EuCAIFCon 2024

EUROPEAN AI FOR FUNDAMENTAL PHYSICS CONFERENCE

Amsterdam, 30 April - 3 May

Anomaly detection search for BSM physics in ATLAS experiment at LHC

Francesco Cirotto[1,3], Francesco Conventi[2,3], Elvira Rossi[1,3]

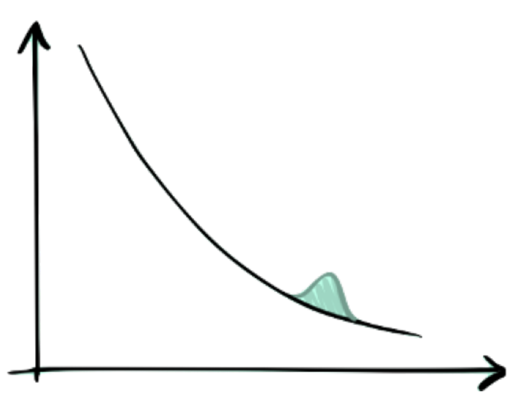
[1] Università degli Studi di Napoli Federico II
[2] Università degli Studi di Napoli Parthenope
[3] INFN Napoli

The lack of evidence for new interactions and particles since the Higgs boson's discovery has motivated the execution of generic searches to complement the existing rigorous, model-dependent analysis program. Unsupervised machine learning can offer a new style of analyses which is completely agnostic to types of new-physics models and to any expectations of scientists. ATLAS collaboration is pursuing this new approach, with a first published result in a search for a heavy resonance Y decaying into a Standard Model Higgs boson H and a new particle X in a fully hadronic final state. Moreover Graph Anomaly Detection (GAD) exploits innovative machine learning algorithms denoted as Graph Neural Networks, which have proved to be more efficient than standard techniques when applied to heterogeneous data naturally structured as graphs.

Finding anomalies in HEP

Anomaly Detection (AD) uses unsupervised Machine Learning architectures to identify outliers in a set of "standard" objects. In High Energy Physics, this means the identification of features of detector data inconsistent with the expected background.

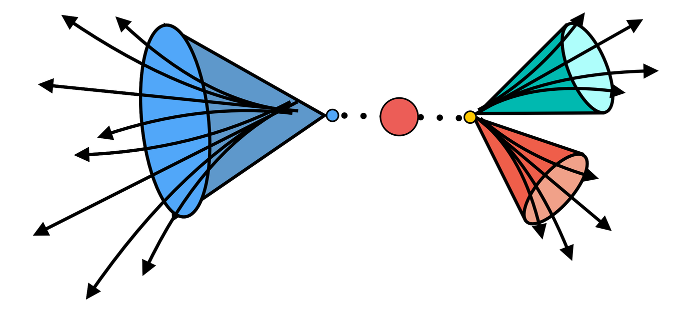
Building a tool to perform model-independent classification of collision events involves training on data events, and therefore requires the ability to cope with a lack of labels indicating whether inputs are signal or background.



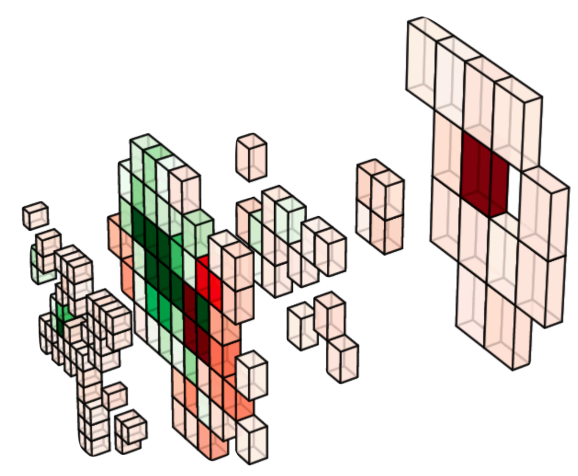
Using Jets for Anomaly Detection

Many Beyond Standard Model theories predict new massive resonances which can decay hadronically, leading to final states involving jets.

For massive particles, their decay products become collimated, or 'boosted', in the direction of the progenitor particle. For massive particles that are sufficiently boosted, it is advantageous to reconstruct their hadronic decay products as a single large-radius (large-R) jet.



Large-R jet (left) and small-R jet (right) reconstruction.



