Advent of Deep Learning

in LHC search and QCD

Partha Konar
Physical Research Laboratory, Ahmedabad, India

Anomalies 2021 [Online] @ IIT, Hyderabad
Nov 10-12

PRELUDE

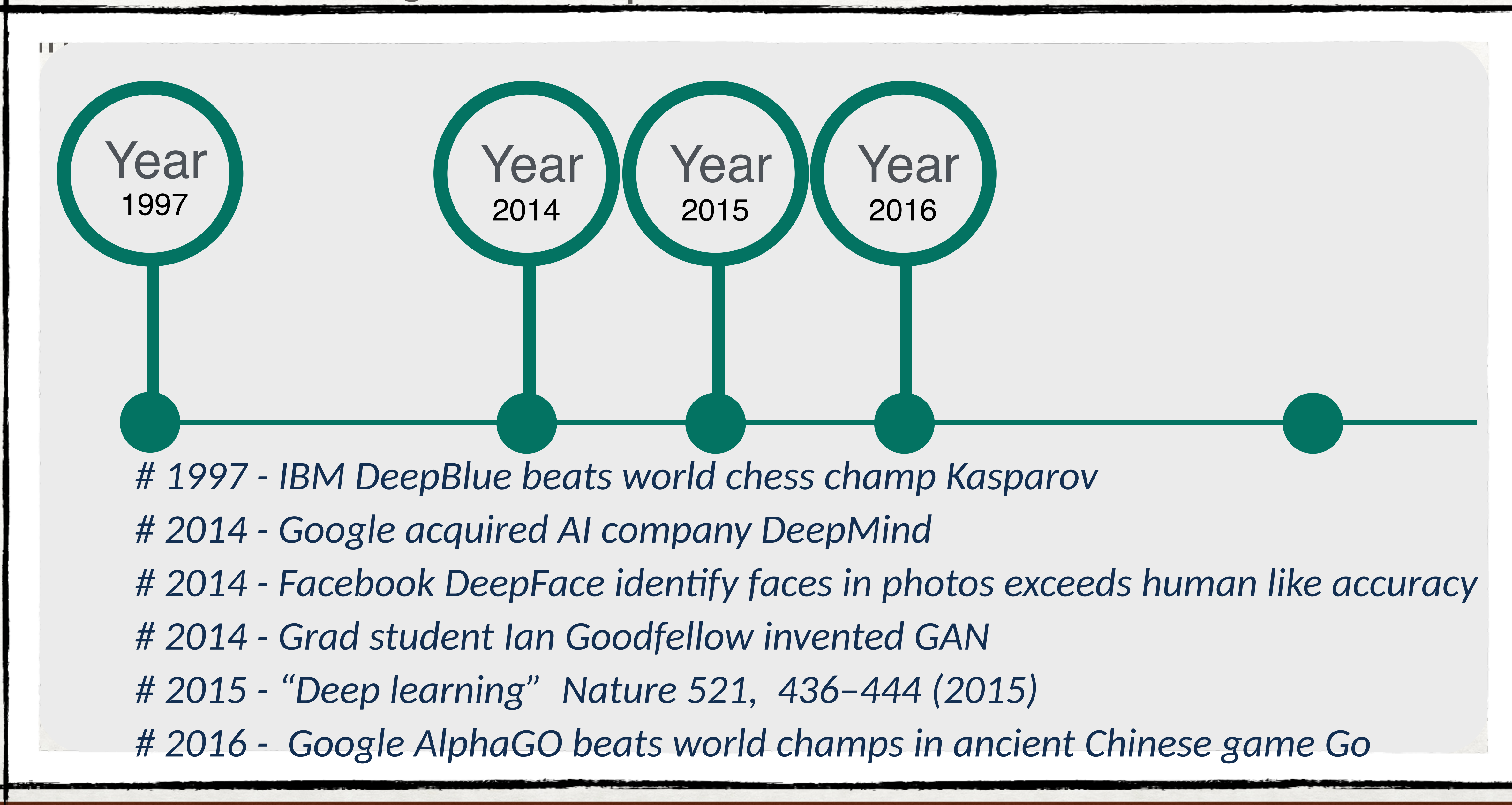
ML@HEP

- Machine learning is not new for HEP community
- Different low to high level experimental measurements with **track finding**, **calorimeter hit reconstruction**, **particle identification**, **reconstruction of energy/ momenta** etc
- Multi Variate Analysis (**MVA**) & Boosted Decision Tree (**BDT**) used extensively on high level variables with primary focus as Classifier
- Today, I focus from the viewpoint of the emergence of **modern deep learning** era that greatly outperformed the previous state of arts during last one decade or so
- Some of the crowning moments that shaped the progress….

PRELUDE

ML@HEP

- Machine learning is not new for HEP community
- Different low to high level experimental measurements with track



MACHINE LEARNING

WHY RELEVANT AT LHC ?

- With around 40 mHz branch crossing LHC taking ~ 40 million snaps/s
- Each snapshot encounter large no of particles
  compounding ~10^8 sensors at different parts of detector
- ML takes role from low level reconstruction, identification, underlying event mitigation to high level identification, extraction, classification and anomaly detection
- Crucial roles in
  (i) Data reduction based on anomaly detection
  (ii) Fast accurate reconstruction, identification with multi-sensor data
  (iii) Significant improvements in classification, regression, goodness fit

MACHINE LEARNING

NURAL NETWORK

- Universal function approximation - NN with a single hidden layer can approximate any continuous fn to any desired precision

- **Search for a function** $f : X \to Y$

  which **optimize** some **loss function** $\mathscr{L}[Y - f_w(x)]$

  X : **Observable space** Y: **Target space [low dim]**

- During training, trainable ***weight parameters*** are ***learned*** by the back-propagation whose aim i ss **function.**

Introduction of non-linearities along with linear terms

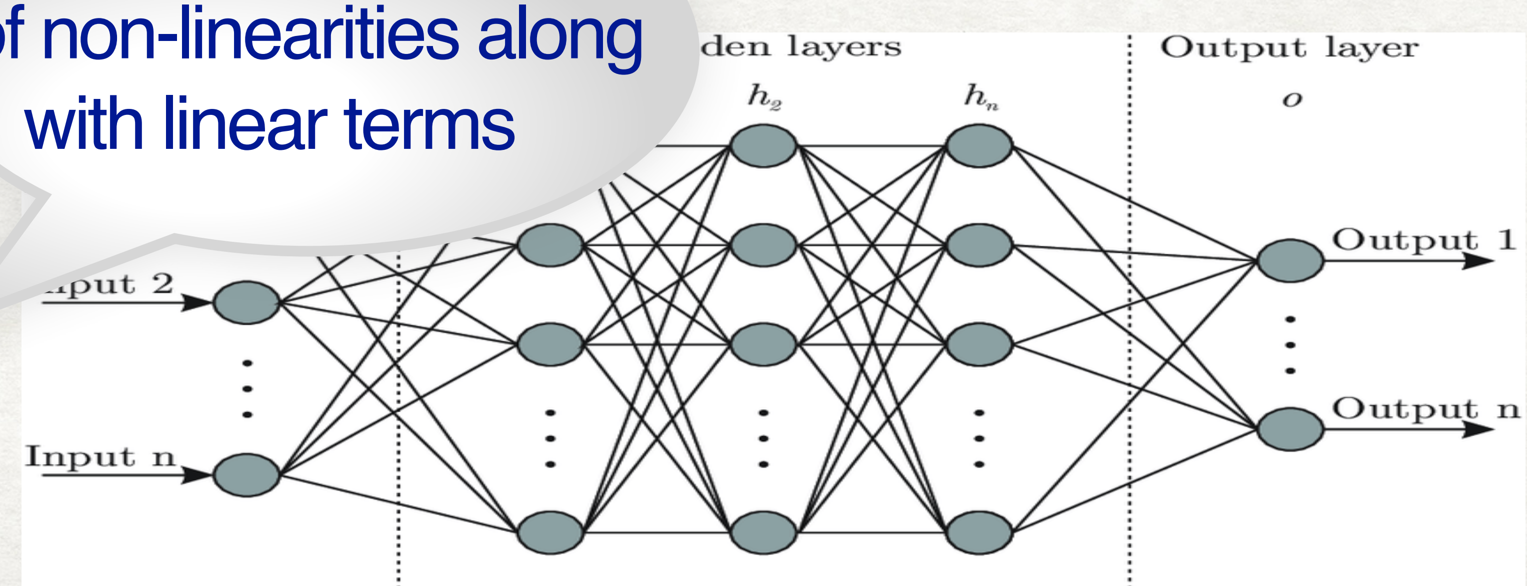
$$h_{i+1} = \sigma_i(\sum_i w_i h_i + b_i)$$

Response fn : $\sigma(W^T x)$

ReLu, sigmoid, softmax

Cost fn : $\mathscr{L}[Y - f_w(x)]$

$\nabla_w \mathscr{L}[Y - f_w(x)] \to 0$



MACHINE LEARNING

GOING DEEPER

- Deep learning models with **multiple hidden layers** solves the need for infinitely large nodes in shallow NN

- Learning **scalable** with data - larger data for better performance

- Deep learning models are now capable of **extracting feature directly from low level data**
  — End of high level variables from domain experts? Not sure!

