How to Train Taggers on Data ...and Other Problems

Jennifer Thompson

Universität Heidelberg

17.07.2018

BOOST 2018

Gregor Kasieczka[1], Tilman Plehn[2], Michael Russell[2], **Jennifer Thompson**[2]

[1] Institut für Experimentalphysik Universität Hamburg [2]ITP Universität Heidelberg

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

Overview

Machine Learning with LoLa

- Machine learning techniques can be used in many physics problems
  $\longrightarrow$ e.g. event classification
- Can we use an NN as a top tagger?
  $\longrightarrow$ Yes! With images

  Kasieczka, Plehn, Russell, Schell arXiv:1701.08784v2 [hep-ph]

  Macaluso and Shih arXiv:1803.00107v1 [hep-ph]

  $\longrightarrow$ and with constituents (LoLa)

  Butter, Kasieczka, Plehn, Russel arXiv:707.08966v3 [hep-ph]

  $\longrightarrow$ We also have trees and graphs

  Henrion et al (NIPS 2017)

  Louppe et al arXiv:1702.00748

Cola and LoLa in Equations

The combination layer (CoLa) acts on 4-momenta $k_{\mu,i}$:

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

and Lorentz layer (LoLa)

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

trainable

Cola and LoLa in Equations

The combination layer (CoLa) acts on 4-momenta $k_{\mu,i}$:

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

and Lorentz layer (LoLa)

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

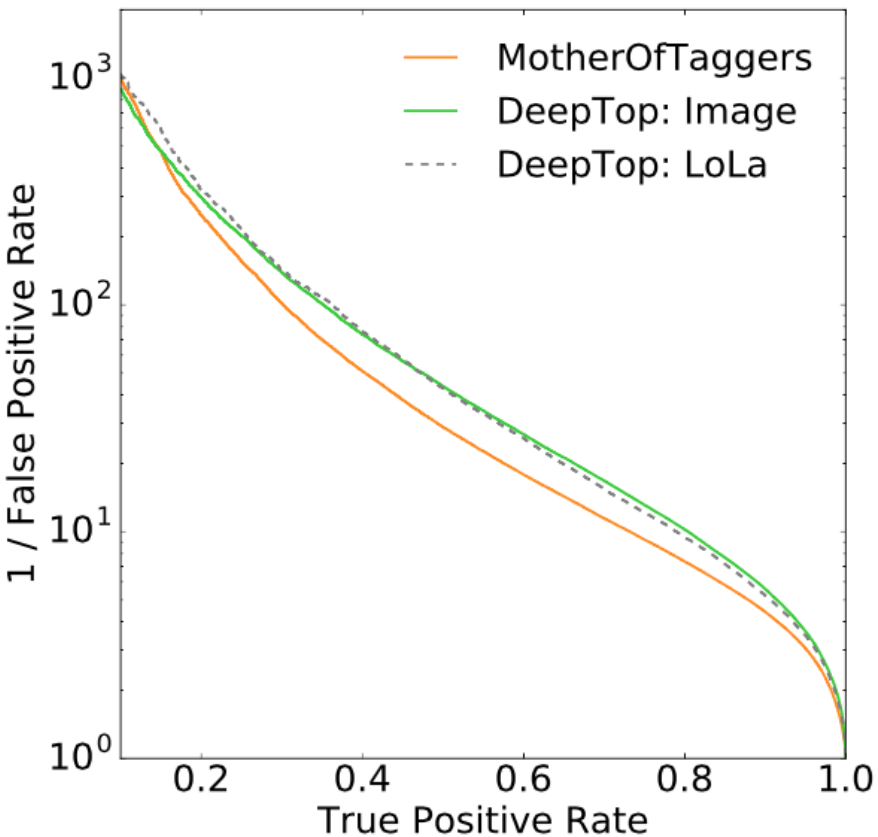
$\longrightarrow$ make d_{jm}^2 trainable:

$$g = \mathrm{diag}(0.99 \pm 0.02, -1.01 \pm 0.01, -1.01 \pm 0.02, -0.99 \pm 0.02)$$

$\longrightarrow$ Minkowski metric learnt! trainable

DeepTop vs. Conventional Taggers

Butter, Kasieczka, Plehn, Russel arXiv:707.08966v3 [hep-ph]

