Identification of new long-lived particles using deep neural networks

Matthias Komm[1,*], for the CMS Collaboration

[1]CERN, Geneva, Switzerland

Abstract. We present the development of a deep neural network for identifying generic displaced jets arising from the decays of exotic long-lived particles in data recorded by the CMS detector at the CERN LHC. Various jet features including detailed information about each clustered particle candidate as well as reconstructed secondary vertices are refined through the use of 1-dimensional convolution layers before being combined with high-level engineered features and passed through a series of fully-connected layers. The proper lifetime of the long-lived particle, $c\tau_0$, is treated as a parameter of the neural network model, which allows for hypothesis testing over several orders of magnitude ranging from $c\tau_0 = 1\,\mu\mathrm{m}$ to $10\,\mathrm{m}$. Domain adaptation by backward propagation is performed to construct domain-independent features at an intermediate layer of the network to mitiage difference between simulation and data. The training is performed by streaming ROOT trees containing $O(100M)$ jets directly into the TensorFlow queue system, which allows for a flexible selection of input features and asynchronous preprocessing. The application of the tagger is showcased in a search for long-lived gluinos as predicted by split supersymmetric models demonstrating significant gains in sensitivity over a reference analysis.

1 Introduction

Machine-learned algorithms are routinely deployed to perform event reconstruction, particle identification, event classification, and other tasks [1] when analysing data samples recorded by experiments at the CERN LHC. For example, the ATLAS [2] and CMS [3] Collaborations have developed numerous algorithms based on boosted decision trees or neural networks to identify jets originating from the hadronisation of bottom quarks with unprecedented performance [4, 5].

This note, based on Ref. [6], summarises the development and application of a novel algorithm for identifying jets originating from the decay of long-lived particles (LLPs). The algorithm is based on a deep neural network (DNN) that is inspired by the CMS DeepJet approach [7], albeit several aspects required an extension of the DeepJet architecture and training procedure. To perform supervised learning a generic definition for displaced jets is introduced. Furthermore, the DNN is parametrised as a function of the proper lifetime, $c\tau_0$, of the long-lived particle to allow hypothesis testing over several orders of magnitude ranging from $c\tau_0 = 1\,\mu\mathrm{m}$ to $10\,\mathrm{m}$. To mitigate uncertainties arising from difference between simulation and data, domain adaptation by backward propagation [8] is applied during the

---

*e-mail: Matthias.Komm@cern.ch

training. The application of the resulting DNN is demonstrated in a search for long-lived gluino production as predicted by split supersymmetric (SUSY) models [9].

2 Simulated samples and jet labelling

The DNN is trained on labelled anti-k_t jets with $p_T > 20$ Gev and $|\eta| < 2.4$ that are clustered from candidates reconstructed with the particle-flow (PF) algorithm [10] using a distance parameter of 0.4. Jets from the hadronisation of gluons, light-flavoured quarks (uds), and charm or bottom quarks are used as background classes in the DNN training. These are taken from simulated samples of multijet events produced via the strong interaction, a manifestation of quantum chromodynamics (QCD), and top quark pair production. Signal jets from LLP decays are taken from various samples of simulated split SUSY events containing pair-produced long-lived gluinos ($\tilde{g}$) that decay into a light quark/antiquark pair and a neutralino ($\tilde{\chi}_1^0$) with varying lifetime and gluino/neutralino masses. At the LLP decay vertex, the two quarks can still interact which each other resulting often in more than only two distinct jets as shown in Fig. 1 for two example gluino decays.
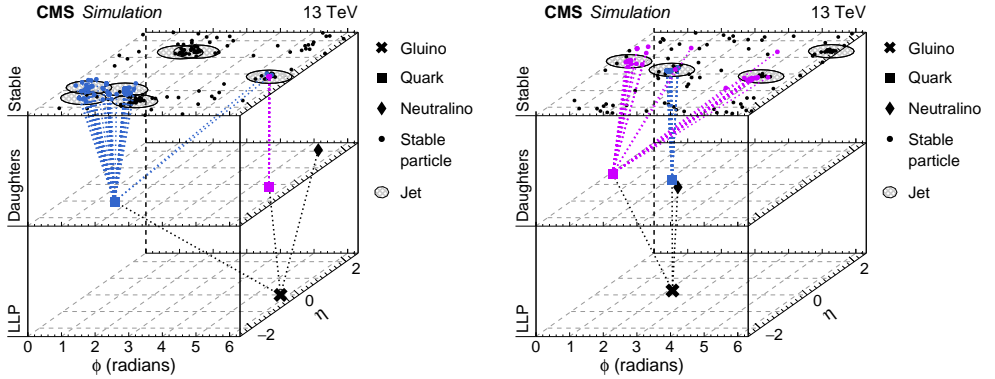

