PROCEEDINGS OF SCIENCE

METNet: A combined missing transverse momentum working point using a neural network with the ATLAS detector

Benjamin Hodkinson*, **on behalf of the ATLAS Collaboration**
University of Cambridge
E-mail: ben.hodkinson@cern.ch

In order to suppress pile-up effects and improve resolution, the ATLAS experiment at the LHC employs a suite of working points for missing transverse momentum ($p_{\mathrm{T}}^{\mathrm{miss}}$) reconstruction, and each is optimal for different event topologies and different beam conditions. A neural network (NN) can exploit various event properties to pick the optimal working point on an event-by-event basis, and also combine complementary information from each of the working points. The resulting regressed $p_{\mathrm{T}}^{\mathrm{miss}}$ (METNet) offers improved resolution and pile-up resistance across a number of different topologies compared to the current $p_{\mathrm{T}}^{\mathrm{miss}}$ working points. Additionally, by using the NN's confidence in its predictions, a machine learning-based $p_{\mathrm{T}}^{\mathrm{miss}}$ significance ('METNetSig') can be defined. This contribution presents simulation-based studies of the behaviour and performance of METNet and METNetSig for several topologies compared to current ATLAS $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction methods.

ATL-PHYS-PROC-2021-057
17 September 2021

PoS(EPS-HEP2021)625

**** The European Physical Society Conference on High Energy Physics (EPS-HEP2021), ****
**** 26-30 July 2021 ****
**** Online conference, jointly organized by Universität Hamburg and the research center DESY ****

*Speaker.

1. Introduction

Missing transverse momentum ($p_{\mathrm{T}}^{\mathrm{miss}}$) [1] is defined as the magnitude of the negative vectorial sum of the transverse momentum of all the visible objects produced in the hard scatter of a proton–proton collision event in the ATLAS detector [2] at the LHC. ATLAS employs several working points for $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, each of which is optimal for different event topologies and levels of pile-up. The working points considered here are: *Loose*, *Tight*, *Tighter* and *Tenacious*. These have increasingly strict selections on hadronic jets. The variable $p_{\mathrm{T}}^{\mathrm{miss}}$ significance [3] is also used to separate processes with 'real' $p_{\mathrm{T}}^{\mathrm{miss}}$ (from genuine invisible particles, such as neutrinos) and 'fake' $p_{\mathrm{T}}^{\mathrm{miss}}$ (from detector mis-measurement and pile-up).

This contribution presents the methods and results of training a neural network (NN) to pick and combine the $p_{\mathrm{T}}^{\mathrm{miss}}$ working points for each event to produce a new $p_{\mathrm{T}}^{\mathrm{miss}}$ definition ('METNet'), and also to define a machine learning-based $p_{\mathrm{T}}^{\mathrm{miss}}$ significance ('METNetSig'). Further details on this study can be found in Ref. [4].

2. Neural network architecture

The NN is a multi-layer perceptron trained on about three million $t\bar{t}$ and di-boson MC events. The NN predicts generator-level $p_{x,y}^{\mathrm{miss}}$ ($p_{x,y}^{\mathrm{miss,\,True}}$) given 60 input features including $p_{\mathrm{x,y,T}}^{\mathrm{miss}}$ predictions for the *Loose*, *Tight*, *Tighter* and *Tenacious* working points, plus additional information characterising pile-up and event topology, such as number of primary vertices and mean number of interactions per bunch crossing.

Results are shown for a NN trained using two loss functions:

1. $\mathcal{L} = \mathcal{L}_{\mathrm{Huber}} = \begin{cases} \frac{1}{2}(y-\hat{y})^2 & ,\, |y-\hat{y}| \le \delta \\ \delta|y-\hat{y}| - \frac{1}{2}\delta^2 & ,\, \text{otherwise} \end{cases}$
2. $\mathcal{L} = \mathcal{L}_{\mathrm{Huber}} + \mathcal{L}_{\mathrm{Sinkhorn}}$

