Machine Learning and Statistics in HEP
part 2

David Rousseau, IJCLab-Orsay

rousseau@ijclab.in2p3.fr @dhpmrou

ICFA 2023 Instrumentation School

Mumbai, Feb 2023

Irène Joliot-Curie
Laboratoire de Physique des 2 Infinis

IN2P3 Les deux infinis

université PARIS-SACLAY

INSTITUT DATAIA Science des données, Intelligence & Société

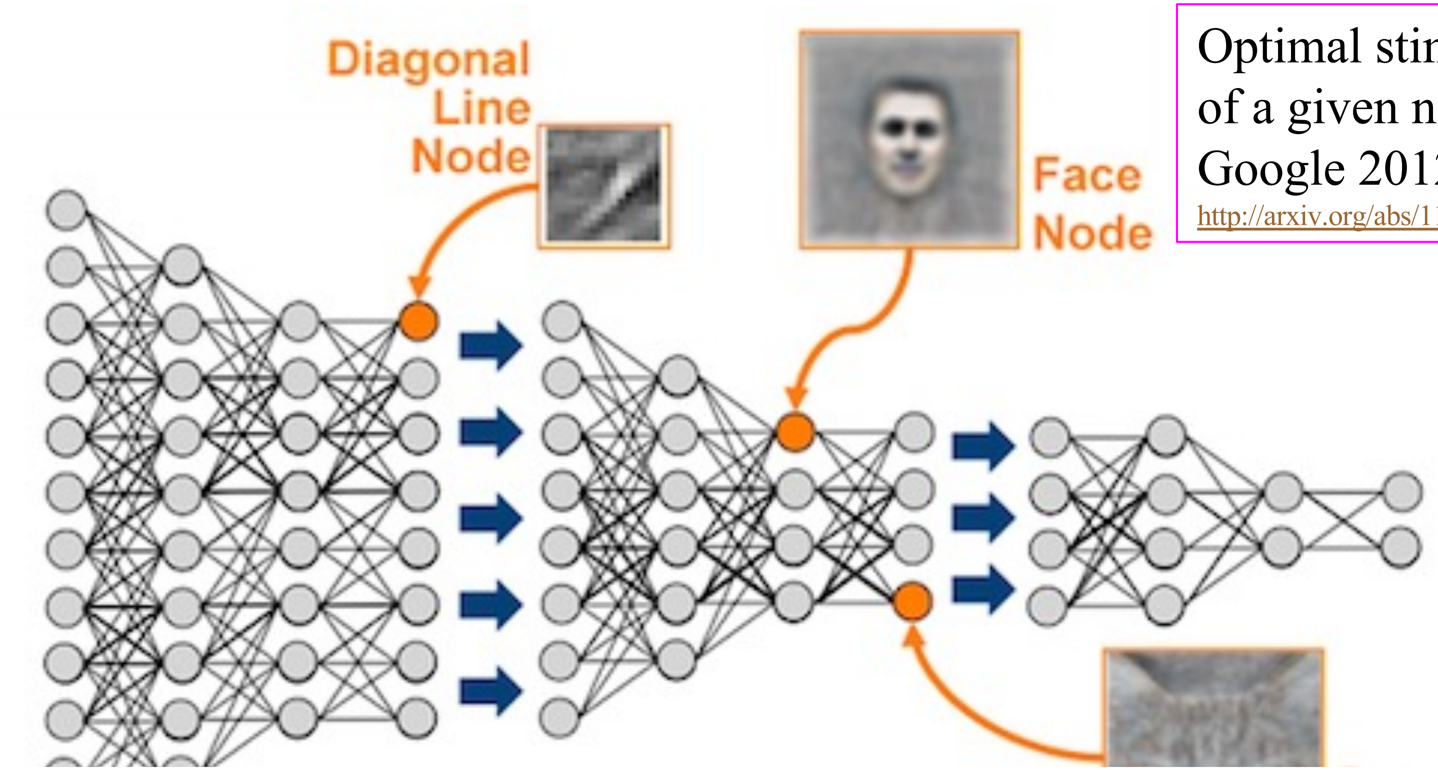
Outline

- Mostly Machine Learning with Statistics interludes
- Part 1 : Overview
  - What is Machine Learning?
  - Specificities of ML in physics
  - Useful concepts
- Part 2 : wider and deeper
  - NN on HEP data
    - various hammers and nails (including wrong ones)
  - Graph NN
  - Anomaly detection
- Part 3 : even wider and deeper
  - ML training tricks
  - Surrogate models
  - Recommendations for ML software and tools

See CERN Inter-Experiment Machine Learning workshop May 2022

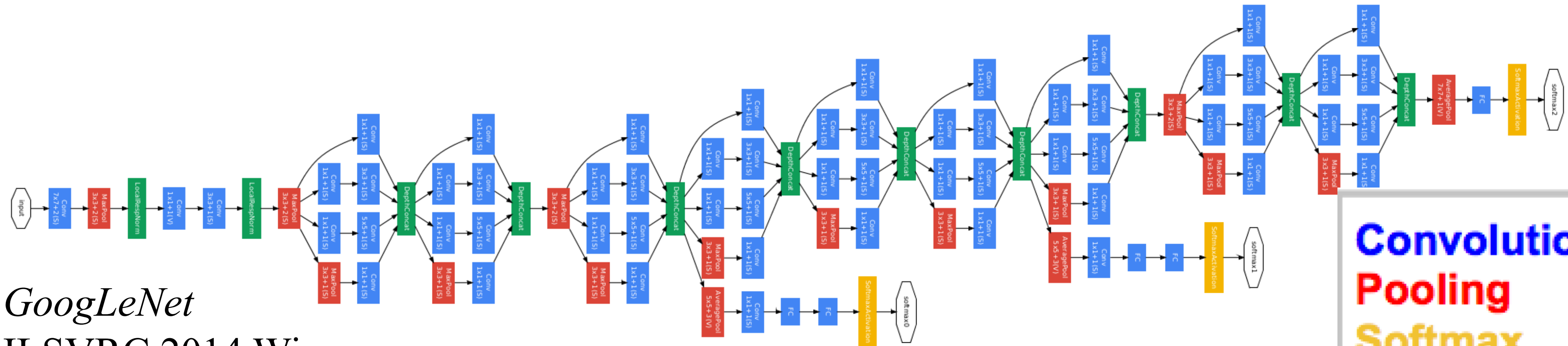
Deep Learning

Deep Learning



Optimal stimulus
of a given neuron
Google 2012
http://arxiv.org/abs/1112.6209

~1950
1990 hand-written digits
2010 deep learning explosion



GoogLeNet
ILSVRC 2014 Winner
4M parameters

'Godfathers of AI' honored with Turing Award, the Nobel Prize of computing

Yoshua Bengio, Geoffrey Hinton, and Yann LeCun laid the foundations for modern AI

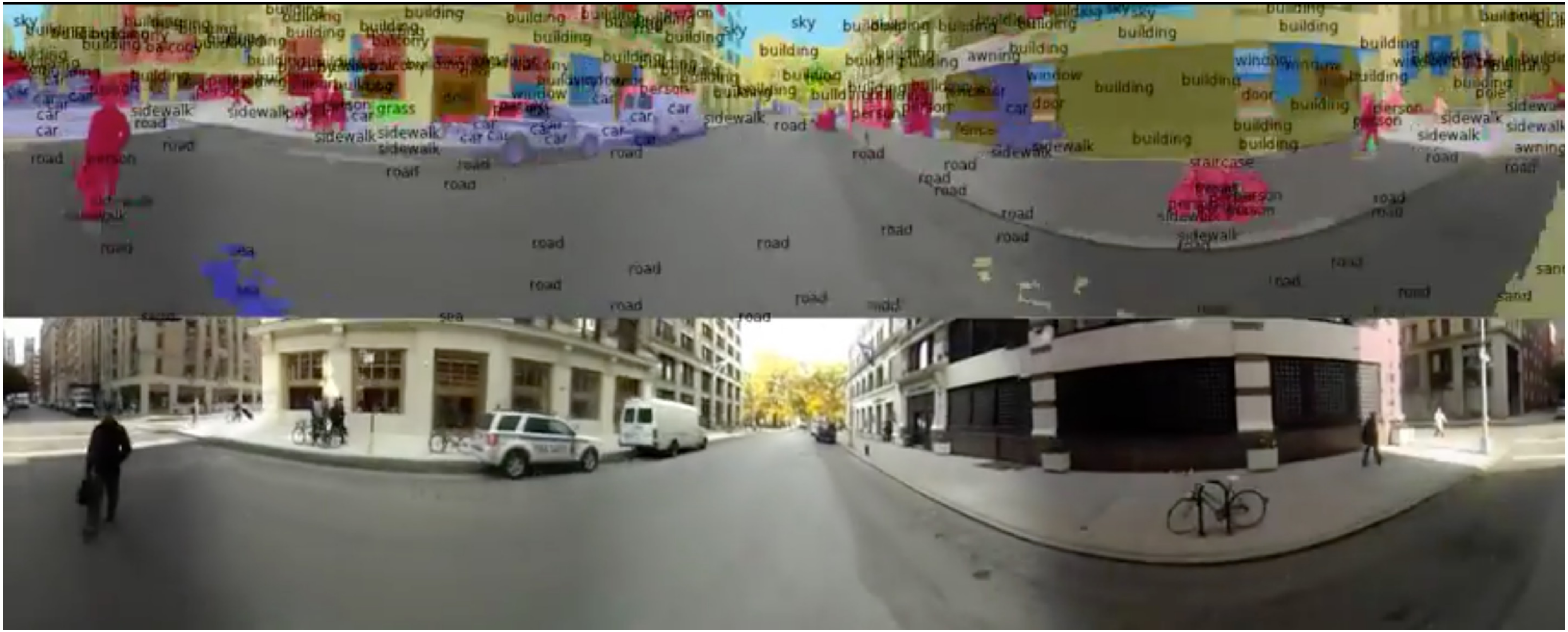
By James Vincent | Mar 27, 2019, 6:02am EDT