Figure 1. Two example decays of a gluino into a quark/antiquark pair and a neutralino. The positions in η and ϕ of the gluino, the quarks and neutralino, and the jets clustered from the stable particles after hadronisation are shown in the lower, middle, and top layer, respectively. The figures are taken from Ref. [6].

A novel definition of a generic LLP jet is introduced by labelling only those jets as 'LLP' for which most of the momentum is carried by clustered particles stemming from the gluino decay vertex, determined from generator-truth information.

3 Deep neural network architecture and training

An overview of the DNN architecture is given in Fig. 2. In total, approximately 600 input features are considered. The features associated with the clustered charged and neutral jet constituents and the secondary vertices are compressed through a series of one-dimensional convolutions with a kernel size of one. These are then combined with the LLP lifetime and global jet features. After a single dense layer with 200 nodes the network is split in two parts. The top part attempts to predict the jet label, whereas the bottom part predicts the domain of a jet, i.e. if a jet stems from data or simulation. Domain adaptation by backward propagation [8] is performed by reversing the gradients of the domain loss with respect to the

network weights in the preceding layers as indicated. When minimising the combined loss $L_{\text{class}} + \lambda L_{\text{domain}}$, the network is forced to only retain features that are domain-invariant. The strength of this effect is controlled through the hyperparameter λ.
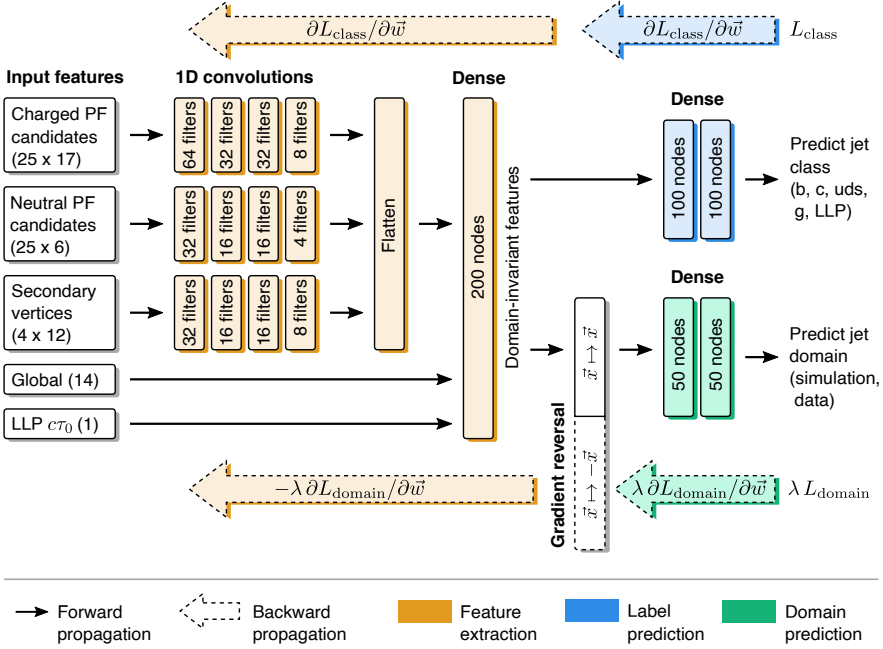
Figure 2. An overview of the DNN architecture, which comprises convolutional and dense layers; the numbers of filters and nodes are indicated. The input features are grouped by object type and $(m \times n)$ indicates the maximum number of objects (m) and the number of features per object (n). The gradients of the class (L_{class}) and domain (L_{domain}) losses with respect to the weights w, used during backward propagation, are shown. The figure is taken from Ref. [6].

