Tagging Leptonically Decaying Boosted Top Quarks Using Residual Networks

DESY Summer Student Programme, 2021

The University of Manchester

Maddalena Mariano

Supervisors:
Soham Bhattacharya, Isabell Melzer-Pellmann

September 8, 2021

Abstract

The aim of this analysis is to study Machine Learning (ML) techniques to perform tagging of boosted top quarks decaying leptonically. In particular, to test the performance of the Residual Convolutional Neural Network ResNet50. The network successfully distinguishes the signal from the background, especially from the QCD background, with background efficiencies at 60% signal efficiency in the order of 10^{-2} for the hadronic top background and in the order of 10^{-3} for the QCD background.

Contents

1 Introduction

The top quark is the heaviest particle in the Standard Model and is important to study for numerous reasons, such as its role in the Higgs mass correction self-coupling loop [1]. Studying the production of top-antitop pairs could shed light on new physics, as some Beyond the Standard Model (BSM) processes with energies in the order of a TeV have boosted top quarks as the final states. Additionally, lower energy phase spaces for those processes have been excluded by ATLAS and CMS searches. Hence the importance of developing tools and techniques to tag such interactions, in particular boosted top quark interactions. At very high energy, these boosted top quarks decay into one single very collimated jet, called a fatjet, so it's important for those techniques to be able to tag boosted top jets as a single fatjet. Hadronically decaying boosted top quarks have been widely studied using various ML techniques, while there is still much demand for techniques that are effective on leptonic boosted top quarks.

2 Methodology

2.1 Jet image formation

The network uses jet images as inputs. The images are histograms of size 50x50 pixels, with the color axis representing the fraction of jet energy of each constituent. First the jets are formed by clustering its constituent using a jet clustering algorithm. In this case, the algorithm used is the sequential jet clustering algorithm Anti-kt [3], with radius parameter set to 1.5. The radius parameter describes the size of the fatjet, and is determined by the rapidities of the jet components, $\Delta\phi$. The jets are then rescaled and Lorentz boosted. All the constituent four-momenta are rescaled such that the mass of the jet is given by a constant m_0, and then a Lorentz boost is performed such that, in the new frame of reference, all jets have a constant energy. E_0 [2]. A Gram-Schmidt transformation is then applied so that the plane of the image is perpendicular to the jet axis, and the two subjets with the highest energy lie along the x-axis of the image plane [4].
For heavily boosted top quarks, the subjects in the image have lower resolutions, and the jet constituent are not distinguishable in the image. The preprocessing helps mitigate this effect and keep the subjects resolved independently from the top's boost in the lab frame. This makes results reliable in a wide range of energies, compared to the usually applied $p_T - \eta$ rotation, as seen in figure 1.

2.2 Convolutional Neural Networks

Convolutional neural networks make use of the convolution to process an input image and classify it. A basic CNN is composed of convolutional layers, pooling layers and fully connected layers. Convolutional layers apply cross correlation between sections of the input and a smaller matrix called "kernel", and do it across the whole input image at intervals defined by the stride of the convolution. Pooling layers, also called subsampling or downsampling layers, reduce the size of the feature map: for example, an average pooling layer will take in input 9 pixels and return the average as a single output. Fully connected layers are the same as in a regular perceptron network, and are used for classification once all the convolution and pooling layers have been applied.