2019

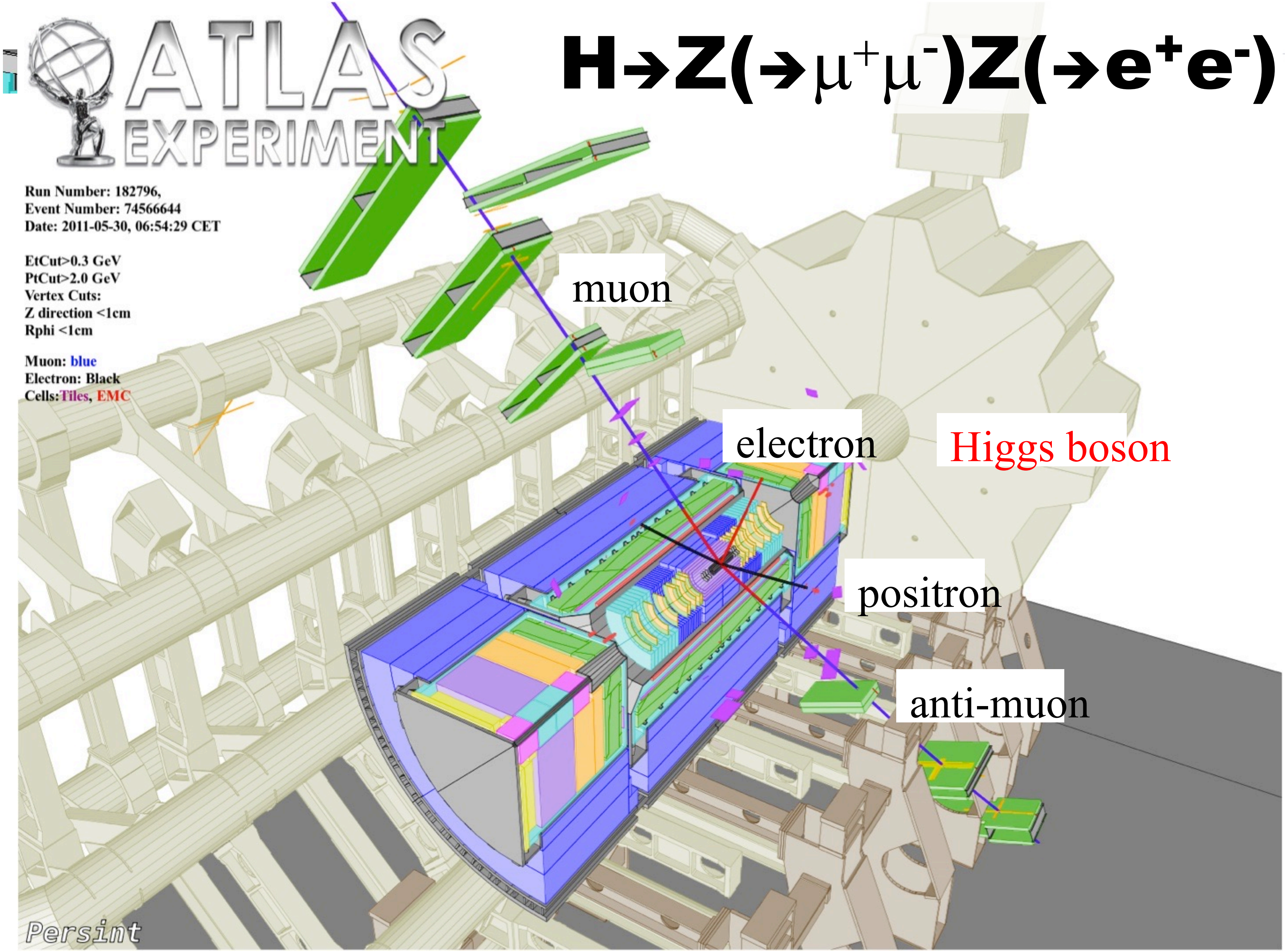
SHARE

HEP data are rarely images

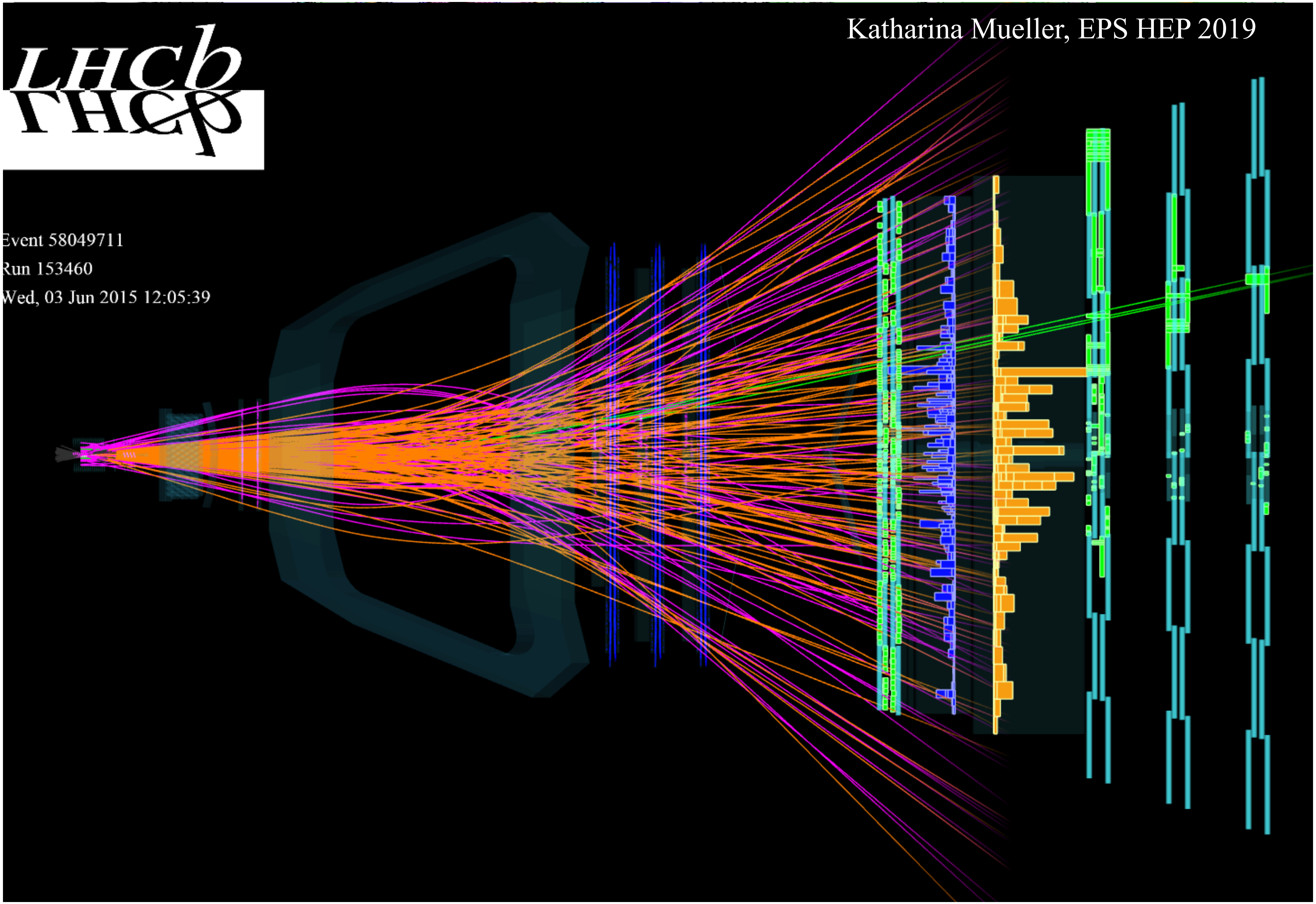
Typical Deep Learning application



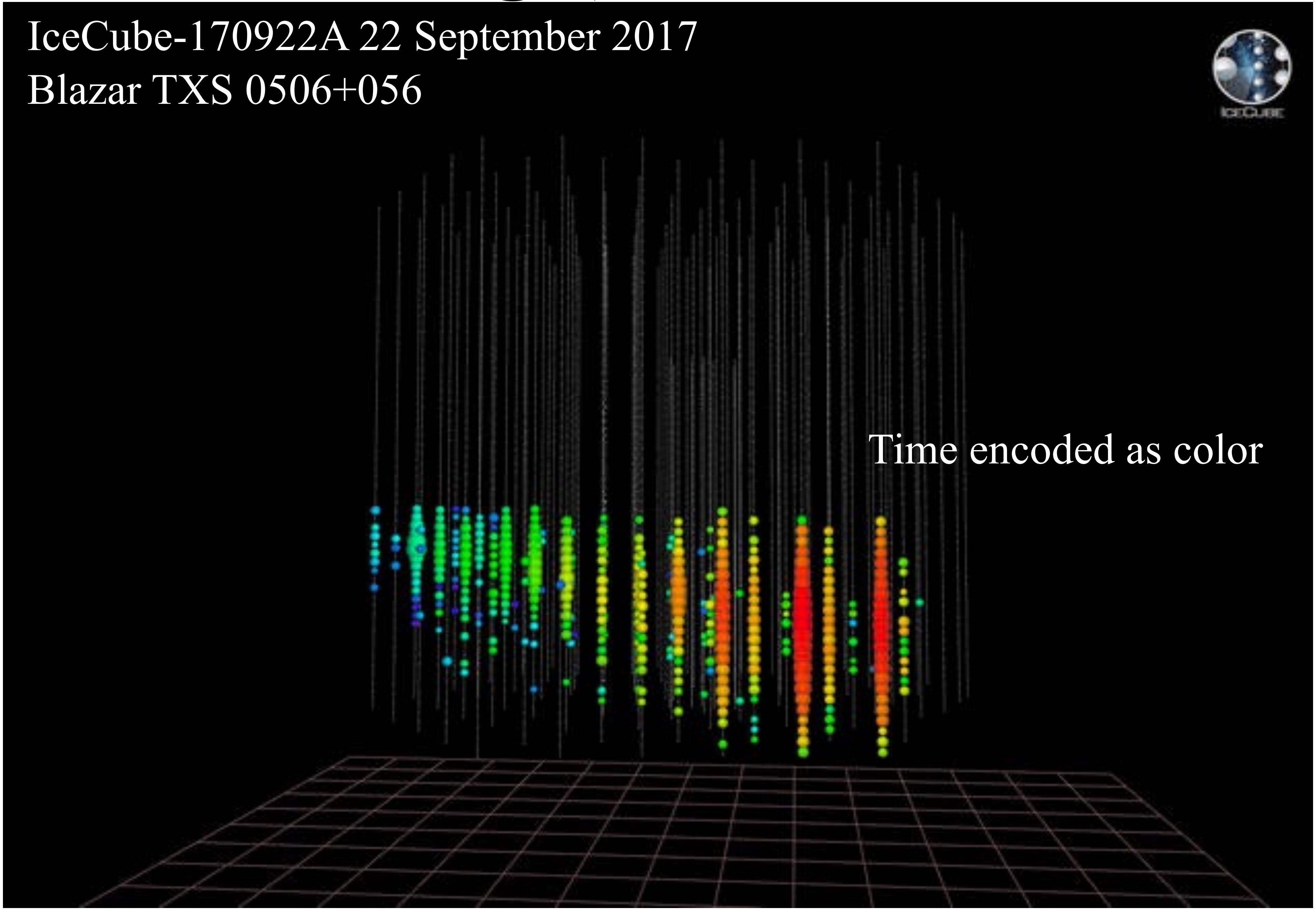
$H \to Z(\to \mu^+\mu^-)Z(\to e^+e^-)$



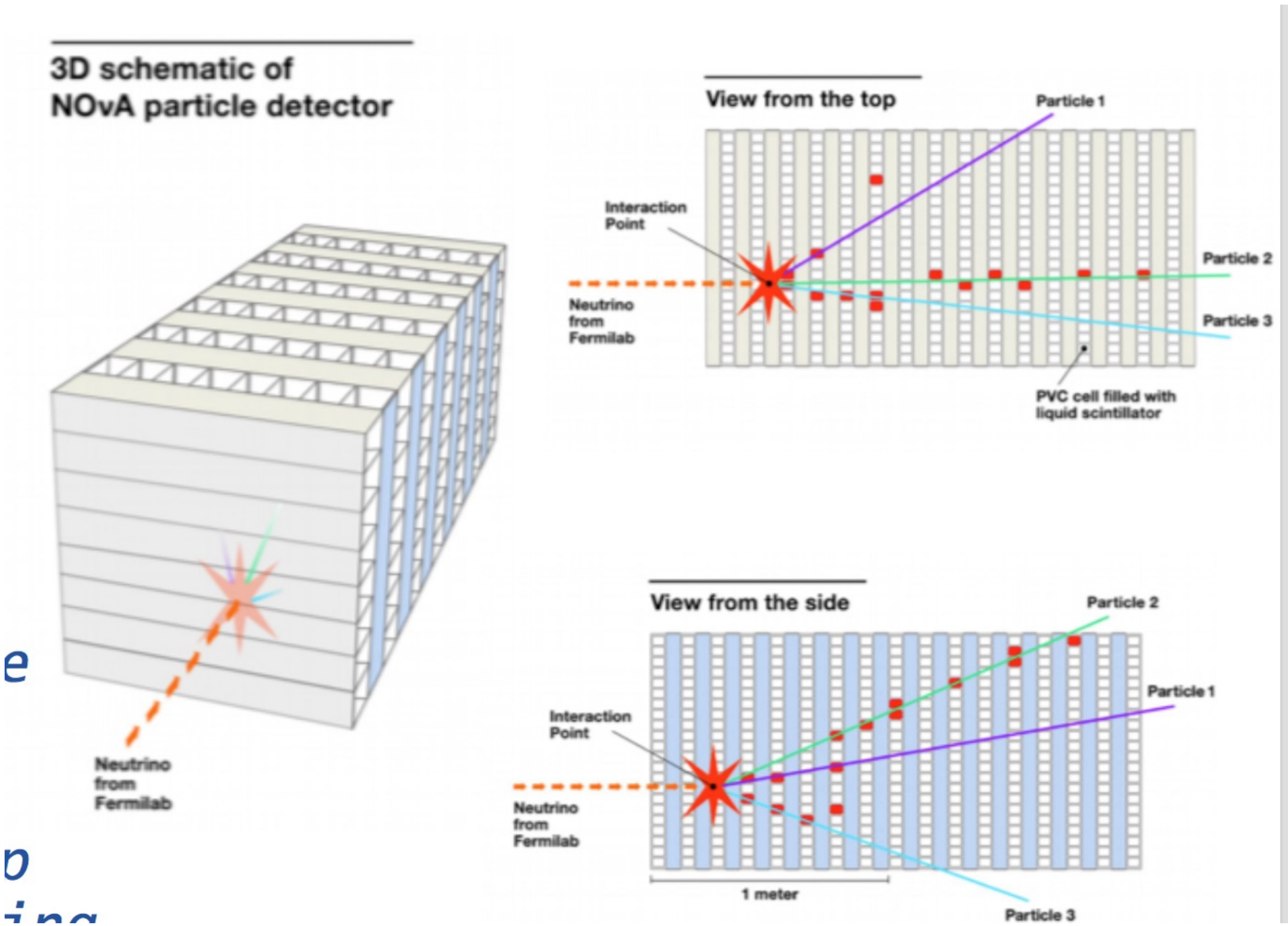
An image, not the data



An image, not the data

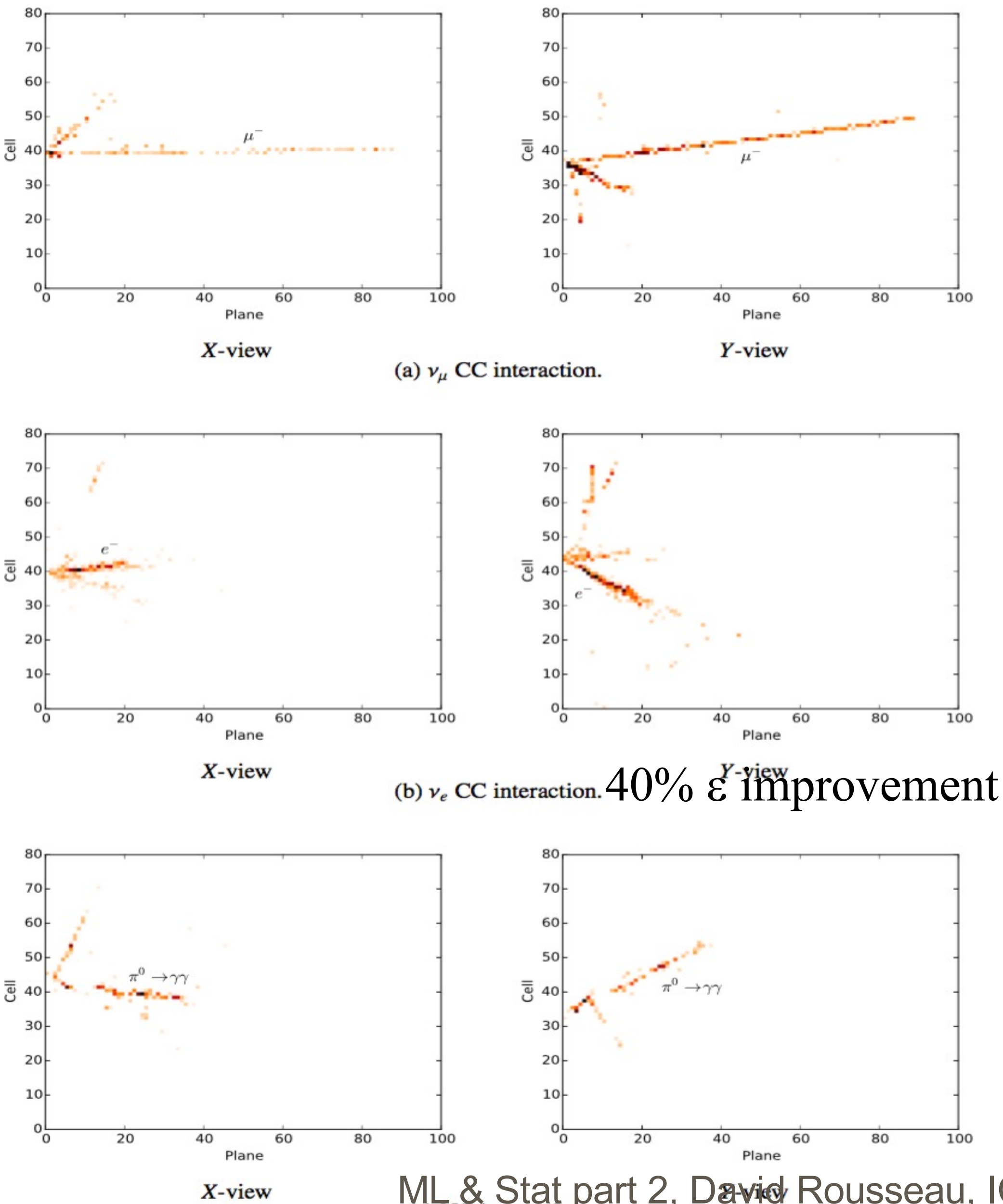


An exception : NOVA



NOVA (2)

arXiv 1604.01444 Aurisano et al



(a) ν_μ CC interaction.

(b) ν_e CC interaction.

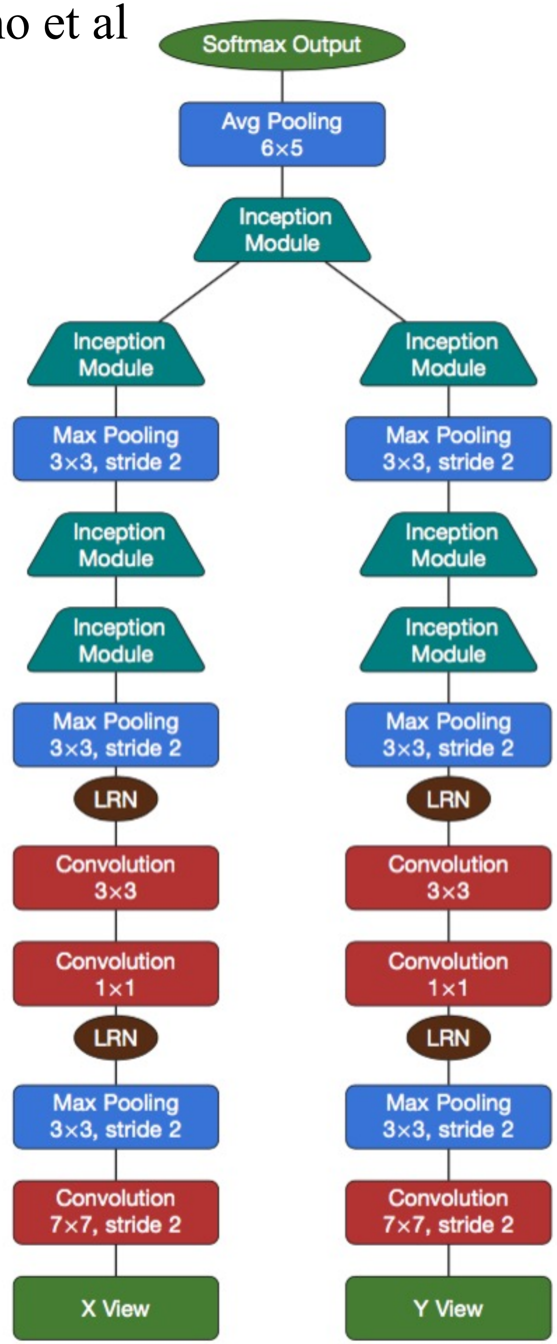
(c) NC interaction.

40% ε improvement

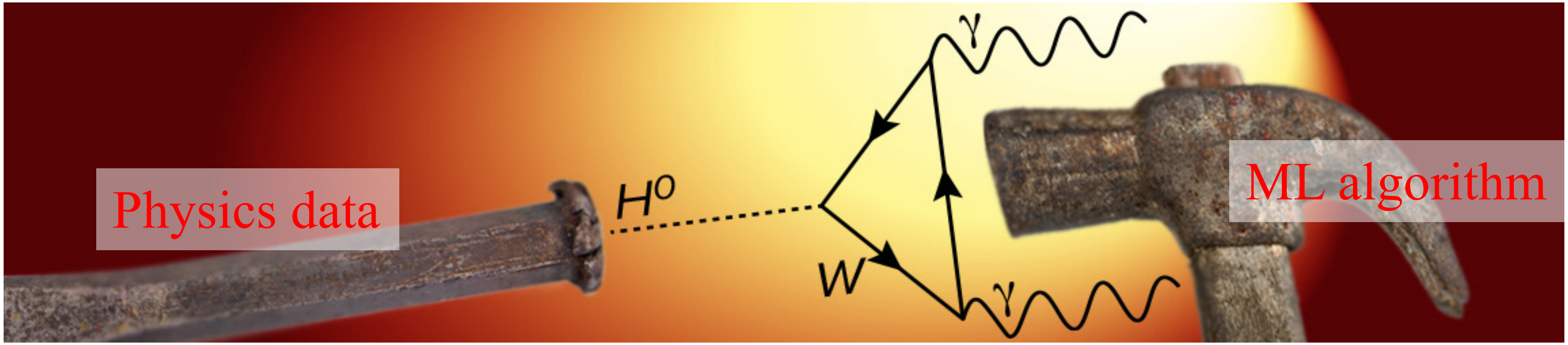
Neutrino interaction classification

Using Convolutional Neural Network (GoogleNet)

Actually used in physics results 1703.03328 and 1706.04592



Hammer and Nail

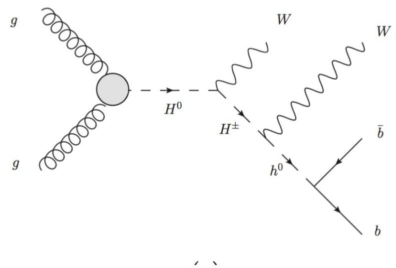


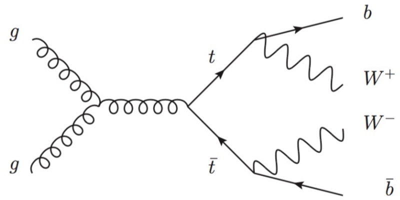
H&N, Weizmann

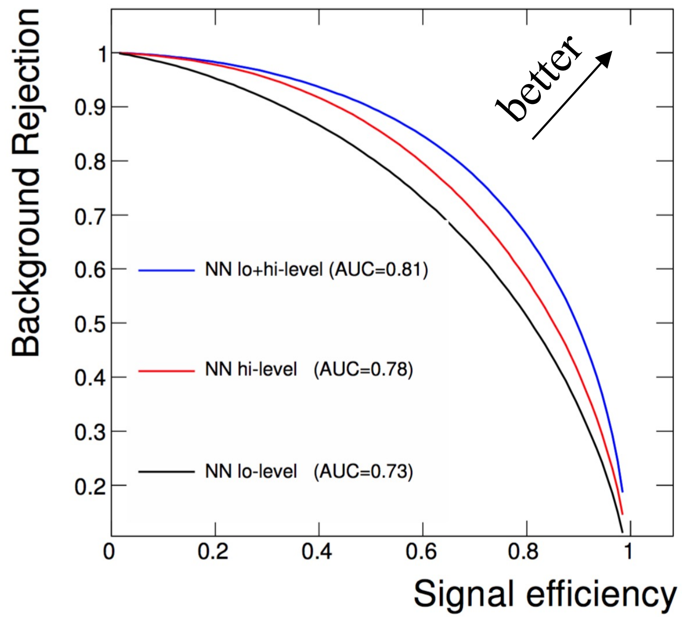
- If it does not fit:
  - Reshape the nail : feature engineering
  - Reshape the hammer : Neural Network architecture