Data from a control region consisting of events with at least two jets and exactly one isolated muon is used in the DNN training for domain adaptation. The resulting agreement between data and simulation is validated in an independent control region in which events are required to contain at least two jets and exactly two isolated muons instead. The distribution of the maximum LLP likelihood predicted by the network over all selected jets within an event is shown in Fig. 3. Applying domain adaptation results in a good agreement with negligible uncertainties remaining, whereas deviations of up to 50% appear in the binned counts between data and simulation when training the DNN without domain adaptation.

To train the DNN efficiently on a large sample of jets ($\mathcal{O}(100M)$), taken from data and simulation, a novel interface between ROOT TTrees and TensorFlow/Keras has been developed in the context of this work. Custom operational kernels are used to read jets from ROOT TTrees and produce corresponding tensors that are preprocessed and streamed into the Tensorflow queue system asynchronously and in parallel to the training cycle. The preprocessing encompasses a resampling of jets such that similar distributions in (p_{T}, η) is achieved for all classes. Furthermore, random $c\tau_0$ values are drawn from the distribution of LLP jets and assigned to non-LLP jets per batch on-the-fly. A demonstration of the workflow can be found in Ref. [11].
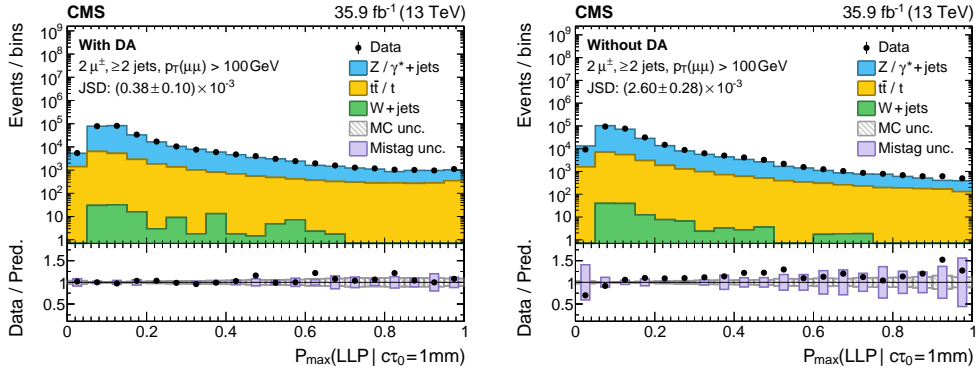
Figure 3. Distributions of the maximum probability of the LLP jet class per event in a control region consisting of two isolated muons and at least two jets with (left) and without (right) domain adaptation applied. The figures are taken from Ref. [6].
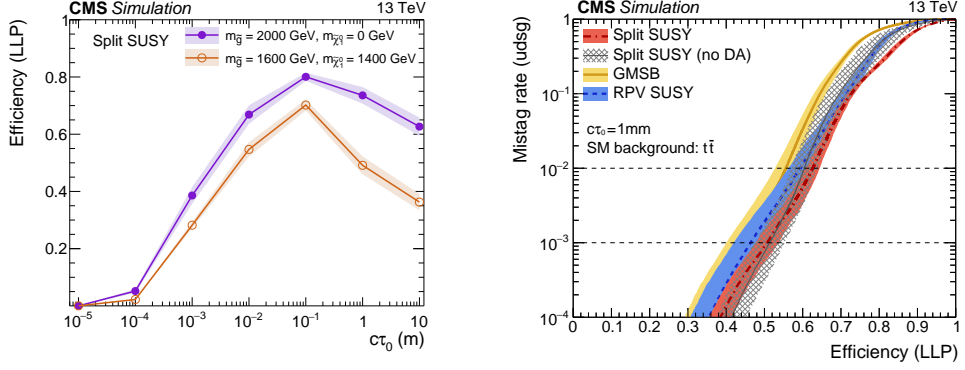
4 Performance