Jet information can be used as input features for neural network architectures. A significant improvement in performances can be achieved by employing a set of features with basic information (**low-level**) [1] such as information coming directly from the detectors. **Jet constituents** represent challenging input features to achieve this goal.

In the architectures described below a pre-processing method is applied which boosts each jet to the same reference mass, energy, and orientation in η − ϕ space, such that all input jets differ only by their substructure [2].

Anomaly Detection in ATLAS in fully hadronic final states

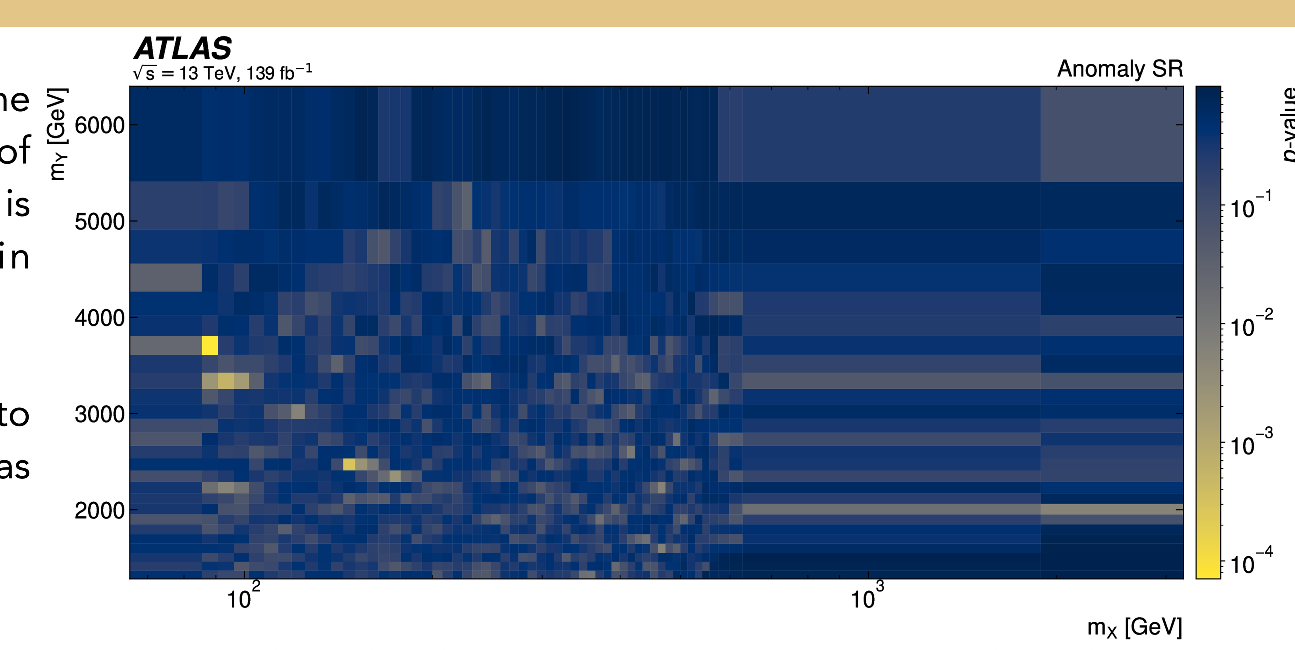
The first application with AD technique in ATLAS is a search for a heavy resonance Y decaying into a Standard Model Higgs boson H and a new particle X in a fully hadronic final state ($Y \to XH \to q\bar{q}b\bar{b}$) using the full Run-2 dataset collected by ATLAS from 2015 to 2018, corresponding to an integrated luminosity of 139 fb^{-1} [3].

The two selected leading large-R jets are identified as H or X. **Higgs tagging** is performed with a neural network approach to separate bosons decaying into b-quarks from top-quark and QCD jet. According to **X tagging** three orthogonal Signal Regions (SR) are defined: **Merged, Resolved** and **Anomaly**.

