DIPS flavour tagging algorithm for variable-radius track jets

Stefan Katsarov

University of Edinburgh

Supervisors: Dr. Paul Philipp Gadow, Dr. Krisztian Peters

September 7, 2022

Abstract

The performance of the DIPS flavour tagging algorithm applied to variable-radius track jets is studied. First, different trainings based on a hybrid sample of jets composed from either $t\bar{t} + Z'$ or $t\bar{t}$ + graviton events are evaluated. They are compared to a reference training based on particle flow jets. The results favour the $t\bar{t}$ + graviton training, which generalises best and outperforms the reference training. Second, an expanded DIPS network architecture is studied, using a L1 regularisation to overcome overtraining. This expanded model outperforms the baseline DIPS model.

Contents

1 Introduction

The identification of jets initiated by bottom-quarks (b-quarks), also referred to as b-tagging, plays a fundamental role within the physics program of the ATLAS experiment in the Large Hadron Collider (LHC).

B-tagging has become a decisive tool in key applications such as observations of Higgs bosons decaying into b-quarks ($H \rightarrow bb$) [1] or in the production of Higgs bosons in association with a pair of top quarks ($pp \rightarrow H + tt$) [2]. For the latter, more explicitly, with the top quarks almost exclusively decaying into W bosons and b-quarks which can then be tagged.

The identification of b-jets is a crucial tool used also in the more general ATLAS physics programme, including Standard Model (SM) [3] precision measurements, studies of Higgs boson and top quark properties [2], and searches for new exotic phenomena beyond the Standard Model (BSM). Out of the potential BSM extensions, b-jet identification is especially important for extensions of the SM with yet undiscovered resonances that preferentially decay into heavy quarks [4].

Therefore, with the extremely rare events involved, combined with the sensitive measurements that are required, it becomes imperative to ensure that b-jets are correctly classified in as many instances as possible and hence for the b-jet tagging algorithms to have optimal performance.

In this document, studies of the Deep Impact Parameter Sets-based algorithm (DIPS) [5] are presented and potential improvements to its network architecture are discussed. The DIPS algorithm is a multivariate machine learning algorithm which is trained and evaluated on Monte Carlo (MC) simulated datasets.

2 Aim and outline of the project

The aim of this project is to present a general overview of the b-jet flavour tagging process and to then carry out two key studies using the DIPS flavour tagging algorithm.

To begin, in section 3 the concept of jets is introduced with a discussion on flavour tagging, the flavour tagging algorithms used in ATLAS and the two main types of jet reconstruction. Section 4 details the DIPS algorithm, our specific implementation and the previously trained DIPS models, while section 5 describes the datasets used to train and evaluate the algorithms.

The first study is based on different trainings the DIPS algorithm on VR track jets. To set the foundation for this set of studies, in section 6 the DIPS model is optimised to ensure that accurate and reliable results can be provided in the later evaluations. Next, in section 7, the DIPS model is trained on the baseline $t\bar{t} + Z'$ variable-radius track jet sample and its performance is compared to the previous DIPS model trained on particle flow jets. This is followed by training the DIPS model on the new $t\bar{t}$+graviton sample

to test its applicability by evaluating its performance compared to the baseline $t\bar{t} + Z'$ sample.

The second study is based on evaluating the performance of a significantly expanded version of the baseline DIPS network. As part of this, the concept of overtraining is introduced, and methods of regularisation are proposed as solutions to this problem in section 8.1. Next, the performance of the expanded network is compared to the baseline network in section 8.2. In section 9.1 the methods of L1 and L2 regularisation are independently implemented on the extended network and their performance is evaluated in comparison with the default extended network. Following this, the best regularised model is chosen to be evaluated in comparison with the baseline DIPS model in section 9.2. Finally, the conclusions of our studies are summarised in section 10.

3 Jets

In proton-proton collisions, the scattering of high-energy partons can lead to the formation of different hadrons. Most frequently, mesons such as π, K, D, or B are formed [6]. In instances where the individual quark and anti-quark of a meson have momenta that are non-parallel, this results in a gradual increase in the separation between the two quarks. As the separation increases the energy in the gluon field becomes large enough to create another quark anti-quark pair [7]. This process can continue if the energy of the initiating meson is large enough. Furthermore, this final state can also be achieved though high-energy colour charged quarks emitting gluons which then split into quark anti-quark pairs. Both of these processes can then lead to the formation of a cascade of hadrons that corresponds to a jet.

With this we can therefore define a jet as a collection of collimated bunches of hadrons, with each bunch originating from a single quark or gluon.

3.1 What is flavour tagging?

In the ATLAS experiment, once jets are formed in collisions and are reconstructed, they then need to have the flavour of their initiating quark determined. This task is referred to as flavour tagging.

In the process of flavour tagging, the possible candidates which need to be considered are b-, c- and light-flavour (u, d, s) quarks, along with gluons. These candidates correspond to the jet flavour classifications of b-, c- and light-flavour jets respectively.

The properties of b-hadrons such as their large mass of $\sim$5 GeV and their characteristically long life-times of $\sim$1.5 ps, produce distinct decay characteristics that make them stand out from the background of c- and light-flavour jets [8]. Out of the b-hadron properties the long life-time creates the biggest distinction by producing a mean flight path of 450 µm which results in at least one displaced vertex from the collision point

[9]. Furthermore, the large mass of b-hadrons results in decay products that have large transverse momenta with respect to the jet axis.

Therefore, unlike c- and light-flavour hadrons, b-hadrons decay at secondary vertices which are the source of tracks with large impact parameters caused by a comparatively large transverse momentum relative to the jet axis.

This results in a wider jet cone which can be identified as the blue cone in Figure 1.

