DEEP LEARNING FOR BOOSTED TOP QUARK JET TAGGING

Suyong Choi, Yunjun Kim, Seung Joon Lee (Korea University)
Mihailo Backovich (CP3) , Maxim Perelstein (Cornell), Jesse Thaler (MIT)

INTRODUCTION

In many beyond the standard model scenarios, heavy particles can decay into top quarks

- these top quarks are boosted and hadronic decay products of top quark may not be resolved
- Top quark-jet tagging algorithms needed and there are already many such tagging algorithms

Boosted top quark jet tagging could be treated as an image recognition problem

- Deep learning methods excel at image recognition problems

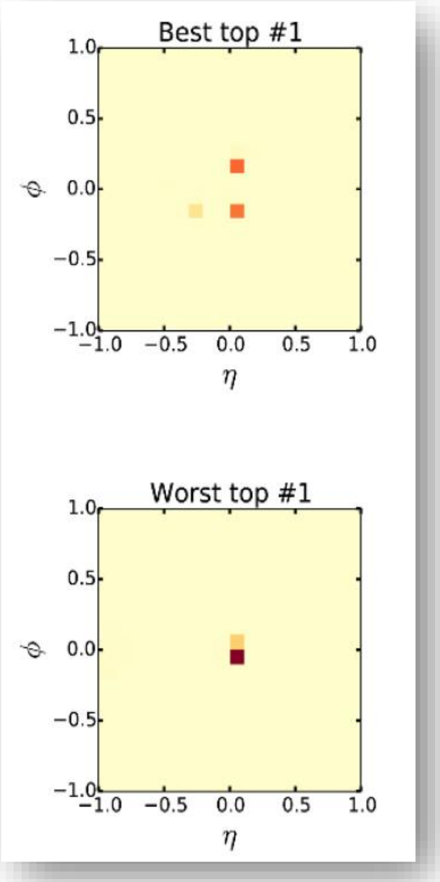
BOOSTED TOP QUARK JET

For top quark $p_T > 1\ \mathrm{TeV}$, hadronic jets fall well within cone of $R = 0.4$ and unresolvable

Energy deposition shown for top quark

- Jet reconstructed with cone size $R = 1.0$
- Madgraph 5 + Pythia 8 simulation
- Using only jets with $p_T = 800 - 900\ GeV$ and $m_J = 130 - 210\ GeV$

3 clumps of energy deposition visible, but not always



JHEP07 (2015) 086

QCD JET

Energy deposition shown for QCD jets

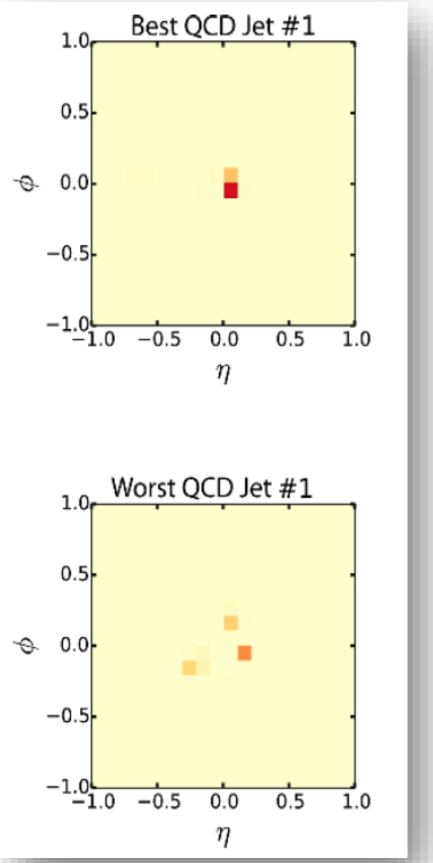
- Madgraph 5 + Pythia 8 events
- Same reconstruction and kinematic range $p_T = 800 - 900\ GeV$ and $m_J = 130 - 210\ GeV$