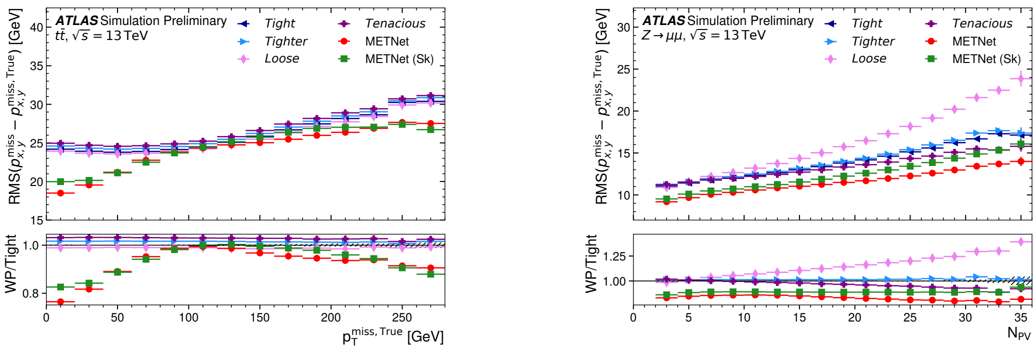
where y and $\hat{y}$ are the NN's prediction and target respectively and δ is set to $\frac{3}{2}$. Here $\mathcal{L}_{\mathrm{Sinkhorn}}$ is the Sinkhorn distance between the output and target batch which is included to reduce a bias in the distribution of the NN's predictions. Sample weights are also used to reduce a bias towards the bulk of the training set. The weight of each event is calculated using the reciprocal of the $p_{\mathrm{T}}^{\mathrm{miss,\,True}}$ histogram of the training set.

The training and testing sets are passed through two pre-processing steps. First, to remove ϕ invariance from the inputs each event is rotated such that $p_{\mathrm{T}}^{\mathrm{miss,\,Tight}}$ points along the x-axis. Then, each input/output feature is standardised by subtracting the mean and dividing by the standard deviation. The NN outputs are transformed by the inverse of these steps.

3. METNet: Resolution and bias

Figure 1 shows the root-mean square of the deviation from truth of the $p_{x,y}^{\mathrm{miss}}$ predictions (ie. the resolution) for the four working points and METNet both with and without the Sinkhorn loss. METNet has improved resolution compared to the $p_{\mathrm{T}}^{\mathrm{miss}}$ working points for both topologies. The performance for $Z \to \mu\mu$ events and stability against pile-up is particularly notable, as this is a topology which was not seen by the NN during training and which contains no real $p_{\mathrm{T}}^{\mathrm{miss}}$.

PoS(EPS-HEP2021)625



(a) $t\bar{t}$ **(b)** $Z \rightarrow \mu\mu$

Figure 1: Root-mean-square of the difference between the predicted value of $p_{x,y}^{\text{miss}}$ and $p_{x,y}^{\text{miss, True}}$ for (a) $t\bar{t}$ events in bins of $p_{\text{T}}^{\text{miss, True}}$, (b) $Z \rightarrow \mu\mu$ events in bins of number of primary vertices. METNet is shown with and without the Sinkhorn (Sk) loss, along with current $p_{\text{T}}^{\text{miss}}$ working points. The lower panel shows the ratio with respect to the *Tight* working point, and the hatched band indicates the statistical uncertainty for the *Tight* working point [4].

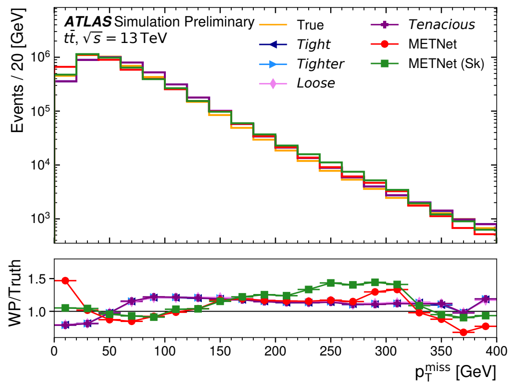


Figure 2: $p_{\text{T}}^{\text{miss}}$ distribution for METNet (with and without the Sinkhorn loss), $p_{\text{T}}^{\text{miss, True}}$ and four $p_{\text{T}}^{\text{miss}}$ working points, for $t\bar{t}$ events. Events are unweighted [4].