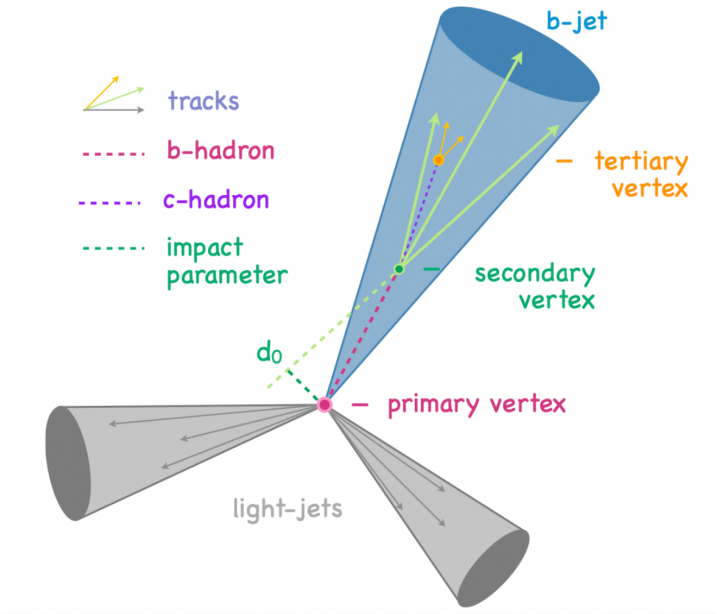


Figure 1: Characteristic properties of b-jets compared to background jets [10].

3.2 Jet flavour tagging algorithms in ATLAS

Multiple dedicated algorithms are involved in the process of jet flavour tagging.

First, a series of specialised low-level algorithms carry out physics-based feature extractions which corresponds to obtaining information about features such as track impact parameters and secondary vertices [9]. These algorithms are designed and fine-tuned with expert knowledge and are hence detector specific.

Features that are extracted by the low-level algorithms are then fed as input to high-level multivariate or machine learning algorithms which provide a jet flavour classification. These algorithms produce a final jet classification based on likelihood ratio estimation through the calculation of a discriminant using the three output jet probabilities [5]. There are multiple discriminants that can be used but for our analysis, but we have used the DL1 b-tagging discriminant which has been denoted as D_b and is in the form:

$$D_b = \log \frac{p_b}{(1 - f_c)\, p_l + f_c p_c} \tag{1}$$

where p_b, p_c and p_l are the probabilities of a jet being a b-, c- or light-jet respectively, and f_c is a free parameter that balances between the rejection of light- vs c-jets.

The flexible nature of the high-level algorithms allows them to be detector agnostic as they can generalise and be applied to any set of extracted input features.

3.3 Jets in ATLAS

To distinguish between the previous (baseline) set of particle flow jets and the new set of variable-radius track jets used for training, a distinction needs to be made between them.

3.3.1 Particle flow jets

Particle flow (PFlow) [11] jets are reconstructed by combining both inner detector and calorimeter cluster data. These jets are reconstructed by searching for tracks within a fixed cone radius of $R = 0.4$. Reconstruction of these jets is enhanced by extrapolating inner detector tracks to calorimeter regions and then subtracting the energy deposits of charged particles to leave the neutral energy deposits of the jets. This means that both charged and neutral components of the jet can be separately reconstructed as charged and neutral particle flow objects. This approach of jet reconstruction is very well suited for general purpose jet reconstruction and allows to produce the most complete representation of a jet that also has accurate energy calibration through the calorimeters.

However, through the involvement of the low-resolution information of the calorimeters, the overall jet resolution is greatly limited. For high-p_{T} decays, b-jets will be highly collimated, resulting in multiple jets having tracks within a cone of $R = 0.4$ and hence being reconstructed as the same jet.

3.3.2 Variable-radius track jets

On the other hand, variable-radius (VR) [12] track jets remedy this issue as they are reconstructed using only the inner silicon pixel detector hits which produces very high-resolution tracking information. Although energy calibration in this reconstruction process is very limited, this does not directly contribute to the ability of correct b-jet classification. With the high-resolution tracking information, the reconstructed tracks can then be clustered within an adaptable jet cone that has a radius as a function of the jet p_{T}.

The radius of the jet cone can therefore be expressed through the equation:

$$R \longrightarrow R_{\mathrm{eff}}\left(p_{\mathrm{T}}\right) = \frac{\rho}{p_{\mathrm{T}}} \tag{2}$$

where ρ is a tuneable parameter.

Therefore VR track jets can be used to greatly improve jet reconstruction over PFlow jets when reconstructing jets from boosted decays where both b-jets are highly collimated and inside a cone of $R = 0.4$.

4 The DIPS neural network

The DIPS architecture consists of two main components: a Φ and an F network which are connected by an intermediate function [5].

The Φ network operates in parallel on each track of a jet and its outputs are concatenated and summed over all the tracks through a permutation invariant sum operation. This allows the tracks to be treated as sets without any specific order, which is also physically motivated. The F network then takes these values as input and outputs the corresponding probabilities for b-, c- and light-flavour jets.

In Figure 2 below, the baseline configuration of the DIPS architecture is shown:

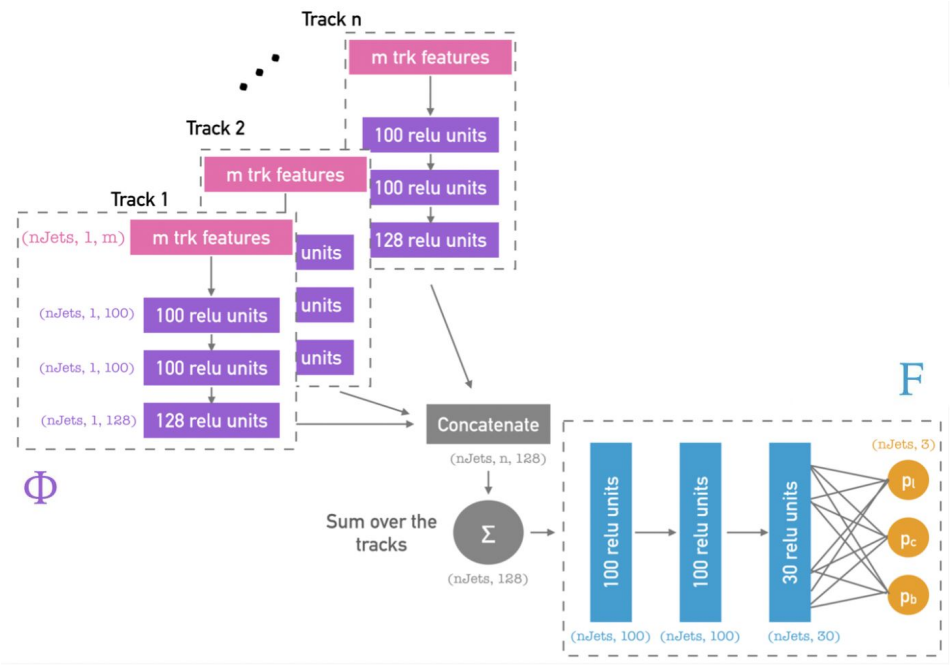


Figure 2: The Baseline DIPS architecture [5].

