New Physics Searches with highly boosted heavy objects in final states with jets.

L. Favilla([1])([2]) on behalf of the CMS experiment

([1]) *Scuola Superiore Meridionale - Napoli, Italy*

([2]) *INFN Sezione di Napoli - Napoli, Italy*

Summary. — The search for new physics in the high Lorentz boosted regime is a crucial part of the LHC physics programme, as it could reveal the presence of new heavy resonances predicted by physics models beyond the Standard Model at high masses. In the CMS Collaboration, several boosted-jet tagging algorithms have been developed to identify hadronic jets stemming from hadronization of the decay prduct of heavy Standard Model particles such as top quarks, Higgs bosons, or vector bosons, and used in a variety of analyses. In this context, machine learning techniques will be described and their application to physics analysis will be discussed. The performance of these algorithms will be showcased in analyses performed on collision data from the LHC's Run-II at a centre-of-mass energy of 13TeV.

1. – Introduction

Several theories *Beyond Standard Model* (BSM) involve non Standard Model (SM) heavy particles with a high Lorentz boost decaying to heavy SM particles, such as W, Z, Higgs bosons, or top quarks. These objects can eventually produce collimated hadrons showers into the detector, which are usually referred to as jets. However, the topology of these induced sprays of particles can vary depending on the particle that originated them (fig. 1) and on the energy at which this phenomenon occurs.

It is important to have a focus on hadronic decays of these particles because they usually show higher Branching Ratios (BR) with respect to the respective leptonic channel, resulting in a larger statistics for a potential discovery. To increase the sensitivity to those BSM rare processes, it is of fundamental importance to have a proper procedure to correctly reconstruct and assign the correct mother particle to either each jet combination of jets. For such a purpose, jet tagging algorithms have been developed within the CMS Collaboration [1], thus enhancing discrimination capabilities.

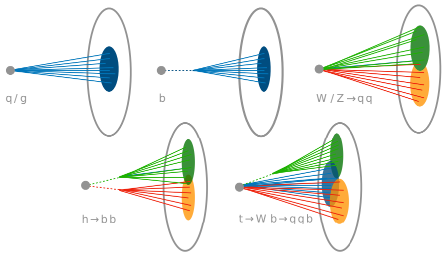


Figure 1.: Cartoon example of hadron shower topologies [2].

2. – How to reconstruct jets

Conceptually, jets can be described as collimated flows of hadrons, and they can be seen as proxies to the high-energy quarks and gluons produced in a collision. However, this simple concept is not sufficient to practically identify the jets in an event. To do that, one relies on a jet definition, *i.e.* a well-defined procedure to reconstruct the jets from the set of hadrons in the final state of the collision.

A jet definition can be seen as made of a few essential building blocks: the jet algorithm, which is the recipe itself, and a set of parameters associated with free knobs in the algorithm. A typical parameter, present in almost all jet definitions used in hadron colliders is the jet radius which essentially provides a distance above which two particles are considered as no longer part of the same jet, *i.e.* no longer considered as collinear.

For analysis purposes, two different kinds of jets are generally used, reconstructed with the anti-K_T algorithm [3] and characterized by a different value for the cone radius R (in the $\eta - \phi$ plane):

- Jets with $R = 0.4$, referred to as $AK4$ or *Narrow* jets.
- Jets with $R = 0.8$, referred to as $AK8$ or *Fat* jets.

For an extensive review on jet definitions and algorithms, refer to [4, 5].

In the following chapter, two algorithms for jet-tagging used in the CMS Collaboration will be briefly presented, as well as analyses which make use of them:

1. *DeepAK8* jet-tagging algorithm, with the analysis for the search of diboson pairs in all-jet final state;
2. *ParticleNet* jet-tagging algorithm, with the analysis for the search of non-resonant H ($\rightarrow b\bar{b}$) pair production.

3. – DeepAK8

One of the most used algorithm for the study of $AK8$ jets is the so called *DeepAK8* [6], which aim is to categorize a hadronically decaying particle using a single large jet, establishing five primary classification classes: W, Z, H, t, and other, mostly coming from single-quark hadronization as results of QCD hard radiation or multijet production.

