Guided Weakly Supervised Machine Learning for Resonance Searches in Low-Energy Regions of Phase Space in the $\gamma\gamma + 1\tau$ Channel at the Large Hadron Collider

Phodiso Maroeshe[1], Bruce Mellado[1,2],
Edward Nkadimeng[1], Mukesh Kumar[1]

[1]School of Physics and Institute for Collider Particle Physics (WITS), [2] Institute of High Energy Physics, Beijing, University of Chinese Academy of Sciences, 19B Yuquan Road, Shijingshan District, Beijing, China

Institute for Collider Particle Physics
University of the Witwatersrand

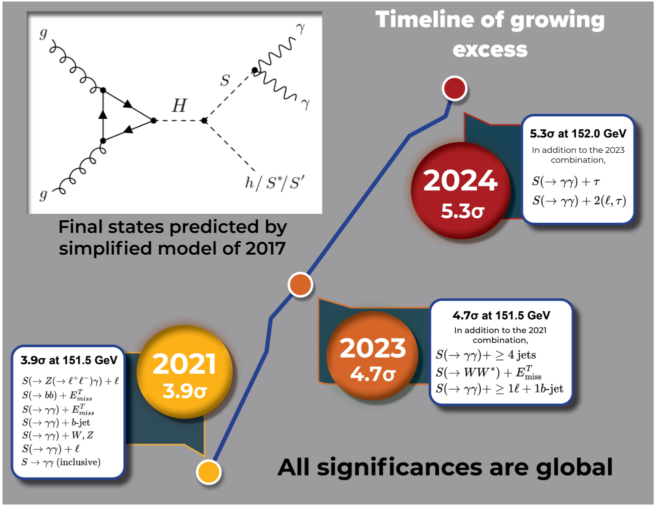
中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

Outline

- Introduction
- Background
- Feynman Diagrams of Physics Process in pp Collision
- Event Simulation and Selections
- Jet to τ Fake Rate
- Events as Graphs
- Feature Distributions
- Correlation between features and the Invariant Mass
- Model Architecture
- Model Performance (Full Supervision & Weak Supervision)
- Conclusion

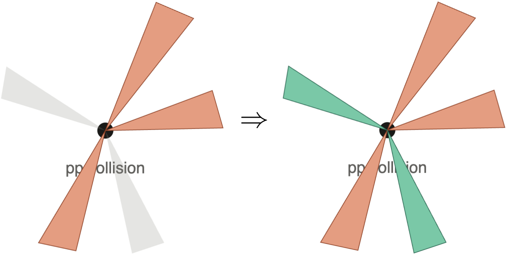
Introduction

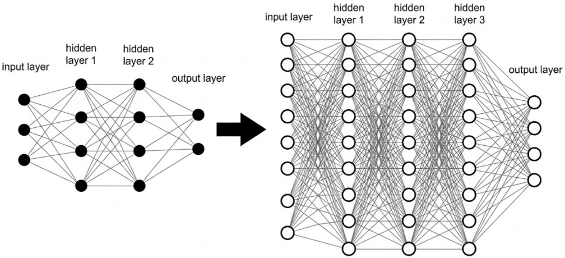
- Multi-lepton anomalies observed in LHC data led to the formulation of the 2HDM+S model.
- This predicted a scalar resonance of mass 152 ± 1 GeV, later observed in the di-photon spectrum.
- The significance of the excess has increased as independent channels are combined, reaching a global significance of 5.3σ at $m_S = 152$ GeV.



Background

- The jet pT threshold is lowered to 10 GeV, extending the search into the soft-jet, low-energy region of phase space that fully supervised analyses typically discard.
- Recovers soft, signal-correlated objects, yielding richer, higher-multiplicity events with more structure per collision.
- Deep learning used to learn directly from the full high-dimensional event rather than a handful of high-pT variables.

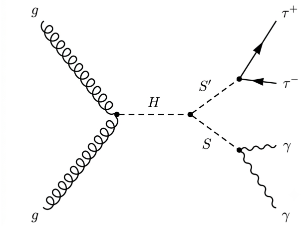




Feynman Diagrams of Physics Processes in pp Collision

Signal

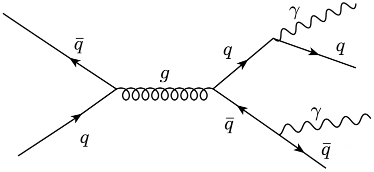
- Gluon-gluon fusion.
- Require 1τ final state.



