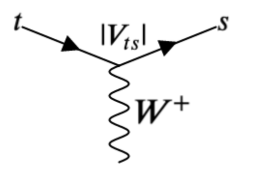
Deep Learning for the Rare Top Decay t→sW at the LHC

Jeewon Heo[1], Woojin Jang[1], Jason Lee[1], Inkyu Park[1], Youn Jung Roh[1], Ian James Watson[1], Seungjin Yang[2]

[1] University of Seoul, [2] Kyung Hee University

Introduction

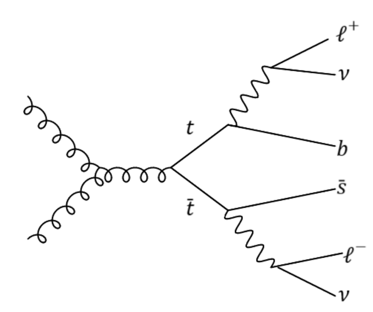
- The Cabibbo-Kobayashi-Maskawa (CKM) matrix describes the flavor-changing charged weak interaction
- Identification of a strange jet originating from top quark decays (t→sW) is important task to achieve a direct measurement of $|V_{ts}|$ [1, 2]
- We propose a novel deep learning model based on the Self-Attention mechanism to find the jets decaying from the t→sW decay in the top pair production with dilepton final state



$$|V_{CKM}| = \begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & \underline{|V_{ts}|} & |V_{tb}| \end{pmatrix} \approx \begin{pmatrix} 0.97435 & 0.22500 & 0.00369 \\ 0.22486 & 0.97349 & 0.04182 \\ 0.00857 & \underline{0.04110} & 0.999118 \end{pmatrix}$$

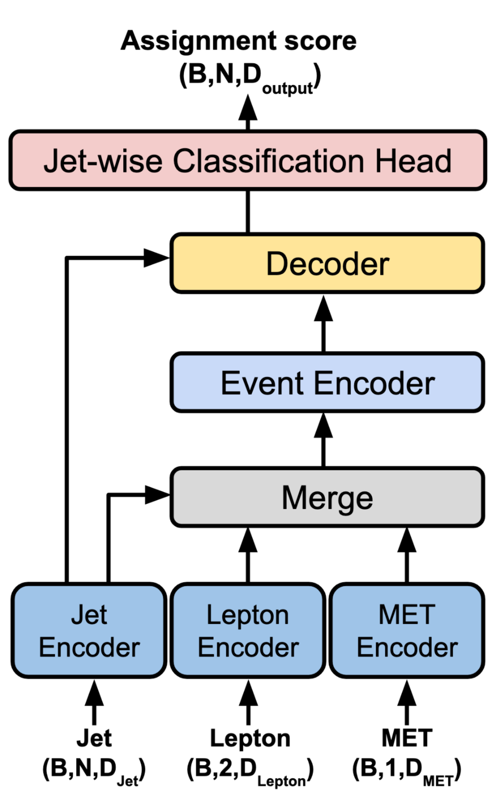
Analysis Setup

- Sample generation
  - Our signal process is $t\bar{t} \rightarrow sWbW$, where both W bosons decay into leptons (e, μ)
  - Our dominant background process is $t\bar{t} \rightarrow bWbW$, we also produce Drell-Yan+jj, Single Top, and diboson processes
  - Samples are generated using MadGraph5_aMC@NLO and PYTHIA 8, followed by simulating the CMS-like detector response with Delphes 3



- Selection for Top pair production with dilepton final state events
  - Electron
    - $p_T \geq 20$
    - $|\eta| \leq 1.442,\ 1.566 \leq |\eta| \leq 2.4$
    - Isolation > 0.12
  - Muon
    - $p_T \geq 20$
    - $|\eta| \leq 2.4$
    - Isolation > 0.15
  - Jet
    - $p_T \geq 30$
    - $|\eta| \leq 2.4$
    - ΔR (j, l) > 0.4
- Event selection
  - Two leptons with opposite charge, M_{ll} > 20 GeV
  - Veto Z boson ($|M_Z - M_{ll}|$ > 15 GeV), p_T^{miss} > 40 GeV in (ee / $\mu\mu$) channel
  - At least 2 jets, Number of b-tagged jet < 2