Given the variability in the signatures produced by the same particle in different detector decay channels, these five main classifications are subdivided into finer categories based on particle decay modes (e.g., Z → bb, Z → cc, and Z → qq). The *DeepAK8* algorithm employs this comprehensive approach to achieve its classification goals. The main architecture involved is the *Convolutional Neural Network* and exploits both particle-level and event-level data.

One of the analyses exploiting such an algorithm is briefly described in the following paragraph.

3˙1. *Diboson pairs in all-jet final state.* – In this paragraph, a search for new heavy resonances, in the all-jets final state, decaying to a VV or VH boson pair with masses between 1.3 and 6TeV and produced via *Drell–Yan* (DY, fig. 2), *gluon fusion* (ggF), or *vector boson fusion* (VBF) is targeted [7].

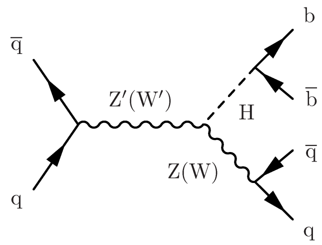


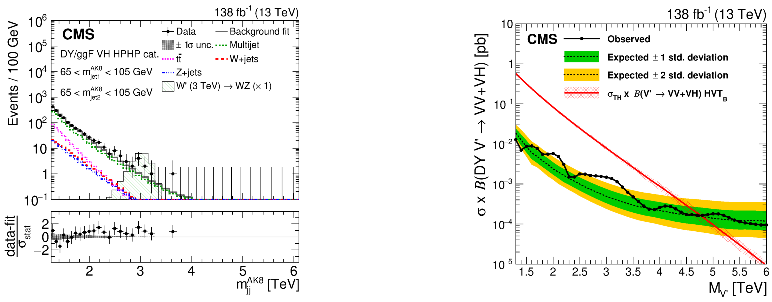
Figure 2.: Feynman diagram of the signal process of new heavy resonances decaying to VH boson pair [7].

Because of the large Lorentz boost of the H, W, and Z bosons from the resonance decay, bosons decay is typically reconstructed into a single large-radius jet, thus producing a final state characterized by two large-radius jets (distance parameter $R = 0.8$) in the case of DY and ggF production, with two additional small-radius ($R = 0.4$) jets in the case of VBF production. Events are selected by requiring at least two AK8 jets with $p_T > 200GeV$, $|\eta| < 2.5$ and a mass $55 < m_j^{AK8} < 215GeV$. They are also required to have $|\Delta\eta^{AK8}| < 1.3$ to reduce the QCD background. Moreover, 2 jet taggers have been used, both based on the previously described *DeepAK8* algorithm:

- $q\bar{q}$ tagger, to discriminate between $W, Z \rightarrow q\bar{q}$ *vs.* q/g induced jets;
- $b\bar{b}$ tagger, to discriminate between $H, Z \rightarrow b\bar{b}$ *vs.* q/g induced jets.

The main SM background processes in this search are QCDmultijet, $\mathrm{t\bar{t}}$, W + jets and Z + jets, the first being the largest one. A maximum likelihood fit of signal and background templates to the data is performed in the $(m_{jj}^{AK8}, m_{j1}^{AK8}, m_{j2}^{AK8})$ space, which are the invariant masses of respectively the system of the two jets, and the invariant masses of the two jets taken separately. This procedure comes from the intrinsic nature of the considered physics processes: the signal is resonant in all the three masses considered, while all the backgrounds are non-resonant in m_{jj}^{AK8} (fig. 3a) and are not (for the QCD) or only partly (for the other backgrounds) resonant in $m_{j_1\, or\, j_2}^{AK8}$.

Observed and expected 95% CL upper limits on the product of the production cross and the branching fraction using a pp collision data set collected by CMS experiment at a centre-of-mass energy of 13TeV in 2016-2018 corresponding to an integrated luminosity of $138fb^{-1}$ are shown in fig. 3b.



(a) (b)

Figure 3.: (a) Projections of data and background post-fit distributions onto the m_{jj}^{AK8} dimension in regions enriched in signal from DY/ggF for the VH channel [7] and (b) Observed and expected 95% CL upper limits on the product of the production cross and the branching fraction using a pp collision data set collected by CMS experiment at a centre-of-mass energy of 13TeV in 2016-2018 corresponding to an integrated luminosity of $138fb^{-1}$ [7].

