Why Does Transformer Not Work For The Higgs Boson?

Wenhao Pan, Maria Giovanna Foti, Shuo Han, Xiangyang Ju, Haichen Wang

Abstract

Covariant Particle Transformer is a recently proposed neural network model for predicting the top quark kinematic properties from reconstructed final state objects has been proposed. When used to predict the Higgs boson kinematic properties, it underpredicts the transverse momentum of the Higgs boson with high transverse momentum values. Three techniques – advanced model training procedure, ratio-pull-based bias correction, and loss function reweighting – are designed to address this issue. Our experiment results show that the best model has significantly better performance for high transverse momentum Higgs bosons with a cost of worse performance for low transverse momentum Higgs bosons.

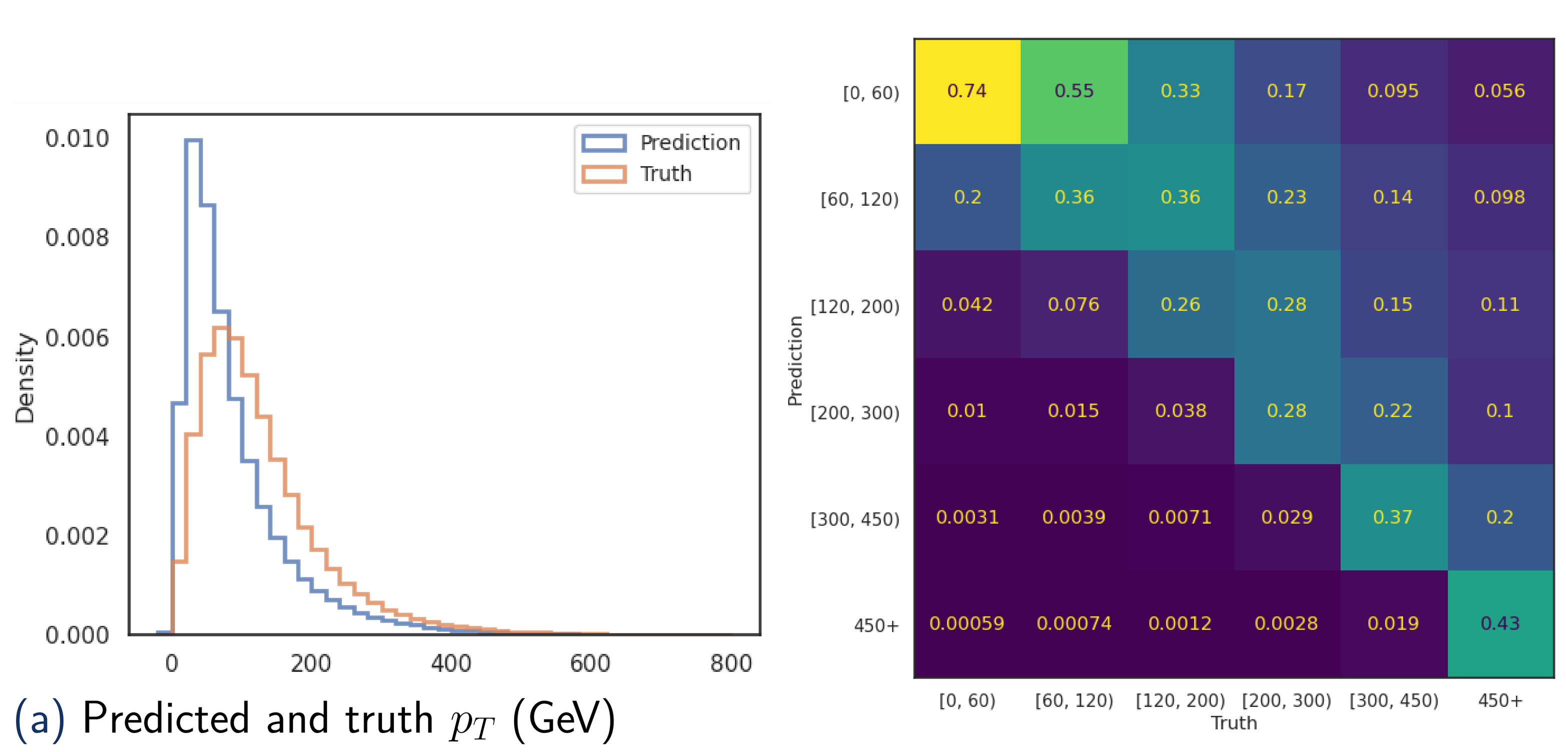
Original Model Performance

The Covariant Particle Transformer (CPT) [1] is a variant of the Graph Transformer model – a combination of graph neural network and transformer architectures. [1] claim the following features of CPT:

- Has covariant attention mechanism to account for symmetry transformation.
- Achieves significantly better resolution than the traditional triplet-based reconstruction method.
- Well reproduces the correlation between top quarks.

Fitted on 7.3 million $t\bar{t}H$ events with lepton $p_T > 10$ GeV, $1 \leq N_{\text{lepton}} \leq 4$, jet $p_T > 25$ GeV, jet $|y| < 4.5$, and tau $p_T > 20$ GeV to predict Higgs boson kinematic properties, the original model predicts rapidity, azimuthal angle, and mass well, but it underpredicts transverse momentum (p_T) of high p_T samples as shown in Figure 1.

Figure 1: p_T (GeV) prediction performance of the original model.



(a) Predicted and truth p_T (GeV) distribution

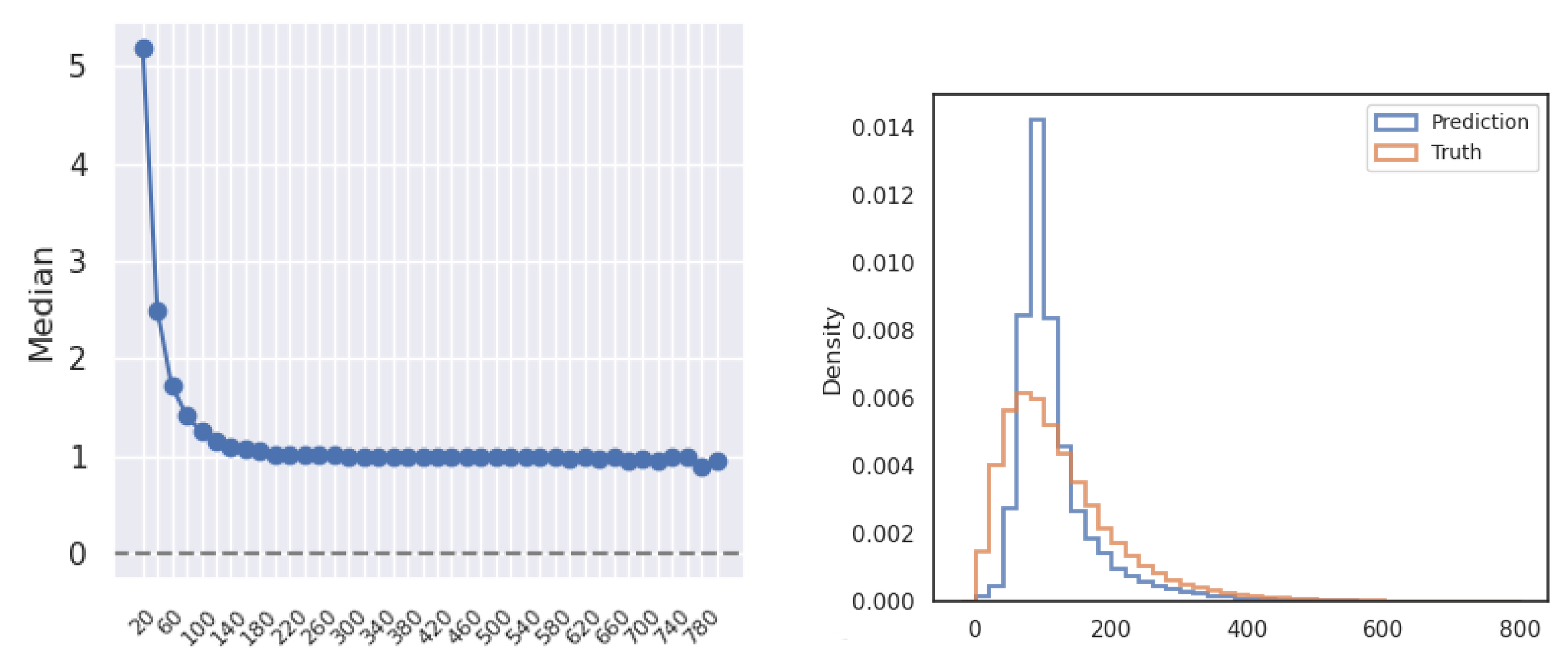
(b) p_T (GeV) migration matrix

Improvements

Advanced training procedure Apply linear warmup and exponential decay learning rate scheduling to stabilize the training process. A sample learning rate curve in Figure 4.

Ratio-pull correction p_T ratio-pulls are ratios of true to predicted values. Figure 2a plots the ratio-pull median of samples in each predicted p_T bin (20 GeV width). Multiply each prediction by the ratio-pull median of its bin and produce a worse performance in Figure 2b.

Figure 2: p_T (GeV) prediction performance of the ratio-pull-corrected model.



(a) p_T (GeV) ratio-pull of the original CPT

(b) Predicted and truth p_T (GeV) distribution

Reweighted loss function The p_T distribution of the training samples in Figure 3 is right-skewed. Reweighting loss function is a potentially effective strategy for imbalanced regression [2]. Weights of samples are computed based on their Gaussian Kernel estimated densities through DenseWeight [3]. The reweighted loss function is equation (1).

$$L(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} w_i \cdot (y_i - \hat{y}_i)^2 \tag{1}$$

where w_i, y_i, and $\hat{y}_i$ are weights, true p_T, and predicted p_T.

Figure 3: p_T (GeV) distribution, estimated density (red), and computed weights (blue)