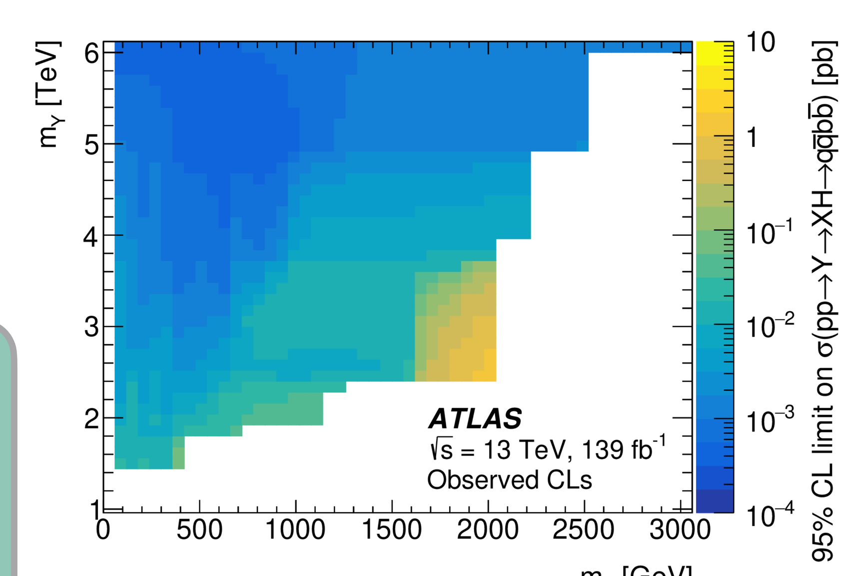
	Signal Regions		
	Merged	Resolved	Anomaly
m_H [GeV]	(75, 145)		
D_{Hbb}	> 2.44		
D_2^{trk}	< 1.2	> 1.2	-
$\lvert\Delta y_{j_1,j_2}\rvert$	-	< 2.5	-
p_T^{bal}	-	< 0.8	-
Anomaly Score	-	-	> 0.5

The observable fitted in the analysis is the m_Y distribution of the data in the SR. The fit is repeated several times in overlapping bins of m_X.

BUMP HUNTER algorithm used to find excesses, with a p-value as goodness-of-fit metric.



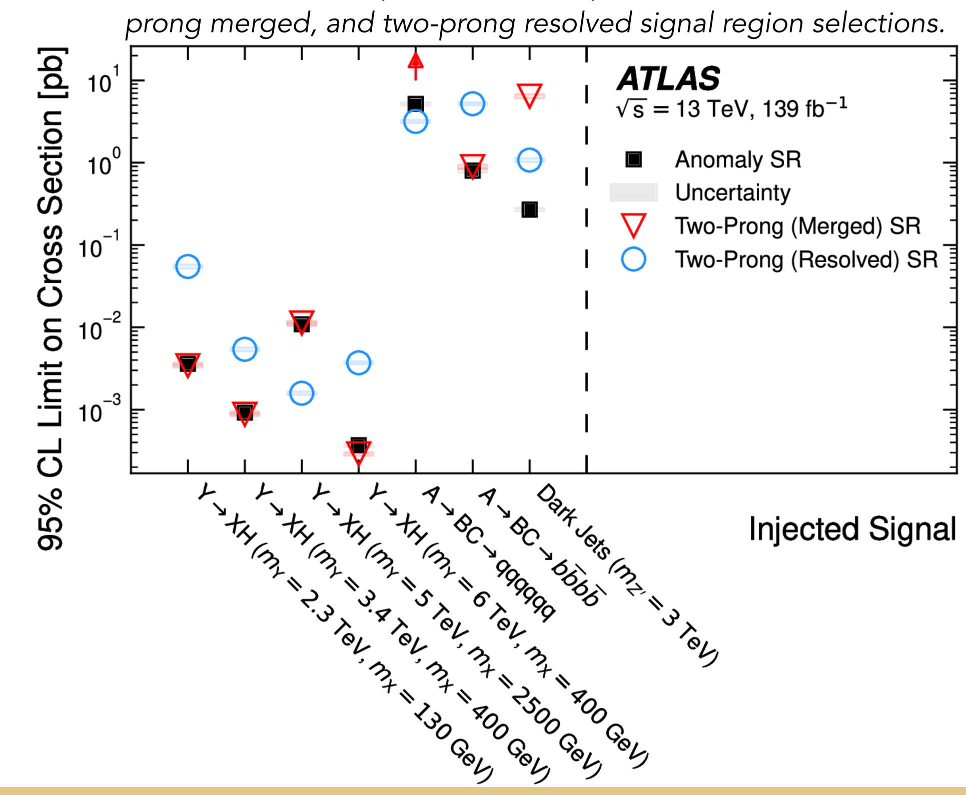
Observed p-values across all m_Y and m_X bins in the anomaly signal region.



