Deep Learning at the LHC: Tagging Resonances & Discovering New Physics

Gregor Kasieczka
(gregor.kasieczka@uni-hamburg.de)

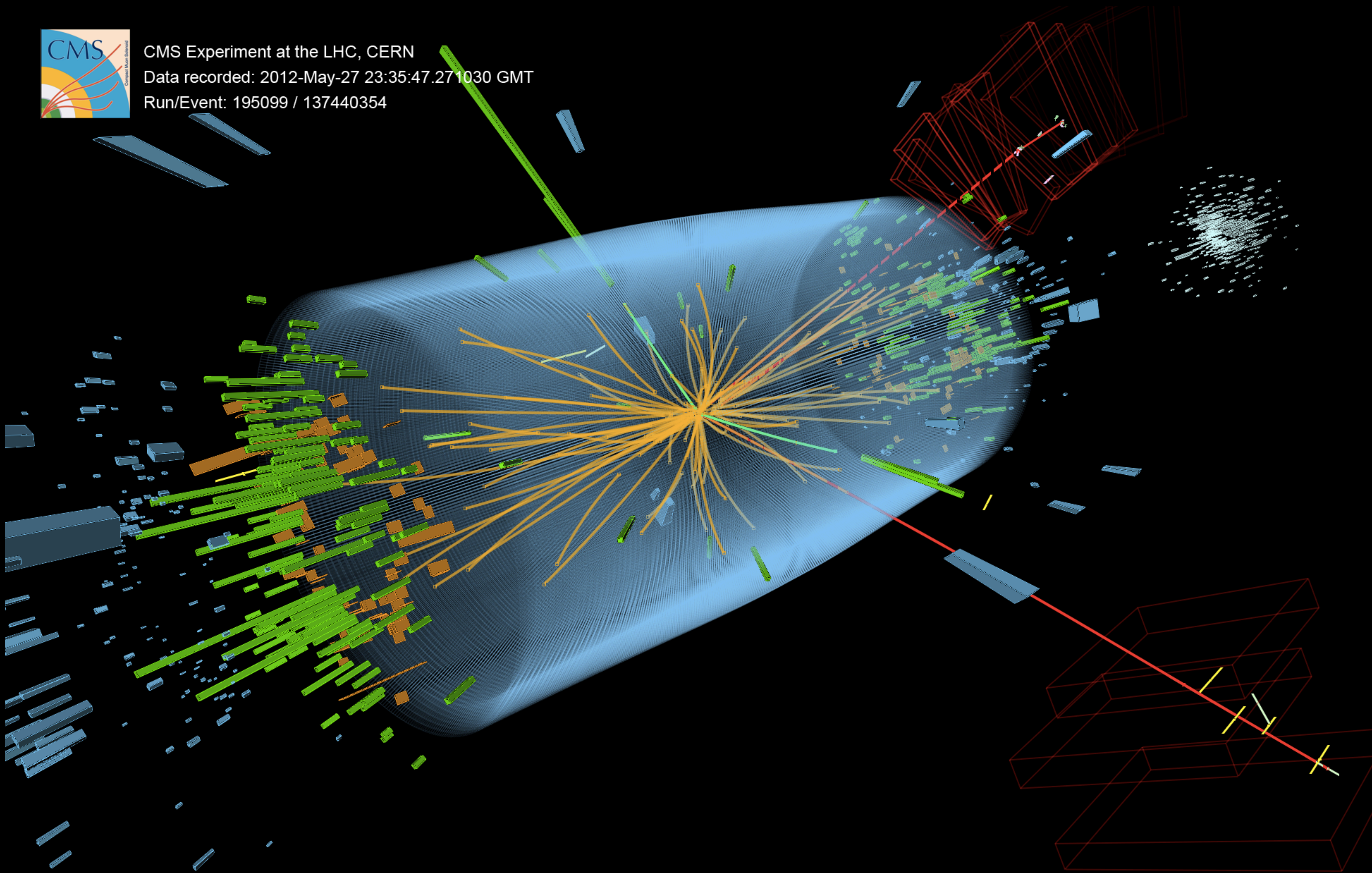
DESY Colloquium

2019-04-08

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

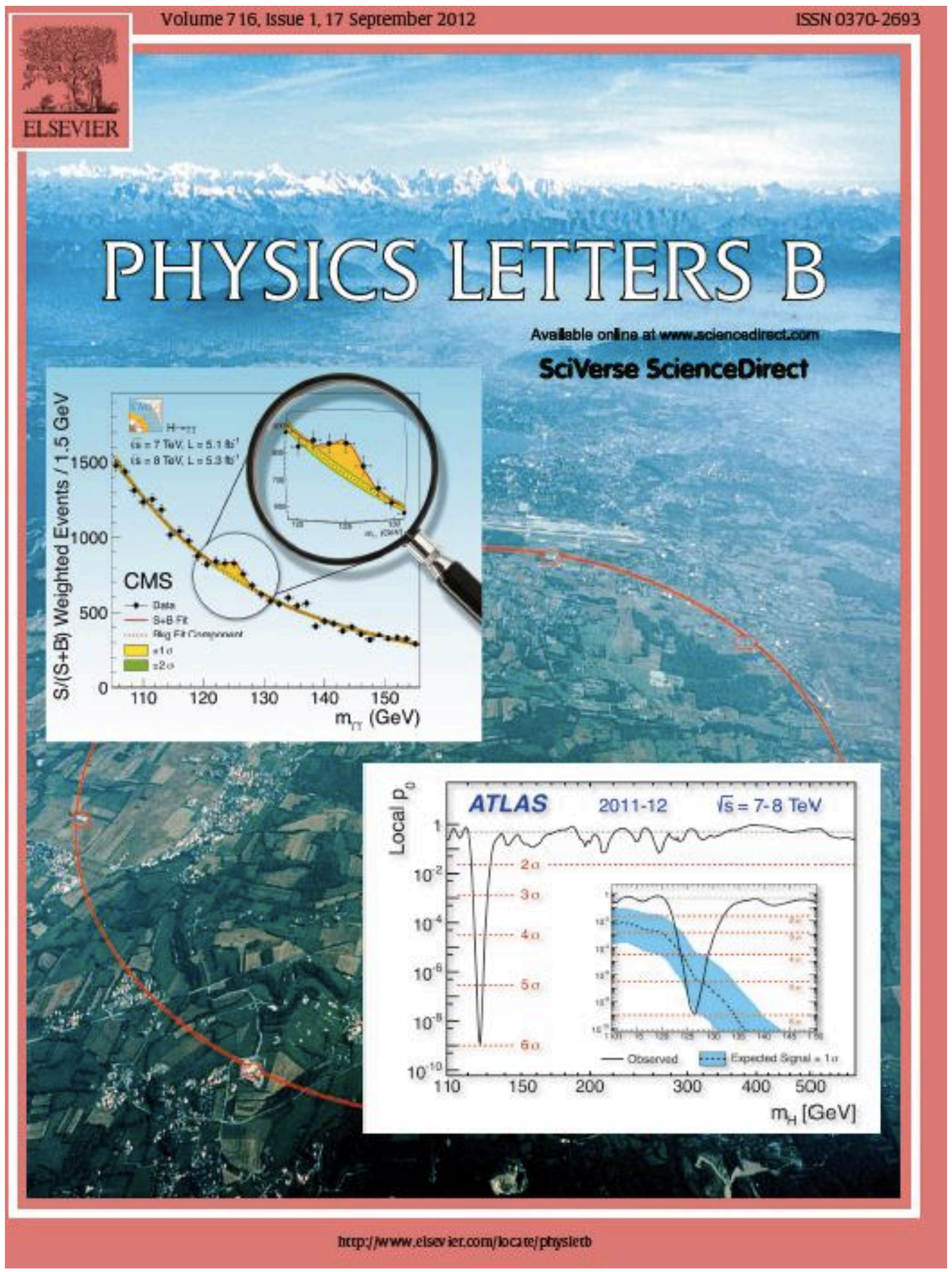
Emmy Noether-Programm
Deutsche Forschungsgemeinschaft DFG

Bundesministerium für Bildung und Forschung

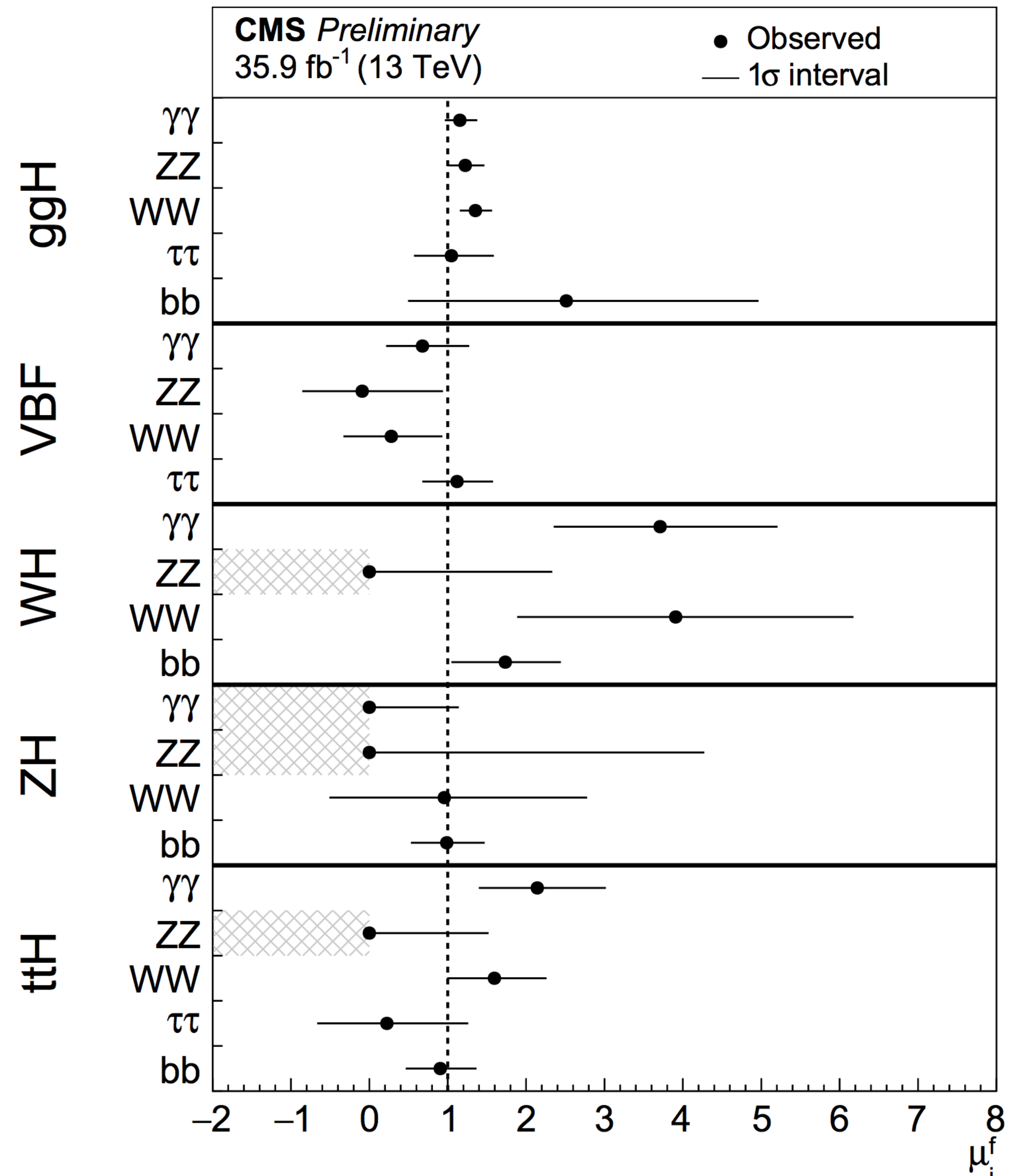


Higgs Boson Discovery in 2012

Discovery to Precision…

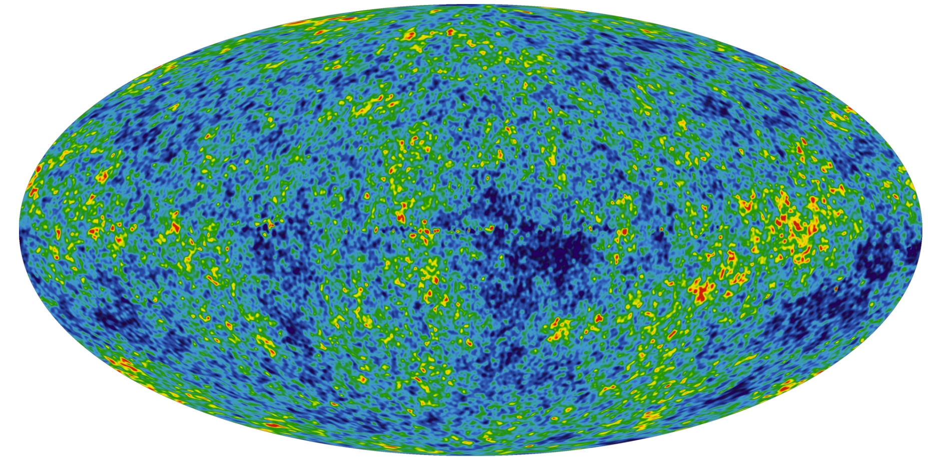


2012



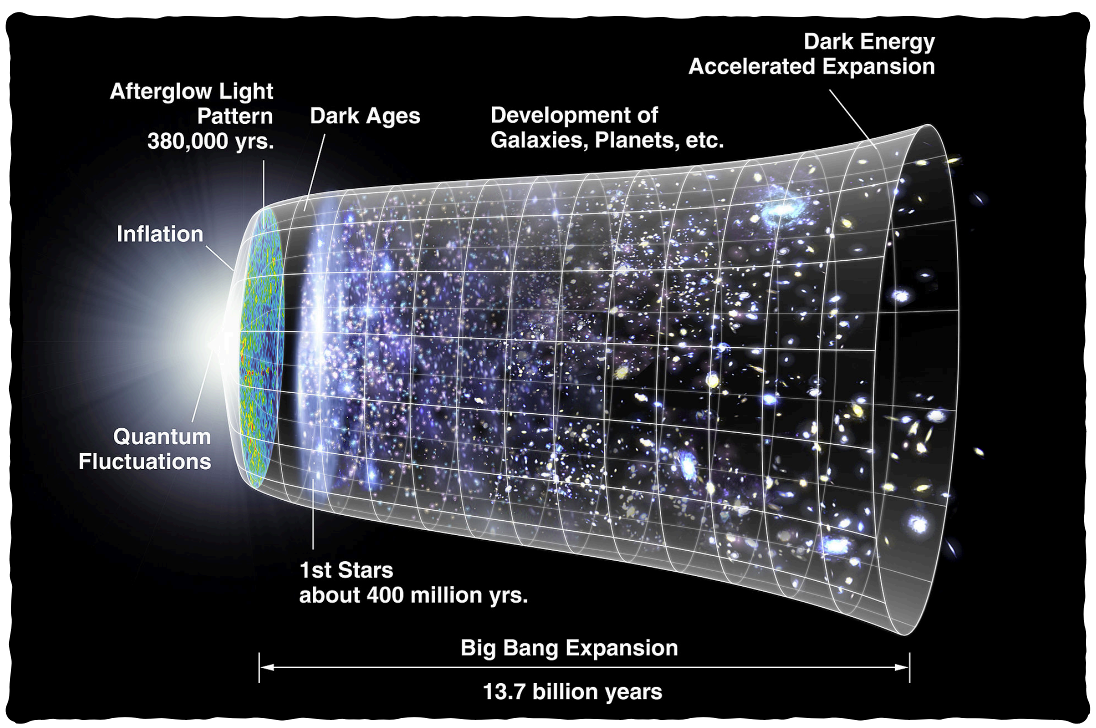
2018

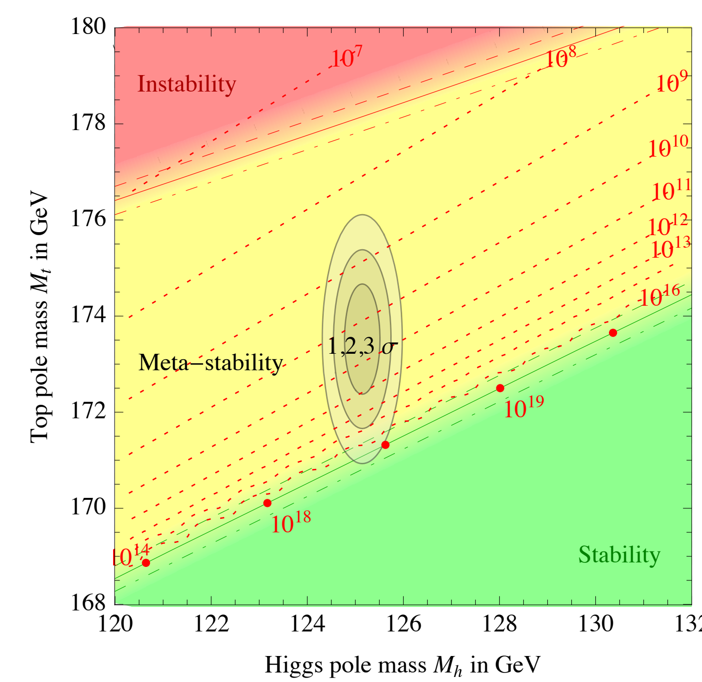
Why are neutrinos massive?



What are the origins of the LHCb flavour anomaly?

What is the nature of dark matter & dark energy?





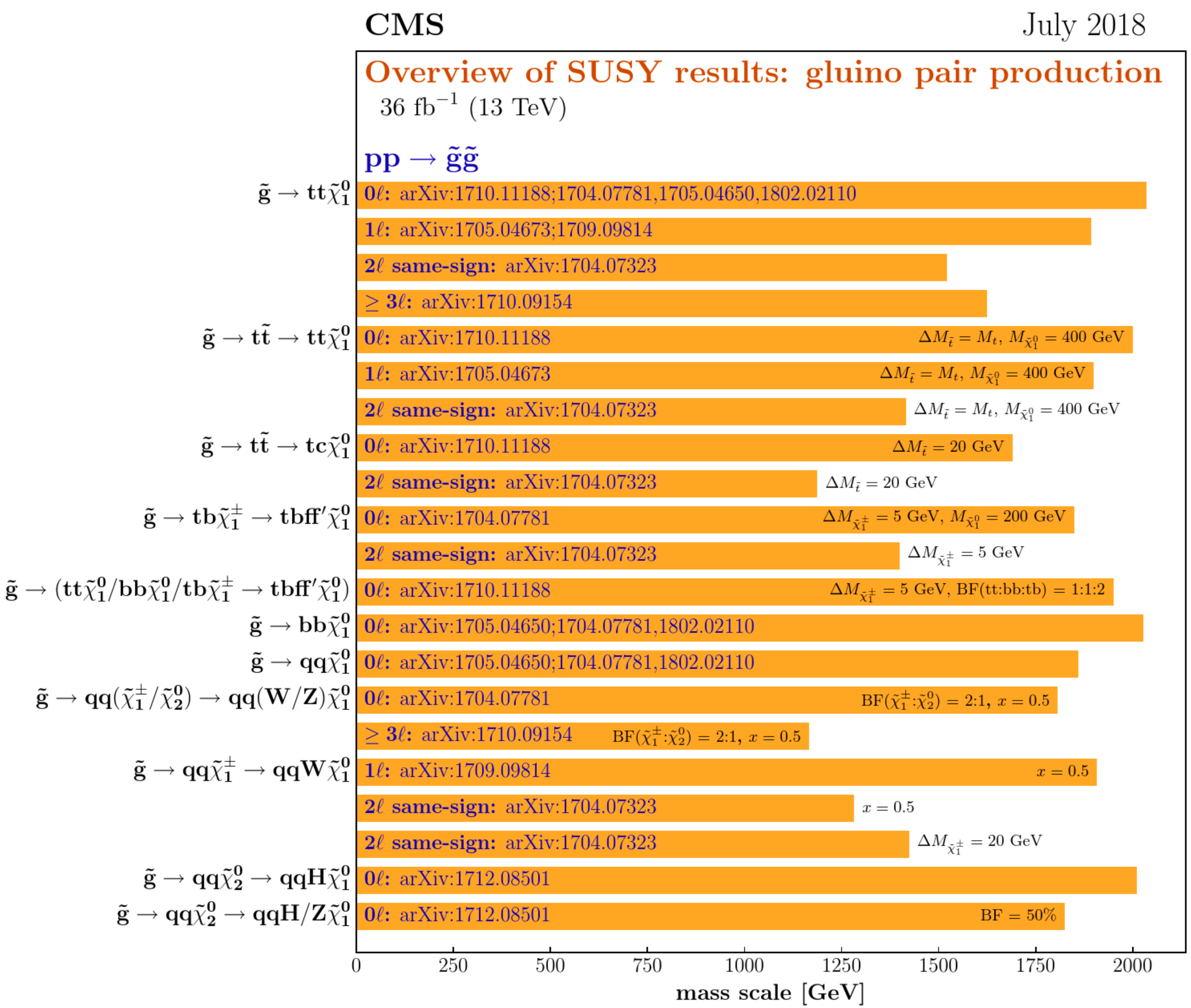
Is the electroweak vacuum stable?

Why is there more matter than anti-matter?

How can the Higgs boson be light when the mass receives large quantum corrections?

What are the details of cosmic inflation?

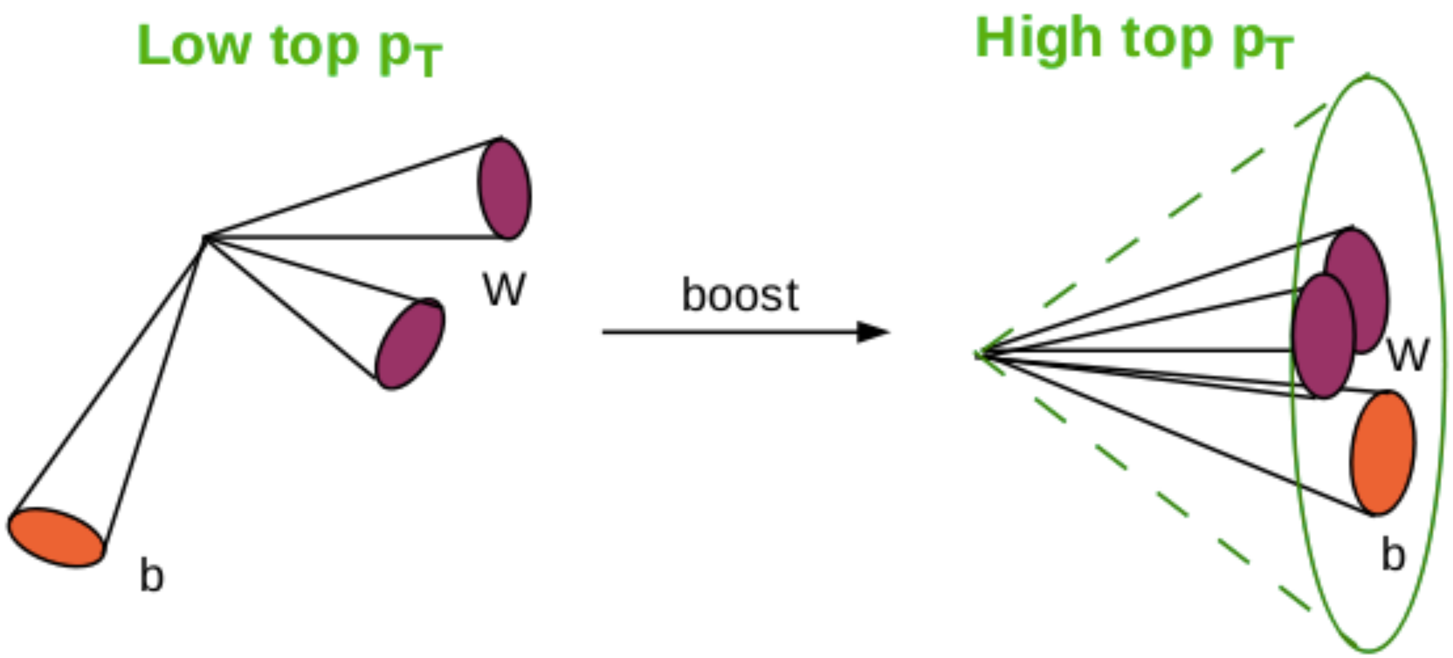
…but no new physics so far



Selection of observed limits at 95% C.L. (theory uncertainties are not included). Probe **up to** the quoted mass limit for light LSPs unless stated otherwise. The quantities ΔM and x represent the absolute mass difference between the primary particle and the LSP, and the difference between the intermediate particle and the LSP relative to ΔM, respectively, unless indicated otherwise.

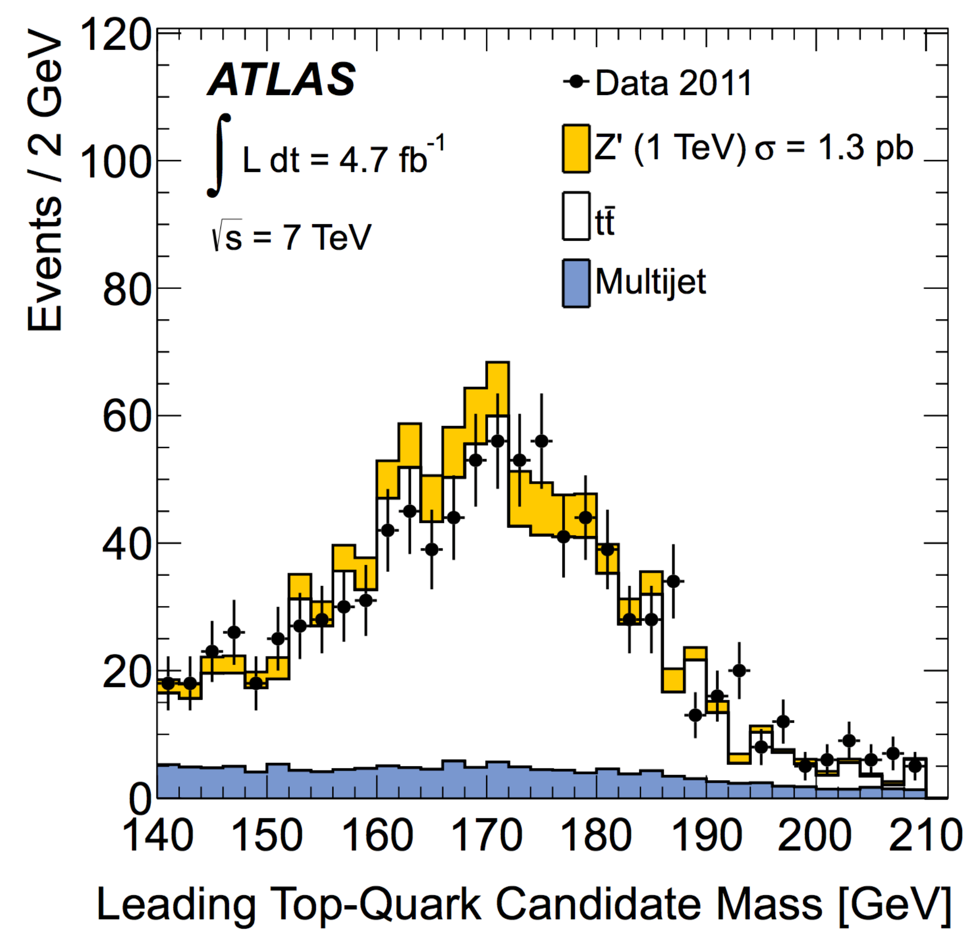
Build better tools to identify known particles

Heavy Resonance Tagging



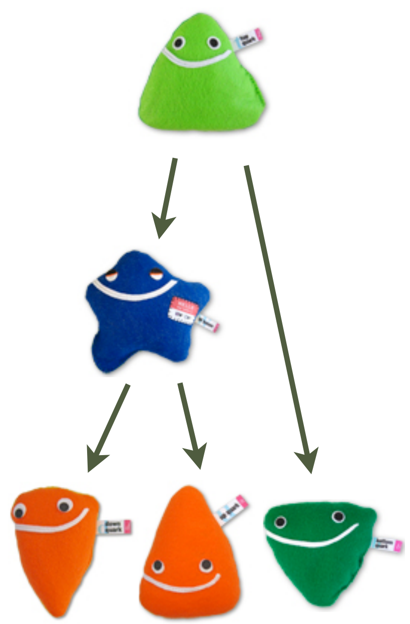
- Hadronically decaying top/Higgs/W/Z
- Contained in one (large-R) jet
- How to distinguish from light quark/gluon jets (and from each other)
- For new physics searches (and SM studies)

- Mass
  Calculate using a grooming algorithm (eg mMDT/softdrop or pruning)
- Centers of hard radiation
  n-subjettiness or energy correlation functions
- Flavour
  b tagging of large-R jets or subjets
- Combinations

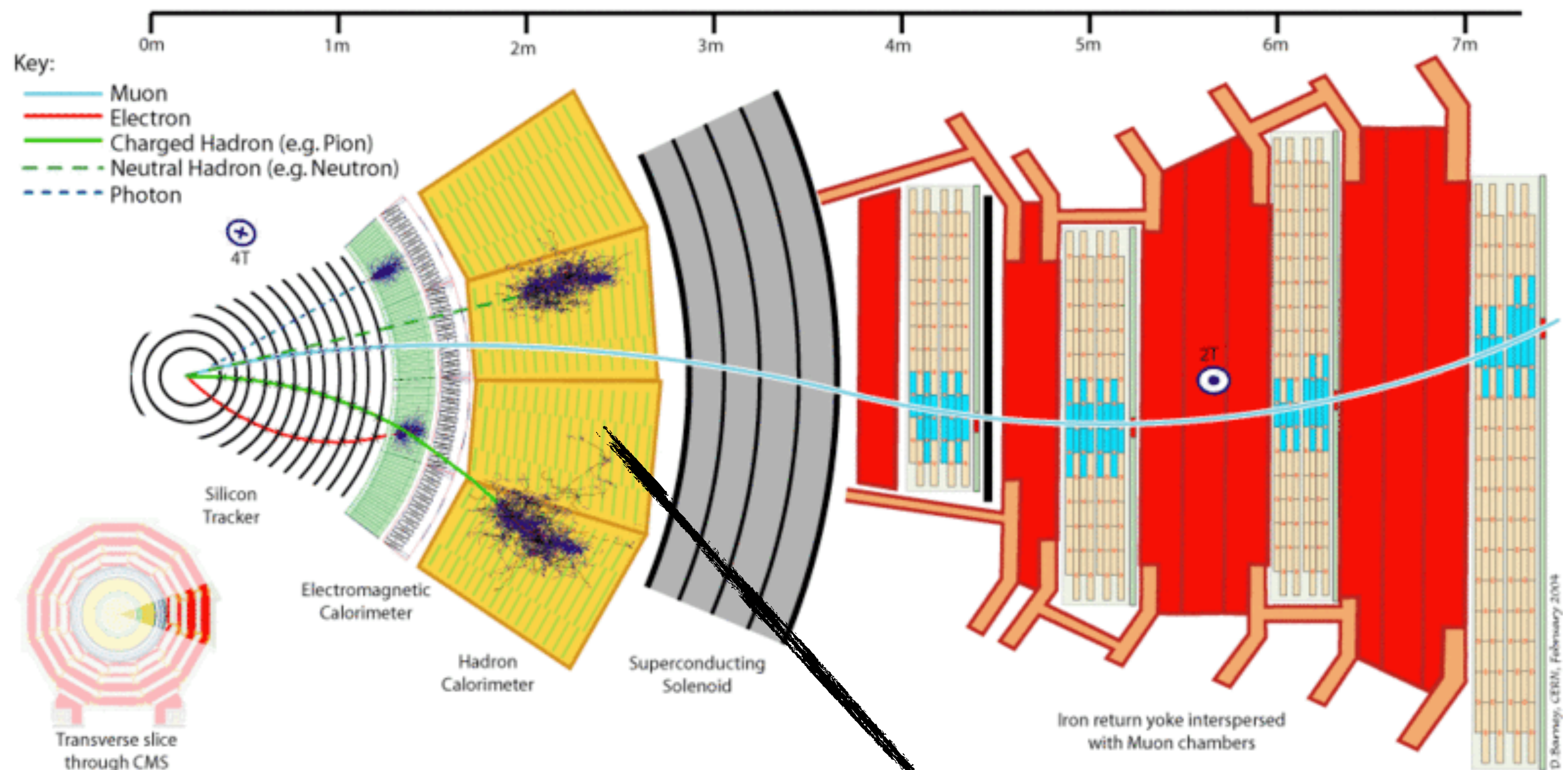


Towards an Understanding of the Correlations in Jet Substructure
D Adams et al (BOOST 2013 Participants), Eur.Phys.J. C75
Top Tagging, T Plehn, M Spannowksy, J.Phys. G39 (2012) 083001
Boosted Top Tagging Method Overview, GK, Proc.Top2017