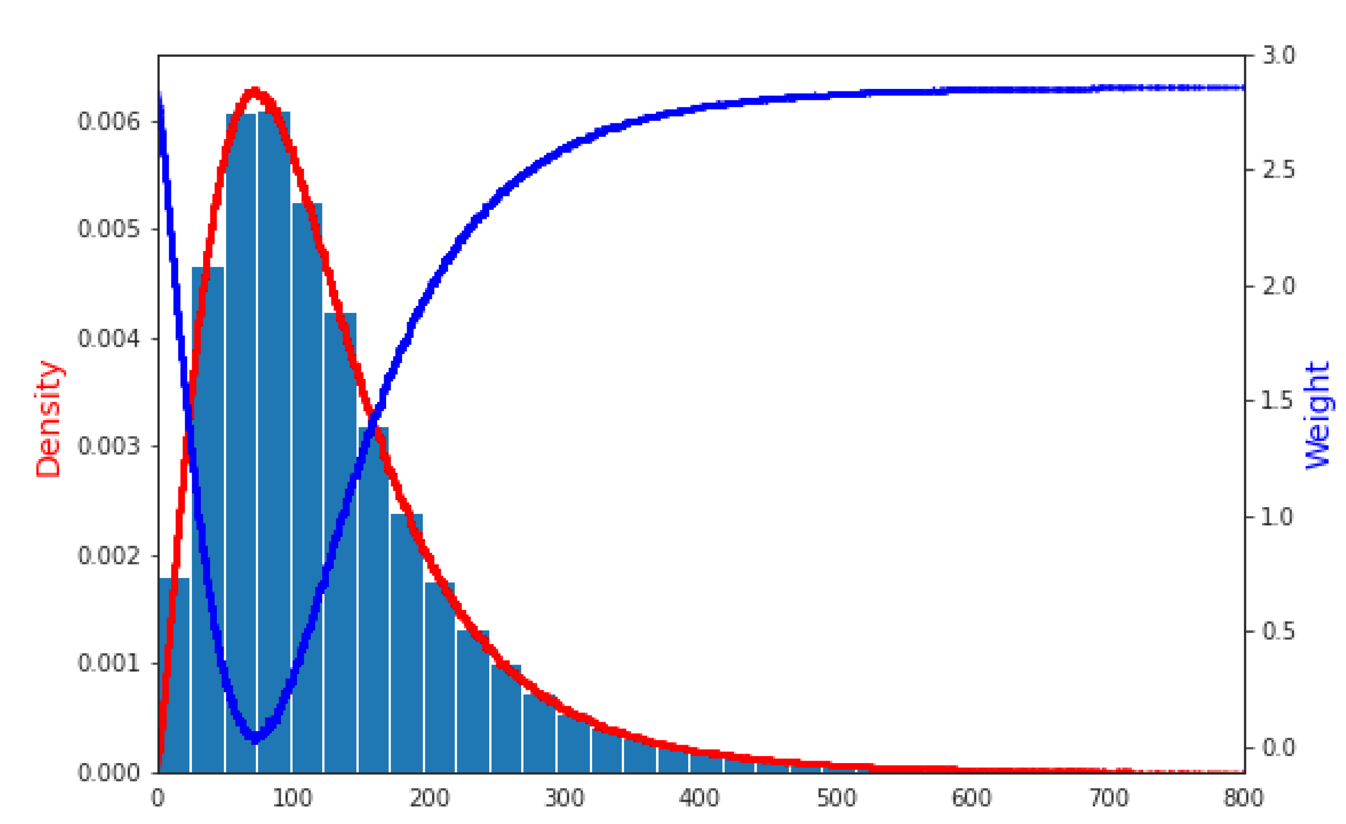