An abstract representation of the network can be expressed through the equation:

$$O\left(\{p_1, \ldots, p_n\}\right) = F\left(\sum_{i=1}^{n} \Phi\left(p_i\right)\right) \tag{3}$$

In this equation p_i are all the input tracks that belong to a given jet and the set O is comprised of the outputs for each jet containing the probabilities of being a b-, c- and light-jet.

The baseline DIPS architecture, as introduced here, consists of a Φ network with 3 layers corresponding to [128,100,100] nodes and an F network also with 3 layers corresponding to [100,100,100] nodes. Each layer in both the Φ and F network uses the RELU activation function.

Later some modified versions of the DIPS architecture will also be displayed which have been trained and evaluated for the purpose of optimisation.

4.1 Previous versions of DIPS trained on PFlow jets

There are already multiple versions of the DIPS algorithm trained on PFlow jets prior to this analysis. The most relevant ones that have been used for comparison are the DIPS loose (DIPS_L) and DIPS loose rescaled (DIPS_L_RS) versions. The loose variant of the DIPS algorithm has the best performance on PFlow jets out of all the DIPS models. This algorithm has increased d_0 and z_0 impact parameter cut-offs compared to the original DIPS and has a minimum track p_{T} acceptance of 0.5 GeV. In order to be applied on VR track jet inputs, the DIPS loose algorithm was adapted by rescaling the PFlow $p_{\mathrm{T}}^{\mathrm{frac}}$ track input variable in training to mimic VR track jet inputs. This adapted version of the DIPS loose algorithm corresponds to the DIPS loose rescaled algorithm and has the best DIPS performance on VR track jets from the algorithms trained on PFlow jets.

5 Training and Evaluation datasets

Two different datasets were used for training and evaluation.

The first set was comprised of a $t\bar{t} + Z'$ hybrid sample and the second of a $t\bar{t}$+graviton hybrid sample. The $t\bar{t} + Z'$ dataset is the baseline dataset that is routinely used as the standard when analysing the performance of the DIPS algorithm. In this study we also include training and evaluation done on a $t\bar{t}$+graviton dataset to examine its general usefulness and compare its generalisability in performance over the $t\bar{t} + Z'$ baseline dataset.

The reason a hybrid sample is used for both datasets is because there is an insufficient number of high-p_{T} jets in the $t\bar{t}$ sample, therefore a BSM sample of Z' or graviton events is needed as a source of high-p_{T} jets.

5.1 Training datasets

During training process, the training set was split into a training and validation set in the ratio 4:1. A validation set is used to monitor the performance of the algorithm with a statistically independent sample. The performance on the validation set can then be used gauge how well the trained model generalises to unseen data and whether or not it is overtraining.

5.1.1 $t\bar{t} + Z'$ hybrid dataset

The $t\bar{t}+Z'$ hybrid sample is made up of 5 million jets, where jets from $t\bar{t}$ events cover the low-p_{T} range while jets from Z' events cover the high-p_{T} range. The $t\bar{t}$ and Z' datasets were downsampled in count, in p_{T} and η together, to ensure that all jet flavours are present in equal proportions and to have a smooth transition between the two samples.

The results of the downsampling can be seen in Figure 3.

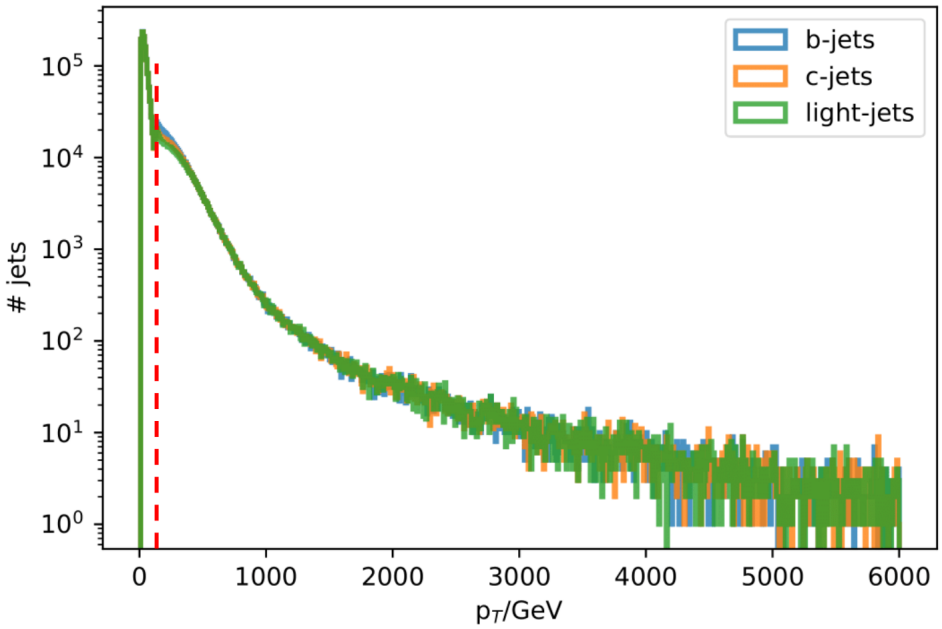


Figure 3: The result of downsampling on the jet p_{T} distribution. The red dashed line indicates the transition from low- to high-p_{T}. On the left side of the red line are the mostly low-p_{T} $t\bar{t}$ events while on the right side are mostly the high-p_{T} hypothetical Z' events.

The features of the combined samples were then standardized to having zero mean and unit variance to improve their handling by the neural network. Finally, all of the jets were shuffled.

5.1.2 $t\bar{t}$+graviton hybrid dataset

For the case of the $t\bar{t}$+graviton hybrid sample 5 million jets were also used to allow for a direct comparison with the $t\bar{t} + Z'$ sample. The $t\bar{t}$+graviton sample was pre-processed in the same way as the $t\bar{t} + Z'$ except with the graviton sample replacing the Z' for high-p_{T} ranges.