- Need for high-end infrastructure GPU, Deep Learning algorithms, large amount of data handling etc

- Problem with **Interpretability** [unlike BDT], lack of physics understanding for feature intersection — work in progress

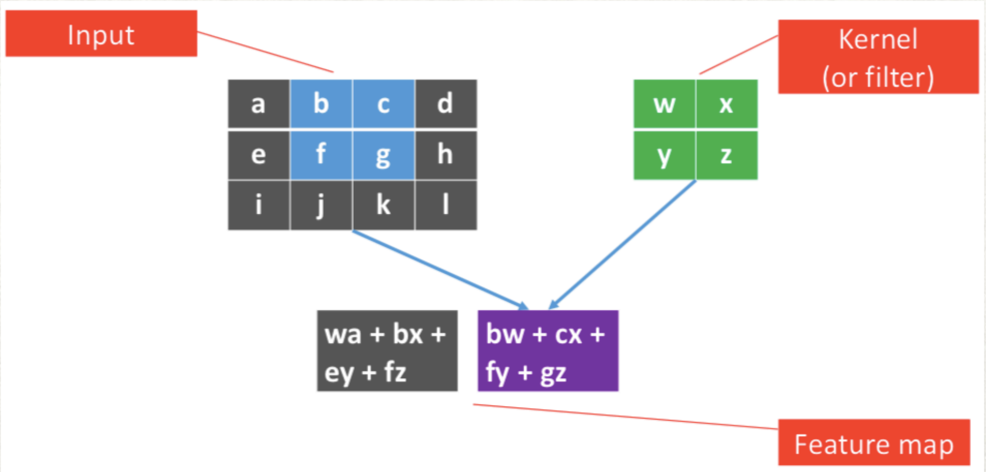
MACHINE LEARNING

WORKING WITH CONVOLUTIONAL NEURAL NETWORK

- Most significant innovation in DNN - **Image processing**
- **Convolution architecture** rely on **local** and **global** features with **translation invariance**
(we can also make it learn scale & rotation invariance)
- Image pixels are convoluted with no. of kernels "k_j"

$$h_{i+1} = \sigma(wx + b) \quad \rightarrow \quad h_{i,j} = \sigma(k_j \,.\, h_i + b_j)$$

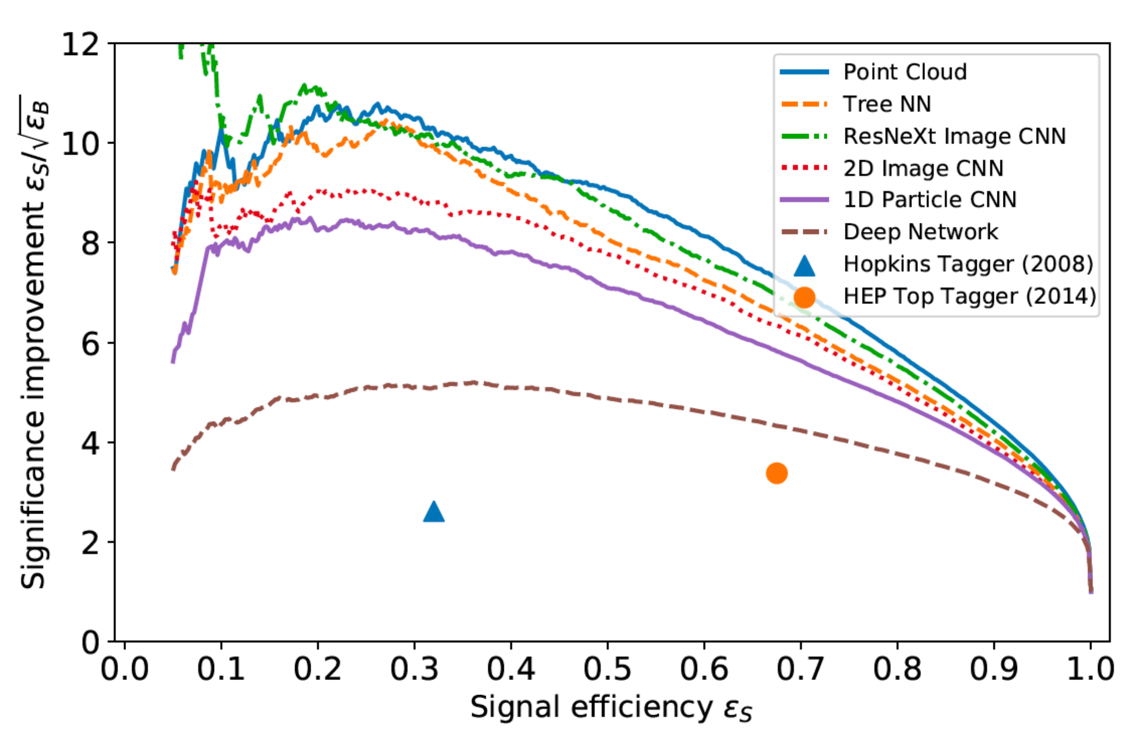
- Same kernel with sharing weight pass through full image, reducing tunable parameter drastically
- Algorithm first learn **edges** and **shapes**
-> more **complex local features**
-> leads to **global features**

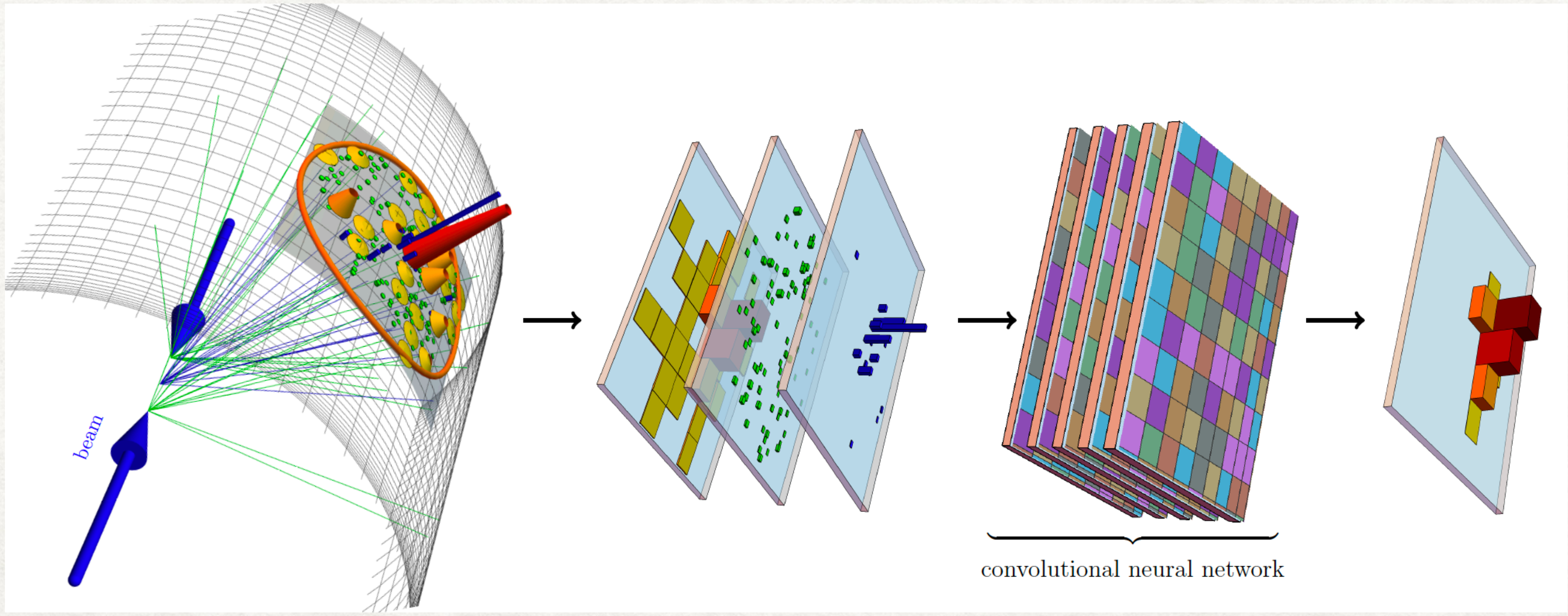


CONVOLUTIONAL NEURAL NETWORK

WORKING PRINCIPLE AT LHC DATA

- Detectors calorimeter tower => pixels of an image
- Powerful image classification network proved to be extremely successful in jet-substructure studies



