Imperial College London

Data Science Institute

Deep learning in Alice and CMS

Markus Stoye on behalf of CMS and ALICE
Imperial College London, DSI

ICHEP, Seoul, 5th , July 2018

Machine learning applied in all aspects of Experiments

L1, DQM, HLT, reconstruction, analysis, interpretation

CMS CHEP talks
- “Fast Boosted Decision Tree inference on FPGAs for triggering at the LHC”
- “Convolutional Neural Network for Track Seed Filtering at the CMS HLT”
- “End-to-end Deep Learning Applications for Event Classification in CMS”

ICHEP:
- “Muon System Monitoring with ML”

9 month old CMS overview DS@CERN seminar

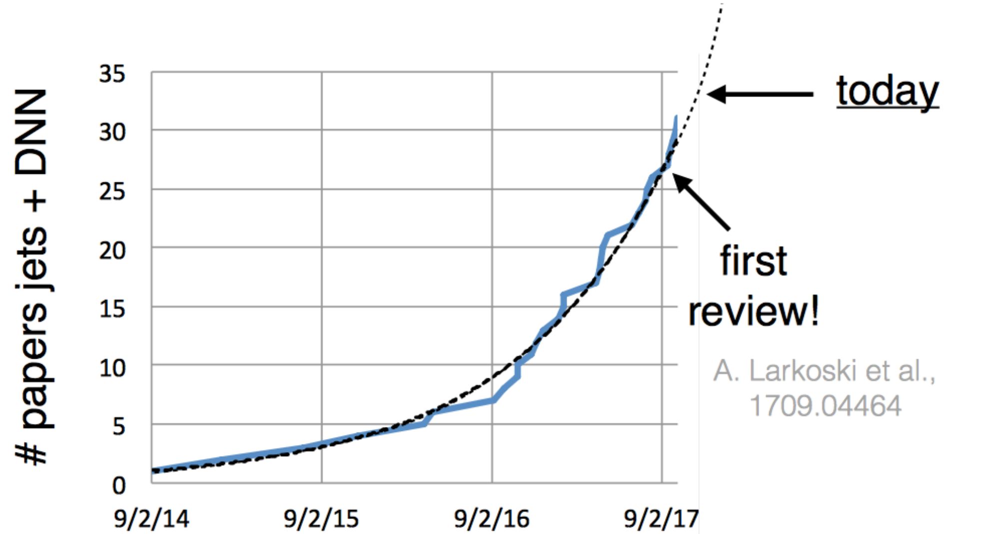
ALICE, recent overview talk

Resources:
- BDTs for $lambda_C$
- Low mass di-electron

IML: LPCC ML working group page

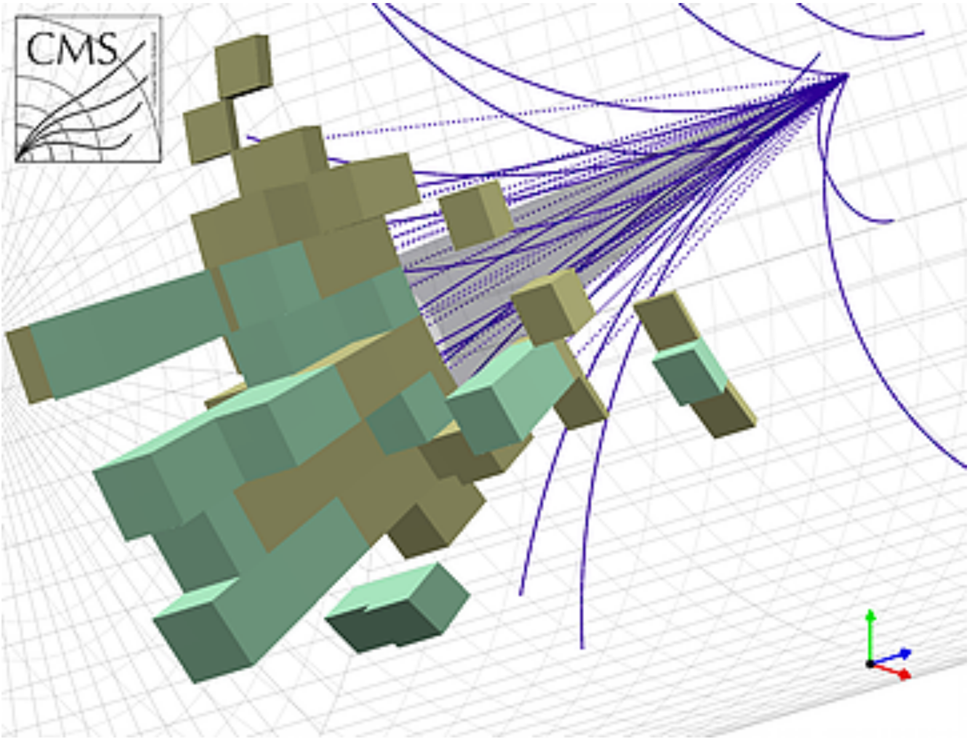
Will not provide extensive overview, but focus on a topic

Deep learning exponential growth



- The driving factor of the ai “boom” is deep learning and big data
- I will focus on deep learning success stories that are implemented in Alice and CMS software:
  - Jet tagging in Alice and CMS
  - Data quality monitoring/certification in CMS

Jet tagging: which parton was that?

