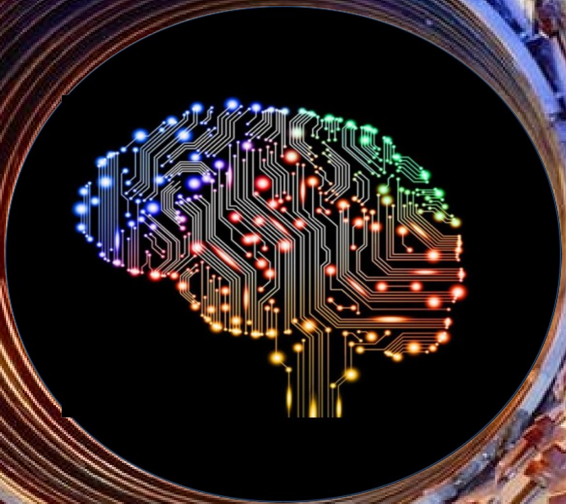
Application of Machine Learning for LHC experiments operations

I Workshop de computing y software de la red española de LHC

Pablo Martínez Ruiz del Árbol

Instituto de Física de Cantabria

EXCELENCIA MARÍA DE MAEZTU



Use of new computing technologies in HEP

- The new computing paradigms are radiply growing inside the HEP community
- Complex and rich environment of new techniques with two clear accelerators:
  - New hardware architectures (GPU, FPGA, TPU)
  - Developments in AI and Deep Learning
- These techniques are being used in almost all aspects of the LHC experiments.
- Focusing here in Machine Learning in the LHC experiments outside of Physics Analysis.
  - I will be showing a personal, biased set of examples that I find interesting

Parallelization (GPU&FPGA)

- Fast Reco (Trigger)
- Fast Simulation

ML for Analysis

- Classification:
  - Binary (BDT, ANN)
  - Multiclass (DNN)
  - Decorrelation
- Anomaly detection

ML for operations

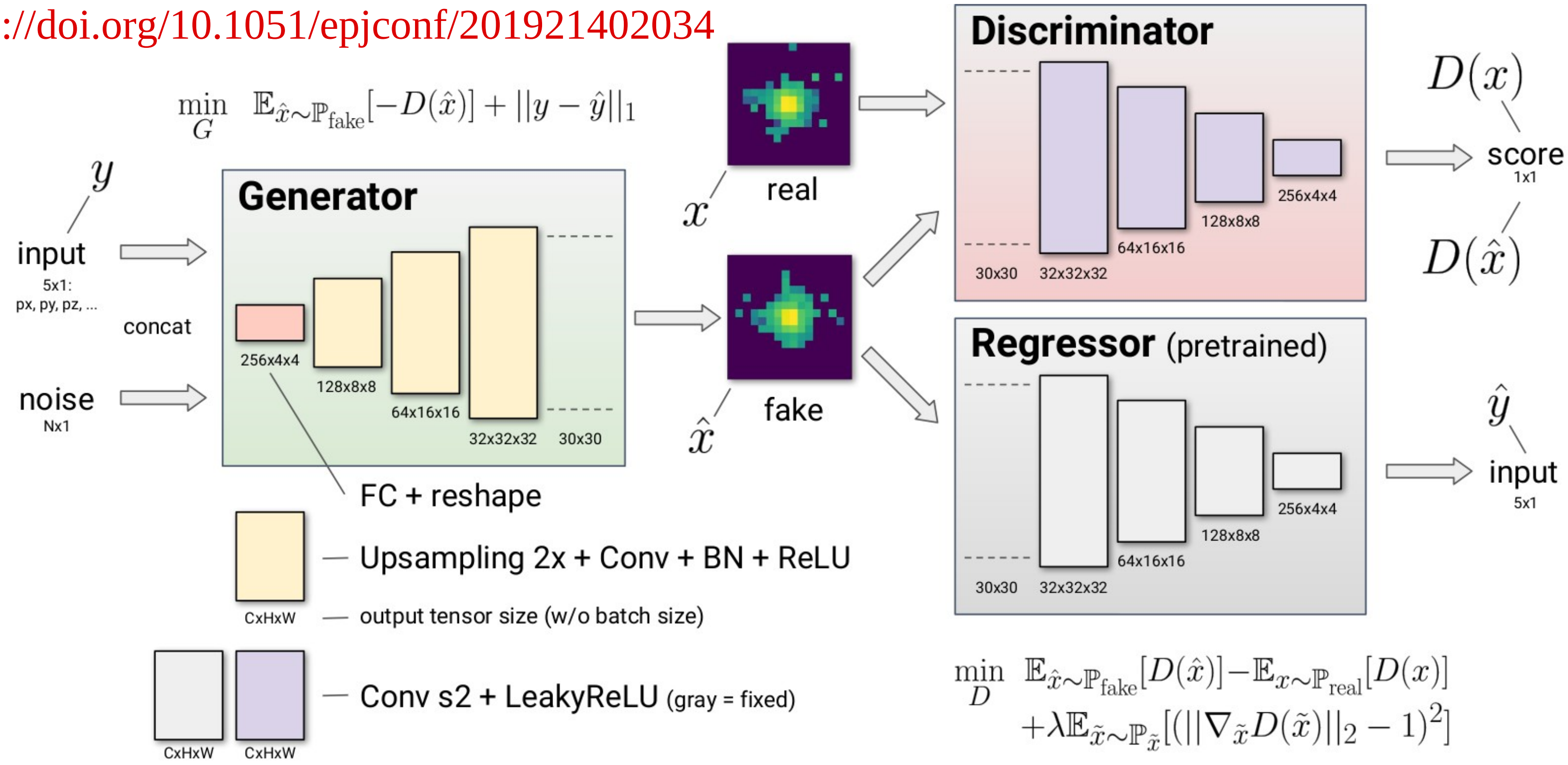
- Fast simulation
  - GAN, DAN
- Reconstruction
  - Vertexing, jets, etc
- Trigger
- DQM
  - Anomaly detection

