Deep-learning top taggers using Lorentz invariance

Based on
Kasieczka, Plehn, MR, Schell '17
Butter, Kasieczka, Plehn, MR '17

Michael Russell
ITP Heidelberg

PHENO workshop, Pittsburgh, May '18

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Top tagging

Modern top reconstruction

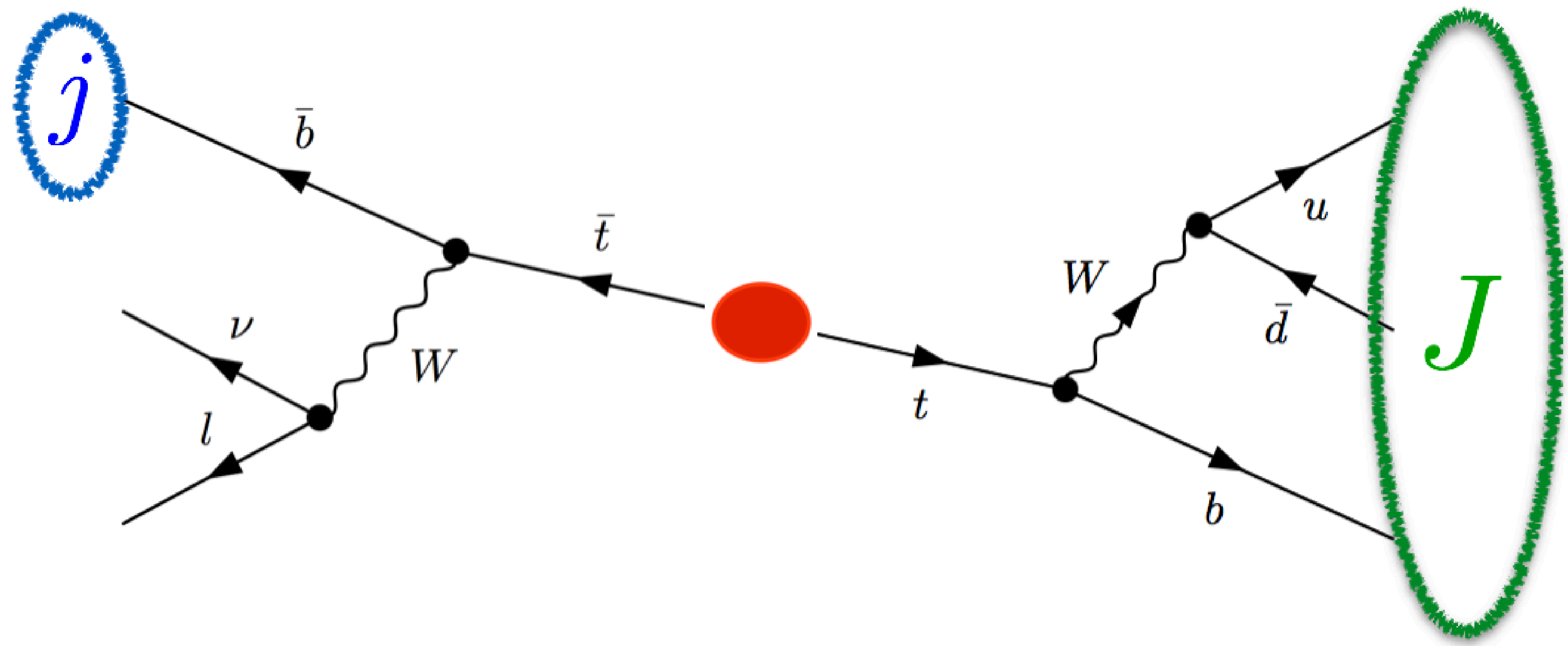
- Easy to reconstruct tops when decay products are well-separated

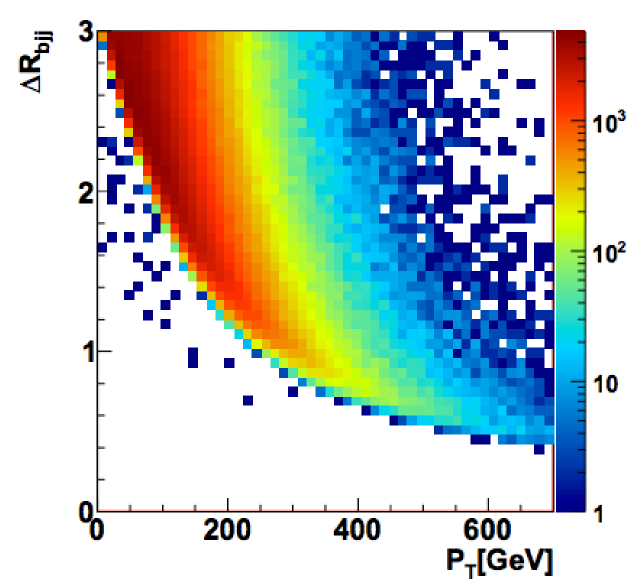
- But standard reconstruction methods fail when tops are highly-boosted

