Distinguishing Vector-Boson Fusion Signatures with Convolutional Neural Networks: Invisible Higgs Decay

Eur. Phys. J. C 80, 1055 (2020) (with A. Bhardwaj, P. Konar, A. K. Nayak)

Physical Research Laboratory, Ahmedabad(India)

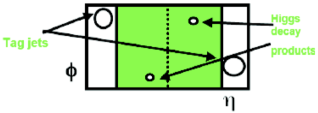
IML Machine Learning Working Group
February 16, 2021

Vishal S. Ngairangbam

Outline

Vector Boson Fusion: A unique signature

- t-channel production of color-singlet particles via fusion of two vector-bosons
  - No central jet activity
  - Large rapidity gap between two jets
  - Large invariant mass of the two jet system
  - Decay products at the central region



- Higher order QCD always below 10% – very stable with scale uncertainty

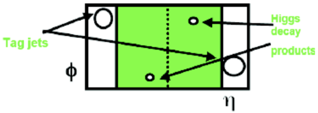
- Very important for BSM searches of color singlet particles.

- Dominant production channel for heavy Higgs at hadron colliders

- **Central-jet veto**: viable to search for lighter Higgs masses

Vector Boson Fusion: A unique signature

- t-channel production of color-singlet particles via fusion of two vector-bosons
  - No central jet activity
  - Large rapidity gap between two jets
  - Large invariant mass of the two jet system
  - Decay products at the central region

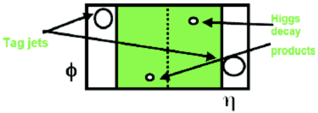


VBF production of $m_h = 125$ GeV Higgs

- Second highest cross-section after gluon-fusion
- Very clean channel for non-hadronic decay of the Higgs
- Most sensitive channel for searching invisible decay of Higgs (Important in many BSM scenario)

Vector Boson Fusion: A unique signature

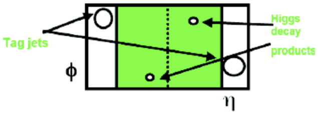
- t-channel production of color-singlet particles via fusion of two vector-bosons
  - No central jet activity
  - Large rapidity gap between two jets
  - Large invariant mass of the two jet system
  - Decay products at the central region



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

Vector Boson Fusion: A unique signature

- t-channel production of color-singlet particles via fusion of two vector-bosons
  - No central jet activity
  - Large rapidity gap between two jets
  - Large invariant mass of the two jet system
  - Decay products at the central region



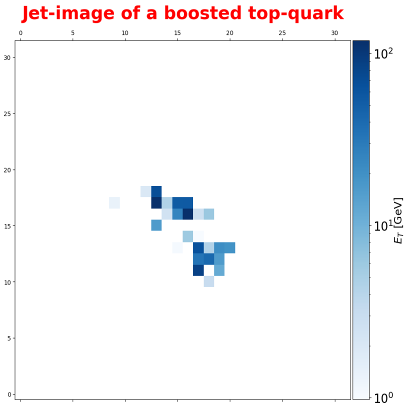
Collider bounds on invisible branching ratio of Higgs much higher than in SM!!

New techniques to reduce the upper limit: Deep learning??

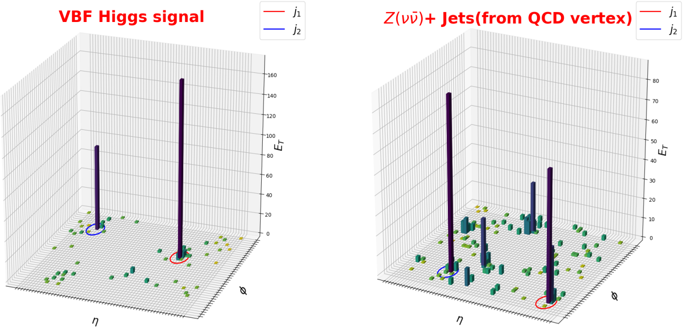
CNNs and jet-images: why do they work?