Fast simulation

GAN architectures for LHCb Fast Simulation

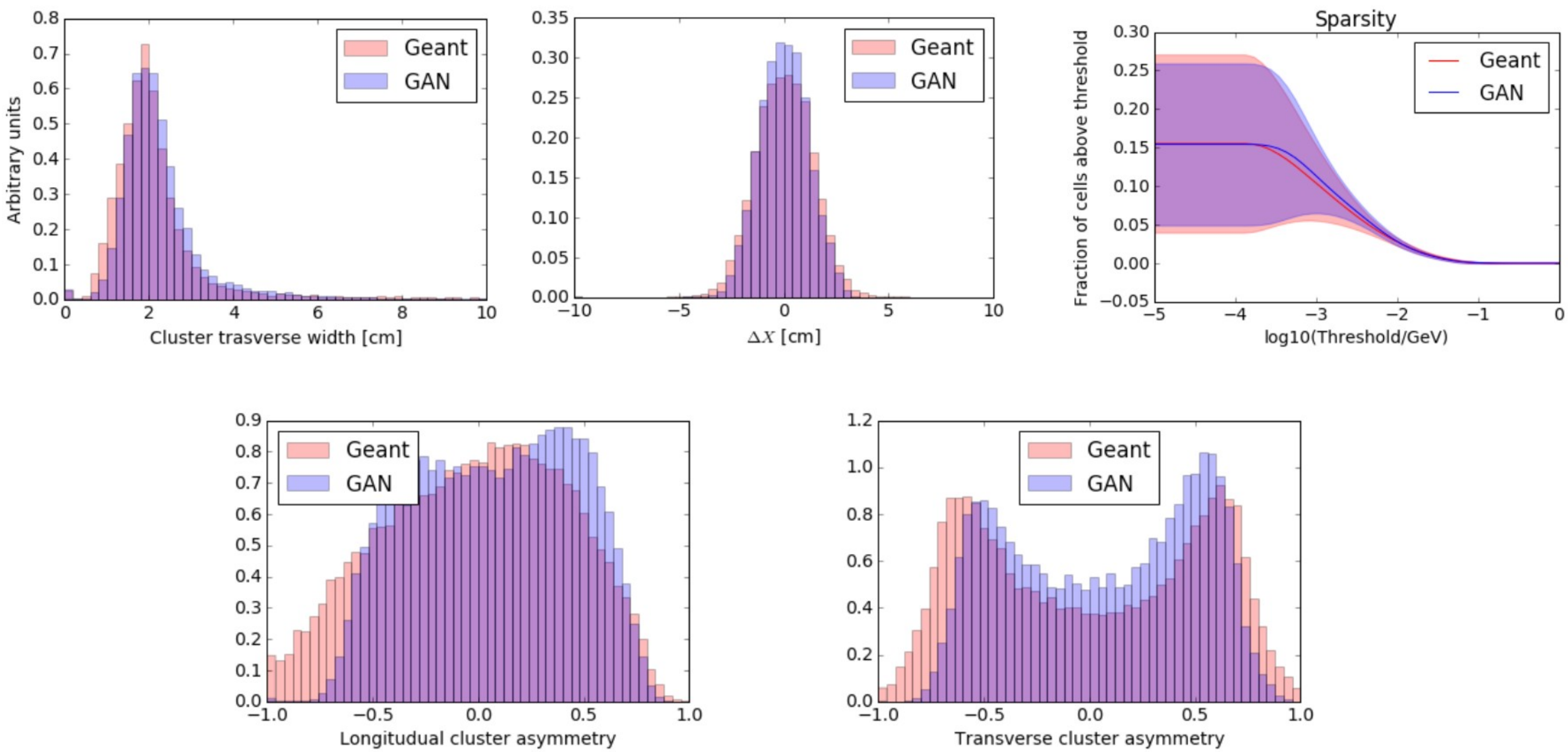
- This machinery aims at simulating the electron interactions in the ECAL of LHCb.
  - The system simulates the energy deposition in a 30x30 matrix of ECAL cells.
- A Wasserstein Generative-Adversarial-Neural Network is used as learning scheme
  - A regressor block is added in order to predict the momentum of the incoming particle.

https://doi.org/10.1051/epjconf/201921402034



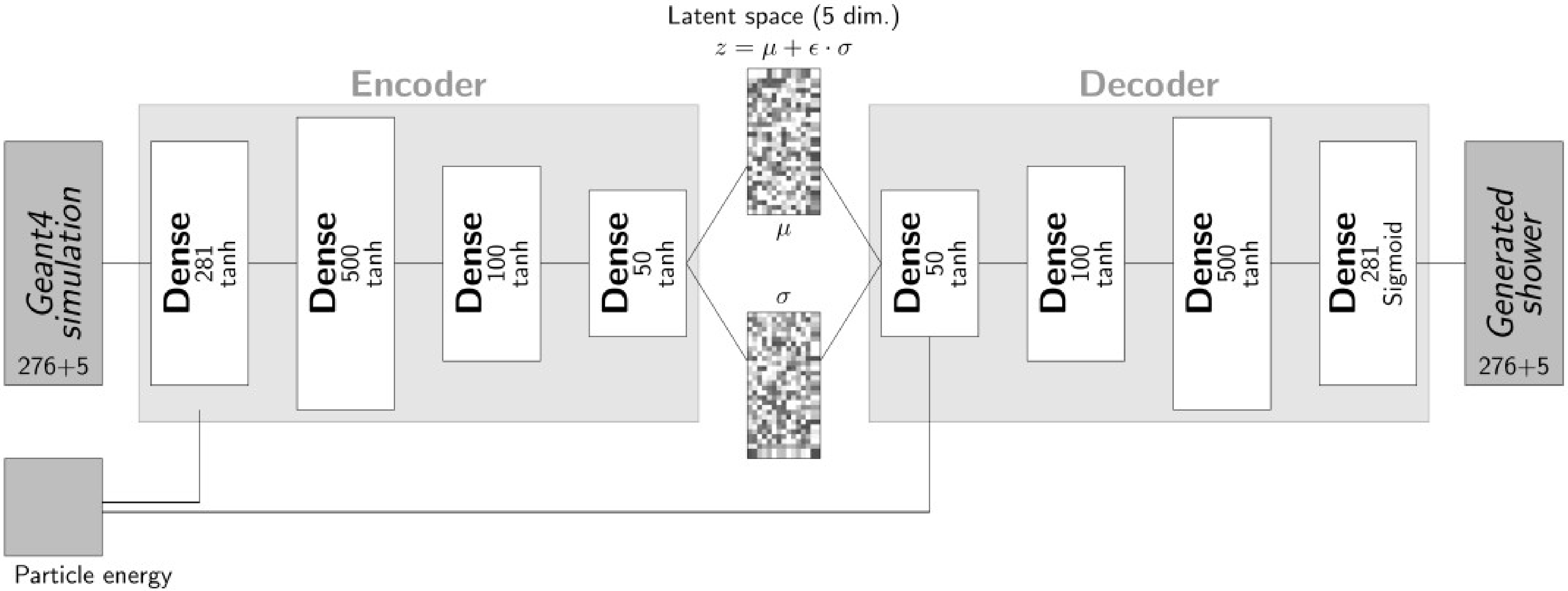
GAN architectures for LHCb Fast Simulation

- The GAN is trained with detailed GEANT4-based simulations
  - A total of 50000 events for the training + 10000 events for the test datasets
- A reasonable agreement between GEANT4 and the GAN is found for the main features
- The speed up in the generation is x10000 with respect to the detailed GEANT4



VAE architectures for ATLAS Fast Simulation

- ATLAS has also studied the simulation of the ECAL showering for photons.
  - Two algorithms: GAN model and a Variational Auto-Encoders (VAE)
- The target (as for the LHCb case) is to generate the energy deposition in a block of cells.
  - A total of 266 ECAL cells are considered from the different ECAL layers.