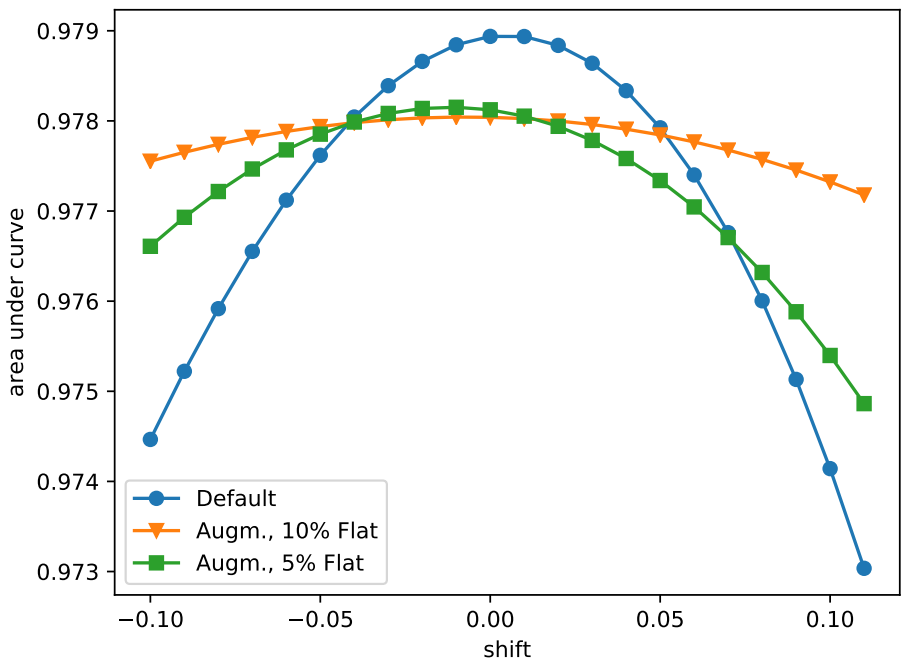


- NN training outperforms traditional taggers
- DeepTop and DeepTopLoLa perform similarly
- This is only the beginning of NN taggers
- Further improvements possible

Systematic Uncertainties

Sven Bollweg, Andriy Borovkov

- Study imact of systematic uncertainties on LoLa
- Start with fully correlated re-scaling of four-vectors (poor-man's jet energy scale)
- Test impact of data augmentation



$\longrightarrow$ **Trade-off between stability and performance**

Quark-Gluon Discrimination

Quark Gluon

Quark-Gluon Discrimination

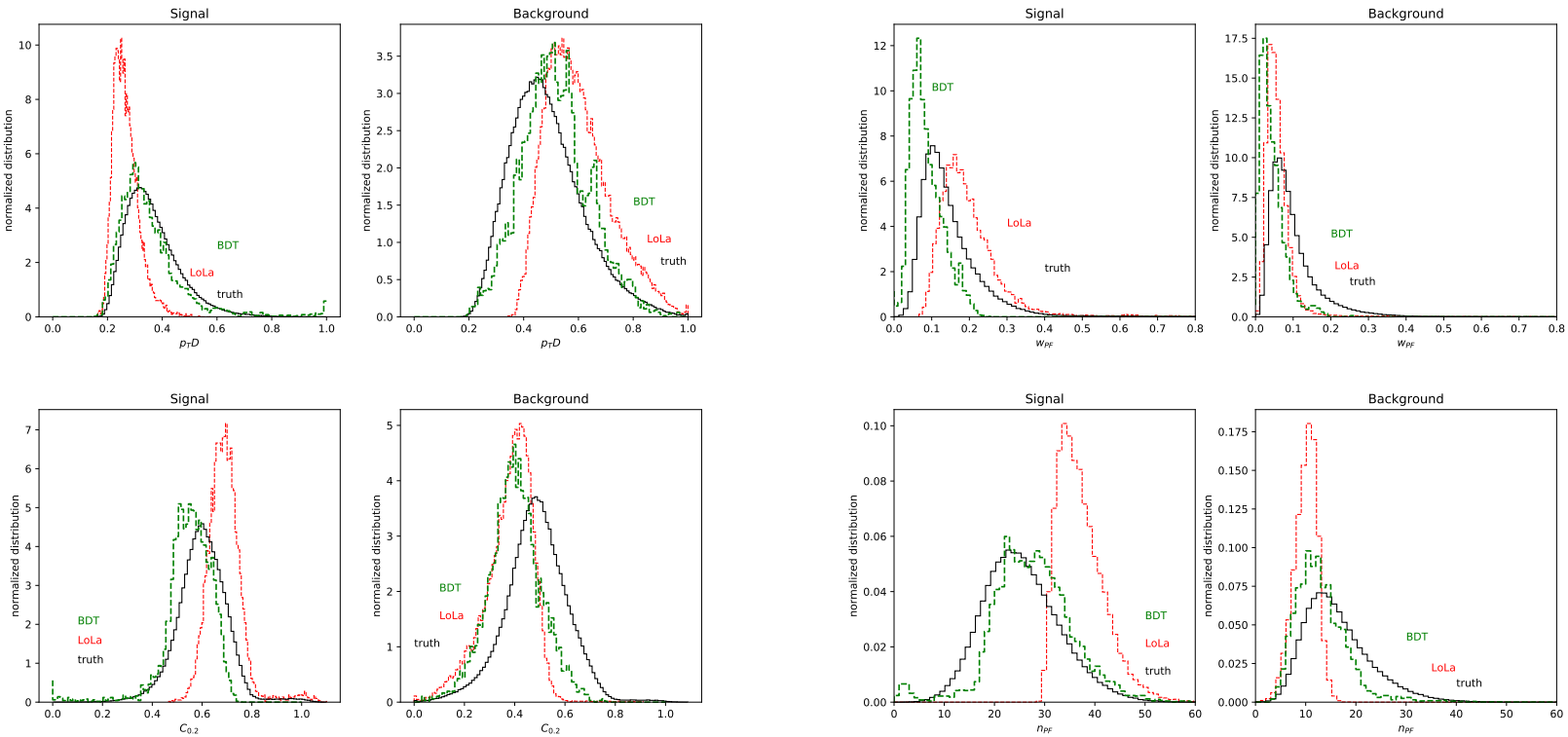
We can distinguish tops. What other problems can we solve?

