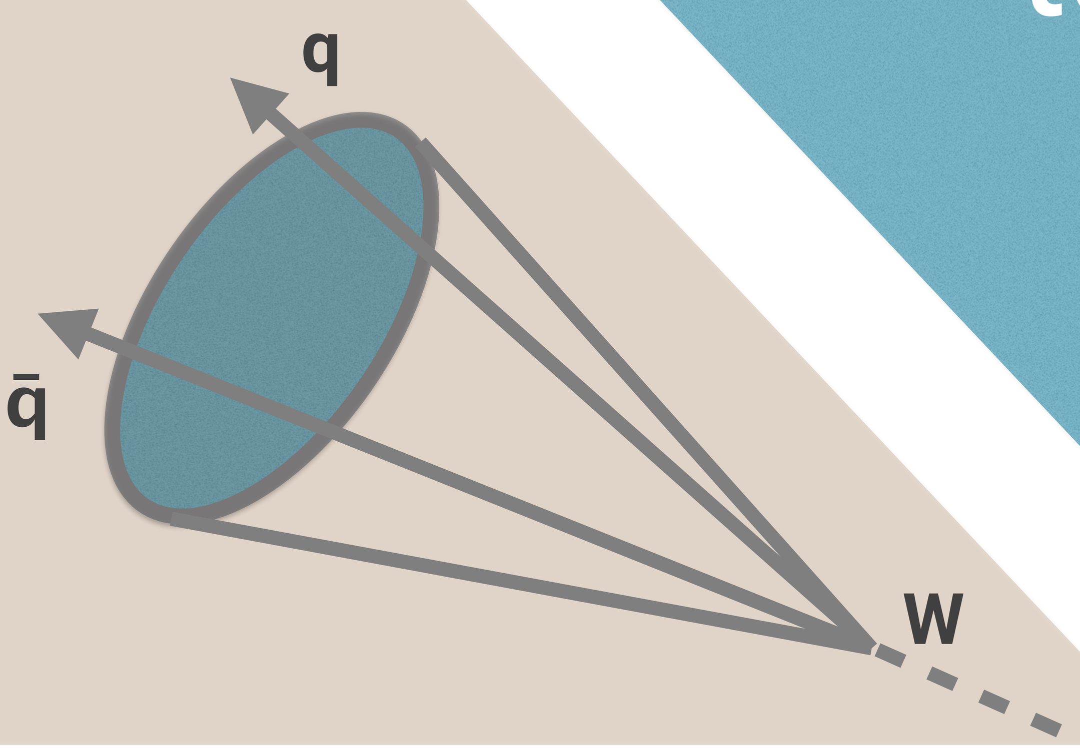
LoLa: Lorentz Invariance Based Deep Neural Network for heavy-resonance tagging

BOOSTED TOPOLOGIES

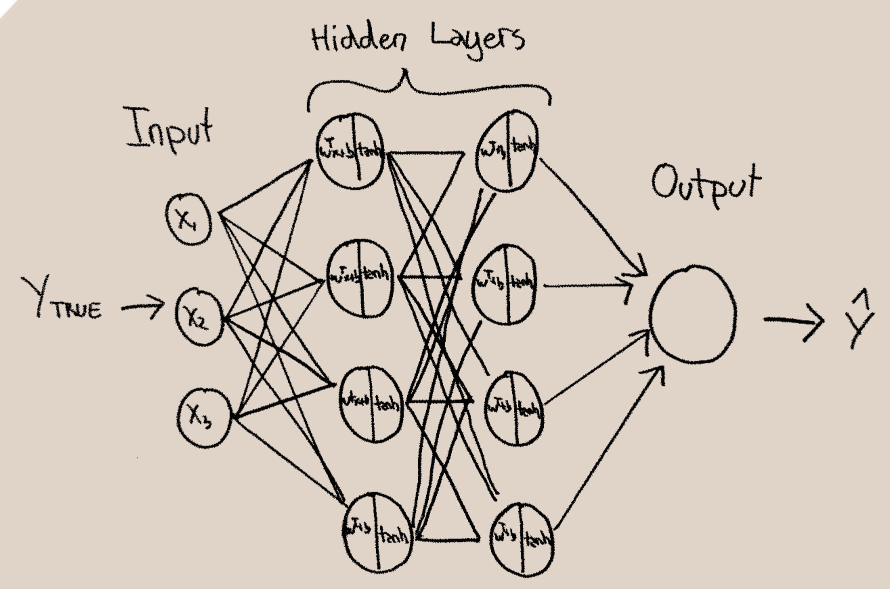
Vector bosons from the decay of TeV resonances usually highly energetic

- Above W p_T = 200 GeV, decay products merge into single massive jet
- “Tag” by identifying jet substructure and mass in order to discriminate against QCD background



DEEP LEARNING

Letting computer estimate true value of a parameter, y, by a series of regression computations starting from a set of inputs, x.