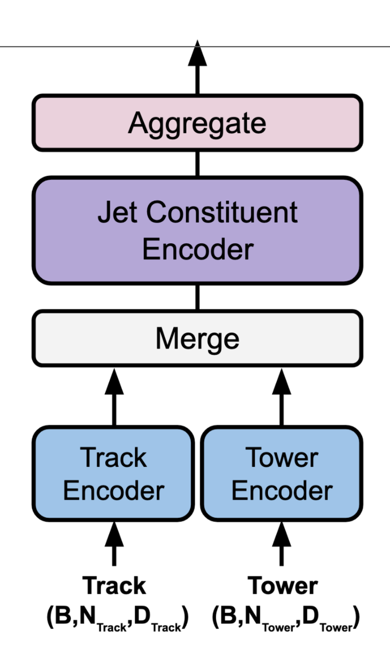
SAJA-Dilepton Model

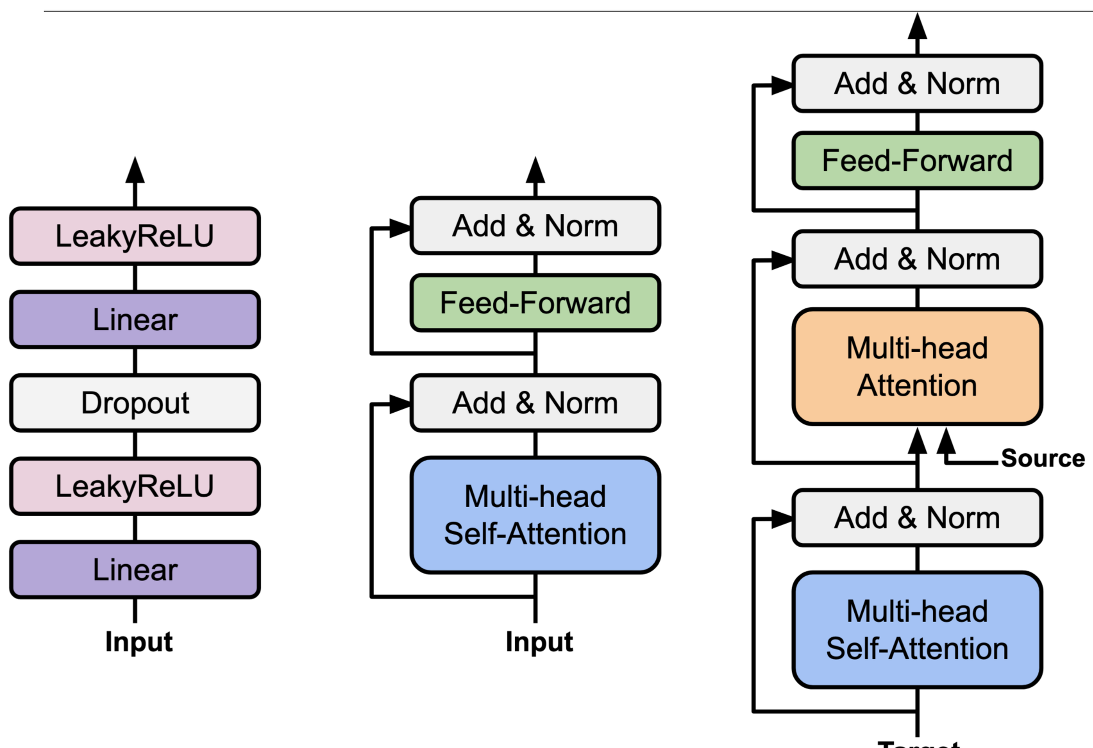


- We modified the Self-Attention for Jet Assignment (SAJA) model [3] for dilepton events
- Object encoder blocks deal with multi-modal inputs
- Information of objects that sent to the same latent space are combined in the merge block
- The event encoder takes output of merge block and calculates their relationship using self-attention mechanism
- The decoder block receives output of event encoder and jet encoder and derives role of jets in event
- Jet-wise classification head outputs categorical scores for jets in events

---

- Jet properties can be derived from jet constituent
  - We can calculate known properties such as number of particles in jet, jet shape, and fragmentation function of jet
  - These high-level features don't capture all information from the constituent
- We propose the model that can take jet constituents
  - The graph on the right can replace jet encoder of SAJA-Dilepton model





- Left: Feed-Forward block
  - Object encoders are feed-forward block
- Middle: Self-Attention block
  - Event encoder and jet constituent encoder are self-attention block
- Right: Decoder block
- In the self-attention block and Decoder block, Dropout is employed to the output of each sublayer