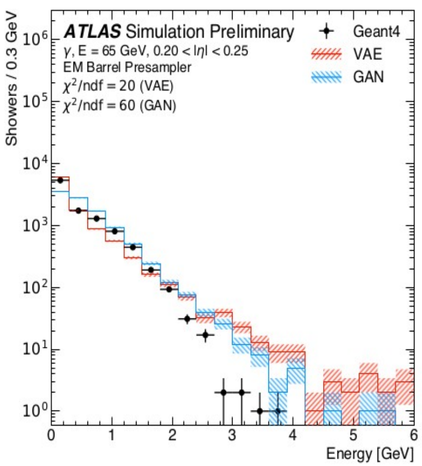


ATL-SOFT-PUB-2018-001

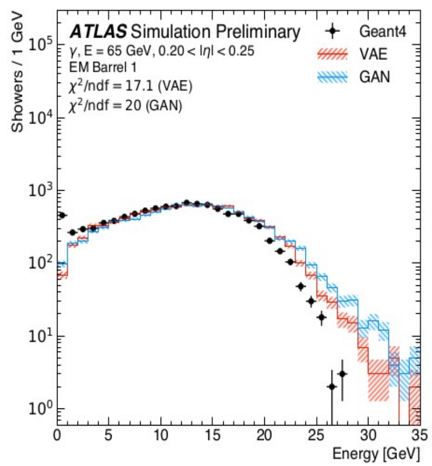
VAE architectures for ATLAS Fast Simulation

- The system is trained using a detailed GEANT4-based dataset with 90000 events.
  - Divided in 9 blocks of 10000 with 9 different incident energies.
  - Only one region of the calorimeter is taken into account (fixed phi and eta)
- The agreement between the VAE and the Geant4 is reasonably good
  - But still far to be used for precision measurements
  - The GAN approach (not explained here) seems to have a better performance.

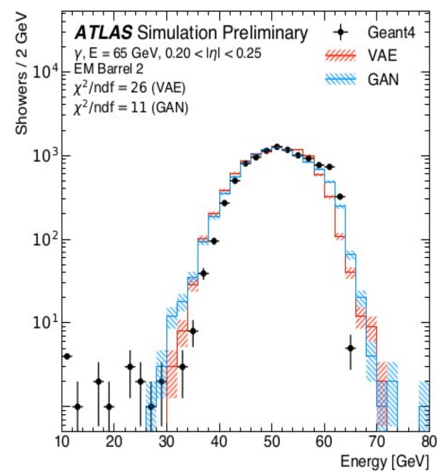
Presampler



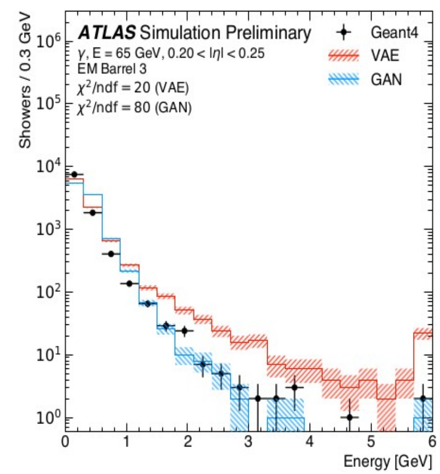
Front layer



Middle layer



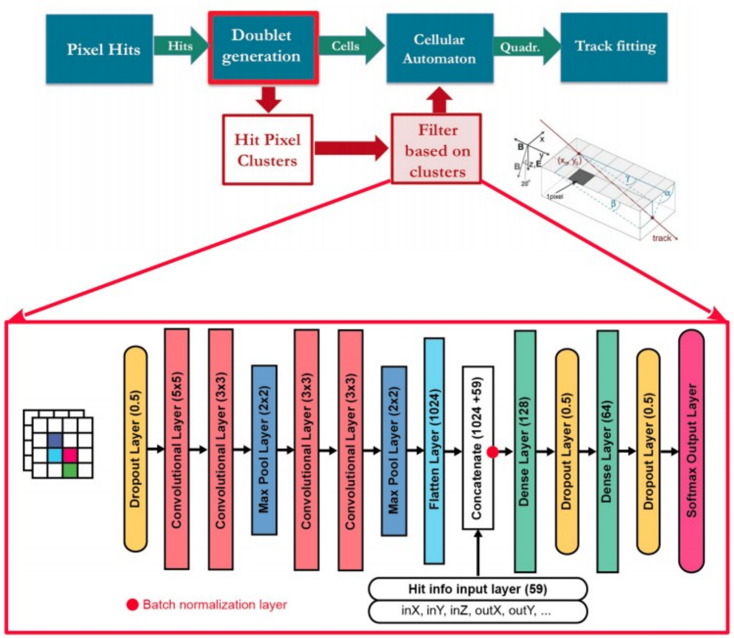
Back layer



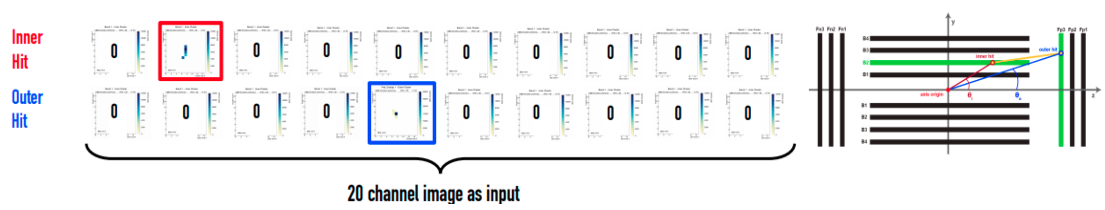
Trigger

CNN for track seed filtering at the CMS HLT

- Tracks reconstructed with the pixel detector are used online for fast tracking and vertexing
- This is a challenge for Phase2 where the PU is expected to scale up to 200
- CMS is devising a full, parallelizable HLT RECO running on GPUs and using CAs
  - Still there is a bottleneck on the number of “doublets” that will be further processed
  - A CNN has been proposed to filter these seeds as a classification problem (Valid or not)
  - Hits represented as 16x16 pixel pads images with colors proportional to deposited charge



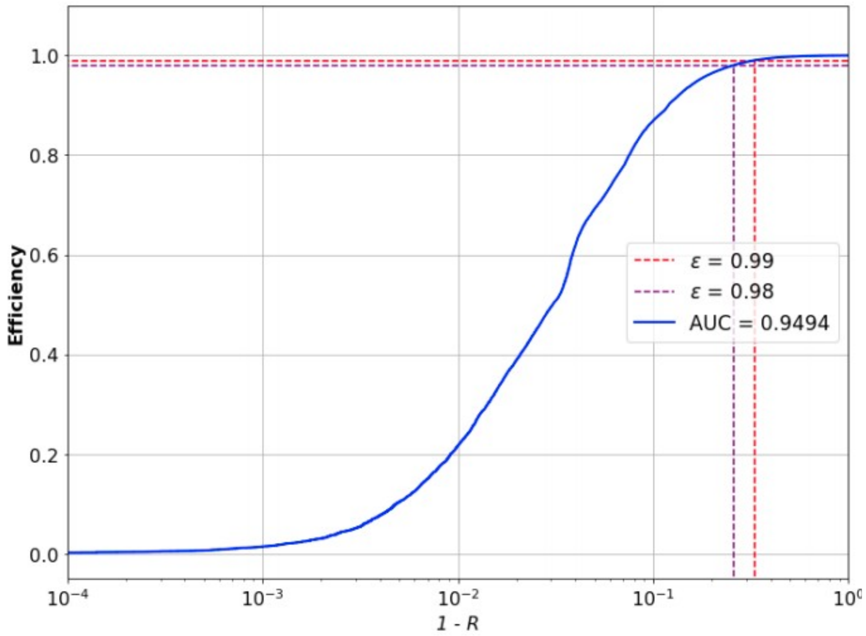
- Images are combined in 20 channels/levels
- Accounting for the different inner/outer layers