- Instead of trying to resolve decay products individually, merge all into “fat jet”

Kaplan, Rehermann, Schwartz, Tweedie ‘08
Plehn, Spannowsky, Takeuchi, Zerwas ‘10

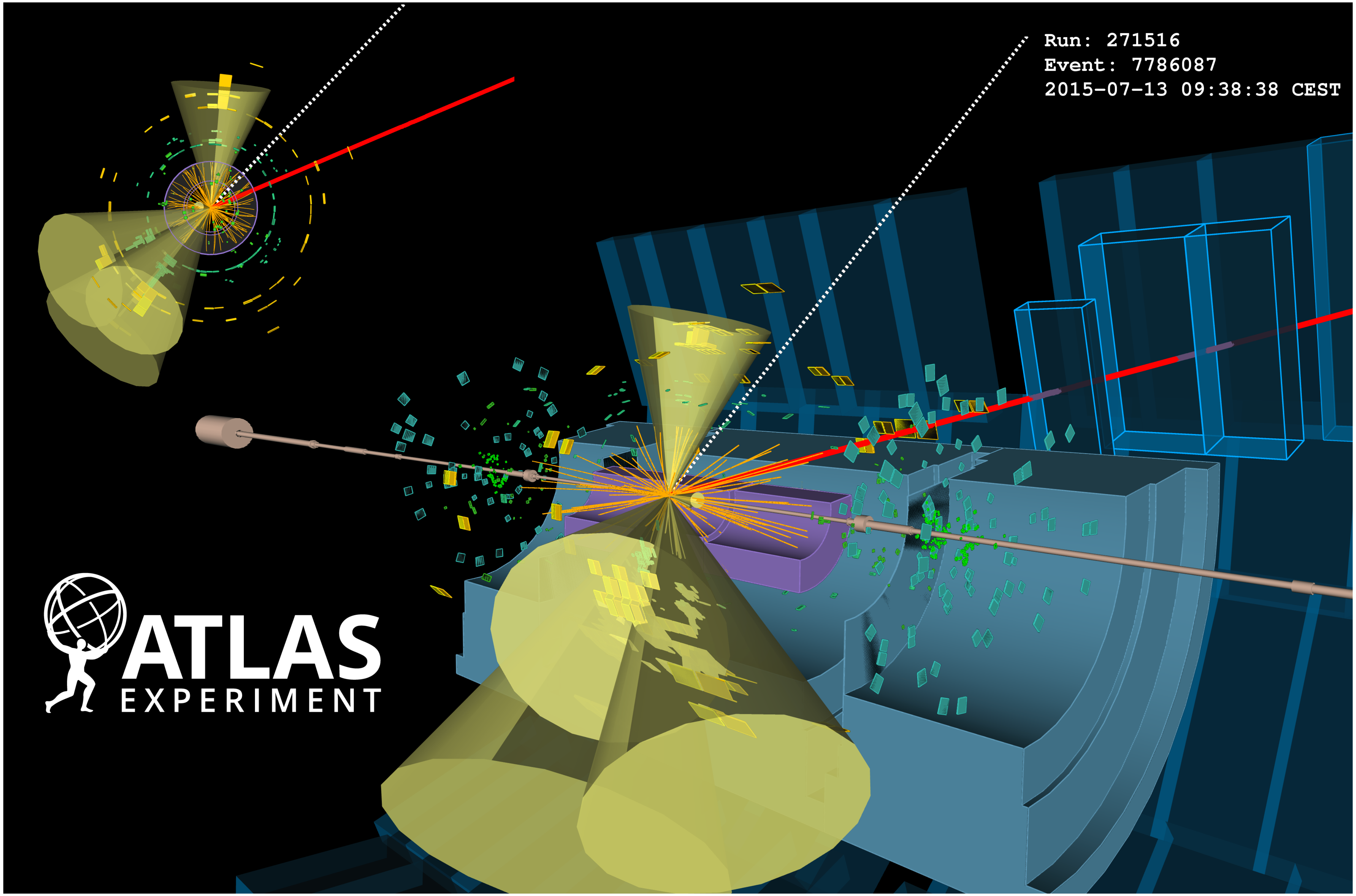
- By now this is well-understood experimentally and theoretically