Mostly 1 clump of energy deposition visible, but some jets look top-like

- Not using b-tagging information here

Separation of top jet from QCD jet

- Various methods exist in the market



JHEP07 (2015) 086

SETUP FOR DEEP NEURAL NETWORK

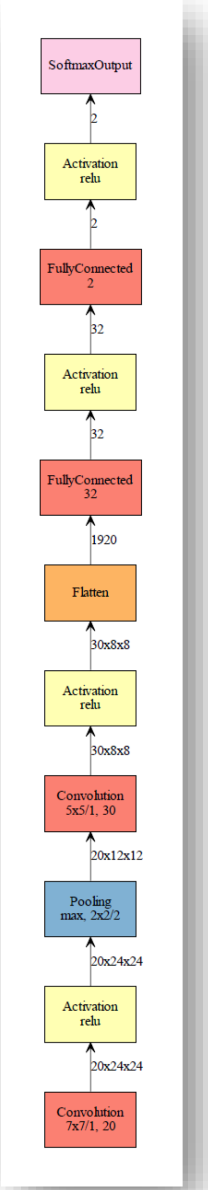
Convolutional Neural Network (CNN) in mxnet framework

- Input: 30 x 30 = 900 dimensional
- 30 convolution filters (7x7)
- 30 convolution filters (5x5)
- 64 fully connected nodes
- 32 fully connected nodes
- Non-linear activation function: RelU