$\longrightarrow$ quark-gluon discrimination

- Harder problem than top vs. QCD
- Could be useful in $H \to E_T^{\text{miss}}$ searches
- Consider q-g discriminating variables on top of usual LoLa set-up.
- First step: pure quark/pure gluon samples

Classification of Signal (Gluon) and Background (Quark)

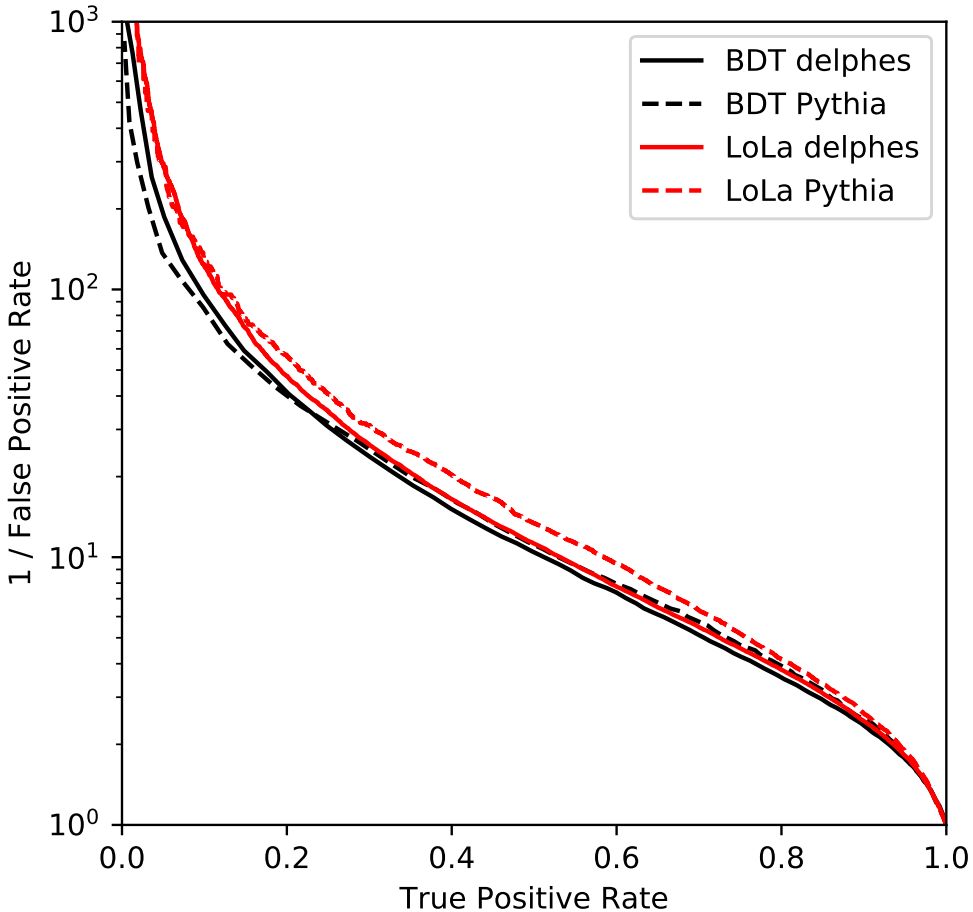
Nicholas Kiefer



$\longrightarrow$ 30% correctly classified, most signal/background-like events

$\longrightarrow$ NN and BDT show good discrimination power

Results: Preliminary



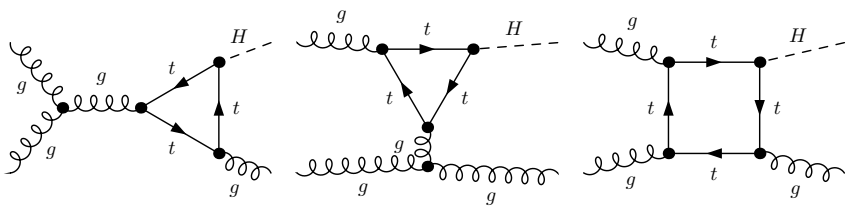
Nicholas Kiefer

- LoLa performs better than standard BDT analysis
- Some loss of performance when detector simulation is included