Top Quark



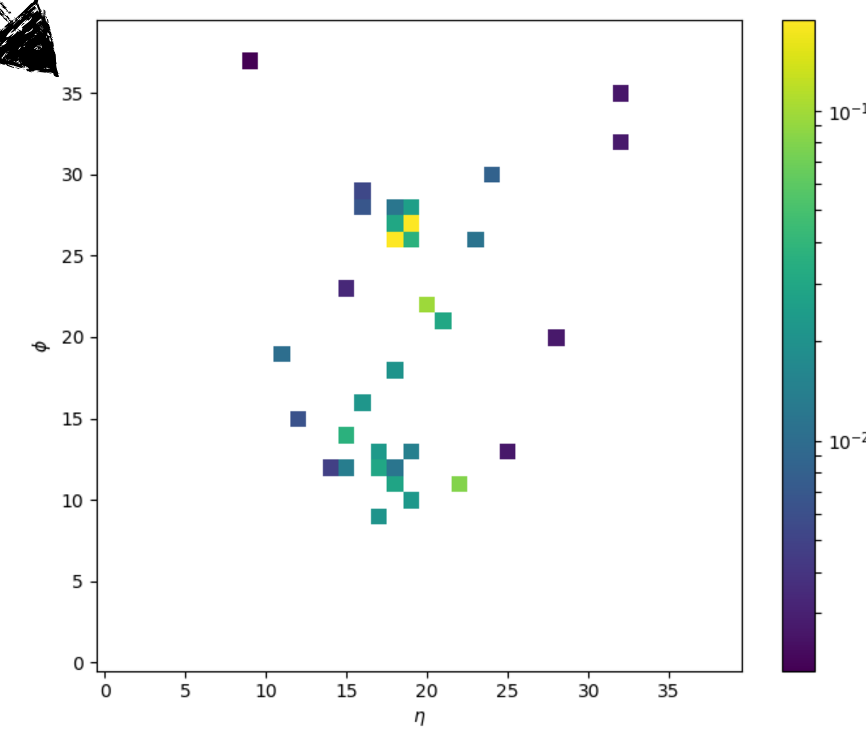
+



(jet images by C Daza)

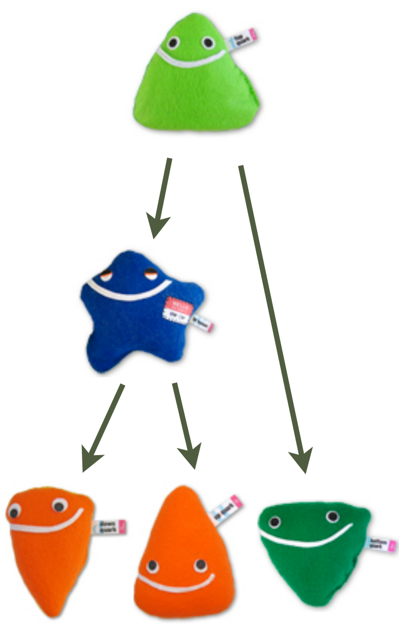
- *Measure particle energies in calorimeter*
- *Reconstruct jet from individual measurements*
- *Image preprocessing*
  - *center, rotate, mirror, pixelate, trim, normalise*

=

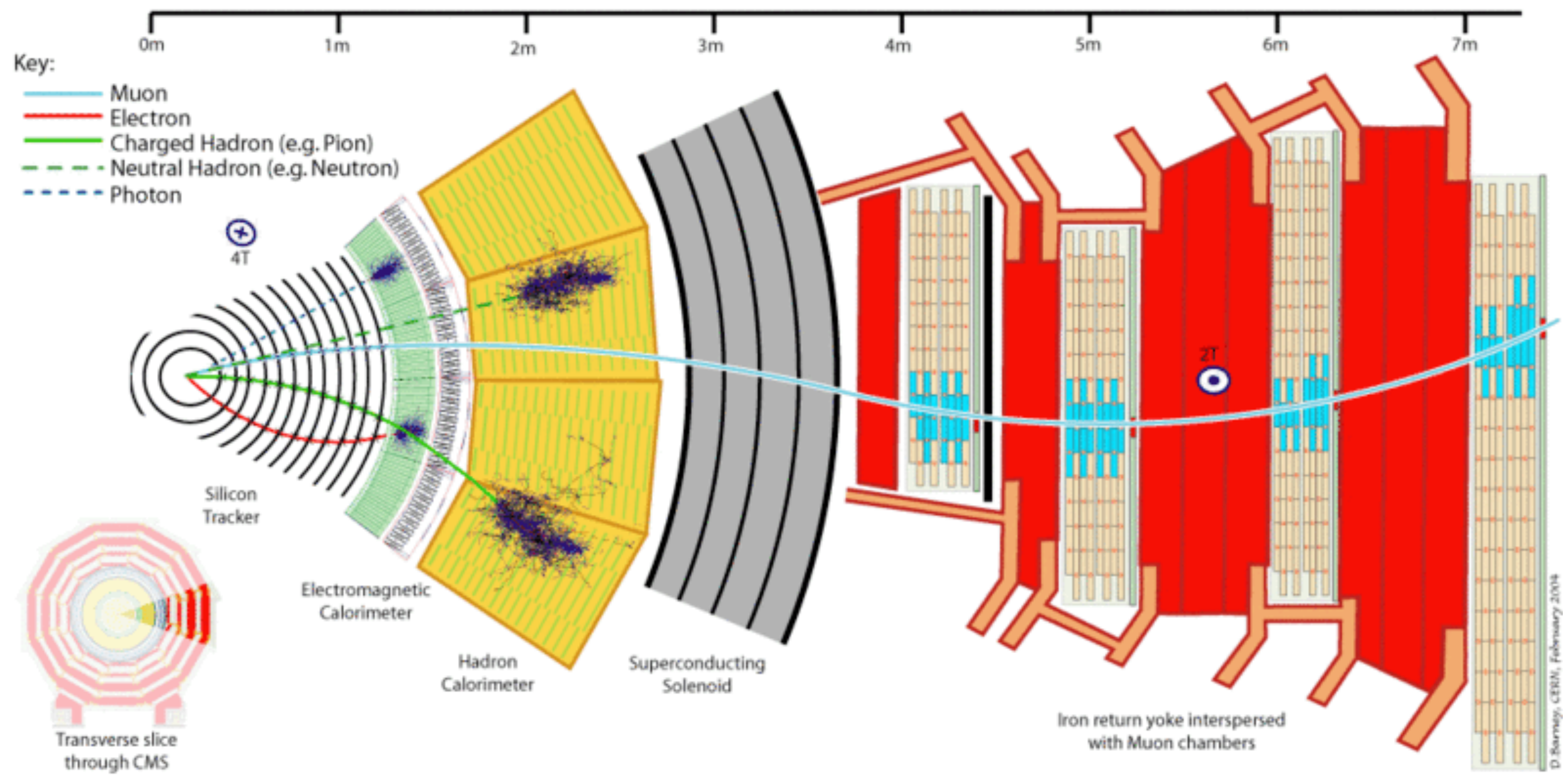


1 image

Top Quark



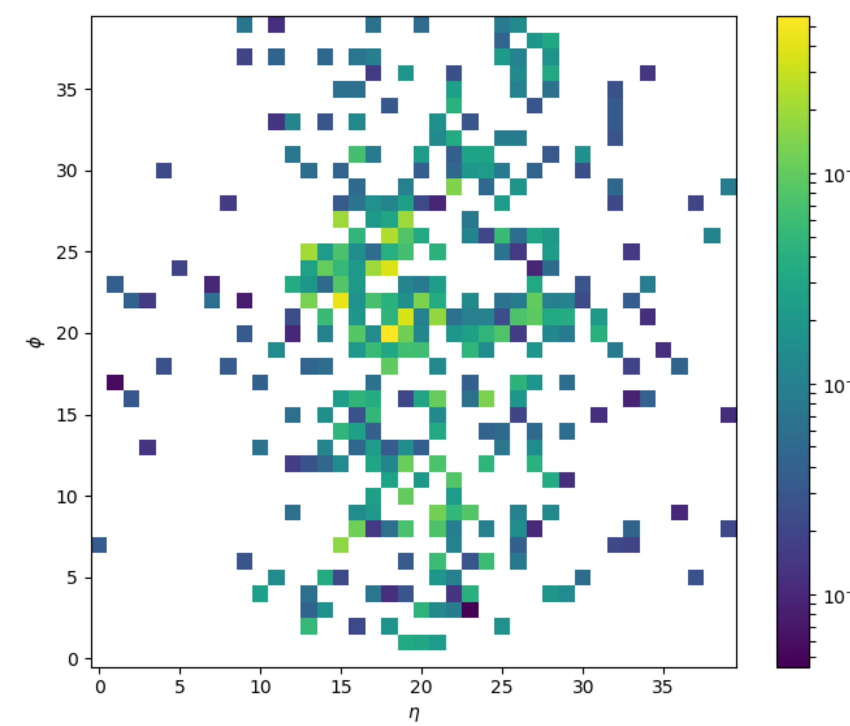
+



(jet images by C Daza)

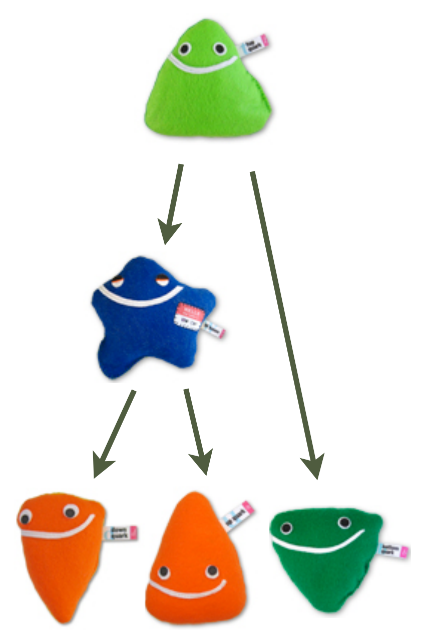
- *Measure particle energies in calorimeter*
- *Reconstruct jet from individual measurements*
- *Image preprocessing*
  - *center, rotate, mirror, pixelate, trim, normalise*

=

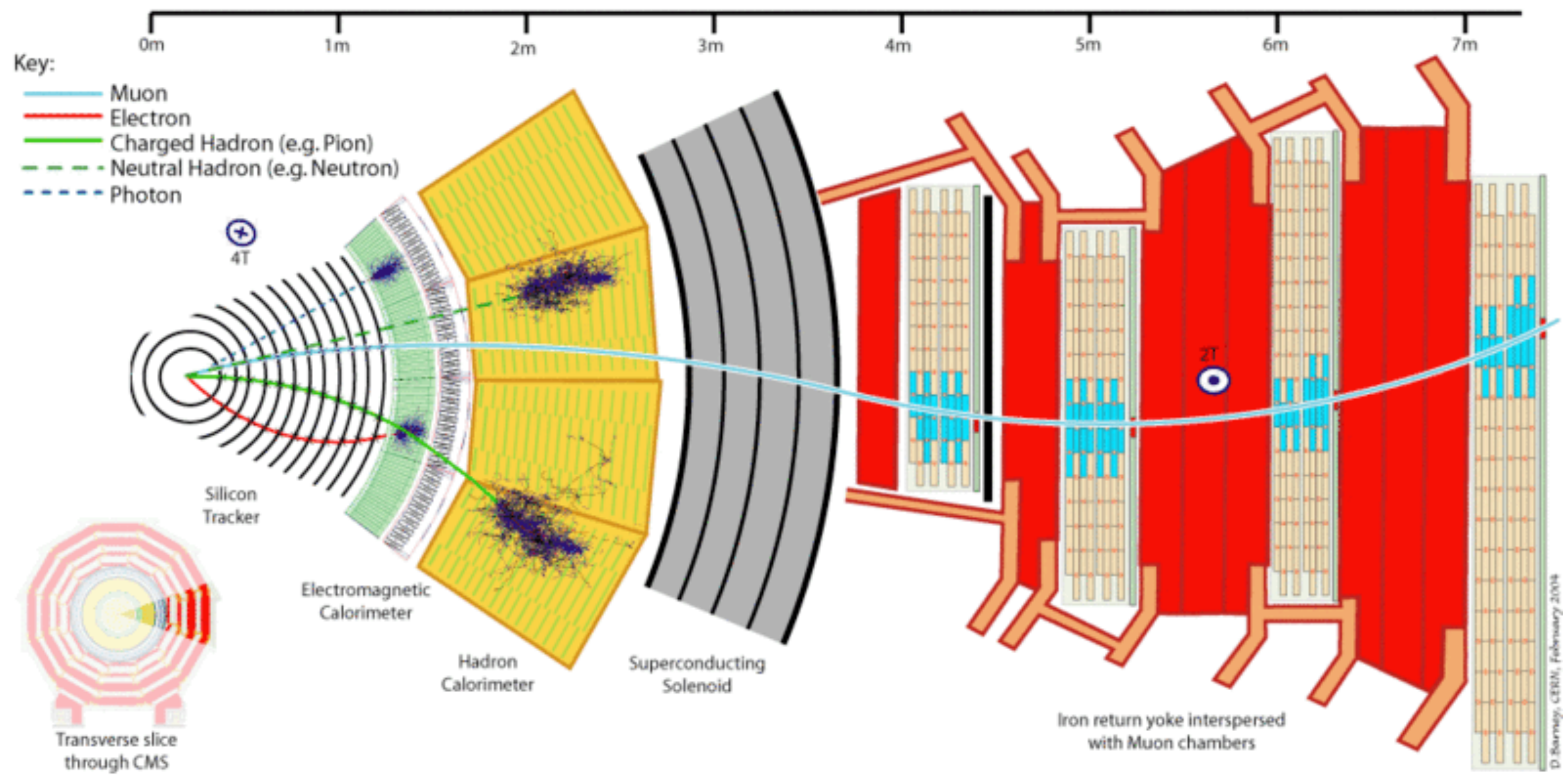


10 image average

Top Quark



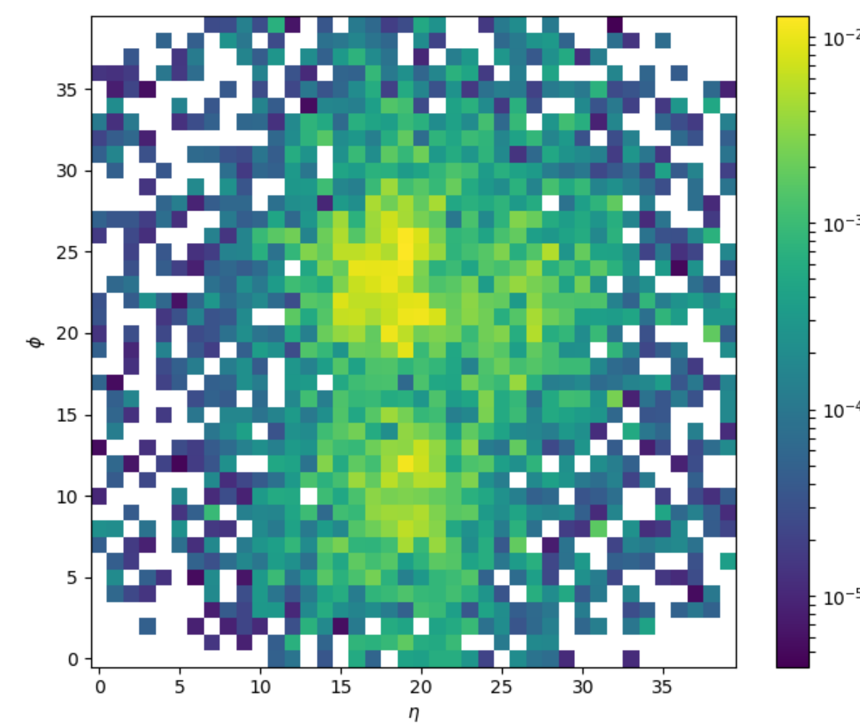
+



- *Measure particle energies in calorimeter*
- *Reconstruct jet from individual measurements*
- *Image preprocessing*
    - *center, rotate, mirror, pixelate, trim, normalise*

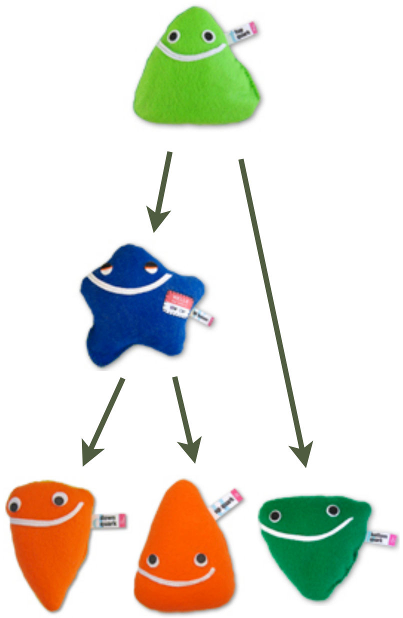
=

(jet images by C Daza)

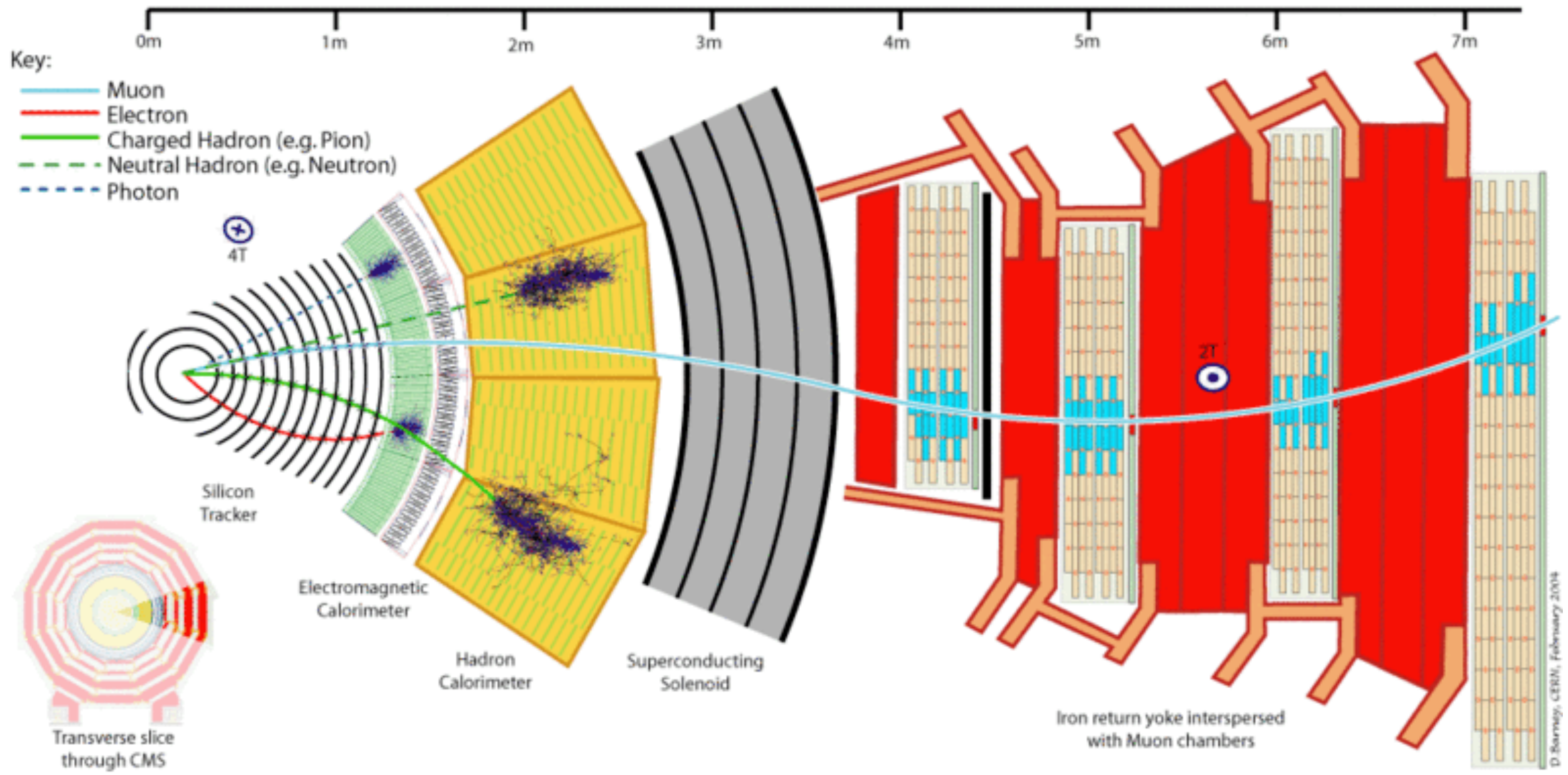


100 image average

Top Quark



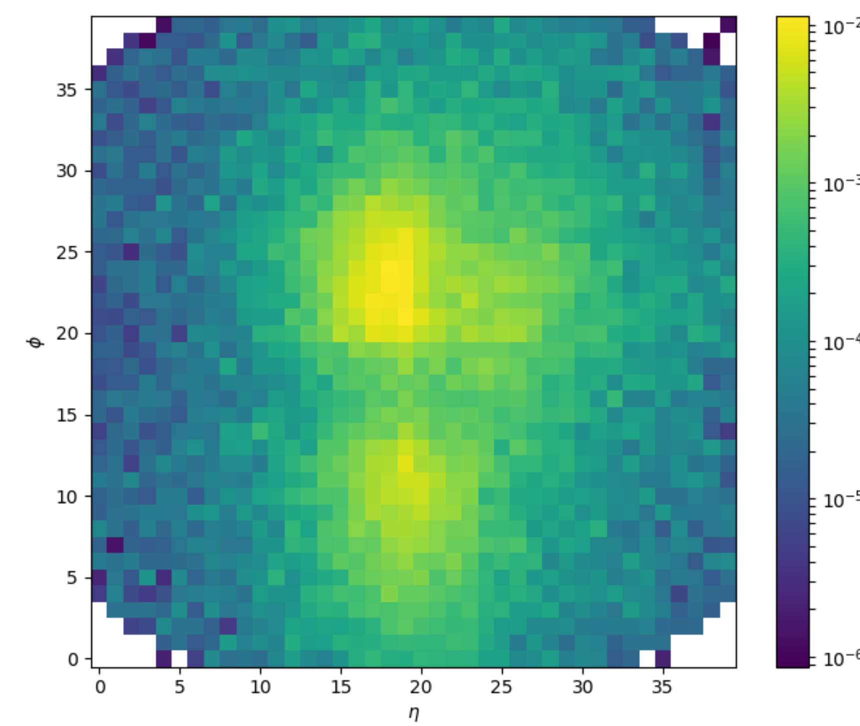
\+



(jet images by C Daza)

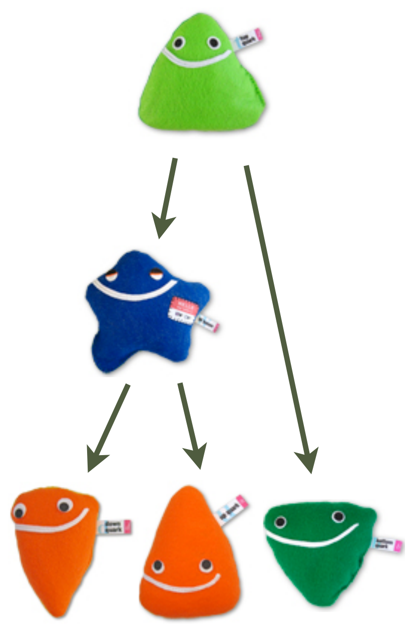
- *Measure particle energies in calorimeter*
- *Reconstruct jet from individual measurements*
- *Image preprocessing*
  - *center, rotate, mirror, pixelate, trim, normalise*

=

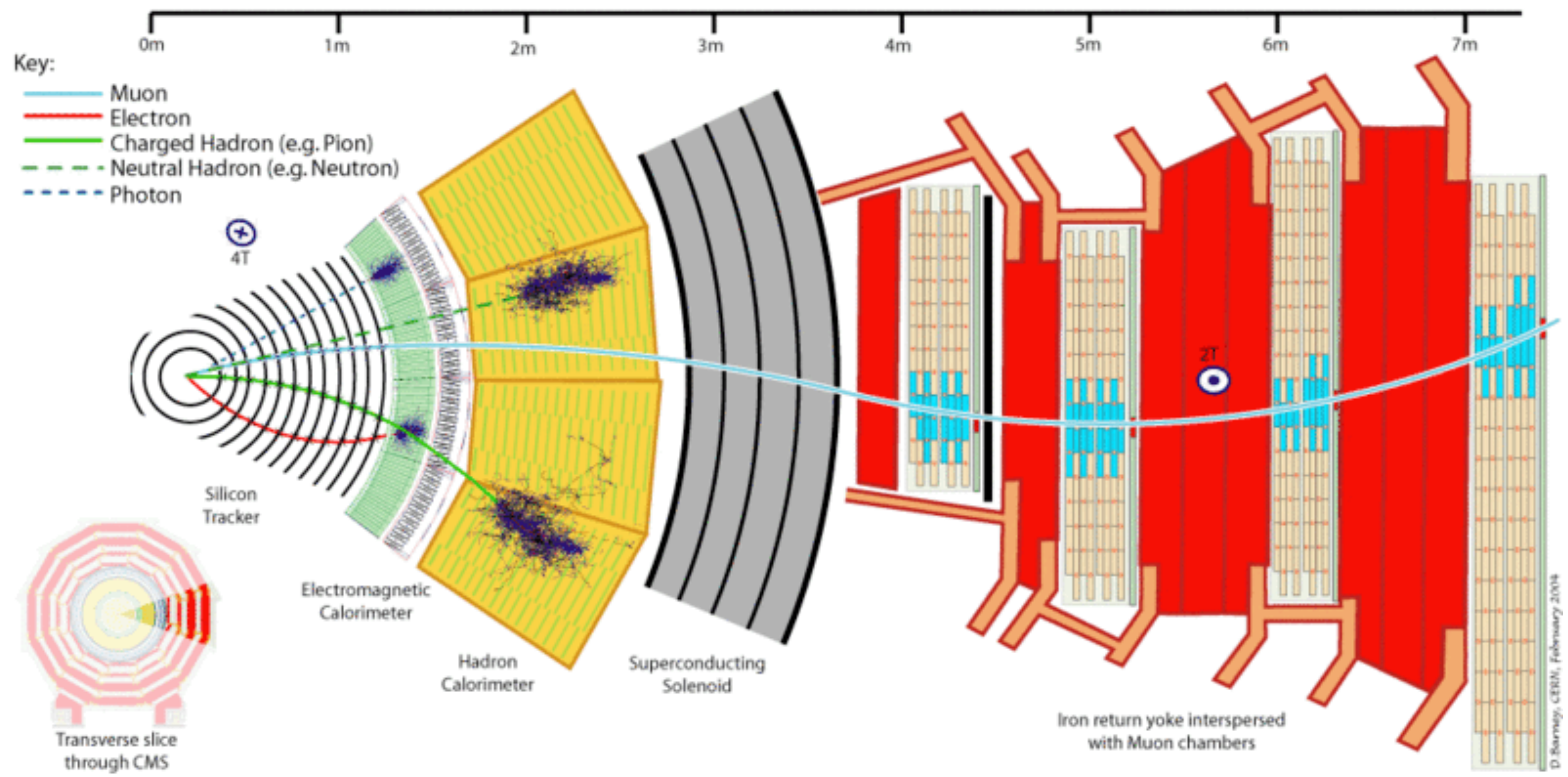


1000 image average

Top Quark



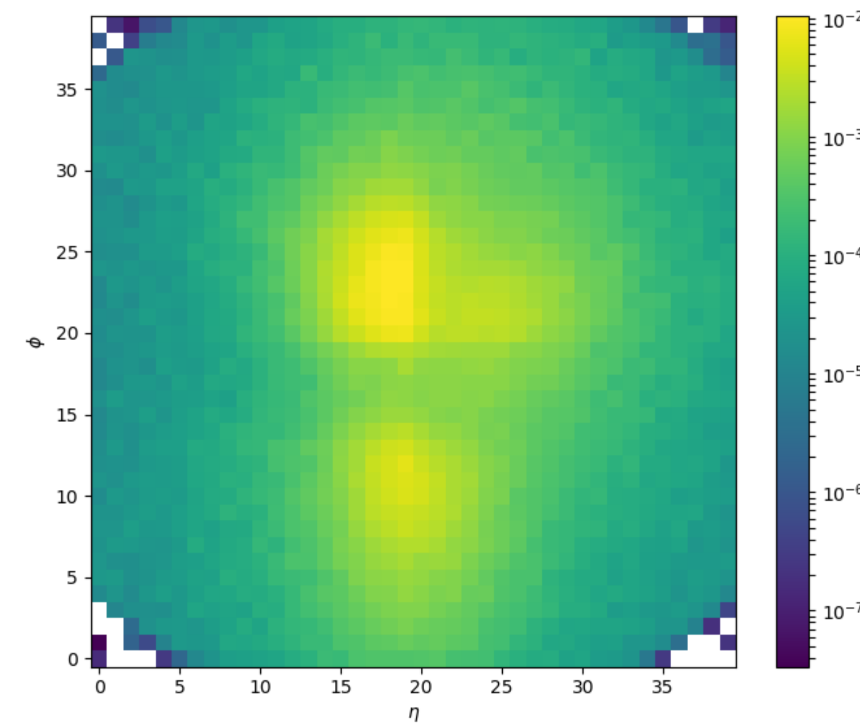
+



- *Measure particle energies in calorimeter*
- *Reconstruct jet from individual measurements*
- *Image preprocessing*
  - *center, rotate, mirror, pixelate, trim, normalise*

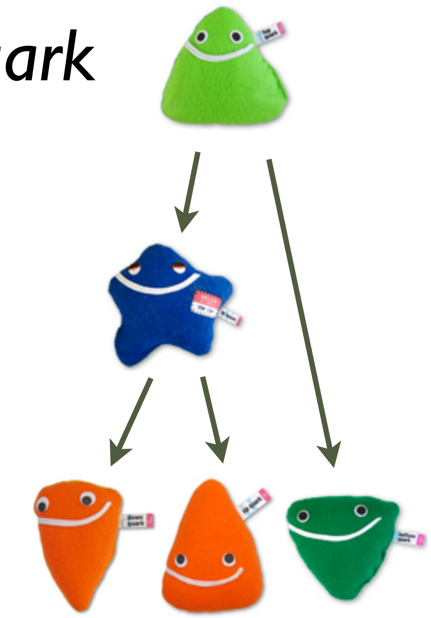
=

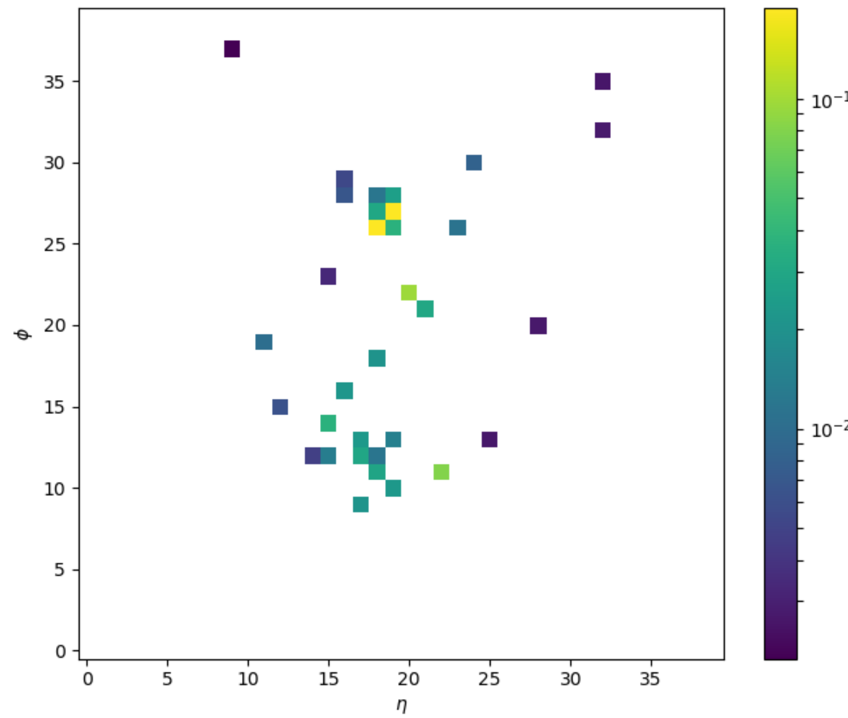
(jet images by C Daza)