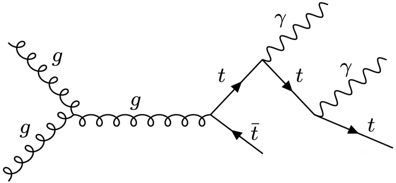
$pp \to H \to SS'$

Background

- Reducible background.
- Dominant is $\gamma\gamma jj$.



$pp \to \gamma\gamma jj$
(dominant)



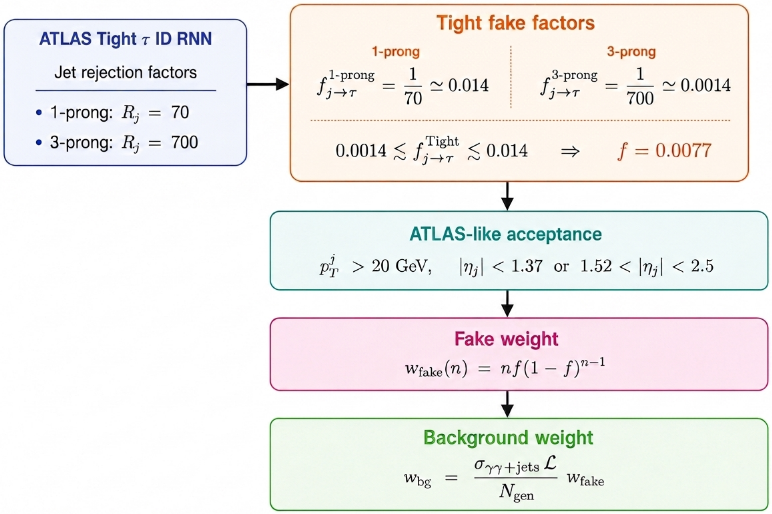
$pp \to \gamma\gamma t\bar{t}$

Event Simulation and Selections

- Used *MadGraph5_aMC@NLO* to generate events.
- Used *PYTHIA 8.2* to simulate parton shower and hadronization.
- Process the simulated events with *DELPHES 3*.

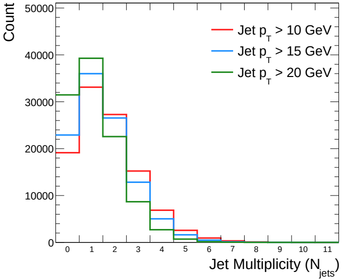
Selection Cut	**Signal Passed**	**Background Passed**
Initial Events	1800000	600000
$N_{\text{leptons}} = 0$	1137189	582963
$N_{\gamma} = 2$	500243	384489
$p_T^{\gamma} > 25$ GeV	449530	366986
Leading Photon $p_T^{\gamma} > 35$ GeV	436731	364358
$N_{\tau\text{-jets}} \geq 1$	115737	246920

Jet $\to \tau$ Fake Rate



Events as Graphs

- Each event has a variable count of jets.
- Jet (physical objects) kinematics as graph nodes.
- Construct fully connected, homogeneous graph.



light jets

τ-jets

$$\mathbf{x}_i = \begin{pmatrix} p_T^{(i)} \\ \eta^{(i)} \\ \phi^{(i)} \end{pmatrix}$$

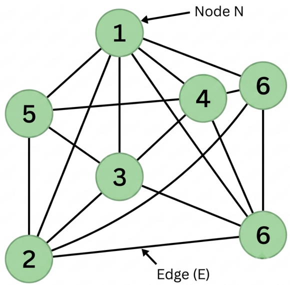
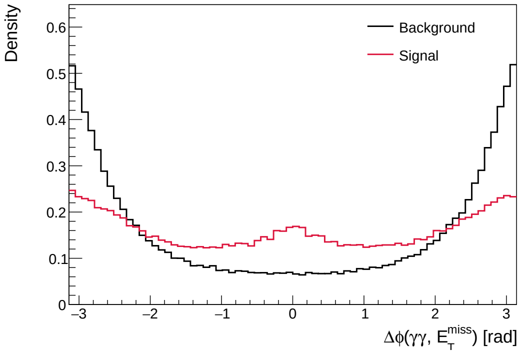
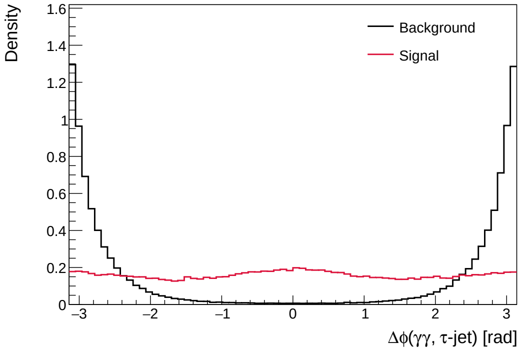