Training sample size: 90k top jets, 90k QCD jets

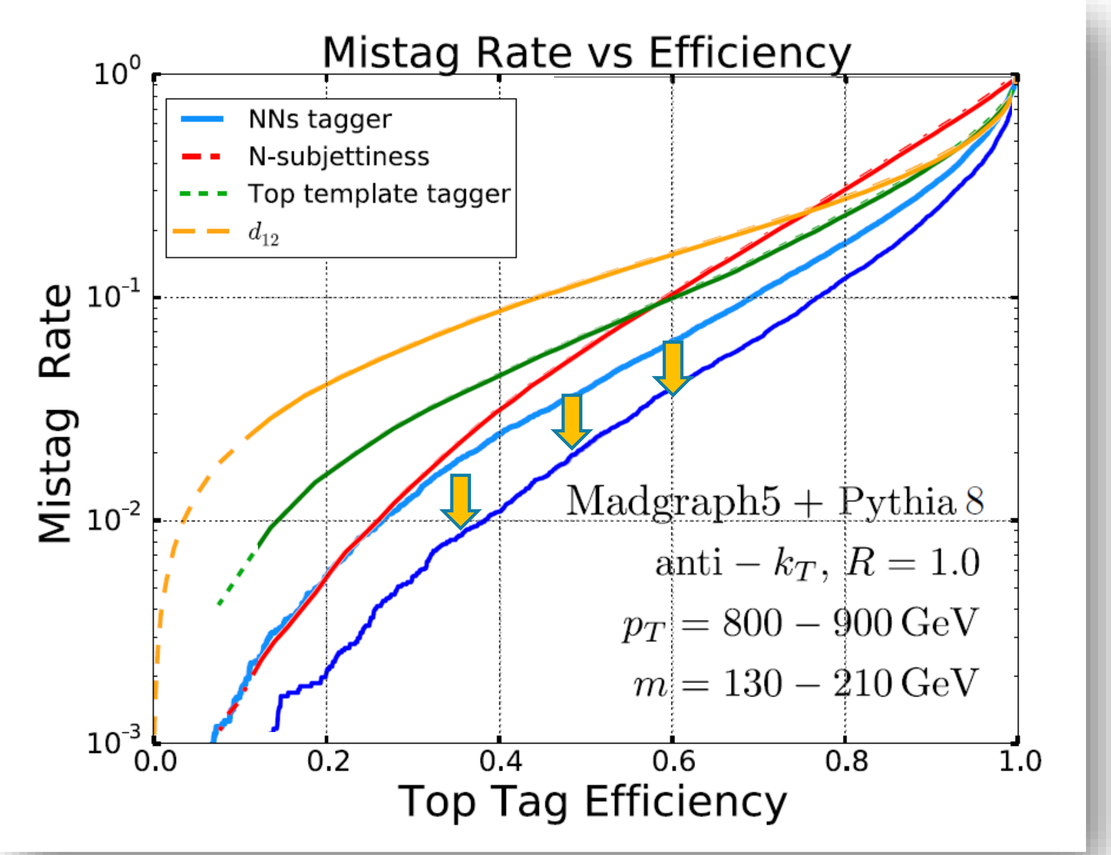
GPU: NVIDIA GTX 1060

- 1280 CUDA cores, 6GB



RESULT OF USING CNN

Factor 1.5~2 gain in rejection using convolutional neural network



USING N-SUBJETTINESS AS DEEP NN INPUT

N-subjettiness observables – You can think of these as moments

$$\tau_N(\beta) = \frac{1}{d_0} \sum_k p_{T,k}^{\beta} \min\{\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k}\}$$

  - $d_0 = \sum_k p_{T,k}^{\beta}$
- Jets composed of N subjets more likely to have small τ_N
- Ratios of τ's
  τ_3/τ_2 - used to separate $t \to bq\bar{q}'$ from QCD jets
  τ_2/τ_1 - used to separate W $\to q\bar{q}'$ or $H \to b\bar{b}$ from QCD jets

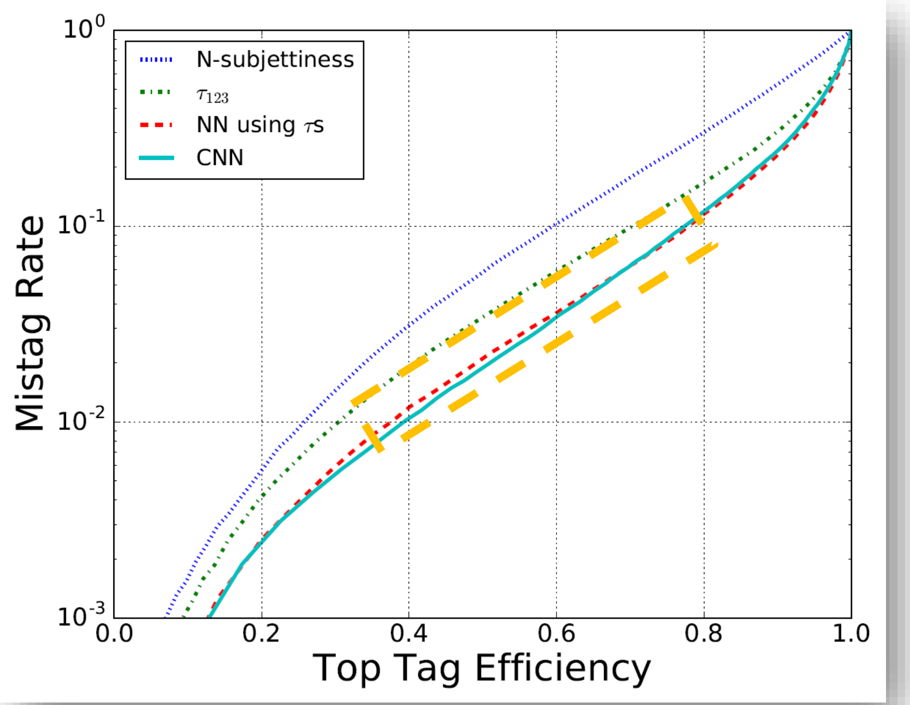
Why not use different τ's as inputs for deep NN training?

- τ's could be considered as moments and space spanned by different (N, β) could be as powerful