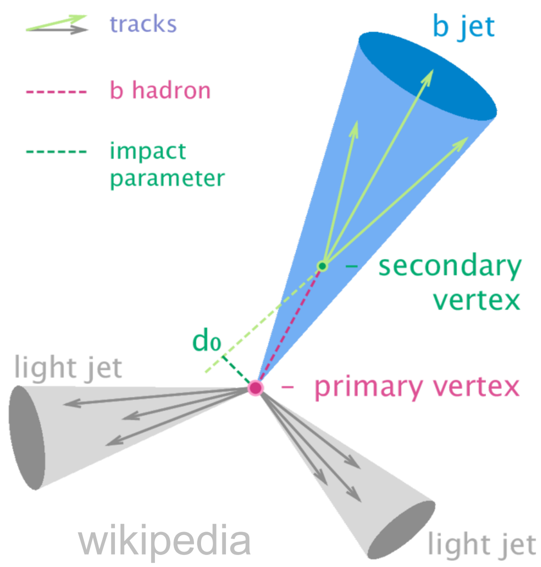


Each line and box represents many measurements

Most commonly used tagger: b-jet tagging

Key features:

- Displaced tracks from longer lifetimes of heavy flavor jets
- Secondary vertex
- Eventually leptons in jets from W* in b → W*c or c → W*s
- Slightly wider jets
- …



- Typically CMS jets have up to 50 particles with detailed information and secondary vertices ~1000 features

Traditional physicist ML

$1000 \rightarrow 200$

Design most discriminating particle-variables: How? Optical?

$200 \rightarrow 30$

Remove unnecessary particles: Which?

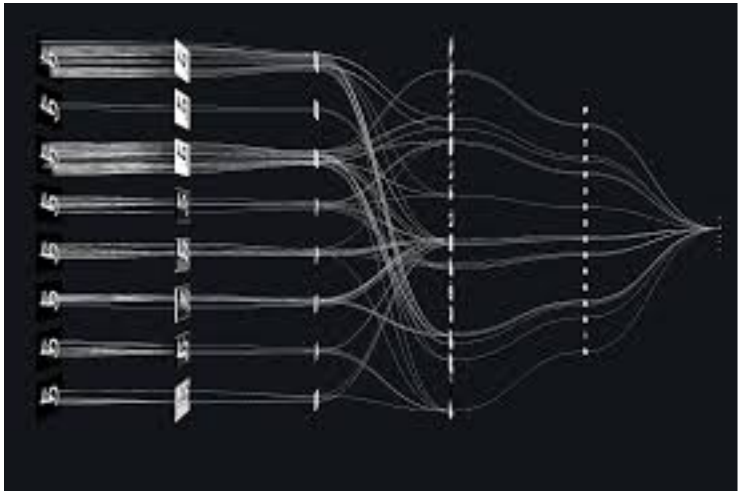
$30 \rightarrow 1$

Run shallow ML: Best performance

The traditional dimensionality reduction includes very difficult questions. Some danger of loosing valuable information.

Deep learning

$1000 \rightarrow 1$



Best performance

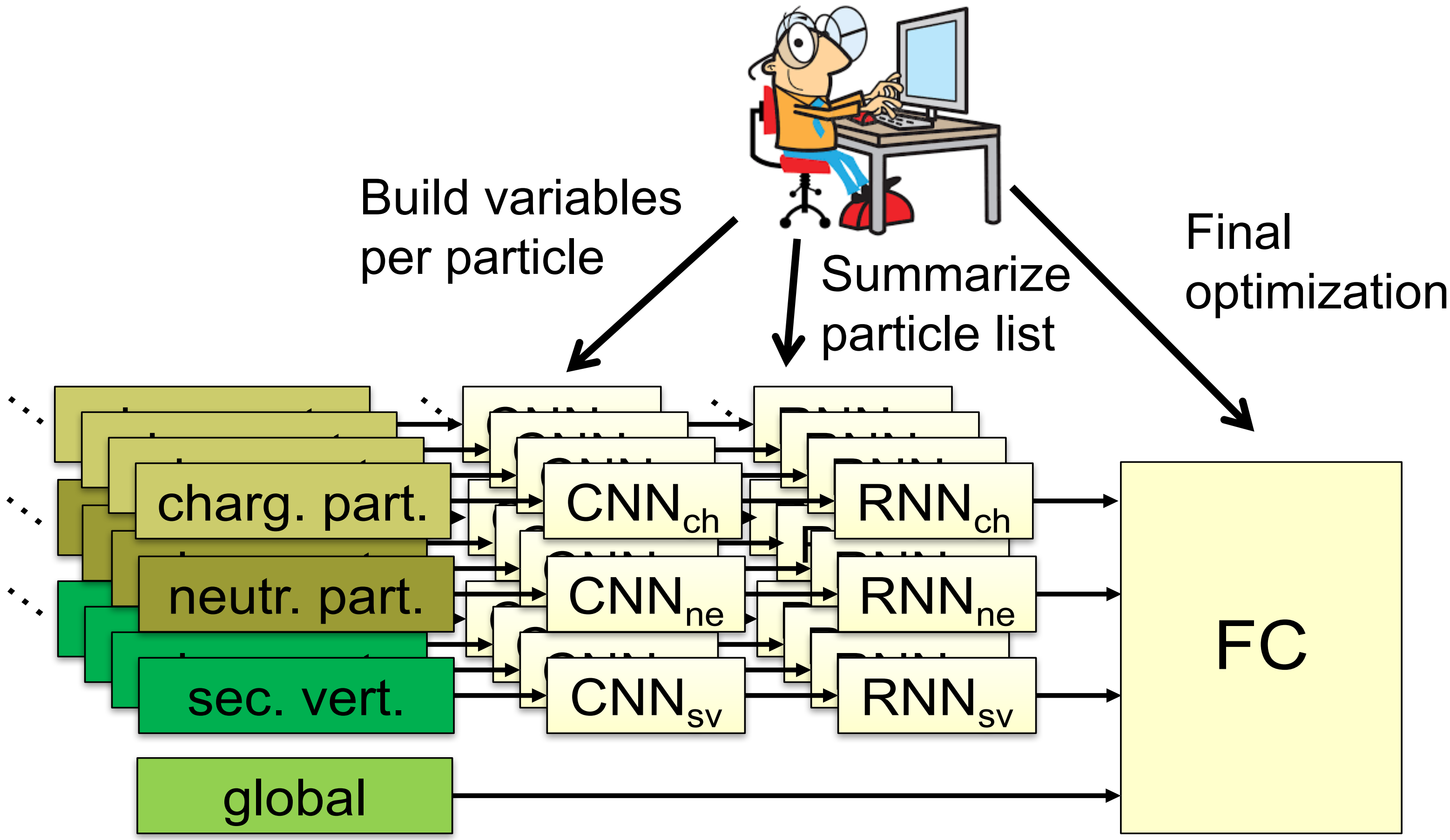
- Deep learning can deal with large input dimensions and reduces dimensionality directly for best performance
- The ***gain*** by deep learning depends on how much information was lost in traditional dimension reduction chain