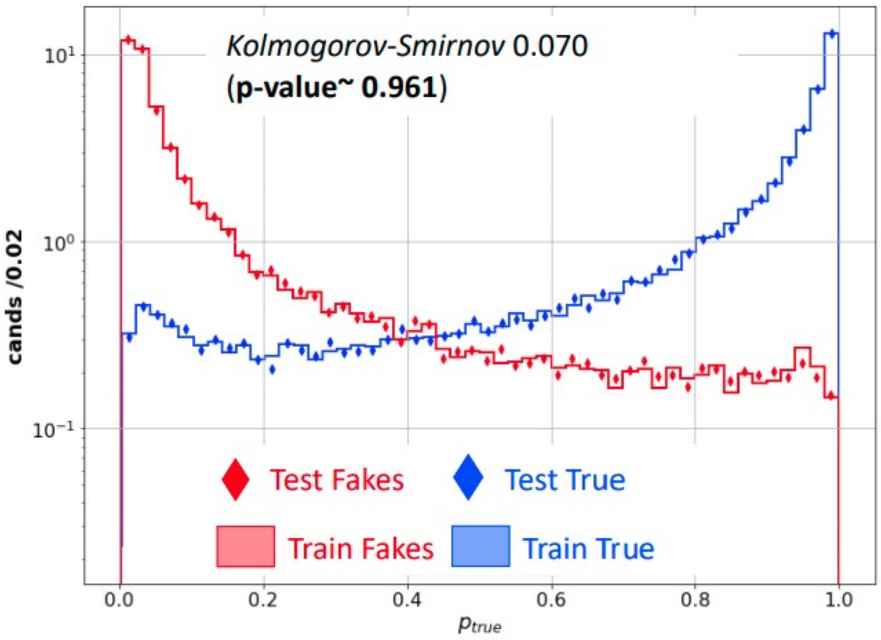


https://indico.cern.ch/event/819693/contributions/3438504/attachments/1858975/3054502/Patatrack_DiFlorio_CMSCalcolo.pdf

CNN for track seed filtering at the CMS HLT

- Test have been done with a training on O(10^7) doublets from RECO simulation
  - Obtained with only O(100) events
  - True doublets are those where the hits can be matched to a same GEN particle
- The system retains about 99% of efficiency while 2/3 of the fake doublets are rejected

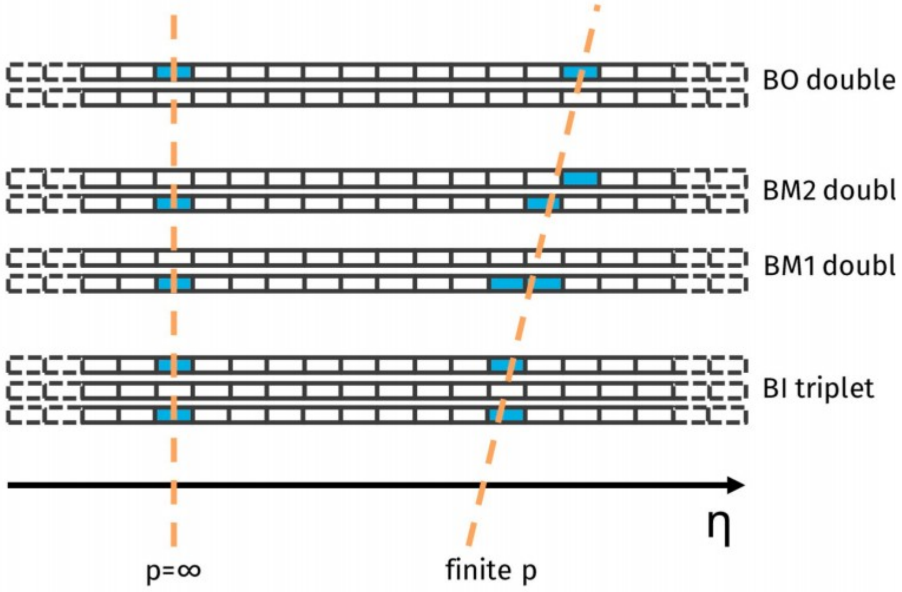


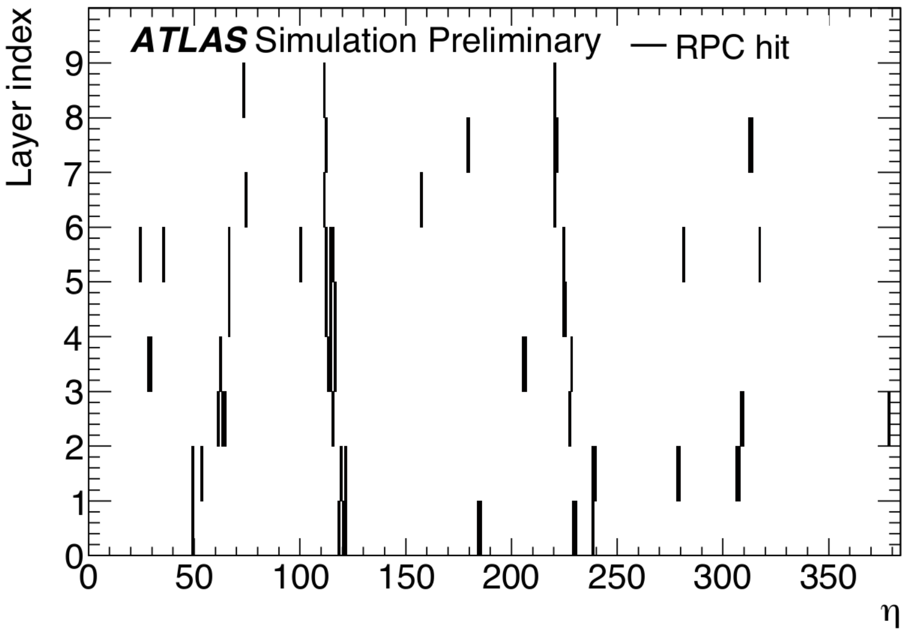


ATLAS Phase-II RPC Muon Barrel Trigger

- The ATLAS collaboration is working on a CNN running on an FPGA for the muon trigger
- Events are interpreted and treated as images that are further fed into a CNN
  - The RPC hits are represented as eta Vs. layer maps in the RPCs
  - Image size is 384 bins in eta x 9 RPC stations
- The CNN performs a regression to 5D space [$p_T^{leading}$, $\eta^{leading}$, $p_T^{leading}$, $\eta^{leading}$, # muons]

