Boosted Jet Identification with Machine Learning

R. M. Cobe, S. F. Novaes, J. C. Ruiz Vargas, T. R. Tomei
UNESP - Universidade Estadual Paulista, Campus de São Paulo

Overview

Hadronic jets are the characteristic signature of quarks in high energy colliders. The identification of jets originating from boson decays and those coming from QCD processes is quite relevant to separate interesting signal events from unwanted backgrounds. Common methods for jet identification involve jet grooming techniques such as trimming, pruning, and soft drop, along with jet substructure variables such as N-subjettiness. In this work, we handle the problem of jet classification in experimental high energy physics using Machine Learning algorithms.

Jet Image Classification

Definitions

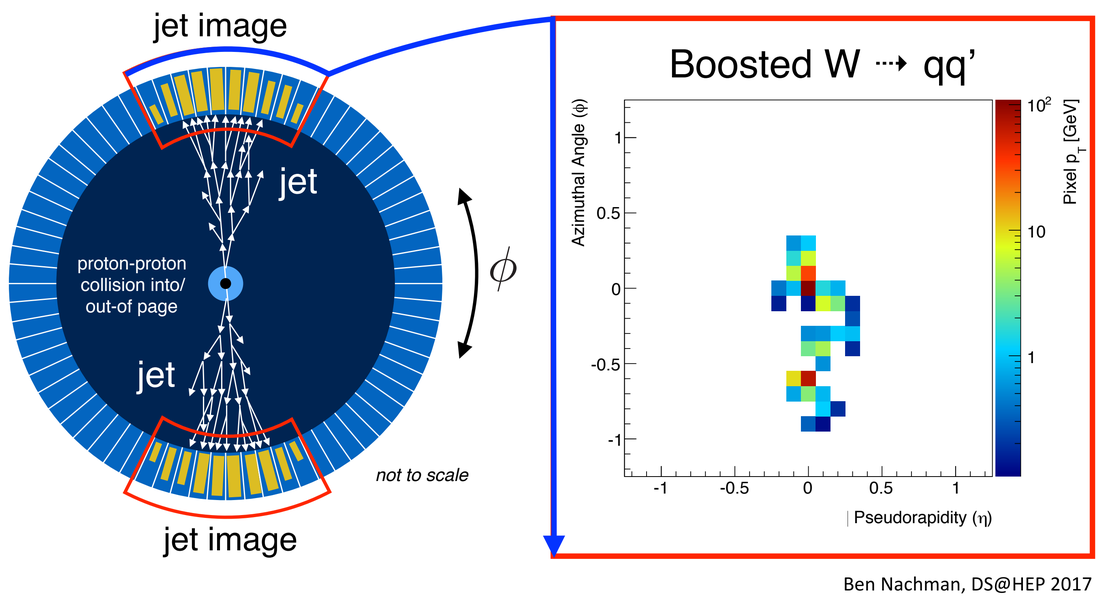
- **Signal**. High energy jets coming from W/Z processes.
- **Background**. Similar jets coming from QCD processes.

Objective

- Classify jet images into signal and background.

Data simulation and analysis strategies

- **Pythia 8** event generator.
- **FastJet** library for jet clusterization.
- **Machine Learning**. Logistic regression, artificial neural nets.
- **Jet substructure**. N-subjettiness made out of jet constituents.