DEEP NN WITH N-SUBJETTINESS

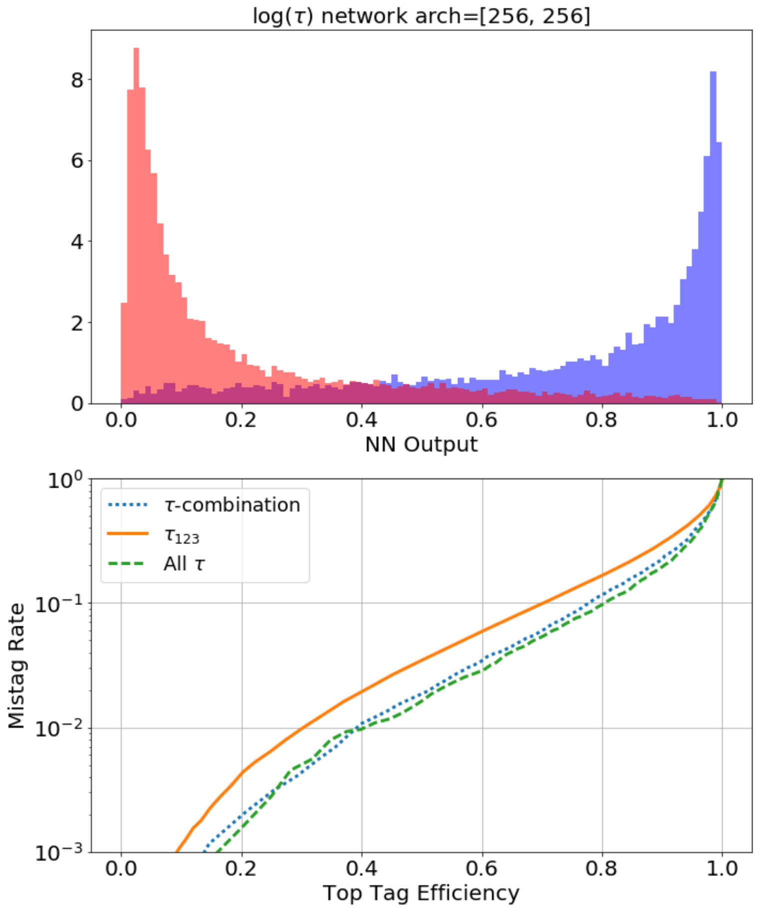
Performance comparable to using jet energy depositions

- 24 inputs - $\tau_1, \tau_2, \tau_3, \tau_4$ with $\beta = 0.1, 0.25, 0.5, 1, 1.5, 2$
- 3 hidden layers (48, 48, 16) used



τ_{123} : a new analytic variable found

SUBSET OF τ'S CLOSE TO OPTIMAL



Subset of τ's almost as powerful as using all of them: $\tau_{123}^{\beta=1}, \tau_4^{\beta=1}, \tau_1^{\beta=2}, \tau_2^{\beta=2}, \tau_3^{\beta=2}$

τ_5 has almost no effect

Labels

- τ-combination: $\tau_{123}^{\beta=1}, \tau_4^{\beta=1}, \tau_1^{\beta=2}, \tau_2^{\beta=2}, \tau_3^{\beta=2}$ with 256x256 hidden layer network
- $\tau_{123} = \tau_1\tau_2/\tau_3$ ($\beta = 1$)
- All τ : τ_N^β with $N = 1, \ldots, 5$ and $\beta = 0.2, 0.4, \ldots, 2.0$ with 256x256x256 hidden layer network

NN JET ALGORITHM PHYSICALLY MEANINGFUL?

Does it exhibit safety from infrared and collinear divergence?

$$O_n(k_1, \dots, k_i, k_j, \dots, k_n) \xrightarrow{k_i\,, k_j\; soft\;\; or\; k_i \| k_j} O_{n-1}(k_1, \dots, k_i + k_j, \dots, k_n)$$

- It should not be sensitive to the presence of soft or collinear gluon radiation
- Fundamental requirement of a jet algorithm to be physically sound