CMS vs. state-of-the-art deep learning

	Training samples	Feat/sample	Model parameter	Samples per parameter
DeepJet	50 M jets	700	0.25 M	100:1
Images	1 M images	0.5 M	50 M	1:50

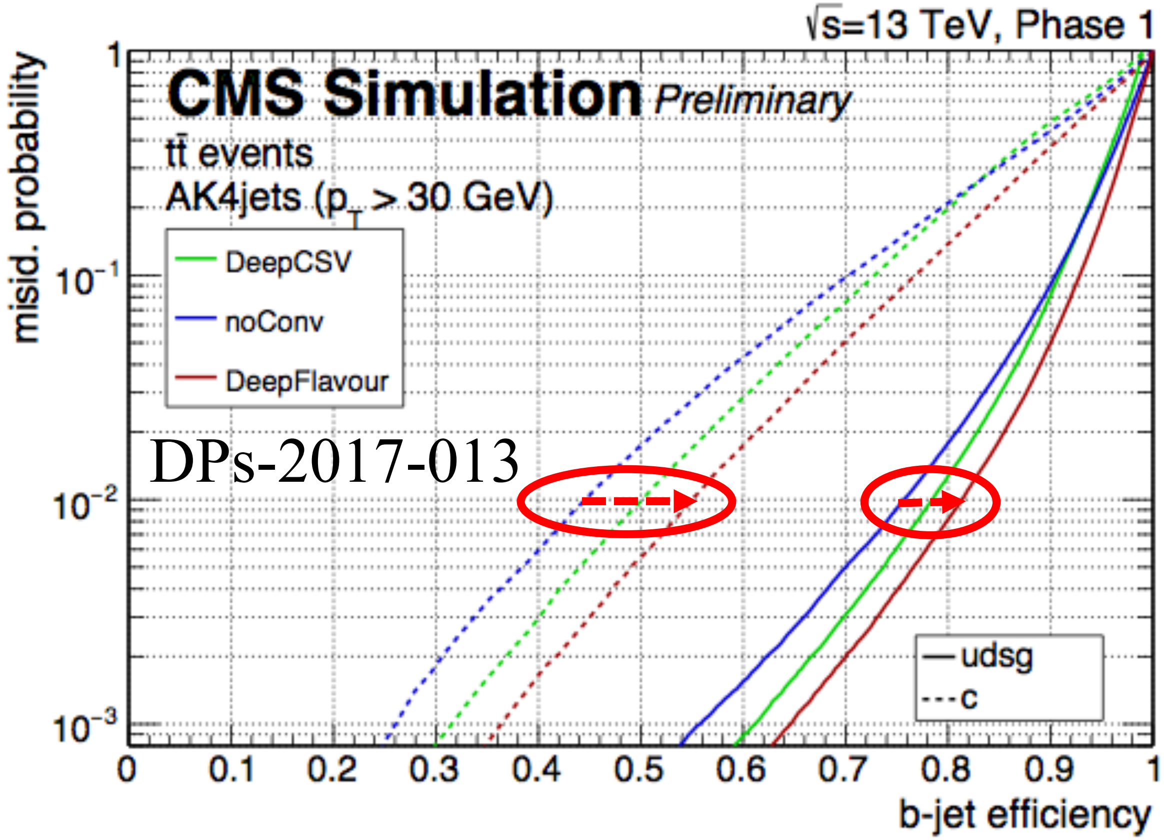
- For tagging we have more samples than model parameters, which is not the norm in deep learning
- Regularization comparably simple in such cases, to be kept in mind when building networks

Designing deep neural networks (DeepJet)