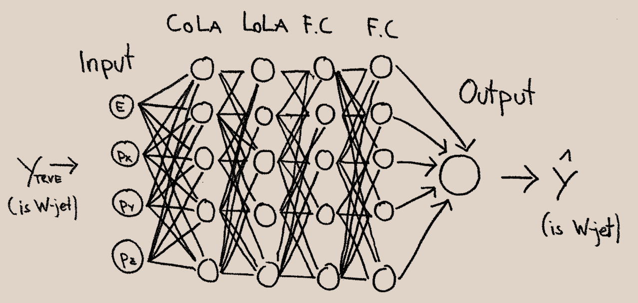


Made extremely easy thanks to Open Source packages like Keras and Tensorflow (https://keras.io/, https://www.tensorflow.org/)

WHO IS LOLA?

Read more about LoLa at arXiv: 1707.08966

LoLa is a four Layer Deep Neural Network attempting to identify hadronically decaying Ws starting from jet constituents (initially designed to tag top quarks). It implements two novel layers; the Combination Layer (CoLa) and Lorentz Layer (LoLa) as well as two fully connected layers.



- CoLa: make linear combinations of particles inside jet
- LoLa: teaches network distances in Minkowski space

THE HEART OF LOLA

First, linear combinations of the input are made in CoLa:

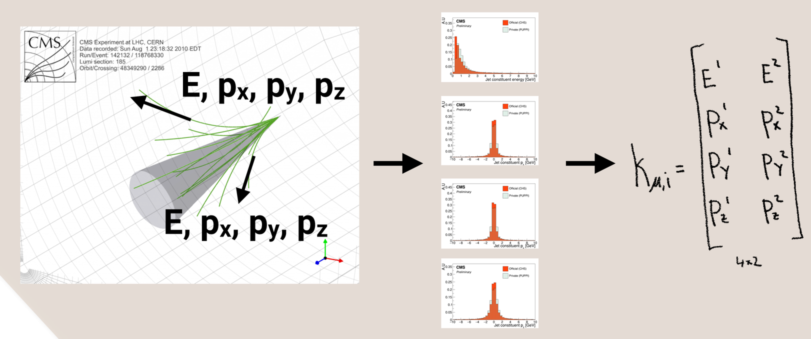
$$k_{\mu,i} = \begin{bmatrix} E^1 & E^2 \\ p_x^1 & p_x^2 \\ p_y^1 & p_y^2 \\ p_z^1 & p_z^2 \end{bmatrix}_{4\times 2} \rightarrow \tilde{k}_{\mu,i} = \underbrace{\begin{bmatrix} E^1 & E^2 \\ p_x^1 & p_x^2 \\ p_y^1 & p_y^2 \\ p_z^1 & p_z^2 \end{bmatrix}}_{k_{\mu,i}} \underbrace{\begin{bmatrix} 1 & 1 & 0 & w_{4,1} \\ 1 & 0 & 1 & w_{4,2} \end{bmatrix}_{2\times 4}}_{C_{ij}\ \text{CoLa}} = \begin{array}{cccc} \text{Sum} & \text{Constituent} & & \text{Linear combination} \\ \end{array} \begin{bmatrix} E^1+E^2 & E^1 & E^2 & w_{4,1}E^1+w_{4,2}E^2 \\ p_x^1+p_x^2 & p_x^1 & p_x^2 & w_{4,1}p_x^1+w_{4,2}p_x^2 \\ p_y^1+p_y^2 & p_y^1 & p_y^2 & w_{4,1}p_y^1+w_{4,2}p_y^2 \\ p_z^1+p_z^2 & p_z^1 & p_z^2 & w_{4,1}p_z^1+w_{4,2}p_z^2 \end{bmatrix}_{4\times 4}$$

Ensures network can 1) learn the sum of momenta in the jet 2) learn the momenta of each constituent 3) make linear combinations of particles during training (e.g for W→qq, $w_{4,1} = w_{4,2}$ because $(k_{\mu,1}+k_{\mu,2})^2 = M_W^2$. Allows network to “groom” away unimportant particles).