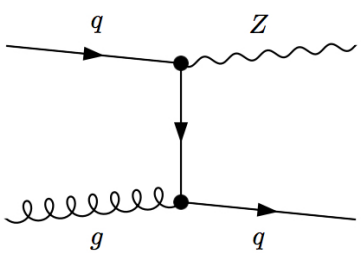
Next Steps: Physics Application

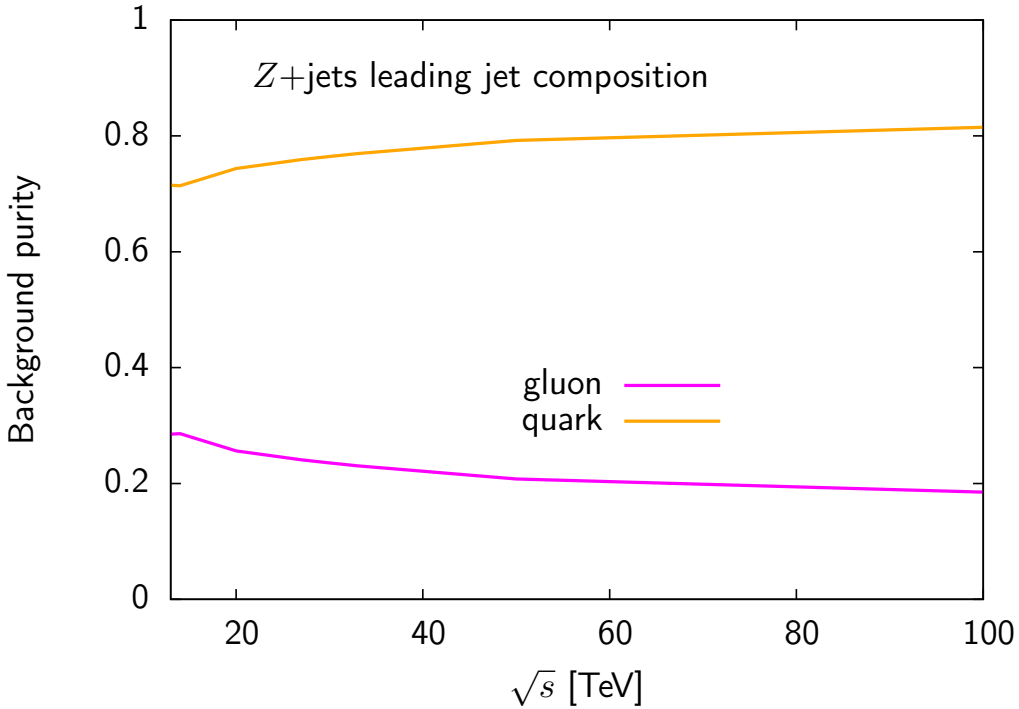
Consider $H(\rightarrow E_T^{\text{miss}})$+jets

Signal∼80% gluon



Background∼80% quark





⟶ Signal plot is similar, with gluon↔quark

Signal Independent Training

Signal Independent Training

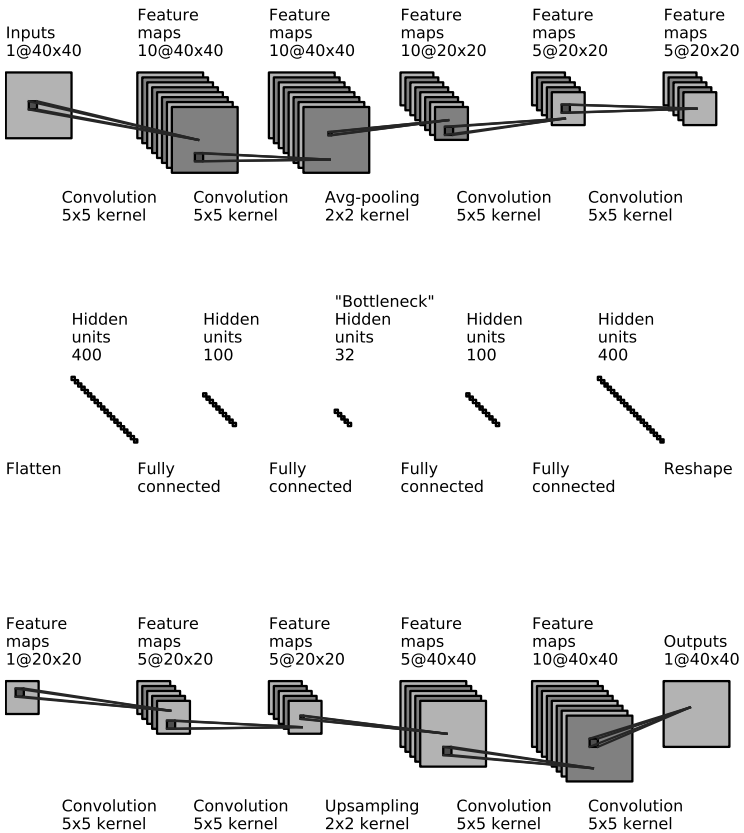
Training on Background

Publication coming soon

- **New way to generically search for new physics: Training on data**
- Allows for generic NP searches
  $\longrightarrow$ No need to understand the signal exactly
  $\longrightarrow$ Dark showers are ideal
- Use an autoencoder
  $\longrightarrow$ encode image and attempt to reconstruct it
  $\longrightarrow$ Present image-based and LoLa-based implementation