$$\Delta R \sim \frac{2m}{p_T}$$

We see something like this



Jets as images

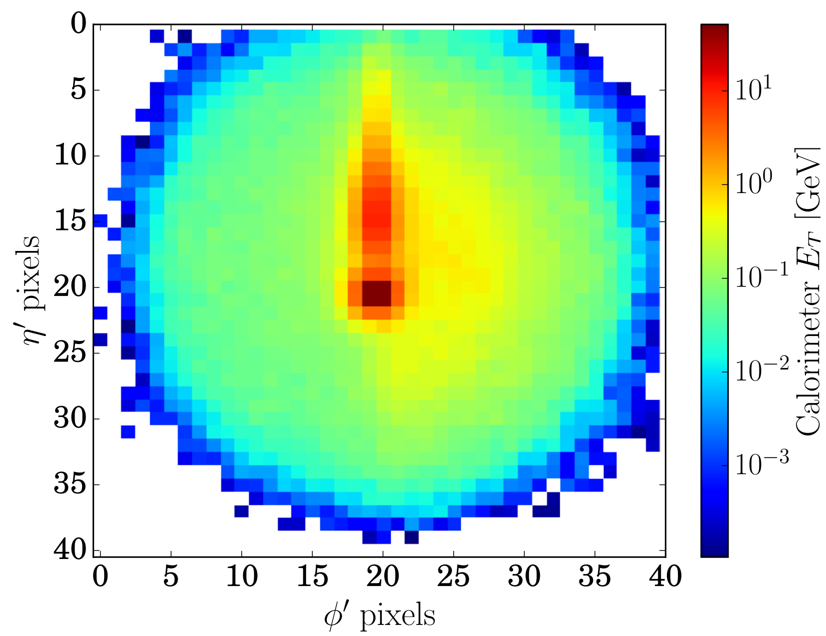
de Olivera, Kagan, Mackey, Nachman, Schwarzmann ‘15

Can recent advances in machine learning benefit particle physics?

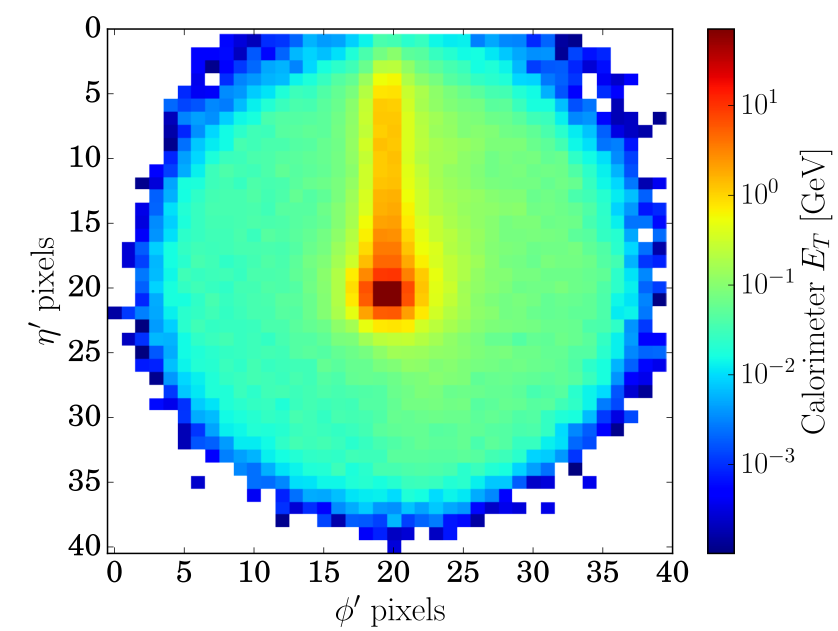
- View calorimeter plane as 2-d “image” with energy deposits as pixels
- After some pre-processing, train a *convolutional neural network* (no details here) on sample of top jets and QCD background