INPUT

In CMS, particles are reconstructed using information from all subdetectors. These are then used as inputs to the jet clustering algorithm.

- LoLa takes as input the 4-vector of the 20 highest-p_T particle candidates in a jet, a 4x20 matrix. E. g. for 2 jet constituents:



Each column of the CoLa matrix is then passed to LoLa where it is mapped onto the following quantities:

LOLA

$$\rightarrow \underbrace{\begin{bmatrix} m^2(\tilde{k}_j) \\ p_T(\tilde{k}_j) \\ w^E_{jm} E(\tilde{k}_m) \\ w^{d}_{jm} \min(d_{jm}) \\ 2w^{d}_{jm} \min(d_{jm}) \\ 3w^{d}_{jm} \mathrm{sum}(d_{jm}) \\ 4w^{d}_{jm} \mathrm{sum}(d_{jm}) \end{bmatrix}}_{\text{LoLa}}_{7\times 4} = \begin{bmatrix} g_{\mu\nu} \tilde{p}^{\mu}_{z p_1} \tilde{p}^{\nu}_{z p_1} & \cdots & \cdots & \cdots \\ \sqrt{\sum p_x^2 + \sum p_y^2} & \cdots & \cdots & \cdots \\ w^E_{jm} \sum E & \cdots & \cdots & \cdots \\ {}_1w^d \min & & & \\ {}_2w^d \min & (\tilde{k}_j-\tilde{k}_m)_\mu g^{\mu\nu} (\tilde{k}_j-\tilde{k}_m)_\nu & \cdots & \\ {}_3w^d \mathrm{sum} & & & \\ {}_4w^d \mathrm{sum} & & & \end{bmatrix}$$

Explicit use Minkowski Metric!

Invariant mass and p_T for each CoLa column. Energy of all particles. Distance between all particles with 4 trainable weights.

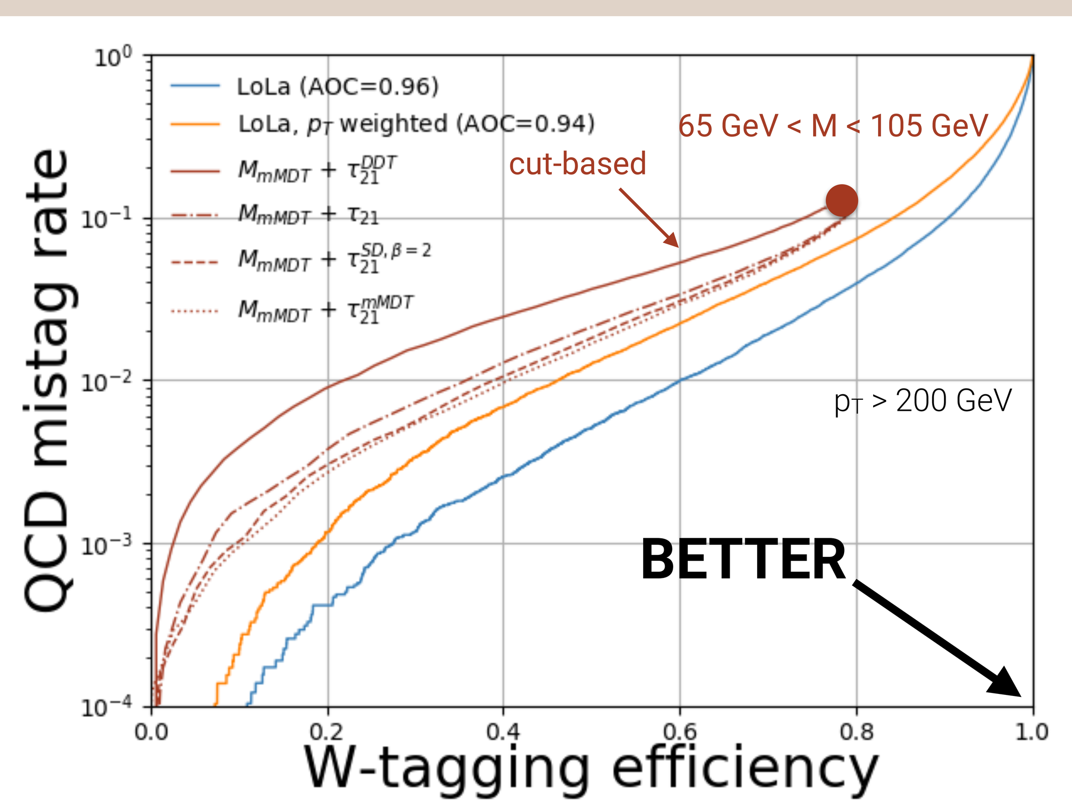
Due to the CoLa 1st column, quantities also computed for “total” jet, e.g distance between all jet constituents and jet axis.