Deep learning for analysis

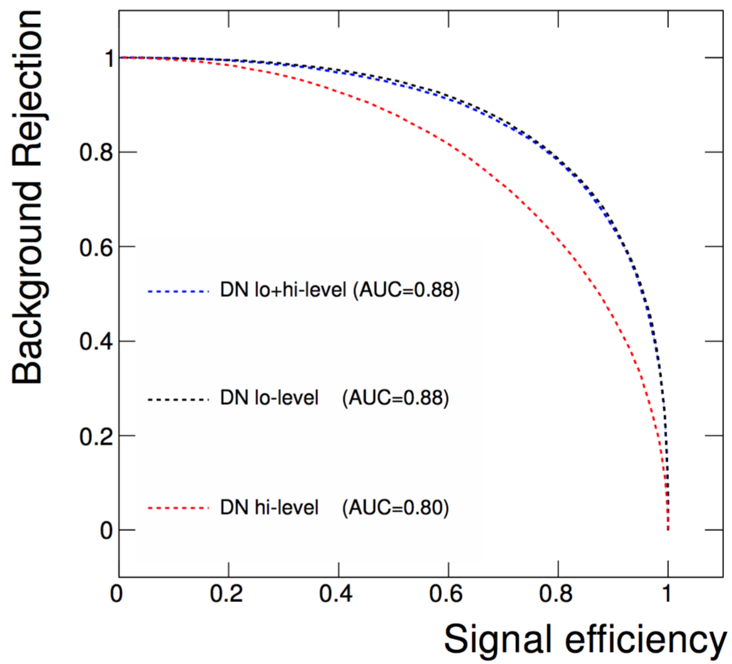
1402.4735 Baldi, Sadowski, Whiteson







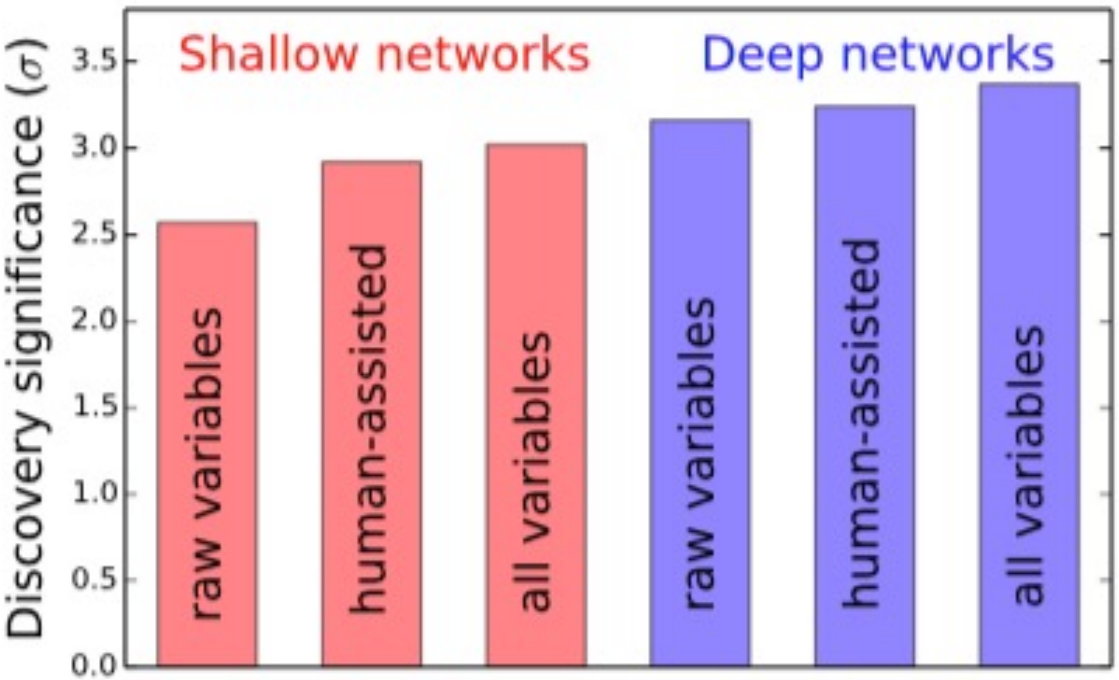
- MSSM at LHC : $H^0 \rightarrow WWbb$ vs $tt \rightarrow WWbb$
- Low level variables:
  - 4-momentum vector
- High level variables:
  - Pairwise invariant masses
- Deep NN outperforms NN, and does not need high level variables
- DNN learns the physics ?



Deep learning for analysis (2)

1410.3469 Baldi Sadowski Whiteson

- H tautau analysis at LHC: H→tautau vs Z→tautau
  - Low level variables (4-momenta)
  - High level variables (transverse mass, delta R, centrality, jet variables, etc…)



- Here, the DNN improved on NN but still needed high level features
- Both analyses with Delphes fast simulation
- ~100M events used for training (>>100* full G4 simulation in ATLAS)

DL for analysis (3)

- Few published LHC analyses using DL
- Recent trend is to feed more (up to 20) variables to classifiers, even low level ones (2/3-vectors of particles) (see recent ATLAS/CMS ttH papers)
- A few NN in top and Higgs physics but no clear advantage wrt BDT
- Not completely clear why, but best guess is the lack of statistics (Baldi et al used ~100M events)
- We'll still need clever variables building like:
  - $\mathbf{M_T(a)} = \sqrt{2|\vec{p}_T^{\,a}||\vec{p}_T^{\,miss}|\,(1-\cos\Delta\phi(\vec{p}_T^{\,a},\vec{p}_T^{\,miss}))}$
  - $\mathbf{M_{T2}(a,b)} = \min_{\vec{p}_{T\mathbf{1}}^{\,miss}+\vec{p}_{T\mathbf{2}}^{\,miss}=\vec{E}_T^{miss}}\left(\max\left[M_T(\vec{p}_T^{\,a},\vec{p}_{T\mathbf{1}}^{\,miss}),\,M_T(\vec{p}_T^{\,b},\vec{p}_{T\mathbf{2}}^{\,miss})\right]\right)$
    $\rightarrow$ well suited for pair production of $X \rightarrow Y(vis.) + Z(inv.)$
  - MHT$=|-\sum_{jets}\vec{p}_T^{\,j}|$
- Neural Net more relevant if large number of unstructured variables (examples later)
- Also developments for physics aware NN, who know that they are dealing with 4-momenta (Lorentz-invariant NN arXiv:2208.07814.)

End to end Learning

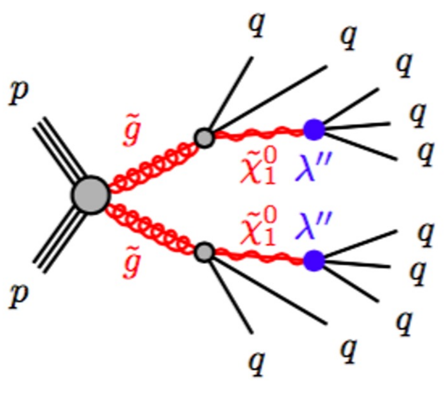
End to end learning

Bhimji et al, 1711.03573

- Train directly for signal on « raw » event ?
- Start from RPV Susy search

ATLAS-CONF-2016-057

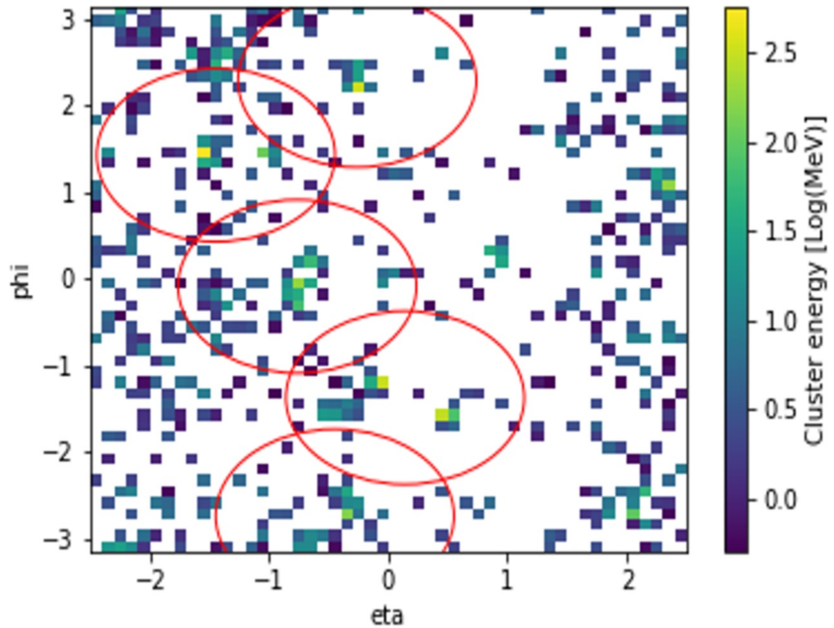
- Fast Simulated events with Delphes



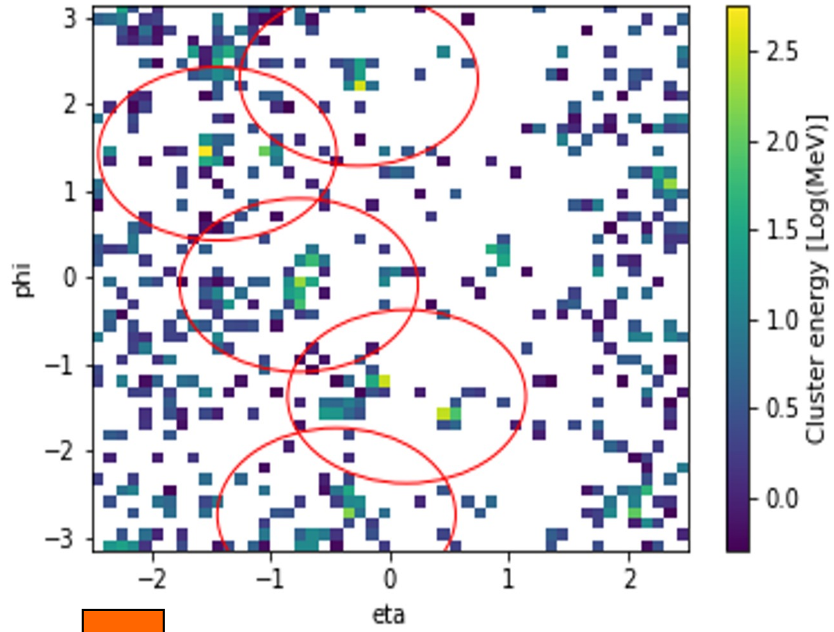
(b) gluino cascade decay

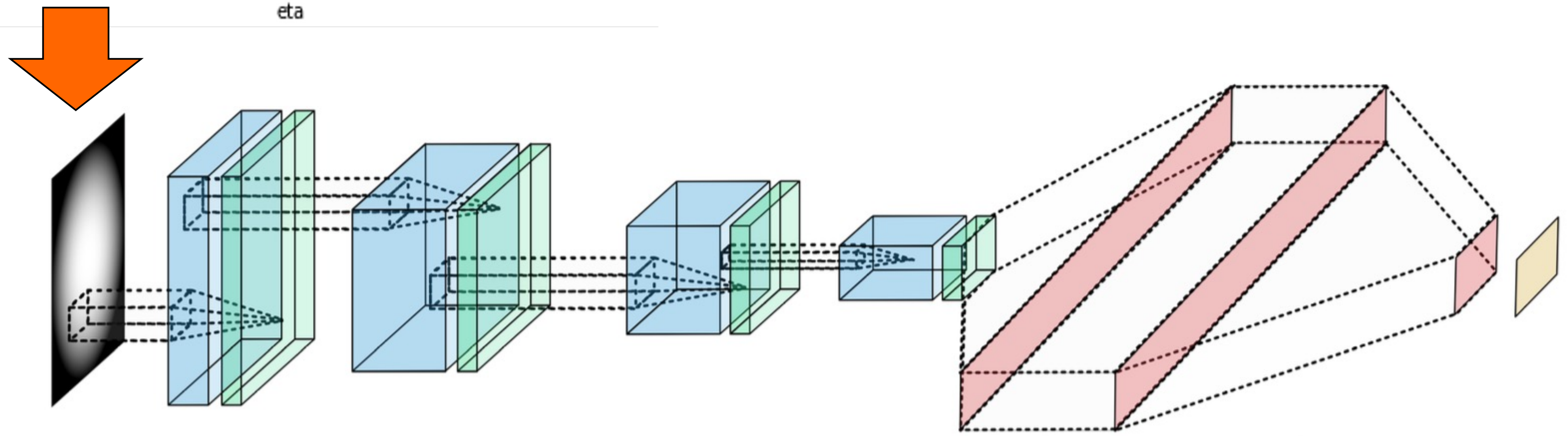
- Project energies on 64x64 ηxϕ grid
- Compare with usual jet Reconstruction and physics Analysis variables such as:

$$M_J^{\Sigma} = \sum_{\substack{p_T > 200\,\mathrm{GeV} \\ |\eta| \leq 2.0}}^{4} m^{\mathrm{jet}}$$

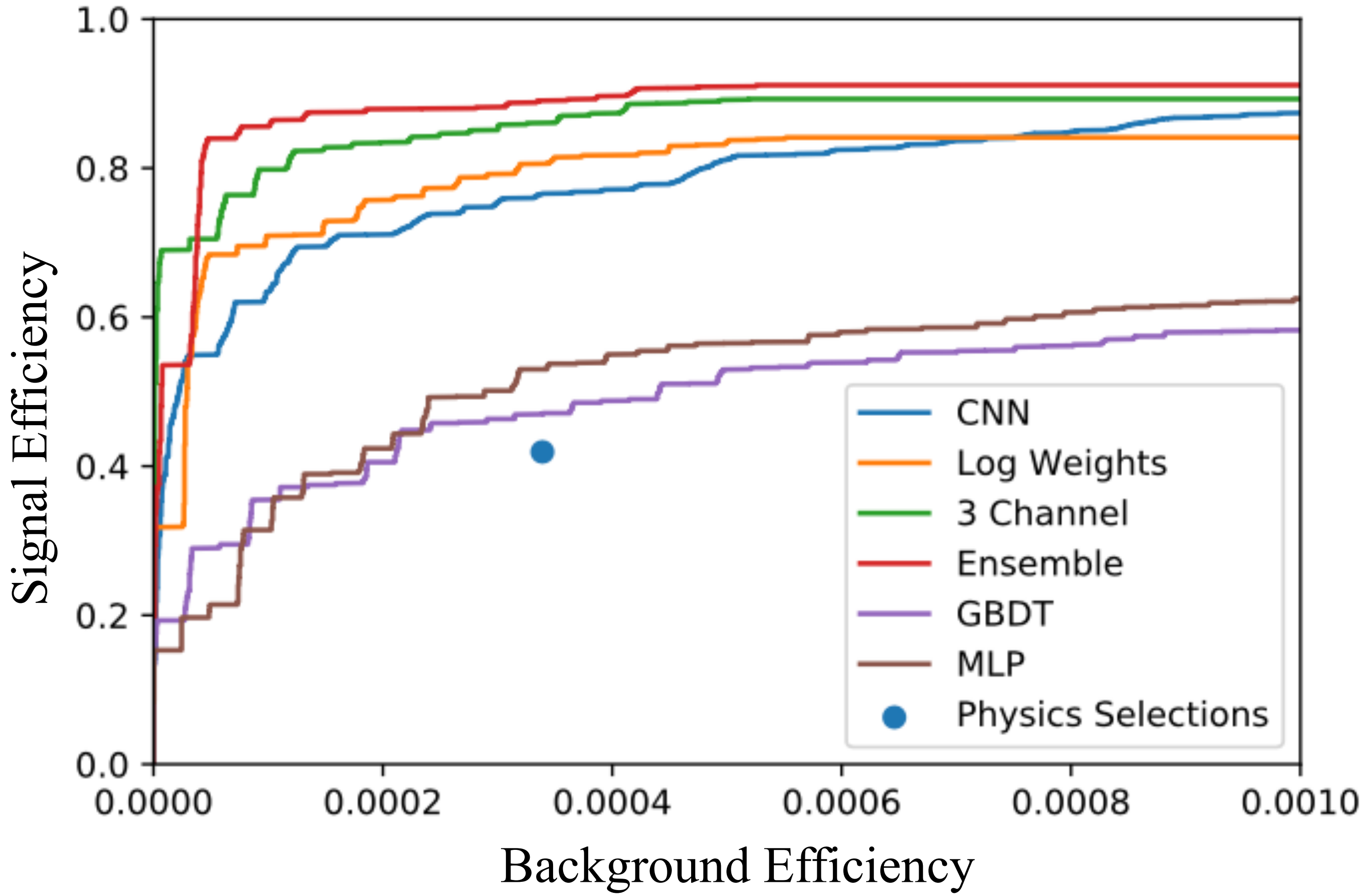


End to end learning (2)





End to end learning (3)

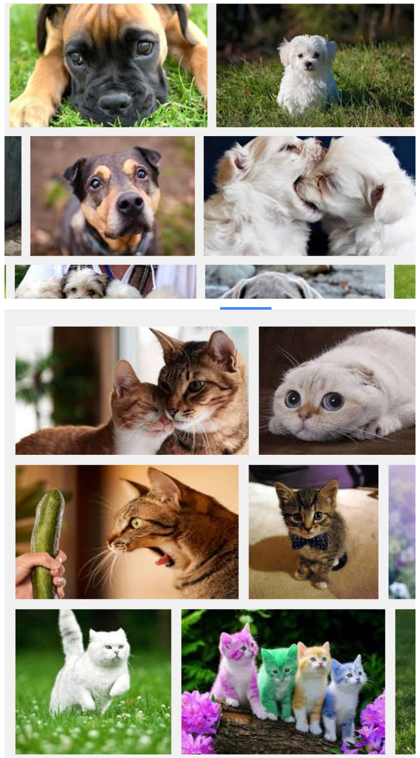


- >x2 gain over BDT/shallow network using physics variable and 5 leading jet 4-momenta
- ➔CNN extract information from energy grid which is lost in the jets ?