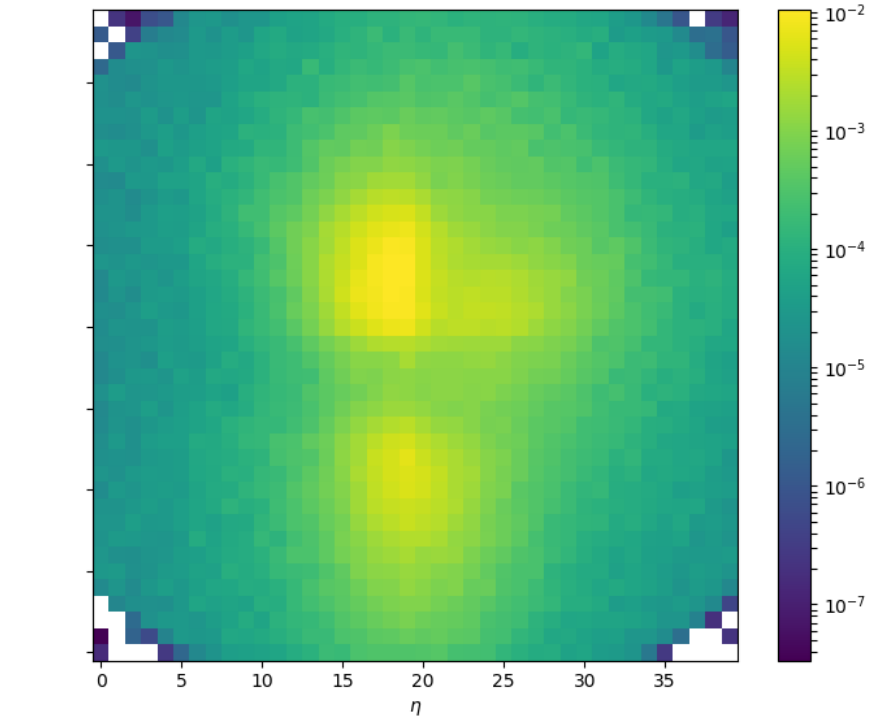


10000 image average

Top Quark Jet

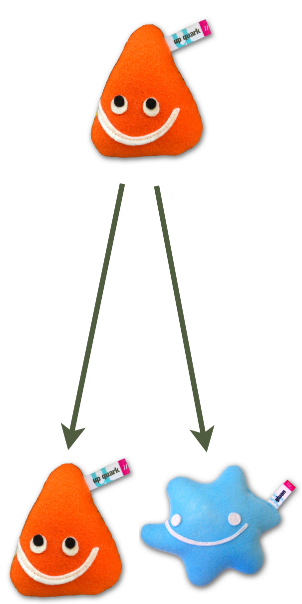


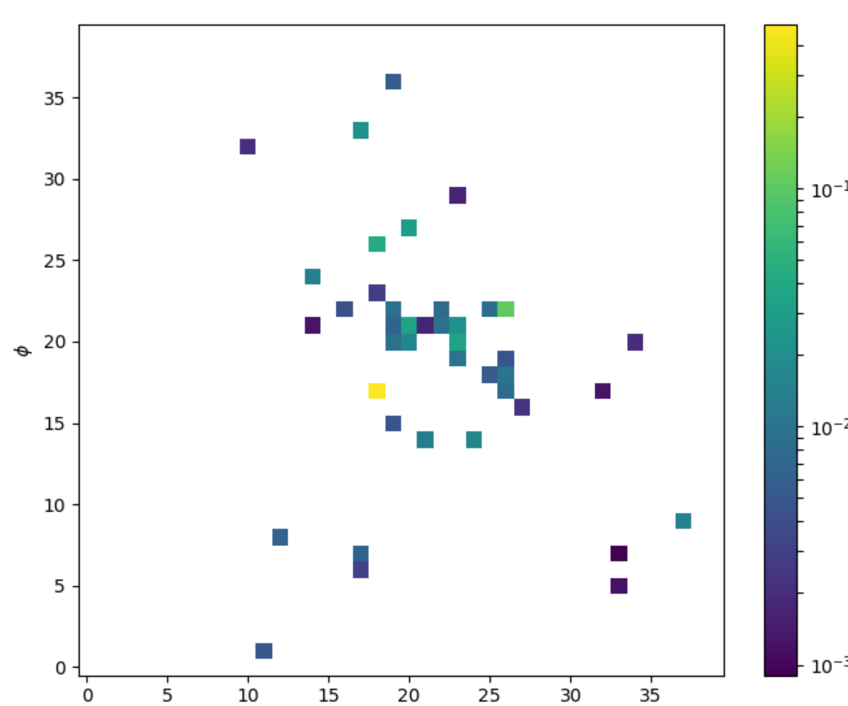


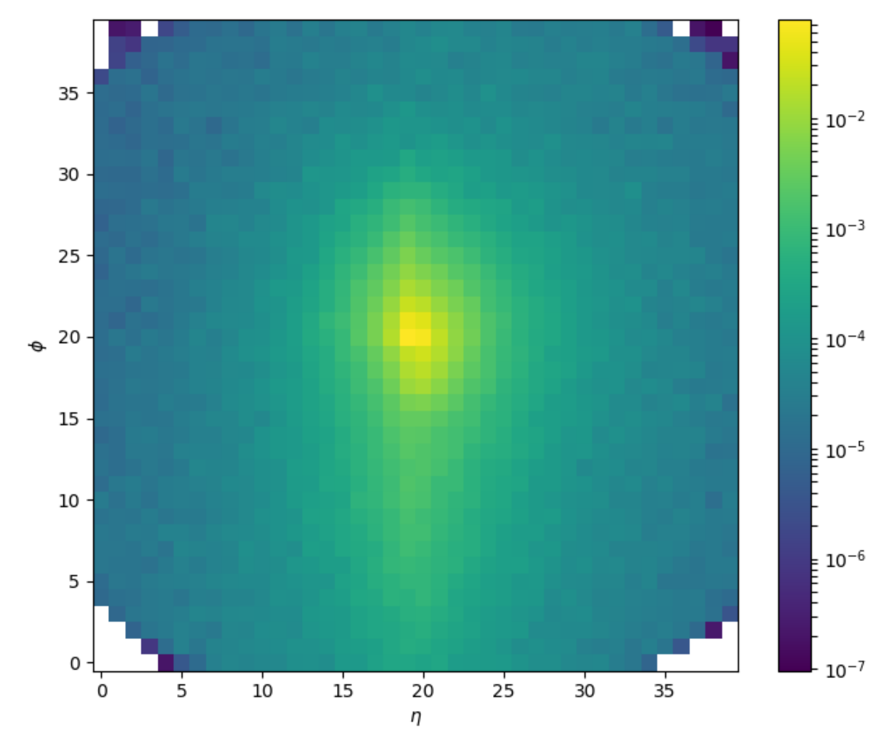


- *Binary classification task*
- *Fully supervised learning (using simulation)*
- *40x40 Pixels, E_T*
- *Perfectly suited for deep learning algorithms*

QCD Jet



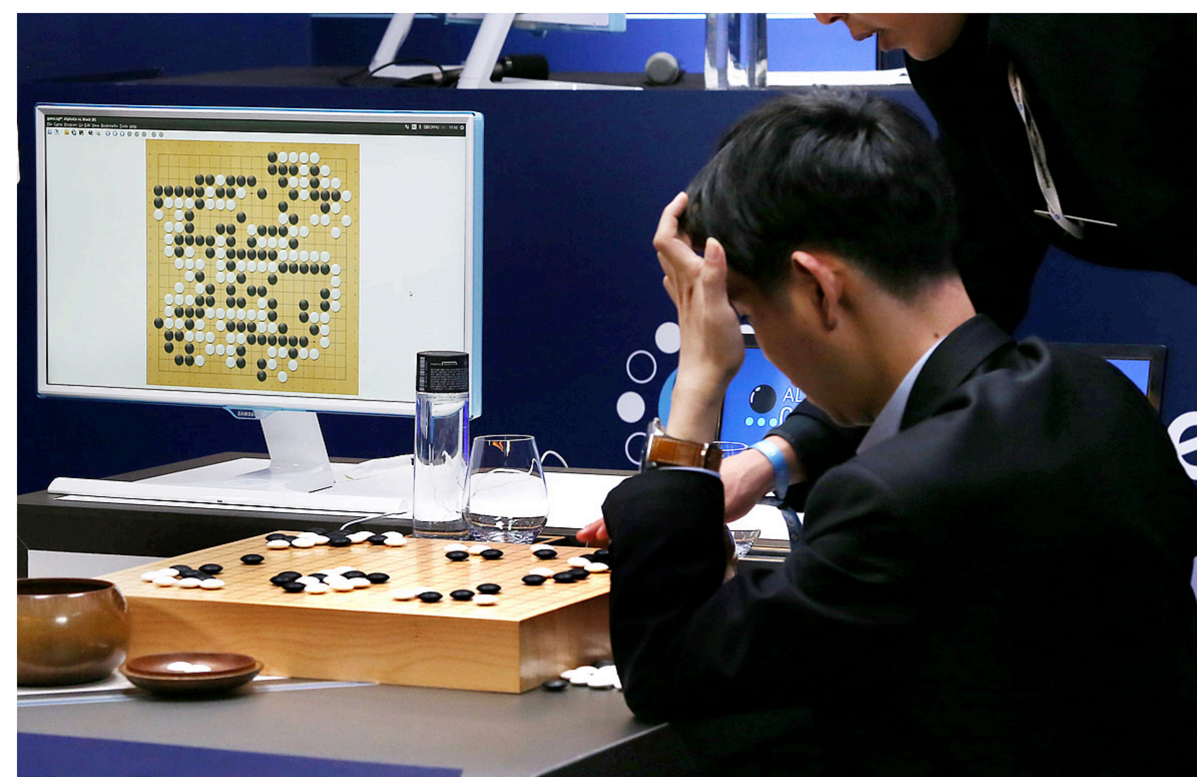




Deep Learning

Humans vs Machines

- 2015 **Image Classification**:
  - K. He et al (Microsoft Research), *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification,* 1502.01852
- 2016 **Go**:
  - Alpha Go (D. Silver et al, *Mastering the game of Go with deep neural networks and tree search*, Nature 529, pp484–489 and D. Silver et al *Mastering the game of Go without human knowledge,* Nature 550, pp354–359)
- 2016 **Speech recognition:**
  - W. Xiong et al (Microsoft Research) *Achieving Human Parity in Conversational Speech Recognition,* 1610.05256
- 2017 **Poker** (heads-up no-limits Texas Hold'em):
  - N Brown and T Sandholm, *Superhuman AI for heads-up no-limit poker: Libratus beats top professionals, Science 359, Issue 6374, pp418-424*
- 2018 ***Translation*** (Chinese-English)
  - H H Awadalla et al (Microsoft AI & Research) *Achieving Human Parity on Automatic Chinese to English News Translation*
- 20?? **Particle Physics**
  - *to be seen*



ImageNet

GT: horse cart
1: horse cart
2: minibus
3: oxcart
4: stretcher
5: half track

GT: birdhouse
1: birdhouse
2: sliding door
3: window screen
4: mailbox
5: pot

GT: forklift
1: forklift
2: garbage truck
3: tow truck
4: trailer truck
5: go-kart

GT: coucal
1: coucal
2: indigo bunting
3: lorikeet
4: walking stick
5: custard apple

GT: komondor
1: komondor
2: patio
3: llama
4: mobile home
5: Old English sheepdog

GT: yellow lady's slipper
1: yellow lady's slipper
2: slug
3: hen-of-the-woods
4: stinkhorn
5: coral fungus

GT: torch
1: stage
2: spotlight
3: torch
4: microphone
5: feather boa

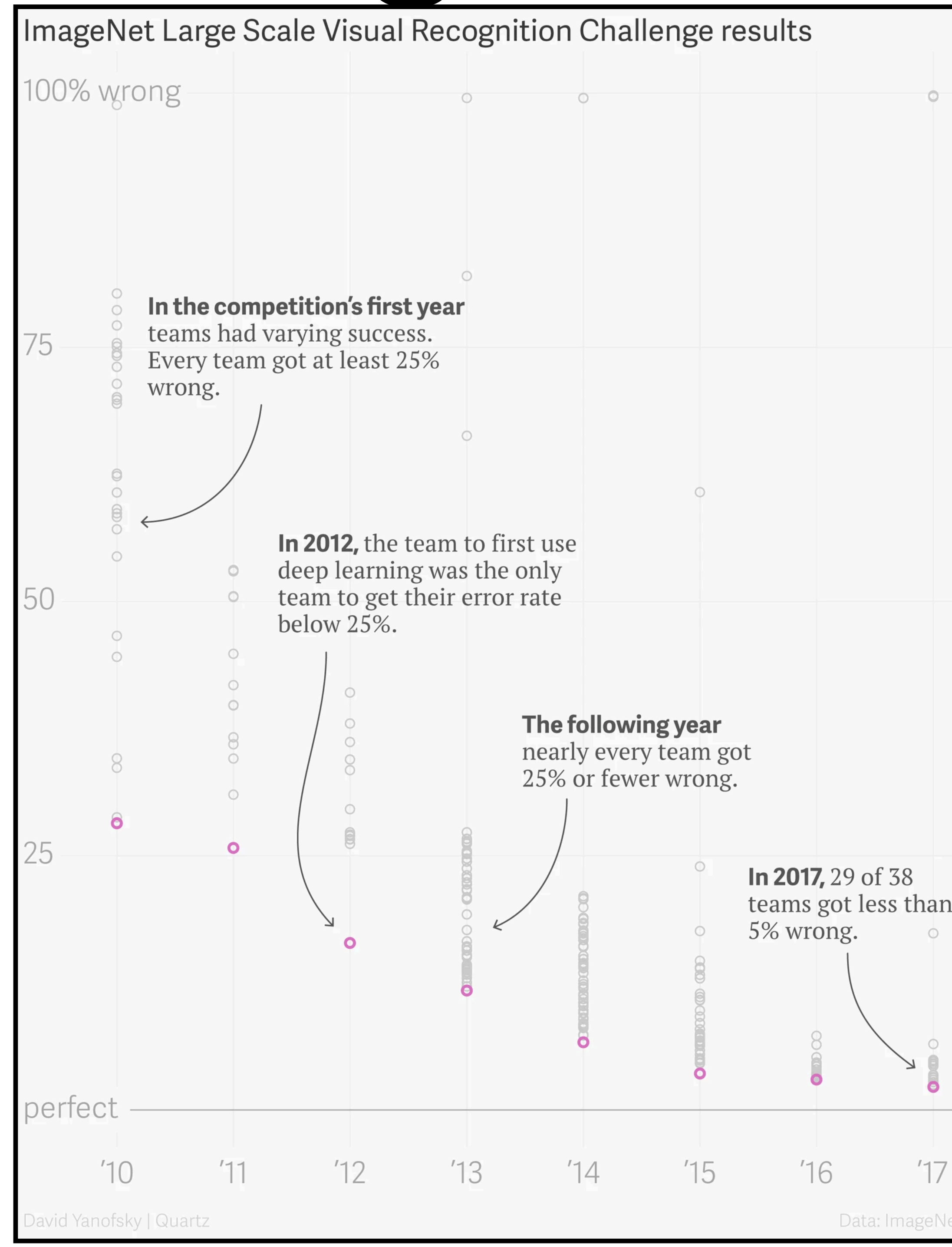
GT: banjo
1: acoustic guitar
2: shoji
3: bow tie
4: cowboy hat
5: banjo

GT: go-kart
1: go-kart
2: crash helmet
3: racer
4: sports car
5: motor scooter

GT: mountain tent
1: sleeping bag
2: mountain tent
3: parachute
4: ski
5: flagpole

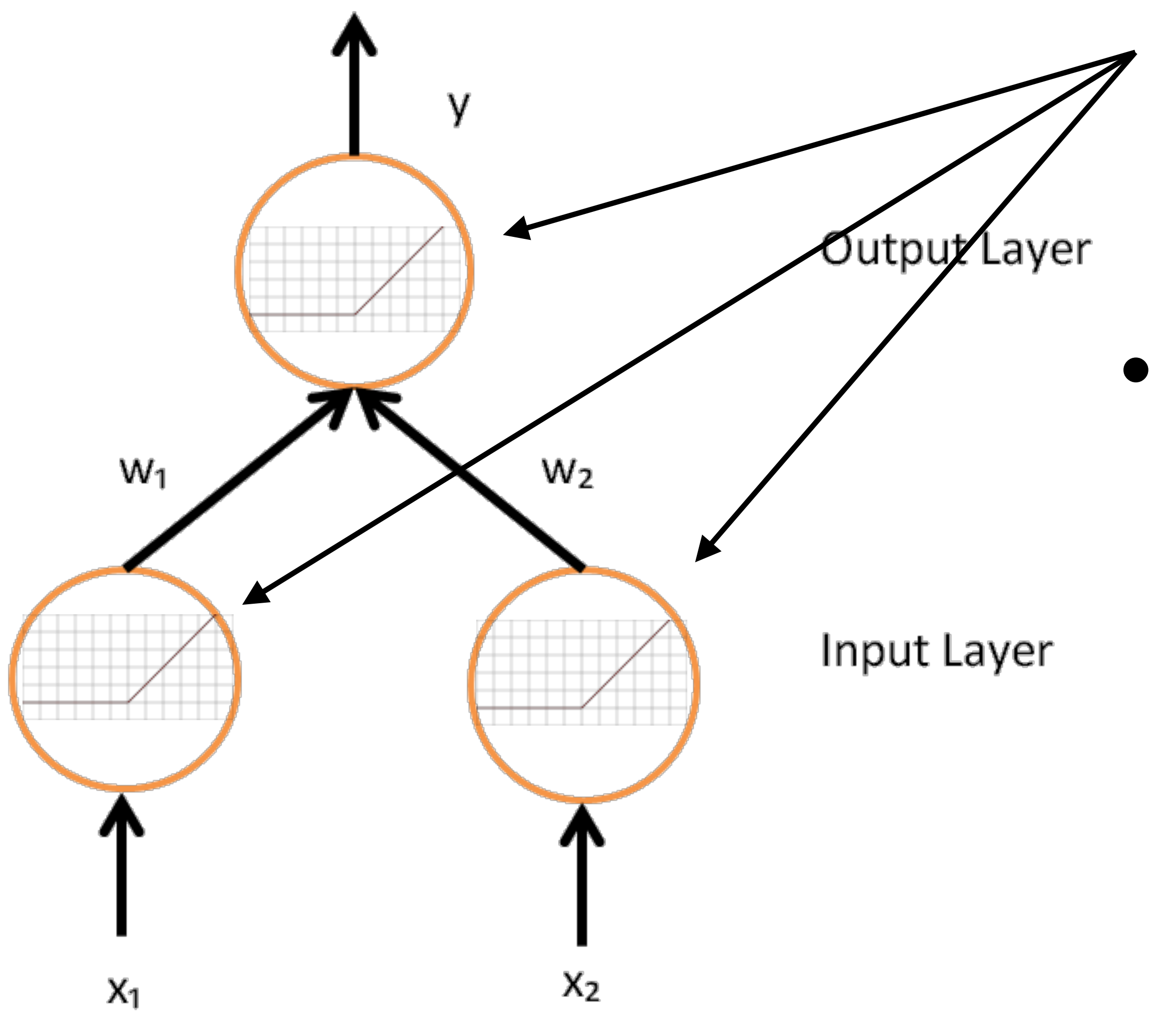
GT: geyser
1: geyser
2: volcano
3: sandbar
4: breakwater
5: leatherback turtle

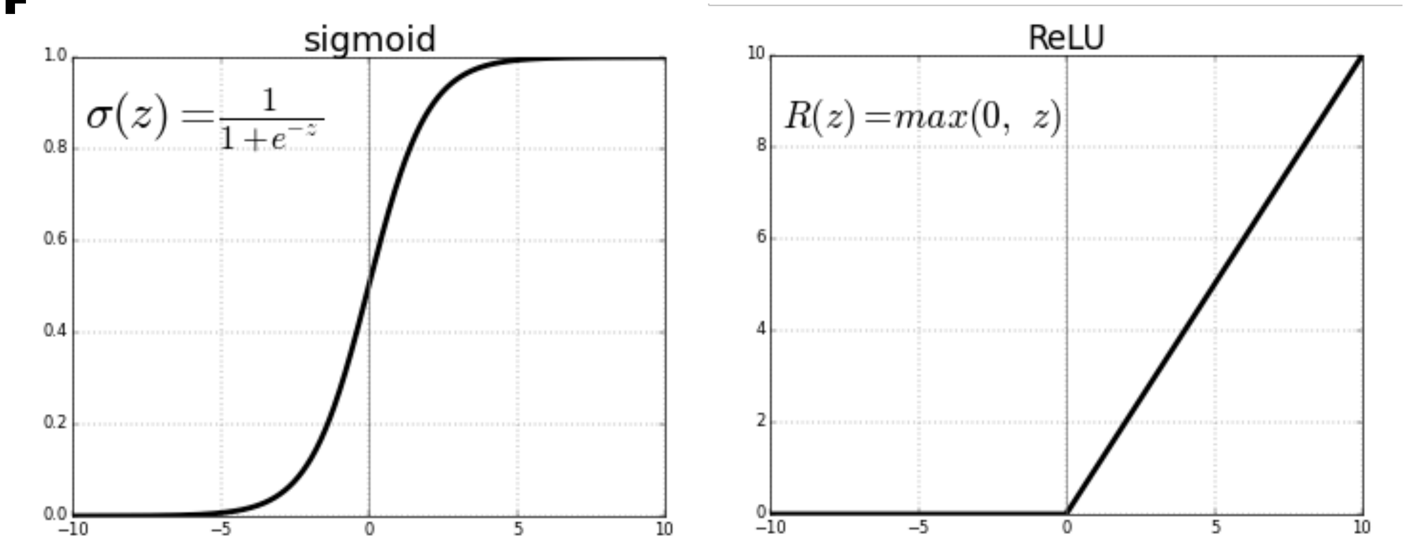
GT: microwave
1: microwave
2: washer
3: toaster
4: stove
5: dishwasher



- 14M labelled images
- 20k categories
- http://image-net.org

A Very Simple Network





towardsdatascience.com

- *Backpropagation + Gradient descent*
  - Pass input (x_1, x_2) to ANN
  - Calculate output (y) and difference to true value ($\hat{y}$)
    This is the loss function L
  - Find gradient of loss function with respect to weights
  - Use gradient to find new weights

$$y = f(f(x_1)w_1 + f(x_2)w_2)$$

$$f(x) = \Theta(x) \cdot x$$

Regression loss function:

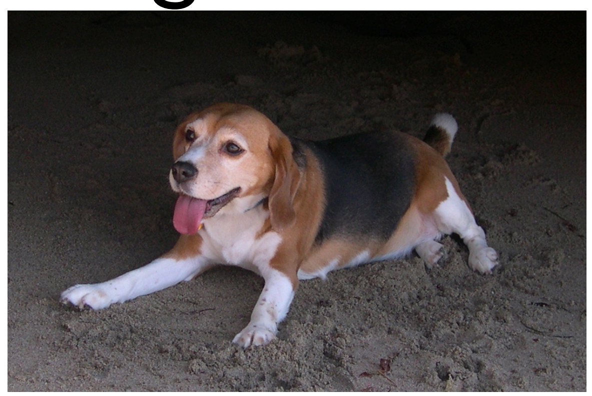
$$L(y, \hat{y}) = (y - \hat{y})^2$$

$$w_i' = w_i + \alpha \cdot \frac{\partial L}{\partial w_i}$$

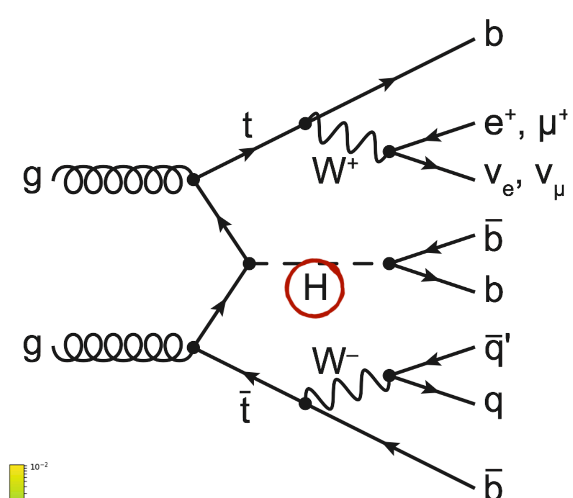
Problem: Classification

Distinguish a pair of classes (binary) or several (multi-class).

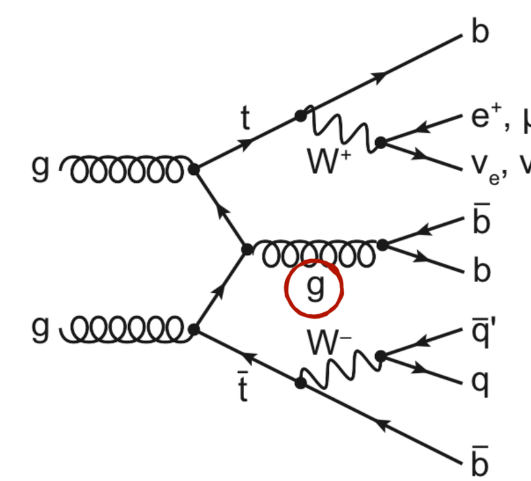
Cat vs Dog



Event Topologies



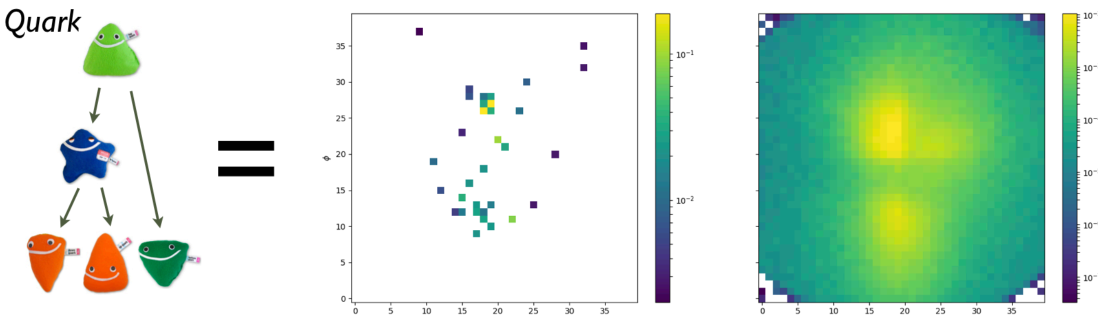
signal



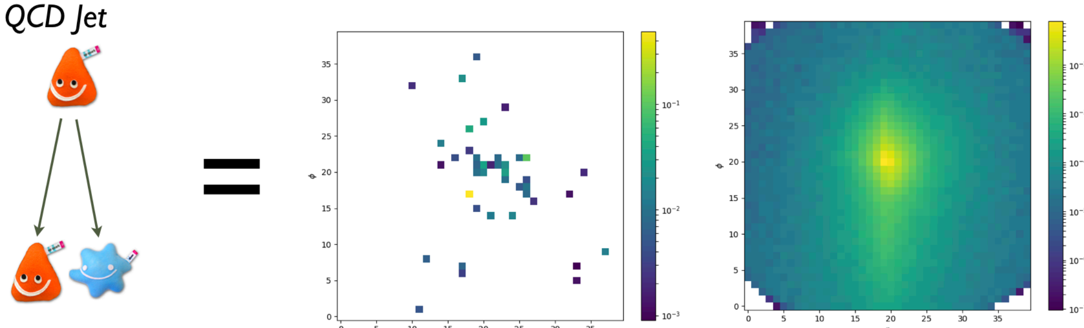
difficult background

Particles

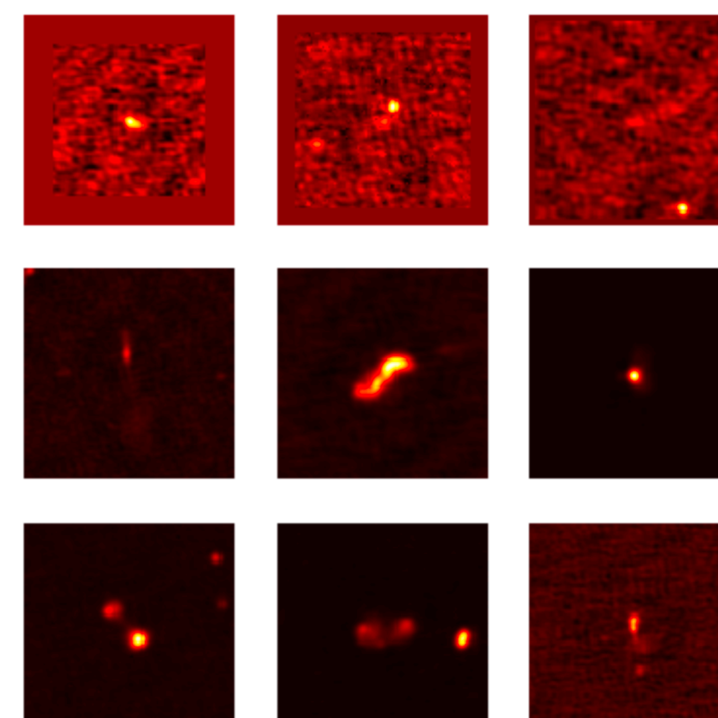
Top Quark Jet



QCD Jet



Galaxies



Classification

$$H(p,q) = -\sum p_i \ln q_i = H(p) + D_{KL}(p||q)$$

- Minimizing cross entropy: For fixed p - minimise difference (KL-divergence) between q and p

$$\frac{1}{N}\ln \Pi_i q_i^{Np_i} = \sum_i p_i \ln q_i = -H(p,q)$$

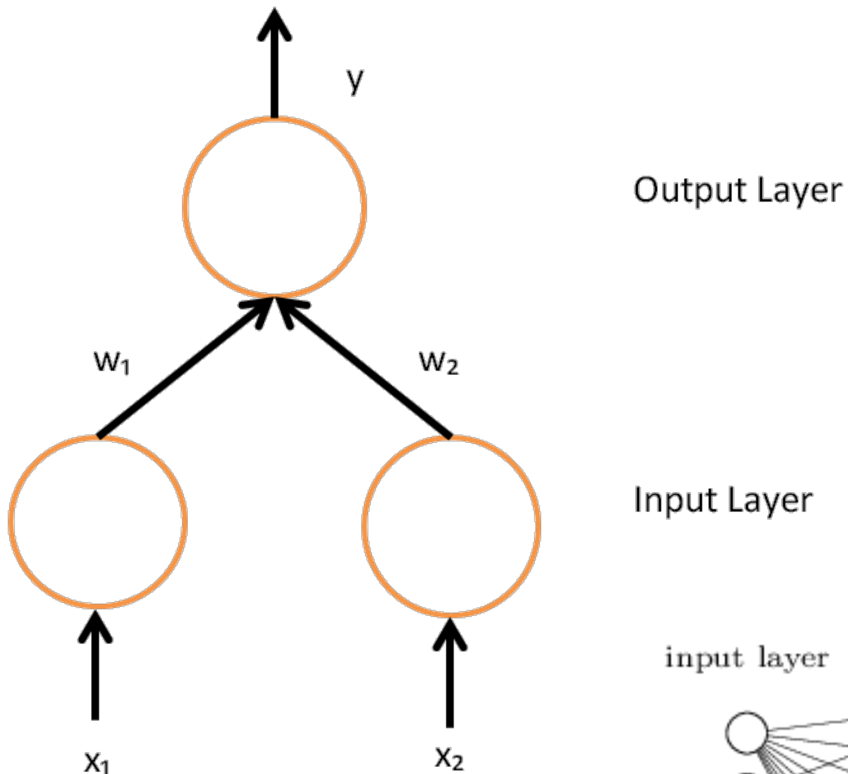
- Minimizing cross entropy: Equivalent to maximising the likelihood

$$L = \sum_{\text{Samples}} -y_s \ln \hat{y_s} - (1-y_s)\ln(1-\hat{y_s})$$

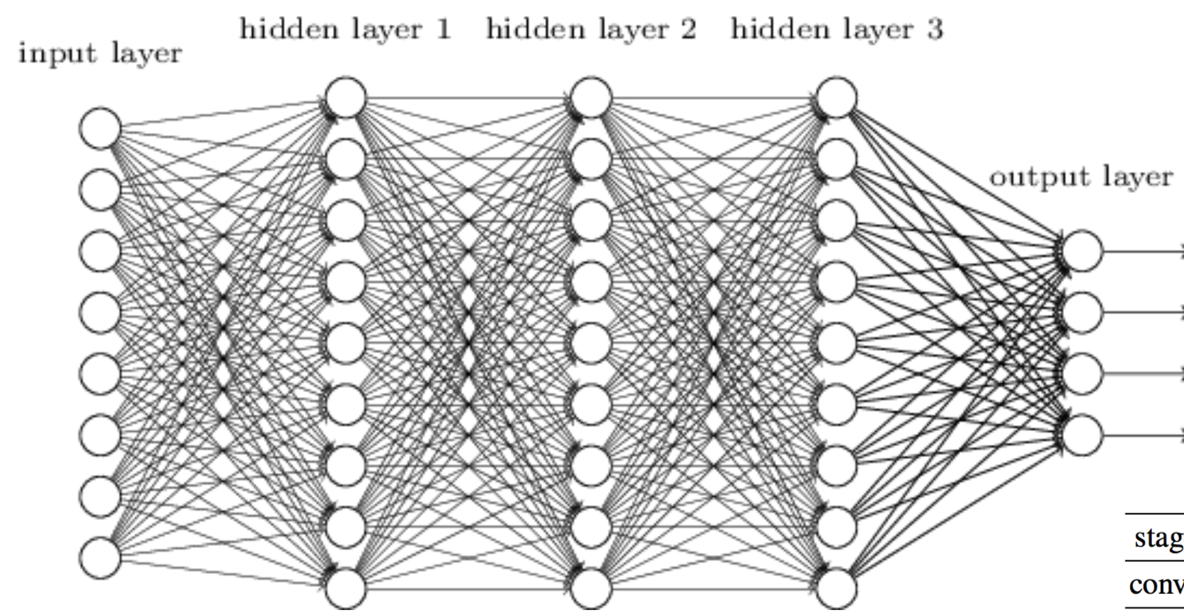
True class
image is cat: 0
image is dog: 1