Table: Introduced global features

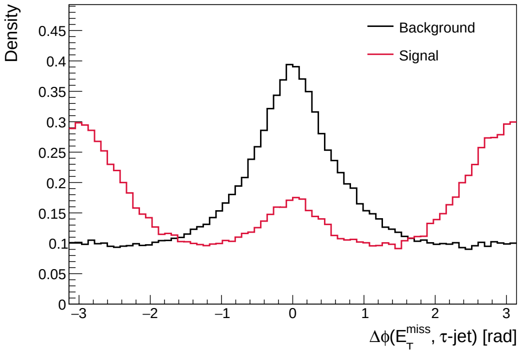
Variable	Description
$E_{\mathrm{T}}^{\mathrm{miss}}$	Missing transverse energy
N_{j}	Number of jets
$\Delta\phi(\gamma\gamma, E_{\mathrm{T}}^{\mathrm{miss}})$	Azimuthal angle between $(\gamma\gamma, E_{\mathrm{T}}^{\mathrm{miss}})$
$\Delta\phi(\gamma\gamma, \tau\text{-jet})$	Azimuthal angle between $(\gamma\gamma, \tau\text{-jet})$
$\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, \tau\text{-jet})$	Azimuthal angle between $(E_{\mathrm{T}}^{\mathrm{miss}}, \tau\text{-jet})$

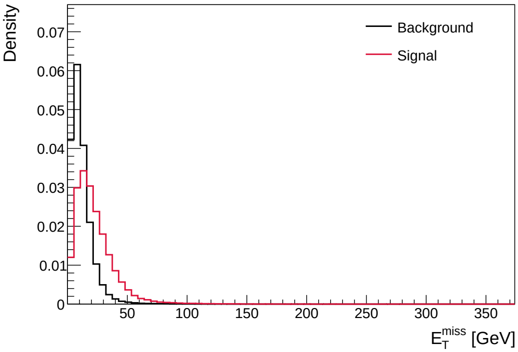
Feature Distributions



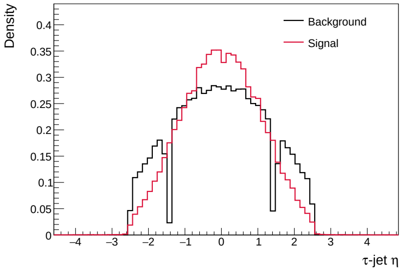


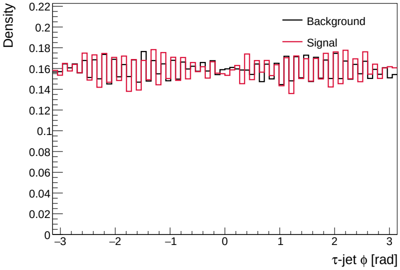
Feature Distributions

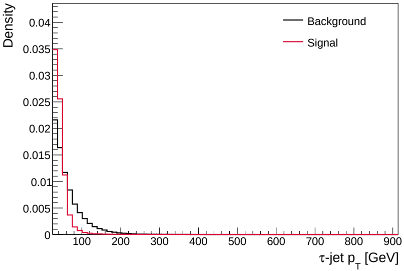




Feature Distributions (τ − jet)