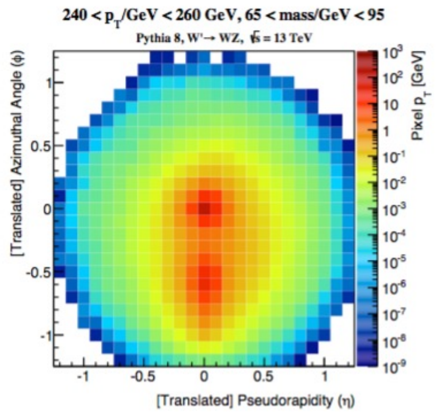
Jet Images

arXiv 1511.05190 de Oliveira, Kagan, Mackey, Nachman, Schwartzman

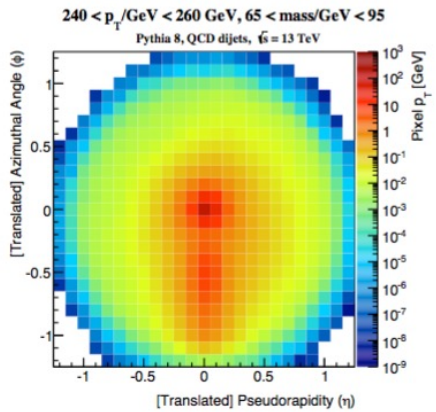
- Distinguish boosted W jets from QCD
- Particle level simulation
- Average images:



Boosted W→qq jet

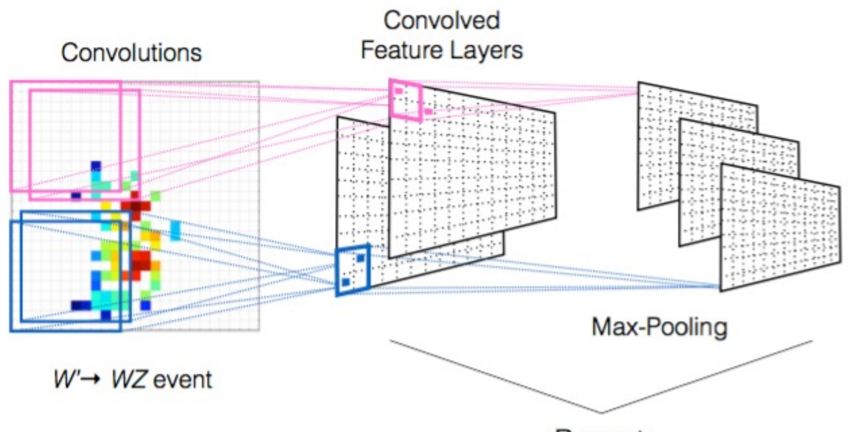


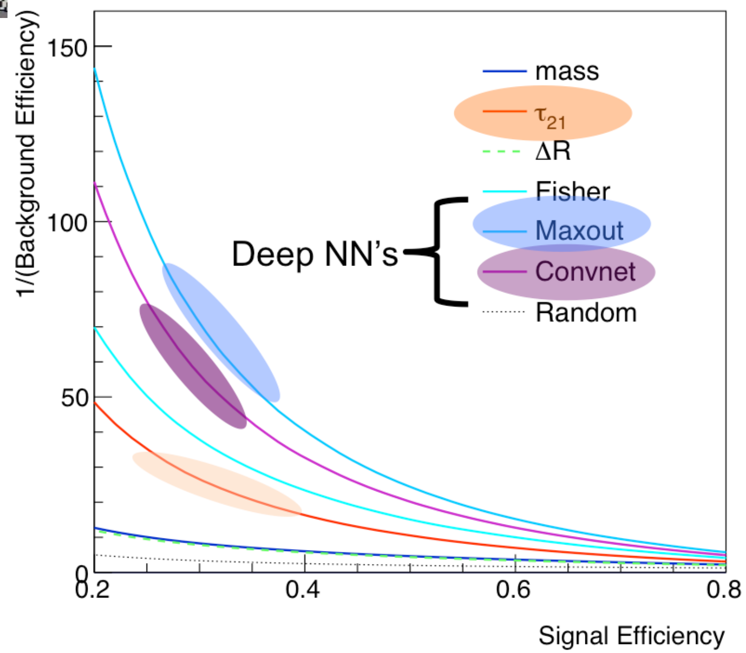
QCD



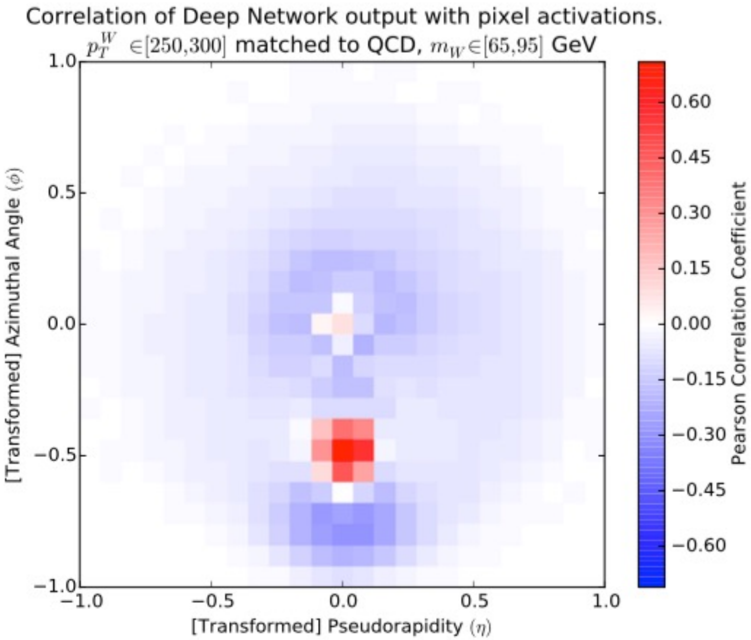
Jet Images : Convolution NN

arXiv:1511.05190



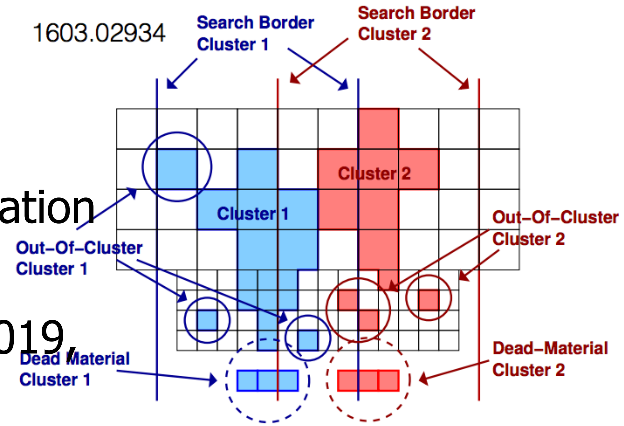


- Projection on calorimeter grid
- Variables build from CNN outperform the more usual ones

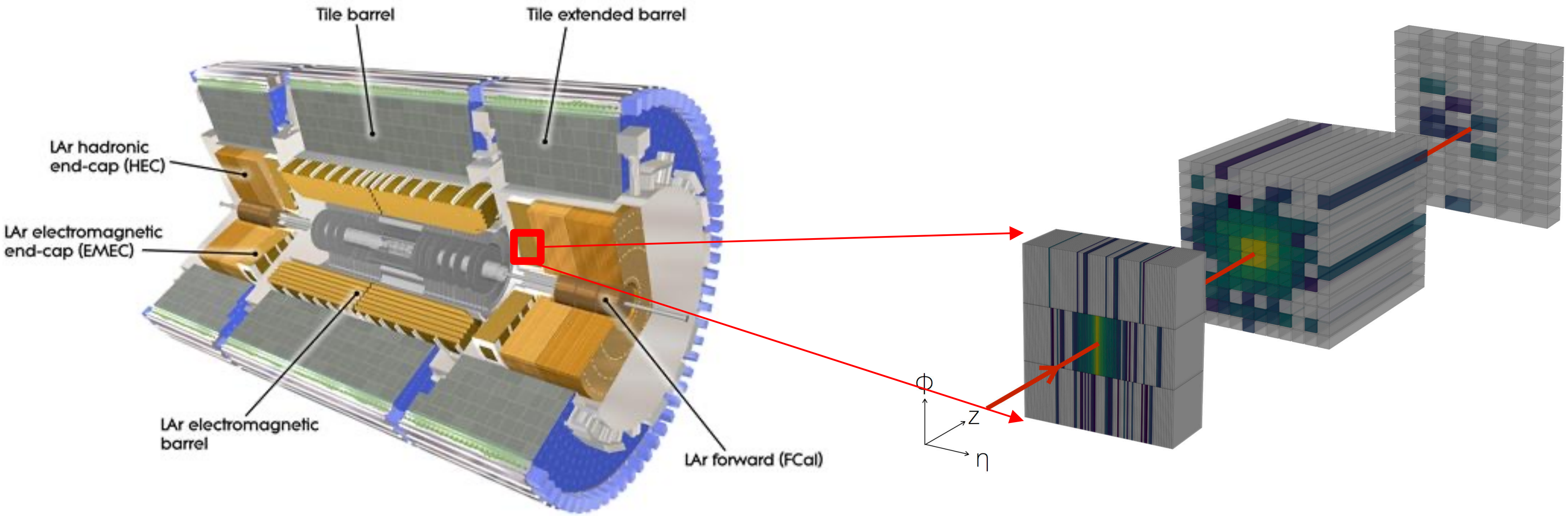


- What the CNN sees (the "cat" neurone")
- Also with proper detector and pileup simulation ATL-PHYS-PUB-2017-017
- ➔3 Dimension ?

- See also GraphNN application to CMS HGCAL, Arabella Martelli, EPS-HEP 2019, and arXiv:1902.07987



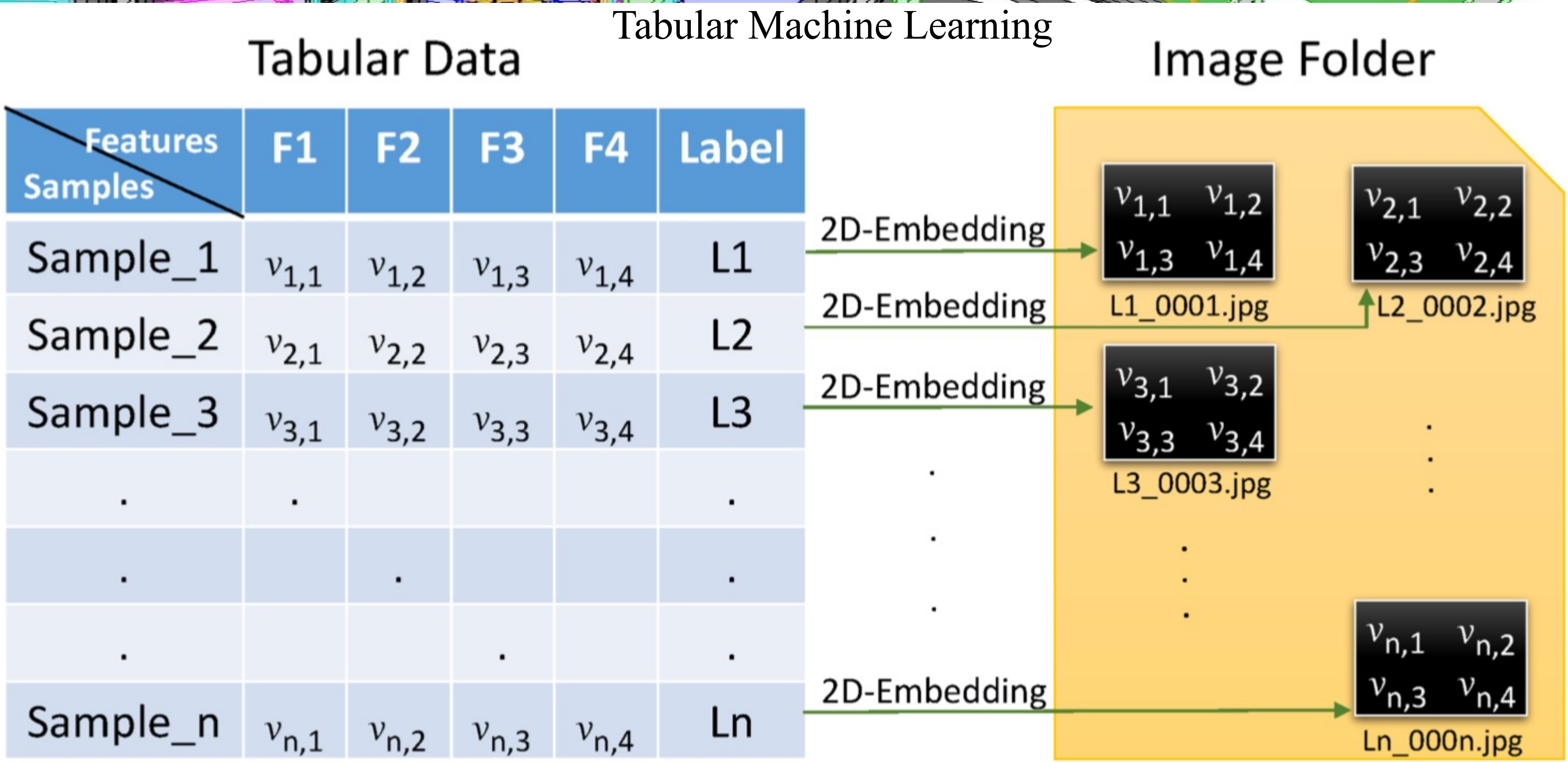
Detectors are complex 3D objects



Not good ideas : SuperTML and Abstract images

Don’t do this at home

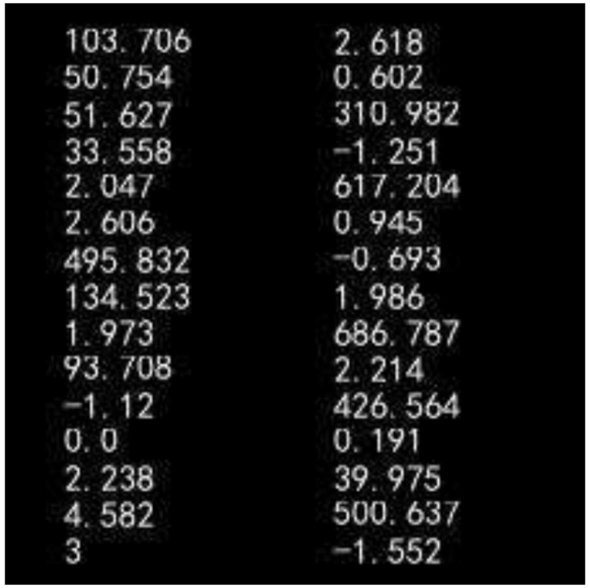
SuperTML : principle



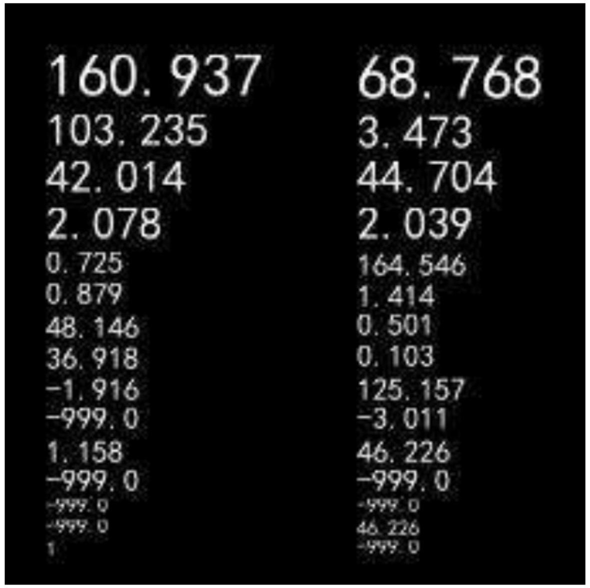
…then analyse image with fine tuned a pre-trained CNN (on cats and dogs)

Super TML on HiggsML dataset

HiggsML : public Higgs Machine Learning challenge dataset with a few 4-momentum features



(a) SuperTML image example for Higgs Boson data. Each feature is given equal importance in this example.



(b) SuperTML image example for Higgs Boson data. Features are given different sizes according to their importance.

SuperTML results

Table 3. Comparison of AMS score on Higgs Boson. The first two rows are winners in the Higgs Boson Challenge.