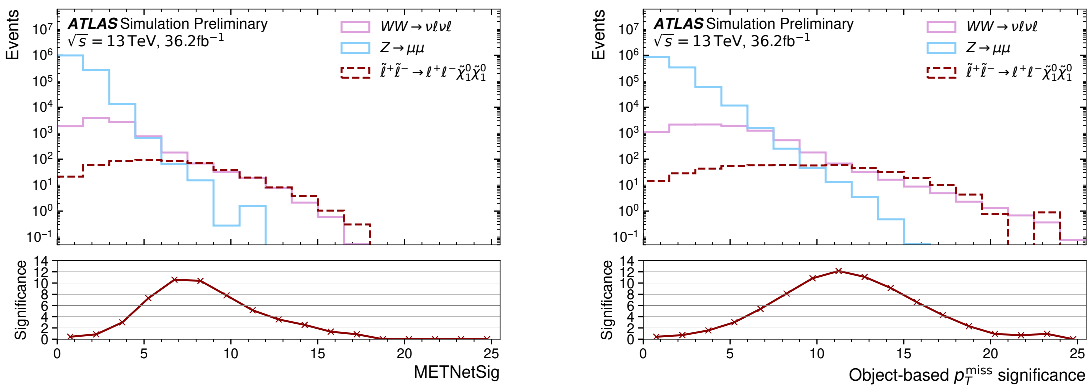
Figure 2 shows the $p_{\text{T}}^{\text{miss}}$ distribution for METNet (both with and without the Sinkhorn loss). Including the Sinkhorn contribution to the loss improves the agreement of the METNet distribution with $p_{\text{T}}^{\text{miss, True}}$ for $0 < p_{\text{T}}^{\text{miss}} < 100$ GeV.

4. METNetSig

The $p_{\text{T}}^{\text{miss}}$ significance is defined by weighting a $p_{\text{T}}^{\text{miss}}$ prediction by a 'confidence', σ. The state-of-the-art ATLAS implementation is an object-based $p_{\text{T}}^{\text{miss}}$ significance [3], for which σ is a function of the p_{T}-dependent resolutions of the objects which enter the $p_{\text{T}}^{\text{miss}}$ calculation. To define a machine-learning based $p_{\text{T}}^{\text{miss}}$ significance, the Huber loss function is replaced with the

Gaussian negative log likelihood (GNLL) loss, $\mathcal{L}_{\text{GNLL}} = \log \sigma_{x,y} + 0.5 \left(\frac{p_{x,y}^{\text{miss, NN}} - p_{x,y}^{\text{miss, True}}}{\sigma_{x,y}} \right)^2$. The resulting NN has four outputs, ($p_x^{\text{miss, NN}}$, σ_x, $p_y^{\text{miss, NN}}$, σ_y), from which METNetSig is defined as METNetSig= $p_{\text{T}}^{\text{miss, NN}}/\sigma$.

Figure 3 shows (a) METNetSig and (b) object-based $p_{\text{T}}^{\text{miss}}$ significance [3] for a supersymmetric signal process plus two Standard Model backgrounds. METNetSig has a similar shape to object-based $p_{\text{T}}^{\text{miss}}$ significance for each topology, but with a lower cut-off point. Note that the NN learns σ independently of the object-based measurements that are inputs to object-based $p_{\text{T}}^{\text{miss}}$ significance. Nonetheless, METNetSig has similar real $p_{\text{T}}^{\text{miss}}$ vs fake $p_{\text{T}}^{\text{miss}}$ separation power.



(a) (b)

Figure 3: Distributions of (a) METNetSig and (b) object-based $p_{\text{T}}^{\text{miss}}$ significance for a supersymmetric signal point and WW and $Z \to \mu\mu$ backgrounds. The lower plot shows the signal significance for a lower-bound cut at each x-axis bin value [4].

5. Conclusion

This contribution presents the performance of METNet and METNetSig - variables defined using the outputs of a neural network - compared to current ATLAS methods. METNet has significantly improved resolution for a range of topologies, including those not seen during training. Including a Sinkhorn contribution to the loss function reduces an observed negative bias in the NN's predictions. METNetSig shows similar behaviour per-topology to object-based $p_{\text{T}}^{\text{miss}}$significance, and can distinguish between real $p_{\text{T}}^{\text{miss}}$ and fake $p_{\text{T}}^{\text{miss}}$.

References

[1] ATLAS Collaboration, *Eur. Phys. J. C* **78** (2018) 903 [1802.08168].

[2] ATLAS Collaboration, *JINST* **3** (2008) S08003.

[3] ATLAS Collaboration, Tech. Rep. ATLAS-CONF-2018-038, CERN, Geneva (Jul, 2018).

[4] ATLAS Collaboration, Tech. Rep. ATL-PHYS-PUB-2021-025, CERN, Geneva (Jul, 2021).