Observed 95% CL upper limit on the cross section for merged and resolved SRs combination.

Given the absence of signal 95% CL upper limit on the cross section of the $Y \to XH$ process have been derived combining results of merged and resolved Signal Regions

The 95% CL upper limit on the cross section for seven benchmark signal processes, comparing the anomaly, two-prong merged, and two-prong resolved signal region selections.

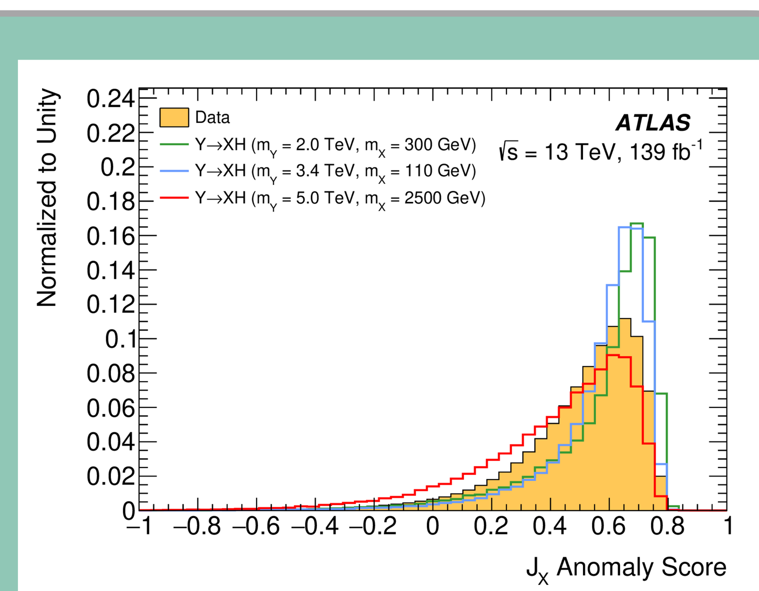


A comparison between anomaly detection and standard analysis can be obtained deriving the 95% CL upper limit on the cross section of several benchmark signals by injecting signals into the data until a significance of 2σ is found. The Anomaly Score shows same results as Merged SR for signals where the X particle is highly boosted.

Building the architecture

Anomaly Score is obtained with a **Variational Recurrent Neural Network** (VRNN), a sequence-modeling architecture which replaces the standard encoder-decoder architecture of a Recurrent Neural Network with a Variational Autoencoder (VAE). This allows the VRNN to perform both sequence modeling in addition to variational inference.

Inputs to the VRNN consist of sequences of **up to 20 jet four-vector constituent** components pT, η, and ϕ, where constituents are assumed to be massless.



Distribution of X candidate anomaly score in data for several signal hypotheses.

The main QCD background is estimated with a fully data-driven method. A DNN has been implemented to estimate background in Signal Region, by reweighting events with a proper function obtained from Control Regions (**direct importance estimation**).

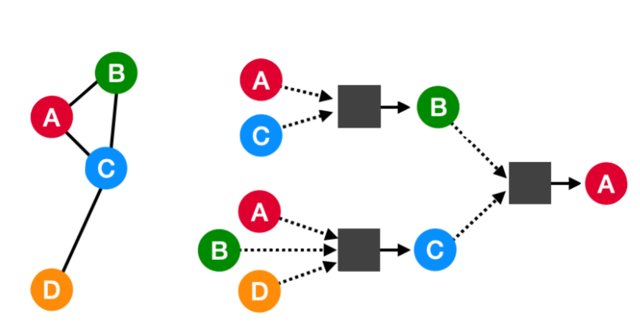
Towards Jets representation as Graphs

Graph Neural Networks (GNN) have proved to be innovative machine learning techniques useful to perform anomaly detection. In the search for resonances in fully hadronic final states jets **complex substructure** can hide New Physics. Due to its sparse structure they result suitable for a **graph representation.** These graphs can be used to represent large-R jets by expressing potentially heterogeneous detector information.

A **graph** is a set of points (**nodes**) that can be connected by **edges**

__what is a node?__
✻each constituent is a node
✻node features: p_T fraction, η, ϕ

__when are nodes connected?__
✻Constituent distance-based edge feature



Message passing: vector messages are exchanged between nodes and updated using networks. At each iteration, every node aggregates information from its local neighborhood. At the end each node embedding contains more and more information from further reaches of the graph.