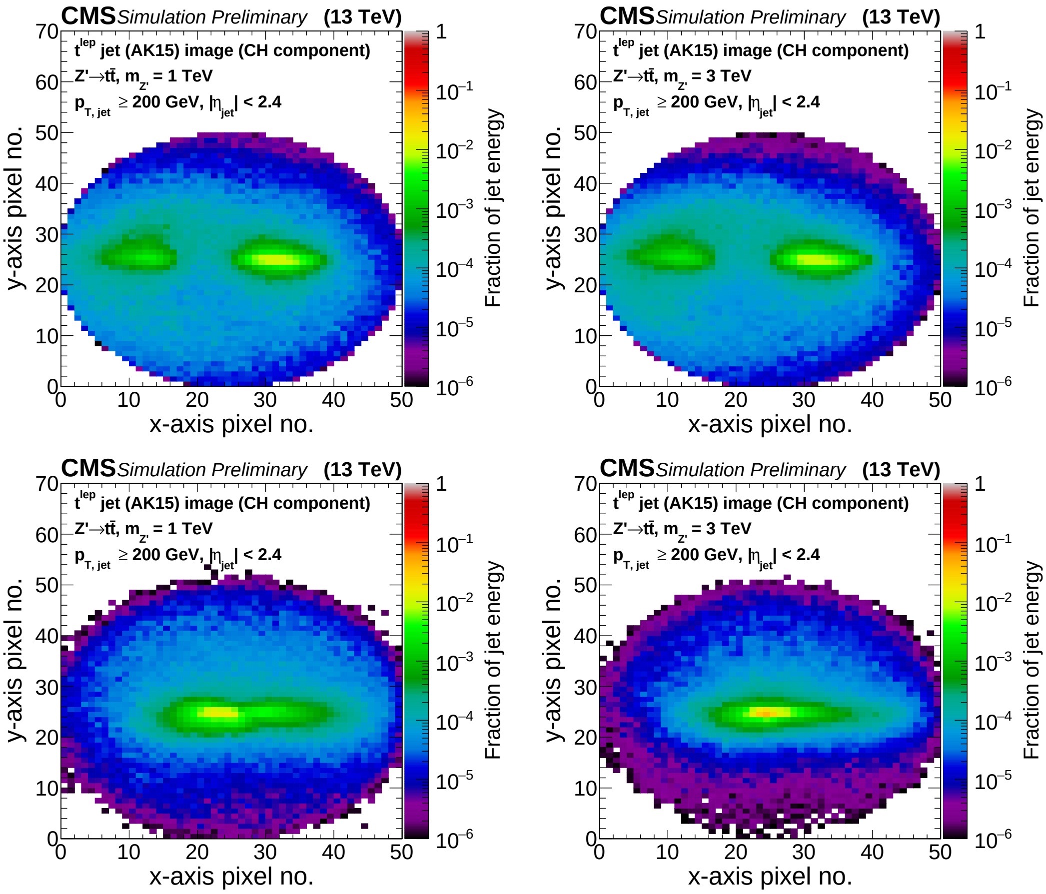

Figure 1: Jet images. At the top, jet images after the process described in the Methodology section. At the bottom, jet images after a $p_T - \eta$ only.

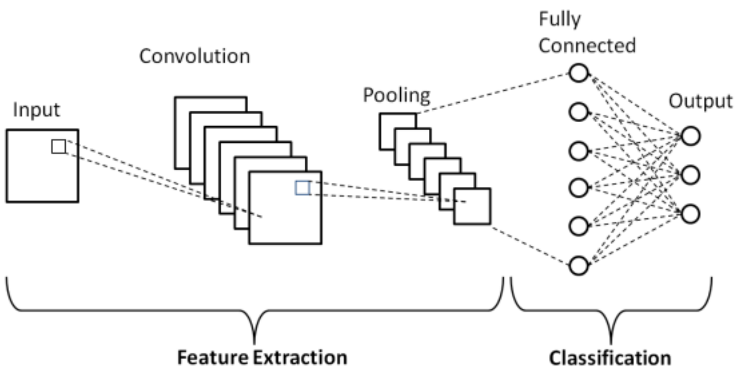

Figure 2: [7] Schematic diagram of a simple Convolutional neural network

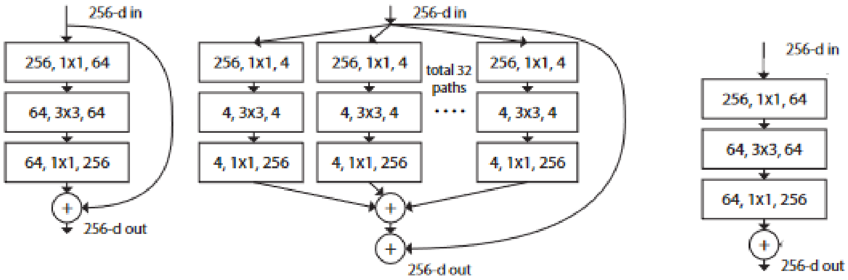

Figure 3: [6] Left: ResNet50 residual block. Center: ResNeXt50 residual block. Right: Regular (non-residual) CNN block.