- Efficiently distinguishes large radius QCD jets from decays of boosted heavy particles ($t, W^{\pm}/Z^0/h^0$)
- Works with data which have an underlying Euclidean-geometry
- Jet-substructure variables are mostly functions of the Euclidean distance $\Delta R_{ij} = \sqrt{\Delta\eta_{ij}^2 + \Delta\phi_{ij}^2}$ in the (η, ϕ) plane, for instance:

$$\mathrm{ECF}(2, \beta) = \sum_{i,j<i\in J} p_T^i p_T^j \, (\Delta R_{ij})^\beta$$



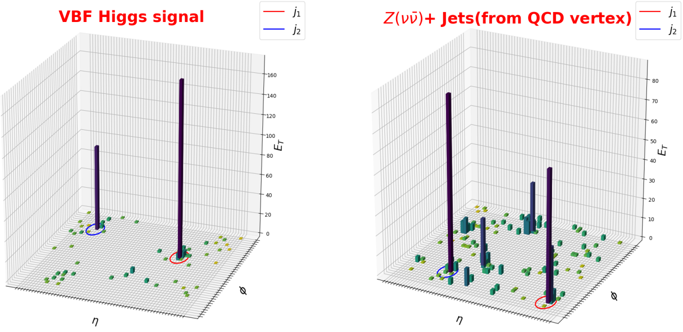
Tower-Image



Salient underlying event structure in Vector-boson fusion(VBF): no color exchanged at LO

Can CNNs leverage information from the full calorimeter tower?

Tower-Image



Salient underlying event structure in Vector-boson fusion(VBF): no color exchanged at LO

Can CNNs leverage information from the full calorimeter tower?

Turns out, we can!

Search for Invisible decays of Higgs at LHC

- Higgs does not couple to ν in SM, couples to dark-matter in many BSM models
- Most recent ATLAS preliminary result[a] puts upper limit on B.R($h \to$inv) < 0.13 at 95% confidence level with $\mathcal{L} = 140$ fb^{-1}.
- Reproduced the shape-analysis of CMS result[b] in our setting, for better comparison of increased sensitivity
  - deliberately weaken cuts in $|\Delta\eta_{jj}|$ and m_{jj}
    $\Rightarrow$Two signals: S_{EW} (VBF) and S_{QCD} (Gluon-fusion)
  - We consider the following major backgrounds:
    - Z_{QCD}: $Z(\nu\bar{\nu}) + jets$
    - W_{QCD}: $W^{\pm}(l^{\pm}\nu) + jets$
    - Z_{EW} : VBF production of $Z(\nu\bar{\nu}) + 2$ *jets*
    - W_{EW} :VBF production of $W^{\pm}(l^{\pm}\nu) + 2$ *jets*

---

[a] ATLAS-CONF-2020-008
[b] Phys. Lett. B 793 (2019) 520 [1809.05937]

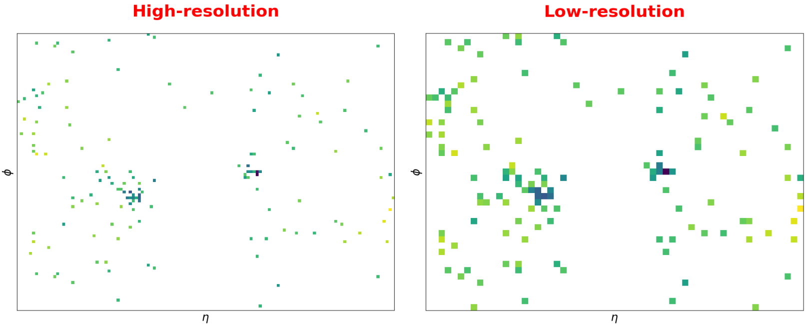
Pre-selection cuts

- **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 $|\eta_{j_i}| < 3$.