Kasieczka, Plehn, MR, Schell ‘17

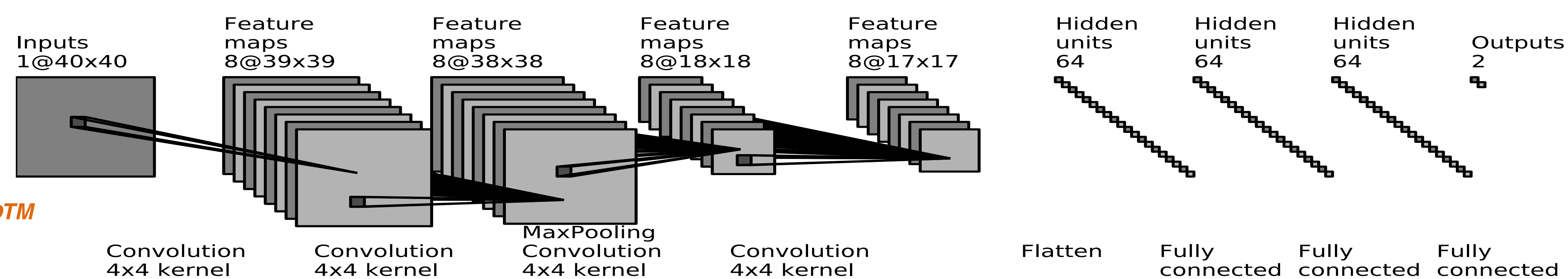
tops



QCD



- Last layer of network converts weights for each image into probability of it being either top or QCD



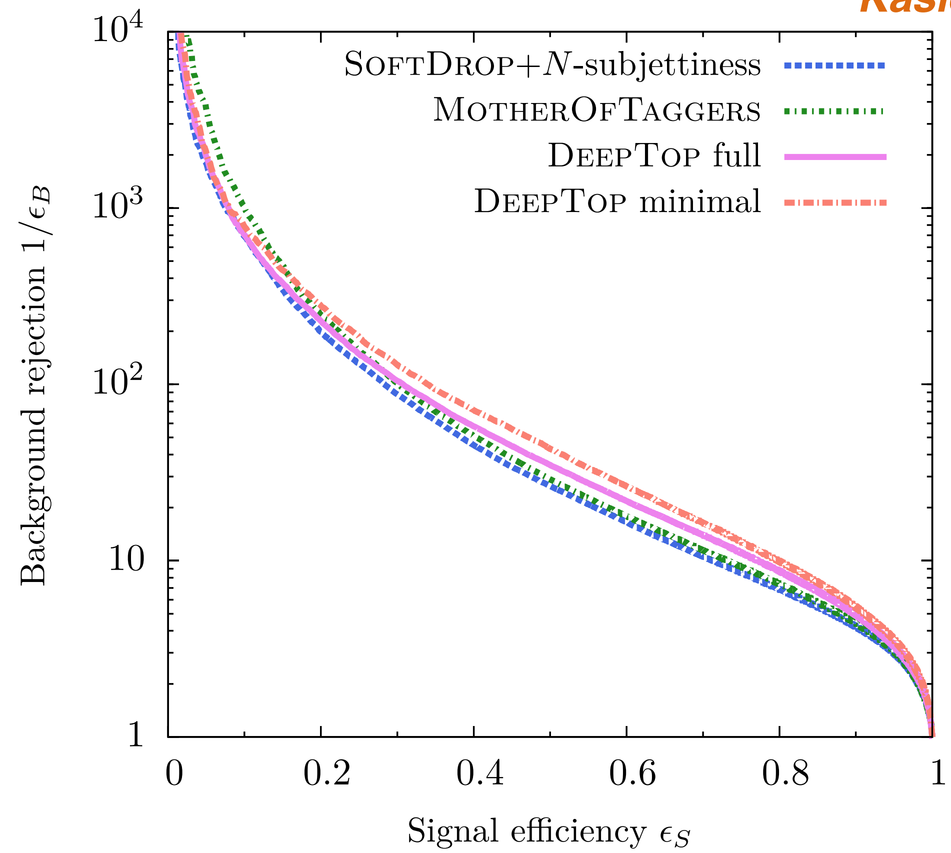
DEEPTOP™

Jets as images

Test performance against traditional taggers and BDT

Kasieczka, Plehn, MR, Schell '17

- Can see up to 50% improvement in background rejection
- Deep neural networks outperform BDT classifiers - logical next step for machine learning on real data
- Preprocessing actually causes loss of information in final classification



But images have many limitations

- Cannot include tracking information
- Not adaptable for non-uniform detectors
- Can we use more physics-motivated inputs, not weird "pixels"?

Butter, Kasieczka, Plehn, MR '17

Beyond images: LoLa

Why not use the jet constituent 4-vectors directly?