- Physics insights needed to design neural network architecture
- Particle and vertex based DNN as b-tagger for CMS

Impact of DeepJet/DeepFlavour architecture



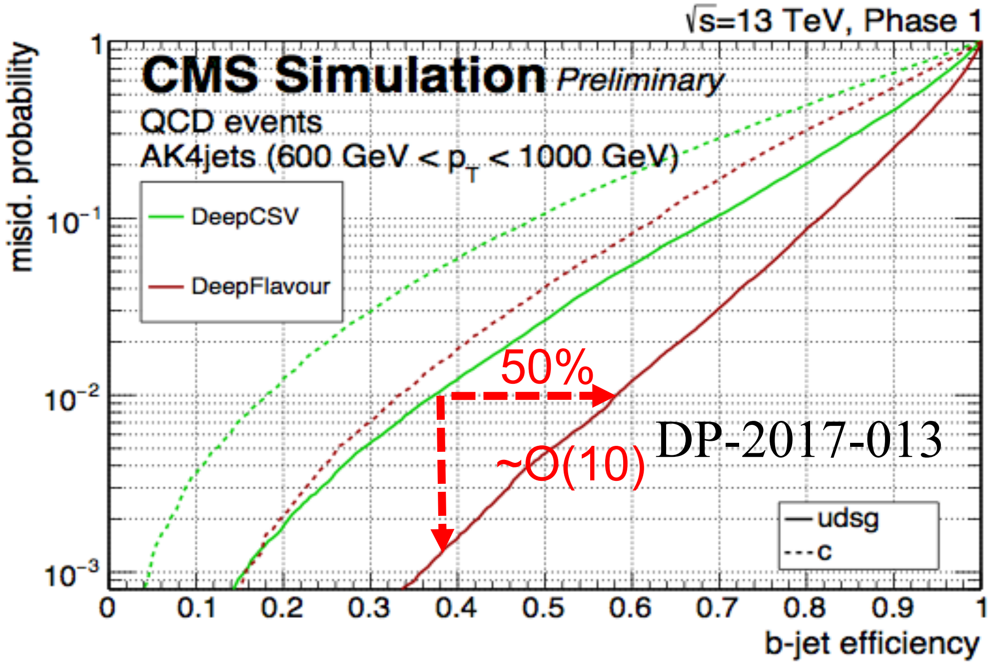
Blue: naive DNN (700 inputs)

Green: CMS tagger (~65 human made inputs)

Red: Physics inspired DNN (700 inputs)

Particle and vertex based DNN performs best

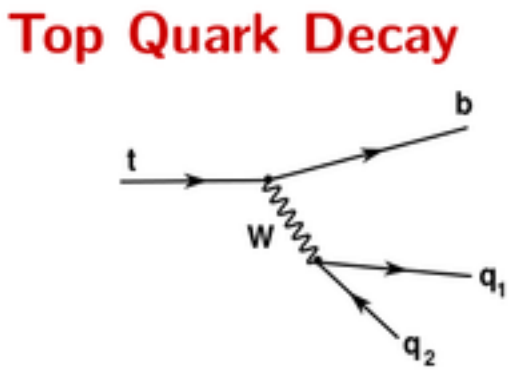
DeepJet results

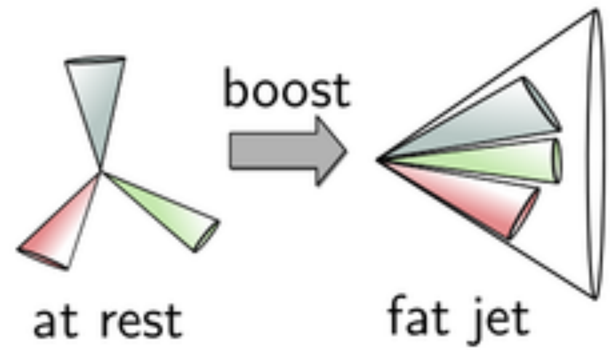


Very significant gain at high p_T

- Main loss of information was identified to be in the particle pre-selection
- Gain not yet confirmed in data, validation ongoing

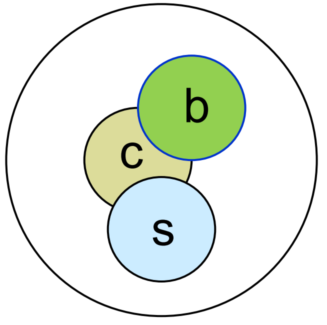
Fat jets





Key features of tops:

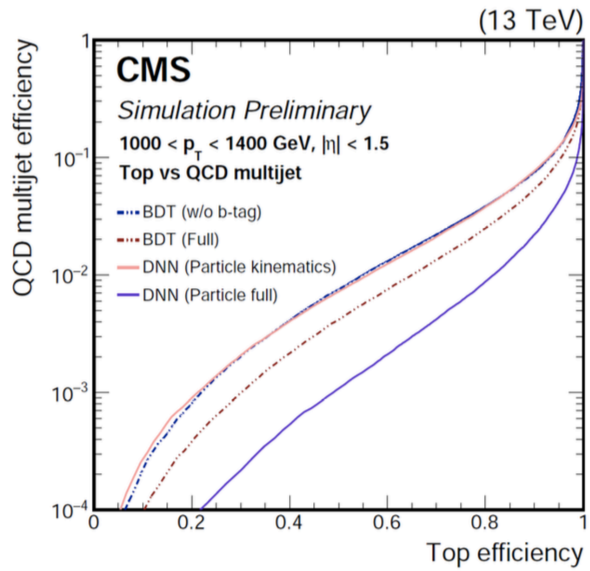
- Masses W, t, W polarization
- PU rejection
- 3 “prong”
- b-subjet and 50% with c-subjet