The performance of the resulting classifier is assessed by comparing the LLP jet selection efficiency against the rate of rejecting udsg jets as found in an inclusive $t\bar{t}$ sample. The LLP jet classification efficiency for a fixed misidentification rate of 0.01% is shown in Fig. 4 on the left as a function of the gluino lifetime for two example split SUSY scenarios. The tagger is able to identify LLP jets with efficiencies of about 40–80% over various orders of $c\tau_0$ ranging from 1 mm to 10 m. The LLP efficiency as a function of the misidentification rate for LLP jets from split SUSY or from two alternative LLP models, gauge-mediated SUSY breaking (GMSB) [12] resulting in displaced gluon jets and weak R-party violation (RPV) [13] resulting in displaced b jets, is presented in Fig. 4 on the right, where the latter two have not been used in the training. Comparable performance is seen within uncertainties for all three models, which demonstrates good generalisation of the tagger and furthermore shows that the LLP identification efficiency depends little on the jet flavour.

5 Showcase search for long-lived gluinos

In the following, the application of the tagger is demonstrated in a search for long-lived gluinos as predicted in split SUSY with $10\,\mu m < c\tau_0 < 10\,m$. Expected limits on the theoretical gluino pair production cross section are determined, while assuming the conditions of the recorded proton-proton collision data set in 2016 at $\sqrt{s} = 13\,TeV$ by the CMS experiment that corresponds to $35.9\,fb^{-1}$.

Signal events are required to contain at least three jets with $p_T > 30\,GeV$ and $|\eta| < 2.4$ and no isolated muons (electrons) with $p_T > 10(15)\,GeV$ within $|\eta| < 2.4$ that also pass loose identification criteria. Furthermore, the scalar and vectorial p_T sum of jets, $H_T = \sum_i^{jets} p_T^i$ and $H_T^{miss} = |\sum_i^{jets} \vec{p}_T^i|$, are required to be both larger than 300 GeV. Multijet events are suppressed to a negligible level by requiring $H_T^{miss}/p_T^{miss} < 1.25$ and that the minimum azimuthal separation between each jet and the H_T of all other jets is greater than 0.2 radians. The remaining events are categorised according to the number of jets, the number of jets that also pass a predefined LLP probability threshold, and H_T. The largest remaining background processes are found to be W+jets, $Z(\to \nu\bar{\nu})$+jets, single top quark, and $t\bar{t}$ production.

Figure 4. Performance of identifying LLP jets while rejecting udsg jets from an inclusive $t\bar{t}$ sample: (left) LLP jet selection efficiency for a fixed background rate of 0.01% as a function of $c\tau_0$; (right) the receiver operator characteristic (ROC) curves for three different LLP models with $c\tau_0 = 1$ mm. The figures are taken from Ref. [6].

The signal and background yields in each category are affected by several sources of systematic uncertainties. This encompasses uncertainties in the background normalisations, the jet energy scale and resolution, the energy scale of the unclustered component of p_T^{miss}, the number of pileup interactions, the renormalisation and factorisation scales, the misidentification of background jets as LLP jets, and the luminosity. The unknown LLP jet identification efficiency in data is estimated in-situ by including an additional nuisance parameter in the likelihood that modifies the scale and shape of the signal template taken from simulation depending on the number of tagged and untagged true LLP jets per category.