Two ingredients:

1. CoLa* - learns the jet clustering history

$$k_{\mu,i} \xrightarrow{\text{CoLa}} \widetilde{k}_{\mu,j} = k_{\mu,i}\, C_{ij}$$

$$C = \begin{pmatrix} 1 & 0 & \cdots & 0 & C_{1,N+2} & \cdots & C_{1,M} \\ 0 & 1 & & \vdots & C_{2,N+2} & \cdots & C_{2,M} \\ \vdots & \vdots & \ddots & 0 & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & C_{N,N+2} & \cdots & C_{N,M} \end{pmatrix}$$

- Test on-shell conditions

$$\tilde{k}^2_{\mu,1} = (k_{\mu,1} + k_{\mu,2} + k_{\mu,3})^2 = m_t^2$$

$$\tilde{k}^2_{\mu,2} = (k_{\mu,1} + k_{\mu,2})^2 = m_W^2 \,.$$

2. LoLa** - learns the kinematics

$$\tilde{k}_j \xrightarrow{\text{LoLa}} \hat{k}_j = \begin{pmatrix} m^2(\tilde{k}_j) \\ p_T(\tilde{k}_j) \\ w^{(E)}_{jm}\, E(\tilde{k}_m) \\ w^{(d)}_{jm}\, d^2_{jm} \end{pmatrix}$$

transform 4-vectors into: invariant mass, pT, energy and Minkowski distance

effectively a rotation in observable space

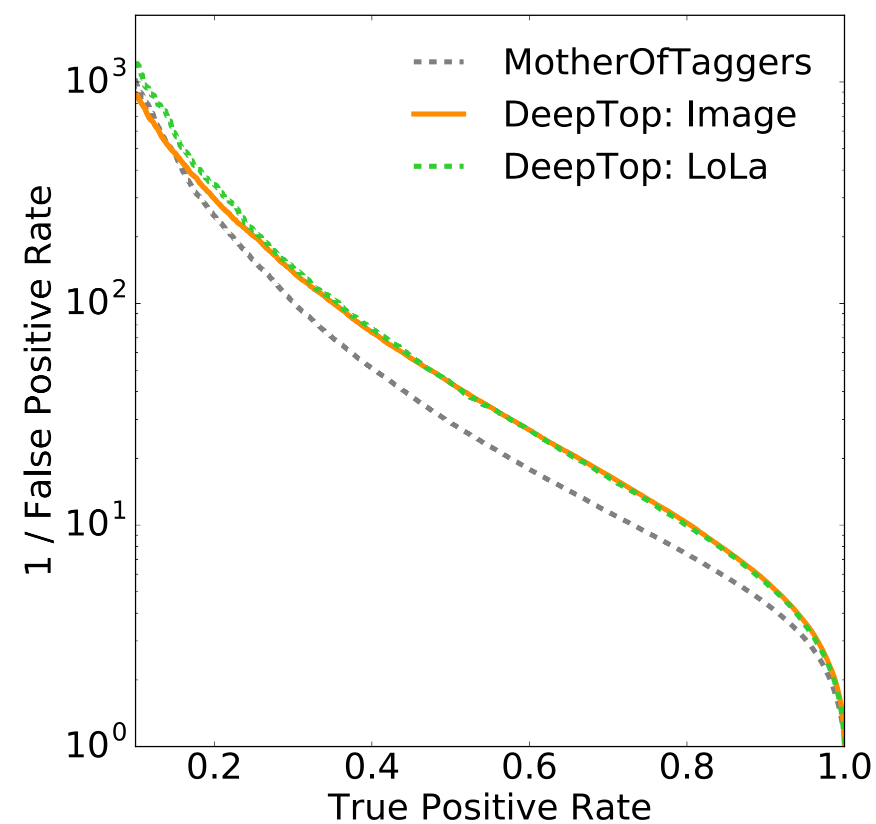
** CoLa = Combination Layer*

*** LoLa = Lorentz Layer*

Performance of LoLa

First test: do we do better than images?

- Using calorimeter information only, no improvement over images (unsurprising)
- Evidence that LoLa learns the same features as image-based approach
- Far less training time, fewer weights, fewer inputs required
- Same performance for much less CPU time = suggests we should move away from images



But how does LoLa improve physics performance?