- Not obvious if these key features factorize or need to be addressed simultaneous.
- Potential gain by deep learning

Large cone jets for boosted objects

CMS-DP-17-049



- DNN=DeepJet (using all particles + vertices) with and without particle displacement information
- Modest gain w.r.t. state of the art features + BDT without particle displacement
- Factor 4 in background rejection for full information and deep learning

DeepJet multi-class classification

slim jets

Label	Sub-label
b	bb
	Leptonic b
	b
c	c
uds	uds
gluon	gluon

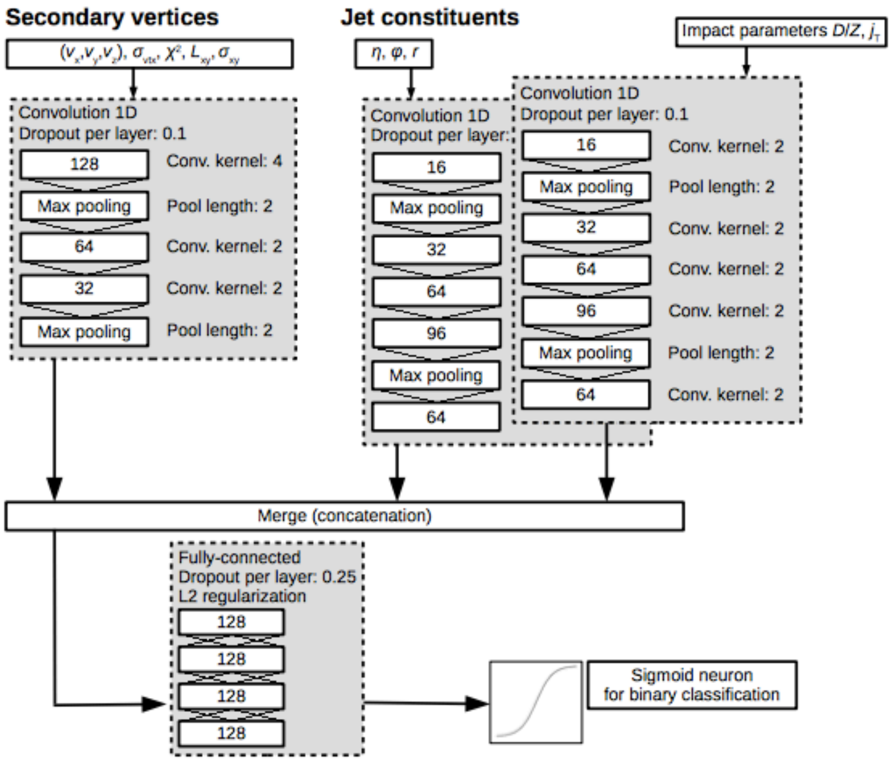
fat jets

Label	Sublabel
Higgs	H (bb)
	H (cc)
	H (VV*→qqqq)
Top	top (bcq)
	top (bqq)
	top (bc)
	top (bq)
W	W (cq)
	W (qq)
Z	Z (bb)
	Z (cc)
	Z (qq)
QCD	QCD (bb)
	QCD (cc)
	QCD (b)
	QCD (c)
	QCD (others)

- DeepJet classifies many categories simultaneously
- Red classes currently in data validation