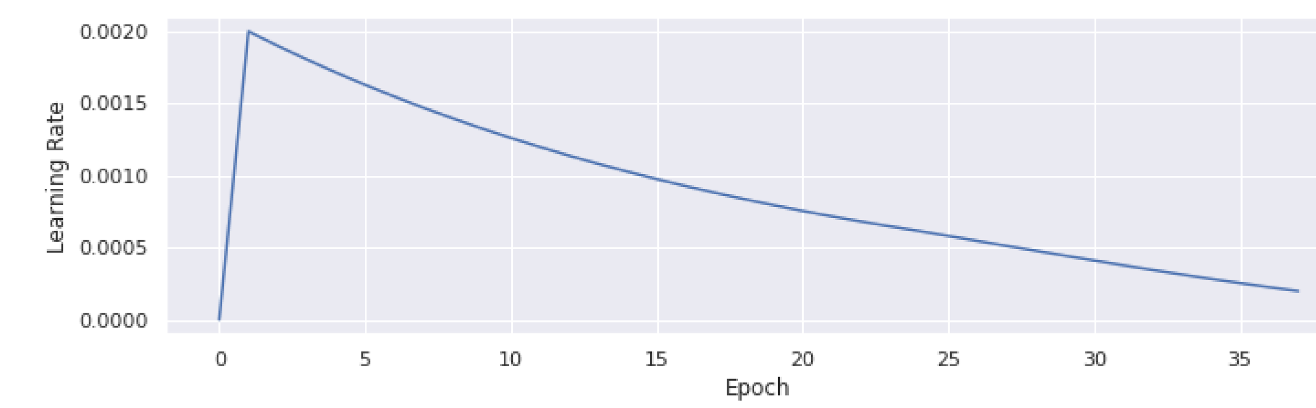