Model Training

- For the training $t\bar{t} \rightarrow sWbW$(signal) and $t\bar{t} \rightarrow bWbW$ (background) are used
  - Targets for signal sample, t→s parton matched events are used
- The task of the SAJA-Dilepton model is a jet-wise classification of events
  - The task is specified by using jet-wise cross-entropy loss
- Training variables
  - Jet: Momentum components, particle multiplicities, jet shape, energy sharing variable [4], b tagging information, jet charge
  - lepton: momentum components, flavor, charge
  - p_T^{miss}, azimuthal angle ϕ of p_T^{miss}
  - Jet constituents: Momentum components of particle, difference of η and ϕ between particle and jet axis, p_T of particle relative to jet p_T, p_T perpendicular to jet axis, p_T perpendicular to jet axis, impact parameter value, charge, EM, hadronic energy

$$\left\{ \begin{pmatrix} J^{(1)} \\ \vdots \\ J^{(N)} \end{pmatrix} + \begin{pmatrix} L^{(1)} \\ L^{(2)} \\ p_{\mathrm{T}}^{\mathrm{miss}} \end{pmatrix} \right\} \rightarrow \mathrm{Model} \rightarrow \begin{pmatrix} y_{t\rightarrow s}^{(1)} & y_{t\rightarrow b}^{(1)} & y_{other}^{(1)} \\ \vdots & \vdots & \vdots \\ y_{t\rightarrow s}^{(N)} & y_{t\rightarrow b}^{(N)} & y_{other}^{(N)} \end{pmatrix}$$

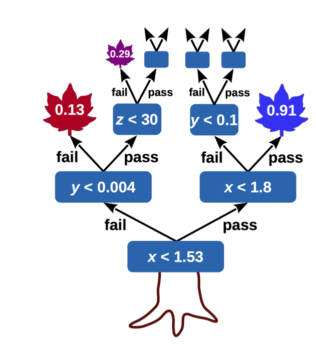
A simple diagram of model input and output structure for an event

$$L(\theta) = \frac{1}{N} \sum_{j=1}^{N} (y_{t\rightarrow s}^{(j)} \log \hat{y}_{t\rightarrow s}^{(j)} + y_{t\rightarrow b}^{(j)} \log \hat{y}_{t\rightarrow b}^{(j)} + y_{other}^{(j)} \log \hat{y}_{other}^{(j)})$$