Alice flavor tagging

arxiv.org:1709.08497

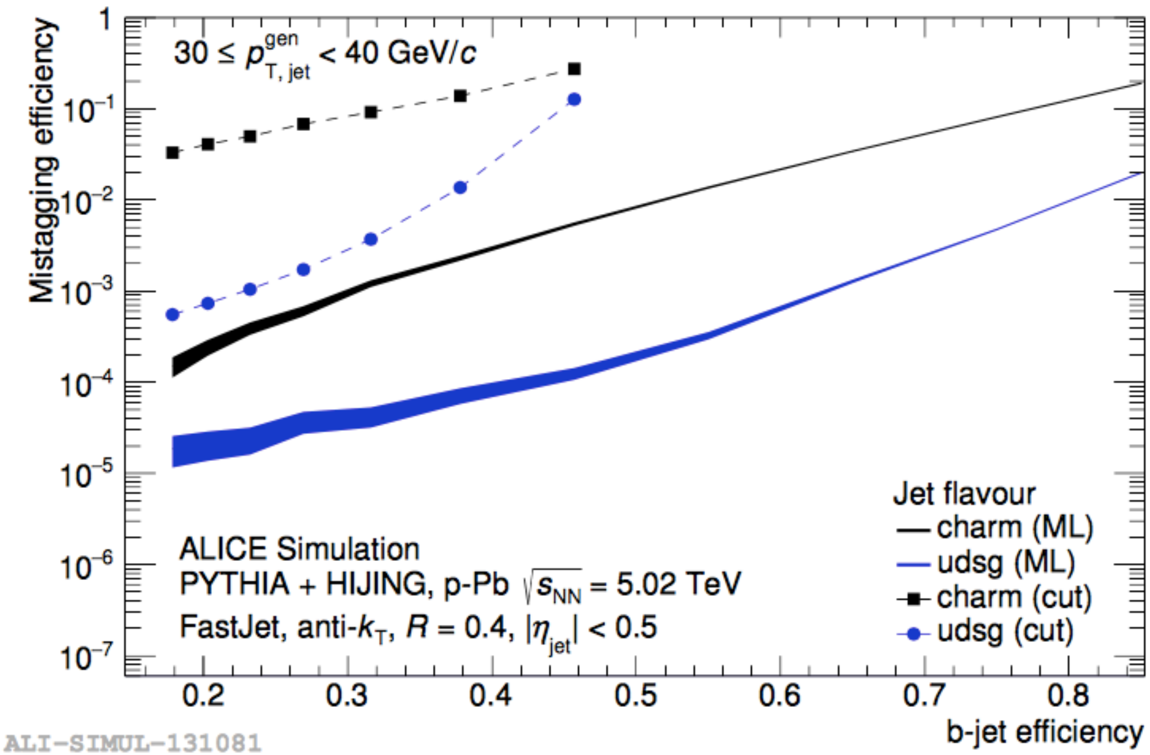


Deep learning architecture based on particles and vertices starting by convolutions on particles

CMS and Alice developed independently related ideas

Alice flavor tagging

arxiv.org:1709.08497



- Significant gain w.r.t. simple cut based approach
- b-tagging performance comparable to CMS

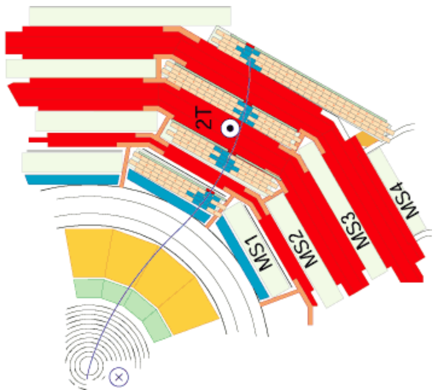
Data Quality Monitoring

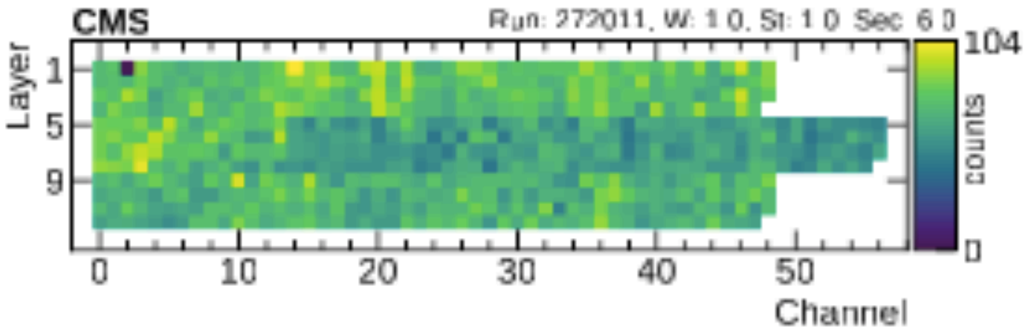
- Spot problems during data taking to react
- Certify data quality offline

Online data quality monitoring in CMS (DQM)

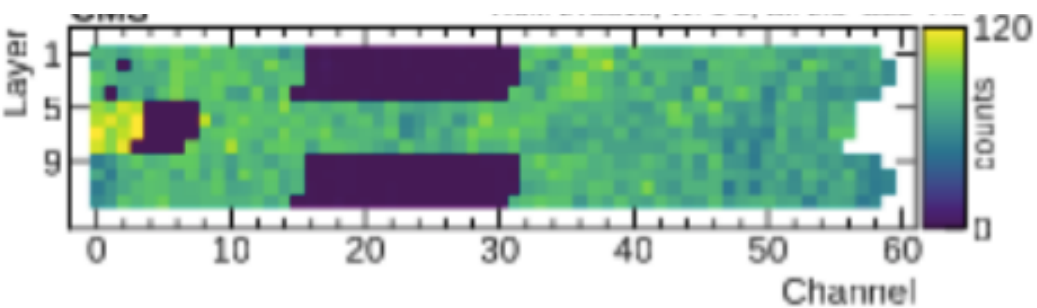
More in: Junghwan Goh talk 16:45

Muon drift tube hit occupancy





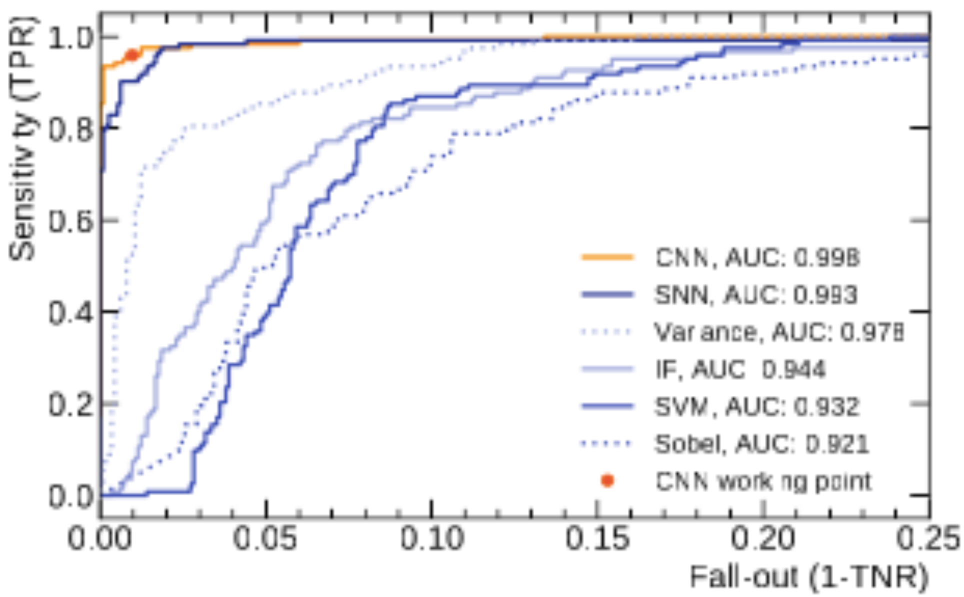
"good"