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

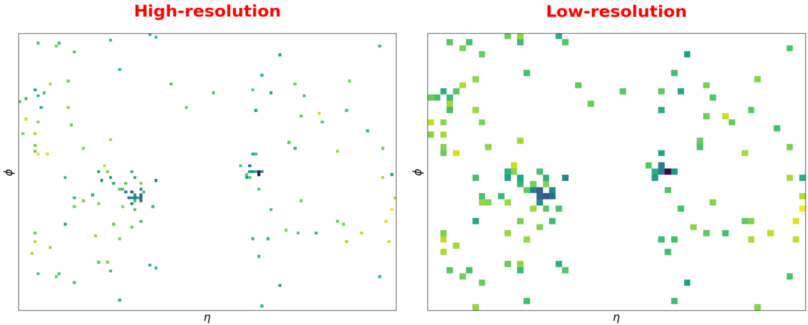
- **Lepton-veto:** No electron(muon) with $p_T > 10$ GeV in the central region, $|\eta| < 2.5(2.4)$.
- **Photon-veto:** No photon with $p_T > 15$ GeV in the central region, $|\eta| < 2.5$
- τ **and b-veto**: no tau-tagged jets in $|\eta| < 2.3$ with $p_T > 18$ GeV, and no b-tagged jets in $|\eta| < 2.5$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

Low-level: Tower-image



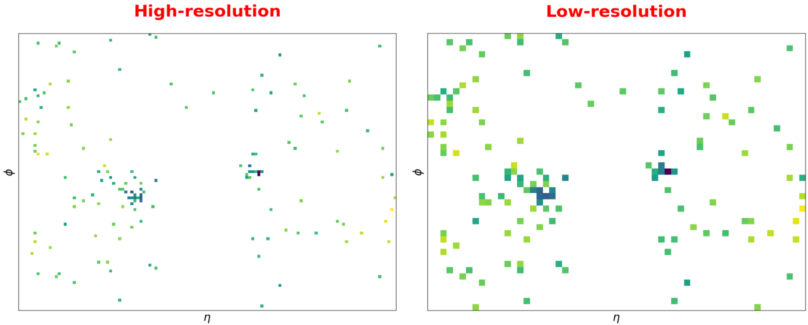
- Pixel wise calorimeter energy deposits (E_T) converted into pictorial description like ‘tower-images’ as input to Convolutional Neural Networks

Low-level: Tower-image



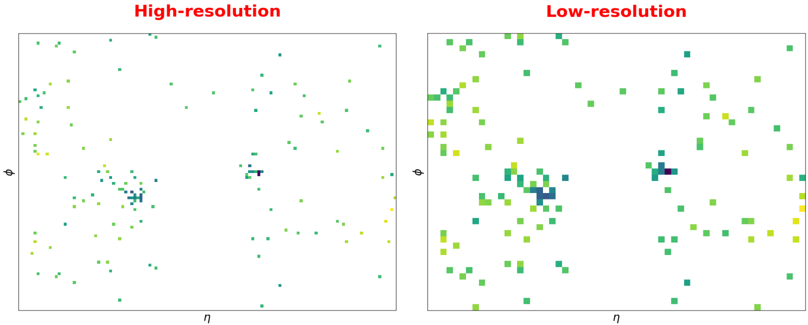
- Different resolution of calorimeter towers in central and forward regions

Low-level: Tower-image



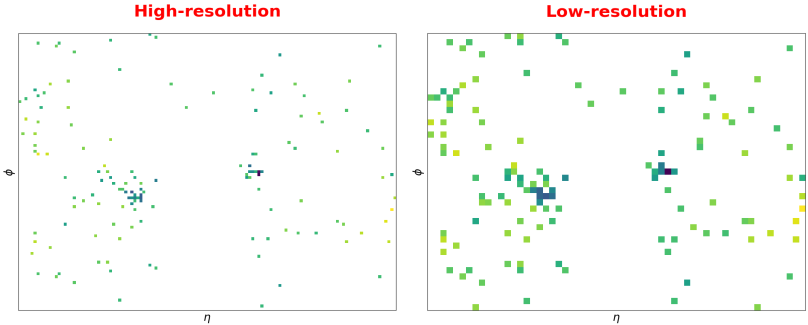
- **Bin-size**: High-resolution(HR) 0.08×0.08 and a low-resolution(LR): 0.17×0.17, $\eta \in (-5, 5)$ and $\phi \in (-\pi, \pi)$