Excesses at $M_{V'} \sim 2.1 - 2.9$TeV have not been found in the semileptonic channels.

4. – Particle-Net: jets as "particle cloud"

A further example of algorithm used for tagging AK8 jets is *Particle Net* [8], that exploits the capability of a *Graph Neural Network* (GNN), through which a jet is represented as a *particle cloud*, having, on the one hand, all the advantages and flexibility of a particle-based representation, and on the other hand, the algorithmic strength of the point-cloud representation of 3D shapes used in computer-vision applications.

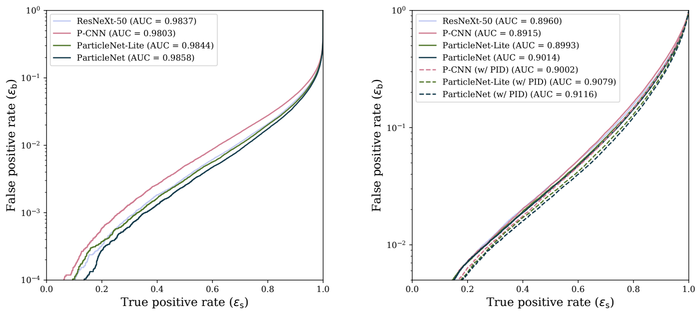
The algorithm exploits particle-level data and is used for:

- top-tagging (as well as W/Z/H tagging), *i.e.*, to identify jets from hadronically decaying top quarks, and for quark-gluon tagging, thus discriminating between jets initiated by quarks and gluons. For the top-tagging algorithm (fig. 4a), only jets with $R = 0.8$ and reconstructed with the anti-k_T algorithm are considered; for each jet, up to 100 constituents with high p_T values are taken into account.

- quark-gluon tagging (fig. 4b), performed on anti-k_T jets with $R = 0.4$. Moreover, two different sets of variables for each particle are used: in the first one, only variables related to four-momentum are taken into account, while in the second one, there is also particle identification information (PID). PID information leads to better performance in jet tagging.

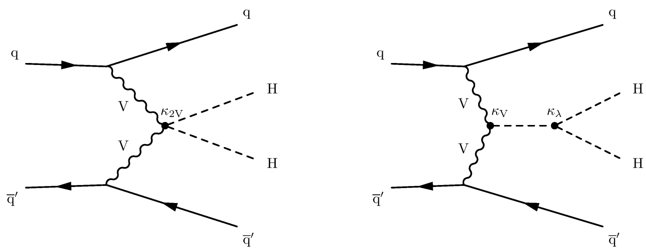
4˙1. *Non-resonant* H *(→* b$\bar{\text{b}}$*) pair production.* – In this paragraph, a search for non-resonant highly boosted HH production via ggF and VBF (fig. 5) modes with H → b$\bar{\text{b}}$ is described [9].

This analysis is among the first to use the ParticleNet GNN algorithm described above to discriminate between H → b$\bar{\text{b}}$ and QCD-induced jets, assigning a score D_{bb} to



(a) (b)

Figure 4.: Performance comparison in terms of ROC curves on (a) top-tagger and (b) quark–gluon tagging [8].



(a) (b)

Figure 5.: HH production via VBF [9].

each AK8 jet. Events containing at least two large-radius jets with $p_T > 300GeV$ and $|\eta| < 2.4$, and no isolated electrons or muons, are selected and grouped into either the ggF or the VBF categories. For the VBF categories (the only one actually presented in this paragraph), the two large-radius jets are required to have a reconstructed mass $50GeV < m_j^{AK8} < 200GeV$. As the VBF process is characterized by the presence of two forward jets with a large dijet invariant mass and a distance in η, we also require two additional small-radius jets, referred to as VBF jet candidates, with $p_T > 25GeV$ and $|\eta| < 4.7$. For ggF events, a tailored BDT has been trained to discriminate between HH and QCD/$\mathrm{t\bar{t}}$ events. A binned maximum likelihood fit in the space $(m_{HH}, D_{bb}, m_{j_{1,2}}^{AK8}, score_{BDT})$ is performed using the sum of the signal and background contributions.