Preliminary results

Tested architecture on **R&D LHC Olympics 2020 dataset [4]**: generated 1M multi-jet background and 100k signal ($Z' \to XY \to q\bar{q}q\bar{q}$) events, where X and Y are reconstructed as large-R jets.

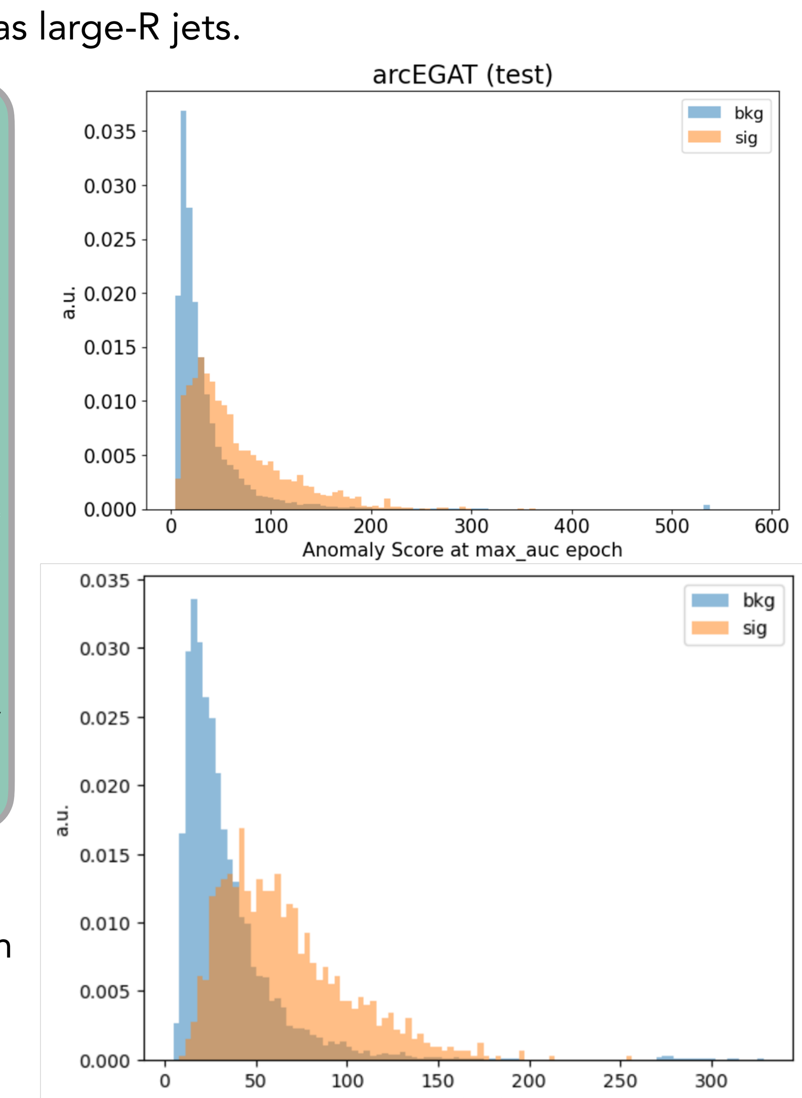
Building the architecture

EGAT (Edge Graph Attention Network): edge focused graph neural network with implemented attention mechanism. Edge features are updated by message passing along with node features at each layer.

Unsupervised minimization of a DeepSVDD objective, results interpreted with discriminating anomaly score → loss value calculated per jet after training.

Event level score obtained by combining the 2 jets anomaly scores.

Very promising results with LHC Olympics dataset, AUC ~0.82 in event-level anomaly score approach.
The architecture will be tested on official ATLAS datasets.



Jet-level (top) and event-level (bottom) anomaly score distribution for signal (blue) and background (orange).

References

1. Baldi, P., Sadowski, P. & Whiteson, D. Searching for exotic particles in high-energy physics with deep learning. Nat Commun 5, 4308
2. Tuhin S. Roy, Aravind H. Vijay, A robust anomaly finder based on autoencoders, arXiv:1903.02032
3. ATLAS Collaboration, Anomaly detection search for new resonances decaying into a Higgs boson and a generic new particle X in hadronic final states using √s = 13 TeV pp collisions with the ATLAS detector, Phys. Rev. D 108, 052009
4. The LHC Olympics 2020 a community challenge for anomaly detection in high energy physics, Rep. Prog. Phys. 84 (2021) 12420