly sampled between 0 and $\pi/3$. The position is randomly set to be within a radius of 20 to 60 cm with respect to the beam axis, which is well within the inner and outer radii of the calorimeter at its front face (15 and
67 cm, respectively). The rotation with respect to the projection axis is also randomly sampled, but constrained such that at least the first and the last layer of the calorimeter are hit. We consider in total 780,000 signal events for training, 8,800 for validation, and 14,400 for evaluating the performance of the proposed algorithm (testing).

To estimate the rate and the effect of multiple interactions per bunch crossing, minimum bias events are produced using Pythia8 [31]. We generate two independent samples: 15.3 M events for training and 4 M for testing and validation. The energy deposits of 200 randomly chosen minimum bias events are added to build a background event and to estimate the effect of the contribution of extraneous pp collisions to the signal. Thus, a signal event contains exactly one displaced photon superimposed with 200 minimum bias events, and a background event contains 200 minimum bias events, only. For training and validation, the ratio of signal to background events is 1:1. For testing, 70 background events are generated for each signal event. The total rate is calculated by normalising the minimum bias events by the LHC revolution frequency of 11 246 Hz and the number of bunches of 2760 [32].

3 Neural network and training

To distinguish between events with and without a displaced photon, we use a CNN architecture, developed for pattern recognition in images or other data that can be described by a regular grid structure. The detector geometry is unrolled to a 2 dimensional image in $\eta$ and $\phi$ with 14 color dimensions, one for each layer. The first 8 columns, containing pixels from $\phi = 0$ to $\phi = 0.4$ are repeated at $\phi = 2 \times \pi$, to account for particles that enter the calorimeter at $\phi \approx 0$. In consequence, the total calorimeter image contains $30 \times 128$ pixels. An example of a displaced photon signature after this preprocessing is shown in figure 1. In this projection, the displaced photon forms a line, while the other particles coming from the primary interaction form points. Moreover, the trajectory of the displaced shower through the layers is distinct from the other particles by a clearly visible color gradient.

![Figure 1](image_url)  
Figure 1. A displaced shower (curved line on the left) and a prompt shower (point on the right) in the $\eta$-$\phi$ plane. The calorimeter layers are illustrated using a rainbow color palette, with the color representing the layer number and the marker size indicating the deposited energy.

The neural network needs to be designed such that it provides a compromise between performance and resource requirements. The latter are particularly stringent if this method should be
applied and implemented in dedicated hardware in the first stages of the trigger. While we do not include dedicated studies of the resource requirements on such hardware in this note, the architecture is nevertheless chosen such that it could be adapted to such a setting e.g. through HLS4ML.

Figure 2. Illustration of the neural network architecture. Layers with learnable weights are indicated in yellow, max pooling layers in green, and input as well as reshaping layers in teal. The input is fed from the left sequentially through the network. Batch normalisation layers, applied after each CNN block, are omitted. For CNN and max pooling layers, the kernel sizes are indicated in parentheses.

An overview of the neural network architecture is shown in figure 2. For each pixel, the 14 color dimensions are reduced to 4, by sequentially applying 3 CNN layers with a kernel size of 1 x 1 pixels. The first two layers have 16 nodes, each, and the third has 4. The resulting image embeds the depth information in these 4 features, as opposed to ref. [28], where only a two dimensional representation of the calorimeter deposits is used. The image containing the encoded depth information is fed through 4 CNN blocks, each containing a CNN layer with a kernel size of 3 x 3 pixels, max pooling and batch normalisation [33]. No padding is applied in the neural network. The CNN layer in the first and second block contains 8 filters, and max pooling is applied with a kernel of 2 x 2 pixels. The last two blocks have 12 and 16 filters, and max pooling is only applied on two pixels in \( \phi \) direction. The output of the convolutional blocks is flattened and fed through one dense neural network with 32 nodes before the final classifier is calculated using a sigmoid activation. In the other layers, we employ ReLu activations [34]. The network contains 10,405 trainable parameters.

The training is performed using tensorflow [35] and keras [36] within the DeepJetCore framework [37] using the Adam [38] optimiser. The first epoch is trained with batch size of 50 and a learning rate of 0.0001. The batch size is increased to 500 for another 30 epochs of training with a learning rate of 0.0003.