Low-level: Tower-image



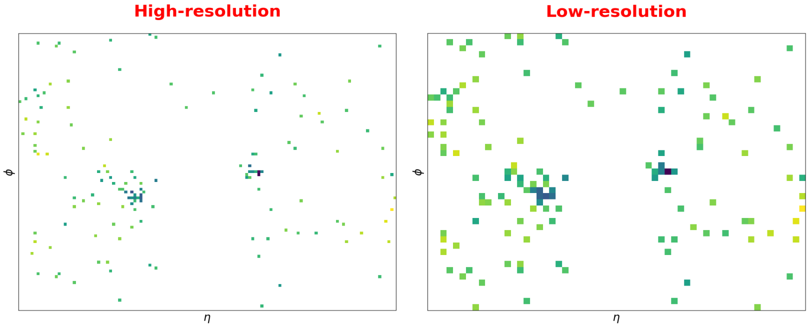
- **Bin-size**: High-resolution(HR) 0.08×0.08 and a low-resolution(LR): 0.17×0.17, $\eta \in (-5, 5)$ and $\phi \in (-\pi, \pi)$
- Periodic in ϕ

Low-level: Tower-image



- **Bin-size**: High-resolution(HR) 0.08×0.08 and a low-resolution(LR): 0.17×0.17, $\eta \in (-5, 5)$ and $\phi \in (-\pi, \pi)$
- **Padding**: padded at each ϕ-boundary with rows from the opposite boundary.

Low-level: Tower-image



- **Bin-size**: High-resolution(HR) 0.08×0.08 and a low-resolution(LR): 0.17×0.17, $\eta \in (-5, 5)$ and $\phi \in (-\pi, \pi)$
- **Padding**: padded at each ϕ-boundary with rows from the opposite boundary.
- **Size** LR: 59×45, and HR: 125×95.

High-level features: Event kinematics and QCD radiation

- **Kinematic**: Information about the event-kinematics from reconstructed objects

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

- **Radiative**: Contains information about the QCD radiation pattern.

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

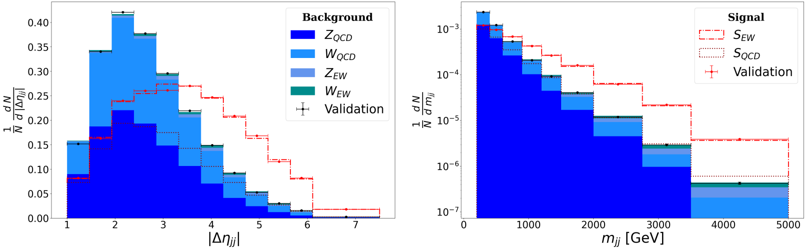
  $\mathcal{E}$: set of chosen η_C's.
  Vary η_C uniformly in the interval [1,5] to get 16 $H_T^{\eta_C}$ variables.