5.2 Evaluation datasets

To evaluate the performance of the trained DIPS models, we used three different pure test samples. The $t\bar{t}$ and graviton samples were the biggest containing 4 million jets each, while the Z' sample was smaller and contained only 1.5 million jets. All of the samples only contained jets with $p_{\mathrm{T}} > 10$ GeV and $\eta < 2.5$.

6 Optimisation of the Baseline DIPS architecture

Some preliminary modifications to the baseline DIPS model were attempted with the goal of improving the network performance and probing the way the network is interacting with the current datasets.

6.1 Training of DIPS on VR track jets

The baseline DIPS model (model 0) was trained multiple times with different network parameters, by mainly focusing on the number of layers in the F network and the batch size. The cross-entropy loss was used as the objective function in training and the validation set loss was used as the key metric to gauge performance. Modifications of the DIPS models showed that the baseline network was already close to the optimal size for the current training sample size (with all modifications giving little improvement) and that regularisation was not beneficial for this model. The only type of regularisation employed in the training was saving the weights of the model which gave the lowest validation loss and ignoring further changes to the weights in continued training. From the parameter optimisation the outcome was that that addition of a single layer in the F network produced the largest reduction in validation loss and the highest validation set accuracy. The network which produced these results corresponded to model 3 and this was set as the new baseline model to be used in the evaluation of the two training sets.

	phi_layers	f_layers	batch_size	loss	acc	val_loss	val_acc	epochs	best_epoch	time_to_best_epoch(mins)	total_training_time(mins)
0	[100, 100, 128]	[100, 100, 100]	1500	0.649946	0.702599	0.664470	0.695118	100	63	38.000000	60.000000
1	[100, 100, 128]	[100, 100, 100]	5000	0.653052	0.701137	0.662369	0.696558	39	34	13.030000	14.950000
2	[100, 100, 128]	[100, 100, 100, 100]	1500	0.652393	0.701346	0.659515	0.697844	27	22	13.492311	16.446092
3	[100, 100, 128]	[100, 100, 100, 100]	1500	0.646824	0.704083	0.656953	0.699167	50	46	27.576576	30.013429
4	[100, 100, 128]	[100, 100, 100, 100, 100, 100]	1500	0.646689	0.704108	0.657612	0.699075	50	39	25.757817	32.754742
5	[100, 100, 128]	[100, 100, 100, 100]	1500	0.640912	0.707719	0.657818	0.699129	100	57	33.648522	59.220437

Figure 4: The DIPS networks with modified parameters and their performance.

7 Evaluating new Baseline DIPS trained on the $t\bar{t} + Z'$ and $t\bar{t}$+graviton hybrid datasets

In this section are the combined results that were obtained by training the new baseline DIPS model on each of the hybrid datasets and then evaluating the trained models on the three pure evaluating samples.

The plots that are used to evaluate the performance of the trained DIPS models are known as receiver operating characteristic (ROC) curves.

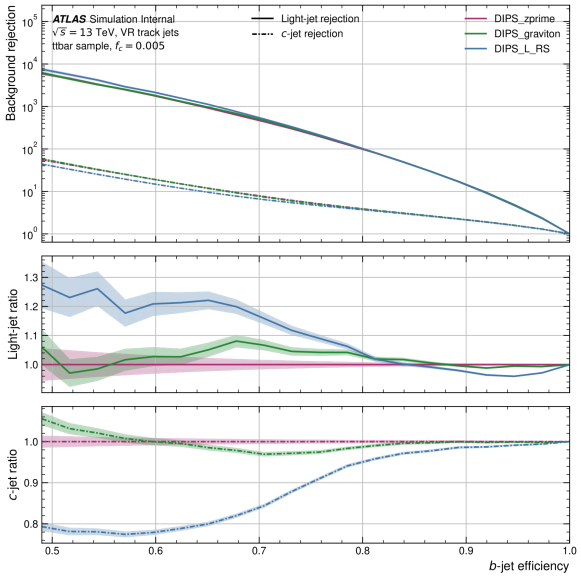
7.1 Evaluation metrics

For ROC curves the x-axis is the fraction of b-jets that have been correctly classified, which is referred to as the b-jet efficiency, and the y-axis is the inverse of the fraction of light- or c- jets that have been incorrectly classified as b-jets, which is referred to as the background rejection.

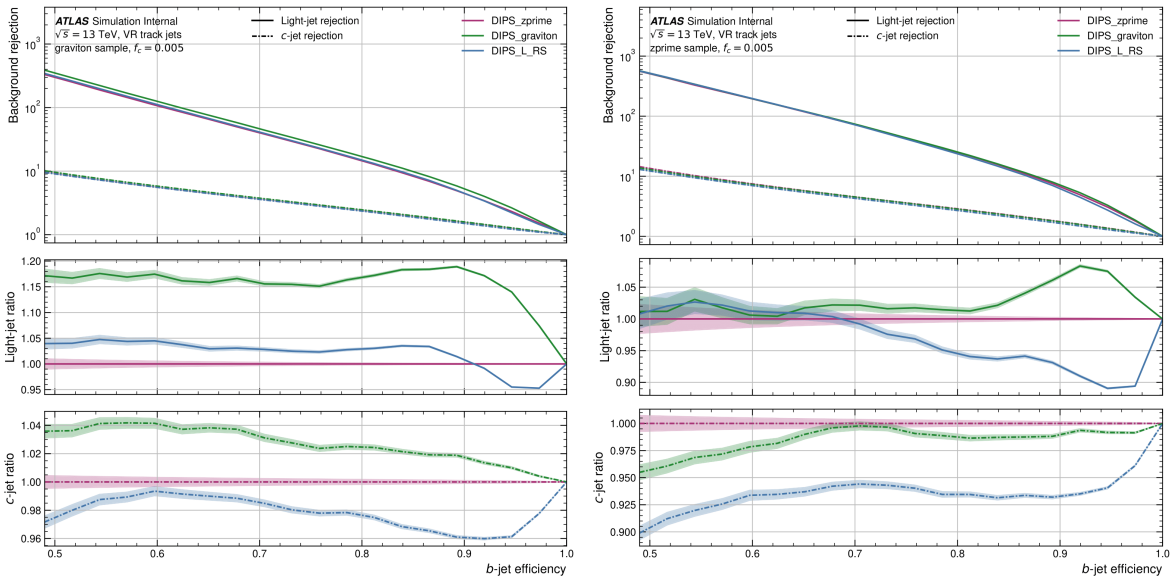
In general, it is desirable for the light- and c-jet rejection to remain high with increasing b-jet efficiency, which means for the ROC curve to be pushed as far as possible to the upper right-hand corner of the plot.