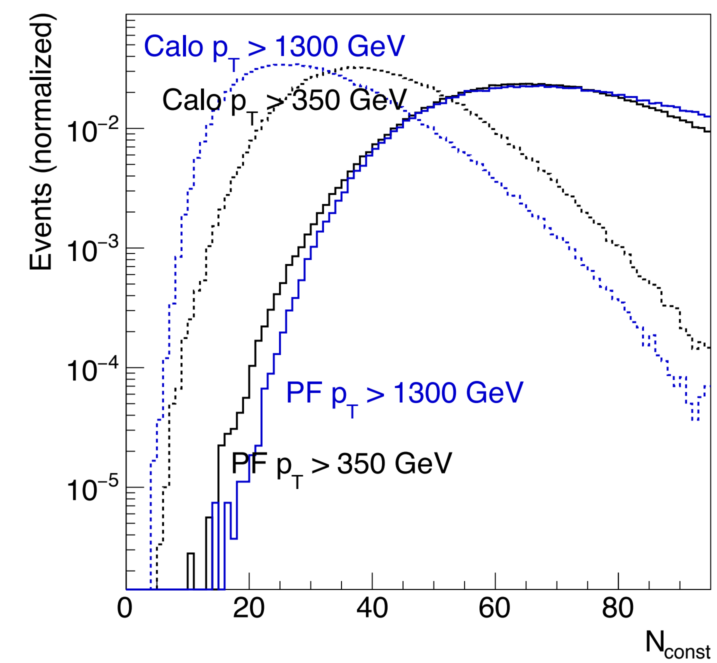
Ultra-boosted tops

Calorimeter resolution degrades for high pT tops

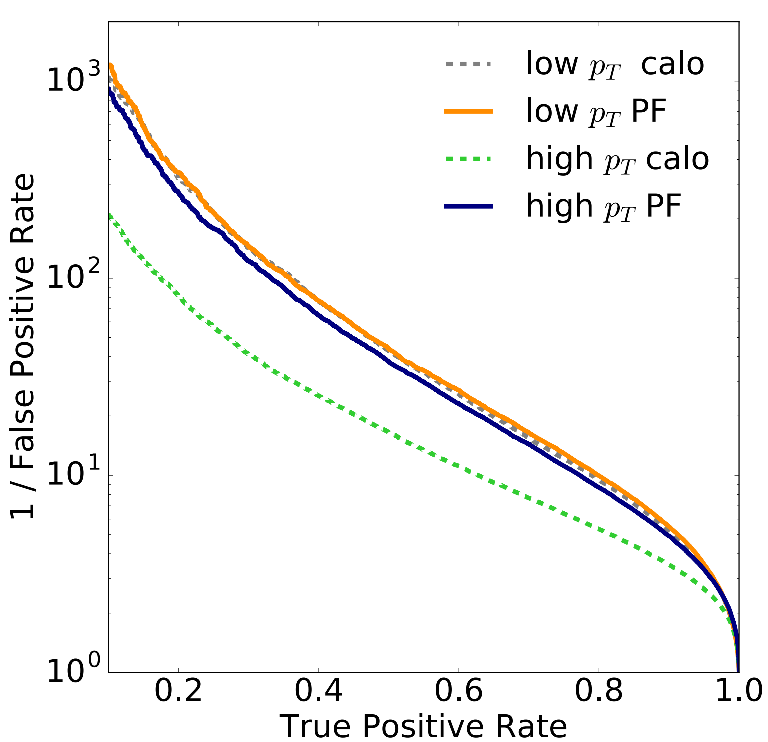
More likely that jet constituents will land on same cell, so loss of information

Use much higher-resolution tracking

Same number of constituents at high pT so no loss of info



Massive increase in performance



Impact on resonance searches?

Conclusions

- Recent developments in machine learning have found novel and exciting applications in top tagging
- Two approaches presented here: image-based and 4-vector based
- Both show excellent ability to identify hadronic top decays
- LoLa-based approach has more physics-motivated inputs, simpler network architecture, less CPU time
- Ability to include tracking and extend to very high pT
- Time to start on real data?

Backup: Analysis details

Signal: all-hadronic ttbar, Background: QCD dijets

(PYTHIA8 + Delphes)

Cluster calorimeter towers or particle-flow objects into fat jets

$$\text{C-A } \Delta R = 1.5$$
$$350 \text{ GeV} < p_{T,J} < 450 \text{ GeV}$$
$$|\eta_J| < 1.0$$

Sort jet constituents by pT, feed four-vectors into NN

300k signal and background events

Train/test/validation split: 60/20/20

Backup: preprocessing

Don't want to waste network parameters on learning special relativity, pre-process to remove this dependence