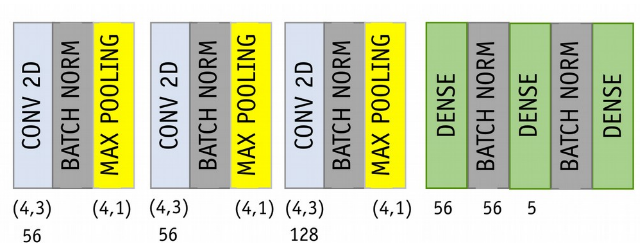
ATLAS-L0-MUON-PUBLIC

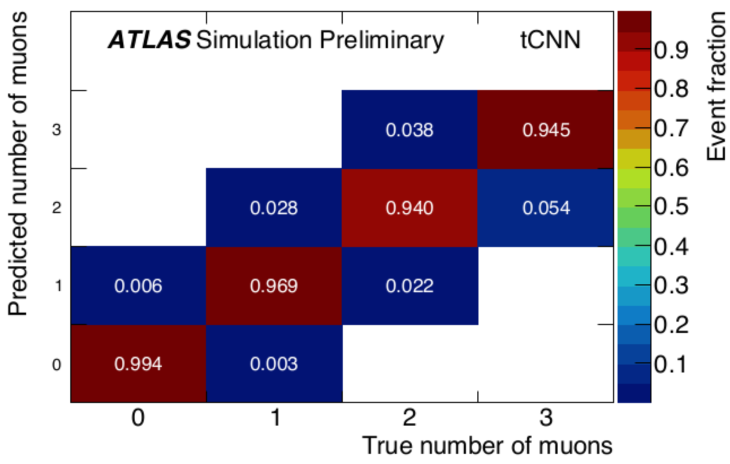


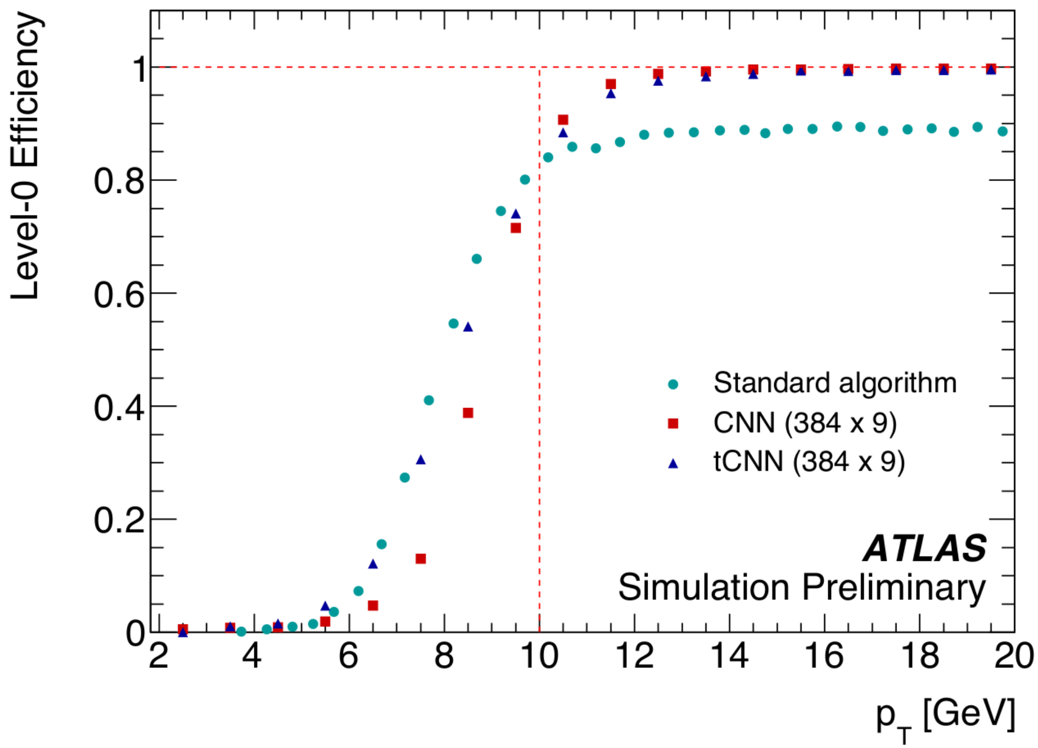


ATLAS Phase-II RPC Muon Barrel Trigger

- The size and architecture of the CNN has to match the characteristics of the FPGA
- In order to reduce memory consumption a “Ternary CNN” is proposed
  - Weights and activations can only take {-1, 0, 1} values instead of floating point.
  - Memory is reduced by a factor 16 thanks to this procedure
- The network outperforms by ~10% the classical algorithm in terms of efficiency.