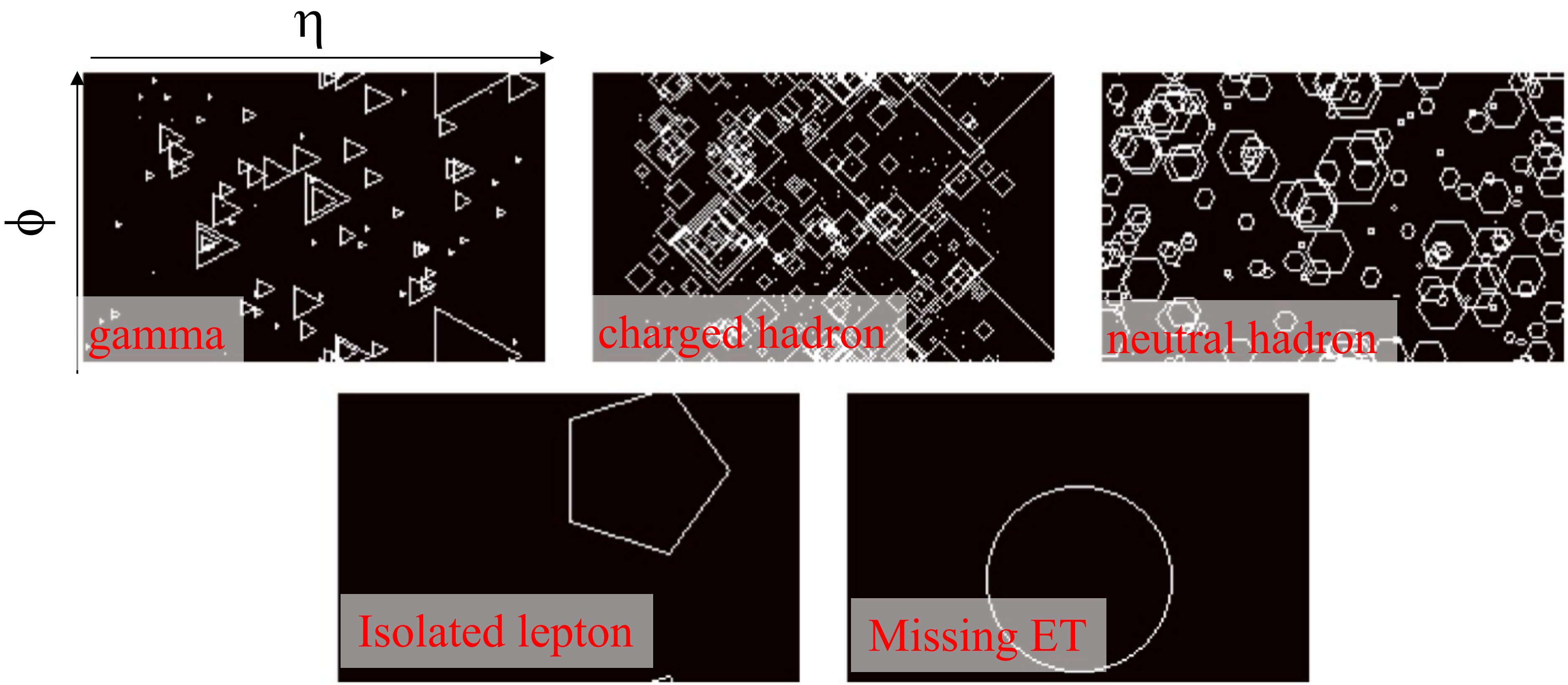
Methods	**AMS**	
DNN by Gabor Meli	3.806	
XGBoost	3.761	
SuperTML_EF(224x224)	**3.979**	No Way!
SuperTML_VF (224x224)	3.838	Overtraining ?

Giles Strong (CMS) tried to reproduce their results without success, contacted them with little success.

➔this was published in a Computer Vision conference
➔Lesson : do not believe whatever appear on arXiv

Abstract Images

arXiv:1708.07034 arXiv:1807.00083



tt event in CMS

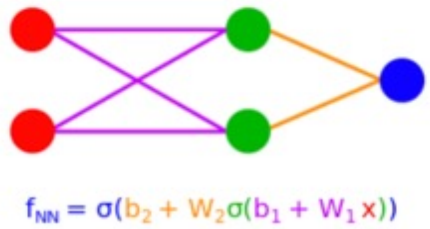
- ➔does not work too well
- Unsurprisingly…. two bad examples of forcing data into one hammer (Convolution Neural Network)
- The physicist' expertise should help the CNN, not obscure it

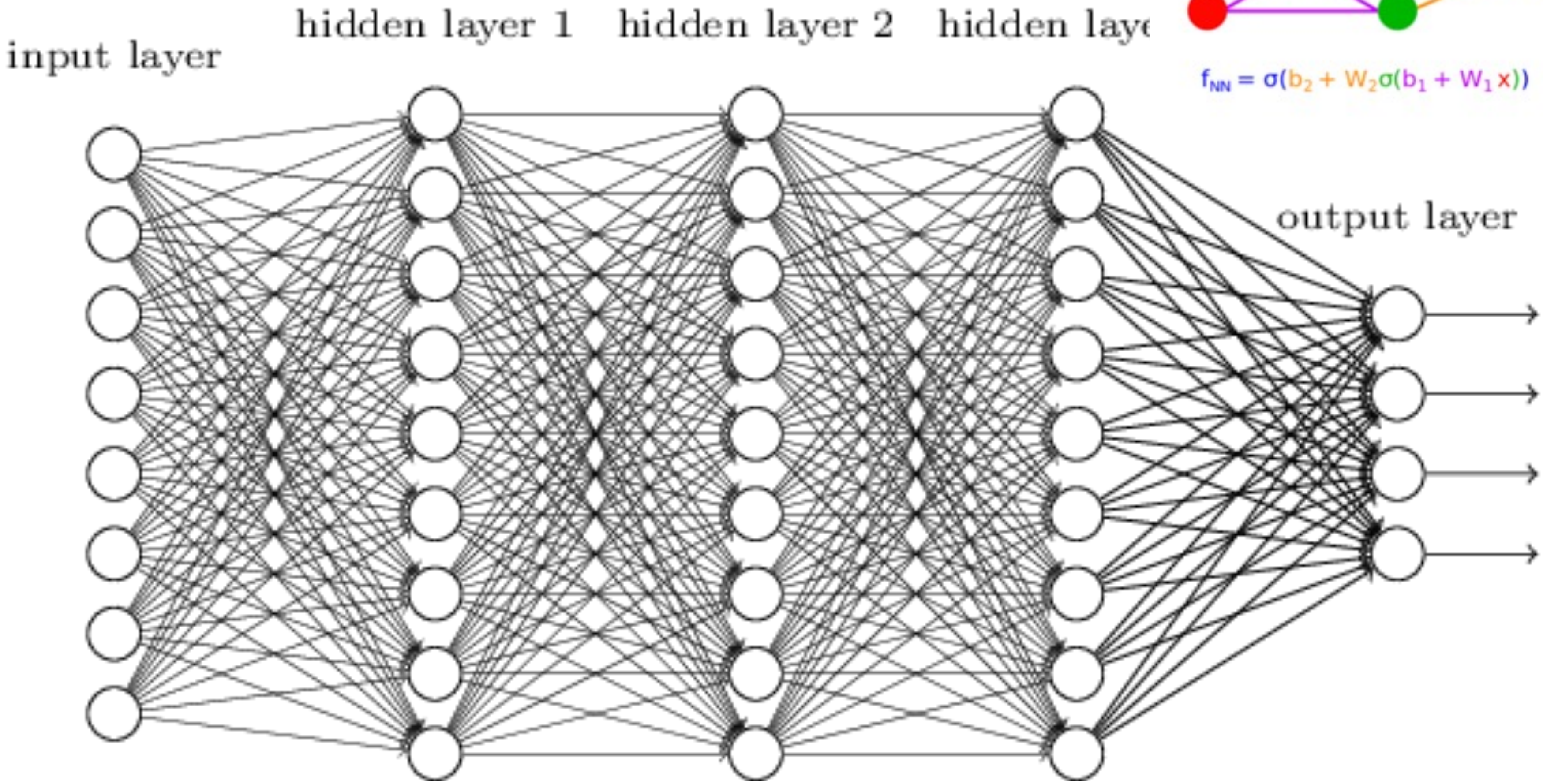
Graph Neural Networks

Graph Neural Networks

Marco Battaglia IML 2020

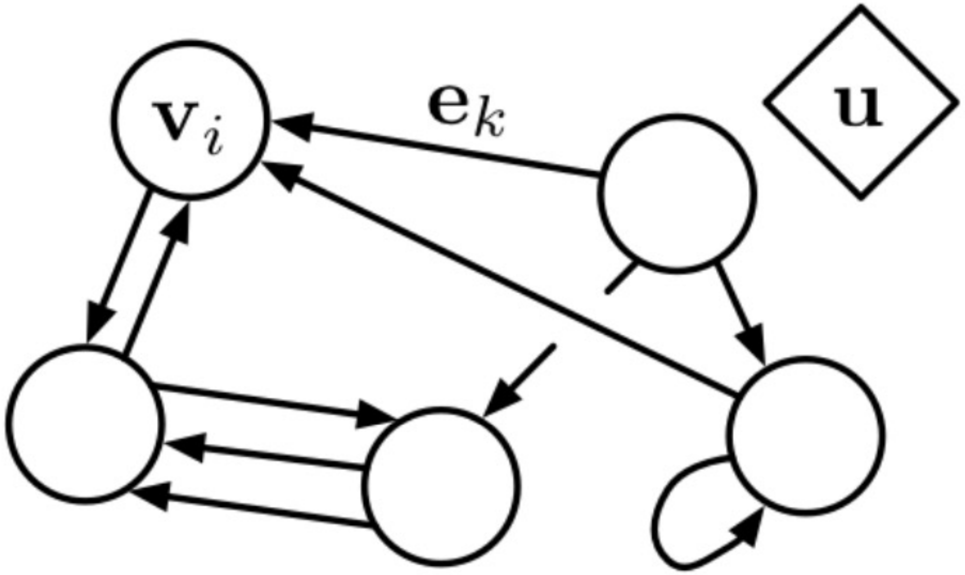
- « Classical » NN : a structure less vector or matrix is transformed from layer to layer (Convolution NN : use neighbouring information)



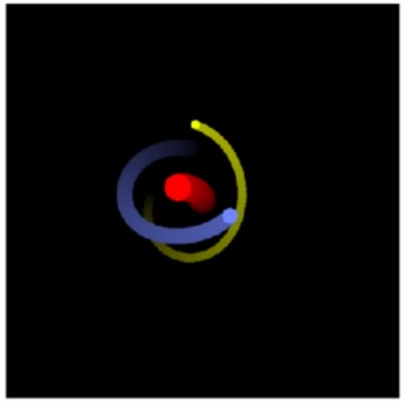


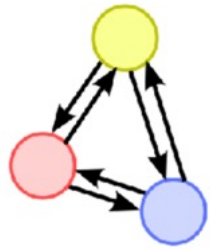
GNN (2)

- Now some structure:
  - v_i : nodes
  - e_k : edges
  - u : global



n-body

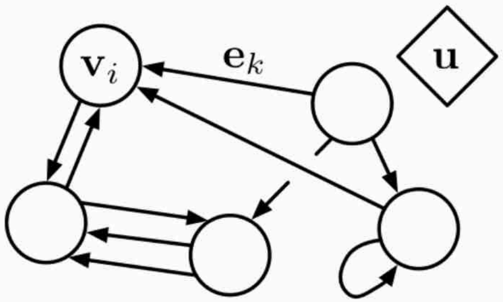


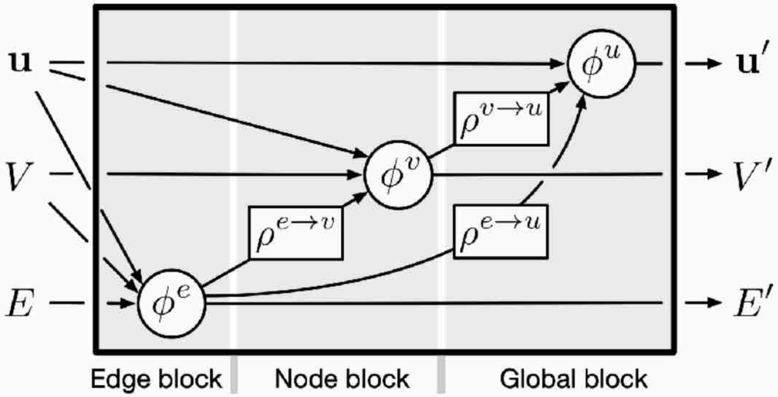


Nodes: bodies
Edges: gravitational forces

Global : potential energy

GNN (3)

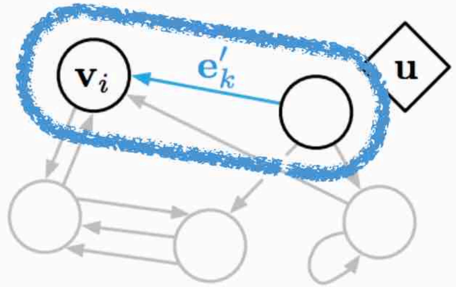




Edge block

For each edge, $\mathbf{e}_k, \mathbf{v}_{s_k}, \mathbf{v}_{r_k}, \mathbf{u}$, are passed to an "edge-wise function":

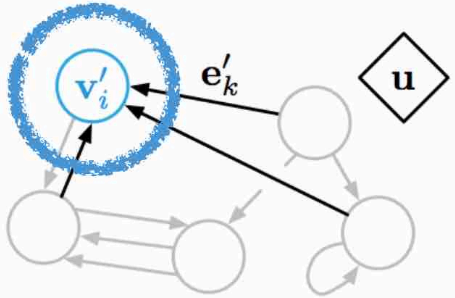
$$\mathbf{e}'_k \leftarrow \phi^e\left(\mathbf{e}_k, \mathbf{v}_{r_k}, \mathbf{v}_{s_k}, \mathbf{u}\right)$$



Node block

For each node, $\bar{\mathbf{e}}'_i, \mathbf{v}_i, \mathbf{u}$, are passed to a "node-wise function":

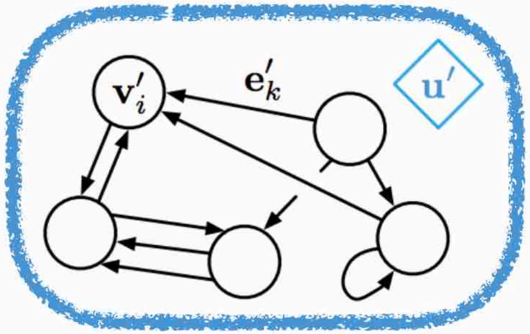
$$\mathbf{v}'_i \leftarrow \phi^v\left(\bar{\mathbf{e}}'_i, \mathbf{v}_i, \mathbf{u}\right)$$



Global block

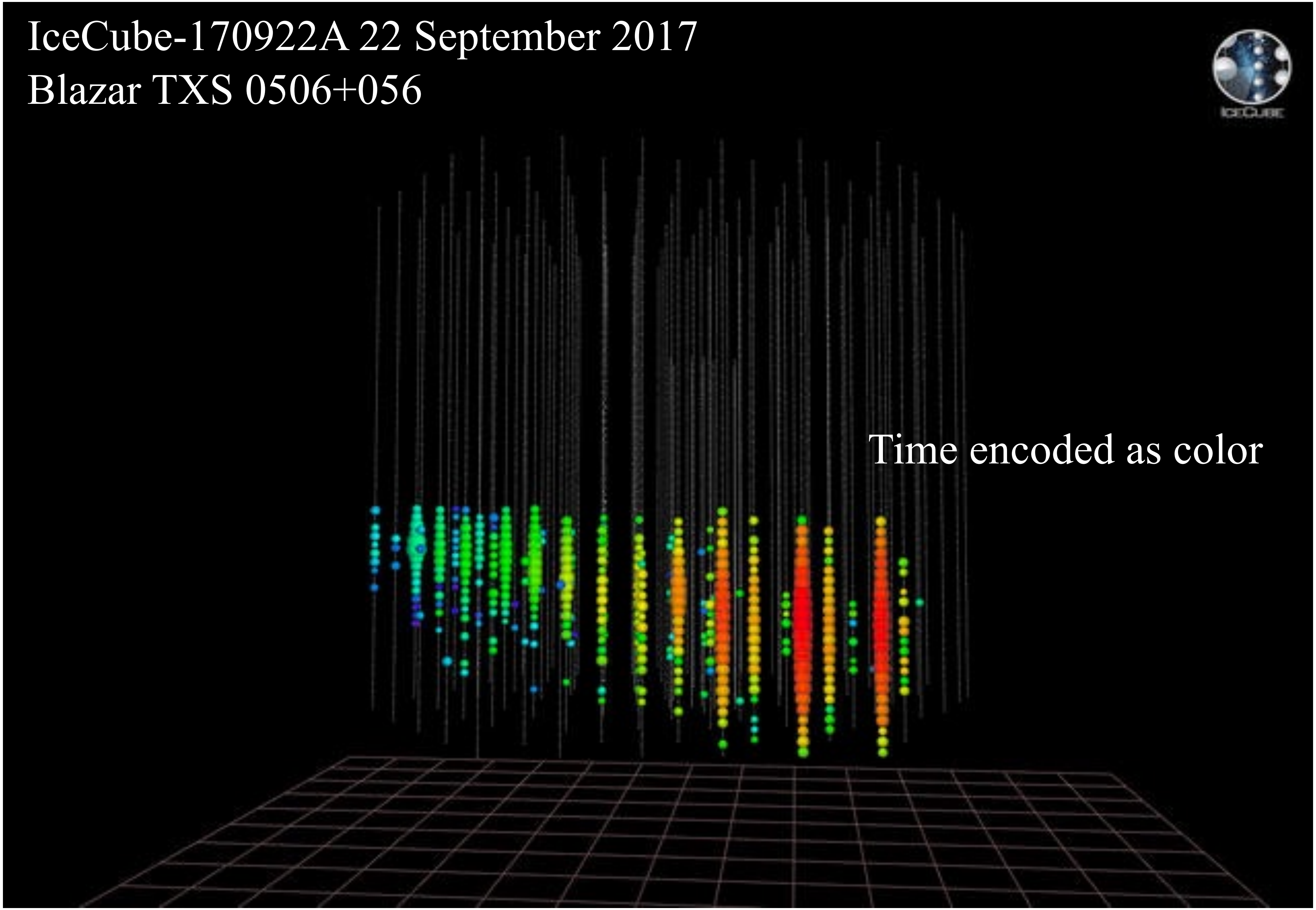
Across the graph, $\bar{\mathbf{e}}', \bar{\mathbf{v}}', \mathbf{u}$, are passed to a "global function":

$$\mathbf{u}' \leftarrow \phi^u\left(\bar{\mathbf{e}}', \bar{\mathbf{v}}', \mathbf{u}\right)$$