"bad"

Use machine learning to learn previous experts ratings

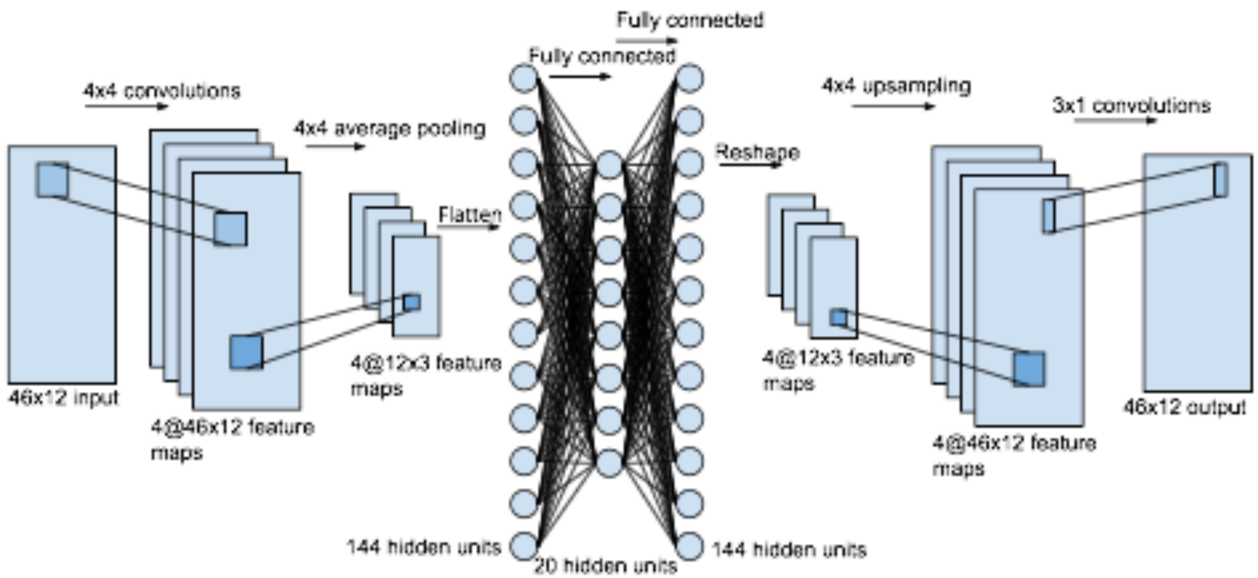
Results for single layers (DQM)



- A convolutional neural network architecture achieves an AUC of 0.998
- Very good automatization of expert
- Implemented real online DQM for test purposes

Semi supervised ML for DQM

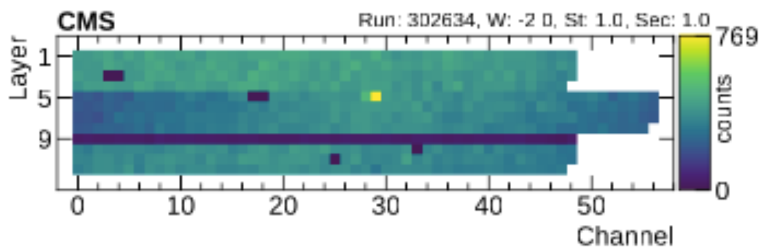
Use machine learning to catch any deviation from “good quality” data



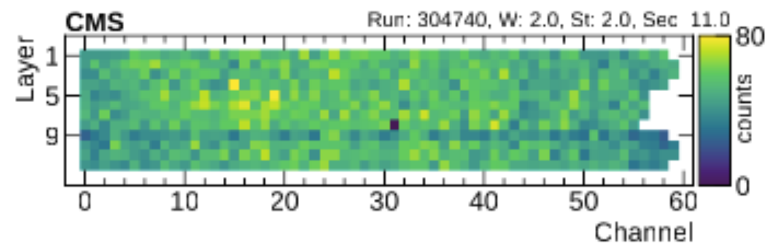
- Auto-encoder: The optimal dimension reduction depends on the ground truth. Only “good quality” data is used for training.
- A large difference between input and output indicates high probability that the data was not “good quality”