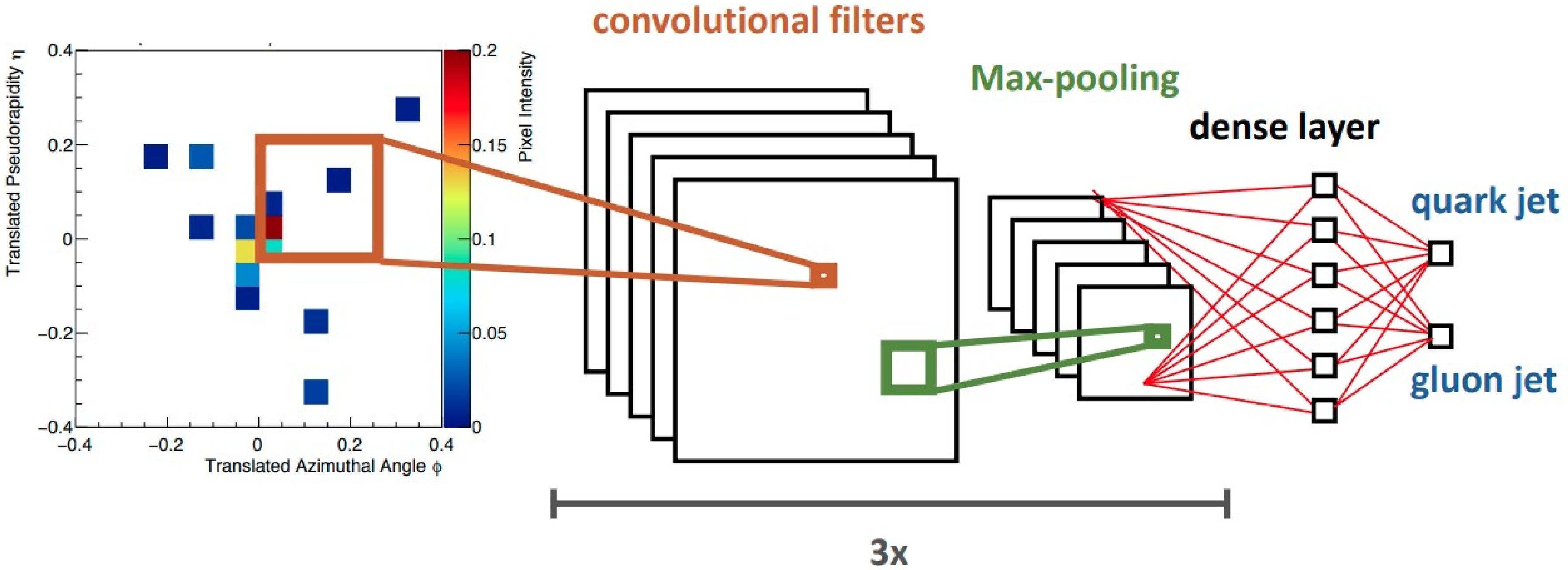




Reconstruction/ Identification

ATLAS quark-gluon discriminator with CNN

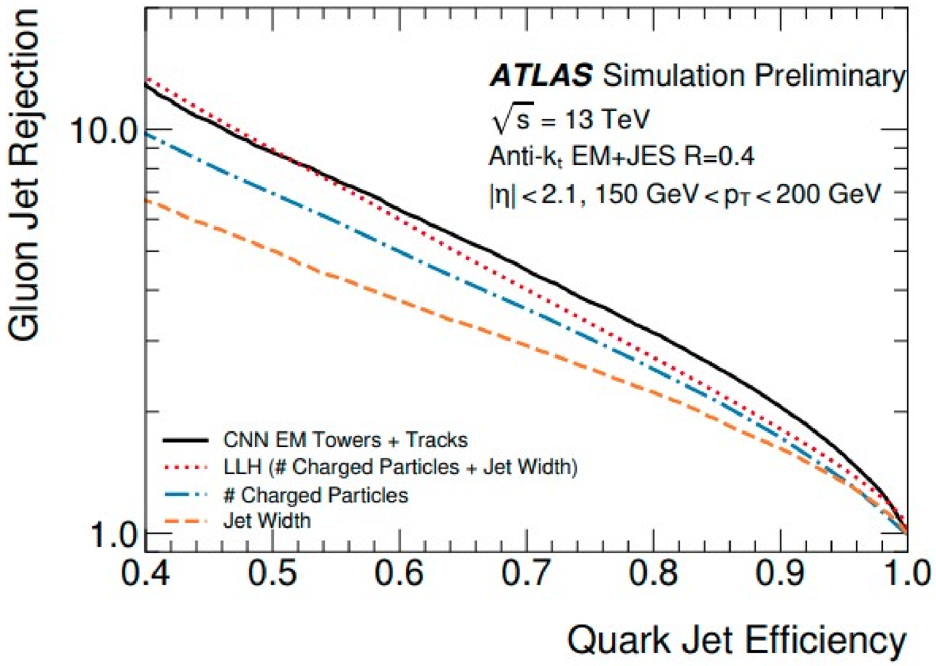
- ATLAS has also explored Convolutional Neural Networks to learn jet substructure
- Jet constituents are represented in eta – phi images with 16x16 binning
  - Tracks and tower or topocluster information are represented in different images
  - The color is proportional to the pt of the constituent (and then normalized)
  - The best performance is found when combining the track + tower/topocluster input

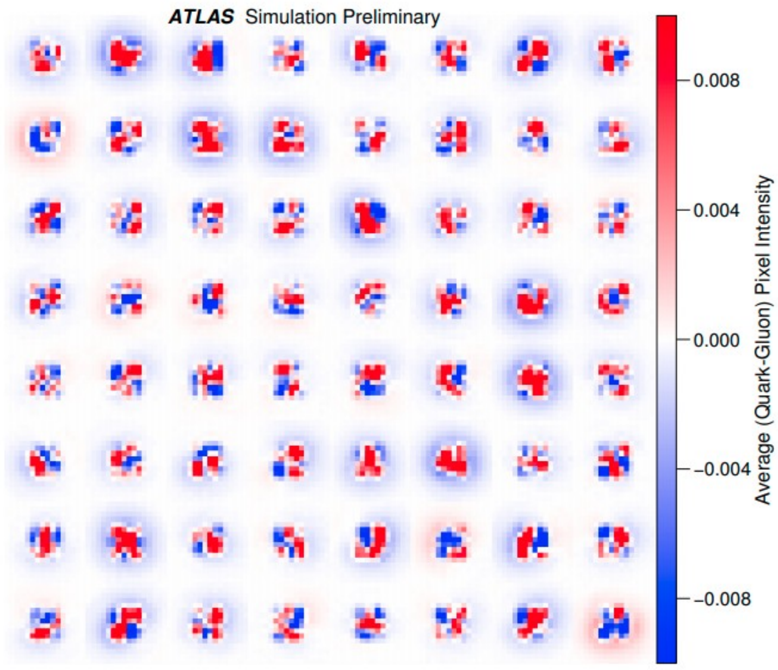


ATL-PHYS-PUB-2017-017

ATLAS quark-gluon discriminator with CNN

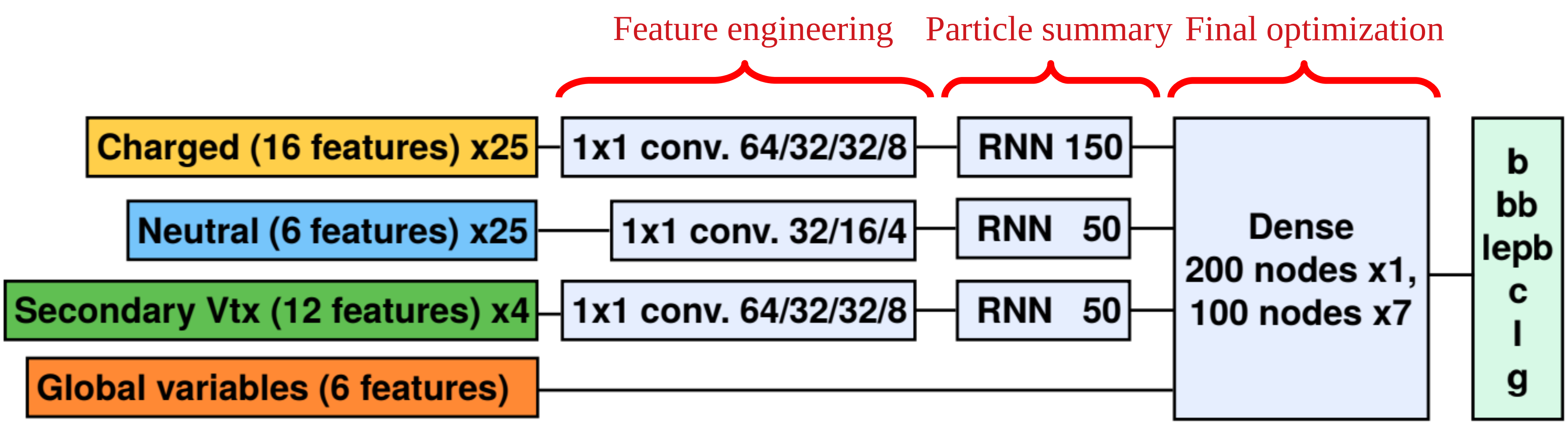
- The network is trained with 2 fragmentation models (Pythia8 and Herwig++) + GEANT4
  - The train dataset is composed of about 224000 images and the test about 56000
- Much better performance than the likelihood based quark-gluon discriminator
- Explainability of the tagger functioning can be also studied by looking at the filters
  - The average jet/quark images are convoluted with the filters and compared