Important : the number of input e and v is not predetermined

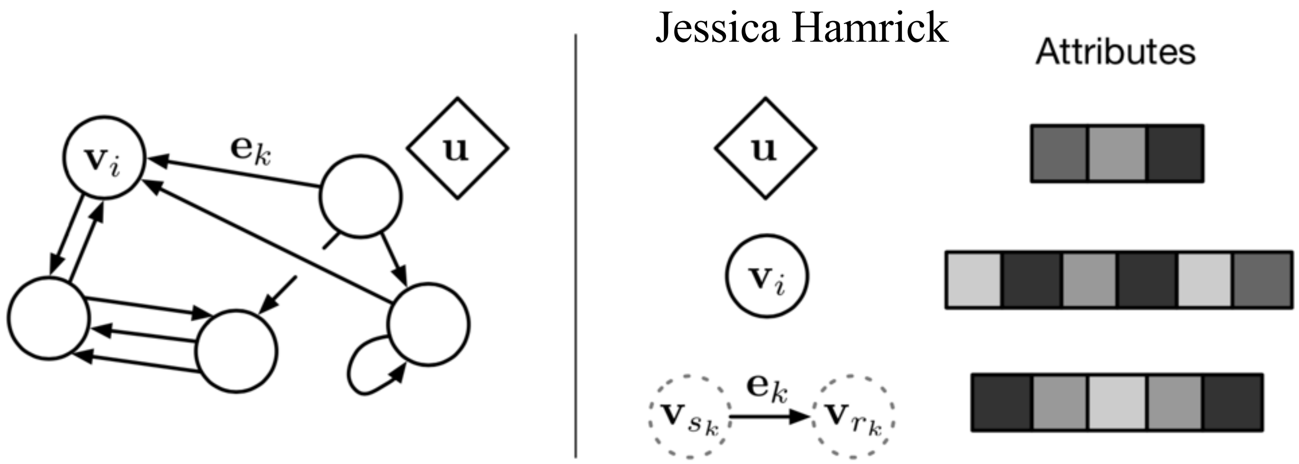
Ice Cube


IceCube-170922A 22 September 2017
Blazar TXS 0506+056
Time encoded as color

Graph NN for Ice Cube

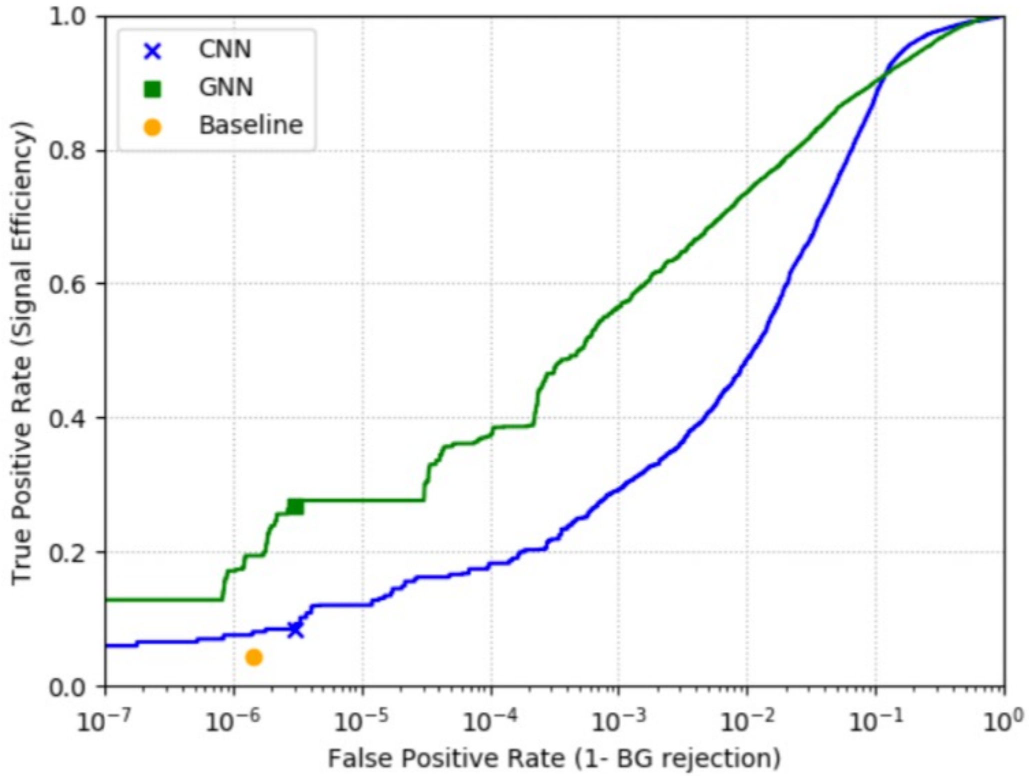
arXiv:1809.06166

- Graph NN: nodes,edges,and Globals
- …allow generalization of neighbouring pixels



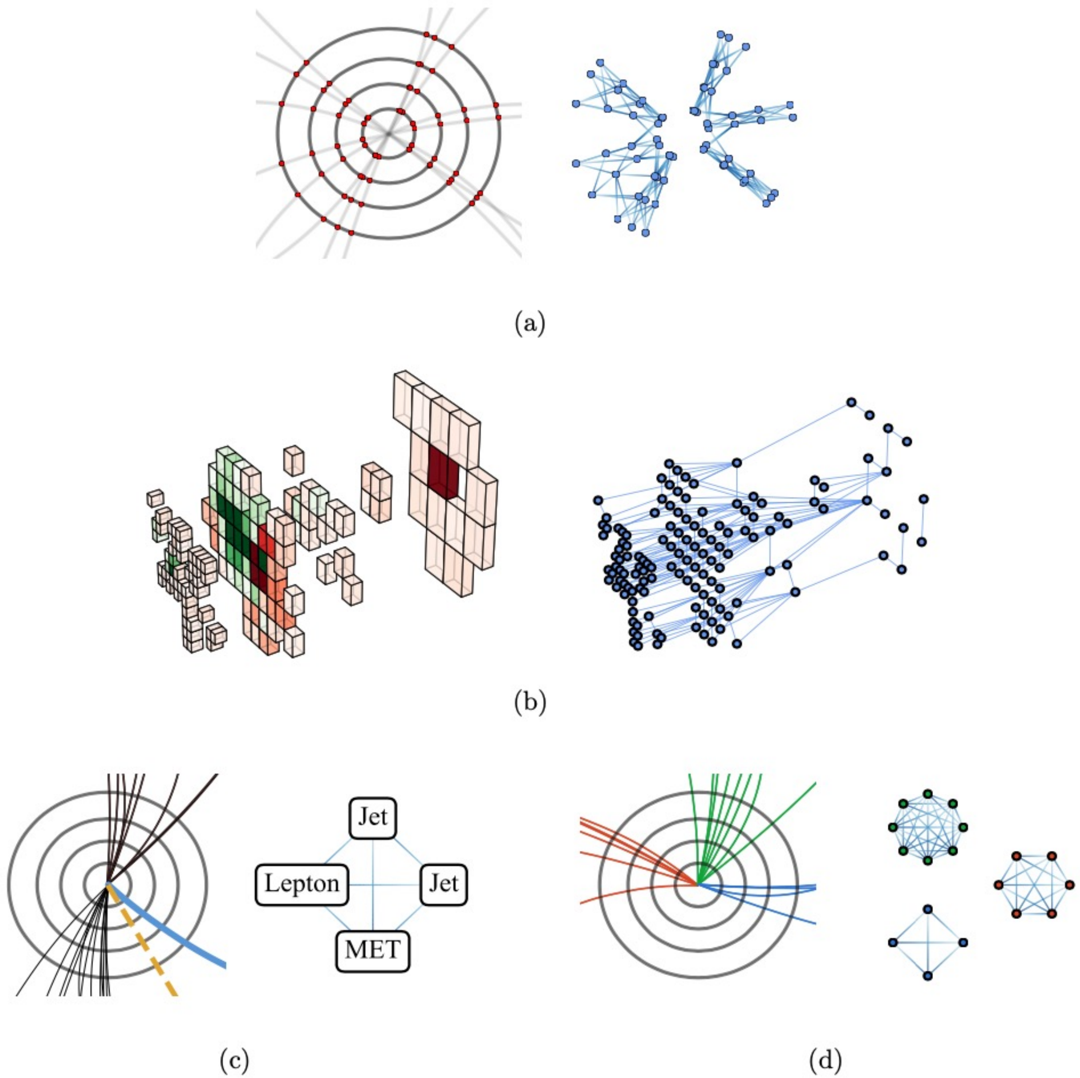
- Application to IceCube, separating downwards muon from neutrino from muon from cosmic rays

- ➔quickly growing interest in Graph NN in HEP



Graph on HEP data

from 2007.13681

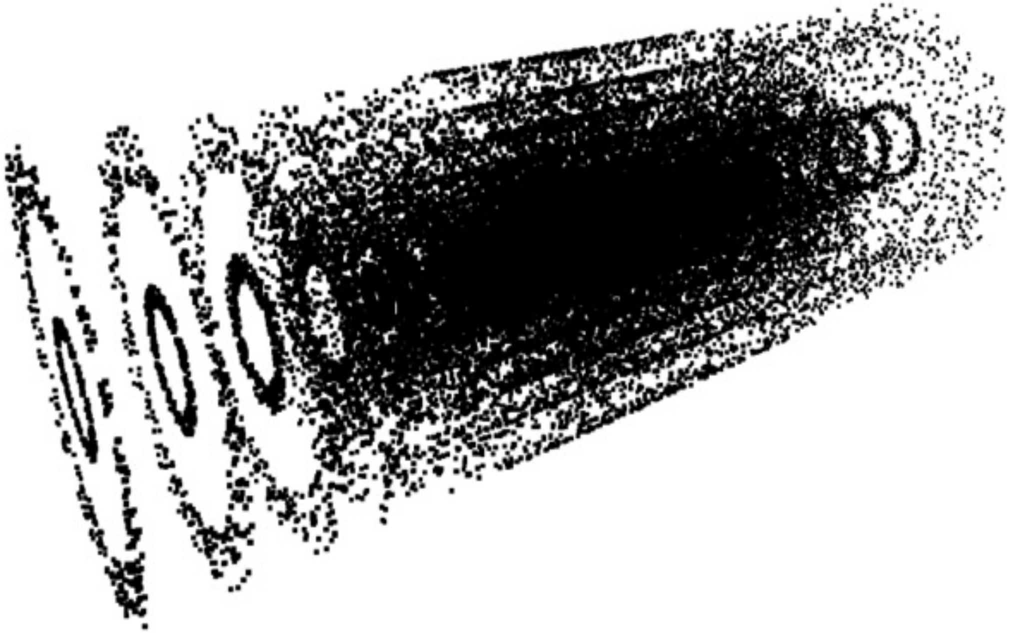


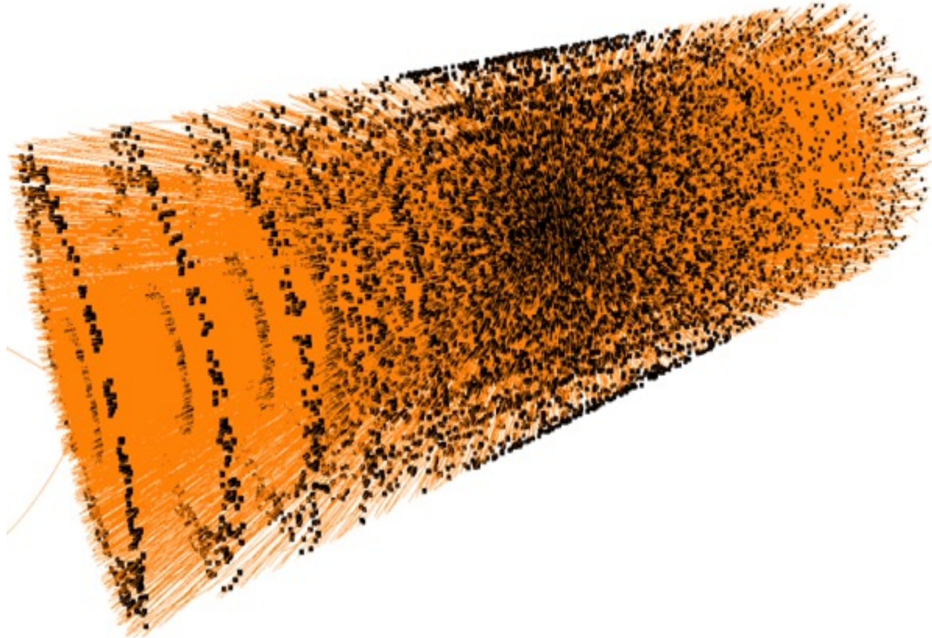
Tracking

TrackML challenge

A 3D point cloud challenge from
Particle Physics on Kaggle and Codalab

TrackML



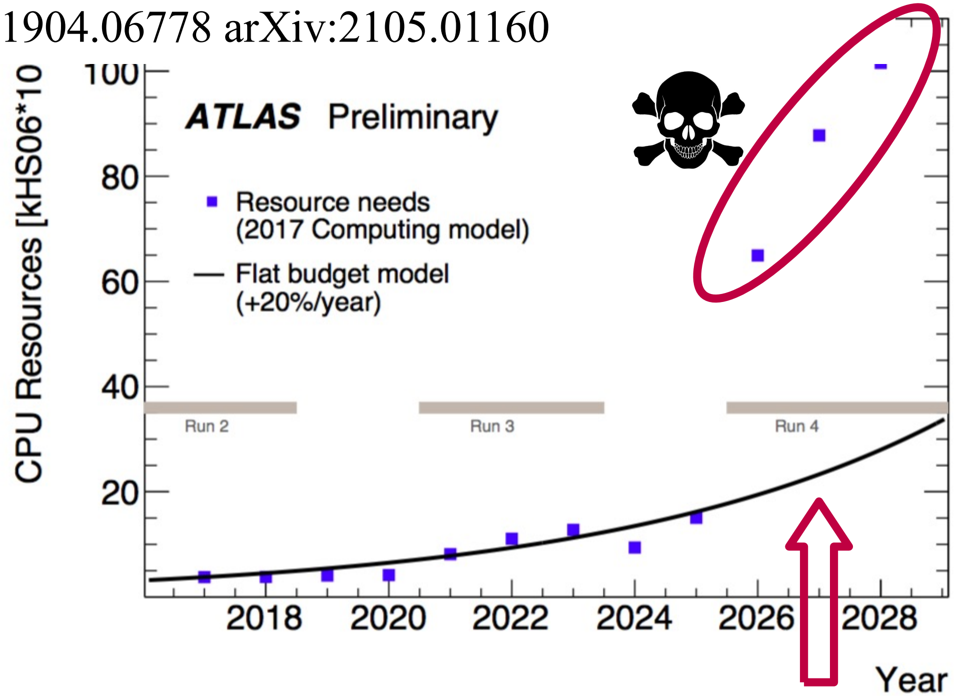


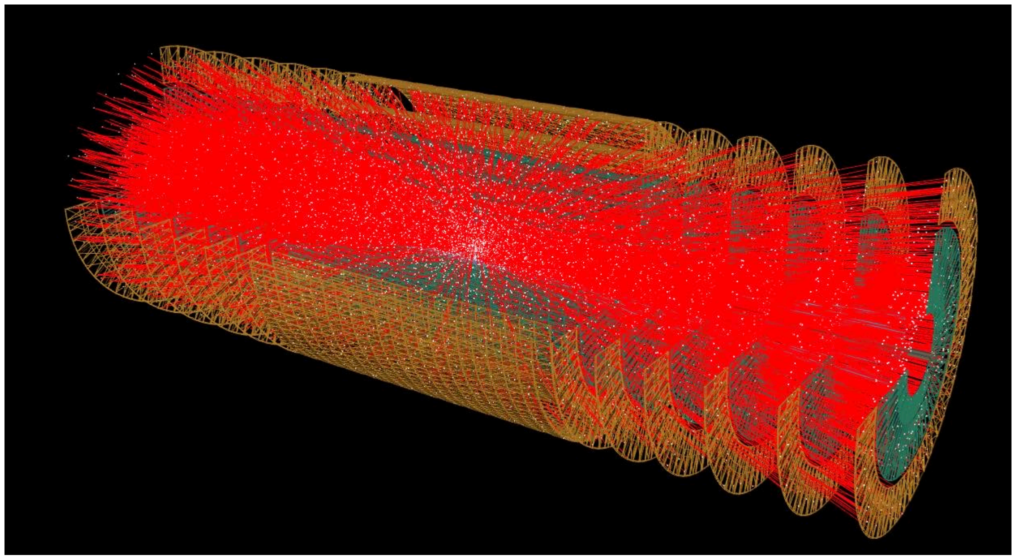
- All-silicon detector ➔ 3D points cloud
- 100’000 points to group into 10’000 tracks of 10 points
  - ➔~$10^{450'000}$ combinations
  - ⇒brute force has (really) no chance
- Precision of the points : ~50μm on a volume ~40 m^3
  - ➔3 10^{14} voxels!
  - 2D projection ➔2 10^9 pixels !
  - ⇒ naïve image recognition algorithm have (really) no chance
- Note : trajectories are approximate arc of helices, originating from the center