Predicted class
DNN output between 0 and 1

Complexity



2 weights



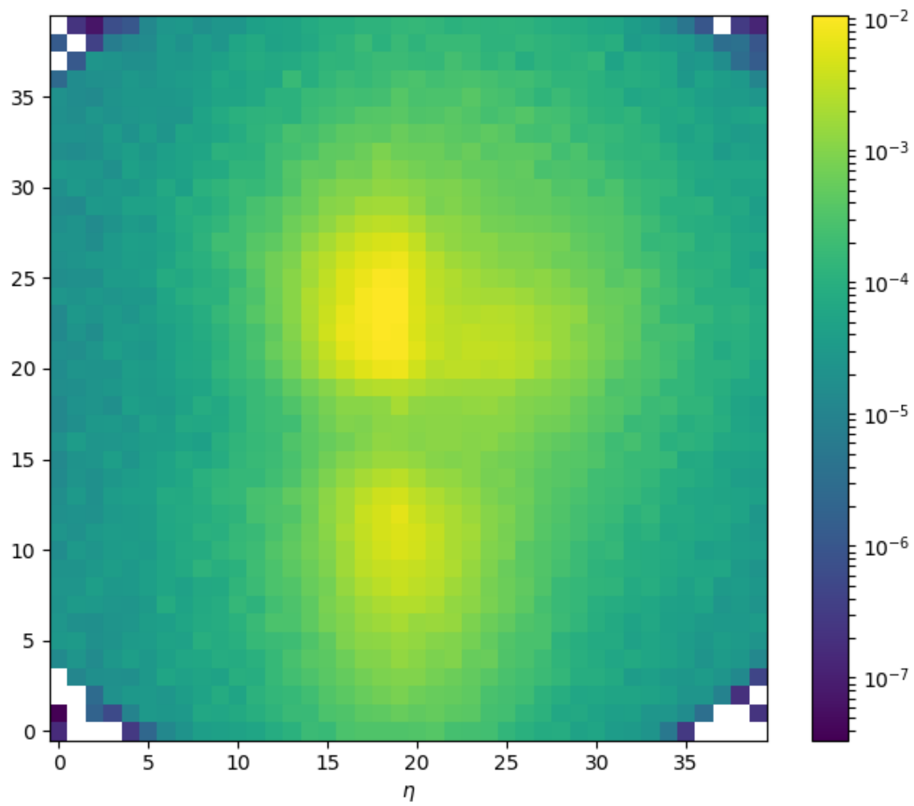
300 weights

25 million weights

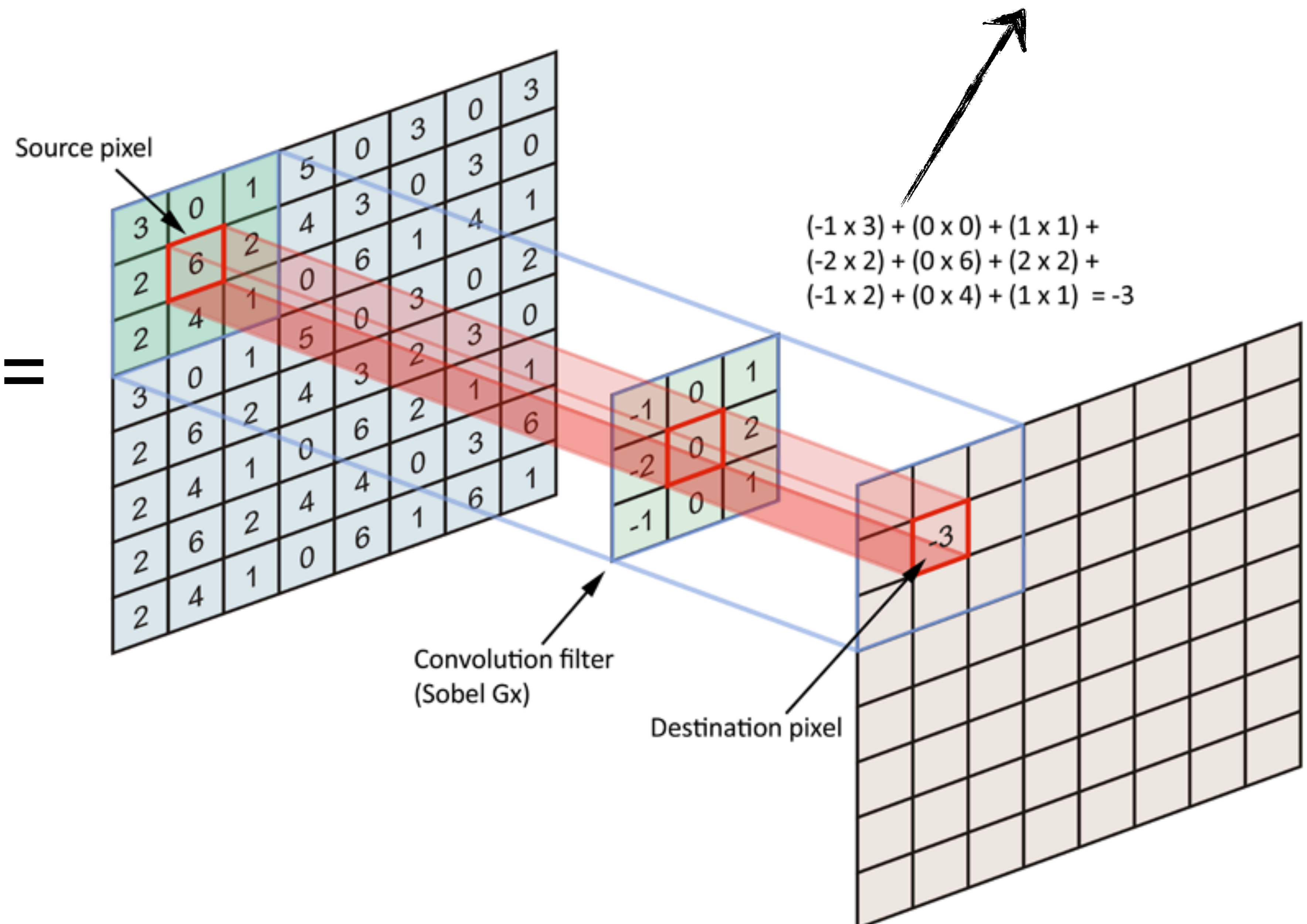
stage	output	ResNet-50	ResNeXt-50 (32×4d)
conv1	112×112	7×7, 64, stride 2	7×7, 64, stride 2
conv2	56×56	3×3 max pool, stride 2 [1×1, 64 3×3, 64 1×1, 256] ×3	3×3 max pool, stride 2 [1×1, 128 3×3, 128, C=32 1×1, 256] ×3
conv3	28×28	[1×1, 128 3×3, 128 1×1, 512] ×4	[1×1, 256 3×3, 256, C=32 1×1, 512] ×4
conv4	14×14	[1×1, 256 3×3, 256 1×1, 1024] ×6	[1×1, 512 3×3, 512, C=32 1×1, 1024] ×6
conv5	7×7	[1×1, 512 3×3, 512 1×1, 2048] ×3	[1×1, 1024 3×3, 1024, C=32 1×1, 2048] ×3
	1×1	global average pool 1000-d fc, softmax	global average pool 1000-d fc, softmax
# params.		$\mathbf{25.5}\times 10^6$	$\mathbf{25.0}\times 10^6$
FLOPs		$\mathbf{4.1}\times 10^9$	$\mathbf{4.2}\times 10^9$

Convolution

Train these weights

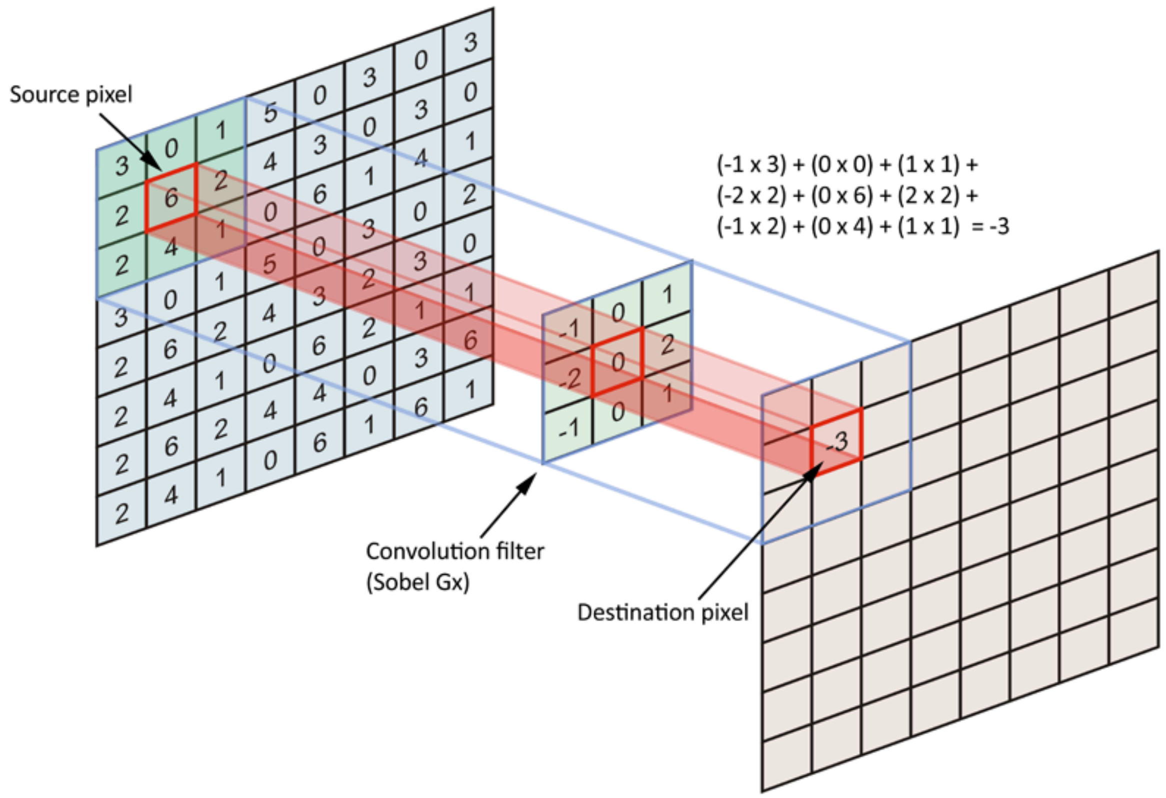


=

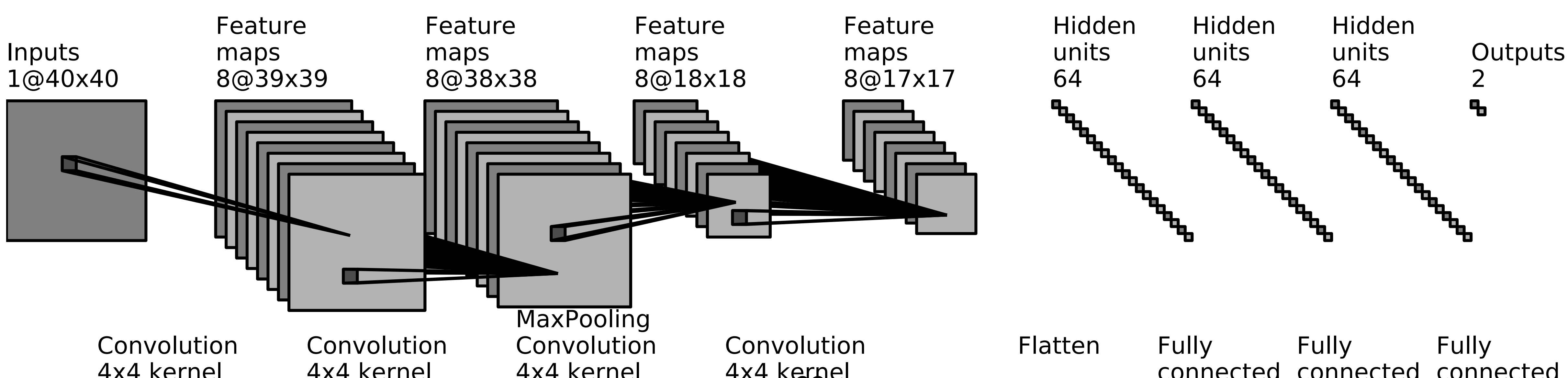


Efficient use of weights and natural encoding of translational symmetry.

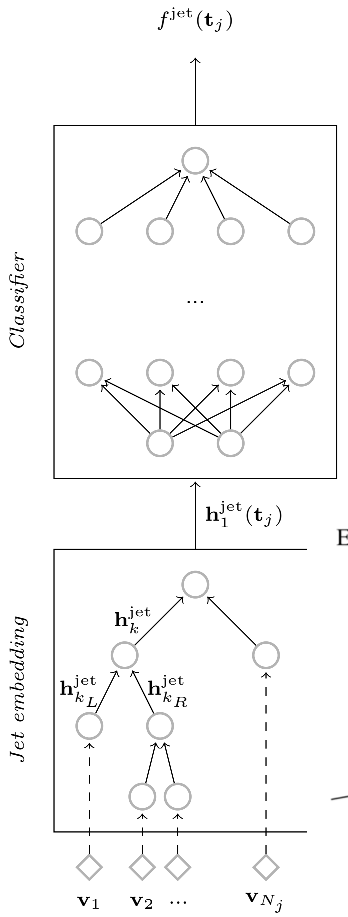
Convolution network

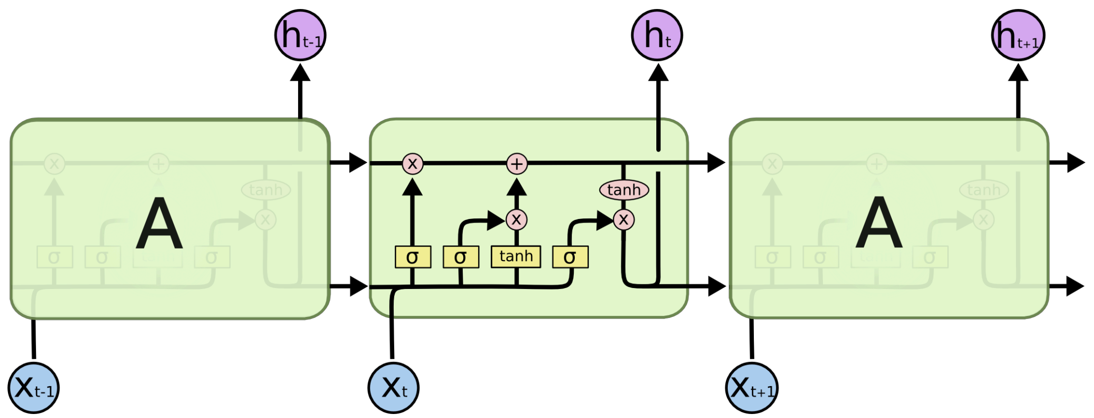


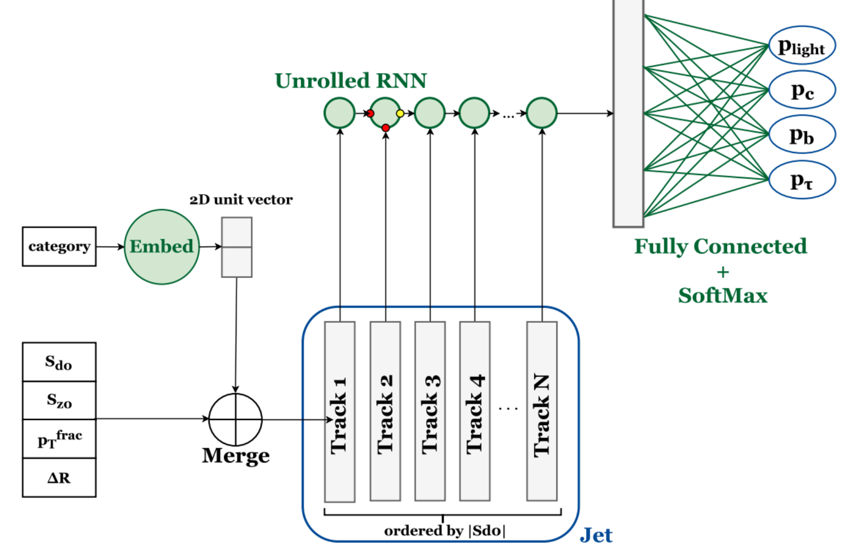
- ***How to build a convolution network?***
  - Multiple parallel and successive convolutions
  - Pooling
  - Simple network in the end

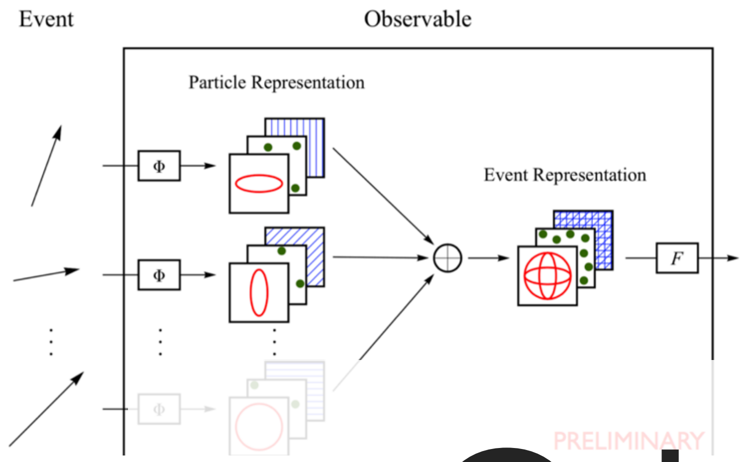


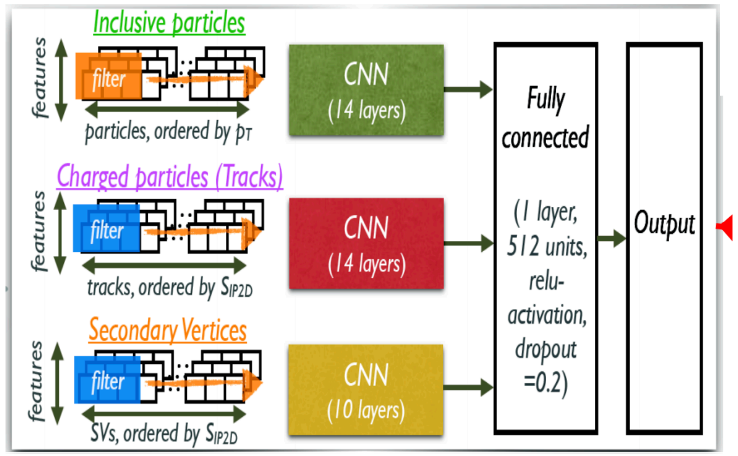
Other methods

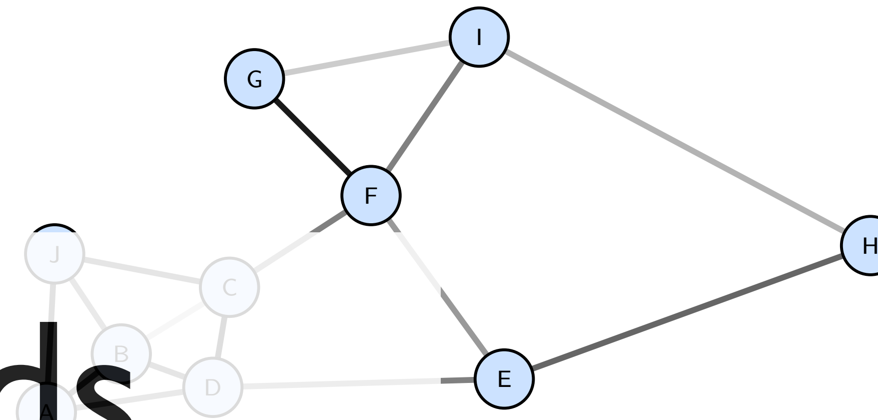


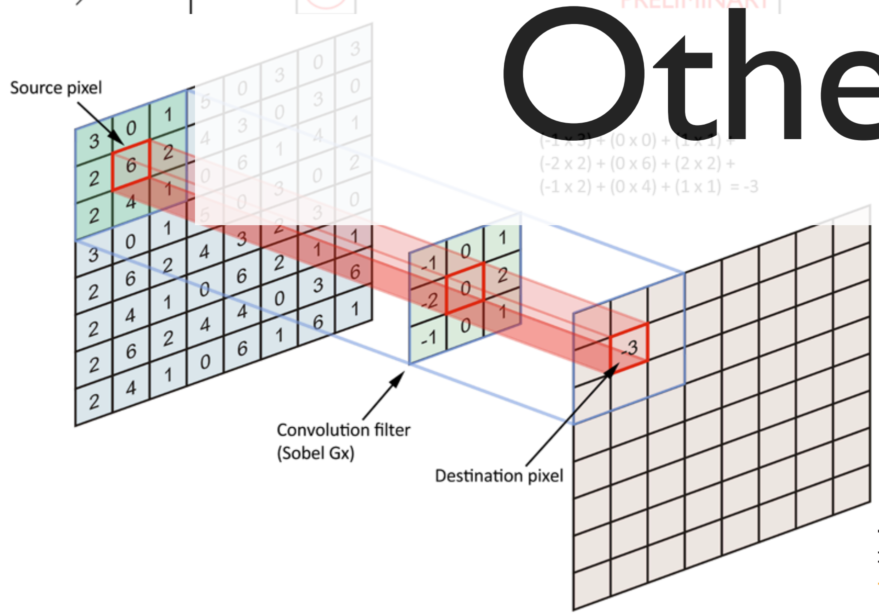


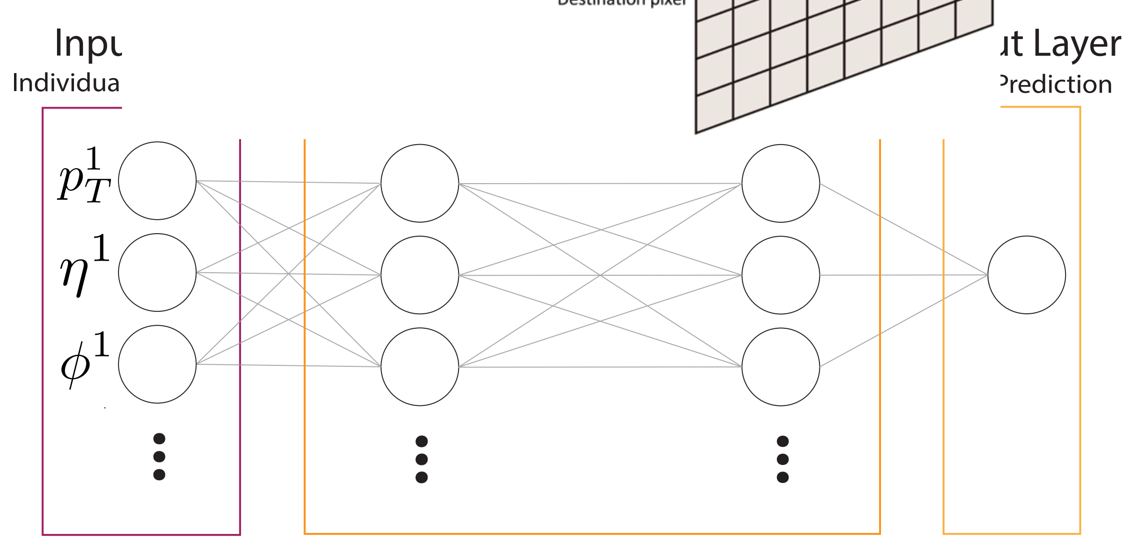












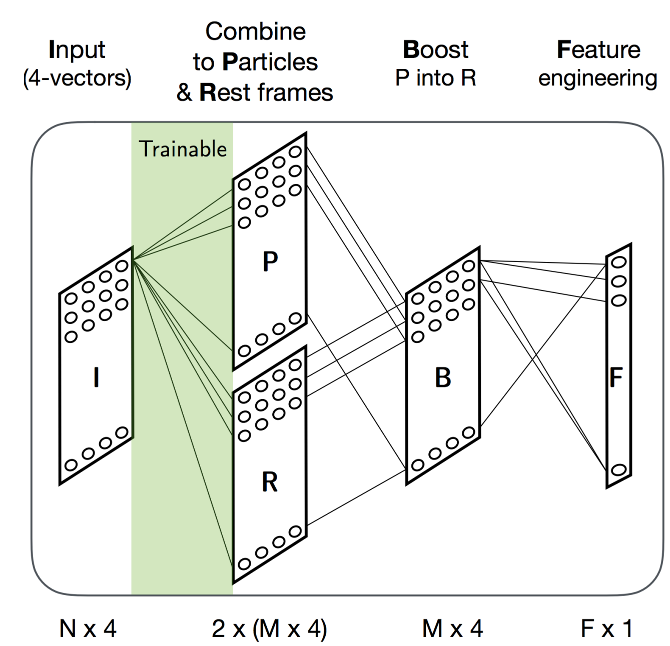
$$k_{\mu,i} = \begin{pmatrix} E_0 & E_1 & \dots & E_N \\ p_{x,0} & p_{x,1} & \dots & p_{x,N} \\ p_{y,0} & p_{y,1} & \dots & p_{y,N} \\ p_{z,0} & p_{z,1} & \dots & p_{z,N} \end{pmatrix}$$

Combination Layer (**CoLa**): create linear combinations:

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

Lorentz Layer (**LoLa**): Use resulting matrix to extract physics features.
Main assumption is the Minkowski metric

$$\widetilde{k}_{\mu,i} \to \sum_j (\widetilde{k}_i - \widetilde{k}_j)_\mu (\widetilde{k}_i - \widetilde{k}_j)_\nu \eta^{\mu\nu} B_{ij}$$



Comparison Studies

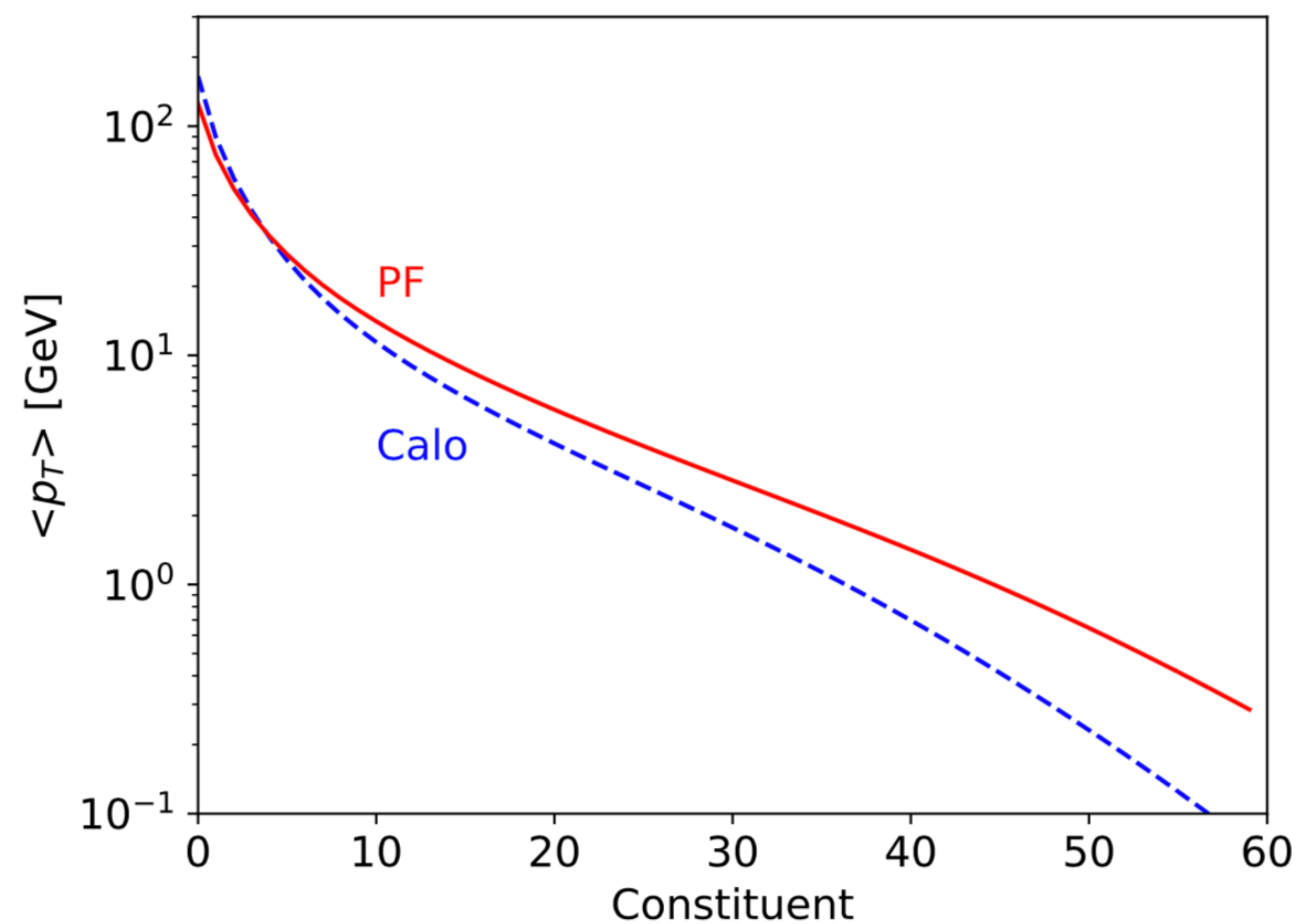
- That's A LOT of different approaches
- How do they compare?