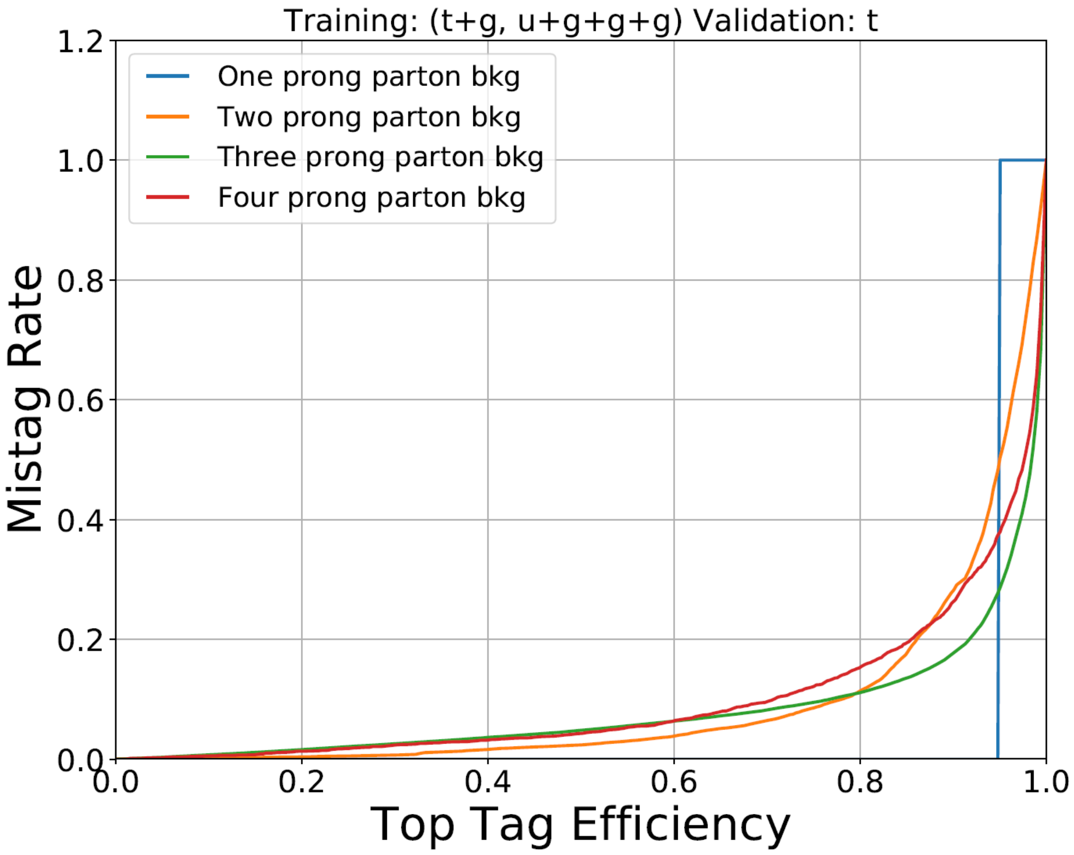
BOOSTED TOP VS QCD

Training:

- Signal: $t + g$
- Background: $uggg$

Validation:

- Signal: t
- Background: $u, ug, ugg, uggg$



CAN YOU TEACH $\sqrt{E^2 - p^2} = m$?

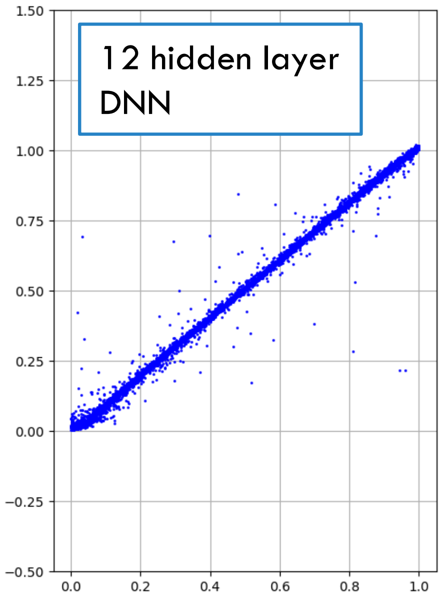
For a mother particle of mass m decaying into $1 + 2$, can a DNN learn to calculate mass m given $\{p_{1,x}, p_{1,y}, p_{1,z}, p_{2,x}, p_{2,y}, p_{2,z}\}$?