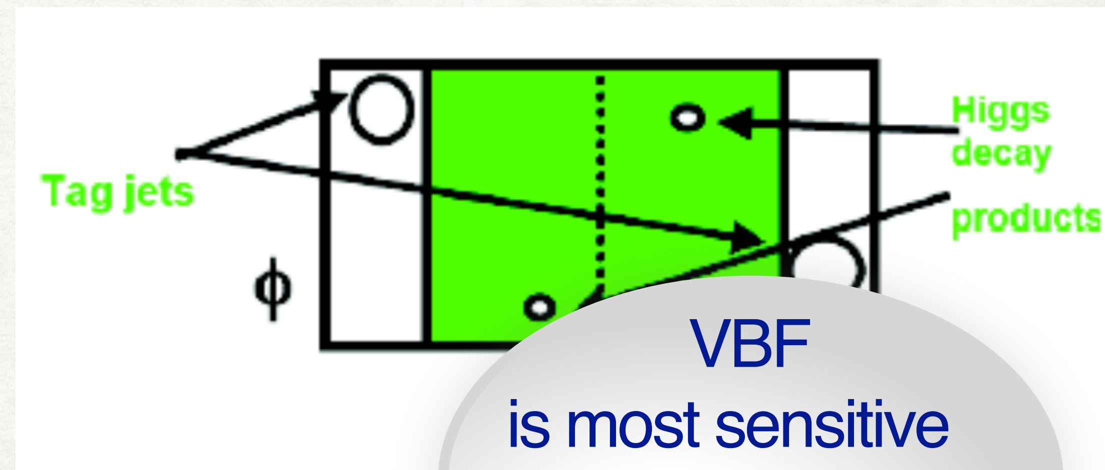


 # Review :1806.11484, 2103.12226

INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

CONVOLUTIONAL NEURAL NETWORK

- ✓ Vector Boson Fusion (VBF) was a novel proposal for Higgs search
- ✓ Interesting topology for a VBF
  Two forward jets + large inv. Mass
  No central jet activity between them
  Decay products at the central region



VBF is most sensitive channel for invisible Higgs search

- Qn. Can CNN learn all the feature for such event se
- Problem is even more difficult if Higgs is decaying invisibly — No additional features from decay product!
- Let us try that!

Collider bounds on invisible branching ratio of Higgs much higher than SM prediction!!

INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

3 SET OF ANALYSIS

A. **CMS analysis with 36 fb^-1 data [Based on expert level VBF feature]**
   Simulated Signal and BG => Reproducing CMS “BR upper limit” result

B. **Analysis with sets of three different high level data [ANN]**
   1. Kinematic data : Event-kinematics from reconstructed objects

   $$\mathcal{K} \equiv (\, |\Delta\eta_{jj}|, \; |\Delta\phi_{jj}| \; , \; m_{jj} \; , \; MET \; , \; \phi_{MET} \; , \; \Delta\phi^{j_1}_{MET} \; , \; \Delta\phi^{j_2}_{MET} \; , \; \Delta\phi^{j_1+j_2}_{MET} \;)$$

   2. Radiative: Contains information about the QCD radiation pattern

   $$\mathcal{R} \equiv (H_T^{\eta_C} | \eta_C \in \mathcal{E}) \quad , \quad H_T^{\eta_C} = \sum_{\eta < |\eta_C|} E_T$$

   3. Combination of above two

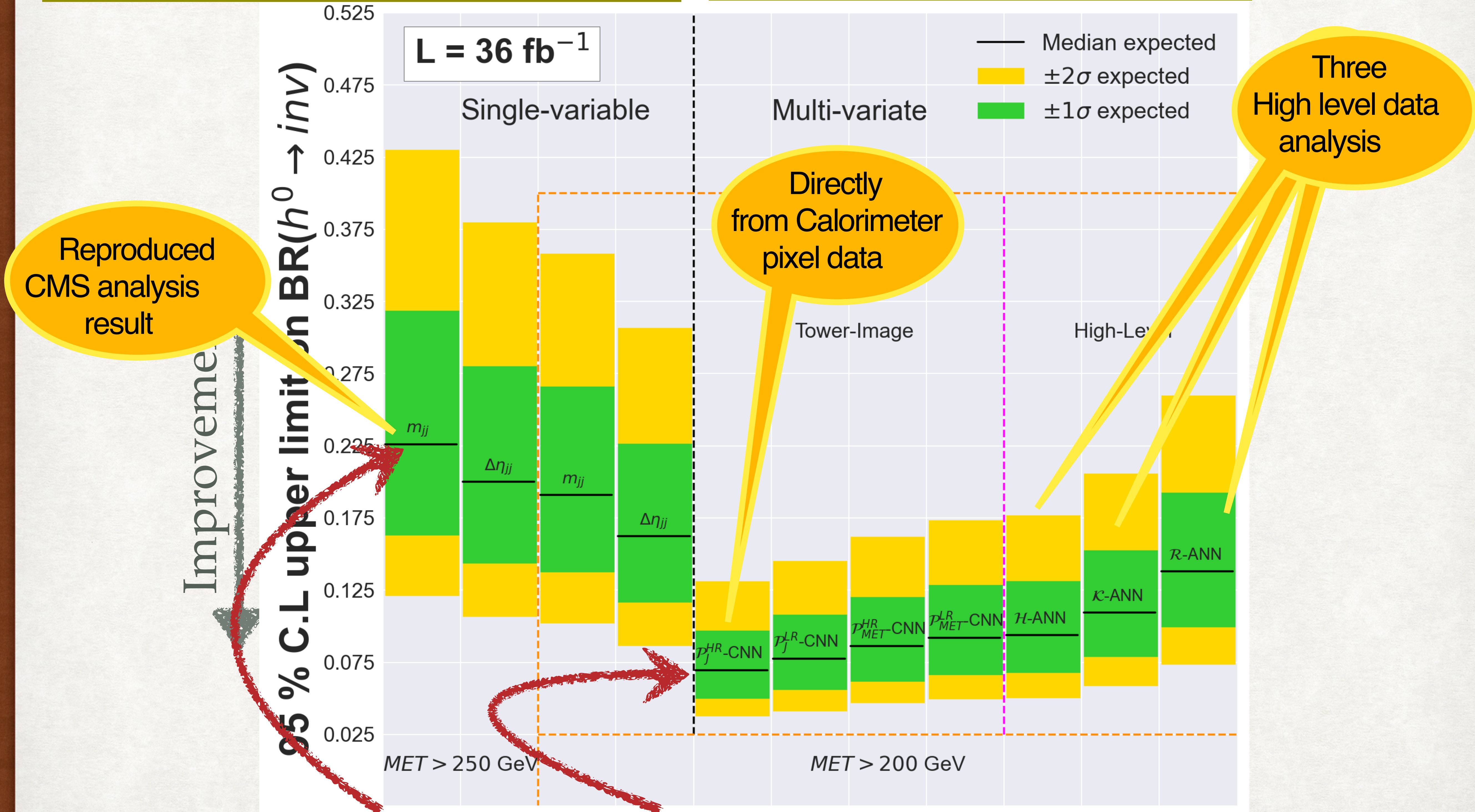
   $$\mathcal{H}$$

C. **Analysis with low level calorimeter input data [CNN]**
   - Hi & Low resolution Calorimetry

 # Eur. Phys. J. C 80, 1055 (2020) - Vishal Ng, Akansha Bhardwaj, PK, Aruna Nayak

INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

-Based on HL variables constructed by experts- *—Based on LL & HL input data——*



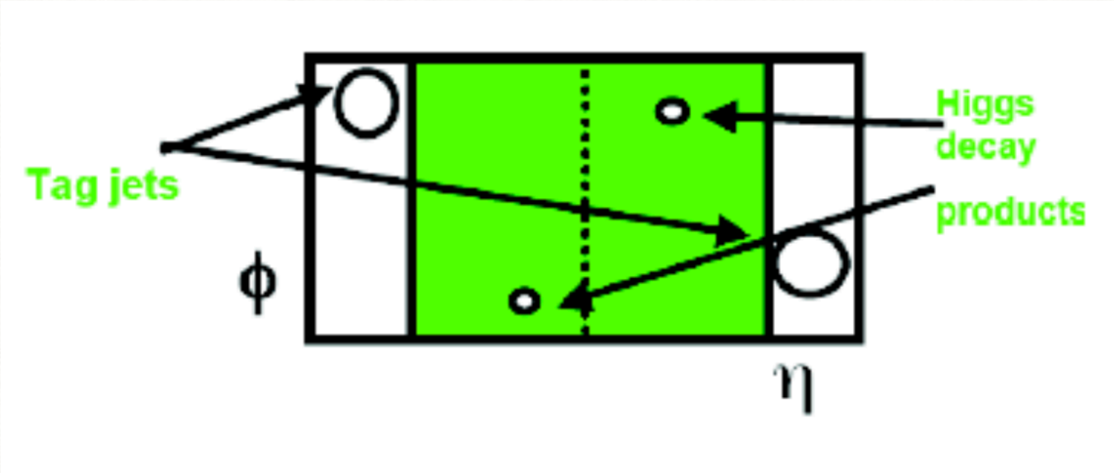
Factor of three improvement using the same data!