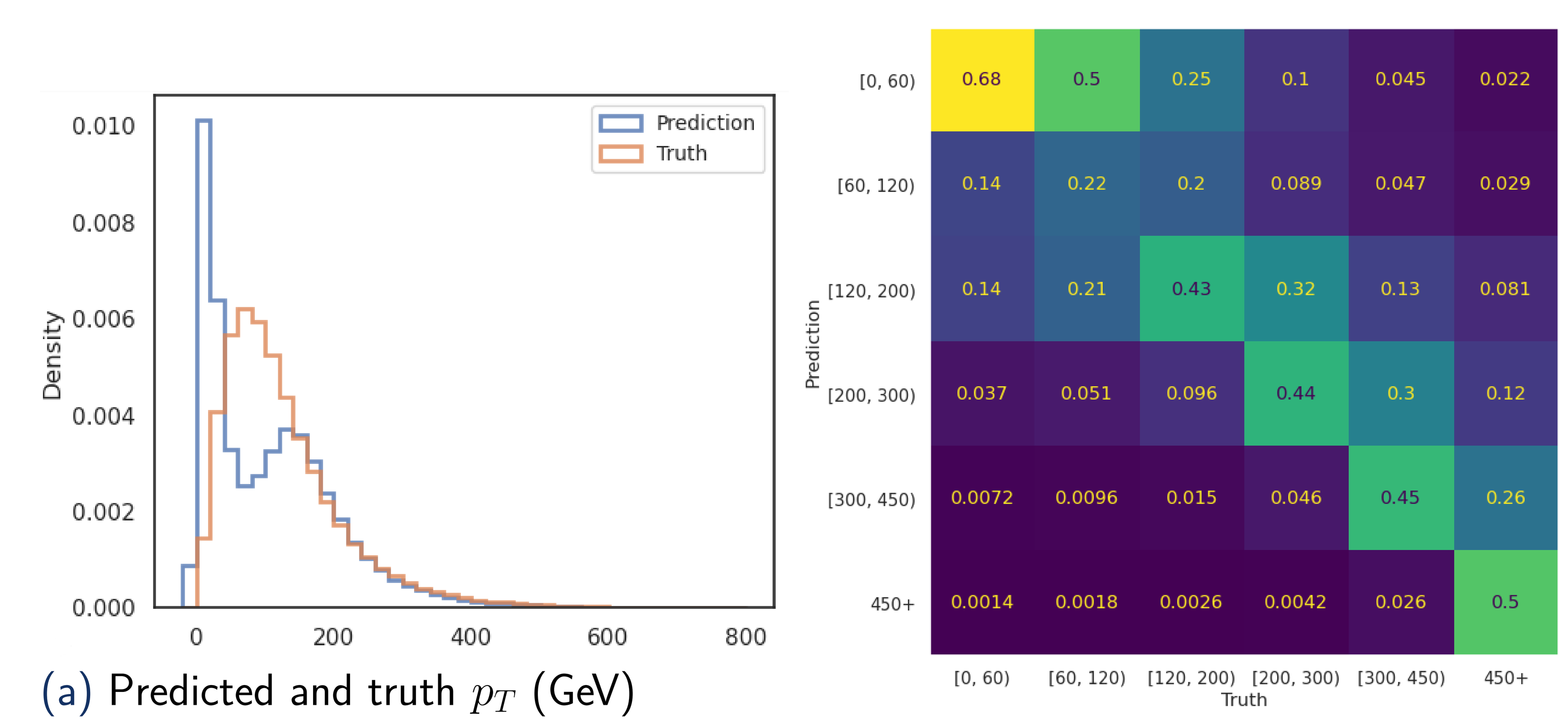
Figure 4: Learning rate curve



Best Model Performance

The best model we have found has the performance shown in Figure 5.

Figure 5: p_T (GeV) prediction performance of the best model.



(a) Predicted and truth p_T (GeV) distribution

(b) p_T (GeV) migration matrix

p_T predictions of high p_T (≥ 120 GeV) samples are significantly improved. p_T predictions of low p_T (< 120 GeV) samples become worse.

Conclusion

We partially address the underprediction issue as we predict high p_T samples better with a loss in low p_T samples prediction. A possible next step is to fix this loss while maintaining the high p_T sample performance.

References

[1] Shikai Qiu, Shuo Han, Xiangyang Ju, Benjamin Nachman, and Haichen Wang. A holistic approach to predicting top quark kinematic properties with the covariant particle transformer. *arXiv preprint arXiv:2203.05687*, 2022.

[2] Yuzhe Yang, Kaiwen Zha, Yingcong Chen, Hao Wang, and Dina Katabi. Delving into deep imbalanced regression. In *International Conference on Machine Learning*, pages 11842–11851. PMLR, 2021.

[3] Michael Steininger, Konstantin Kobs, Padraig Davidson, Anna Krause, and Andreas Hotho. Density-based weighting for imbalanced regression. *Machine Learning*, 110(8):2187–2211, 2021.

Acknowledgements

We thank Shikai Qiu for explaining his design of CPT and providing feedback on our improvement of CPT and Weiming Yao for providing the $t\bar{t}H$ samples for training CPT.