- Many aspects - let's start with a very simple problem
  - Top Jets vs QCD jets
  - Only use four-vectors
  - Ignore detector effects, pile-up and, uncertainties

- Available at:
  - http://tinyurl.com/yxq8q3uk

Dataset

- Comparison study of different top tagging algorithms on common sample.
- Pythia + Delphes, AntiKt (R=0.8) top jets with p_T in [550,650] GeV vs QCD
- 1.2M training events, 400k each for validation and testing
- Up to 200 constituent 4-vectors per jet



The Machine Learning Landscape of Top Taggers

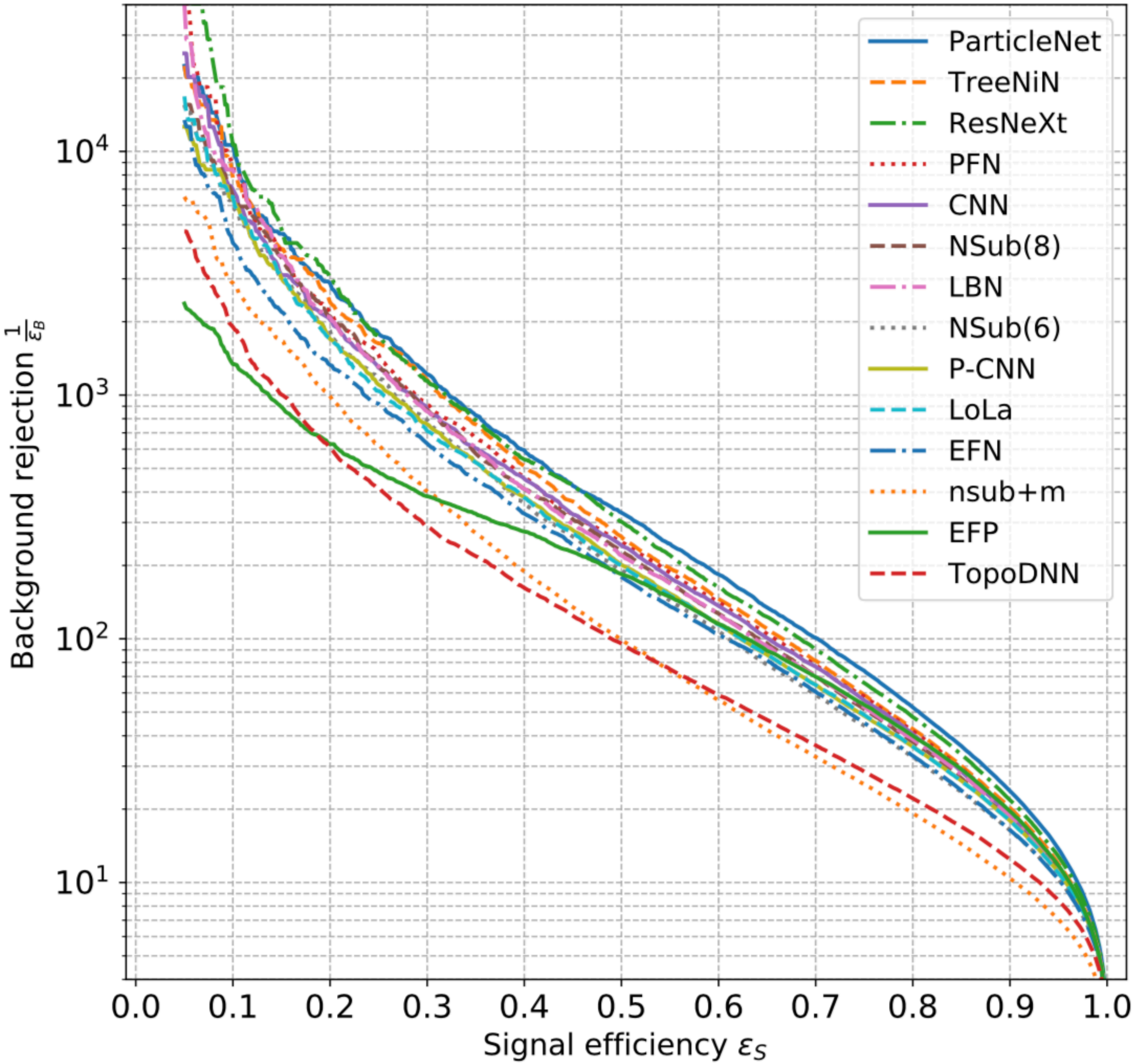
G. Kasieczka (ed)[1], T. Plehn (ed)[2], A. Butter[2], K. Cranmer[3] D. Debnath[4], M. Fairbairn[5], W. Fedorko[6], C. Gay[6], L. Gouskos[7], P. T. Komiske[8], S. Leiss[1], A. Lister[6], S. Macaluso[3,4], E. M. Metodiev[8], L. Moore[9], B. Nachman,[10,11], K. Nordström[12,13], J. Pearkes[6], H. Qu[7], Y. Rath[14], M. Rieger[14], D. Shih[4], J. M. Thompson[2], and S. Varma[5]

1 Institut für Experimentalphysik, Universität Hamburg, Germany
2 Institut für Theoretische Physik, Universität Heidelberg, Germany
3 Center for Cosmology and Particle Physics and Center for Data Science, NYU, USA
4 NHECT, Dept. of Physics and Astronomy, Rutgers, The State University of NJ, USA
5 Theoretical Particle Physics and Cosmology, King's College London, United Kingdom
6 Department of Physics and Astronomy, The University of British Columbia, Canada
7 Department of Physics, University of California, Santa Barbara, USA
8 Center for Theoretical Physics, MIT, Cambridge, USA
9 CP3, Université Catholique de Louvain, Louvain-la-Neuve, Belgium
10 Physics Division, Lawrence Berkeley National Laboratory, Berkeley, USA
11 Simons Inst. for the Theory of Computing, University of California, Berkeley, USA
12 National Institute for Subatomic Physics (NIKHEF), Amsterdam, Netherlands
13 LPTHE, CNRS & Sorbonne Université, Paris, France
14 III. Physics Institute A, RWTH Aachen University, Germany

gregor.kasieczka@uni-hamburg.de
plehn@uni-heidelberg.de

arXiv: 1902.09914

Overview



Background rejection $\frac{1}{\varepsilon_B}$
Signal efficiency ε_S
10^4
10^3
10^2
10^1
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
ParticleNet
TreeNiN
ResNeXt
PFN
CNN
NSub(8)
LBN
NSub(6)
P-CNN
LoLa
EFN
nsub+m
EFP
TopoDNN

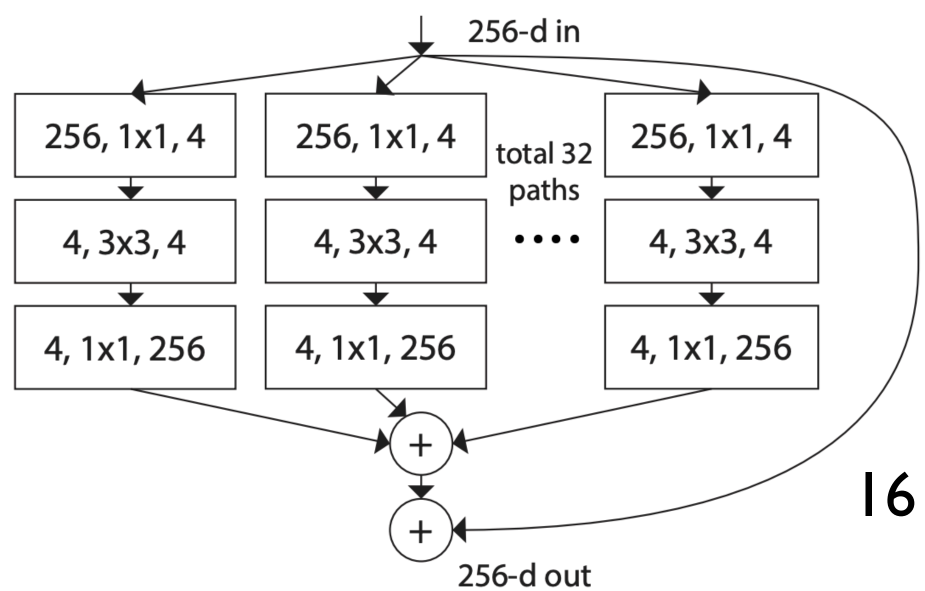
	AUC	Accuracy	$1/\epsilon_B$ ($\epsilon_S = 0.3$) mean	median	#Parameters
CNN [16]	0.981	0.930	995±15	966±18	610k
ResNeXt [30]	0.984	0.936	1246±28	1286±31	1.46M
TopoDNN [18]	0.972	0.916	378± 5	391 ± 8	59k
Multi-body N-subjettiness 6 [24]	0.979	0.922	802±12	783±13	57k
Multi-body N-subjettiness 8 [24]	0.981	0.929	926±20	886±18	58k
TreeNiN [43]	0.982	0.933	1209±23	1167±24	34k
P-CNN	0.980	0.930	838±13	841±14	348k
ParticleNet [47]	0.985	0.938	1383±45	1374±41	498k
LBN [19]	0.981	0.931	852±67	971±20	705k
LoLa [22]	0.980	0.929	768±11	751±11	127k
Energy Flow Polynomials [21]	0.980	0.932	380		1k
Energy Flow Network [23]	0.979	0.927	734±13	729±11	82k
Particle Flow Network [23]	0.982	0.932	1005±21	1005±29	82k
GoaT (see text)	0.985	0.939	1440		25k

Images

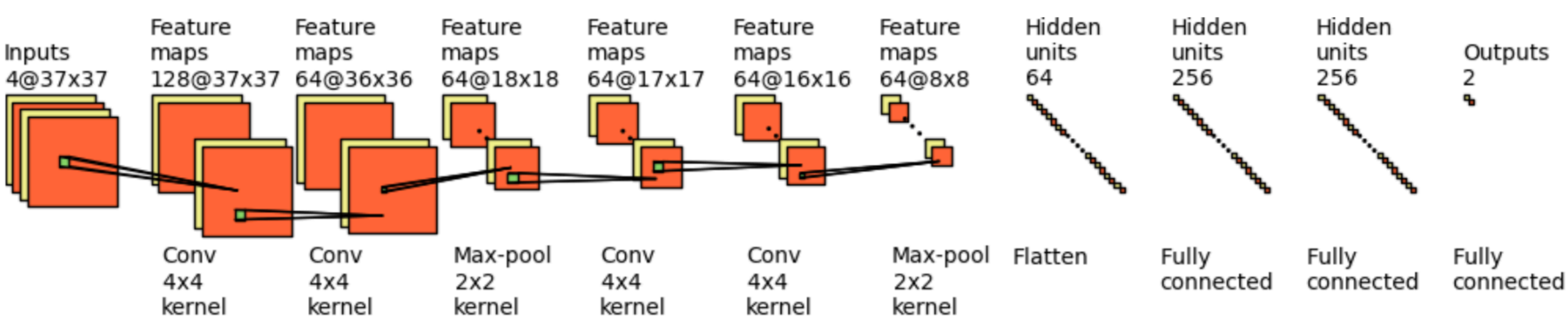
Pulling Out All the Tops with Computer Vision and Deep Learning, S Macaluso, D Shih, 1803.00107

	AUC	Accuracy	$1/\epsilon_B$ ($\epsilon_S = 0.3$) mean	median	#Parameters
CNN [16]	0.981	0.930	995±15	966±18	610k
ResNeXt [30]	0.984	0.936	1246±28	1286±31	1.46M
TopoDNN [18]	0.972	0.916	378± 5	391 ± 8	59k
Multi-body N-subjettiness 6 [24]	0.979	0.922	802±12	783±13	57k
Multi-body N-subjettiness 8 [24]	0.981	0.929	926±20	886±18	58k
TreeNiN [43]	0.982	0.933	1209±23	1167±24	34k
P-CNN	0.980	0.930	838±13	841±14	348k
ParticleNet [47]	0.985	0.938	1383±45	1374±41	498k
LBN [19]	0.981	0.931	852±67	971±20	705k
LoLa [22]	0.980	0.929	768±11	751±11	127k
Energy Flow Polynomials [21]	0.980	0.932	380		1k
Energy Flow Network [23]	0.979	0.927	734±13	729±11	82k
Particle Flow Network [23]	0.982	0.932	1005±21	1005±29	82k
GoaT (see text)	0.985	0.939	1440		25k

stage	output	ResNet-50	**ResNeXt-50 (32×4d)**
conv1	112×112	7×7, 64, stride 2	7×7, 64, stride 2
conv2	56×56	3×3 max pool, stride 2	3×3 max pool, stride 2
		[1×1, 64 3×3, 64 1×1, 256] ×3	[1×1, 128 3×3, 128, C=32 1×1, 256] ×3
conv3	28×28	[1×1, 128 3×3, 128 1×1, 512] ×4	[1×1, 256 3×3, 256, C=32 1×1, 512] ×4
conv4	14×14	[1×1, 256 3×3, 256 1×1, 1024] ×6	[1×1, 512 3×3, 512, C=32 1×1, 1024] ×6
conv5	7×7	[1×1, 512 3×3, 512 1×1, 2048] ×3	[1×1, 1024 3×3, 1024, C=32 1×1, 2048] ×3
	1×1	global average pool 1000-d fc, softmax	global average pool 1000-d fc, softmax
# params.		**25.5**×10^6	**25.0**×10^6
FLOPs		**4.1**×10^9	**4.2**×10^9



1611.05431

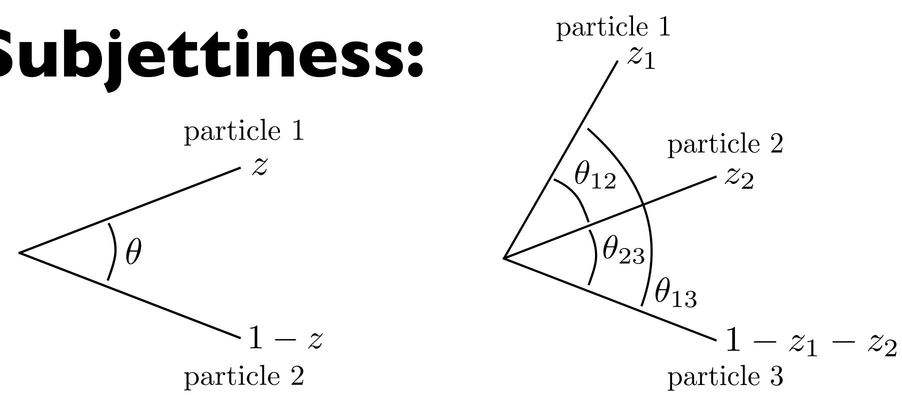


Basis

How Much Information is in a Jet?
K Datta, A Larkoski, 1704.08249
Reports of My Demise Are Greatly Exaggerated: N-subjettiness Taggers Take On Jet Images,
L Moore et al 1807.04769
Energy flow polynomials: A complete linear basis for jet substructure,
PT Komiske, ER Metodiev, J Thaler, 1712.07124

	AUC	Accuracy	$1/\epsilon_B$ ($\epsilon_S = 0.3$) mean	median	#Parameters
CNN [16]	0.981	0.930	995±15	966±18	610k
ResNeXt [30]	0.984	0.936	1246±28	1286±31	1.46M
TopoDNN [18]	0.972	0.916	378± 5	391 ± 8	59k
Multi-body N-subjettiness 6 [24]	0.979	0.922	802±12	783±13	57k
Multi-body N-subjettiness 8 [24]	0.981	0.929	926±20	886±18	58k
TreeNiN [43]	0.982	0.933	1209±23	1167±24	34k
P-CNN	0.980	0.930	838±13	841±14	348k
ParticleNet [47]	0.985	0.938	1383±45	1374±41	498k
LBN [19]	0.981	0.931	852±67	971±20	705k
LoLa [22]	0.980	0.929	768±11	751±11	127k
Energy Flow Polynomials [21]	0.980	0.932	380		1k
Energy Flow Network [23]	0.979	0.927	734±13	729±11	82k
Particle Flow Network [23]	0.982	0.932	1005±21	1005±29	82k
GoaT (see text)	0.985	0.939	1440		25k

N-Subjettiness:



$$\tau_N^{(\beta)} = \frac{1}{p_{TJ}} \sum_{i\in \mathrm{Jet}} p_{Ti} \min\left\{R_{1i}^{\beta}, R_{2i}^{\beta}, \ldots, R_{Ni}^{\beta}\right\}$$

2-body: $\tau_1^{(1)}, \tau_1^{(2)}$

3-body: $\tau_1^{(0.5)}, \tau_1^{(1)}, \tau_1^{(2)}, \tau_2^{(1)}, \tau_2^{(2)}$

4-body: $\tau_1^{(0.5)}, \tau_1^{(1)}, \tau_1^{(2)}, \tau_2^{(0.5)}, \tau_2^{(1)}, \tau_2^{(2)}, \tau_3^{(1)}, \tau_3^{(2)}$

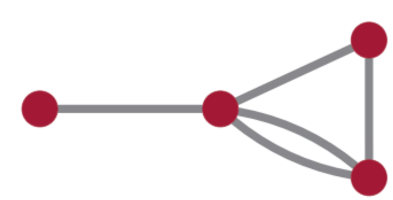
5-body: $\tau_1^{(0.5)}, \tau_1^{(1)}, \tau_1^{(2)}, \tau_2^{(0.5)}, \tau_2^{(1)}, \tau_2^{(2)}, \tau_3^{(0.5)}, \tau_3^{(1)}, \tau_3^{(2)}, \tau_4^{(1)}, \tau_4^{(2)}$

6-body: $\tau_1^{(0.5)}, \tau_1^{(1)}, \tau_1^{(2)}, \tau_2^{(0.5)}, \tau_2^{(1)}, \tau_2^{(2)}, \tau_3^{(0.5)}, \tau_3^{(1)}, \tau_3^{(2)}, \tau_4^{(0.5)}, \tau_4^{(1)}, \tau_4^{(2)}, \tau_5^{(1)}, \tau_5^{(2)}$

Energy Flow Polynomials:

$$\Longleftrightarrow \sum_{i_j=1}^{M} z_{i_j}$$

$$\Longleftrightarrow \theta_{i_k i_\ell}$$



$$= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M}\sum_{i_4=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} \theta_{i_1 i_2} \theta_{i_2 i_3} \theta_{i_2 i_4}^2 \theta_{i_3 i_4}$$

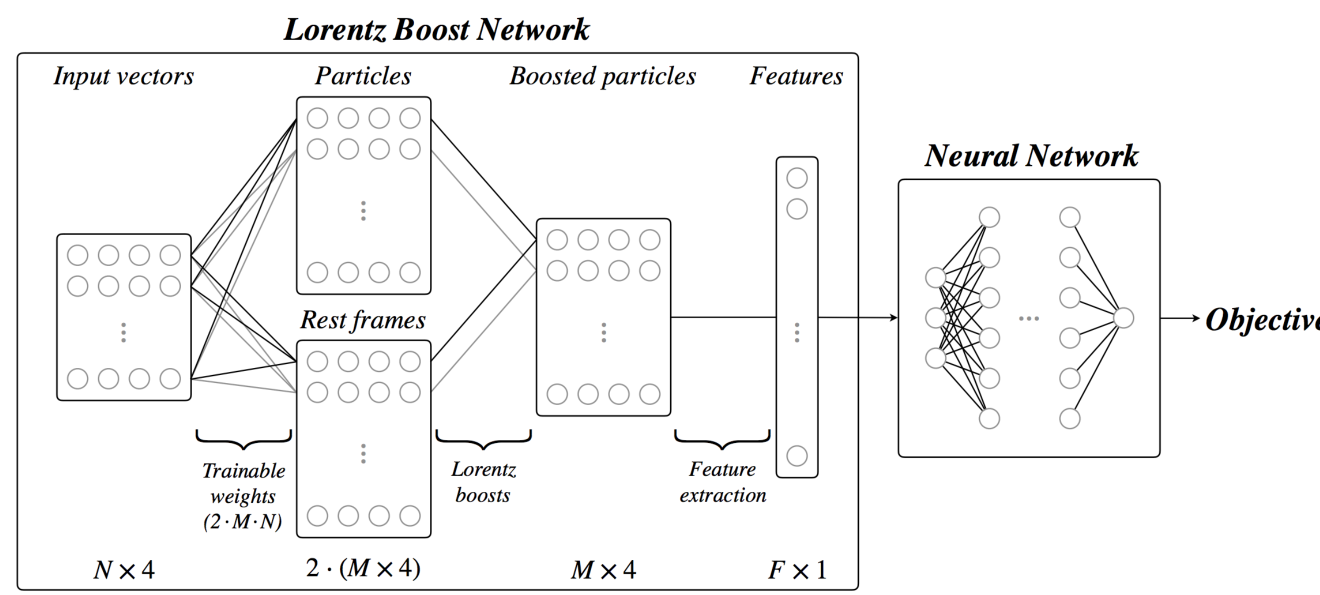
1000 graphs used in linear model

Lorentz Layer and Lorentz Boost Network

	AUC	Accuracy	$1/\epsilon_B$ ($\epsilon_S = 0.3$) mean	median	#Parameters
CNN [16]	0.981	0.930	995±15	966±18	610k
ResNeXt [30]	0.984	0.936	1246±28	1286±31	1.46M
TopoDNN [18]	0.972	0.916	378± 5	391 ± 8	59k
Multi-body N-subjettiness 6 [24]	0.979	0.922	802±12	783±13	57k
Multi-body N-subjettiness 8 [24]	0.981	0.929	926±20	886±18	58k
TreeNiN [43]	0.982	0.933	1209±23	1167±24	34k
P-CNN	0.980	0.930	838±13	841±14	348k
ParticleNet [47]	0.985	0.938	1383±45	1374±41	498k
LBN [19]	0.981	0.931	852±67	971±20	705k
LoLa [22]	0.980	0.929	768±11	751±11	127k
Energy Flow Polynomials [21]	0.980	0.932	380		1k
Energy Flow Network [23]	0.979	0.927	734±13	729±11	82k
Particle Flow Network [23]	0.982	0.932	1005±21	1005±29	82k
GoaT (see text)	0.985	0.939	1440		25k

Deep-learning Top Taggers & No End to QCD
A Butter, GK, T Plehn, M Russell
1707.08966

Lorentz Boost Networks: Autonomous Physics-Inspired Feature Engineering
M. Erdmann, E. Geiser, Y. Rath, and M. Rieger
1812.09722



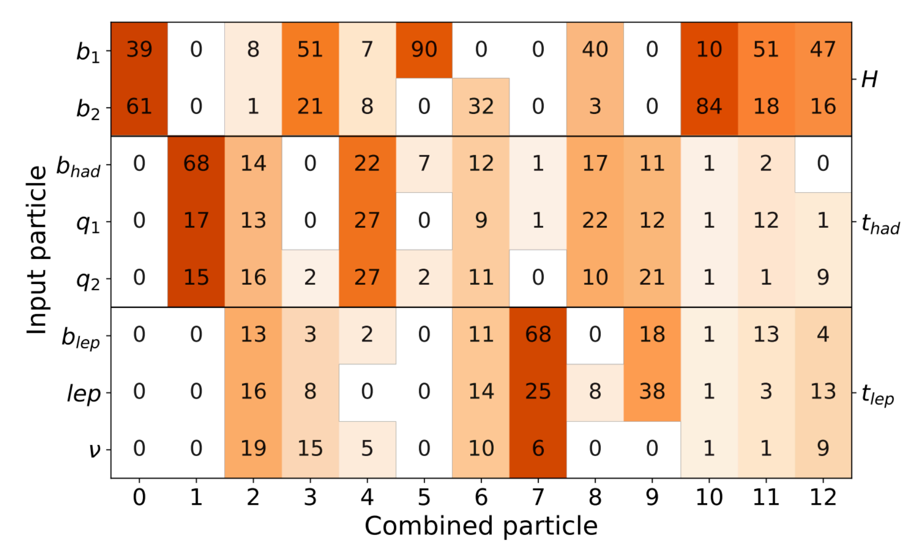
Input is a p_T sorted list of Lorentz four-vectors: (calo towers or particle flow objects)

$$k_{\mu,i} = \begin{pmatrix} E_0 & E_1 & \dots & E_N \\ p_{x,0} & p_{x,1} & \dots & p_{x,N} \\ p_{y,0} & p_{y,1} & \dots & p_{y,N} \\ p_{z,0} & p_{z,1} & \dots & p_{z,N} \end{pmatrix}$$

Combination Layer (**CoLa**): create linear combinations: $k_{\mu,i} \xrightarrow{\text{CoLa}} \widetilde{k}_{\mu,j} = k_{\mu,i}\, C_{ij}$

Lorentz Layer (**LoLa**): Use resulting matrix to extract physics features. Main assumption is the Minkowski metric

Fully connected layers for final output

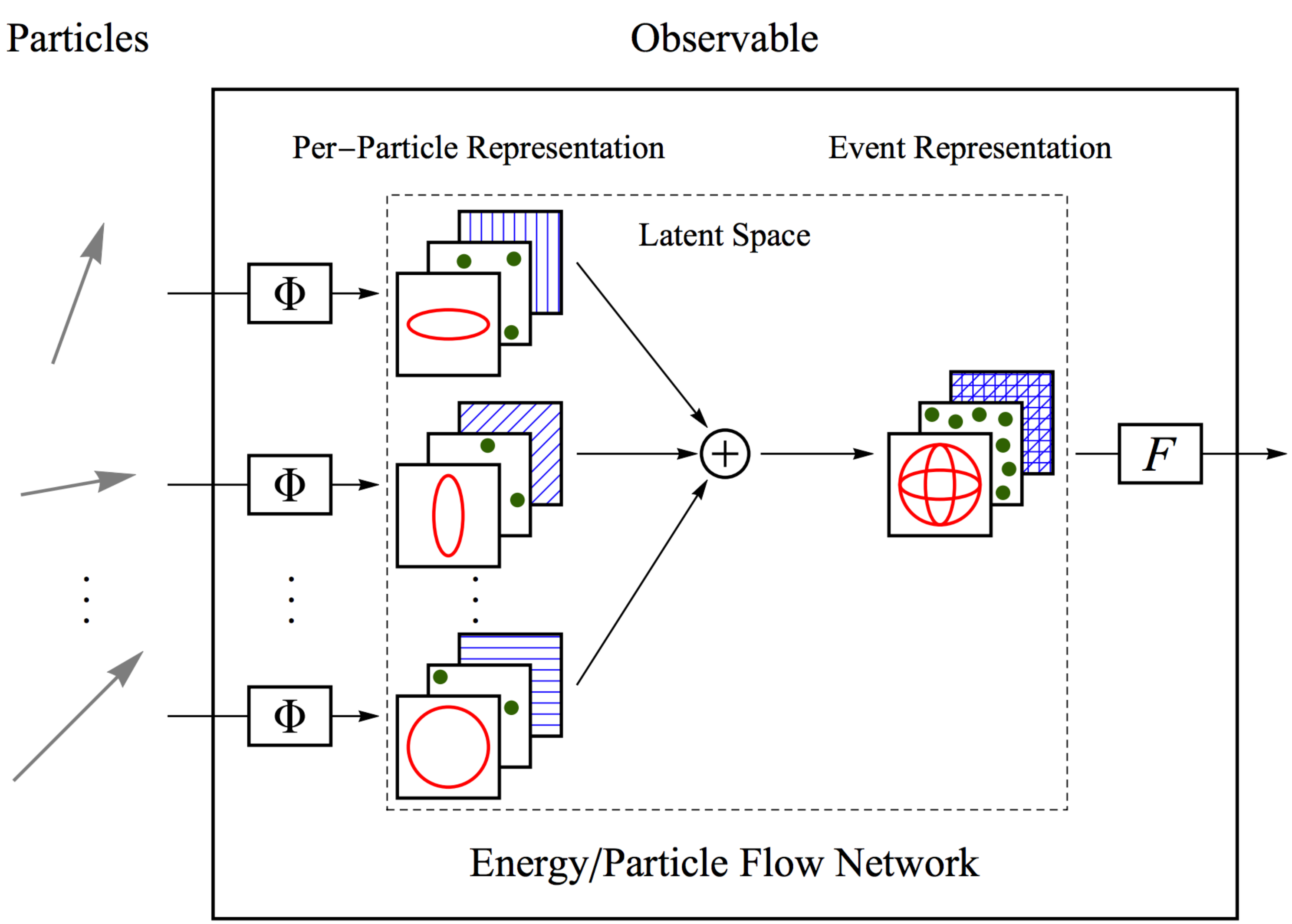


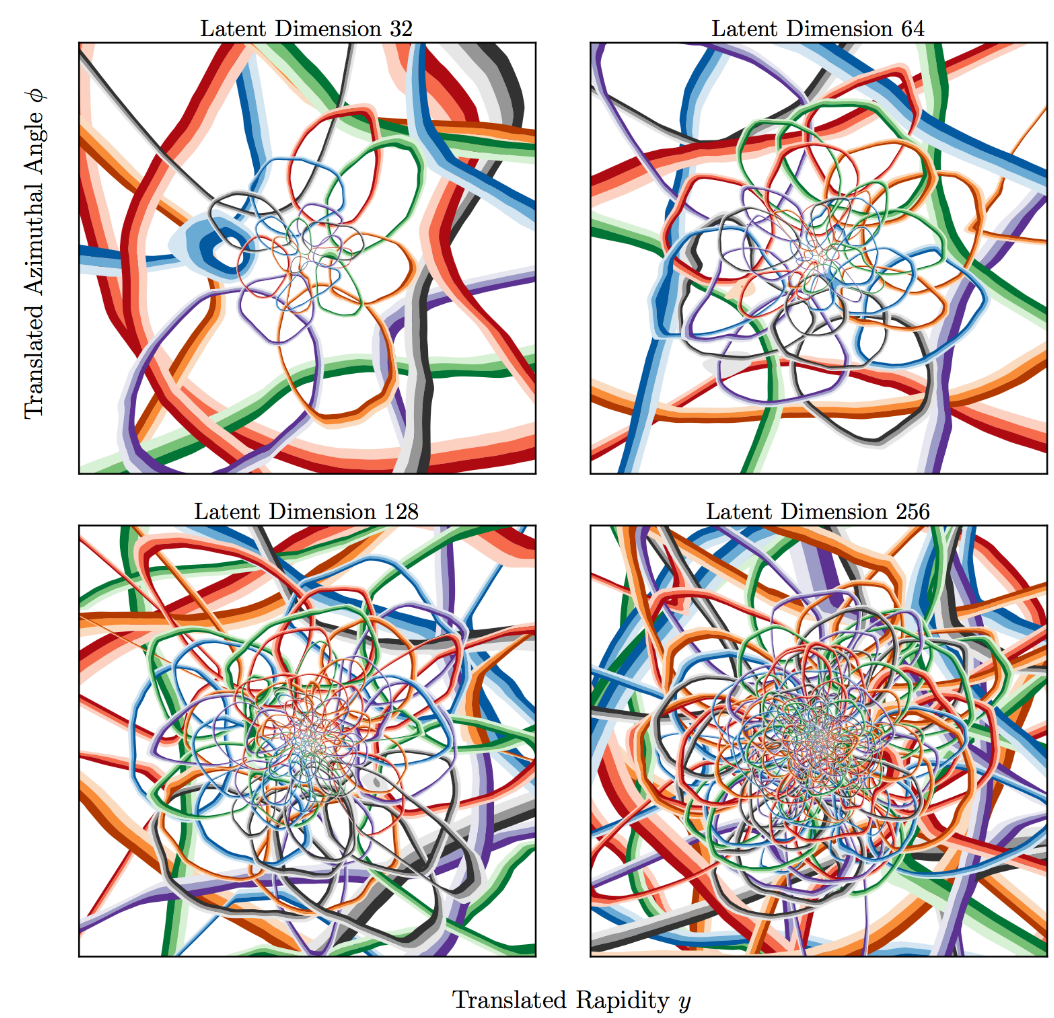
Deep Sets

Energy Flow Networks: Deep Sets for Particle Jets, PT Komiske, EM Metodiev, J Thaler, 1810.05165

$$\text{PFN:} \quad F\left(\sum_{i=1}^{M} \Phi(p_i)\right)$$

	AUC	Accuracy	$1/\epsilon_B$ ($\epsilon_S = 0.3$) mean	median	#Parameters
CNN [16]	0.981	0.930	995±15	966±18	610k
ResNeXt [30]	0.984	0.936	1246±28	1286±31	1.46M
TopoDNN [18]	0.972	0.916	378± 5	391 ± 8	59k
Multi-body N-subjettiness 6 [24]	0.979	0.922	802±12	783±13	57k
Multi-body N-subjettiness 8 [24]	0.981	0.929	926±20	886±18	58k
TreeNiN [43]	0.982	0.933	1209±23	1167±24	34k
P-CNN	0.980	0.930	838±13	841±14	348k
ParticleNet [47]	0.985	0.938	1383±45	1374±41	498k
LBN [19]	0.981	0.931	852±67	971±20	705k
LoLa [22]	0.980	0.929	768±11	751±11	127k
Energy Flow Polynomials [21]	0.980	0.932	380		1k
Energy Flow Network [23]	0.979	0.927	734±13	729±11	82k
Particle Flow Network [23]	0.982	0.932	1005±21	1005±29	82k
GoaT (see text)	0.985	0.939	1440		25k



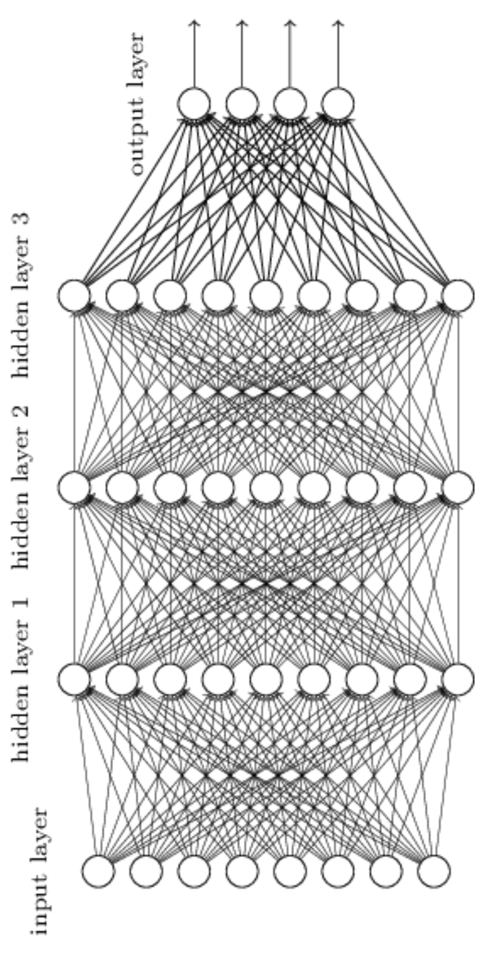


N-Vectors

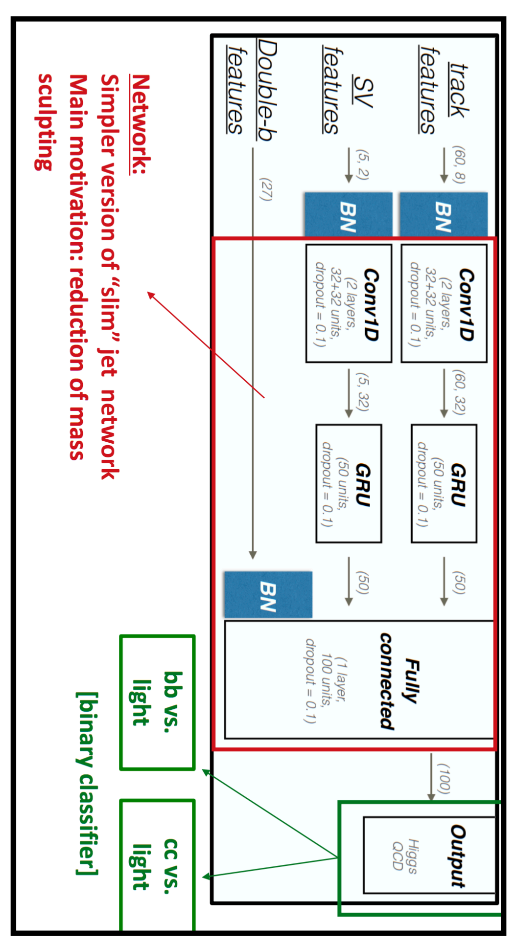
Jet Constituents for Deep Neural Network Based Top Quark Tagging, J Pearkes et al, 1704.02124
QCD-Aware Recursive Neural Networks for Jet Physics, G Louppe et al, 1702.00748
ParticleNet: Jet Tagging via Particle Clouds, H Qu, L Gouskos, 1902.08570