Hours of CNN training just extracted the relevant underlying feature better than our decades of research!

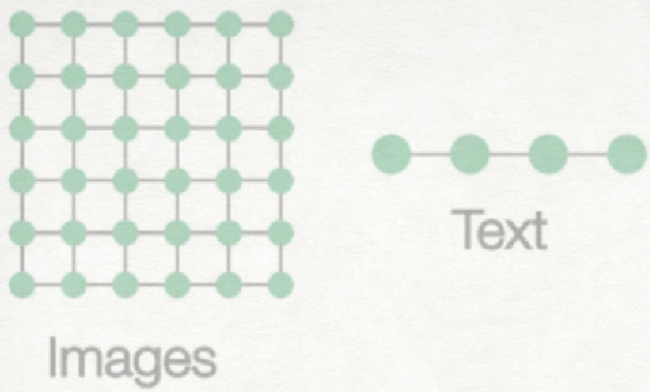
INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

ROLE OF PARTON SHOWER

- ★ In this simple setup with just two jets : NN minutely learned the **kinematic relation** & **radiation pattern** from the data

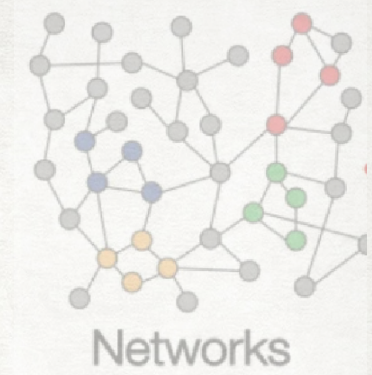


- ★ Extra QCD radiation between two tag jets extremely significant!!
- ★ Central-jet Veto:
  Efficiently rejects large QCD backgrounds by vetoing events with additional central jet
- ★ Qn. How *faithful the distribution function which NN learn?*
- ★ True potential unfolds if theoretical predictions are accurate enough.



BEYOND CNN

GRAPH NEURAL NETWORK



- ➡ Detectors calorimeter hits are typically very sparse and unstructured
- ➡ Varying number of reconstructed constituents
- ➡ Large number of tunable parameters
- ✓ Euclidean image (CNN) to general non-Euclidean domain (GNN) : Geometric deep learning
- Graph: Event as point cloud with each entry containing a vector composed of observables
- Graph == Nodes (data point) + Edges (connections are as important as the data itself)
- Through “Message passing operation” nodes features and edge features are exchanged and provide a sophisticated feature extraction
- GNN is very powerful recent concept - mostly unexplored!!
- Infra-red and collinear safe GNN mechanism is constructed for QCD jet study

 # 2109.14636 - PK, Vishal Ng, Michael Spannowsky — See talk by Vishal

EPILOGUE

WAY AHEAD

- ➡ Deep Learning success story already in NOVA & Ice Cube
- ➡ LHC: time of booming effort both from exp & tho community
- ➡ Significant progress underway in
  - Jet (substructure) study
  - Event simulation
  - Reconstruction, identification with multi-sensor data
  - Improved trigger mechanism
  - Anomaly detection
  - Analysis - classification, regression, goodness fit

- ✓ CNN based model shown excellent efficiency for invisible Higgs search from VBF, using low-level calorimeter image to study the event topology
- ✓ Accurate simulation of QCD radiation is imperative

INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

PRESELECTING CUTS

- **VBF Jet tag**: At least two jets with leading(sub-leading) jet $p_T > 80$ (40) GeV with $|\eta| < 4.7$. At least one of the jets to have $|\eta_{j_i}| < 3$.

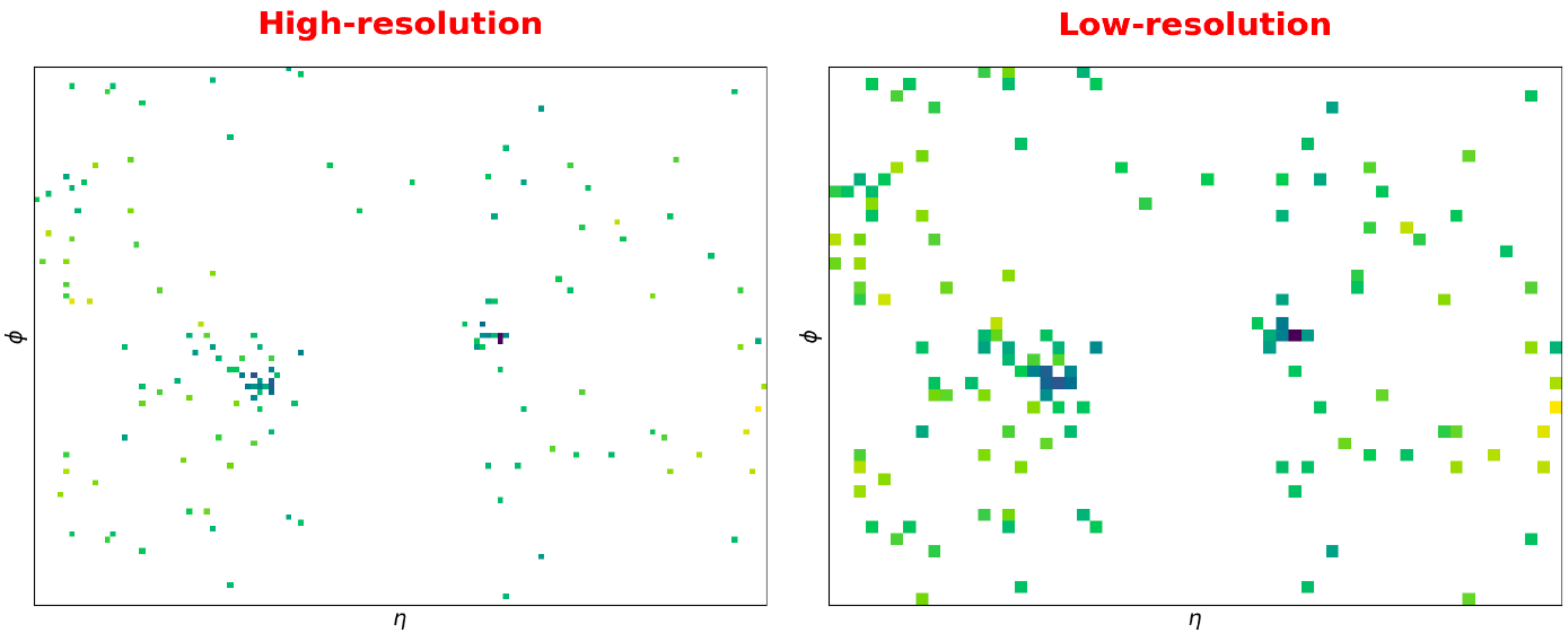
$$\eta_{j_1}\,\eta_{j_2} < 0 \quad , \quad |\Delta\phi_{jj}| < 1.5 \quad , \quad |\Delta\eta_{jj}| > 1 \quad , \quad m_{jj} > 200 \text{ GeV}$$

- **Lepton-veto:** No electron(muon) with $p_T > 10$ GeV in the central region, $|\eta| < 2.5(2.4)$.
- **Photon-veto:** No photon with $p_T > 15$ GeV in the central region, $|\eta| < 2.5$
- τ **and b-veto**: no tau-tagged jets in $|\eta| < 2.3$ with $p_T > 18$ GeV, and no b-tagged jets in $|\eta| < 2.5$ with $p_T > 20$ GeV.
- **Missing** E_T(MET): $MET > 200$ GeV (250 GeV for CMS shape-analysis)
- **MET jet alignment**: $\min(\Delta\phi(\vec{p}_T^{\,\mathrm{MET}}, \vec{p}_T^{\,\mathrm{j}})) > 0.5$ for upto four leading jets with $p_T > 30$ GeV with $|\eta| < 4.7$.