Autoencoder Architecture for Images

Theo Heimel

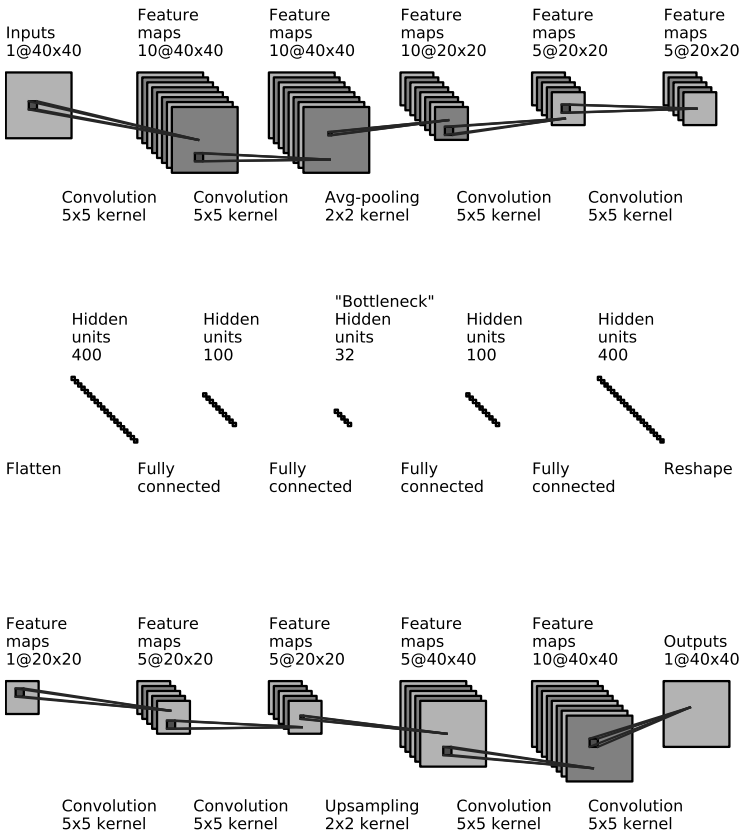


- Convolutional layers with average pooling
- Dense layers narrowing to a bottleneck
- Reconstruction: Build dense layers back up
- Undo pooling

$\longrightarrow$ good representation of background: fails for signal.

Autoencoder Architecture for Images

Theo Heimel



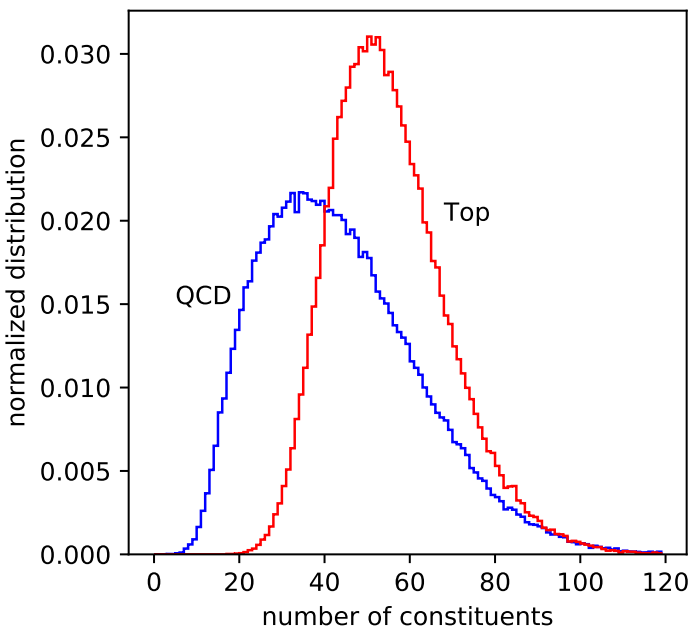
- Convolutional layers with average pooling
- Dense layers narrowing to a bottleneck
- Reconstruction: Build dense layers back up
- Undo pooling