7.2 Results

Figure 5 shows the results that were obtained from training the new baseline DIPS model on the two hybrid datasets and then evaluating the trained models on each of the testing sets.



(a)



(b) (c)

Figure 5: The results obtained from training the new baseline DIPS model on the two hybrid datasets and evaluating the trained models on the three testing sets. The DIPS_L_RS model is used as a baseline comparison.

In these plots the DIPS_zprime curve is the DIPS model[1] trained on the $t\bar{t} + Z'$ hybrid

[1]From this point onwards, referring to the DIPS model now corresponds to the new baseline DIPS model.

sample, the DIPS_graviton curve is the DIPS model trained on the $t\bar{t}$+graviton hybrid sample and the DIPS_L_RS curve is the already trained version of DIPS on rescaled PFlow track features.

From these ROC curves it is evident that the DIPS_graviton model outperforms the DIPS_zprime model in light-jet rejections across all the three testing samples, with this being slightly less apparent in the $t\bar{t}$ sample. In terms of the c-jet rejection for the DIPS_zprime and DIPS_graviton models, there are no clear-cut results for the $t\bar{t}$ and Z' sample as the models behave similarly. On the other hand, on the graviton sample the DIPS_graviton model shows significantly c- and light-jet rejection than the DIPS_zprime model.

In general, from these results it can be established that both VR DIPS models are consistently outperforming the DIPS_L_RS model is c-jet rejection throughout the three testing samples. However, it is clear also that the VR DIPS models are only outperforming DIPS_L_RS in the Z' and graviton light-jet rejection, with DIPS_L_RS showing considerably better $t\bar{t}$ light-jet rejection.

Therefore, what can be taken away from this evaluation is that the DIPS_graviton model can generalise better on the Z' sample than the DIPS_zprime model can on the graviton sample. This clarifies that the $t\bar{t}$+graviton hybrid training sample can be used to produce coherent and reliable results. Furthermore, these results show that the DIPS model trained on the graviton hybrid sample identifies b-jets with less discrimination to their source of origin compared to the DIPS model trained on the Z' hybrid sample.

All of this stands as strong motivation for further studies on the applicability of using the $t\bar{t}$+graviton hybrid sample over the baseline $t\bar{t} + Z'$ hybrid sample in training of the DIPS algorithm for b-jet flavour tagging.

8 Expanded DIPS architecture

Following our analysis between the performance of the $t\bar{t} + Z'$ and $t\bar{t}$+graviton hybrid training samples, further attempts in improving the DIPS algorithm were made. Given that the fine tuning of parameters in section 6 did not result in significant improvements to the network, the completely opposite approach was then taken. Instead, a full-scale restructuring of the network architecture was carried out by substantially increasing the number of layers and nodes per layer in the baseline DIPS network.

To elucidate the changes that were made to the network, a comparison is shown between the baseline network architecture and the new expanded network architecture in Figure 6 below.

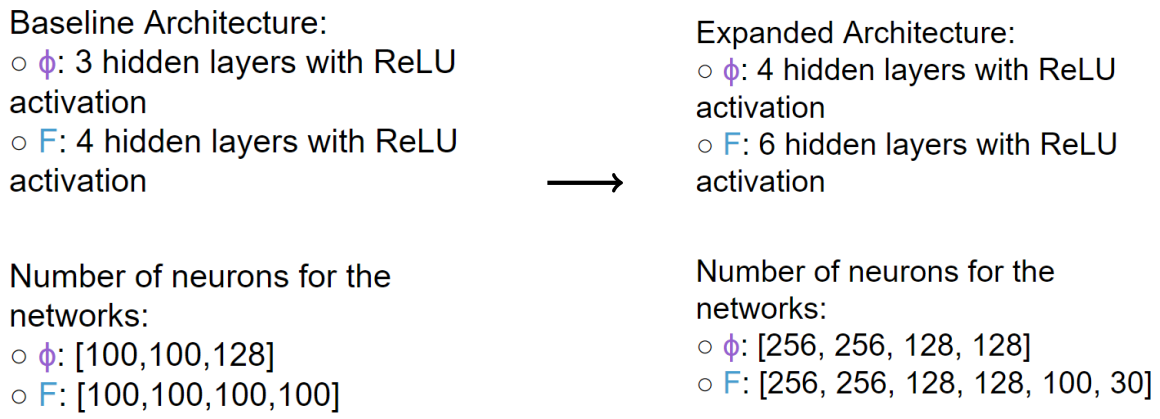


Figure 6: The baseline versus extended DIPS network architecture parameters.

To summarise the changes: the Φ network layers were increased from 3 to 4 and the F network layers from 4 to 6, along with this, the number of neurons in each layer were also increased.

Finally, another important difference to be acknowledged is that the expanded DIPS model has layers of decreasing nodes when going in the direction of input to output nodes.

8.1 Over-training and Regularisation

However, increasing the size of the DIPS architecture was accompanied by undesirable consequences.

By expanding the DIPS model this increases its ability to create a more complex fit of the training data. For the limited training data that is available, this results in overfitting the training data and hence overtraining the model. This happens through the use of the extra node weights as storage of the specific features present in the training dataset, which improves training loss, but without this having generalisability to the validation or testing sets [13]. The effect of overtraining can be seen in Figure 7 below through the convex shape of the validation loss curve compared to the monotonically decreasing training loss curve. This indicates that after a certain amount of training the model becomes worse rather than better.