- Combined high-level feature space: $\mathcal{H}$

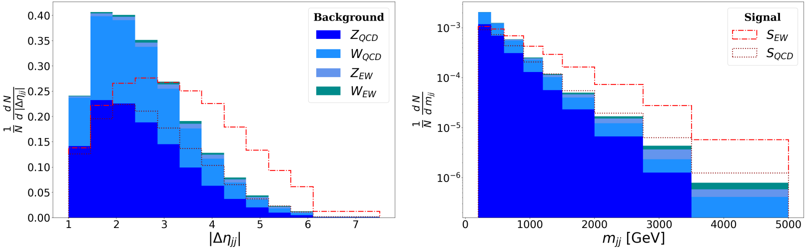
High-level features: Kinematic

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

$MET > 200$ GeV

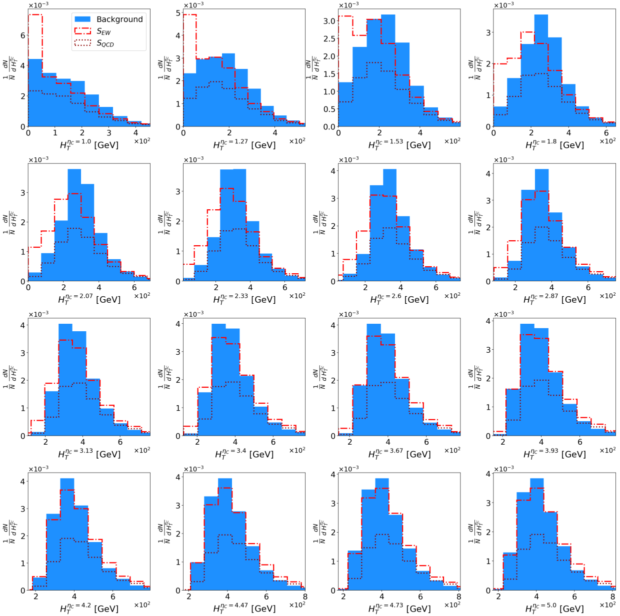


$MET > 250$ GeV

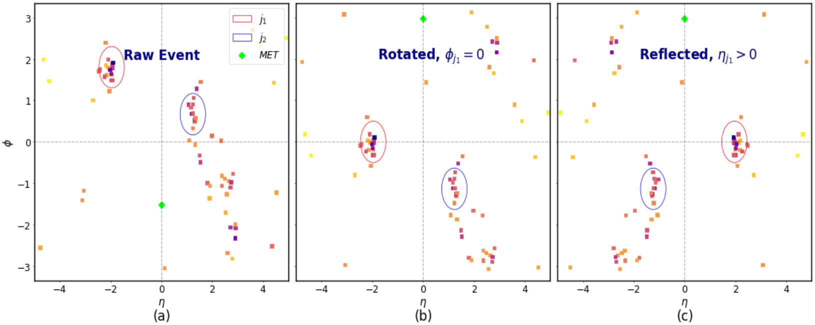


High-level features: QCD-Radiative

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

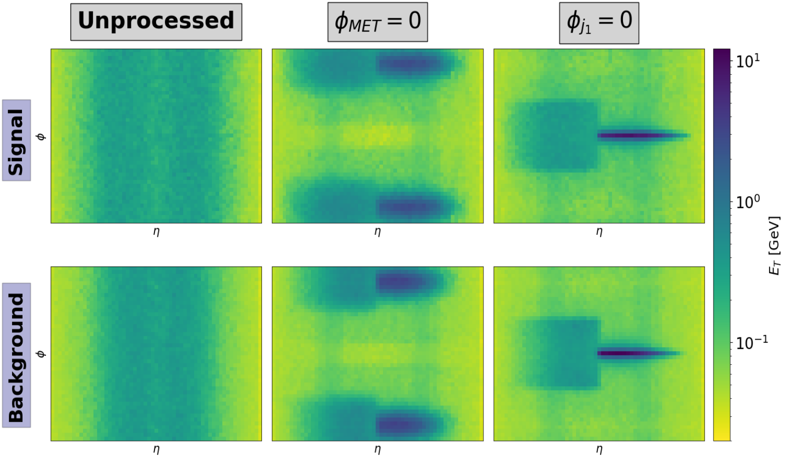


Low-level: Event-preprocessing



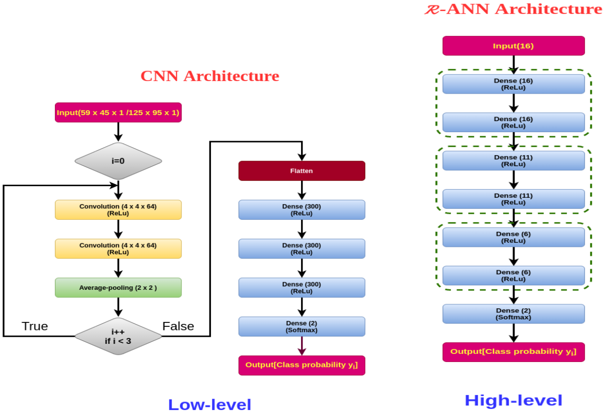
- Rotate along z-axis such that $\phi_0 = 0$. Two instances of $\phi_0 \in \{\phi_{MET}, \phi_{j_1}\}$.
- Reflect along the xy-plane, such that the leading jet's η is always positive.
- After binning (E_T) and padding in LR and HR : $\mathcal{P}^{LR}_{MET}$, $\mathcal{P}^{HR}_{MET}$, $\mathcal{P}^{LR}_{J}$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