Expected 95% CL upper limits on the gluino mass as a function of $c\tau_0$ for two example scenarios are shown in Fig. 5. The results are compared to an inclusive search for SUSY [14] that has been performed on the same data set. The usage of the developed LLP tagger allows to excluded significantly larger gluino masses of up to 500 GeV for $c\tau_0 \gtrsim 1$ mm. The results are also found competitive with respect to a dedicated search with an optimised reconstruction technique [15].
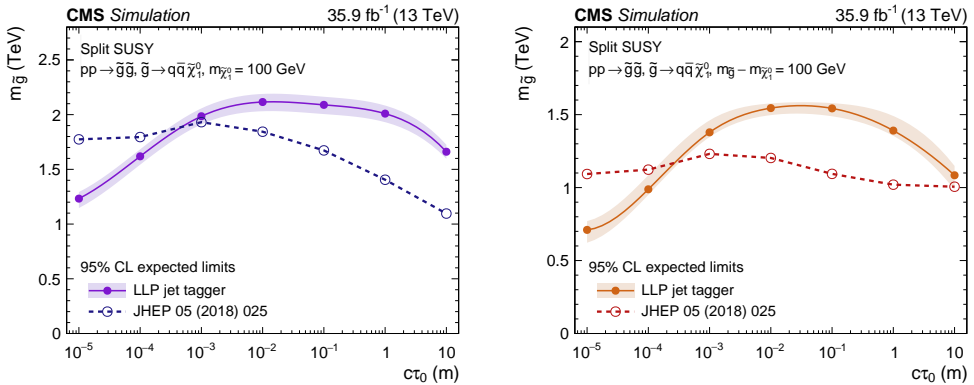
Figure 5. Expected 95% CL lower limit on the gluino mass as a function of $c\tau_0$ for two examples of split SUSY. The figures are taken from Ref. [6].

6 Summary

In this note, the development and application of a novel algorithm based on a deep neural network is presented to identify displaced jets from the decay of exotic long-lived particles as predicted by various extensions of the standard model of particle physics. To accomplish this task various new techniques are used. A generic definition of a displaced jet is introduced based on generator-truth information that allows to perform supervised learning. In the training, differences between simulation and data are mitigated by applying domain adaptation by backward propagation. The network is parametrised as a function of the lifetime of the long-lived particle, which allows for hypothesis testing over a broad range of lifetimes with just a single network.

The resulting jet tagger is able to reject 99.99% of light-flavoured jets while retaining 30–80% of LLP jets for a lifetime range of $1\,\text{mm} < c\tau_0 < 10\,\text{m}$. Lastly, the application of the tagger is demonstrated in a search for split SUSY in which competitive expected limits are determined.

References

[1] K. Albertsson, et al., J. Phys. Conf. Ser. **1085** (2018), 022008, arXiv:1807.02876.

[2] ATLAS Collaboration, JINST **3** (2008) S08003, doi:10.1088/1748-0221/3/08/S08003.

[3] CMS Collaboration, JINST **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[4] ATLAS Collaboration, Eur. Phys. J. C **79** (2019) 970, doi:10.1140/epjc/s10052-019-7450-8.

[5] CMS Collaboration, JINST **13** (2018) P05011, doi:10.1088/1748-0221/13/05/P05011.

[6] CMS Collaboration, submitted to MLST, (2019), arXiv:1912.12238.

[7] CMS Collaboration, CMS-DP-2018-058, (2018), cds.cern.ch/record/2646773.

[8] Y. Ganin, et al., arXiv e-prints (2014), arXiv:1409.7495.

[9] J. L. Hewett et al., JHEP **09** (2004) 070, doi:10.1088/1126-6708/2004/09/070.

[10] CMS Collaboration, JINST **12** (2017) P10003, doi:10.1088/1748-0221/12/10/P10003.

[11] M. Komm, et al., Zenodo (2019), doi:10.5281/zenodo.3571366.

[12] G. F. Giudice, et. al, Phys. Rept. **322** (1999) 419, doi:10.1016/S0370-1573(99)00042-3.

[13] R. Barbier, et. al, Phys. Rept. **420** (2005) 1, doi:10.1016/j.physrep.2005.08.006.

[14] CMS Collaboration, JHEP **05** (2018) 025, 10.1007/JHEP05(2018)025.

[15] ATLAS Collaboration, Phys. Rev. **D97** (2018) 5, 10.1103/PhysRevD.97.052012.