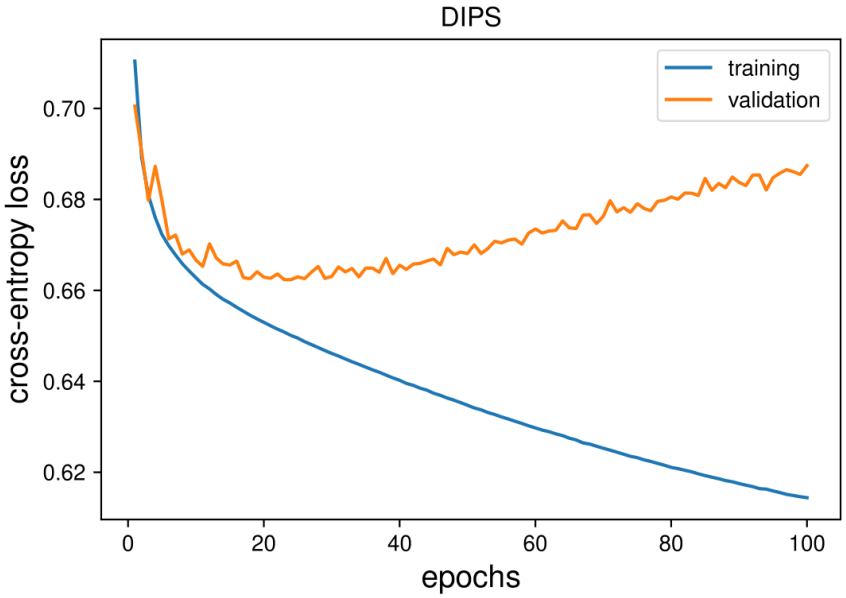


Figure 7: Cross-entropy loss functions for the training and validation datasets showing overtraining taking place through the convex shape of the validation loss curve compared to the monotonically decreasing training loss curve.

A good solution to remedy overtraining is through implementing regularisation. Although there are many methods of regularisation, in this analysis the focus will specifically be placed on L1 and L2 regularisation. These types of regularisation act by constraining the size of node weights within the neural network to reduce biased weight distributions and hence over training [13, 14]. This regularisation is implemented through the addition of a weight penalty term to the loss function so that when minimising the loss function the penalty term is also taken into account.

Looking more closely, L1 regularisation corresponds to LASSO[2] regression and has a penalty term in the form [14]:

$$+\lambda \sum_{i=1}^{N} |\theta_i| \,, \tag{4}$$

where λ is a tuneable parameter and θ_i are the node weights.

This regularisation results in a form of feature selection, whereby the weight of less important/redundant features can become zero during the regularised training and hence for these features to be completely ignored. Furthermore, L1 regularisation benefits from being more robust to outliers due to the linear form of the penalty term.

[2]Least absolute shrinkage and selection operator

On the other hand, L2 regularisation corresponds to ridge regression and behaves as a form of feature prioritisation by having a "softer" penalty term in the form:

$$+\lambda \sum_{i=1}^{N} \theta_i^2, \tag{5}$$

where λ is a tuneable parameter and θ_i are the node weights, once again.

This means that during regularised training with L2, the weights of important features are increased while the weights of less important features are decreased, but never to zero. Additionally, L2 regularisation is less computationally expensive than L1 but is not robust to outliers.

The effects of regularisation can be seen in Figure 8 through the closer relationship between the training and validation loss curves, where the validation loss plateaus with increasing number of epochs instead of being "U" shaped.

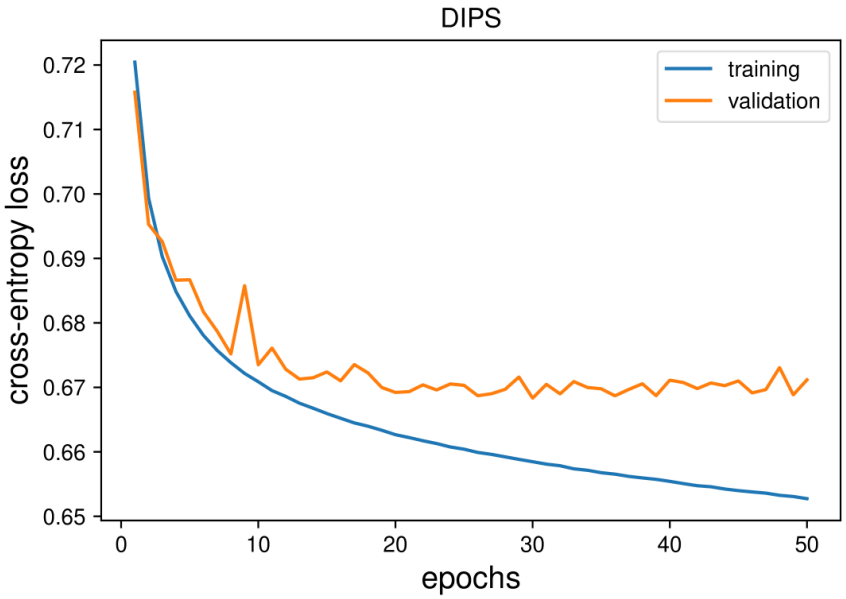
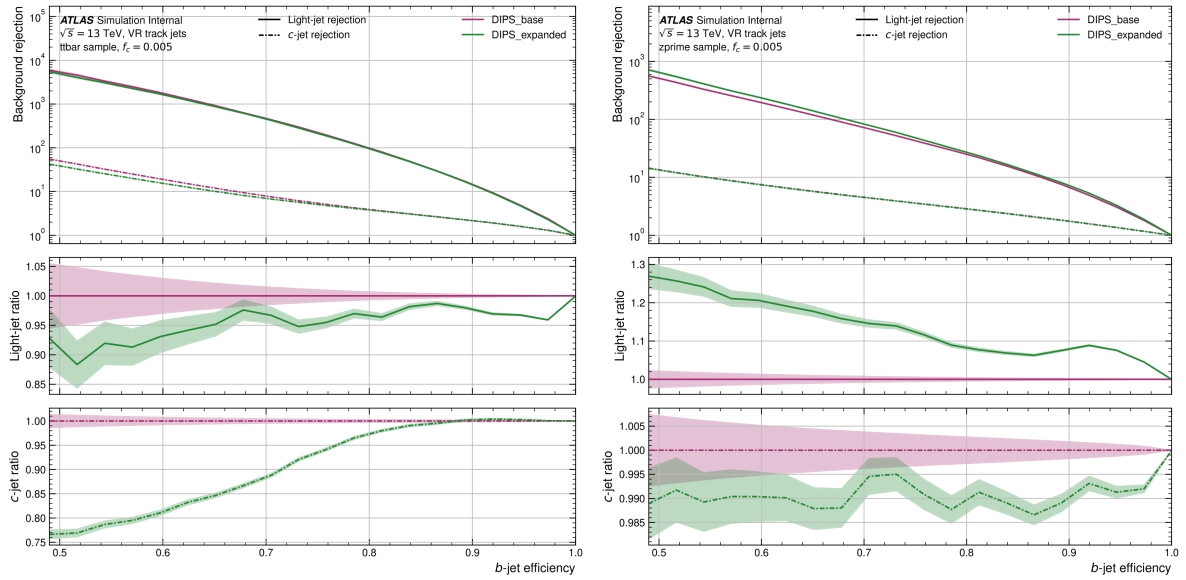