	AUC	Accuracy	$1/\epsilon_B$ ($\epsilon_S = 0.3$) mean	median	#Parameters
CNN [16]	0.981	0.930	995±15	966±18	610k
ResNeXt [30]	0.984	0.936	1246±28	1286±31	1.46M
TopoDNN [18]	0.972	0.916	378± 5	391 ± 8	59k
Multi-body N-subjettiness 6 [24]	0.979	0.922	802±12	783±13	57k
Multi-body N-subjettiness 8 [24]	0.981	0.929	926±20	886±18	58k
TreeNiN [43]	0.982	0.933	1209±23	1167±24	34k
P-CNN	0.980	0.930	838±13	841±14	348k
ParticleNet [47]	0.985	0.938	1383±45	1374±41	498k
LBN [19]	0.981	0.931	852±67	971±20	705k
LoLa [22]	0.980	0.929	768±11	751±11	127k
Energy Flow Polynomials [21]	0.980	0.932	380		1k
Energy Flow Network [23]	0.979	0.927	734±13	729±11	82k
Particle Flow Network [23]	0.982	0.932	1005±21	1005±29	82k
GoaT (see text)	0.985	0.939	1440		25k

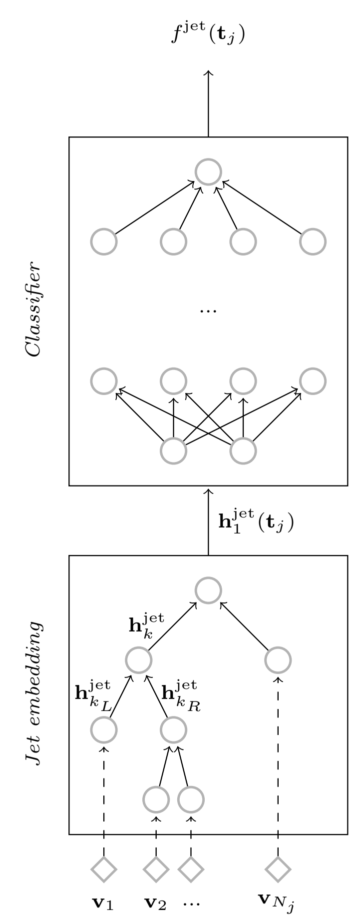
Fully connected



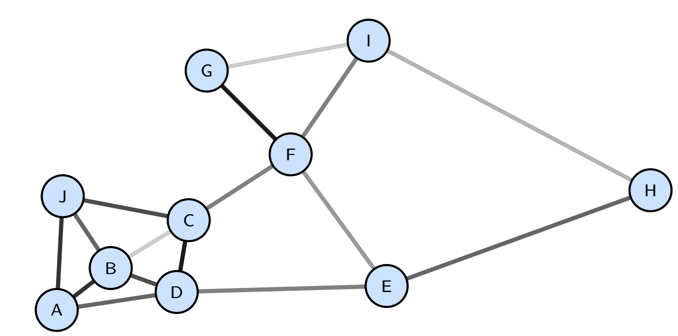
1D Convolution

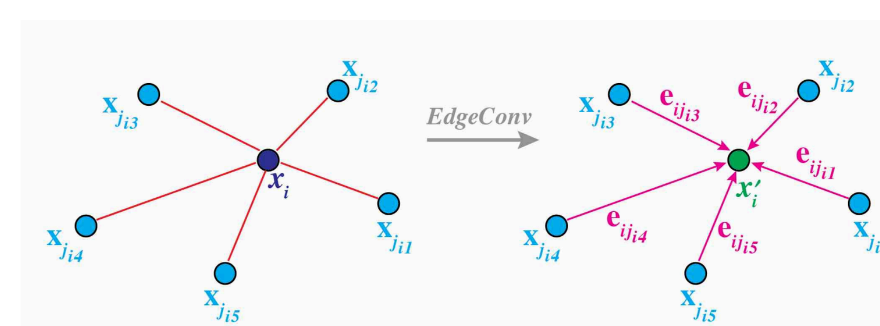


Trees

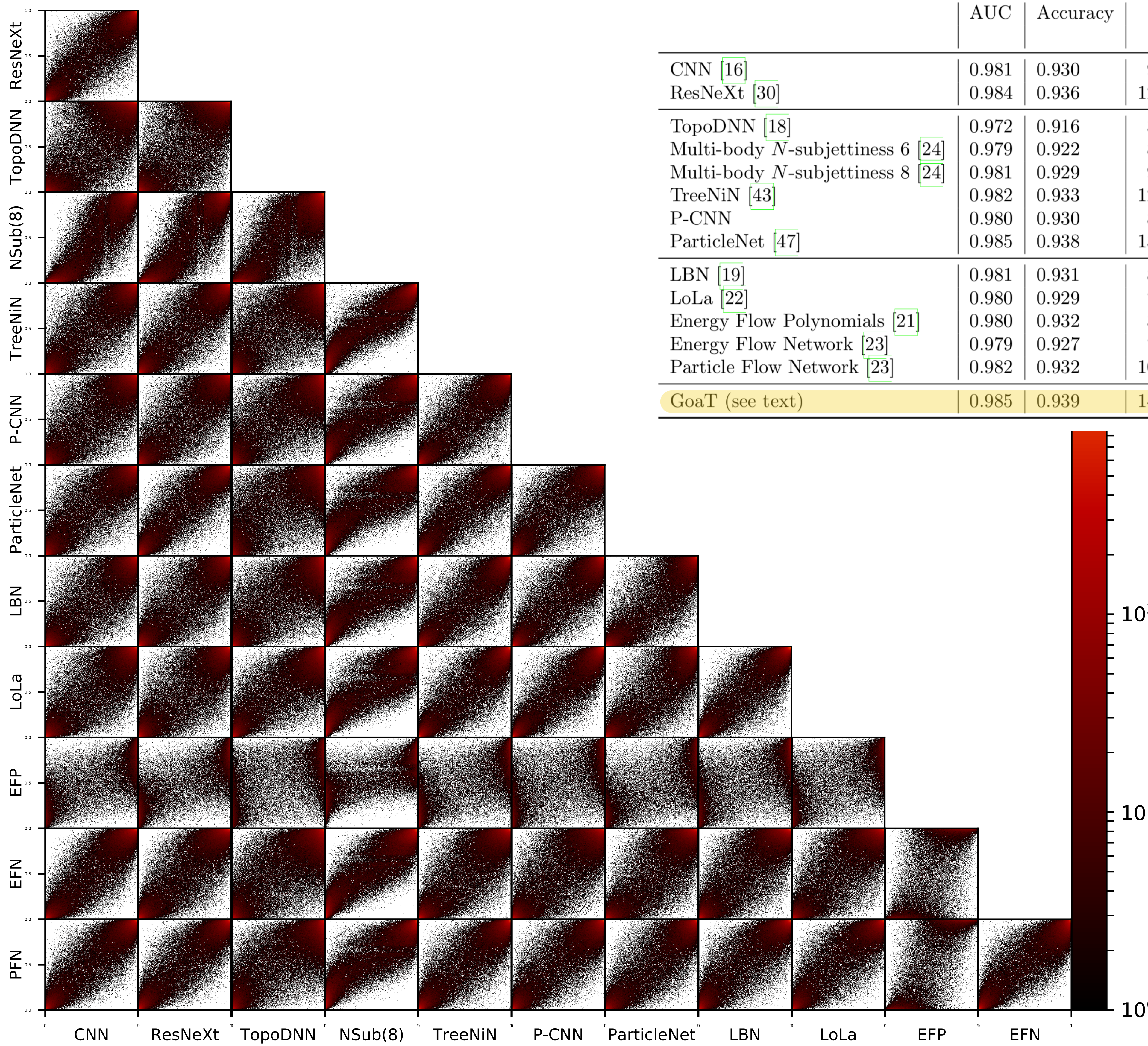


Graphs





	AUC	Accuracy	$1/\epsilon_B$ ($\epsilon_S = 0.3$) mean	median	#Parameters
CNN [16]	0.981	0.930	995±15	966±18	610k
ResNeXt [30]	0.984	0.936	1246±28	1286±31	1.46M
TopoDNN [18]	0.972	0.916	378± 5	391 ± 8	59k
Multi-body N-subjettiness 6 [24]	0.979	0.922	802±12	783±13	57k
Multi-body N-subjettiness 8 [24]	0.981	0.929	926±20	886±18	58k
TreeNiN [43]	0.982	0.933	1209±23	1167±24	34k
P-CNN	0.980	0.930	838±13	841±14	348k
ParticleNet [47]	0.985	0.938	1383±45	1374±41	498k
LBN [19]	0.981	0.931	852±67	971±20	705k
LoLa [22]	0.980	0.929	768±11	751±11	127k
Energy Flow Polynomials [21]	0.980	0.932	380		1k
Energy Flow Network [23]	0.979	0.927	734±13	729±11	82k
Particle Flow Network [23]	0.982	0.932	1005±21	1005±29	82k
GoaT (see text)	0.985	0.939	1440		25k



Rediscovered Ensembling!

Transfer Learning

- There exist very powerful architectures for image classification
  - InceptionResNet V2 weights available (55M weights, 572 layers)
  - Trained on 1000 classes of "real" photographs
  - Why not just apply it to jet images?
  - Preprocessing

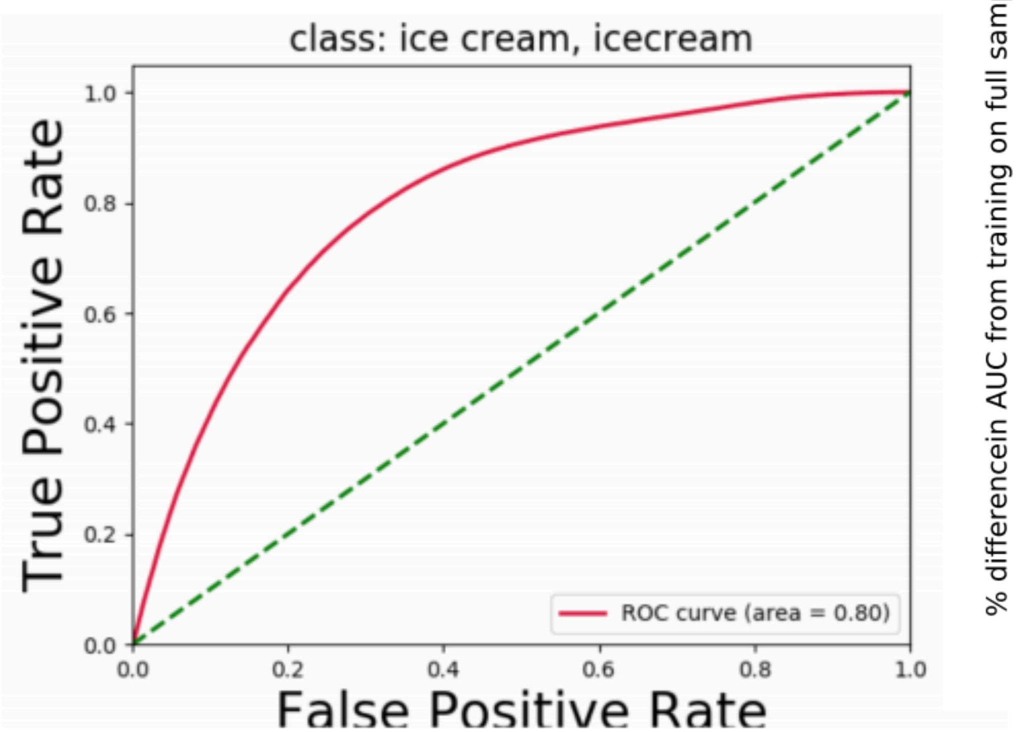
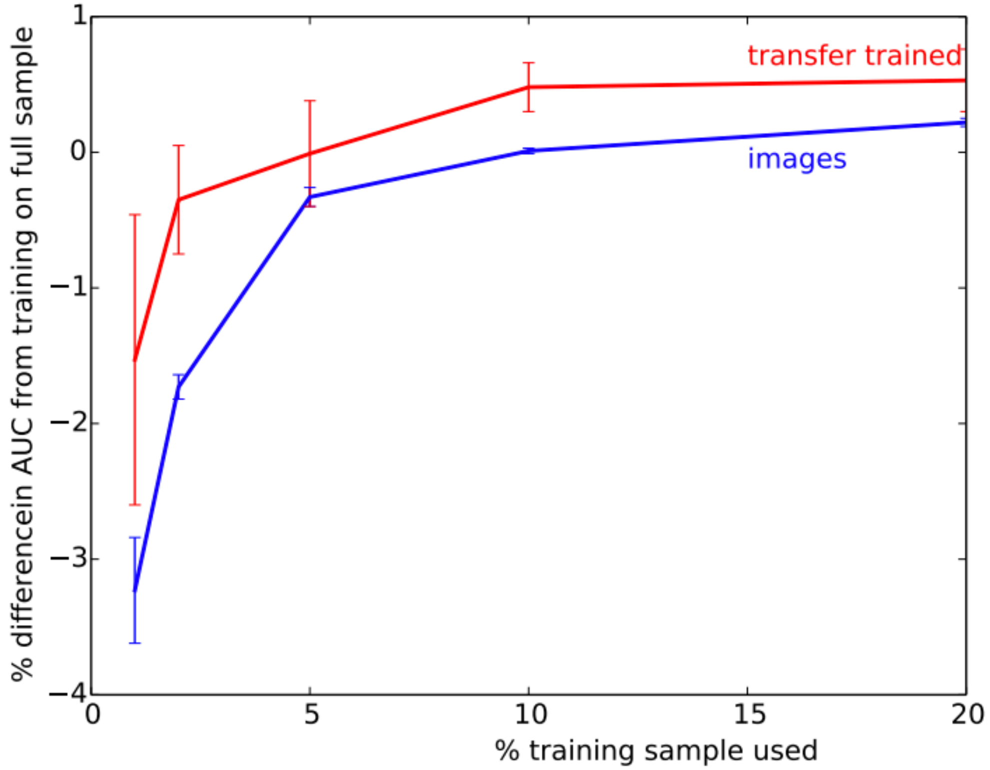


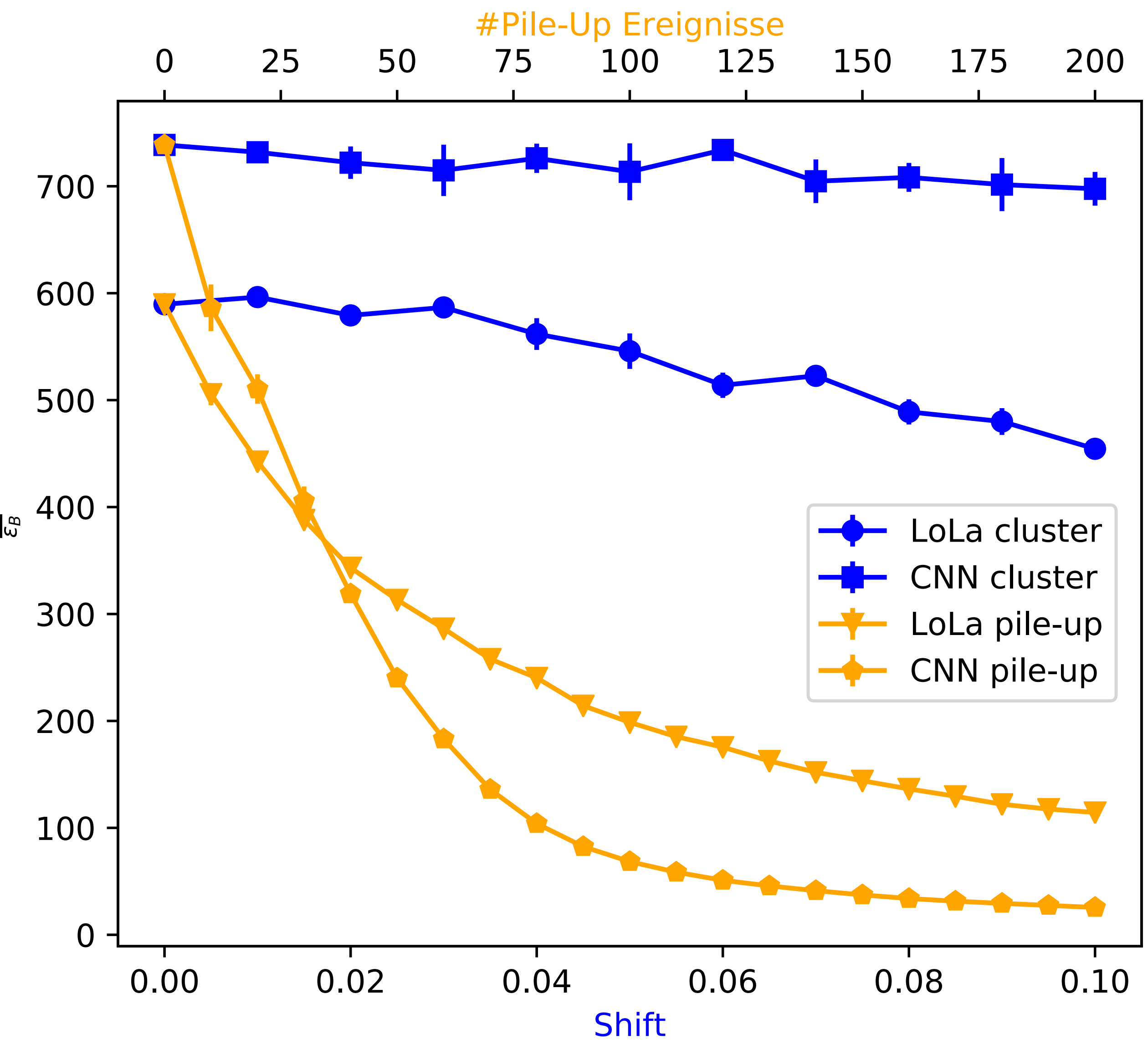
Image class "ice cream" identifies QCD jets



Transfer learning strong classifiers

(plots by Lisa Benato and Jennifer Thompson)

Systematic Uncertainties



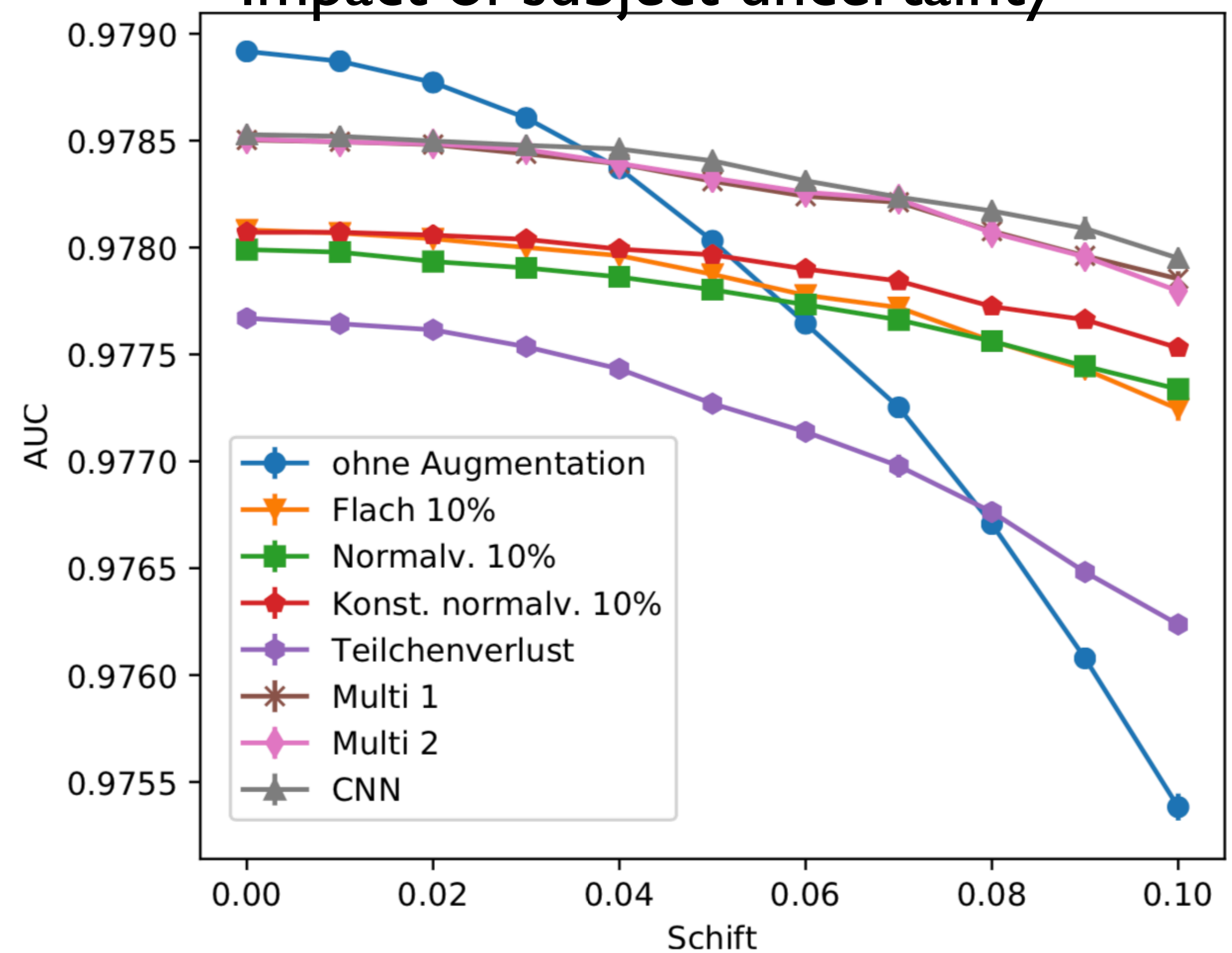
- Simulate systematic differences between training MC and collision data
- Test network response under
  - rescaling of 4-vector inputs (mimic jet energy scale)
  - adding Pile-Up
- Test mitigation with data augmentation/ adversarial training

Plot by Sven Bollweg (BSc thesis), work with Heidelberg (Plehn),

Data Augmentation

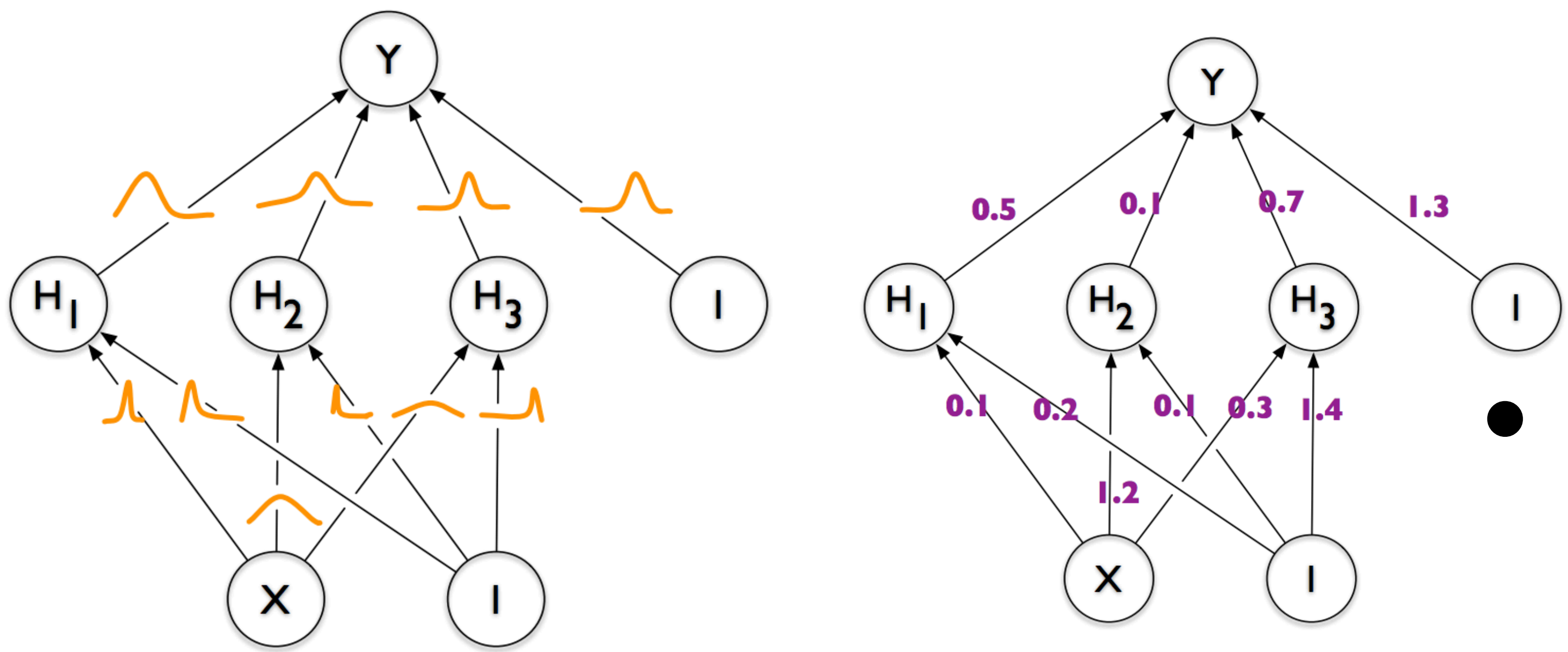
- Test network response under global rescaling of all 4-vector inputs (simimilar to jet energy scale)
- Re-train network using shifted samples as well.
  - So the network sees multiple (shifted) copies of the event = *data augmentation*
- Trade off performance and stability
- Now looking into multiple simultaenous uncertainties
  - resolution
  - pile up
  - lost particles
  - …
- Can adversarial training help further?

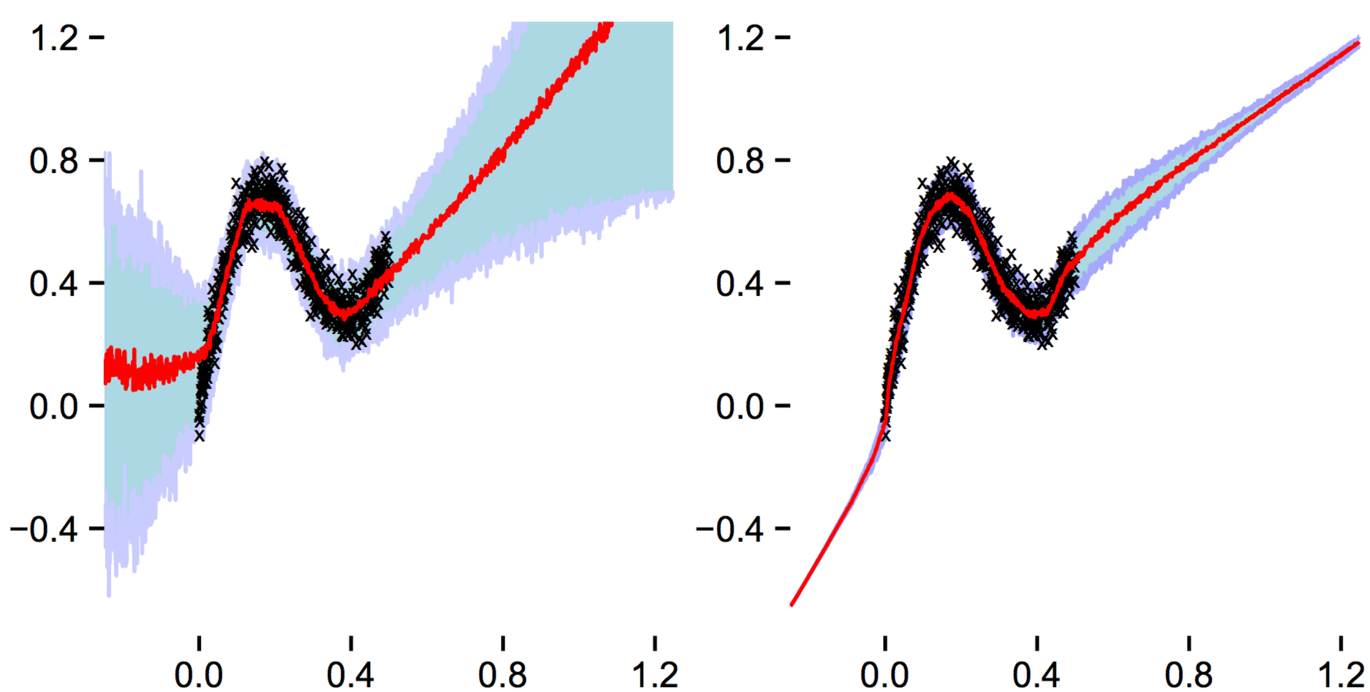
Impact of subject uncertainty



Plot by Sven Bollweg

Bayesian Networks





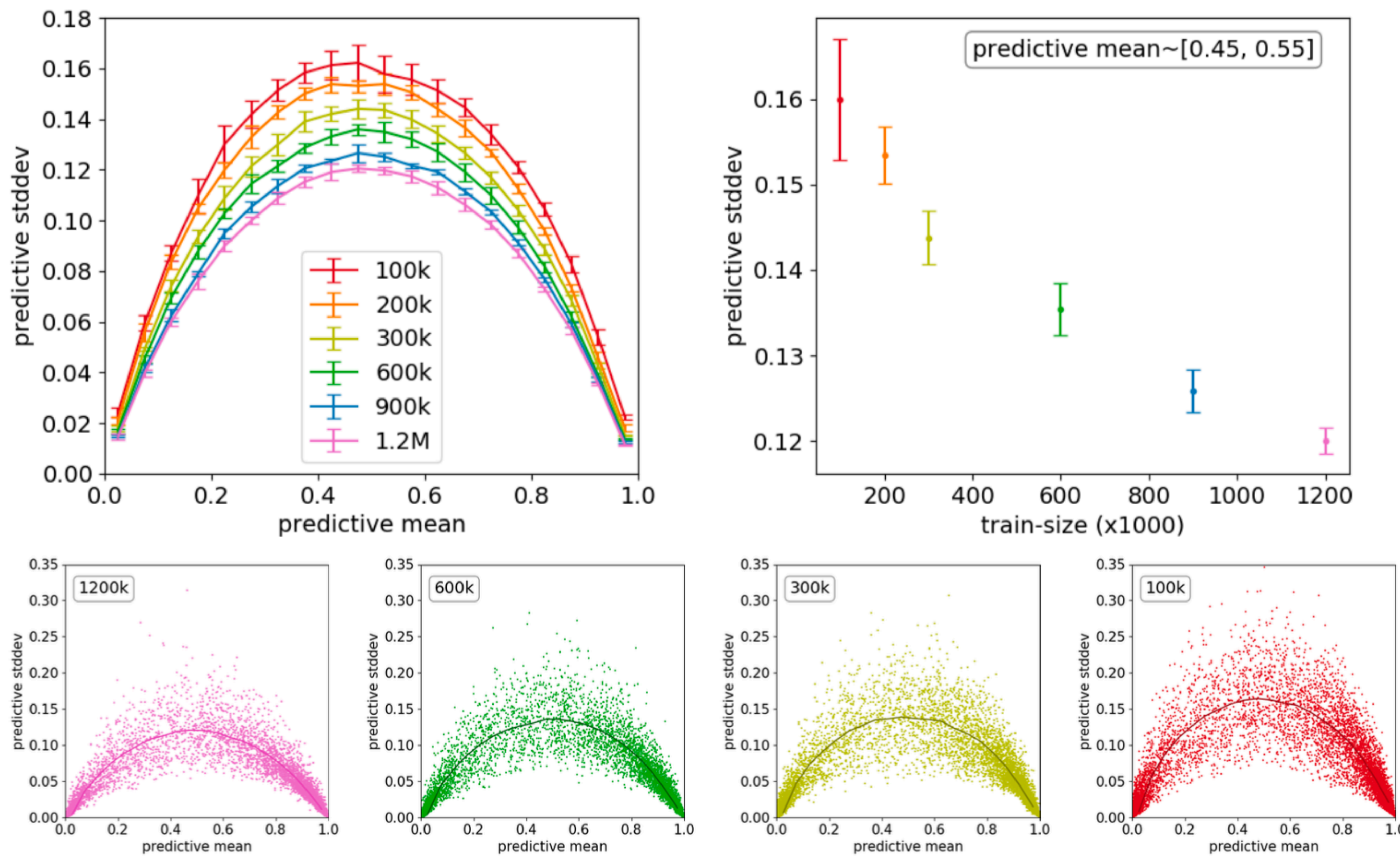
- So far discussed handling uncertainties on the inputs
- How can we with training dat not fully covering the phase space?
- Sampling over Gaussian distribution for weights