4 Results

We study the efficiency as a function of rate, for different photon energies and angles \( \alpha \) with respect to the projection axis. The trigger rate is calculated as the total rate multiplied by the background efficiency for a certain classifier output. As described in section 2, the photon energy and angle are sampled from a uniform distribution. This way of presenting the results is model-independent, whereas any choice of displacement would be inherently model-dependent. As shown in figure 3, the efficiency rapidly increases with the photon energy for a fixed rate, and reaches values above 60% for a rate of 10 kHz already for energies larger than 30 GeV.
As opposed to a trigger that is based on energy thresholds only, the proposed DNN trigger depends critically on the energy and the angle $\alpha$. The trigger efficiency as a function of the energy for a trigger rate of 15 kHz is shown in figure 4 left. Particles entering the calorimeter with angles of $\alpha > 0.2$ provide a sufficiently distinct signature to be detected already at relatively low energies, while for smaller angles, the efficiency remains moderate up to high energies. The dependence of the trigger efficiency on $\alpha$, shown in figure 4 right, does not follow the same pattern. Here, the efficiency increases with $\alpha$ for all energies, but decreases slightly beyond approximately $\alpha = 0.5$. This behavior is dependent on the DNN architecture and geometry. Starting from a certain angle, the cells hit by a particle are no longer adjacent pixels, but leave a sparse image that can only be resolved by a DNN with sufficient complexity and a larger receptive field.

5 Summary

The first proof of concept of using pattern recognition with fast convolutional neural networks to trigger on displaced calorimeter signatures is presented. In particular, displaced signatures in a forward calorimeter can be identified with good efficiency and low background rate. For a target trigger rate of 15 kHz, individual particles with angles with respect to the projection axis greater
than 0.2 can be detected with good efficiency at low particle energy. This study indicates a potential increase in sensitivity to low mass, forward-moving long-lived particles.

Acknowledgments

We thank our colleagues in the CERN EP/CMG group for their support using their GPU cluster for training and the preparation of the data sets.

References

[1] ATLAS collaboration, Search for displaced vertices of oppositely charged leptons from decays of long-lived particles in pp collisions at \( \sqrt{s} = 13 \) TeV with the ATLAS detector, Phys. Lett. B 801 (2020) 135114 [arXiv:1907.10037].

[2] ATLAS collaboration, Search for heavy neutral leptons in decays of W bosons produced in 13 TeV pp collisions using prompt and displaced signatures with the ATLAS detector, JHEP 10 (2019) 265 [arXiv:1905.09787].

[3] CMS collaboration, Search for long-lived particles using delayed photons in proton-proton collisions at \( \sqrt{s} = 13 \) TeV, Phys. Rev. D 100 (2019) 112003 [arXiv:1909.06166].

[4] CMS collaboration, Search for long-lived particles using nonprompt jets and missing transverse momentum with proton-proton collisions at \( \sqrt{s} = 13 \) TeV, Phys. Lett. B 797 (2019) 134876 [arXiv:1906.06441].

[5] LHCb collaboration, Search for \( A^+ \rightarrow \mu^+\mu^- \) decays, Phys. Rev. Lett. 124 (2020) 041801 [arXiv:1910.06926].

[6] CMS collaboration, The CMS trigger system, 2017 JINST 12 P01020 [arXiv:1609.02366].

[7] G. Apollinari et al., High-Luminosity Large Hadron Collider (HL-LHC), CERN Yellow Rep. Monogr. 4 (2017) 1.

[8] CMS collaboration, Particle-flow reconstruction and global event description with the CMS detector, 2017 JINST 12 P10003 [arXiv:1706.04965].

[9] J. Duarte et al., Fast inference of deep neural networks in FPGAs for particle physics, 2018 JINST 13 P07027 [arXiv:1804.06913].

[10] D. Guest, K. Cranmer and D. Whiteson, Deep learning and its application to LHC physics, Ann. Rev. Nucl. Part. Sci. 68 (2018) 161 [arXiv:1806.11484].

[11] L. de Oliveira, B. Nachman and M. Paganini, Electromagnetic showers beyond shower shapes, Nucl. Instrum. Meth. A 951 (2020) 162879.

[12] T.Q. Nguyen et al., Topology classification with deep learning to improve real-time event selection at the LHC, Comput. Softw. Big Sci. 3 (2019) 12 [arXiv:1807.00083].