2.3 Residual Networks

Residual neural networks include skip connections in their architecture. These connections bypass the deeper layer and the output of such layer is added to the output of the previous one. In networks that employ gradient-based learning, the skip connection allows the gradient information to be preserved through the layer, solving the vanishing gradient problem that deep networks tend to have [5]. The layers between two nodes of a skip connection are called, a residual block.

2.4 Training the networks

The network used for the analysis is a ResNet50 residual network, adapted for a problem with 3 categories and 1.5M training parameters. The original network was designed for much larger problems. The network and training were implemented in Python using Tensorflow. The output is a multiclassifier with 3 nodes, one for each category of jets. The training sample consists in three different jet categories: leptonically decaying top quarks, hadronically decaying top quarks and other QCD events (light quarks and gluons). The signal and background events were, respectively, approximately 2M and 6M. The ResNet50 network was trained over a number of epochs between 20 and 60, with learning rates ranging from 10^{-4} to 10^{-8}. An epoch is defined as a complete cycle through the training dataset. Each configuration of learning rates has been tested both including and excluding secondary vertex layers, to test any difference in performance between the two approaches. Those trainings will respectively be referred to as "All Layers" and "Energy Fraction" trainings.

3 Results

The most successful run, the one that reached the best epoch accuracy, had learning rates of 10^{-5} for epochs 0 to 29, 10^{-6} 30 to 59, and 10^{-7} epoch 60 and over. The results are taken from such run. Both in the training with all layers and the one with energy fraction layers only, the epoch accuracy starts reaching a plateau around epoch 25, so all the relevant values and results have been measured at epoch 25. In the epoch accuracy graph, the Energy Fraction and All Layers training performed very similarly, both reaching accuracies of 0.80 for the validation and 0.81 for the training, as seen in figure 4.
However, using all the layers leads to better ratios between signal and background efficiency,

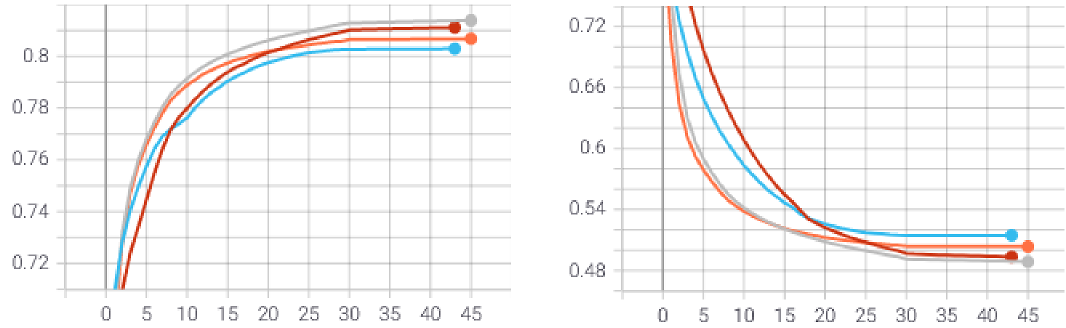

Figure 4: Epoch accuracy (left) and loss (right). Red: All Layers train. Blue: All Layers validation. Grey: energy fraction train. Orange: Energy Fraction validation.