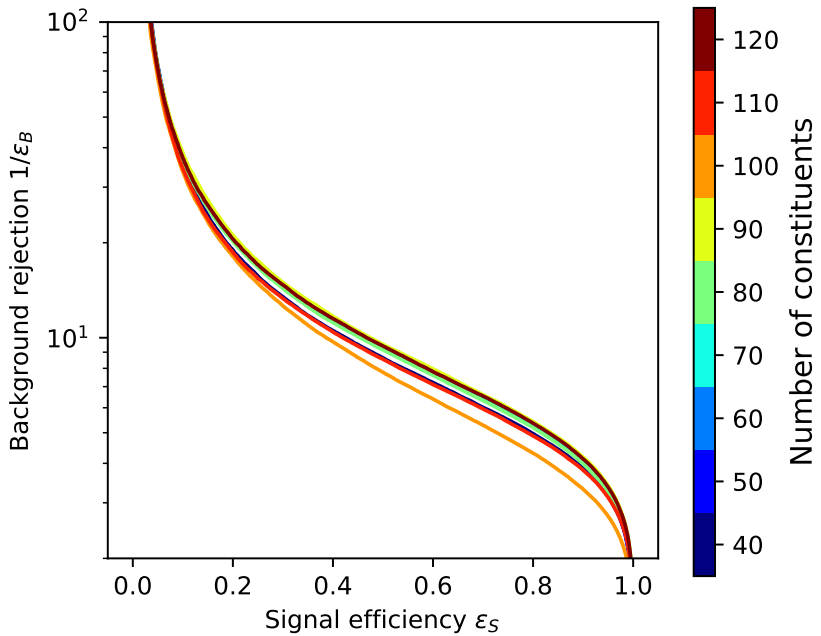
Relevant layers for LoLa

$\longrightarrow$ good representation of background: fails for signal.

LoLa Top-Tagging Results

Theo Heimel





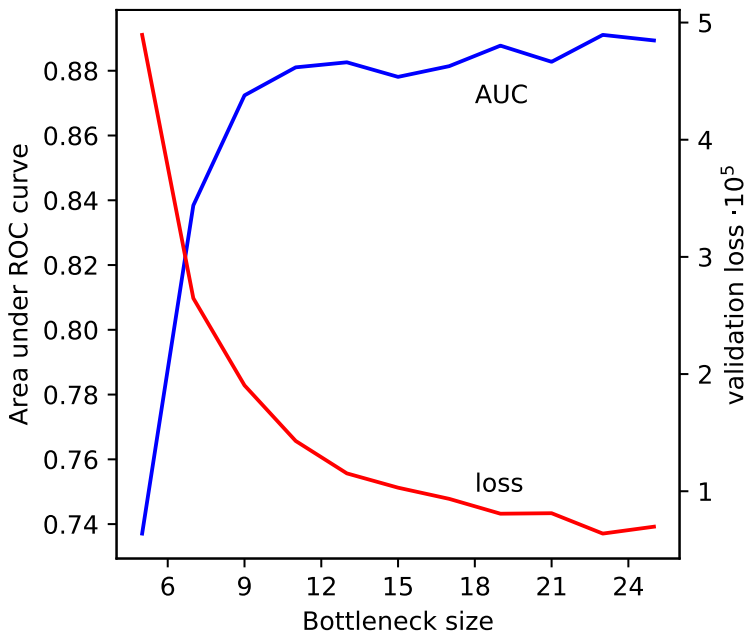
$\longrightarrow$ Apply to a familiar problem: can it find tops?

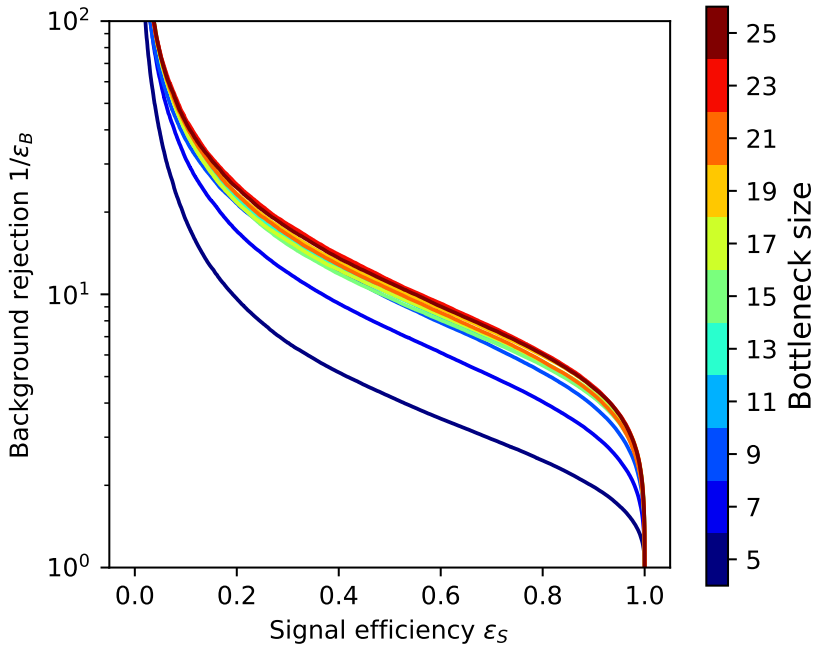
$\longrightarrow$ Train entirely on QCD background