DeepJet tagging in CMS

- CMS has devised a system to perform jet tagging combining CNN, RNN and Dense Layers.
- The network uses 4 levels of features:
  - Charged particles: 16 features per particle x 25 charged particles
  - Neutral particles: 6 features x 25 neutral particles
  - Secondary vertex: 12 features x 4 secondary vertices
  - Global variables: 6 features (number of vertices, jet pt, eta, etc.).
- The CNN creates features per particle while the RNN (LSTM) summarizes sequentially

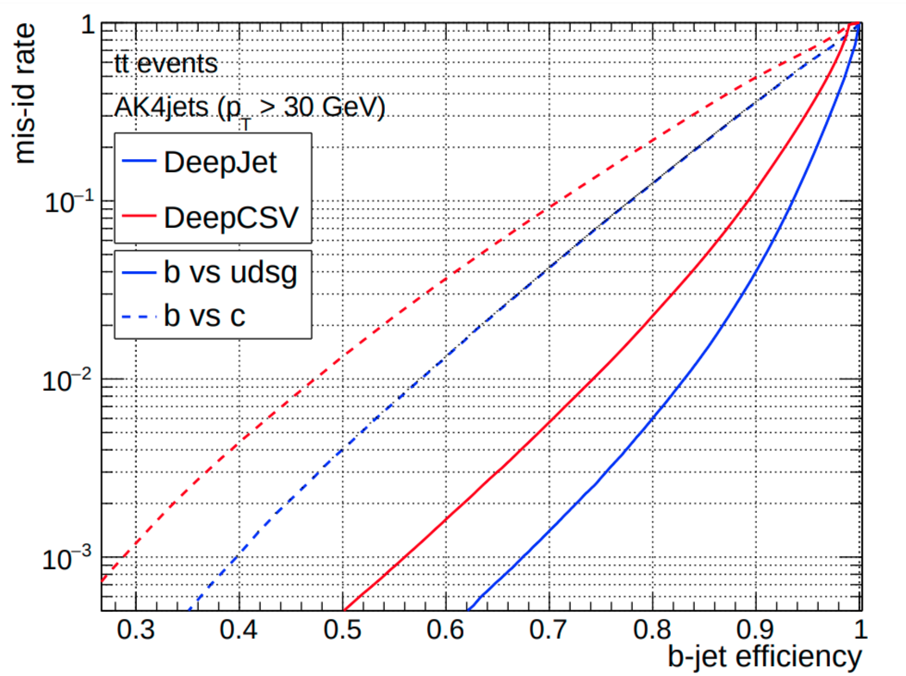


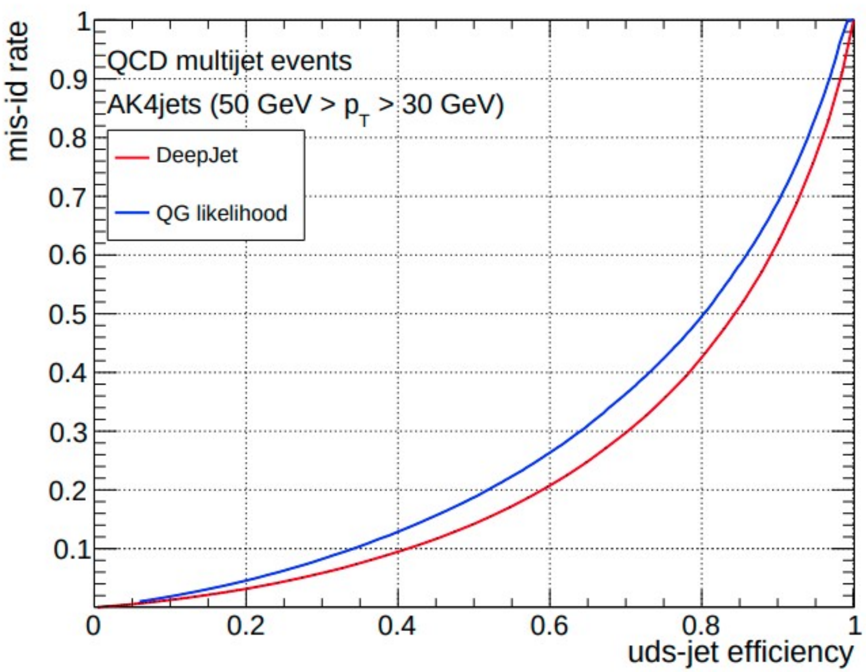
https://arxiv.org/pdf/2008.10519.pdf

DeepJet tagging in CMS

Instituto de Física de Cantabria

- The algorithm is trained with 130 million jets coming from simulated QCD and ttbar
- The performance is compared to the CMS DeepCSV algorithm based on a fully dense ANN
  - DeepJet outperforms by ~12% the b-tagging efficiency for 0.001 misidentification rate
- Also the performance is compared to the likelihood-based quark-gluon discriminator
  - DeepJet outperforms by ~10% the quark-gluon discriminator for 0.3 fake rate