Low-level: Event-preprocessing



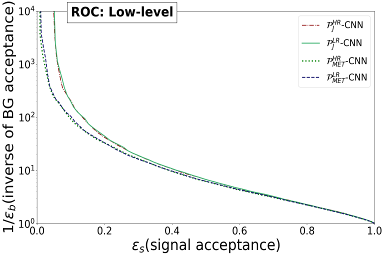
Averaged Images

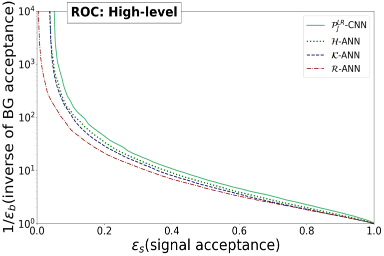
Brief detail of networks

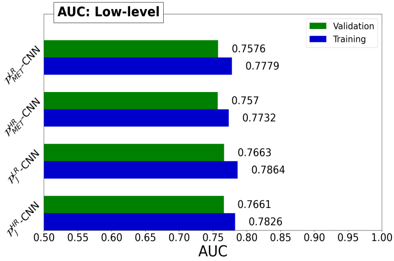


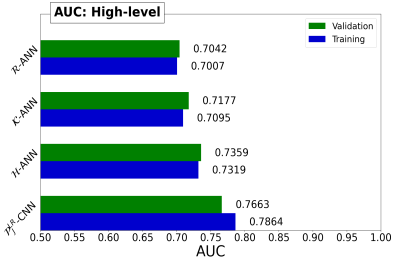
- After training for 20-1000 epochs, best performing network on the validation data choosen (for each of the 7 networks).
- ANN architectures are inspired by the information bottleneck principle, closely related to coarse-graining in RG evolution.

Network Performance

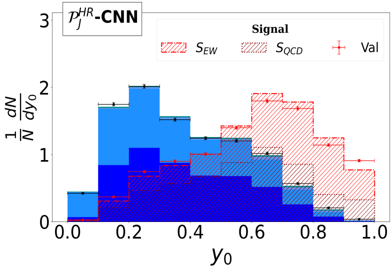


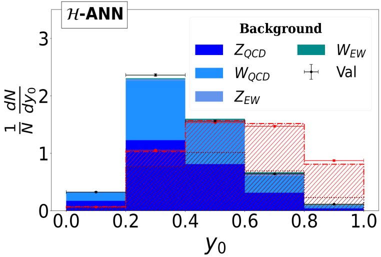






Network Performance: Channel-wise outputs

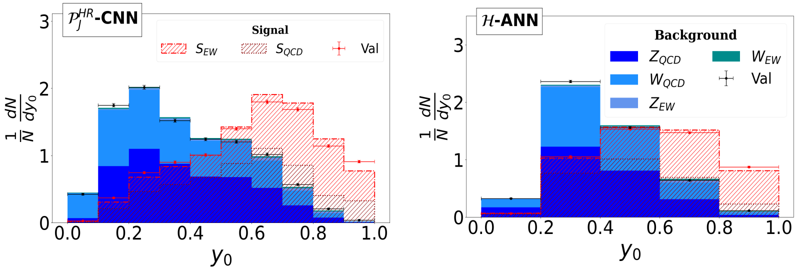




- Harder to distinguish S_{QCD} from the QCD dominated ($\sim 95\%$) background class (significant S_{QCD} contamination in traditional analysis too)

- For the CNN, W_{QCD} dominates over Z_{QCD} in the first bin??