Per chamber more problem can be spotted

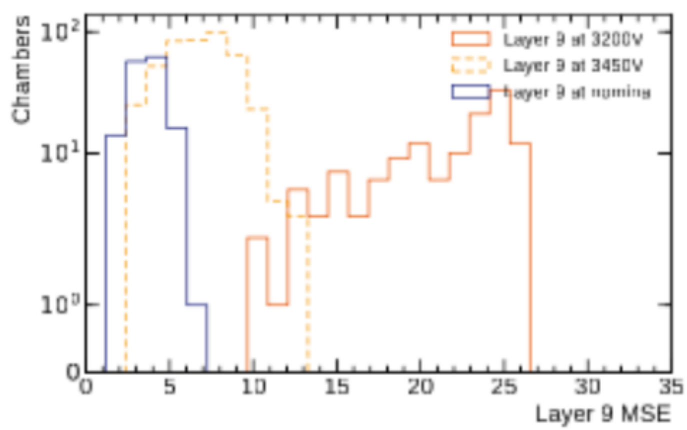
Low voltage layer



good layer



Mean Squared Error of Auto-encoder

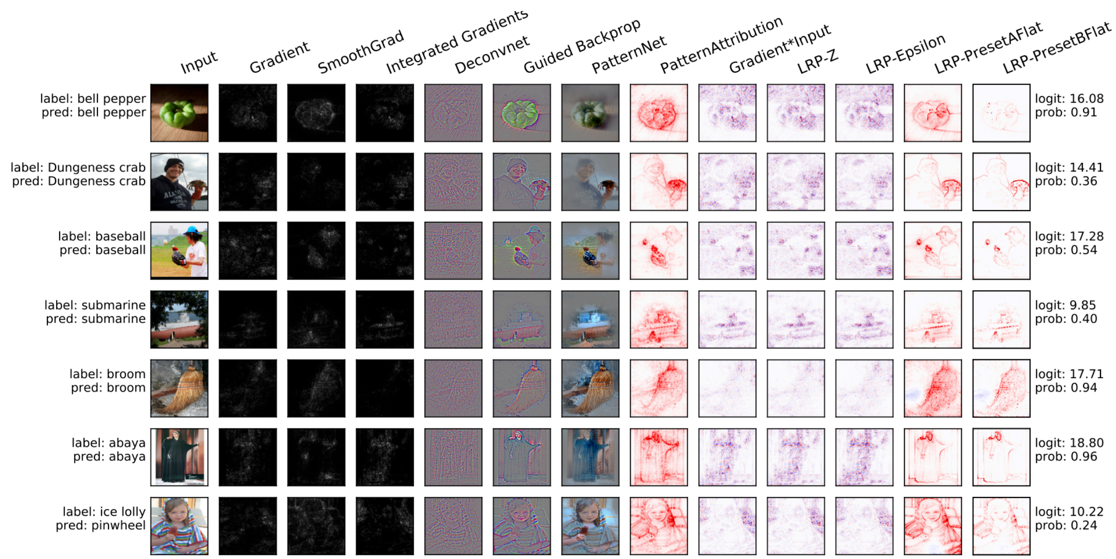


Auto-encoder indicated problems for low voltage

Black box ⟷ Interpretability

The actual analytical function ($\mathcal{O}$(million) parameters) is difficult to represent for deep learning, but input output relation can be studies to interpret the “black box”

Interpretability tools a "booming" research area



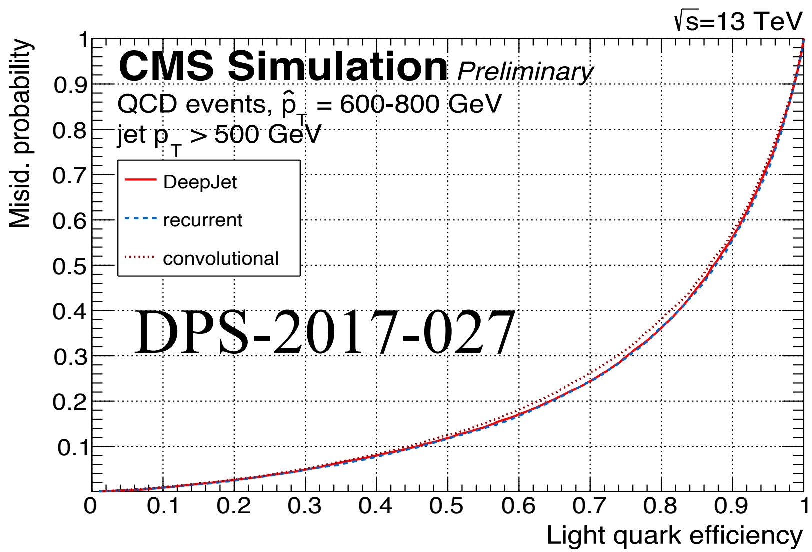
- Gradient or simple cuts are often used in physics to study the network behaviors
- Increasing number of more sophisticated tools on the market and implemented into standard libraries

Summary

- Deep learning on the rise in HEP
- Deep learning successfully applied to tagging and DQM in Alice and CMS
- Currently these deep learned methods are validated in real data
- Deep learning promises gains. Potential gains critically depend on how good the previous methods were and how much data is available
- Still many opportunities for deep learning
- Join and follow Inter-experimental LHC ML meetings!

Comparisons of DNNs

We filter on ***generator*** level only light quarks and gluons that did **NOT** split to heavy flavor.



→ Generic DeepJet and custom quark vs. gluon DNN (2D convolutions) gave very similar results!

→ **Data is multi-class, without heavy flavor removed DeepJet was clearly best**