$\longrightarrow$ Network stable w.r.t. number of constituents

LoLa Top-Tagging Results

Theo Heimel

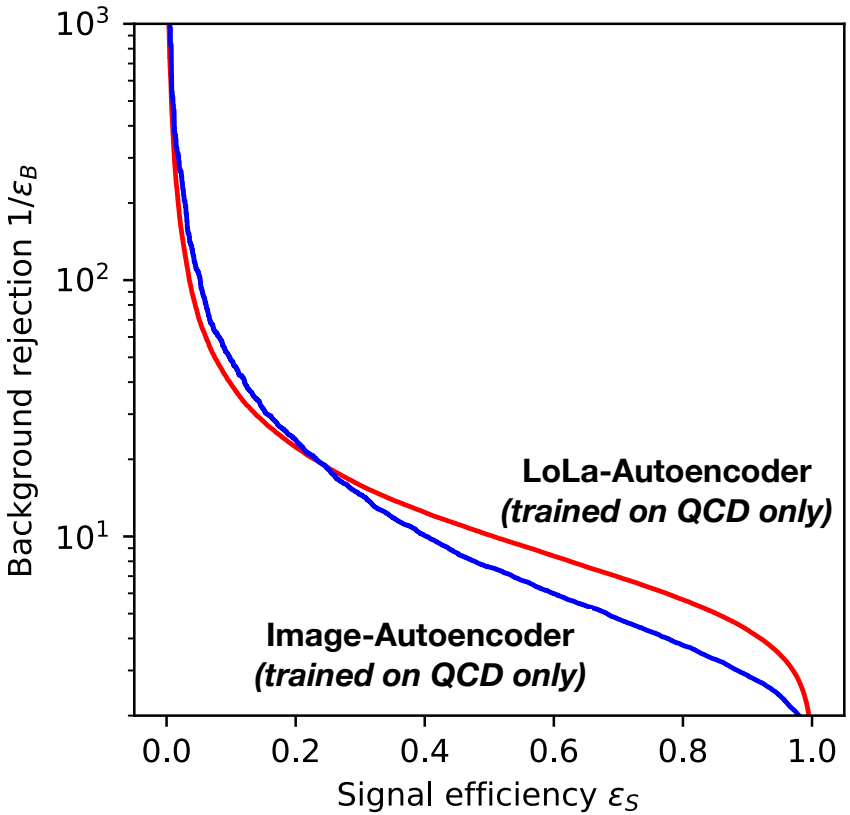




$\longrightarrow$ Loss of information at small bottlenecks, no chance to learn

$\longrightarrow$ stable from ~ 12

LoLa vs. Image: Preliminary



Theo Heimel

- The network can find tops when trained only on QCD
- Good performance of LoLa at high signal efficiency
- Top as a benchmark for comparison, but **we are asking what it DOESN'T look like**
- **AUC=0.88 $\rightarrow$ with a network that has never seen signal**

Transfer Learning

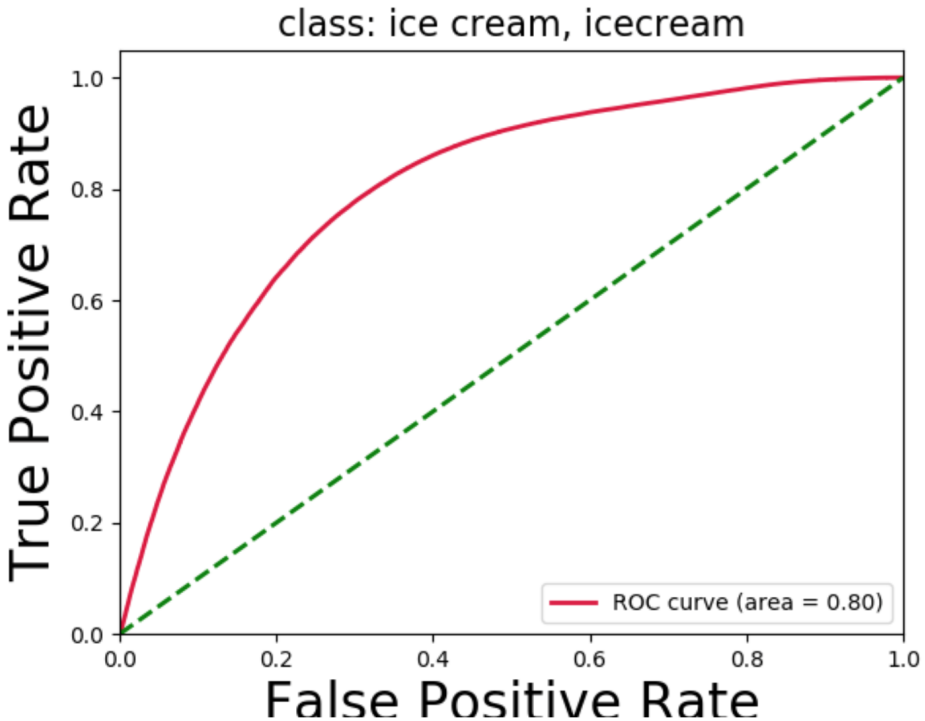
Another interesting idea:

- We know we can classify jet images as top/QCD events
- Google has developed an image analyser
- Architecture and weights are publically available
  $\longrightarrow$ InceptionResNetV2 https://keras.io/applications/#inceptionresnetv2
- How does this perform for top vs QCD images?
  $\longrightarrow$ Out of the box
  $\longrightarrow$ Retrained on jet images

Results: What is the Flavour of QCD?

Lisa Benato

⟶ Choose the best Google image class for QCD background:



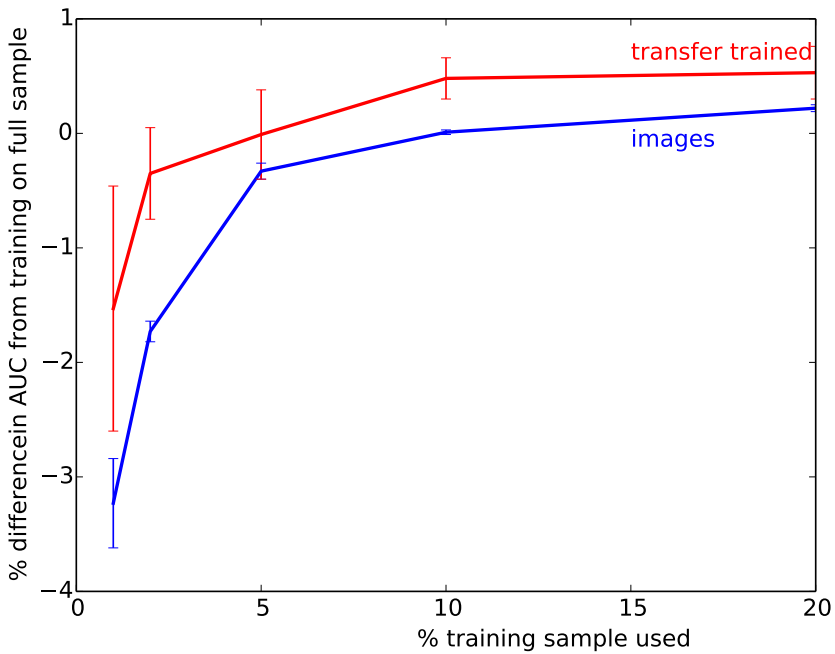
⟶ Turns out to be ice-cream! So: anti-ice-cream tagger

⟶ **This tagger has never seen signal or background**

Results: If we train the network

Lisa Benato

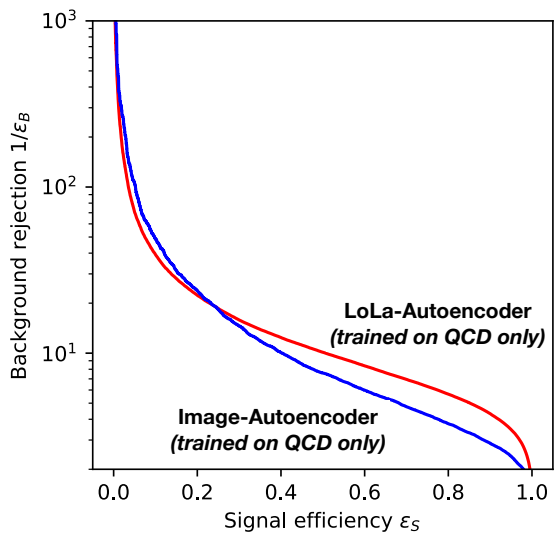
- What if we retrain the network?
- Very flexible framework for a tagger
- More stable, better AUC than dedicated images



$\longrightarrow$ **Very powerful tool for classification**

Conclusions

- LoLa is a powerful physics-based tagging architecture, works for tops vs QCD and q/g tagging
- **First unsupervised training of tagger for new physics:**
  Autoencoder approach
  - Works for Images and Lola
  - Can be trained using only background data
- InceptionResNetV2 can discriminate top and QCD out of the box
  - Transfer learning with less training data.



Conclusions

Samples are available: https://goo.gl/XGYju3

- LoLa is a powerful physics-based tagging architecture, works for tops vs QCD and q/g tagging
- **First unsupervised training of tagger for new physics:**
  Autoencoder approach
  - Works for Images and Lola
  - Can be trained using only background data
- InceptionResNetV2 can discriminate top and QCD out of the box
  - Transfer learning with less training data.