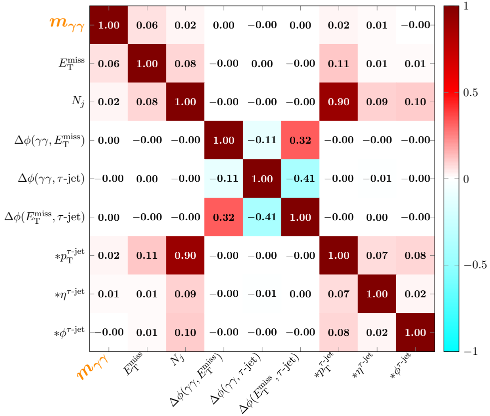


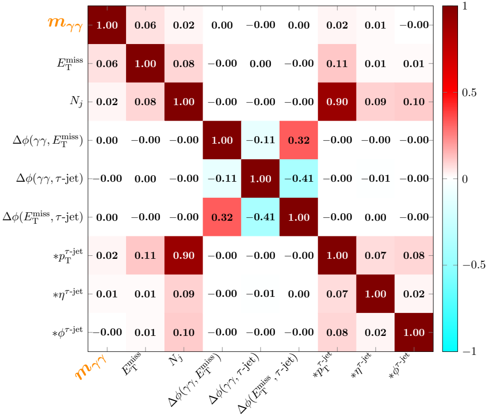


Correlation Between Features

- Spearman monotonic correlation.
- Invariant Mass uncorrelated with global features ('*' $\tau-$jet).

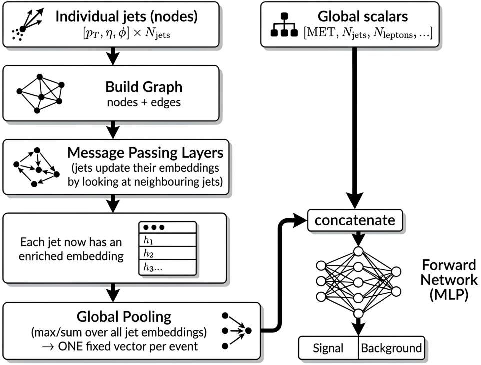
$$\rho_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$





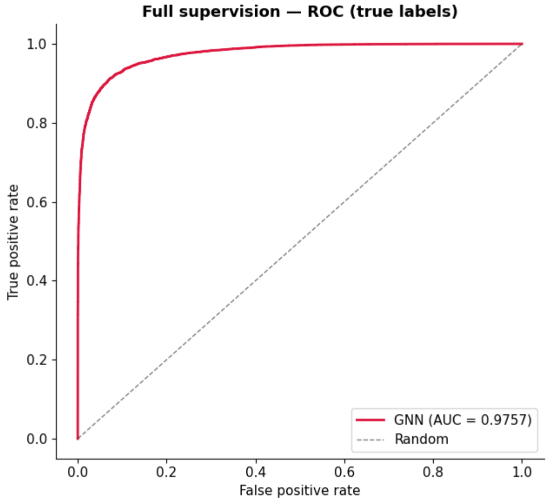
Model Architecture

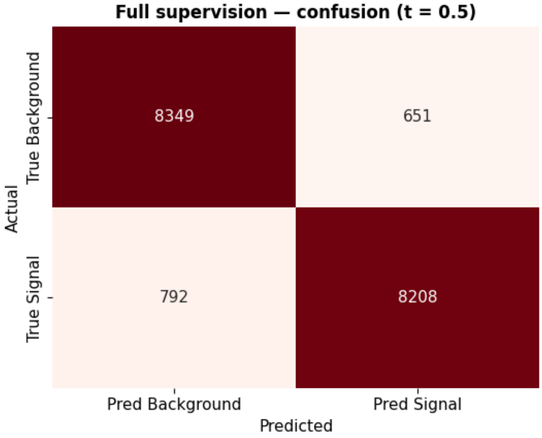
- ***EdgeConv*** message passing over jet nodes.
- Pool jet embeddings into one event vector.
- Combine with global features.
- Multilayer feedforward Artificial Neural Network with ***ReLU*** activation function.
- Final classification layer with ***sigmoid*** activation function.