Weigh

C Blundell et al, ICML Proc's 2015

Bayesian Cont'd

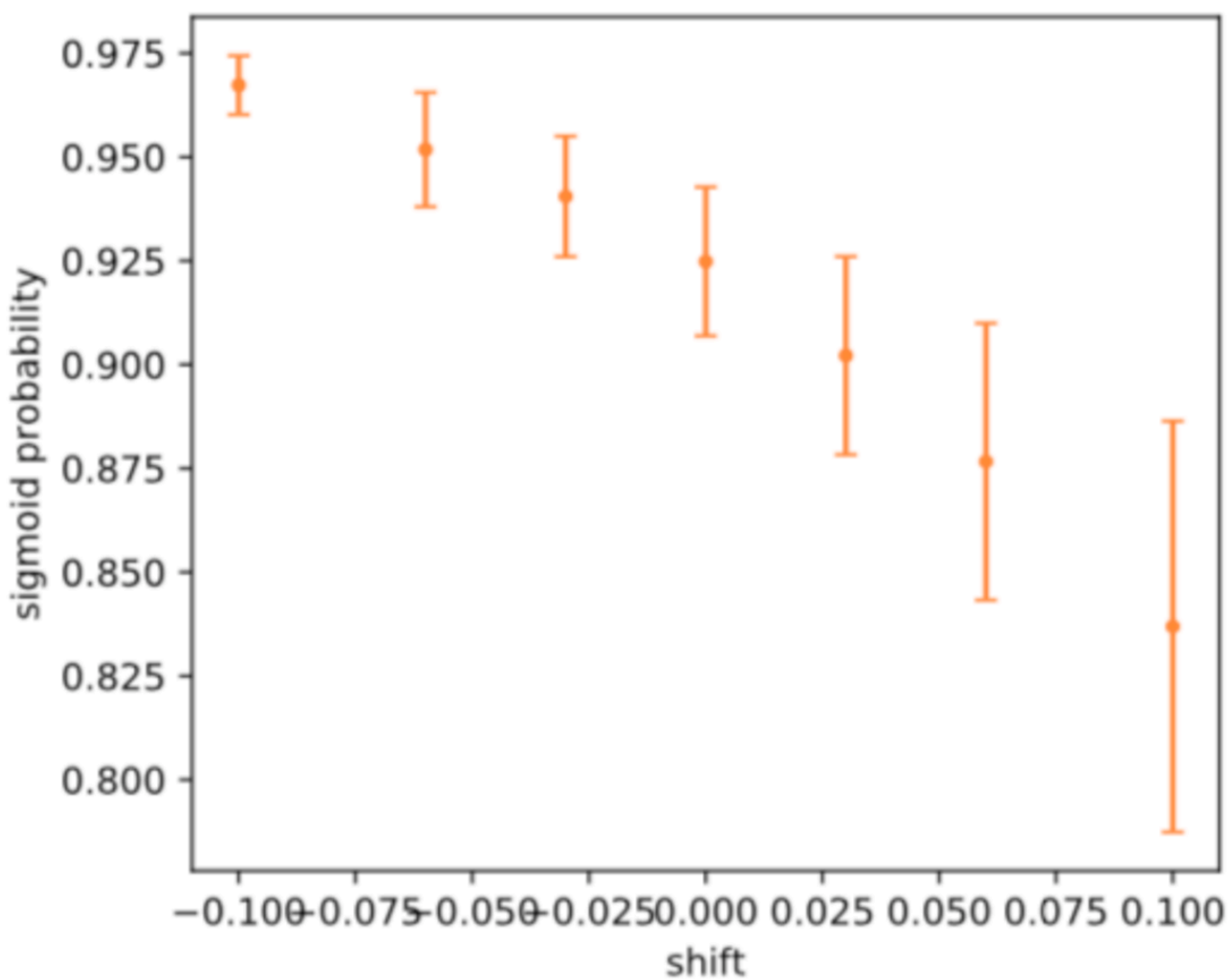
- learn classification output and uncertainty
- (60 ± 30)% top is very different from (60 ± 1)% top
- tagger calibration part of the network training
- for instance: effect of MC statistics



Plehn - Aspen 2019
Bollweg, Haussmann, GK, Luchmann, Pehn, Thompson: Upcomimg

Preview

- First look at impact of Jet Energy Scale uncertainty
- Only rescale leading subset

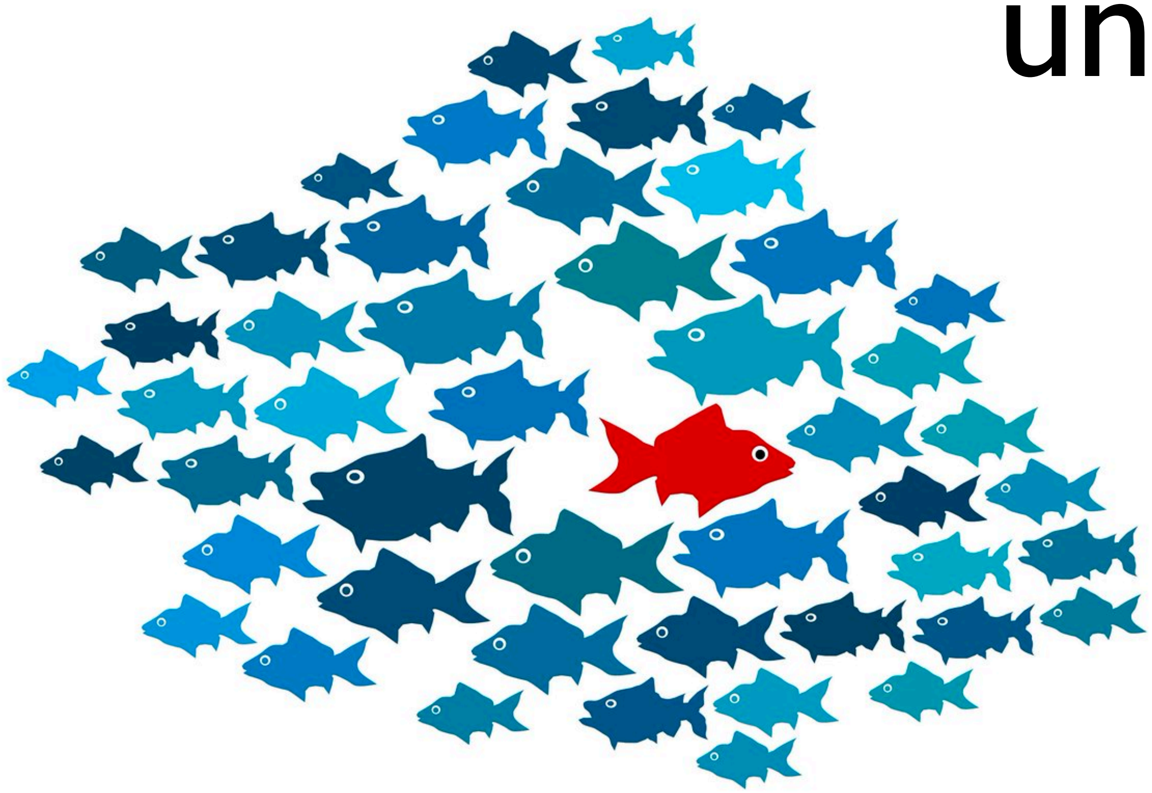


Bollweg, Haussmann, GK, Luchmann, Pehn, Thompson: Upcomimg

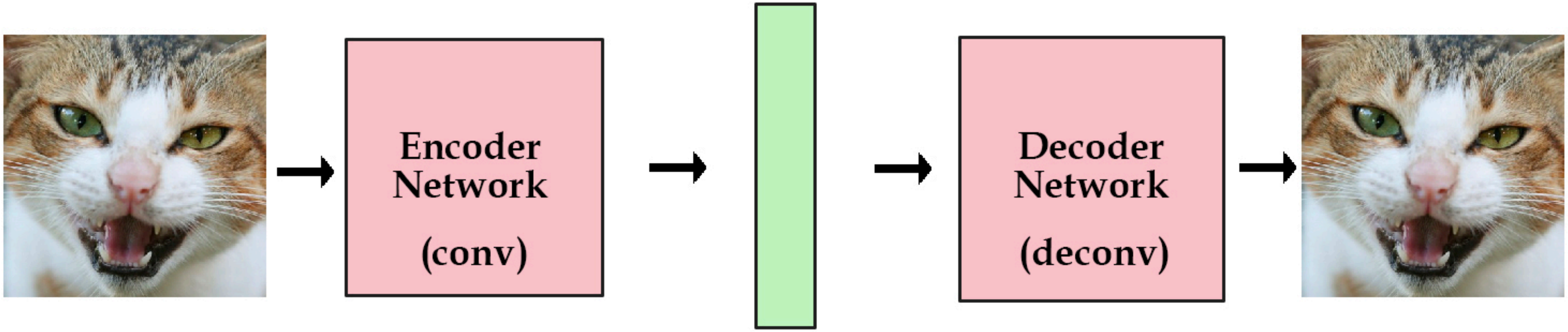
Look for unknown signatures

Can we look for new physics,
without knowing what to look for?

Can we avoid systematic
uncertainties in searches?



Autoencoder

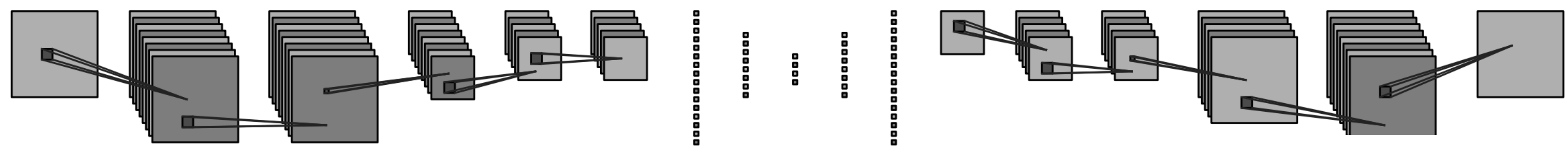


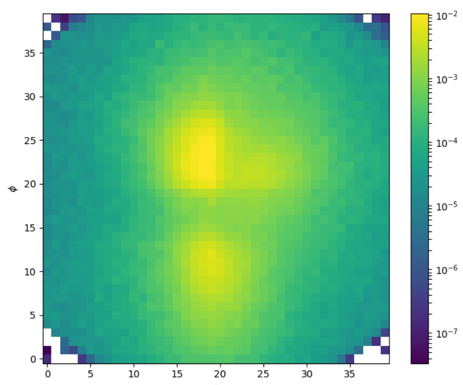
$$L = (\hat{y} - g(f(x)))^2$$

- Self-supervised learning
- *Latent space/bottleneck* with compressed representation
- Dimension reduction
- Denoising

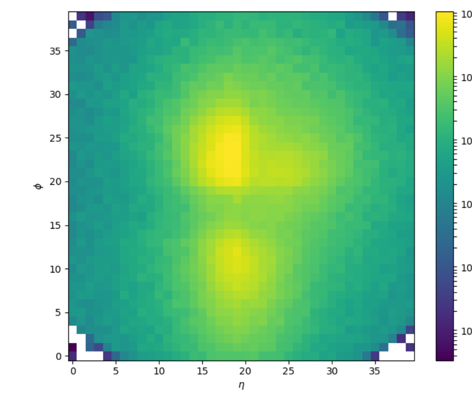
kvfrans
deeplearningbook.org

Autoencoder for Physics

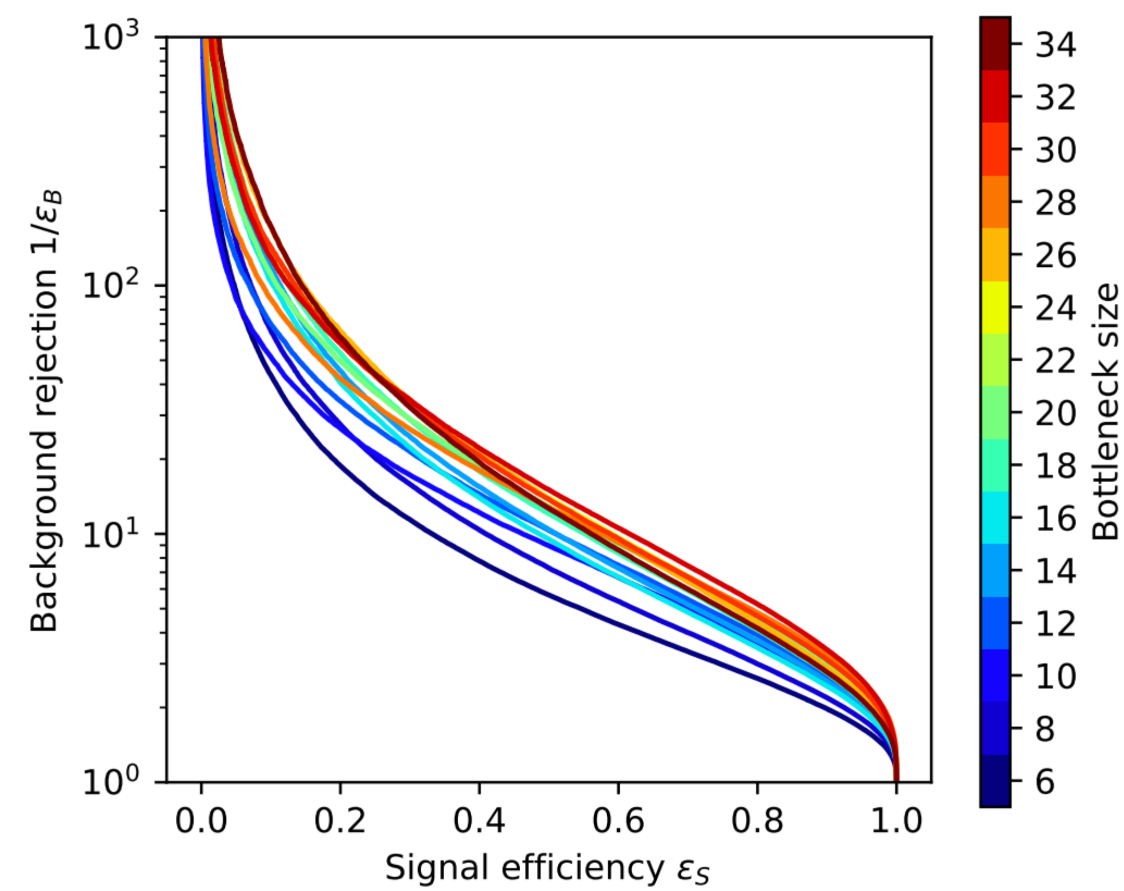




$$L_{\text{auto}} = \sum_{1600 \text{ pixels}} \left(k_T^{\text{norm,in}} - k_T^{\text{auto}} \right)^2$$



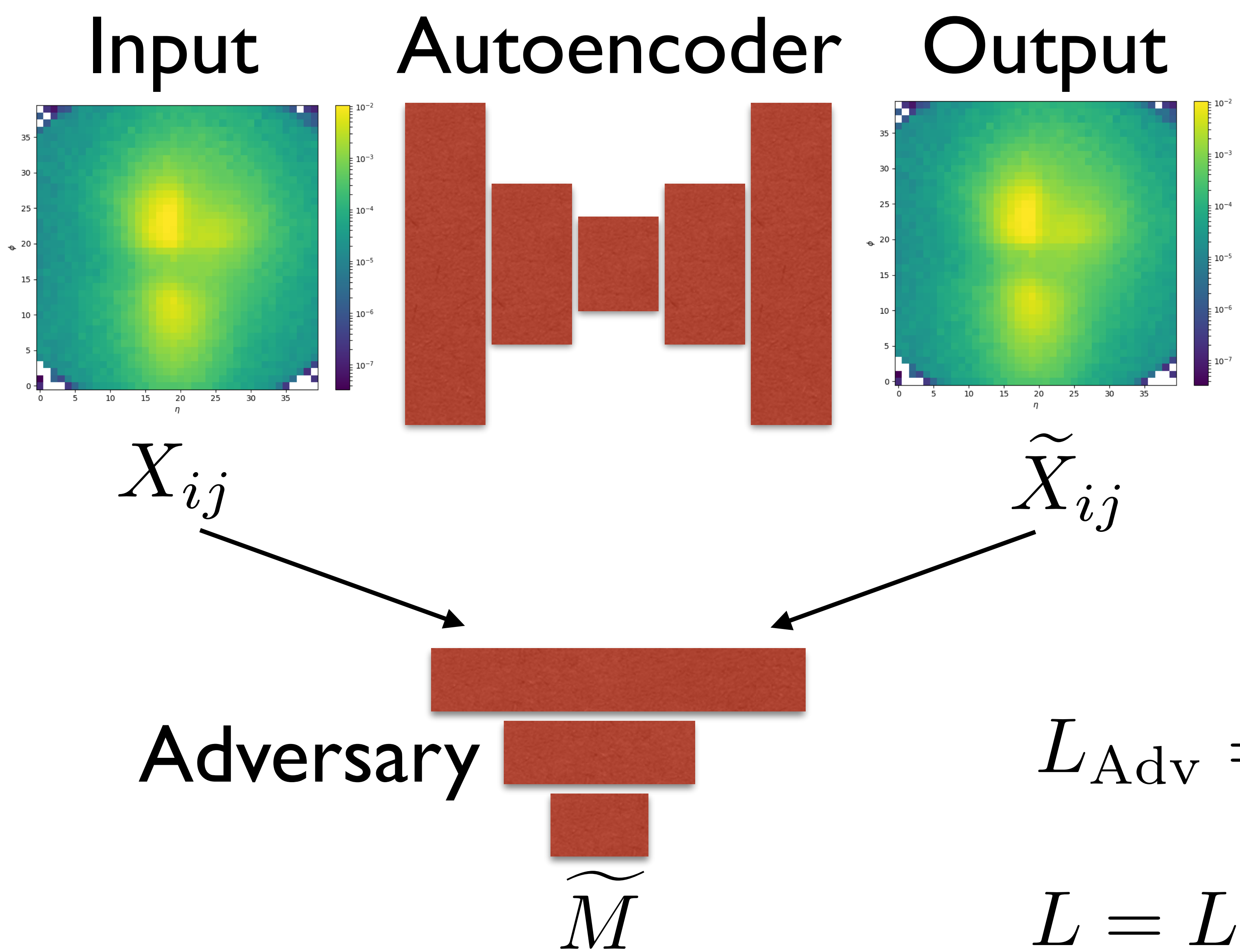
- Can we find new physics without knowing what to look for?
- Train on pure QCD light quark/gluon jets and apply to top tagging
- Top quarks/ new physics identified as anomaly



QCD or What?
T Heimel, GK, T Plehn, JM Thompson, 1808.08979
Searching for New Physics with Deep Autoencoders
M Farina, Y Nakai, D Shih, 1808.08992

Combined Setup

- Autoencoder alone will also learn mass distribution
- Counteract with adversary:



$$L_{\text{Auto}} = \sum_{\text{Pixels } ij} \left(X_{ij} - \widetilde{X}_{ij} \right)$$

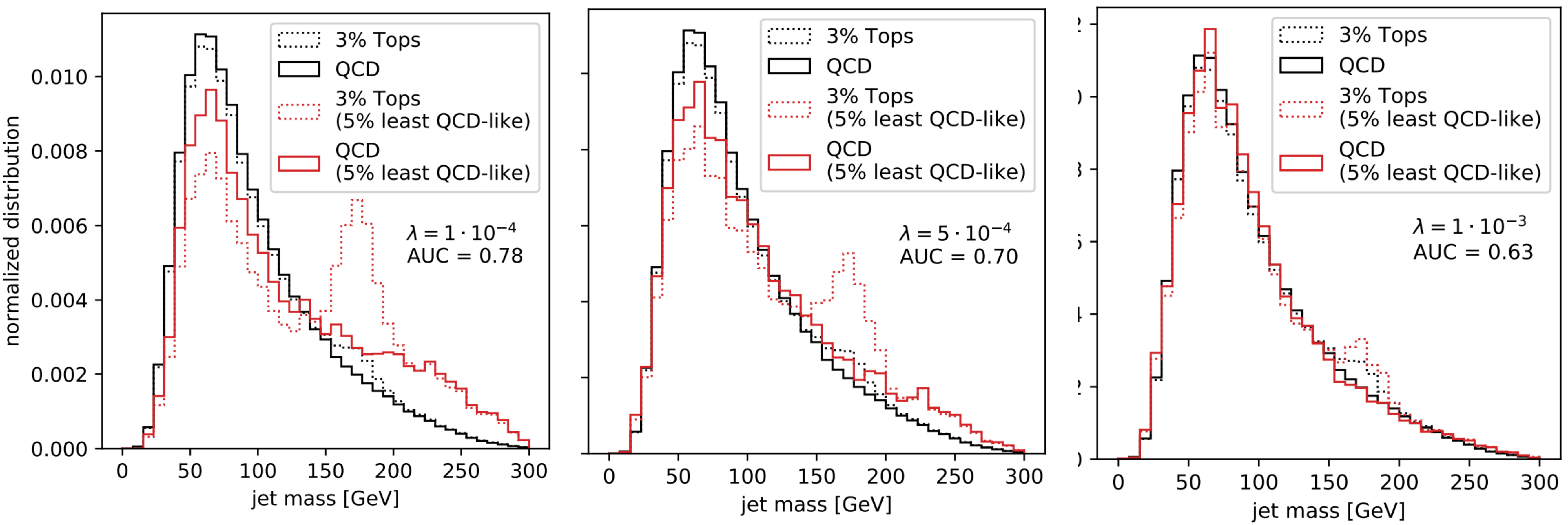
$$L_{\text{Adv}} = \text{CCE}\left(M, \widetilde{M}(X_{ij} - \widetilde{X}_{ij}) \right)$$

$$L = L_{\text{Auto}} - \lambda L_{\text{Adv}}$$

Mass Sculpting

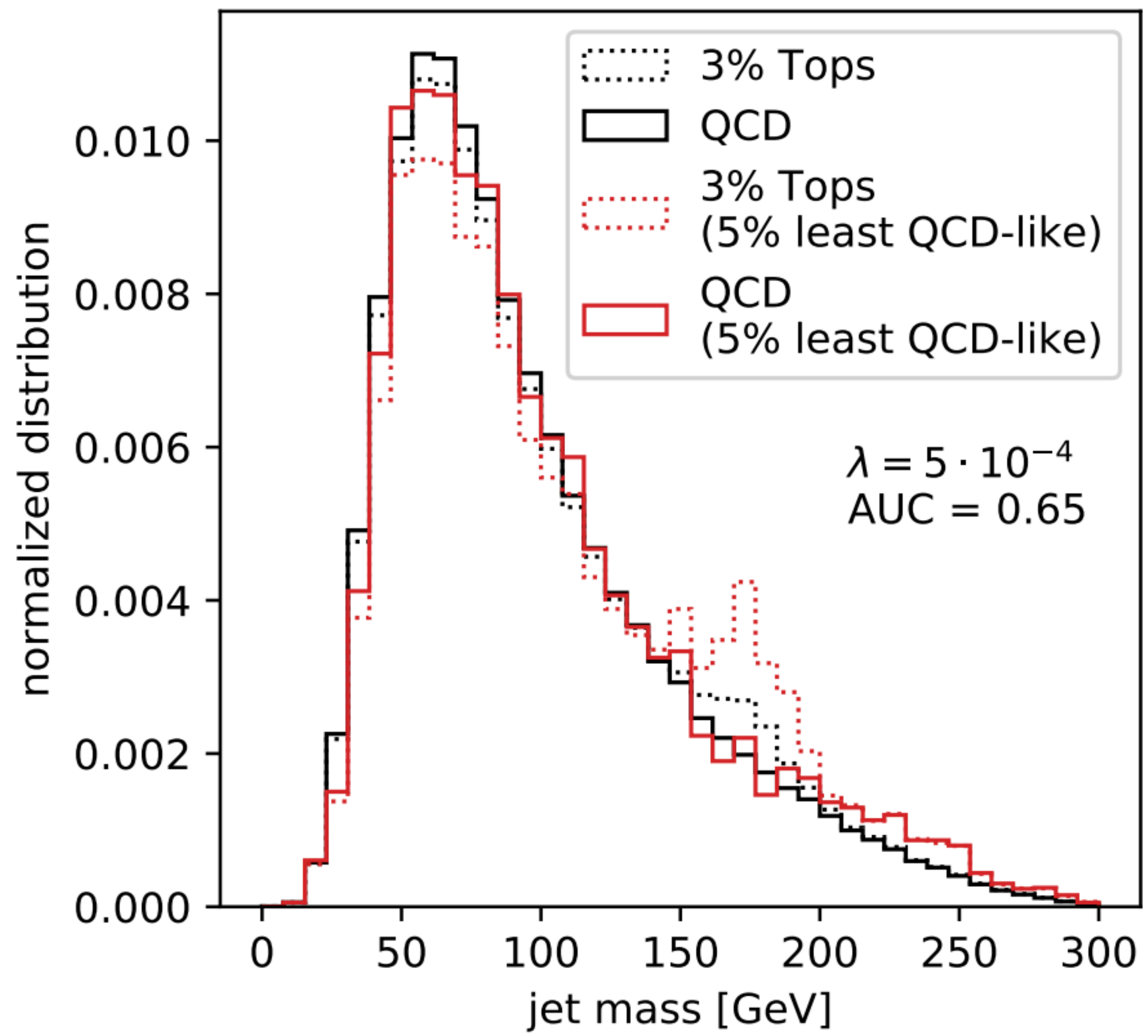
- Tune mass dependency with Lagrange multiplier:

$$L = L_{\text{auto}} - \lambda \, L_{\text{adv}}(M)$$

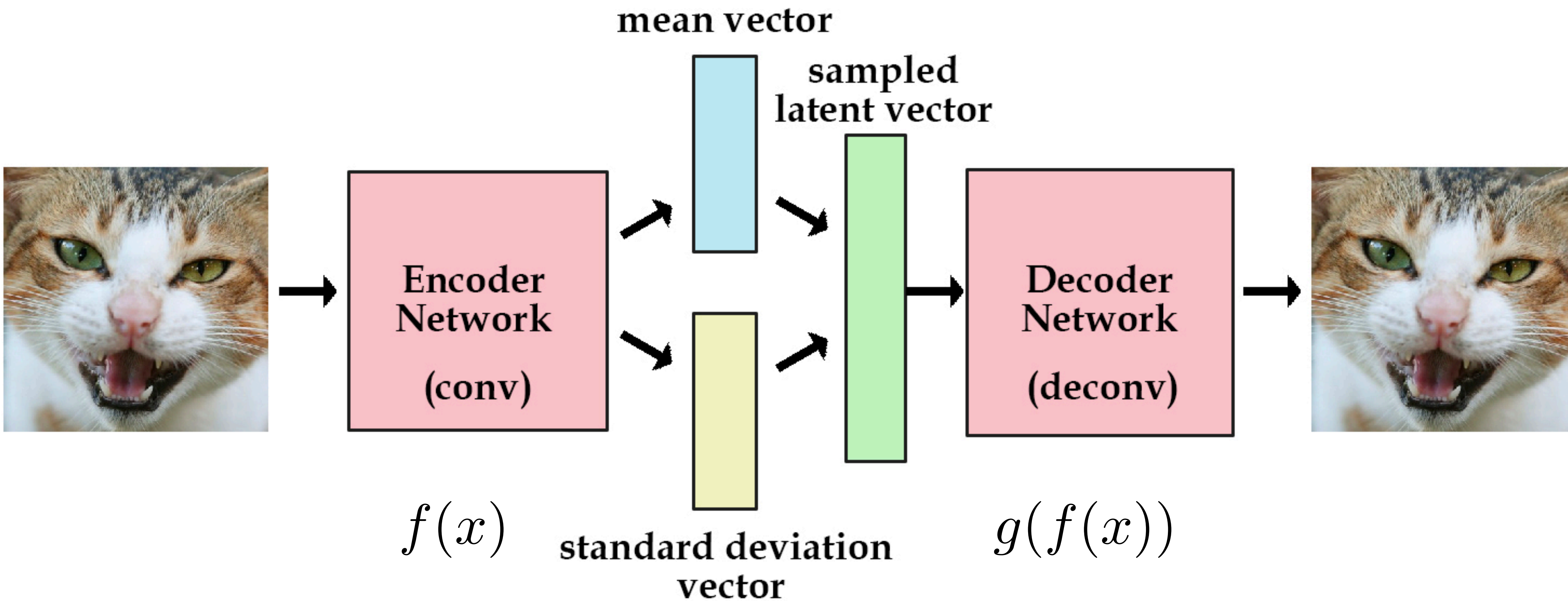


Signal contamination

- Procedure works also when signal is present in training data
- We now have a versatile tool to search for new physics (anomalies) in a purely data driven and unsupervised way
  - Apply to LHC collision data!
  - Potential trigger!