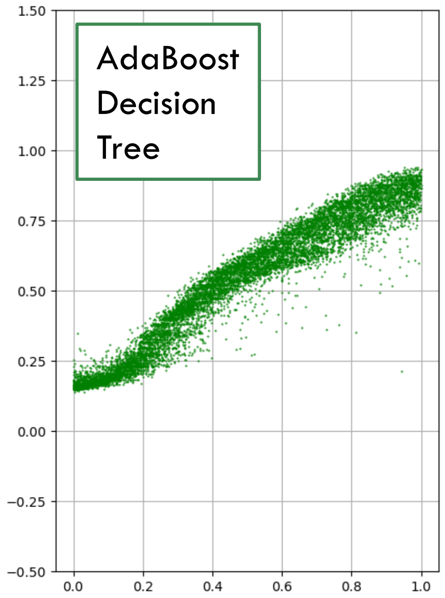
- Performance depends on whether the target is m or m^2

Deep NN with 12 hidden layers

- # of hidden nodes: [512, 128, 256, 32, 512, 512, 64, 128, 256 , 512, 256, 256]

With DNN, we can probably use four vector components directly for complicated events



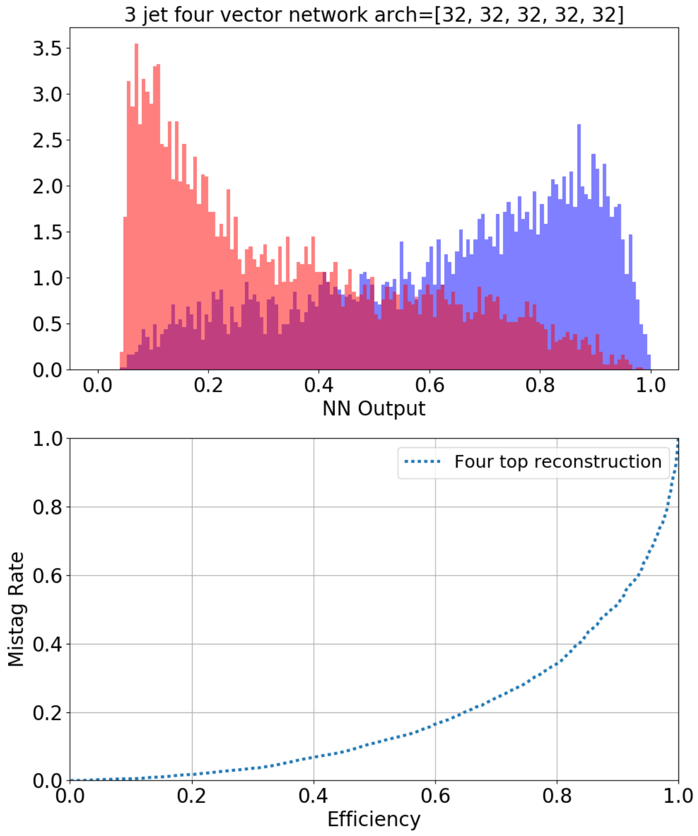


HADRONIC FOUR TOP

Four top quarks decaying hadronically produce 12 hadronic jets

Finding 3 jets from the same top quark is challenging

- Large number of possible combinations,
- Usually wrong combination gets selected when using kinematic reconstruction on events with jet energy resolutions



TOWARDS SEMI-SUPERVISED METHODS

Data problem

- For deep learning we need lots of data, but for some, we cannot generate enough simulated data using GEANT due to resource constraints
- For some backgrounds, our simulation model may not be good enough

Anomaly detection

- What if we do not know what the New Physics looks like?
- In a supervised learning, we need the signal sample

Semi-supervised training methods to learn $P(\{x_i\})$

- Generate augmented data – to solve the data problem
- Detect anomalies in the data – possible to use as a trigger?
- How to do it? This is where research is needed.

SUMMARY AND OUTLOOK

Deep learning applied to top quark jet tagging looks promising

- Energy depositions as image recognition problem
- Using DNN with many hidden layers in reduced N-subjettiness variables space is as good
- Qualitatively looks IR and collinear safe

Deep learning could be used in other problems with complicated final states

- multi-top quarks, multi-Higgs, etc.
- Four vectors can be used directly

Future directions

- Development of semi-supervised methods
- Interpreting DNN results to gain insight – DNN inspired physical observables. we probably need combination with analytical methods