The data are found to agree with the background-only hypothesis, and an observed (expected) upper limit at 95% confidence level is set to 9.9 (5.1) relative to the standard model cross section, using pp collision data at 13TeV corresponding to $138fb^{-1}$ collected by the CMS experiment. Upper limits on the production cross section are set as a function of the coupling modifier parameters k_λ and k_{2V} (fig. 6), which parametrize the strengths of the Higgs boson self-coupling, and the quartic VVHH couplings, respectively. This analysis was able to exclude $k_{2V} = 0$ for the first time with a significance of 6.3 standard deviations.

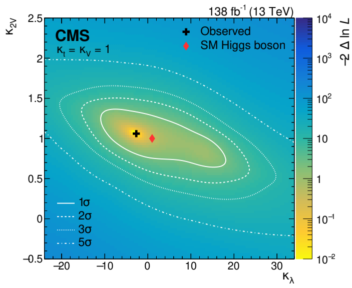


Figure 6.: Two-parameter profile likelihood test statistic ($-2\Delta\ln\mathcal{L}$) scan in data as a function of κ_λ and κ_{2V} [9].

5. – Conclusion

The quest for higher sensitivity to rare processes has demanded a huge effort to keep on finding new ways and algorithms able to correctly reconstruct and identify heavy and highly-boosted SM particles decaying into quarks, which give rise to energy sprays in the detector, usually referred to as *hadronic jets*. Hence, a brief summary of two of the most used algorithms for jet tagging inside the CMS collaboration has been presented, as well as the conceptual view of two associated analysis performed using pp-collision data coming from Run2 phase of LHC at a centre-of-mass energy of 13TeV.

REFERENCES

[1] CMS Collaboration, "The CMS Experiment at the CERN LHC", JINST **3** (2008), S08004 doi:10.1088/1748-0221/3/08/S08004

[2] E. A. Moreno, T. Q. Nguyen, J. R. Vlimant, O. Cerri, H. B. Newman, A. Periwal, M. Spiropulu, J. M. Duarte and M. Pierini, "Interaction networks for the identification of boosted $H \to b\bar{b}$ decays", Phys. Rev. D **102** (2020) no.1, 012010 doi:10.1103/PhysRevD.102.012010 [arXiv:1909.12285 [hep-ex]].

[3] M. Cacciari, G. P. Salam and G. Soyez, "The anti-k_t jet clustering algorithm", JHEP **04** (2008), 063 doi:10.1088/1126-6708/2008/04/063 [arXiv:0802.1189 [hep-ph]].

[4] G. P. Salam, "Towards Jetography", Eur. Phys. J. C **67** (2010), 637-686 doi:10.1140/epjc/s10052-010-1314-6 [arXiv:0906.1833 [hep-ph]].

[5] S. Marzani, G. Soyez and M. Spannowsky, "Looking inside jets: an introduction to jet substructure and boosted-object phenomenology", Lect. Notes Phys. **958** (2019), pp. Springer, 2019, doi:10.1007/978-3-030-15709-8 [arXiv:1901.10342 [hep-ph]].

[6] CMS Collaboration, "Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques", JINST **15** (2020) no.06, P06005 doi:10.1088/1748-0221/15/06/P06005 [arXiv:2004.08262 [hep-ex]].

[7] CMS Collaboration, "Search for new heavy resonances decaying to WW, WZ, ZZ, WH, or ZH boson pairs in the all-jets final state in proton-proton collisions at s=13TeV", Phys. Lett. B **844** (2023), 137813 doi:10.1016/j.physletb.2023.137813 [arXiv:2210.00043 [hep-ex]].

[8] H. Qu and L. Gouskos, "ParticleNet: Jet Tagging via Particle Clouds", Phys. Rev. D **101** (2020) no.5, 056019 doi:10.1103/PhysRevD.101.056019 [arXiv:1902.08570 [hep-ph]].

[9] CMS Collaboration, "Search for Nonresonant Pair Production of Highly Energetic Higgs Bosons Decaying to Bottom Quarks", Phys. Rev. Lett. **131** (2023) no.4, 041803 doi:10.1103/PhysRevLett.131.041803 [arXiv:2205.06667 [hep-ex]].