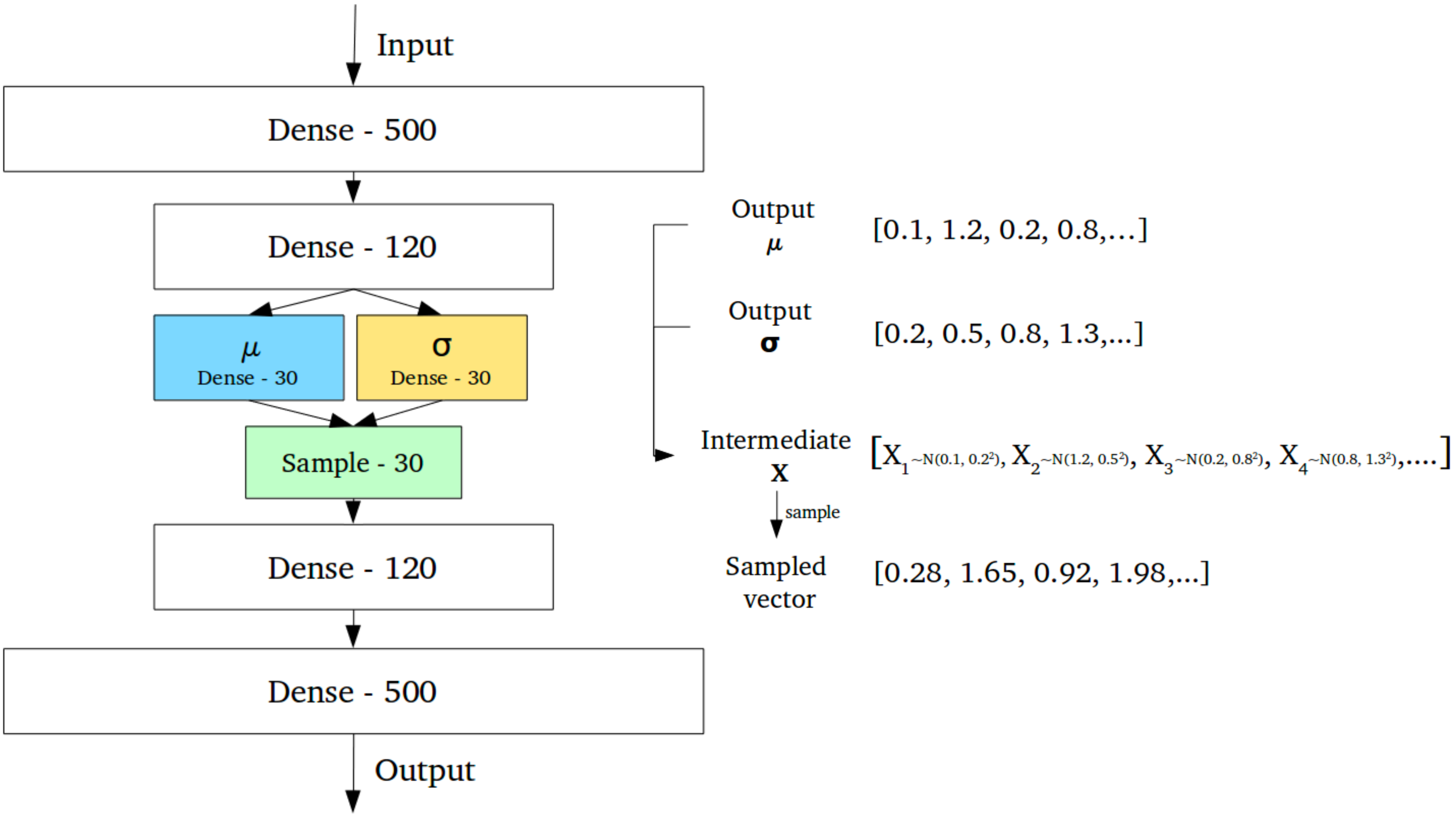
Variational Autoencoder



- We want to sample from latent space
- Split into mean and standard deviation
- Add penalty term (Kullback-Leibler divergence) so mean/std are close to unit Gaussian

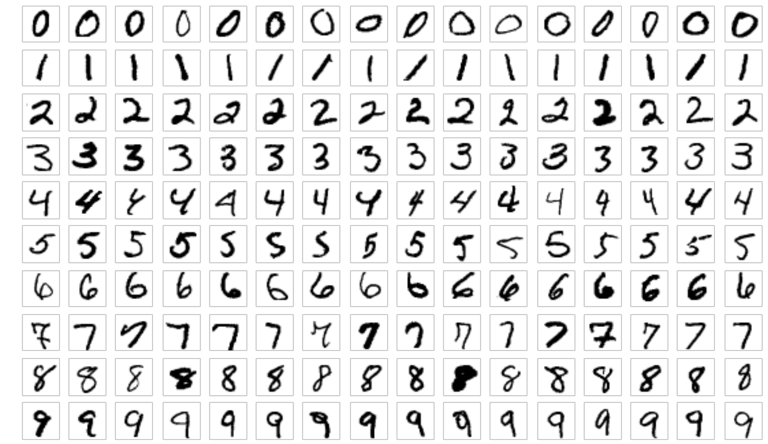
kvfrans
towardsdatascience.com

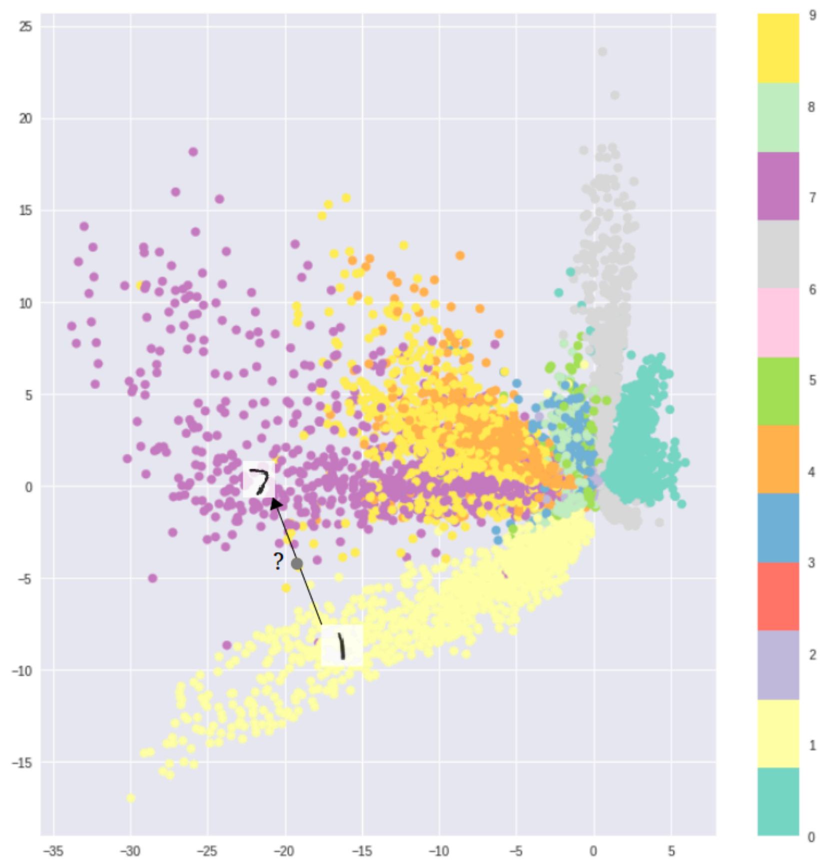
Concrete



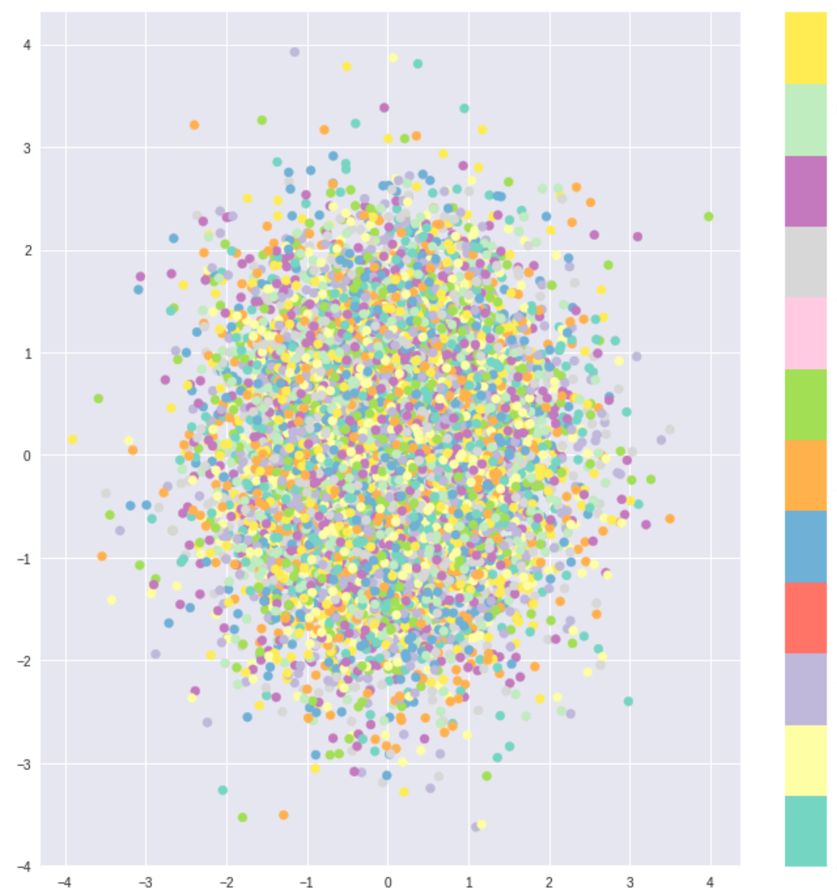
towardsdatascience.com

Variational Autoencoder

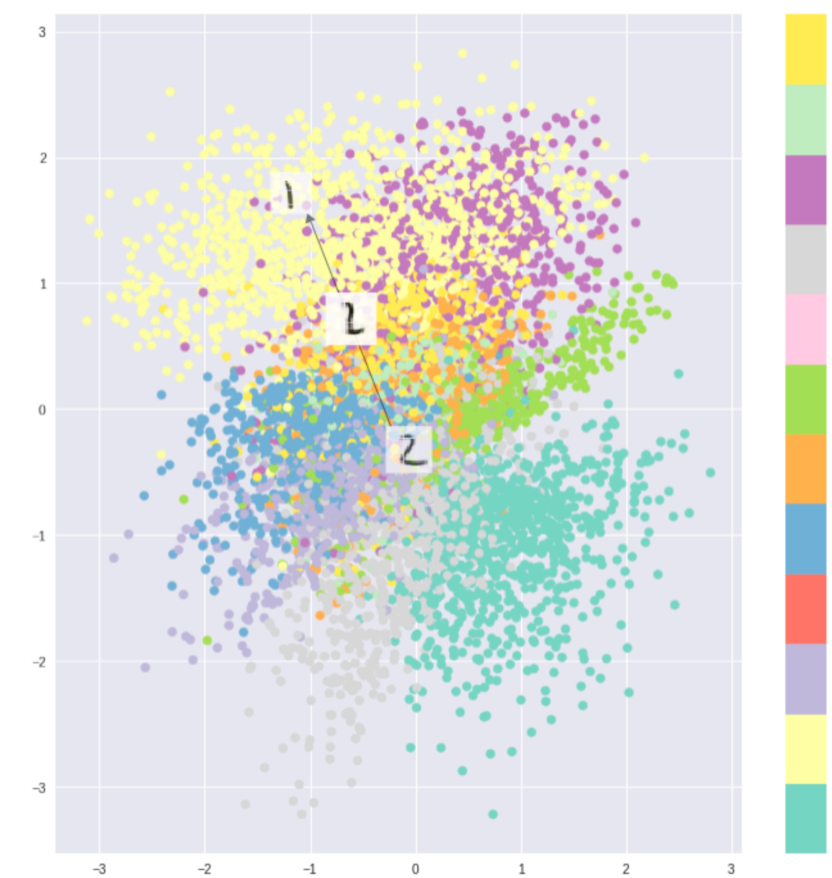




Reconstruction Loss Only



KL Loss Only

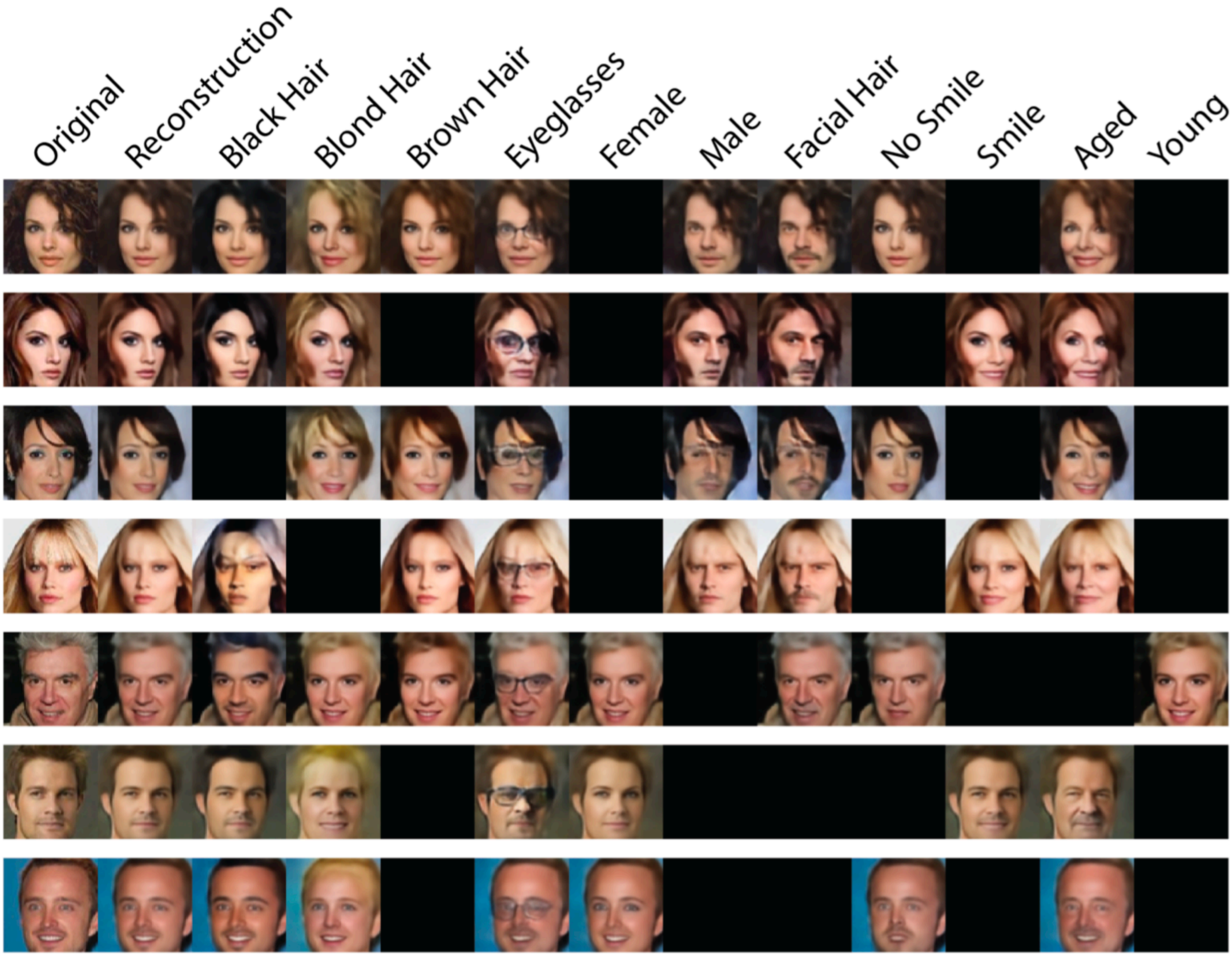


Combined Loss

$$L = (\hat{y} - g(f(x)))^2$$

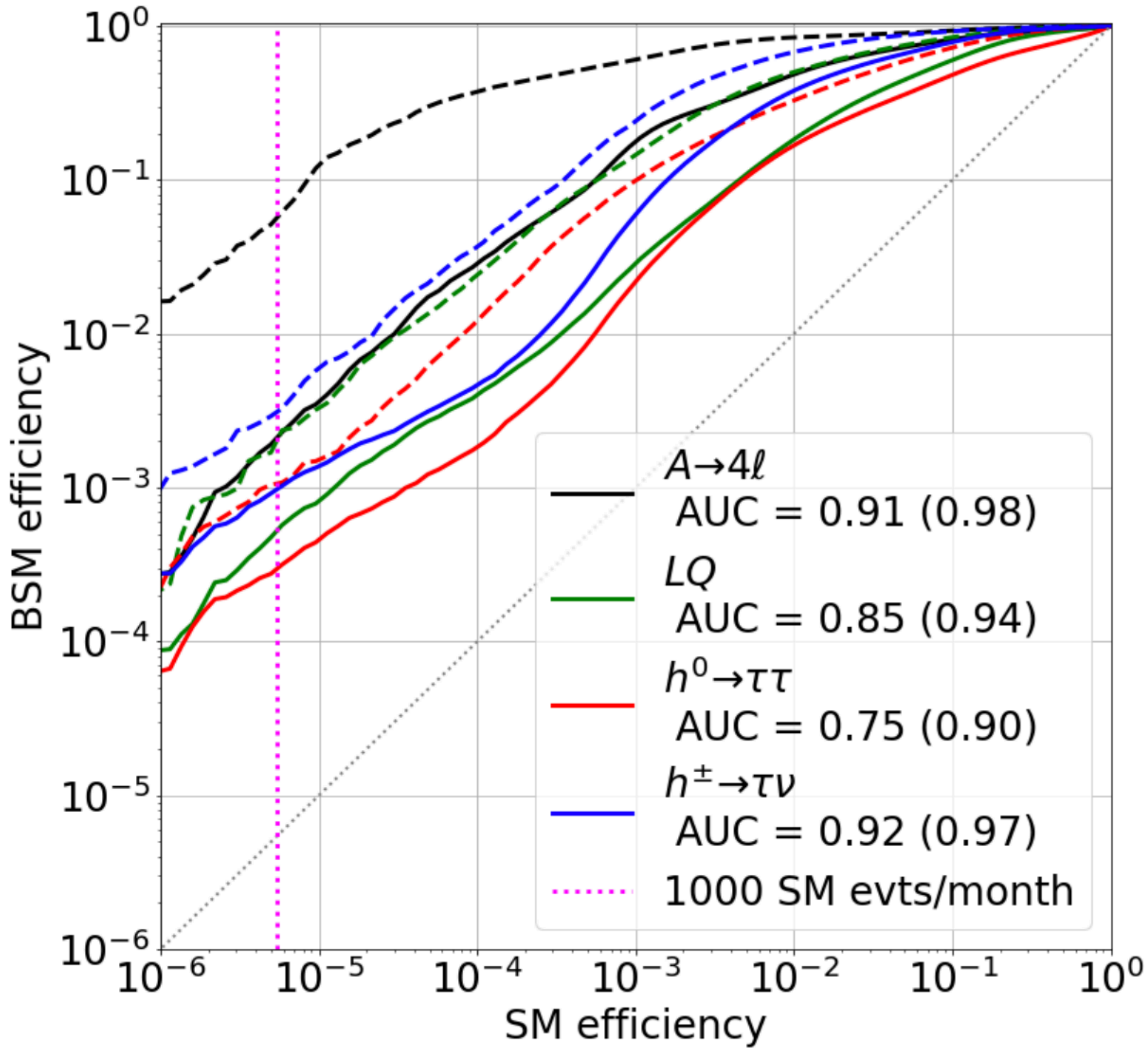
$$\sum_{i=1}^{n} \sigma_i^2 + \mu_i^2 - \log(\sigma_i) - 1$$

kvfrans
towardsdatascience.com



Latent Constraints: Learning to Generate Conditionally from Unconditional Generative Models, J Engel, M Hoffman, A Roberts, 1711.05772

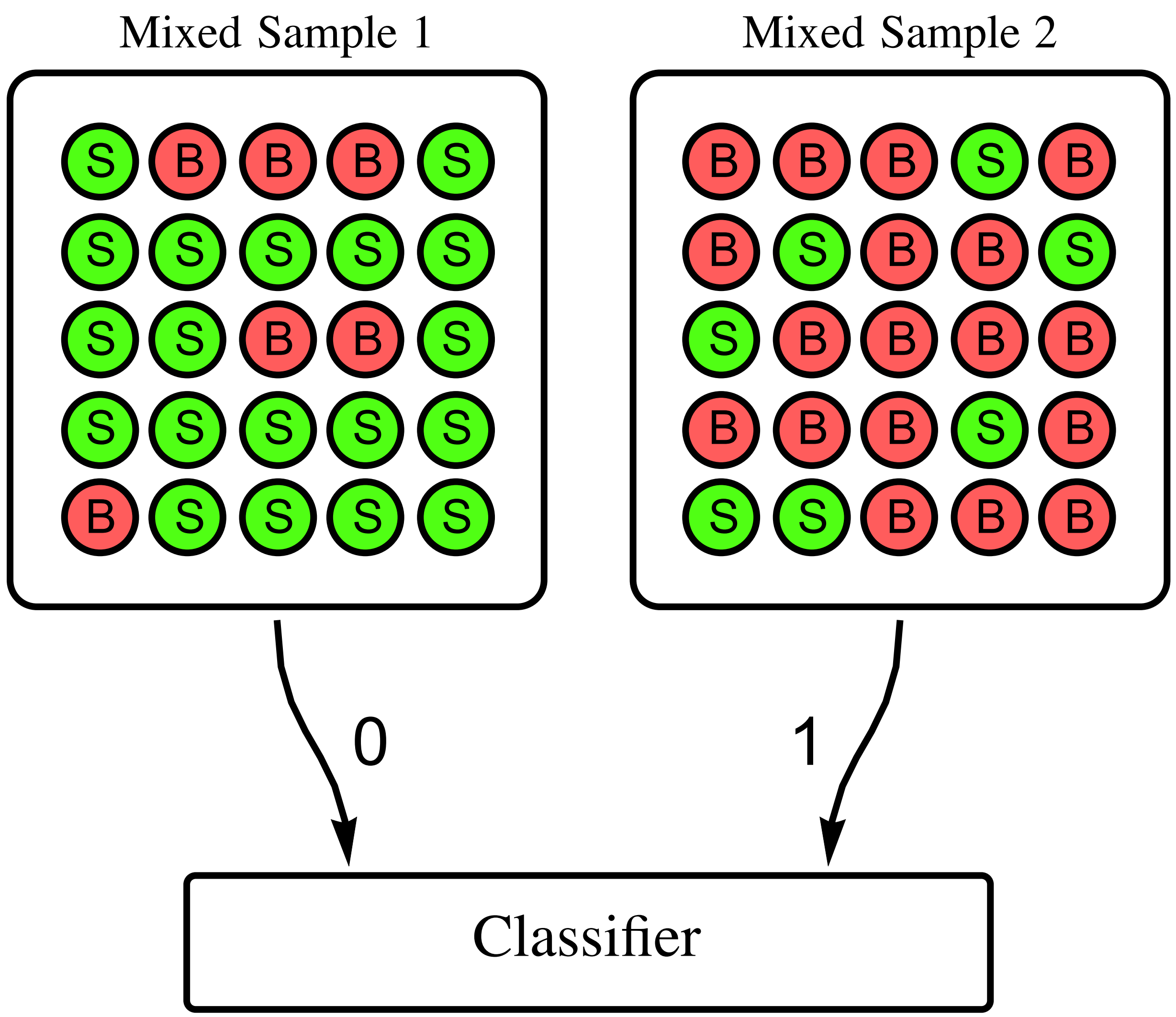
Extend to event tagging

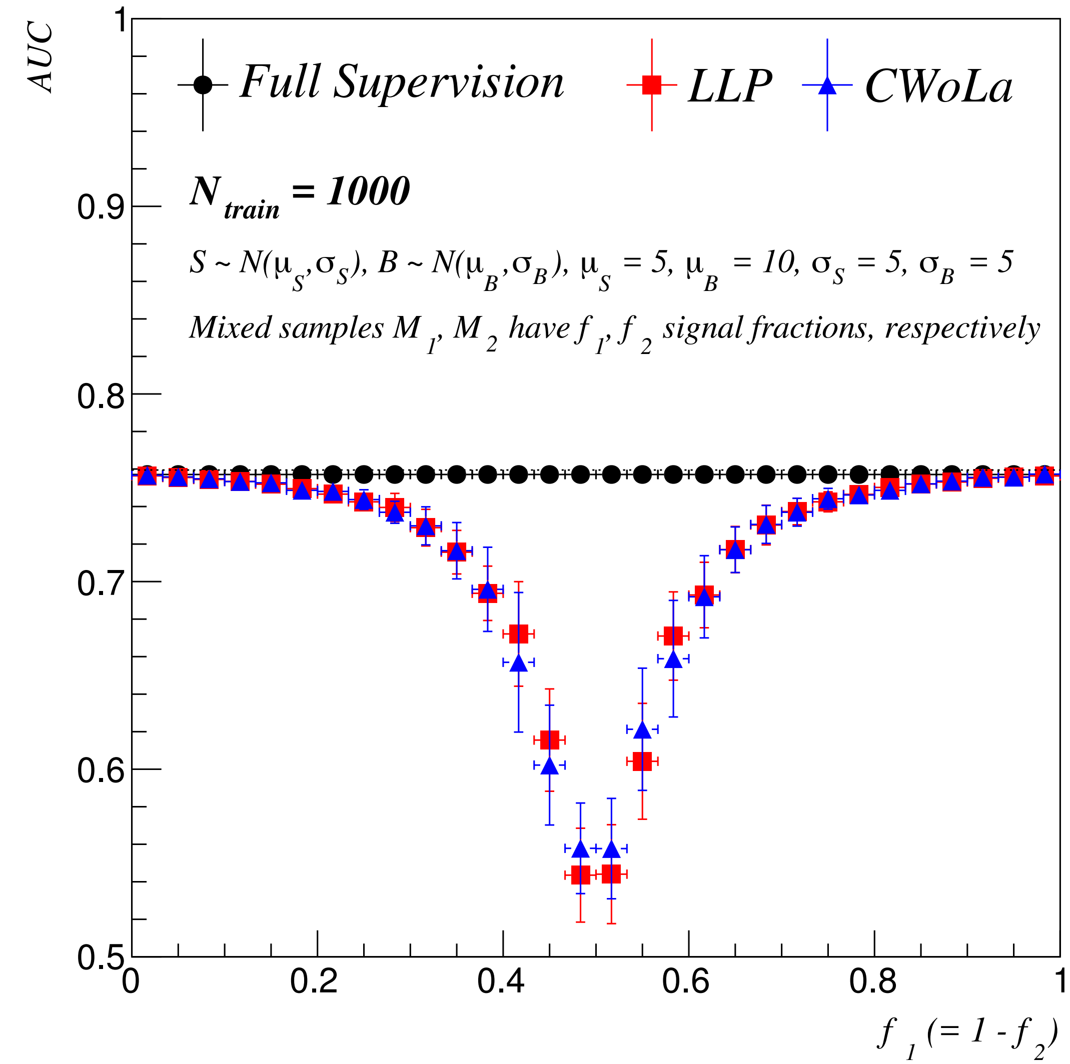


- Variational auto encoder
- Train on SM event cocktail
- Test sensitivity to different BSM models
- Opportunity for future trigger?

Variational Autoencoders for New Physics Mining at the Large Hadron Collider
O. Cerri et al, 1811.10276

Alternative: CWoLa



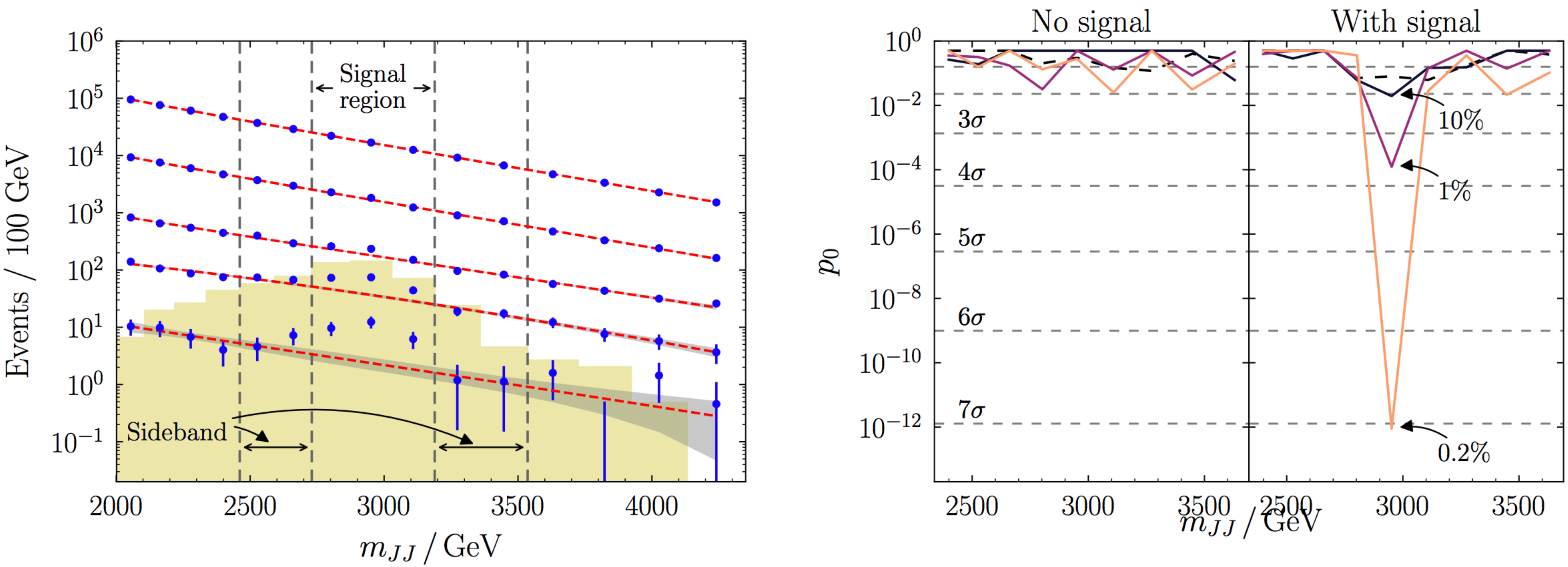


$$L_{M_1/M_2} = \frac{p_{M_1}}{p_{M_2}} = \frac{f_1\, p_S + (1-f_1)\, p_B}{f_2\, p_S + (1-f_2)\, p_B} = \frac{f_1\, L_{S/B} + (1-f_1)}{f_2\, L_{S/B} + (1-f_2)}$$

Weakly Supervised Classification in High Energy Physics
LM Dery, B Nachman, F Rubbo, A Schwartzman, 1702.00414
Learning to Classify from Impure Samples
PT Komiske, EM Metodiev, B Nachman, MD Schwartz, 1801.10158
Classification without labels: Learning from mixed samples in high energy physics, EM Metodiev, B Nachman, J Thaler, 1708.02949

Distinguishing mixed samples is equivalent to signal/background classification!

Alternative: CWola Hunting



- Assume signal is resonant in one variable
- Define signal region and sidebands
- Train classifier and look for excess

Anomaly Detection for Resonant New Physics with Machine Learning
JH Collins, K Howe, B Nachman
1805.02664

Spectrum of MC Reliance

Fully Supervised Learning

Fully Supervision, Mixed signals

Weak Supervision

CWoLa

Autoencoder

Dependence on Simulation

Assumes a classifier trained on MC will be correct on data as well

Assume MC gets the fractions of different classes right on average

Assume that we can tune the adversarial setup such that an anomaly based on data is credible.

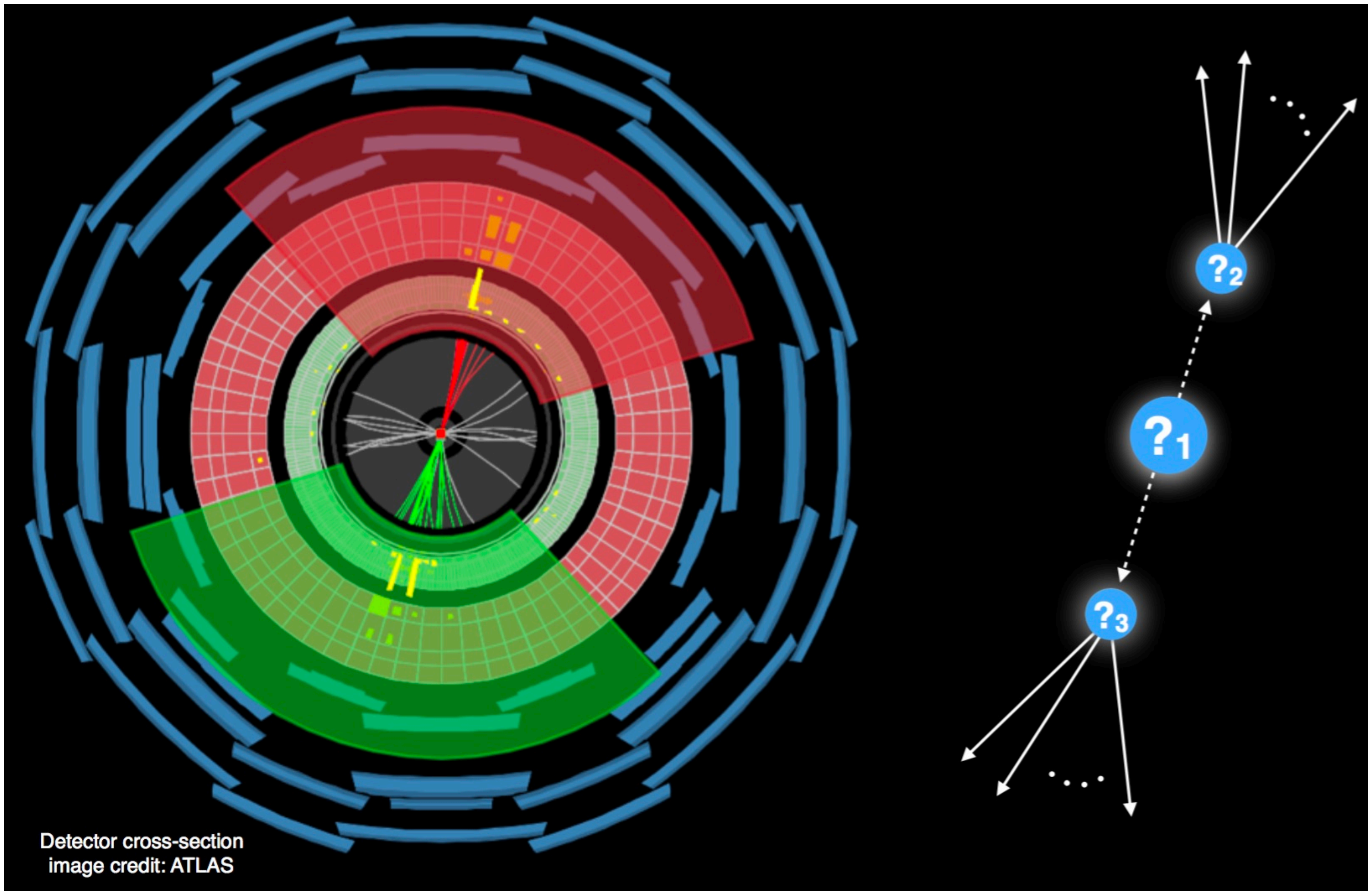
Assume simulation is good. Do not depend on specific model (but still limited by which models are in the mix)

Assume interpolation between phases spaces in data.

What else?

- Autoencoding is an exciting new tool in our box
- What else can we do?
  - Better understand what's going on in the latent space?
  - Event level?
  - Other things than searches?
  - …

LHC Olympics 2020



https://indico.cern.ch/event/809820/page/16782-lhcolympics2020

LHC Olympics 2020 Dataset

- Challenge to find new physics in simulated events
- Whatever approach you like - BUT we will not tell what the signal is
- Warm up phase: now until ~1.1.2020
  - 1M QCD events + Signal (labelled) available, single jet 1.3 TeV trigger
  - Use to develop methods to find new physics.
- Challenge Phase: ~1.1.2017 - ~15.1.2017:
  - Find new physics (yes/no, mass, x-sec) in the dataset
  - No labels provided
- At ML4Jets (15.1.-17.1.2020):
  - Identify winners, discuss strategies and write-up

Get the data here:

https://zenodo.org/record/2629073#.XKyG0-szbh9

Conclusions

- Deep learning and its applications to physics is a lively and exciting research area
- Images are a powerful tool to represent physics/detector/other information (although they are not perfect)
- Potential for new physics searches from:
    - Better reconstruction of known particles
    - Detection techniques for long-lived objects
    - New model-independent searches

Thank you!