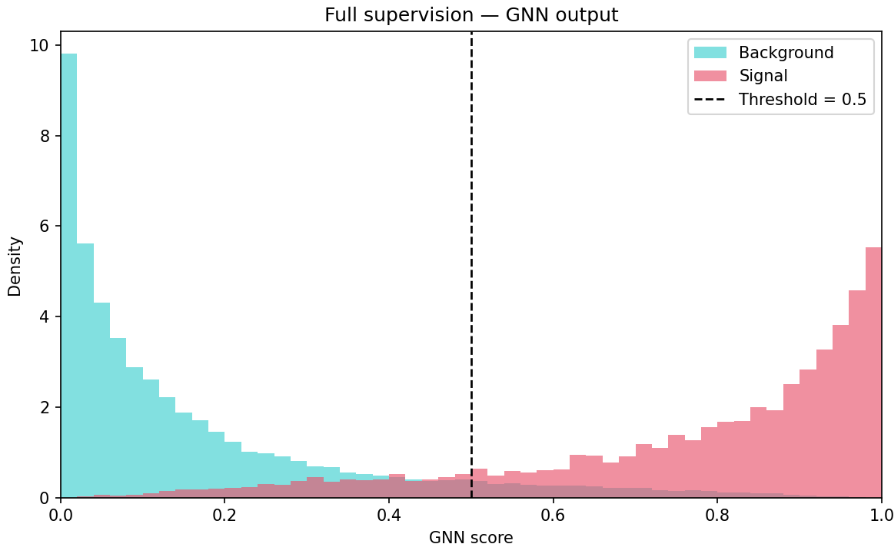
$$\mathbf{h}_i^{(l+1)} = \gamma^{(l)} \left(\mathbf{h}_i^{(l)}, \bigoplus_{j \in \mathcal{N}(i)} \psi^{(l)} \left(\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)} \right) \right)$$

Model Performance (Full Supervision)





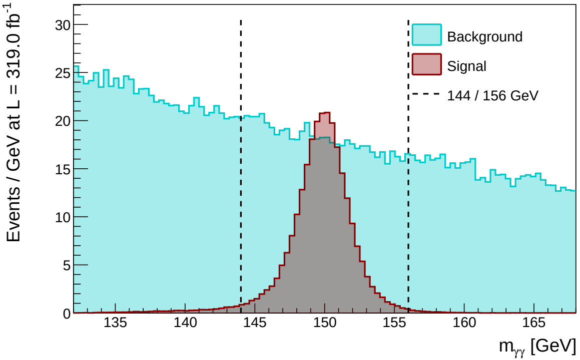
Model Performance (Full Supervision)

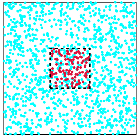


Weak Supervision : Population Discrimination Instead of Instance Labels

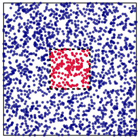
- Model-independent search.
- Weak labels formed by sideband and signal region.
- Di-photon invariant mass ($m_{\gamma\gamma}$) is the weak supervision model guide

$$
\begin{aligned}
R_S &= \{e : 144 < m_{\gamma\gamma}(e) < 156\}, \\
R_B^{\text{low}} &= \{e : 132 < m_{\gamma\gamma}(e) < 144\}, \\
R_B^{\text{high}} &= \{e : 156 < m_{\gamma\gamma}(e) < 168\}
\end{aligned}
$$





True Labels

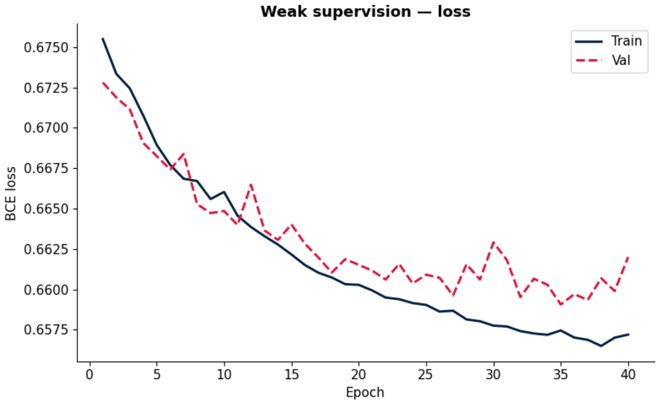


Weak Labels

Weak Supervision (Preliminary Results)

- Hard learning problem.
- AUC greater than 0.54 hints at a present learning signal

Derived Metrics	
Accuracy	0.618
Precision (SigRegion)	0.805
Recall (SigRegion)	0.087
Specificity (Sideband)	0.985
ROC AUC	0.5475



Conclusion

- Excesses at the electroweak scale motivate searches that reach into soft, low-energy regions of phase space.
- The ability of deep learning to handle high dimensional data offers great opportunity to extend the search in previously unexplored regions.
- Fully supervised methods are highly trainable for these searches however they are dependent on per event labels
- Weak supervision is a hard learning problem however removes the need for event-level truth labels, allowing for a model-independent search
- Model performance is still to be optimized. Hyper-parameter tuning and drop-out shall be explored, then a cut on the GNN score will be exercised to suppress background to reveal the resonance.

THANK YOU!