[13] D. Belayneh et al., Calorimetry with deep learning: particle simulation and reconstruction for collider physics, Eur. Phys. J. C 80 (2020) 688 [arXiv:1912.06794].

[14] P.T. Komiske, E.M. Metodiev, B. Nachman and M.D. Schwartz, Pileup Mitigation with Machine Learning (PUMML), JHEP 12 (2017) 051 [arXiv:1707.08600].

[15] CMS collaboration, CMS phase 1 heavy flavour identification performance and developments, Tech. Rep. CMS-DP-2017-013, CERN, Geneva, Switzerland (2017).

[16] CMS collaboration, New developments for jet substructure reconstruction in CMS, Tech. Rep. CMS-DP-2017-027, CERN, Geneva, Switzerland (2017).
[17] ATLAS collaboration, Identification of jets containing $b$-hadrons with recurrent neural networks at the ATLAS experiment, Tech. Rep. ATL-PHYS-PUB-2017-003, CERN, Geneva, Switzerland (2017).

[18] A. Butter et al., The machine learning landscape of top taggers, SciPost Phys. 7 (2019) 014 [arXiv:1902.09914].

[19] S.R. Qasim, J. Kieseler, Y. Iiyama and M. Pierini, Learning representations of irregular particle-detector geometry with distance-weighted graph networks, Eur. Phys. J. C 79 (2019) 608 [arXiv:1902.07987].

[20] E.A. Moreno et al., JEDI-net: a jet identification algorithm based on interaction networks, Eur. Phys. J. C 80 (2020) 58 [arXiv:1908.05318].

[21] H. Qu and L. Gouskos, ParticleNet: jet tagging via particle clouds, Phys. Rev. D 101 (2020) 056019 [arXiv:1902.08570].

[22] F.A. Di Bello et al., Towards a computer vision particle flow, arXiv:2003.08863.

[23] J. Kieseler, Object condensation: one-stage grid-free multi-object reconstruction in physics detectors, graph and image data, Eur. Phys. J. C 80 (2020) 886 [arXiv:2002.03605].

[24] CMS collaboration, The phase-2 upgrade of the CMS endcap calorimeter, Tech. Rep. CERN-LHCC-2017-023, CERN, Geneva, Switzerland (2017).

[25] Y. Bai and P. Schwaller, Scale of dark QCD, Phys. Rev. D 89 (2014) 063522 [arXiv:1306.4676].

[26] P. Schwaller, D. Stolarski and A. Weiler, Emerging jets, JHEP 05 (2015) 059 [arXiv:1502.05409].

[27] J. Liu, Z. Liu, L.-T. Wang and X.-P. Wang, Enhancing sensitivities to long-lived particles with high granularity calorimeters at the LHC, JHEP 11 (2020) 066 [arXiv:2005.10836]

[28] B. Bhattacherjee, S. Mukherjee and R. Sengupta, Study of energy deposition patterns in hadron calorimeter for prompt and displaced jets using convolutional neural network, JHEP 11 (2019) 156 [arXiv:1904.04811].

[29] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, Gradient-based learning applied to document recognition, in Intelligent signal processing, Proc. IEEE 86 (1998) 2278.

[30] GEANT4 collaboration, GEANT4 — a simulation toolkit, Nucl. Instrum. Meth. A 506 (2003) 250.

[31] T. Sjöstrand, S. Mrenna and P.Z. Skands, A brief introduction to PYTHIA 8.1, Comput. Phys. Commun. 178 (2008) 852 [arXiv:0710.3820].

[32] L. Evans and P. Bryant eds., LHC machine, 2008 JINST 3 S08001.

[33] S. Ioffe and C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6–11 July 2015, pg. 448.

[34] V. Nair and G.E. Hinton, Rectified linear units improve restricted Boltzmann machines, in Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML’10, Madison, WI, U.S.A., Omnipress, (2010), pg. 807.

[35] M. Abadi et al., TensorFlow: large-scale machine learning on heterogeneous systems, https://www.tensorflow.org/, (2015).

[36] F. Chollet et al., Keras, https://github.com/fchollet/keras, (2015).

[37] J. Kieseler et al., Deepjetcore, Zenodo, (2020).

[38] D.P. Kingma and J. Ba, Adam: a method for stochastic optimization, in 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, U.S.A., 7–9 May 2015 [arXiv:1412.6980].