Network Performance: Channel-wise outputs

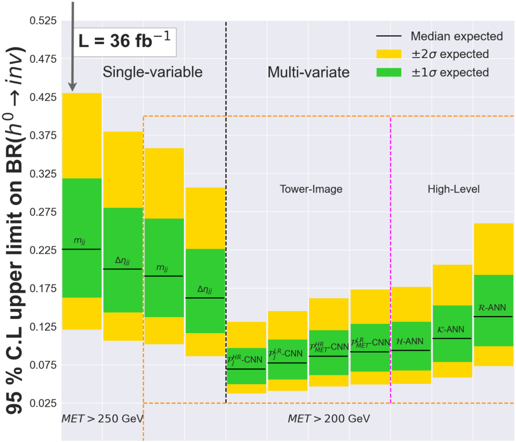


- Harder to distinguish S_{QCD} from the QCD dominated ($\sim 95\%$) background class (significant S_{QCD} contamination in traditional analysis too)

- 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 $\tau^{\pm}$).

Bounds on B.R($h^0 \to$ inv)

Reproduced CMS result at 36 fb^{-1} (actual : BR < 0.25)



Expected 95% C.L median upper limit on the invisible branching ratio of SM Higgs with one and two sigma sidebands.

Bounds on B.R($h^0 \to$ inv)

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	$0.094^{+0.038}_{-0.026}$	$0.065^{+0.026}_{-0.018}$	$0.057^{+0.022}_{-0.015}$

- **factor of three improvement, utilising the same amount of data**.
- It can constrain many different BSM models severely.

Bounds on B.R($h^0 \to$ inv)

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	$0.094^{+0.038}_{-0.026}$	$0.065^{+0.026}_{-0.018}$	$0.057^{+0.022}_{-0.015}$

- Pileup increases the upper-limit within 1σ errors for $\mathcal{P}_J^{HR}$-CNN.

Conclusion

- Posibility to replace decades old dependence on central-jet veto for the reduction of non-VBF backgrounds, in the meantime gaining significantly in performance.

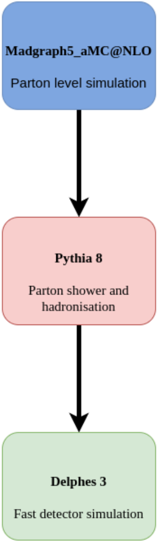
- Low-level calorimeter image outperforms high-level physics motivated features.

    - High-level variables need reconstruction of events.
      $\Rightarrow$ Feasibility of CNN/ANN triggers for VBF?

- Minimally affected by pileup even without any mitigation.

Event simulation details

- Modified version of Higgs Effective Field theory model
  $\Rightarrow$ Higgs decays at parton level to two scalar dark matter particles for signal
- Finite top-mass: Reweight the Missing $E_T(MET)$ distribution
- After preselection cuts: unweighted for Neural Network training
- Parton level cross-sections matched upto 4 and 2 jets for Z_{QCD} and W_{QCD}, respectively



Details of data used in analysis

- Signal and background classes formed by mixing the channels with the expected proportions: $k \times \sigma \times \epsilon_{baseline}$
- **Shape-analysis**($MET > 250$ GeV):
  - Signal: 39% S_{EW} and the 61% S_{QCD}
  - Background: 54.43% Z_{QCD} , 40.92% W_{QCD}, 3.05% Z_{EW} and 1.58% W_{EW}
  - Expected number of background events at 36 fb^{-1} integrated luminosity, scaled for other luminosities.
- **Neural Network analysis**($MET > 200$ GeV):
  - Signal: 44.8% S_{EW} and the 55.2% S_{QCD}
  - Background: 51.221% Z_{QCD} , 44.896% W_{QCD}, 2.295% Z_{EW} and 1.587% W_{EW}
  - 100,000 training and 25,000 validation events for each class
  - Models completely agnostic to validation data
  - Further statistical analysis uses validation data scaled by different luminosities.
- Performed shape-analysis for $MET > 200$ GeV, for a better comparison.