Figure 8: Cross-entropy loss functions for the training and validation datasets showing a regularised training where the validation loss more closely follows the training loss curve.

8.2 Baseline vs Expanded DIPS

The extended DIPS model was trained and evaluated on the $t\bar{t} + Z'$ hybrid sample and was then compared to the baseline DIPS model that is used as a reference. In the following evaluation only the $t\bar{t} + Z'$ sample will be considered as it is the established

baseline training and evaluation sample. The results of the evaluation can be seen in Figure 9 below.



(a) (b)

Figure 9: The results obtained by training and evaluating the expanded DIPS model on the $t\bar{t} + Z'$ hybrid sample, compared with the performance of the baseline DIPS model.

From these results it is clear that the extended DIPS underperforms in all ROC curves except in the light-jet rejection of the Z' samples. This reduction in performance is a direct consequence of overfitting to the training data. In this case the algorithm has become too focused on the high-p_{T} light-jet rejection.

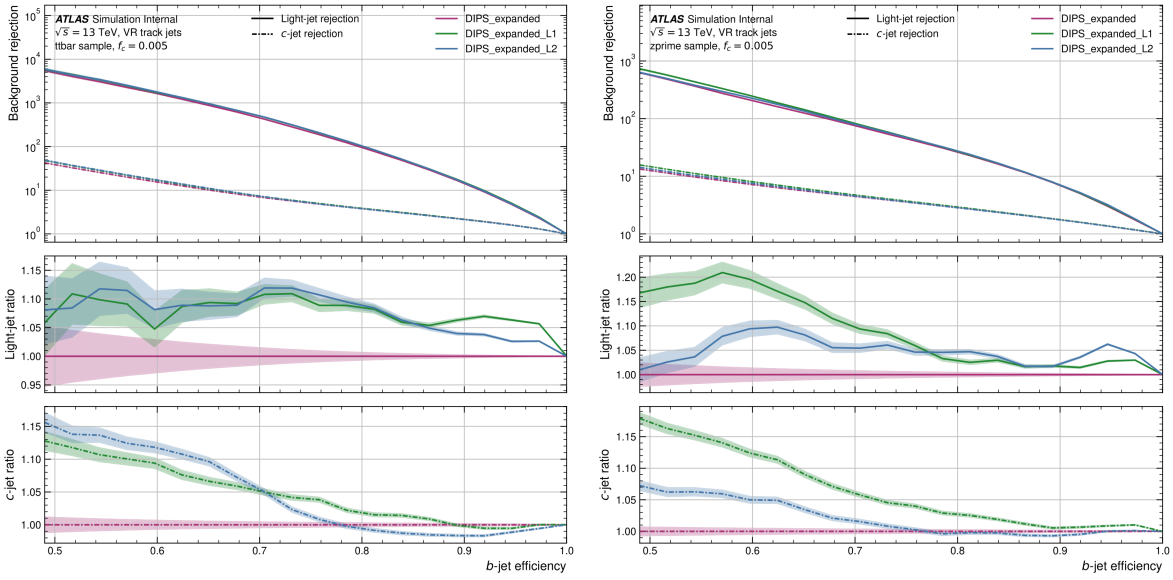
Finally, these results show that the extended DIPS has much worse overall light- and c-jet rejection on the $t\bar{t}$ sample compared to the Z' sample. This further indicates that there is a general bias towards the high-p_{T} jets from the Z' sample.

9 Evaluating Regularised DIPS

9.1 Expanded vs Regularised DIPS

L1 and L2 regularisation were independently implemented to the training of the extended DIPS network once it was determined to be overtraining. A lambda value of 0.00001 was used for the L1 and L2 penalty terms during the regularised training. This lambda value was found to give the best results as it allowed the regularisation to reduce over-fitting without overly restricting the networks ability to learn new features. The performance

of the regularised DIPS models against the extended DIPS model can be seen in Figure 10 below.



(a) (b)

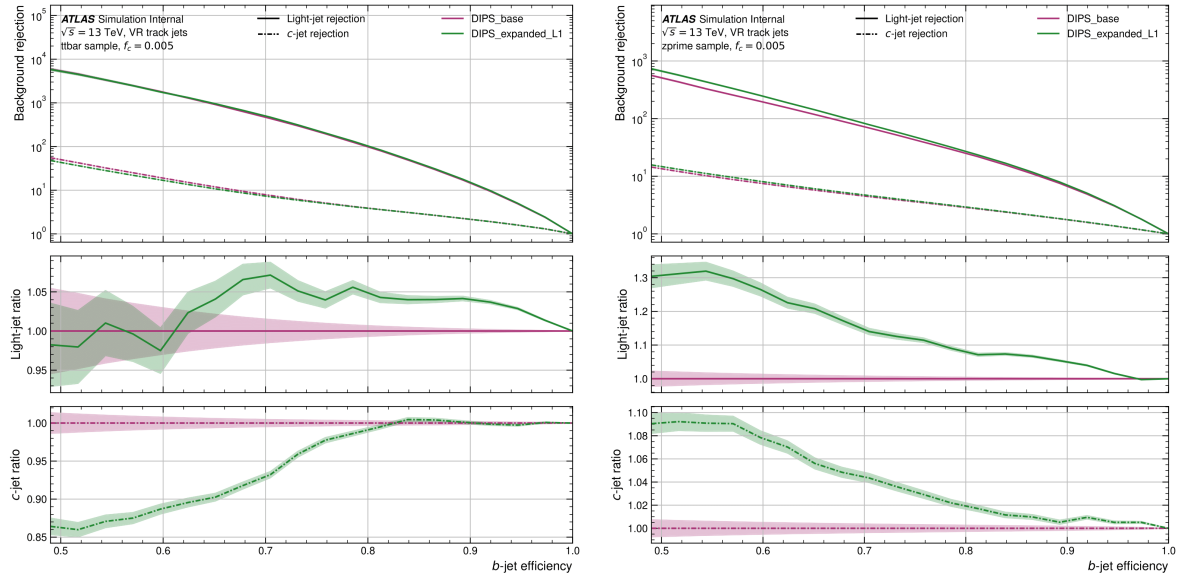
Figure 10: The results obtained through the training and evaluation of the L1 and L2 regularised expanded DIPS models on the $t\bar{t} + Z'$ hybrid sample, compared to the performance of the default expanded DIPS model.