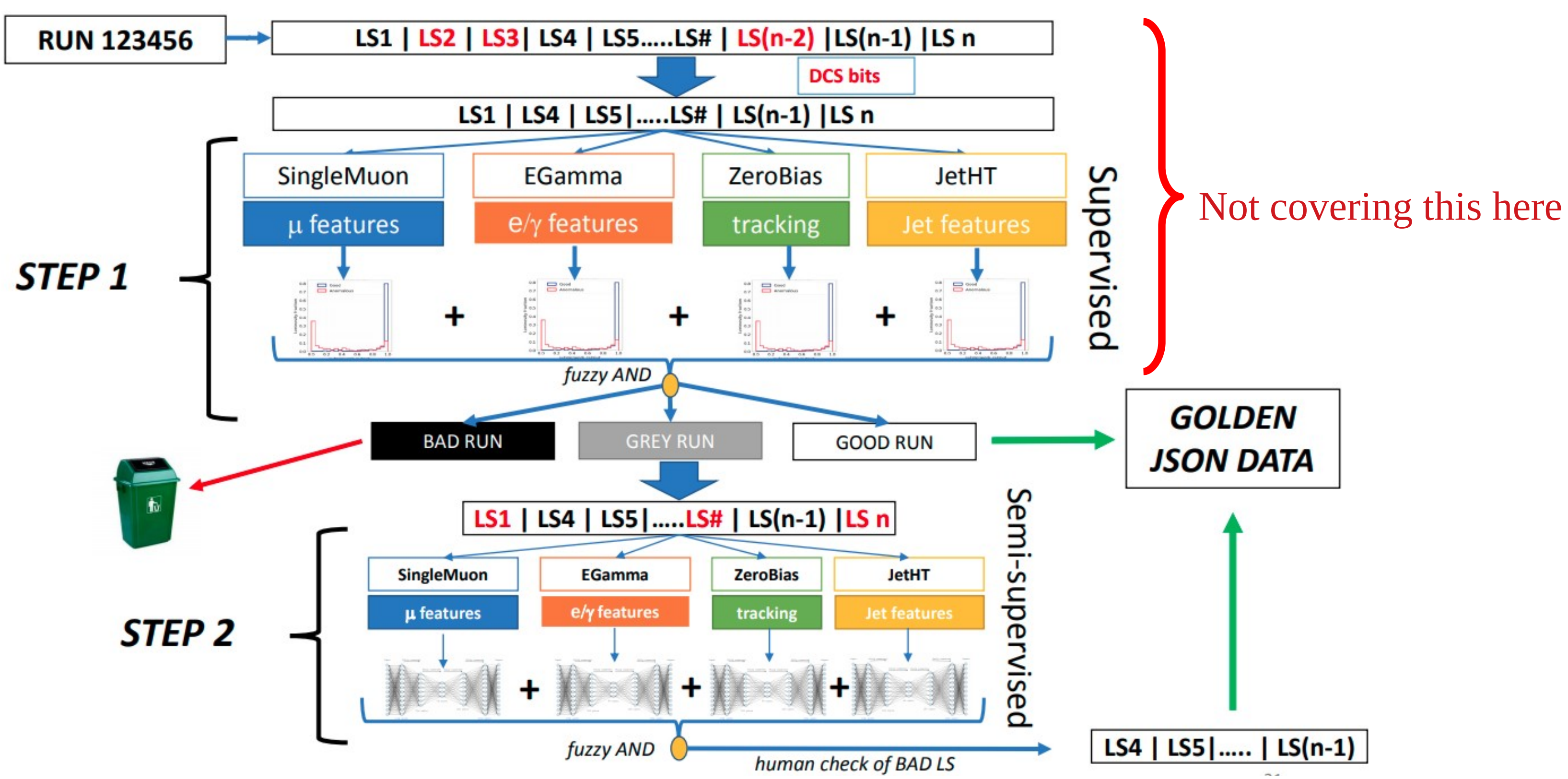




DQM

Anomaly Detection in CMS Data Certification

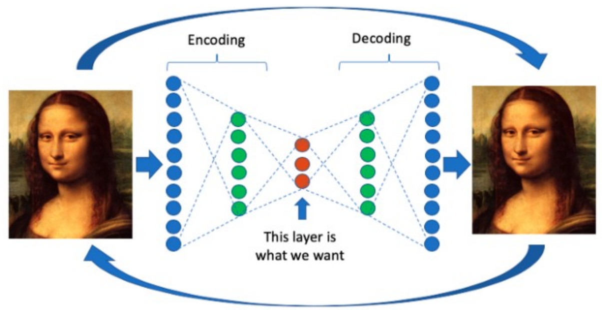
- Data Certification: subsystem experts assign a quality flag to runs and lumisections
  - Tedious and time consuming task (for example rarely DC really occurs per lumisection)
- CMS setting up a 2-step DC procedure combining supervised & unsupervised ML methods

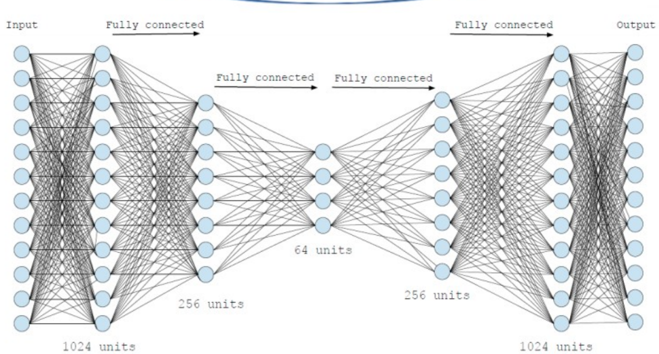


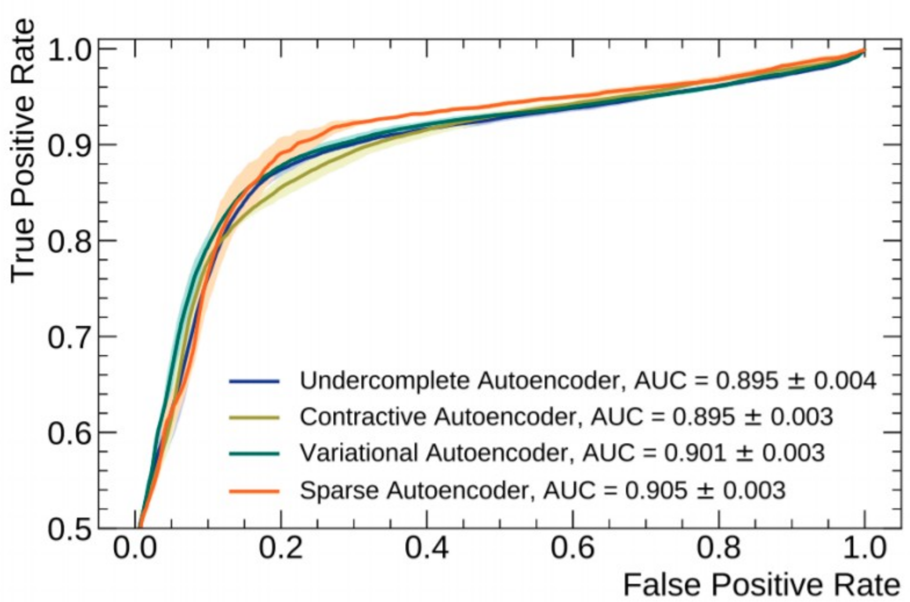
10.1088/1742-6596/1085/4/042015

Anomaly Detection in CMS Data Certification

- The second step uses Variational Auto-Encoders to assign the quality flags
  - No need of BAD data for training (good since fortunately most of the data is GOOD)
  - The source of a given anomaly can be traced back (interpretability of results)
- The VAE learns to compress and uncompress the internal structure of the GOOD data
  - This process does not work for anomalies resulting in an output very different to the input
- The input to the autoencoder are the 5-quantile + mean + RMS of key histograms







Conclusions

- Machine Learning techniques are starting to have a large impact on the experiment operation
- A large plethora of different algorithms are being used at different places of the experiments
- A few examples have been shown on Generation, Trigger, Reco/Identification and DQM
- A set of different algorithms discussed but be aware that many new algorithms are coming
  - Graph Neural Networks (tracking), Autoregressive networks, ...
- Also some steps are being given in the direction of improving explainability of the systems
- Large gain and in some cases impressive results
  - But remember that usually in the talks only the successful examples are shown :-)
  - The large gain usually comes with a large effort in understanding the details
- This is clearly a growing field with a lot of room for ideas and creativity