TrackML challenge

https://sites.google.com/site/trackmlparticle/ arXiv:1904.06778 arXiv:2105.01160

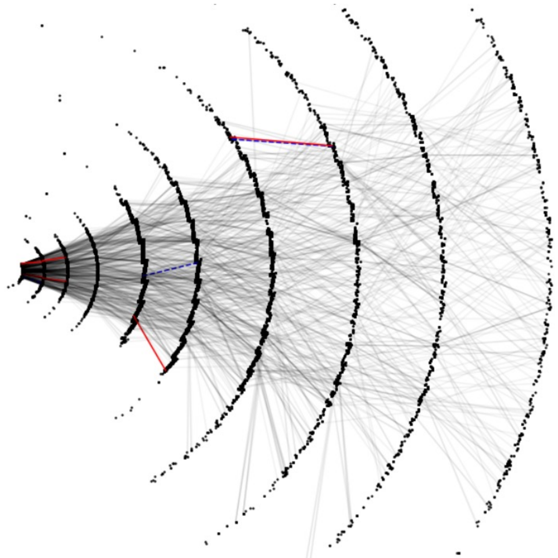
- Tracking (in particular pattern recognition) dominates reconstruction CPU time at LHC
- CPU time quadratic/exponential extrapolation (difficult to quote any number)
- Large effort within HEP to optimise software and tackle micro and macro parallelism. Not sufficient for future conditions
- >20 years of LHC tracking development. Everything has been tried?
  - Maybe no, brand new ideas from ML (i.e. Convolutional NN)
- ➔Tracking challenge 1st May 2018 to March 2019
- Conclusion : "ML-assisted" combinatorial tracking

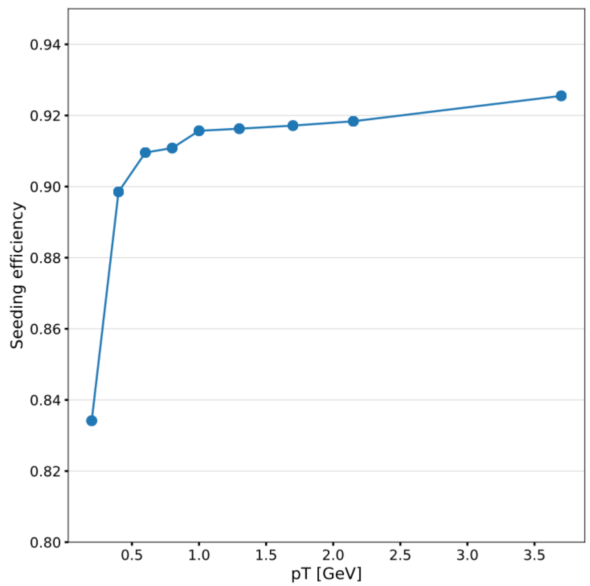




Track Seeding with GNN

[2007.00149](2007.00149)



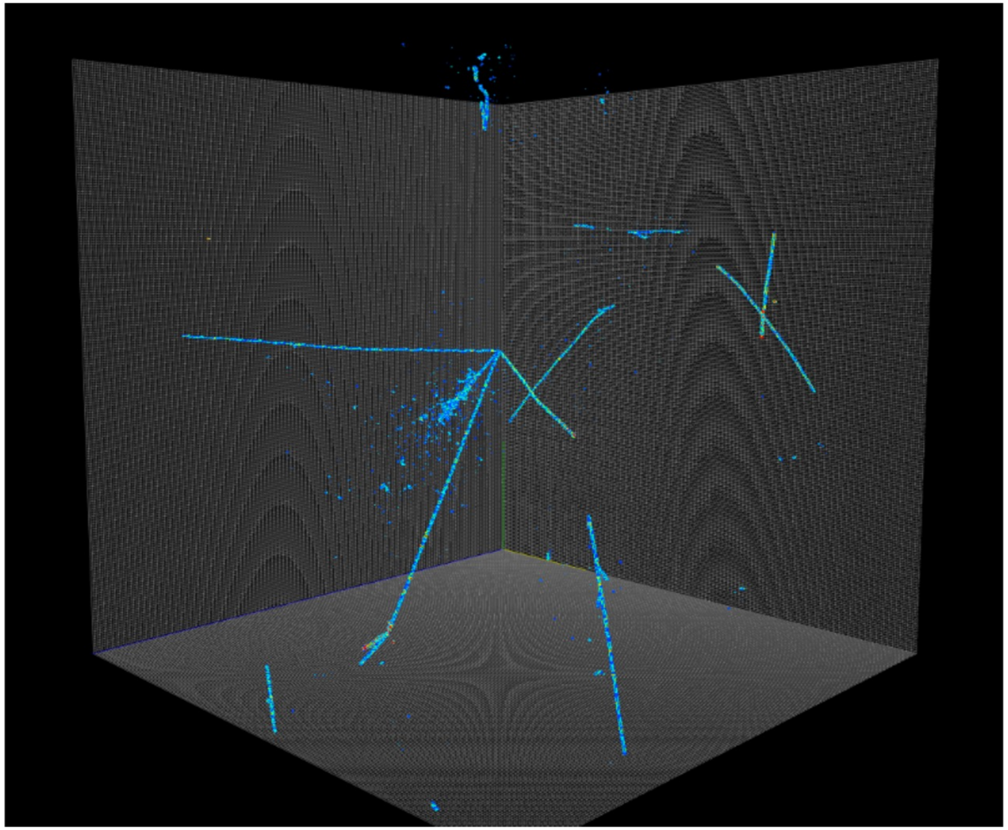


- Build edges between neighbour
- Then GNN trained to classify double and triplet
- High efficiency reached with subsecond computing time (also very parallelisabled)
- ➔can be used as a filtering stage before traditional Kalman filter

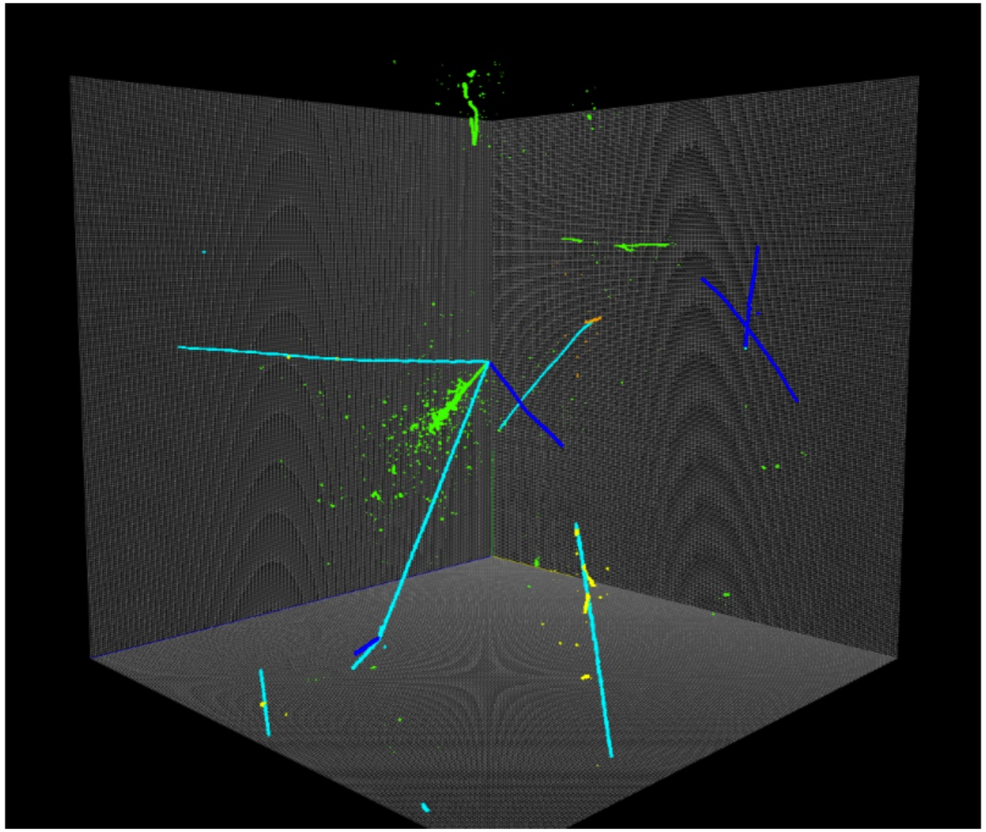
Cosmic ray in LArTPC

1903.05663

Simulated signal



Truth label



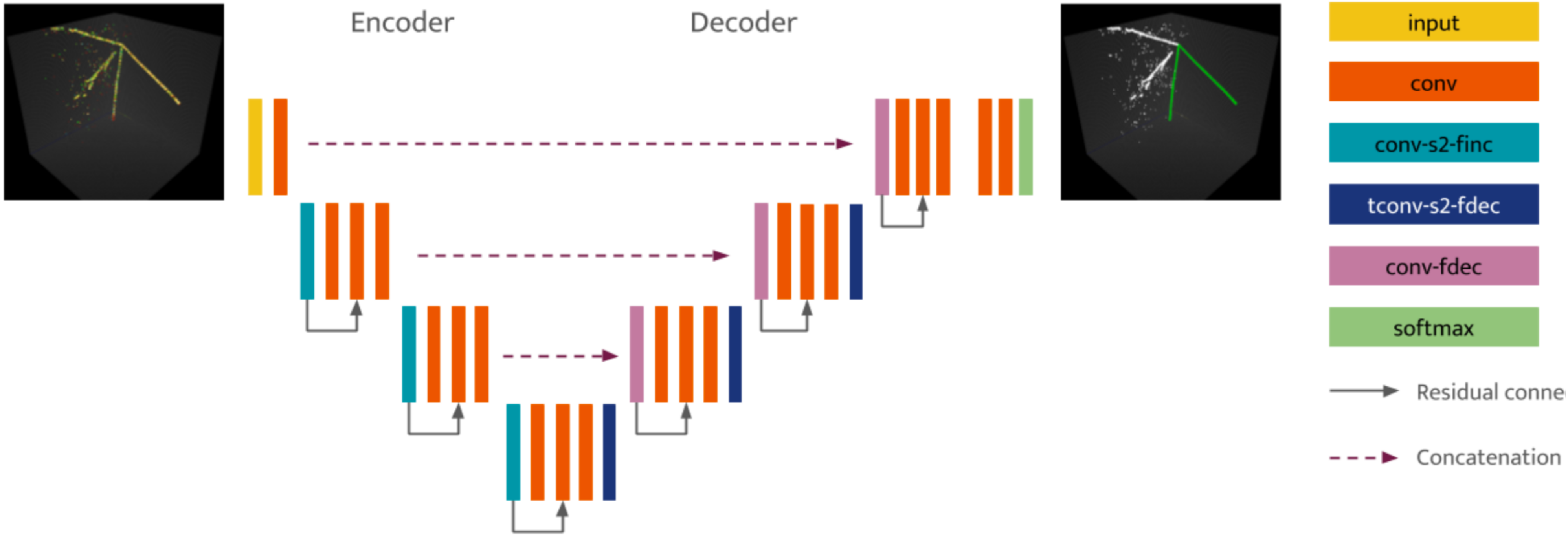
Full 3D information, sparse (non zero voxels <1%)

U-Resnet for semantic segmentation using Submanifold Sparse Convolutional Network

100 times less memory, 10 times faster than normal CNN, high accuracy

Cosmic ray (2)

U-Resnet for semantic segmentation using Submanifold Sparse Convolutional Network



100 times less memory, 10 times faster than normal CNN, high accuracy

Aparté on ML in HEP history

Computer Physics Communications 49 (1988) 429–448
North-Holland, Amsterdam

NEURAL NETWORKS AND CELLULAR AUTOMATA IN EXPERIMENTAL HIGH ENERGY PHYSICS

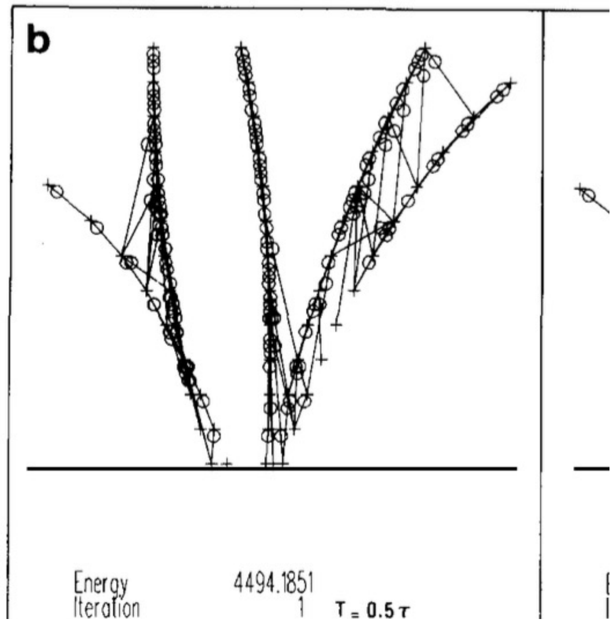
B. DENBY

Laboratoire de l'Accélérateur Linéaire, Orsay, France

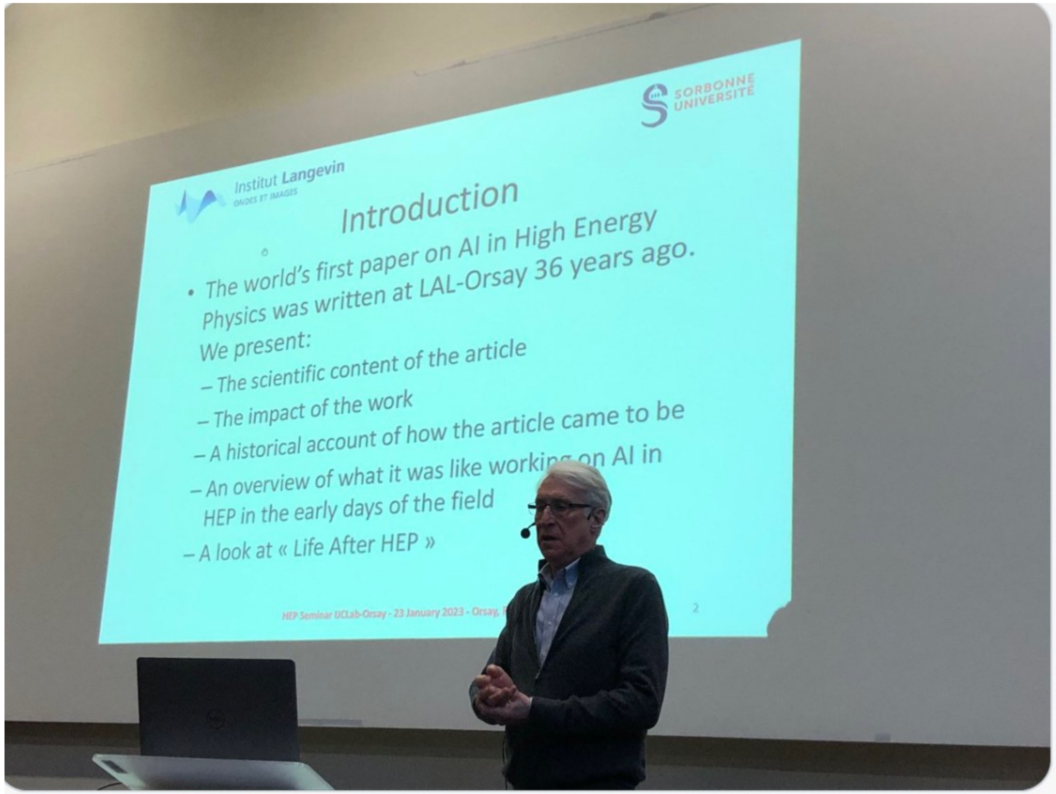
Received 20 September 1987; in revised form 28 December 1987

Bruce Denby

- 1987 Very first Neural Net in HEP paper known
- NN for tracking and calo clustering
- B. Denby then moved from Delphi at LEP to CDF at Tevatron. He still active outside HEP (professor at USPCI Paris): 2017 analysis of ultrasonic image of the tongue
- 1992 JetNet Carsten Peterson, Thorsteinn Rognvaldsson (Lund U.) , Leif Lonnblad (CERN) (~500 citations) really started NN use in HEP



❑ 23 Jan 2023 : Bruce Denby back at IJCLab 34 years later!