The results that were obtained show a consistent and substantial increase in performance across all of the ROC curves for both testing samples. These plots show that L1 regularisation produces similar results to L2 regularisation on the $t\bar{t}$ sample, whereas on the Z' sample, L1 regularisation gives better results over a larger b-efficiency range. Furthermore, these results show that L1 regularisation produces more consistent light- and c-jet rejections compared to the trends observed in L2 regularisation.

Given this analysis, it can be concluded that the extended DIPS model with L1 regularisation has the best performance out of the extended DIPS models. This improved model can then be directly evaluated together with the baseline DIPS model to compare how it performs.

9.2 Baseline vs Regularised DIPS

The extended DIPS model with L1 regularisation was evaluated and directly compared with the baseline DIPS performance. These results are presented in Figure 11 below.



(a) (b)

Figure 11: The comparison of the results of obtained with the L1 regularised expanded DIPS model versus the baseline DIPS model. Both models are only trained and evaluated on the $t\bar{t} + Z'$ hybrid sample.

The extended DIPS model now performs considerably better on the Z' sample, producing significantly increased background rejections compared to the baseline DIPS model. Performance improvements extend also to the low-p_{T} jets, corresponding to the $t\bar{t}$ sample, where light-jet rejection is now higher for the extended DIPS model. However, the bias towards the Z' sample is still apparent with $t\bar{t}$ c-jet rejection remaining notably worse.

In general these results provide good motivation for additional optimisation of the regularisation strength hyperparameter, λ, with the goal of further increasing the $t\bar{t}$ c-jet rejections of the regularised extended DIPS models.

10 Conclusions

With the results that have been obtained in this analysis, it can be concluded that the DIPS algorithm trained on VR tracks jets have noticeably improved performance over the previous DIPS algorithms trained on PFlow jets in the majority of the cases. This shows strong evidence that even the unoptimized VR DIPS network will outperform the previously trained DIPS model given that it had the same amount of training samples.

Furthermore, in this analysis it has been established that the DIPS model trained on the $t\bar{t}$+graviton hybrid sample generalises better than the model trained on the $t\bar{t} + Z'$ sample. Therefore, the $t\bar{t}$+graviton sample looks more promising for better results.

This stands as strong motivation for further studies on the applicability of using the $t\bar{t}$+graviton hybrid sample over the baseline $t\bar{t} + Z'$ hybrid sample.

As part of this analysis a significantly expanded version of the baseline DIPS algorithm was implemented which was evaluated in comparison with the baseline DIPS model. The expanded DIPS model was too complex for the limited training data and was found to be significantly overtraining. However, when coupled with L1 regularisation the extended DIPS model was able to considerably outperform the baseline DIPS model in all background rejection curves except in the c-jet rejection for low-p_{T} jets.

With this, the results obtained with the extended DIPS network provide a strong argument for additional regularisation hyperparameter optimisation for achieving increased performance in future studies.

References

[1] ATLAS Collaboration, 2018. Observation of H→b$\bar{\mathrm{b}}$ decays and VH production with the ATLAS detector. *Physics Letters B*, 786, pp.59-86.

[2] ATLAS Collaboration, 2018. Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector. *Physics Letters B*, 784, pp.173-191.

[3] Cottingham, W. and Greenwood, D., 2010. *An introduction to the standard model of particle physics*. Cambridge: Cambridge University Press.

[4] ATLAS Collaboration, 2020. Search for new resonances in mass distributions of jet pairs using 139 fb^{-1} of pp collisions at $\sqrt{\mathrm{s}}$ = 13 TeV with the ATLAS detector. *Journal of High Energy Physics*, 2020(3).

[5] ATLAS Collaboration, 2020. *Deep Sets based Neural Networks for Impact Parameter Flavour Tagging in ATLAS*. ATL-PHYS-PUB-2020-014

[6] Webber, B., 2011. Parton shower Monte Carlo event generators. *Scholarpedia*, 6(12):10662.

[7] Kar, D., 2019. *Experimental Particle Physics*. IOP Publishing Ltd.

[8] P.A. Zyla et al. (Particle Data Group), *Prog. Theor. Exp. Phys.* 2020, 083C01 (2020).

[9] ATLAS Collaboration, 2019. ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{\mathrm{s}} = 13$ TeV. *The European Physical Journal C*, 79(11).

[10] Bartosik, N., 2016. File:B-tagging diagram.png - Wikimedia Commons. [online] Commons.wikimedia.org. Available at:

https://commons.wikimedia.org/wiki/File:B-tagging_diagram.png [Accessed 7 September 2022].

[11] ATLAS Collaboration, 2017. Jet reconstruction and performance using particle flow with the ATLAS Detector. *The European Physical Journal C*, 77(7).

[12] ATLAS Collaboration, 2017. *Variable Radius, Exclusive-kT, and Center-of-Mass Subjet Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS*. ATL-PHYS-PUB-2017-010

[13] Erdmann, M., Glombitza, J., Kasieczka, G. and Klemradt, U., 2021. *Deep Learning For Physics Research*. World Scientific.

[14] Pykes, K., 2022. Fighting Overfitting With L1 or L2 Regularization: Which One Is Better? - neptune.ai. [online] neptune.ai. Available at: *https://neptune.ai/blog/fighting-overfitting-with-l1-or-l2-regularization* [Accessed 7 September 2022].