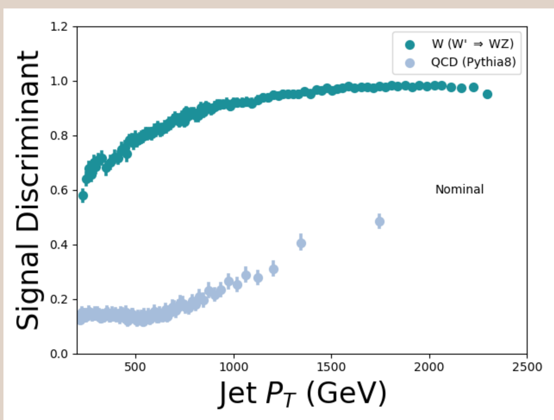
PERFORMANCE

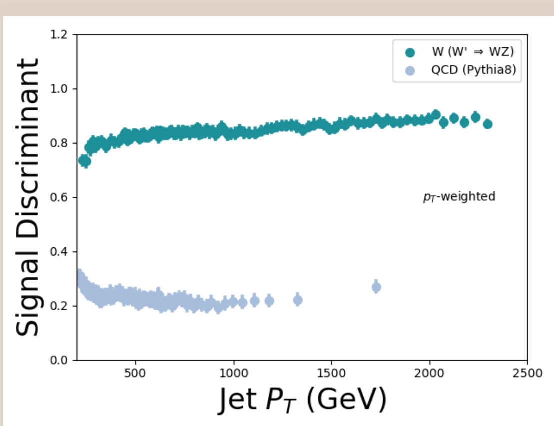
LoLa has a higher signal efficiency at any given mistag rate compared to the current baseline W-taggers in CMS.

Drop in performance observed when reweighting to jet p_T during training, but decreased overall p_T-dependence could still make this the better tagger in analysis.

Work ongoing to teach LoLa about tracks and vertices too, allowing it to learn b-tagging. Many exciting studies await!



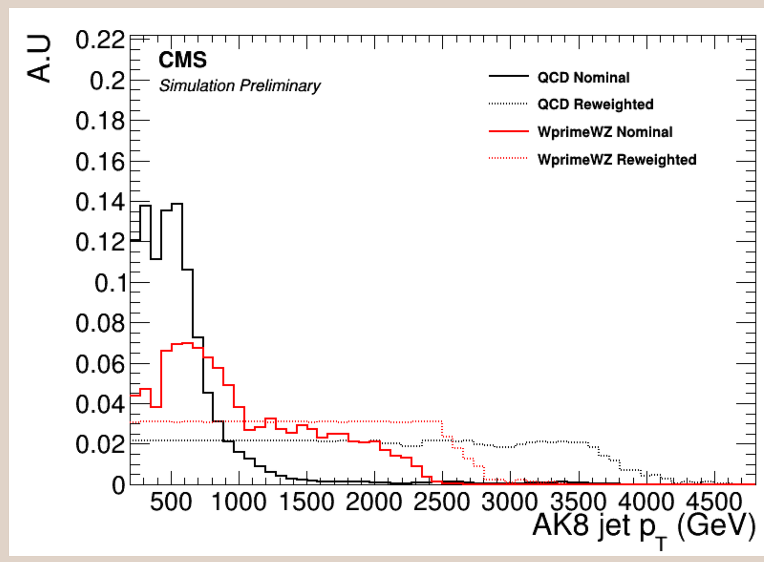




WHAT ABOUT p_T?

The jet transverse momentum spectrum of signal and background jets differ significantly. Do not want performance to depend on p_T.

Reweight training to be flat in p_T!



Thea Aarrestad (thea.aarrestad@cern.ch)
Gregor Kasieczka
Ben Kilminster

Universität Zürich UZH