 # Eur. Phys. J. C 80, 1055 (2020) - Vishal Ng, Akansha Bhardwaj, PK, Aruna Nayak

INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

TOWER IMAGE



- Pixel wise calorimeter energy deposits (E_T) converted into pictorial description like 'tower-images' as input to Convolutional Neural Networks

Eur. Phys. J. C 80, 1055 (2020) - Vishal Ng, Akansha Bhardwaj, PK, Aruna Nayak

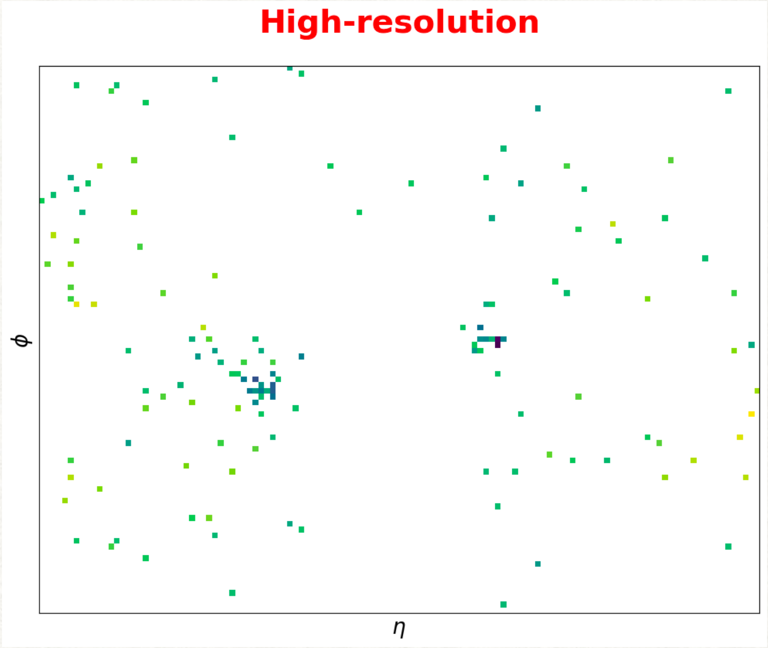
INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

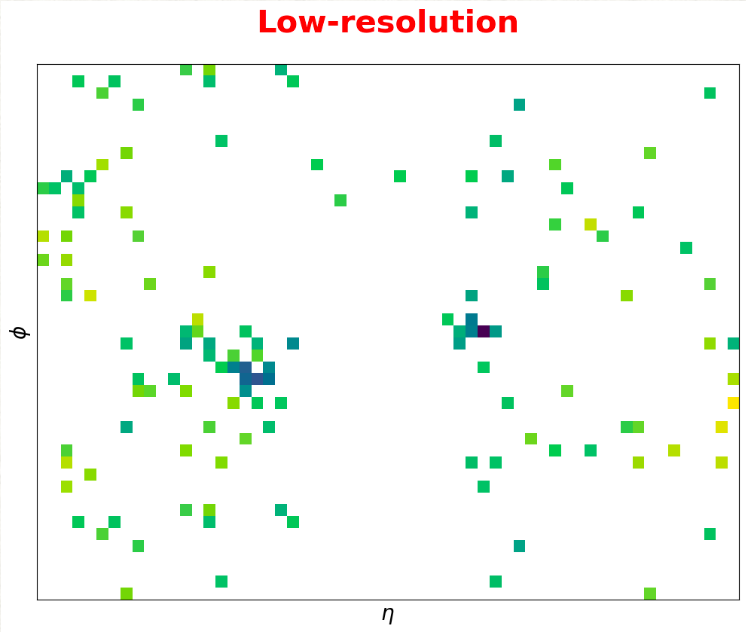
TOWER IMAGE

- Pixel wise calorimeter energy deposits (E_T) converted into pictorial description like `tower-images' as input to CNN

- Bin-size:
  High-resolution (HR) 0.08×0.08
  Low-resolution (LR): 0.17×0.17
  $\eta \in (-5,5)$ & $\phi \in (-\pi, \pi)$

- Periodicity in ϕ implemented with padded at each boundary with rows from the opposite boundary