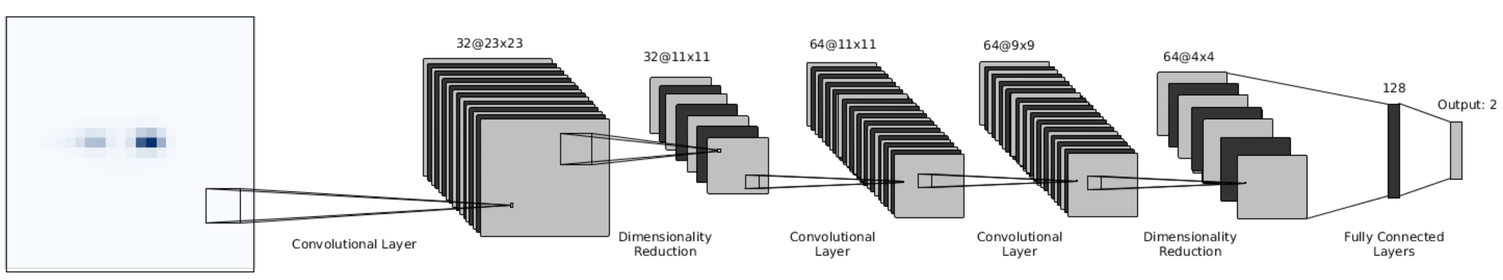
Description of the Algorithms

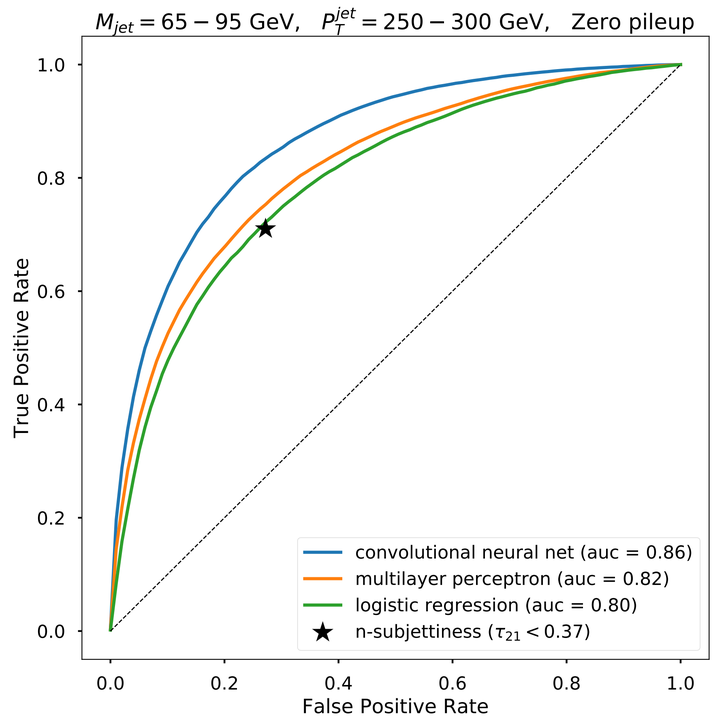
High Energy Physics tools

- **N-subjettiness**. $\tau_{21} < 0.37$ provides optimal discrimination.

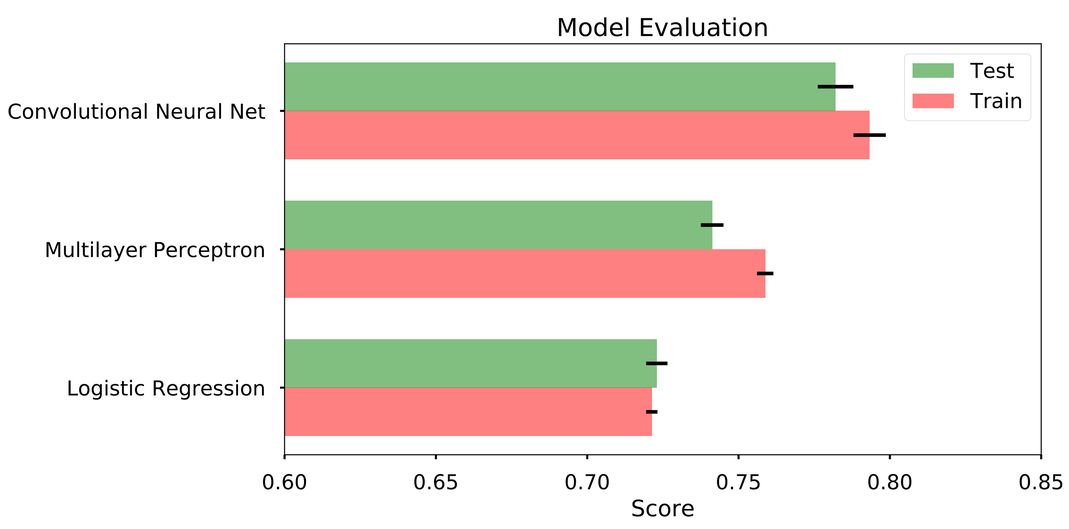
Machine Learning algorithms

- **Logistic regression**. Calculate probability of a jet image to be signal.
- **Multilayer perceptron**. Artificial neural network; 1 hidden layer, 5 neurons.
- **Convolutional Neural Net**. Architecture described below:





Model Evaluation



Cross-validation for determining the score of correct classifications. Training (Test) sample corresponds to 70% (30%) of the full dataset.

Conclusions and Outlook

Our results confirm the good performance of convolutional neural networks to handle the problem of classification of jet images, demonstrated by the area under the ROC curve (auc = 0.86).
Deep learning applications in the field of high energy physics must improve data analysis techniques in the coming years, for the benefit of experiments like the high luminosity LHC.

Acknowledgements

Supported by the São Paulo Research Foundation (FAPESP) under Grant Nº 2013/01907-0, in cooperation agreement with Intel®.