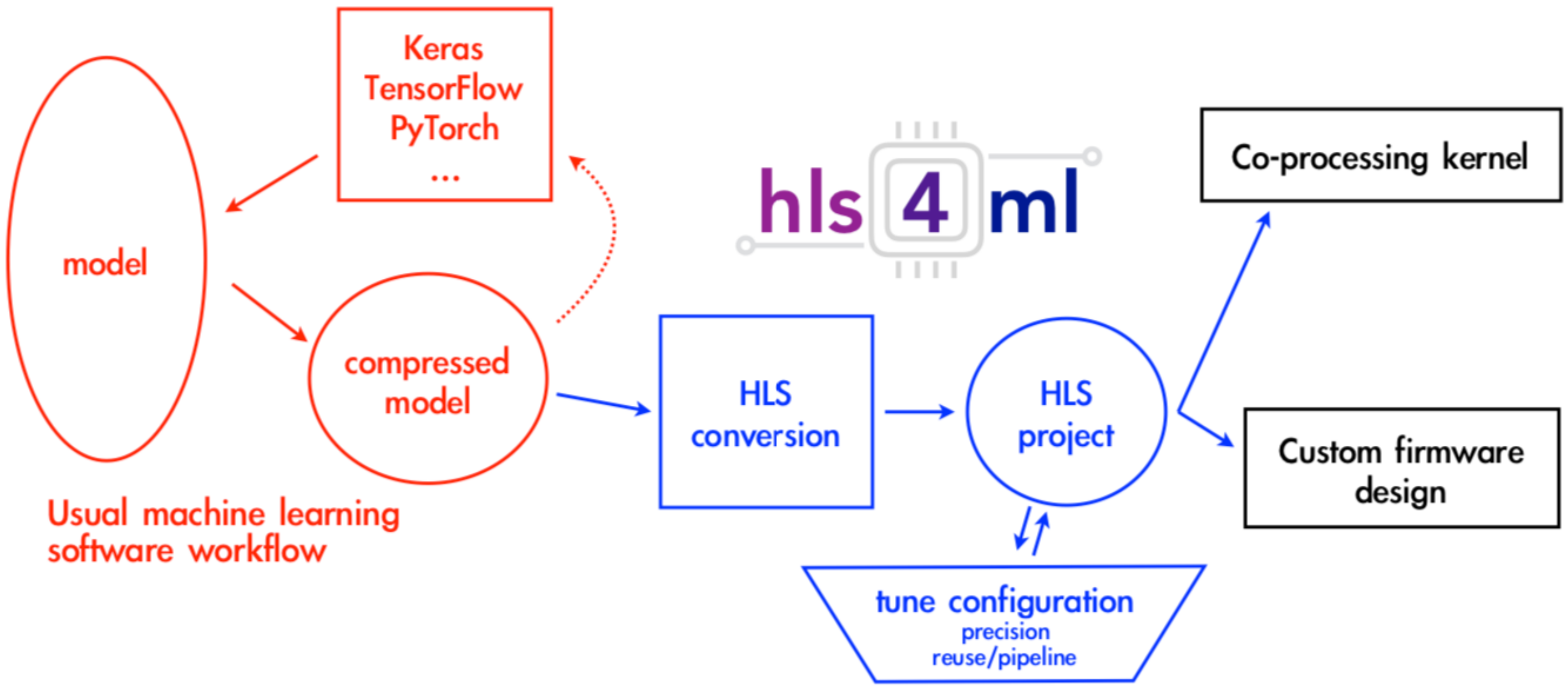


Useful to read popular science magazine!

ML for trigger

ML for trigger

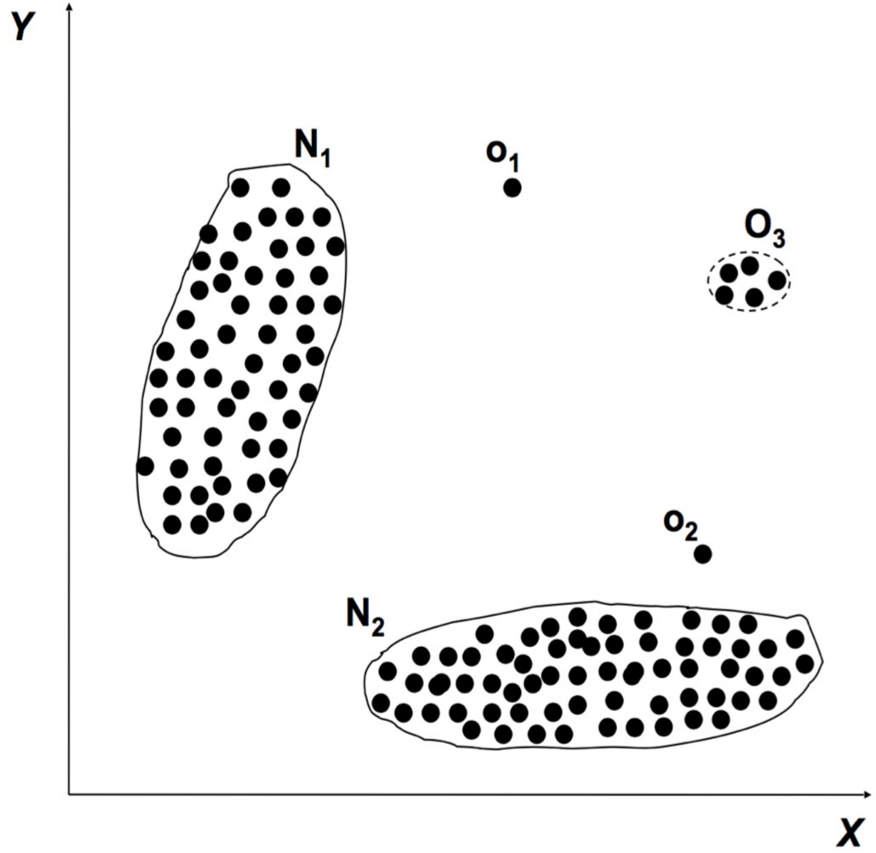
- Growing use of ML in triggers
- Classifiers (BDT or NN) maybe long to train but evaluation is very fast
- Can be implemented on GPU or FPGA
  - GBDT on FPGA (e.g. Yu Nakazawa, EPS-HEP 2019 conference)
  - NN on FPGA : hls4ML (arXiv:1804.06913)
  - See also Sridhara Dasu's talk yesterday
- Manpower efficient, few experts can code directly for GPU or FPGA



Anomaly detection

Anomaly detection

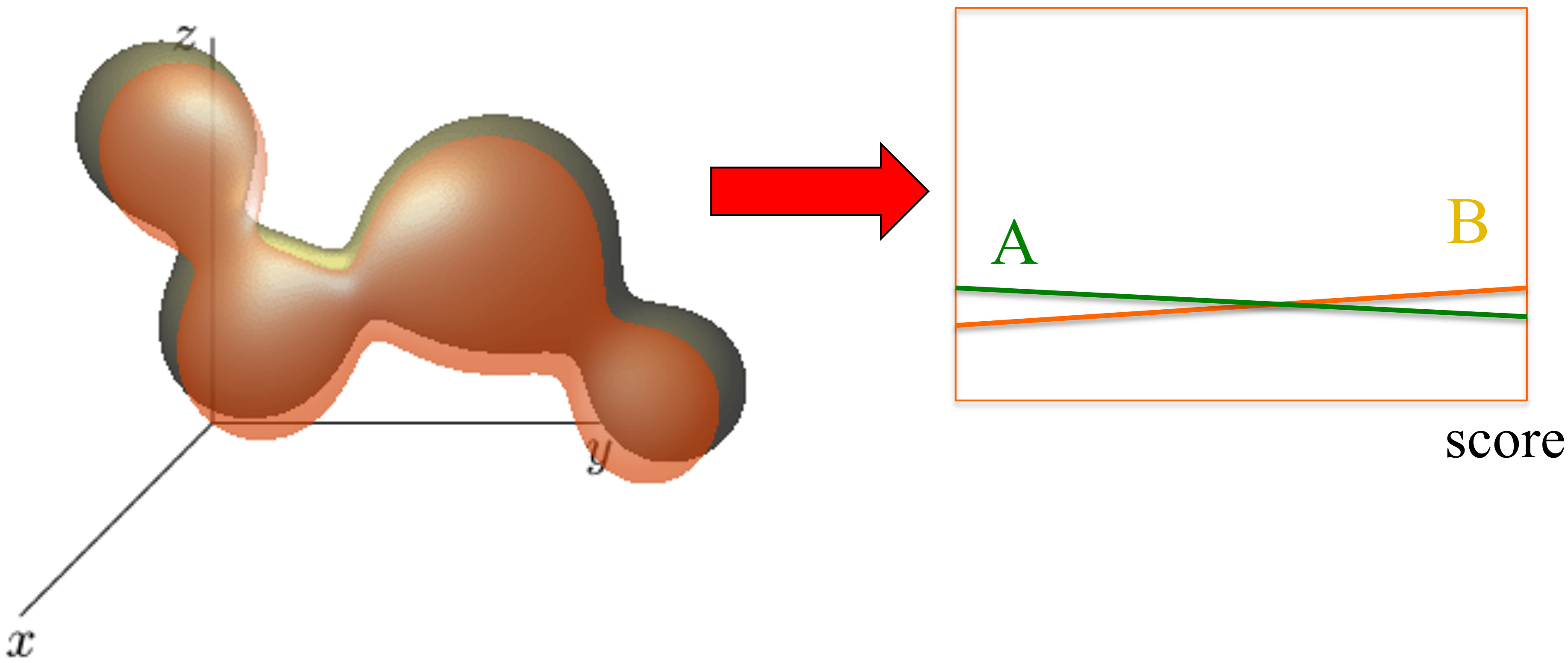
- Three approaches:
  - Supervised : model for O and N
  - Semi-supervised : model for N, O is non-N
  - Unsupervised : give the full data, ask the algorithm to cluster N and find the lone entries : o1, o2, O3



Anomaly detection: supervised

- Suppose you have two independent samples A and B, *supposedly* statistically identical. E.g. A and B could be:
  - MC prod 1, MC prod 2
  - MC generator 1, MC generator 2
  - Geant4 Release 20.X.Y, release 20.X.Z
  - Production at BNL, production at Lyon
  - Data of yesterday, Data of today
- How to verify that A and B are indeed identical ?
- Standard approach : overlay histograms of many carefully chosen variables, check for differences (e.g. KS test)
- One supervised ML approach (not the only one): ~~ask an artificial scientist~~, train your favorite classifier to distinguish A from B, histogram the score, check the difference (e.g. AUC or KS test)
  - →only one distribution to check
- Being developped for accelerator monitoring, experiment Data Quality monitoring

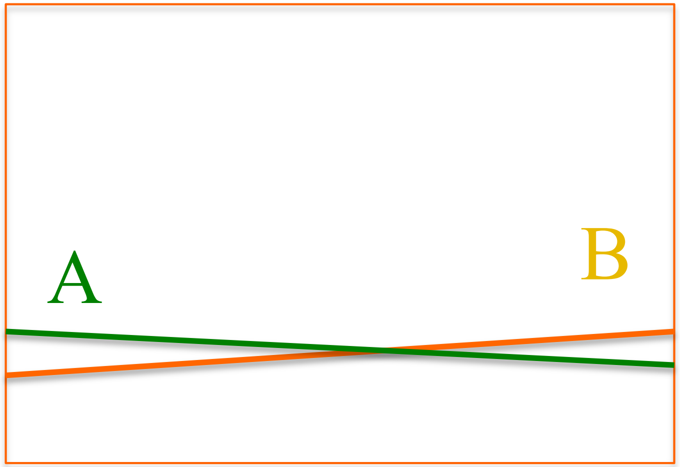
Recall what does a classifier do?

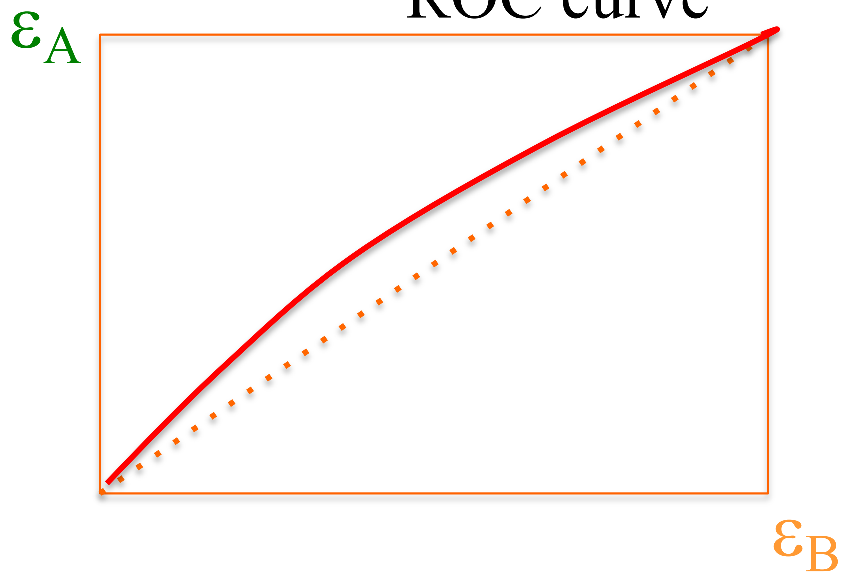


- The classifier "compresses" the two multidimensional "blobs" maximising the difference, without (ideally) any loss of information

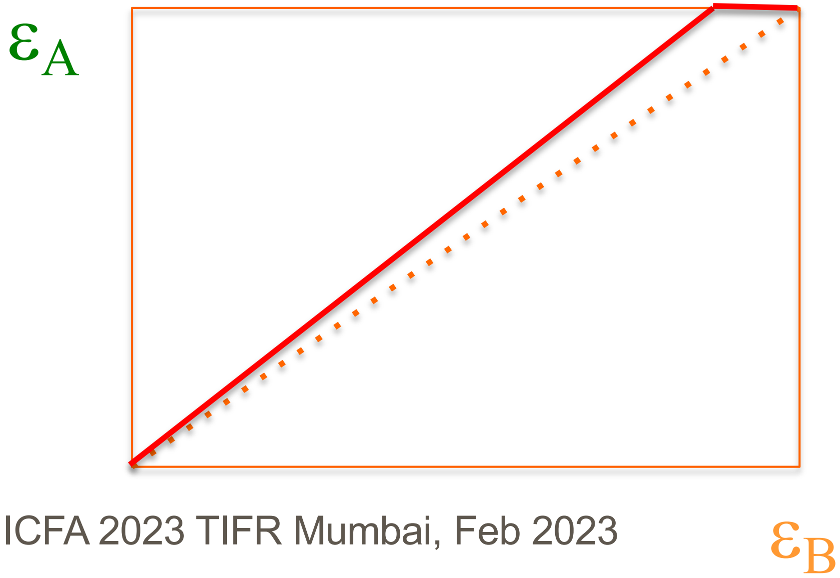
Small non-local difference

ROC curve





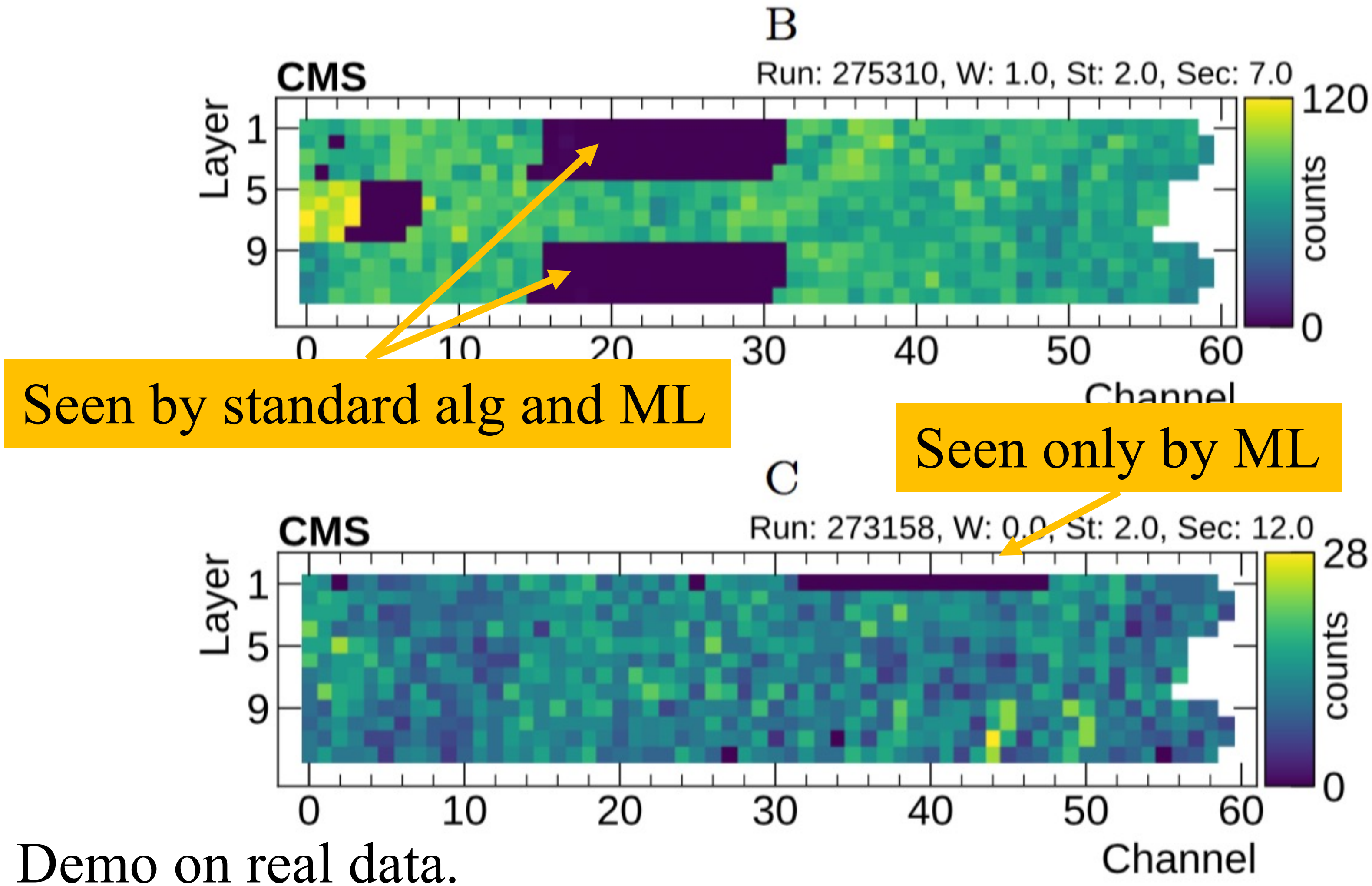
Local big difference (e.g. non overlapping distribution, hole)



Anomaly Detection : Data Quality Monitoring

Semi-supervised: DQM application

- Example application CMS muon chamber monitoring (with Convolutional NN)



Demo on real data.

Application to new physics

Maurizio Pierini

End of Part 2