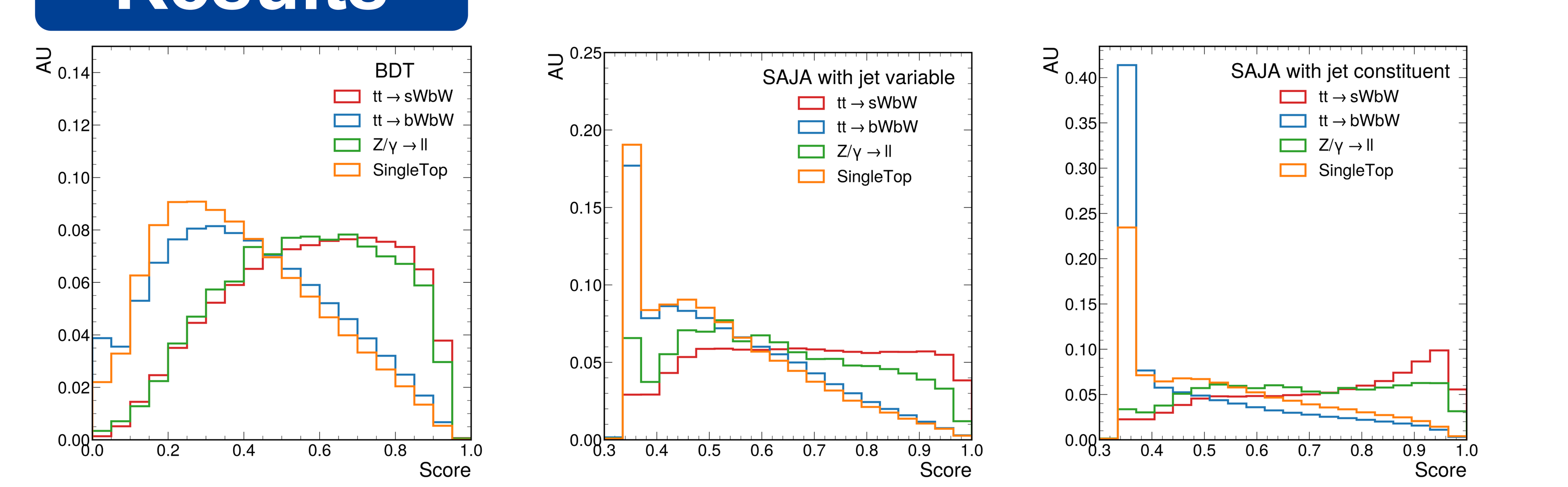
Jet-wise cross-entropy loss

Baseline

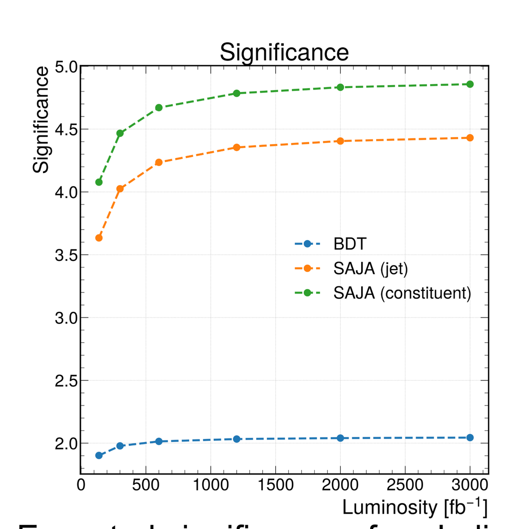
- Boosted Decision Trees (BDT) are used as a baseline model
- For the implementation of BDT, the XGBoost library is used
- BDT is trained to classify jets from t→sW decay jet-wisely
- Jet variables listed in training of SAJA-Dilepton model are used



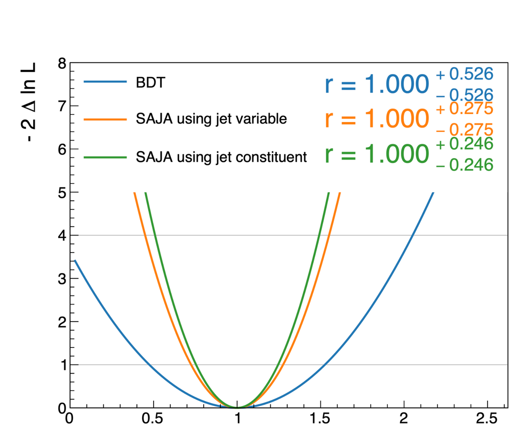
Results



- We use t→s score of models to discriminate signal and backgrounds
- Score distribution is used as input to the binned likelihood fit
- Expected limits and significances are calculated with toy dataset (Asimov)
- Only MC statistics error is considered as a systematic



Expected significance of excluding scenarios with $|V_{ts}| = 0$



Expected Limits of $|V_{ts}|$

- Expected significances are obtained from Run 2 to HL-LHC luminosities with lumi projection
- Expected limits are calculated with Run 2 luminosity
- We obtained expected limits of $0.0293 < |V_{ts}| < 0.0502$ @ 95% CL with SAJA-Dilepton using jet constituent model

Conclusion

- We introduced the models using self-attention mechanism that can apply to various types of input objects
- We compared SAJA-Dilepton models with the baseline model and SAJA-Dilepton models show better performance
- In this study, we can exclude scenarios with $|V_{ts}| = 0$ up to a significance level of ~4 σ at the LHC Run 2 luminosity, considering MC statistics only

Reference

[1] Ahmed Ali, Fernando Barreiro, and Theodota Lagouri. Prospects of measuring the CKM matrix element |Vts| at the LHC. Phys. Lett. B, 693:44–51, 2010.

[2] Woojin Jang, Jason Sang Hun Lee, Inkyu Park, and Ian James Watson. Measuring |Vts| directly using strange-quark tagging at the LHC. J. Korean Phys. Soc., 81(5):377–385, 2022.

[3] Jason Sang Hun Lee, Inkyu Park, Ian James Watson, and Seungjin Yang. Zero-Permutation Jet-Parton Assignment using a Self-Attention Network. arXiv:2012.03542

[4] CMS. Performance of quark/gluon discrimination in 8 TeV pp data. Technical report, CERN, 2013.

6th Inter-experiment Machine Learning Workshop
28th Jan. - 2nd Feb. 2024

Jeewon Heo
E-mail: jeewon.heo@cern.ch