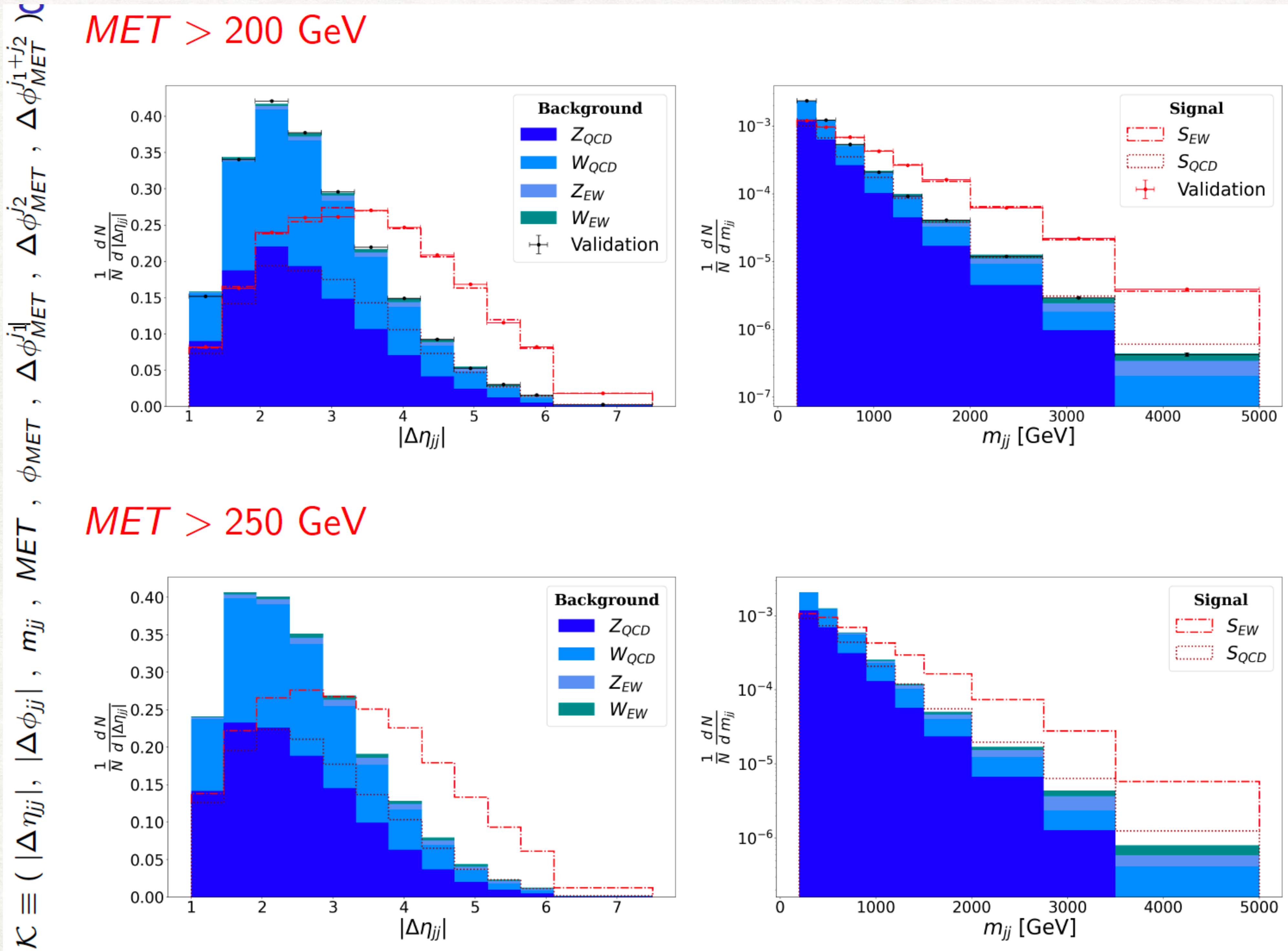




Eur. Phys. J. C 80, 1055 (2020) - Vishal Ng, Akansha Bhardwaj, PK, Aruna Nayak

INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

HI LEVEL - "KINEMATIC" VARIABLE

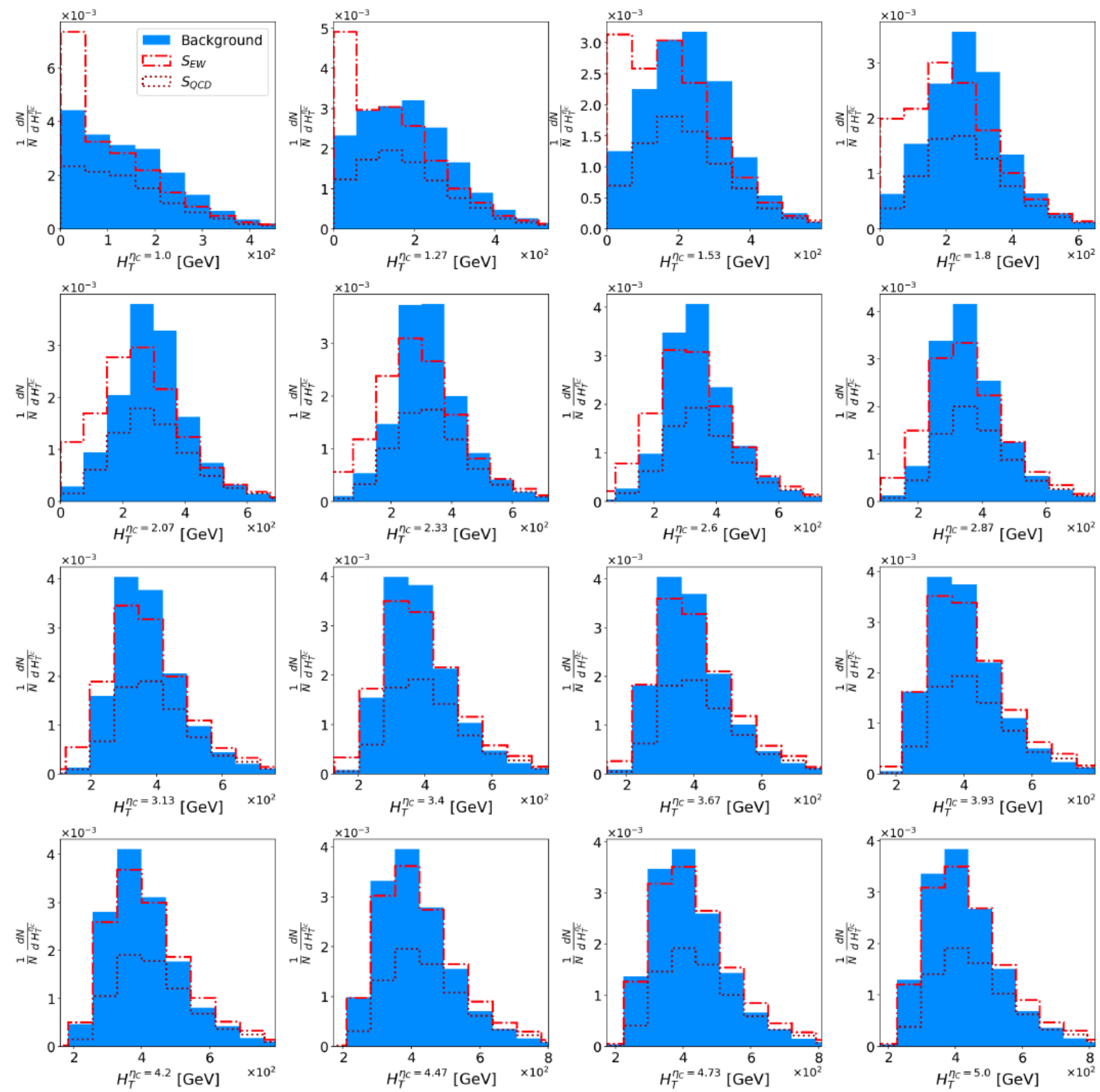


 # Eur. Phys. J. C 80, 1055 (2020) - Vishal Ng, Akansha Bhardwaj, PK, Aruna Nayak

INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

HI LEVEL - “RADIATION” VARIABLE

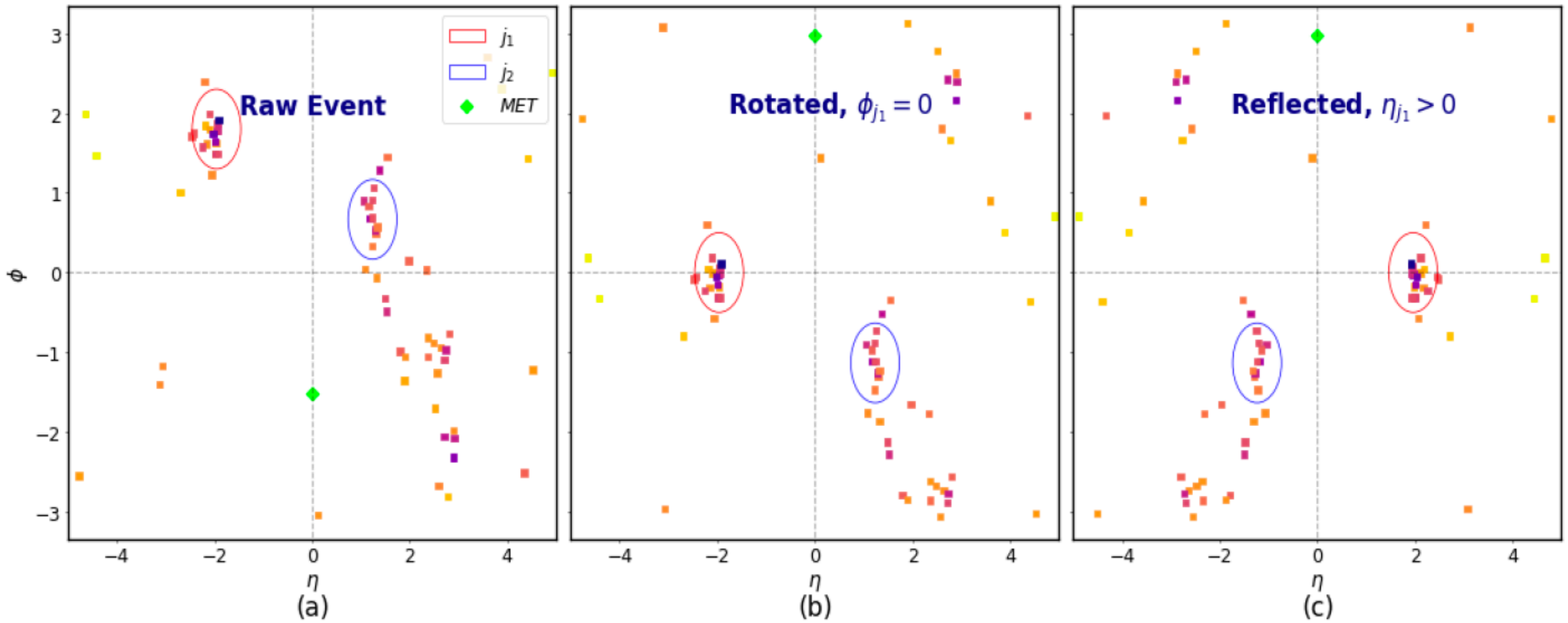
$$\mathcal{R} \equiv (H_T^{\eta_C} | \eta_C \in \mathcal{E}) \quad , \quad H_T^{\eta_C} = \sum_{\eta<|\eta_C|} E_T$$



Eur. Phys. J. C 80, 1055 (2020) - Vishal Ng, Akansha Bhardwaj, PK, Aruna Nayak

INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

LOW LEVEL - EVENT PROCESSING

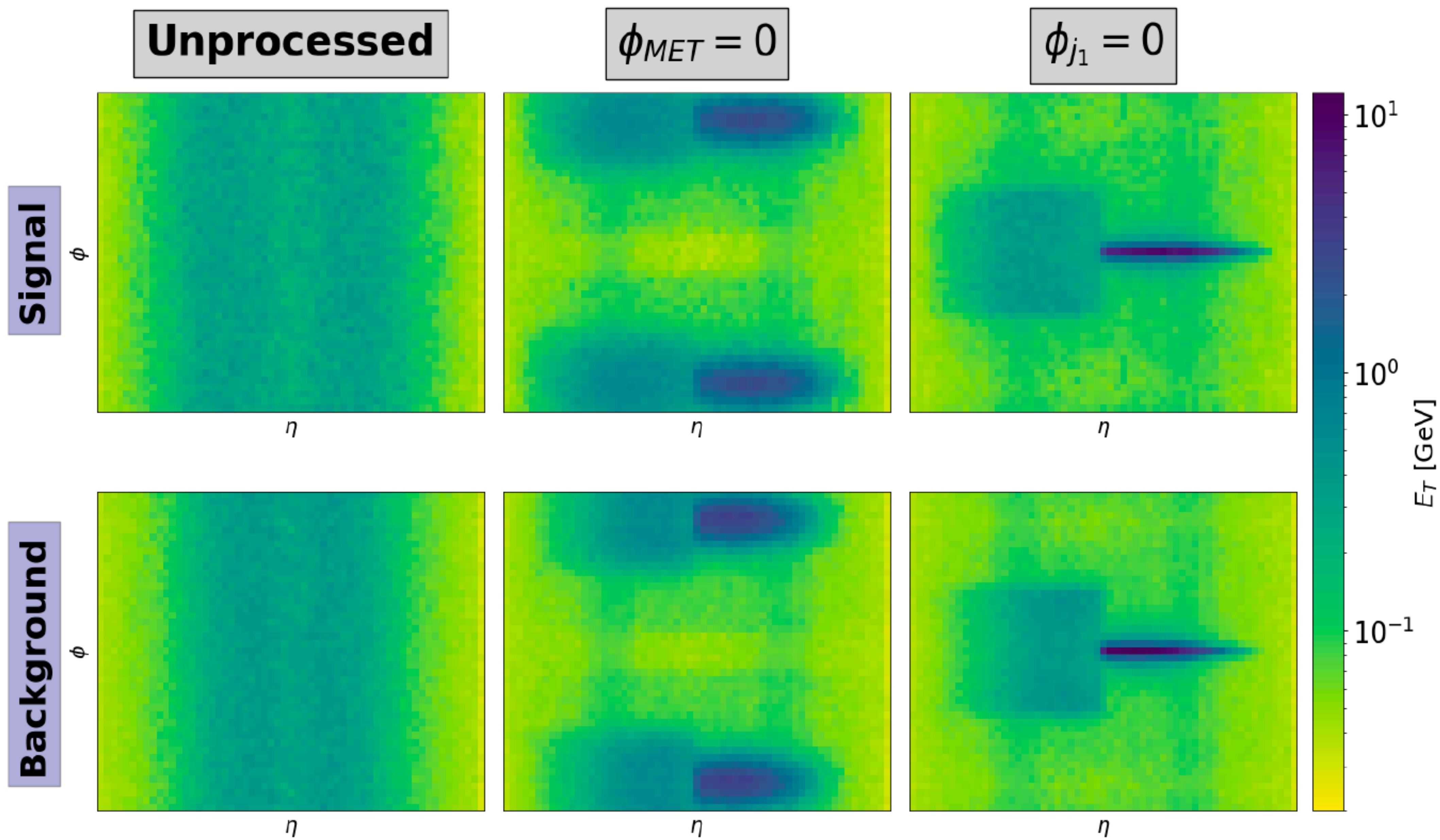


- Rotate along z-axis such that $\phi_0 = 0$.
  Two instances of $\phi_0 \in \{\phi_{MET}, \phi_{j_1}\}$.
- Reflect along the xy-plane, such that the leading jet's η is always positive.
- After binning (E_T) and padding in LR and HR : $\mathcal{P}^{LR}_{MET}$, $\mathcal{P}^{HR}_{MET}$, $\mathcal{P}^{LR}_{J}$ and $\mathcal{P}^{HR}_{J}$

Eur. Phys. J. C 80, 1055 (2020) - Vishal Ng, Akansha Bhardwaj, PK, Aruna Nayak

INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

LOW LEVEL - EVENT PROCESSING

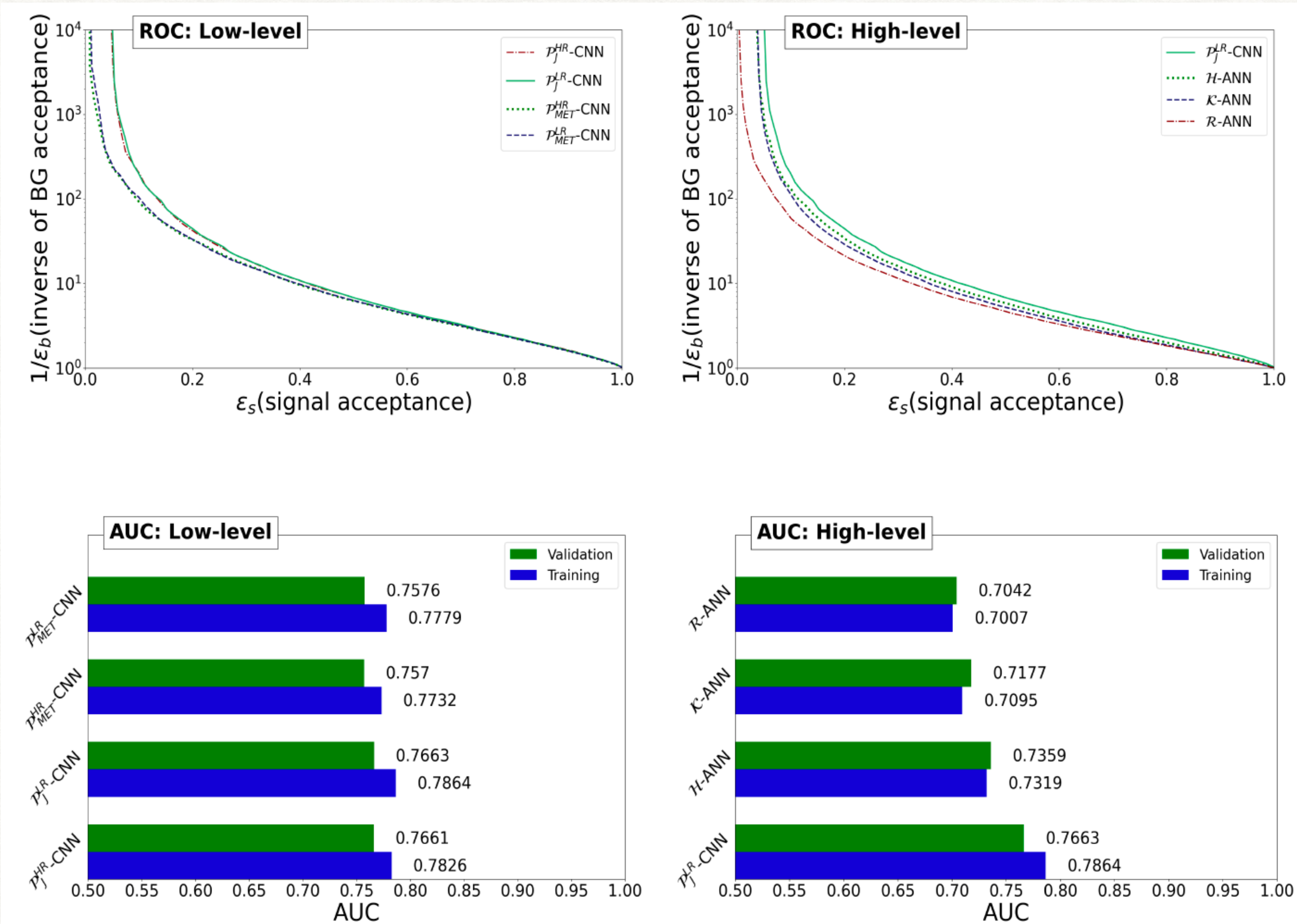


Averaged Images

 # Eur. Phys. J. C 80, 1055 (2020) - Vishal Ng, Akansha Bhardwaj, PK, Aruna Nayak

INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

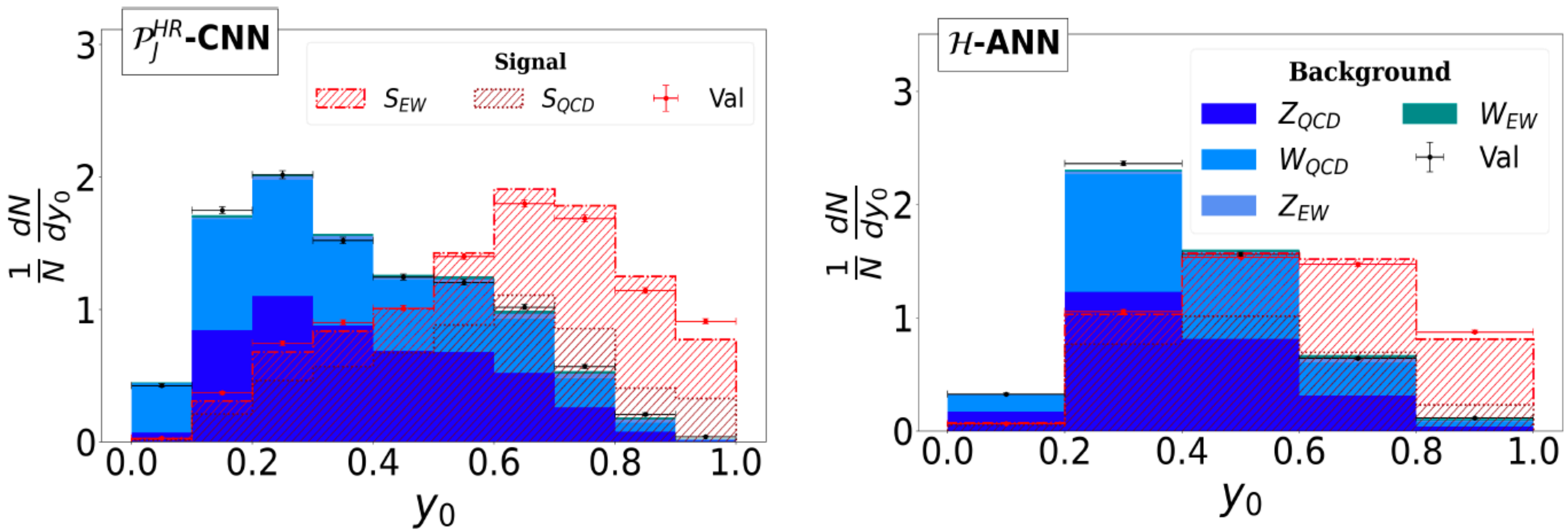
NETWORK PERFORMANCE



Eur. Phys. J. C 80, 1055 (2020) - Vishal Ng, Akansha Bhardwaj, PK, Aruna Nayak

INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

NETWORK PERFORMANCE



- Harder to distinguish S_{QCD} from the QCD dominated ($\sim$ 95%) background class (significant S_{QCD} contamination in traditional analysis too)
- For the CNN, W_{QCD} dominates over Z_{QCD} in the first bin?? $\Rightarrow$ Presence of calorimeter deposits of lepton in regions $|\eta| > 2.5$ or in the central regions when it is misidentified (including $\tau^{\pm}$).

Eur. Phys. J. C 80, 1055 (2020) - Vishal Ng, Akansha Bhardwaj, PK, Aruna Nayak

INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

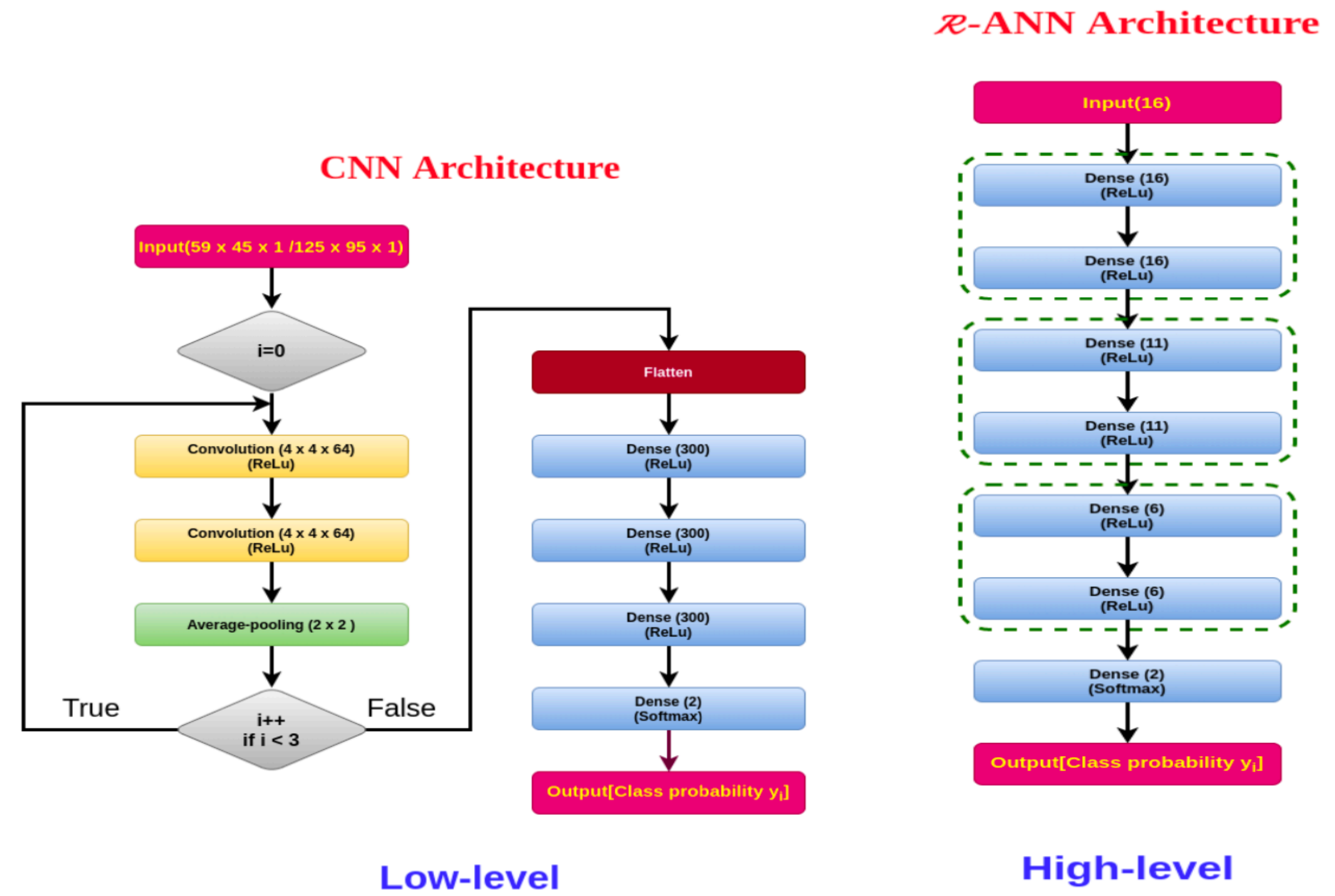
UPPER BOUND ON BRANCHING RATIO

Sl.No	Name	Description	Expected median upper-limit on B.R($h^0 \to$ inv)		
			L = 36 fb^{-1}	L = 140 fb^{-1}	L = 300 fb^{-1}
1.	$m_{jj}(MET > 250$ GeV)	reproduced CMS shape analysis	$0.226^{+0.093}_{-0.063}$	$0.165^{+0.082}_{-0.056}$	$0.130^{+0.089}_{-0.027}$
2.	$\|\Delta\eta_{jj}\|(MET > 250$ GeV)	$\|\Delta\eta_{jj}\|$ analysis with CMS shape-cuts	$0.200^{+0.080}_{-0.056}$	$0.128^{+0.050}_{-0.036}$	$0.106^{+0.041}_{-0.025}$
3.	$m_{jj}(MET > 200$ GeV)	m_{jj} shape analysis with weaker cut	$0.191^{+0.075}_{-0.053}$	$0.116^{+0.071}_{-0.036}$	$0.101^{+0.037}_{-0.045}$
4.	$\|\Delta\eta_{jj}\|(MET > 200$ GeV)	$\|\Delta\eta_{jj}\|$ analysis with weaker cut	$0.162^{+0.065}_{-0.045}$	$0.105^{+0.042}_{-0.029}$	$0.087^{+0.034}_{-0.025}$
5.	$\mathcal{P}_J^{LR}$-CNN	Low-Resolution, $\phi_0 = \phi_{j_1}$	$0.078^{+0.030}_{-0.022}$	$0.051^{+0.020}_{-0.014}$	$0.045^{+0.017}_{-0.013}$
6.	$\mathcal{P}_J^{HR}$-CNN	High-Resolution, $\phi_0 = \phi_{j_1}$	$0.070^{+0.027}_{-0.020}$	$0.043^{+0.017}_{-0.012}$	$0.035^{+0.013}_{-0.010}$
7.	$\mathcal{P}_{MET}^{LR}$-CNN	Low-Resolution, $\phi_0 = \phi_{MET}$	$0.092^{+0.037}_{-0.025}$	$0.062^{+0.024}_{-0.017}$	$0.053^{+0.023}_{-0.014}$
8.	$\mathcal{P}_{MET}^{HR}$-CNN	High-Resolution, $\phi_0 = \phi_{MET}$	$0.086^{+0.035}_{-0.024}$	$0.058^{+0.023}_{-0.016}$	$0.051^{+0.020}_{-0.014}$
9.	$\mathcal{K}$-ANN	8 kinematic-variables	$0.101^{+0.052}_{-0.022}$	$0.075^{+0.029}_{-0.021}$	$0.063^{+0.027}_{-0.017}$
10.	$\mathcal{R}$-ANN	16 radiative $H_T^{\eta c}$ variables	$0.138^{+0.055}_{-0.039}$	$0.094^{+0.036}_{-0.027}$	$0.079^{+0.032}_{-0.022}$
11.	$\mathcal{H}$-ANN	Combination of $\mathcal{K}$ and $\mathcal{R}$ variables	$0.094^{+0.038}_{-0.026}$	$0.065^{+0.026}_{-0.018}$	$0.057^{+0.022}_{-0.015}$

 # Eur. Phys. J. C 80, 1055 (2020) - Vishal Ng, Akansha Bhardwaj, PK, Aruna Nayak

INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

BRIEF DESCRIPTION OF NETWORKS



- After training for 20-1000 epochs, best performing network on the validation data choosen (for each of the 7 networks).

 # Eur. Phys. J. C 80, 1055 (2020) - Vishal Ng, Akansha Bhardwaj, PK, Aruna Nayak