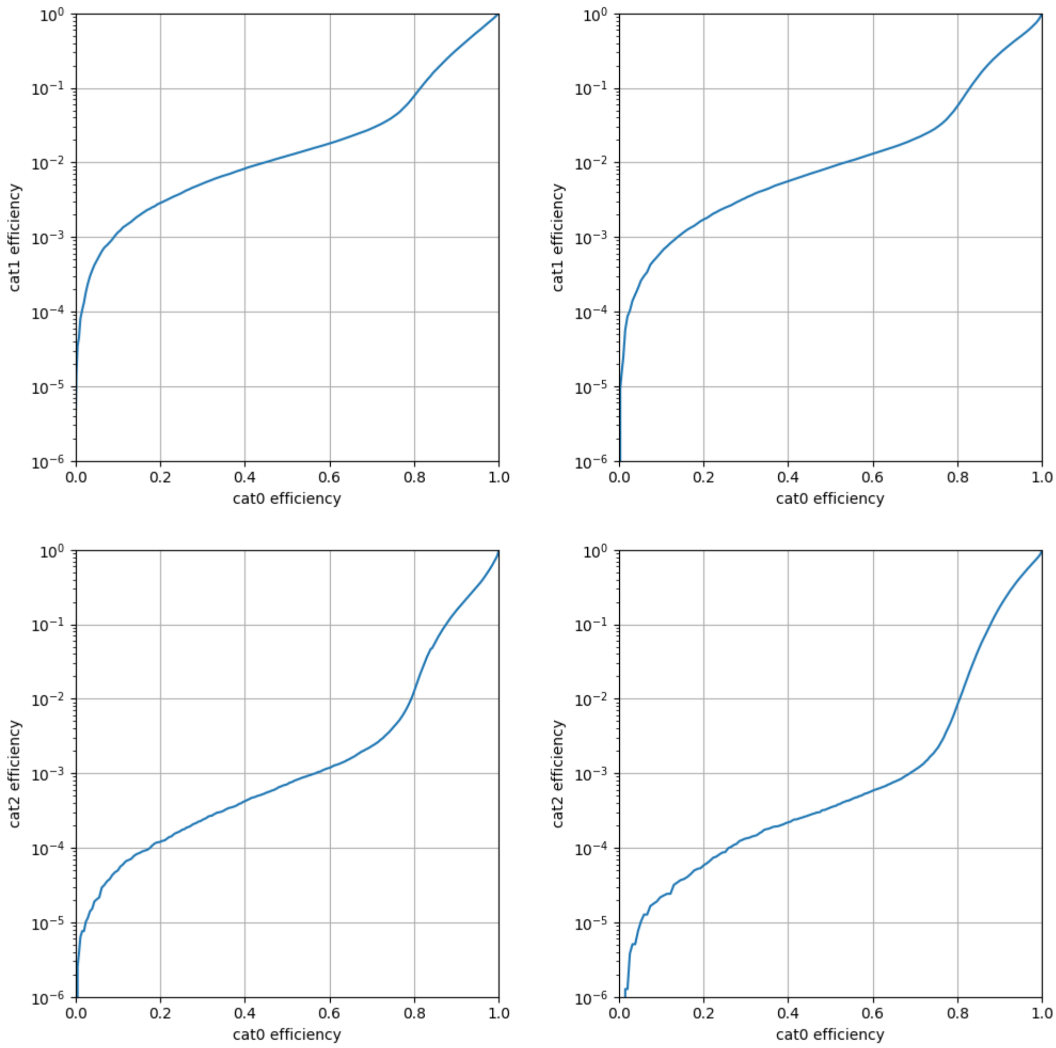

Figure 5: Receiver Operating Characteristics (ROC) curves for network validation cat0=leptonic top events, cat1=hadronic top events, cat2=QCD events. Top and bottom left are the curves for all layers, top and bottom right are the curves for energy fraction layers only.

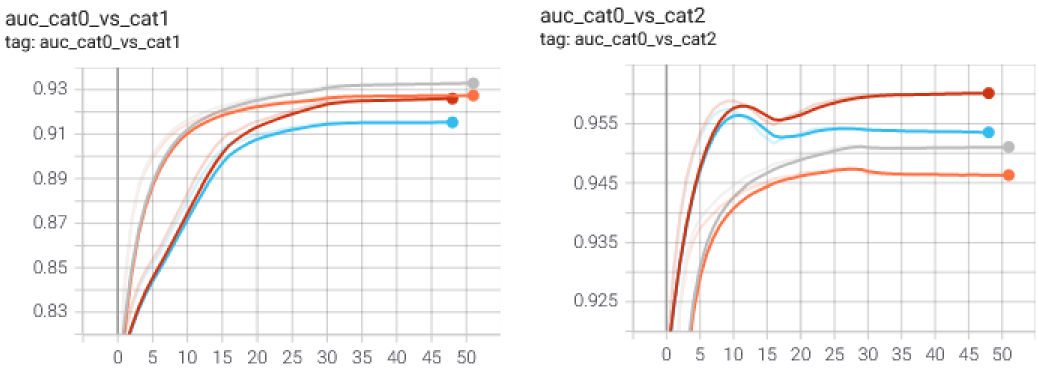

Figure 6: Area under the curve (AUC) for leptonic top vs hadronic top (left) and leptonic top vs QCD (right). Red: All Layers train. Blue: All Layers validation. Grey: energy fraction train. Orange: Energy Fraction validation.

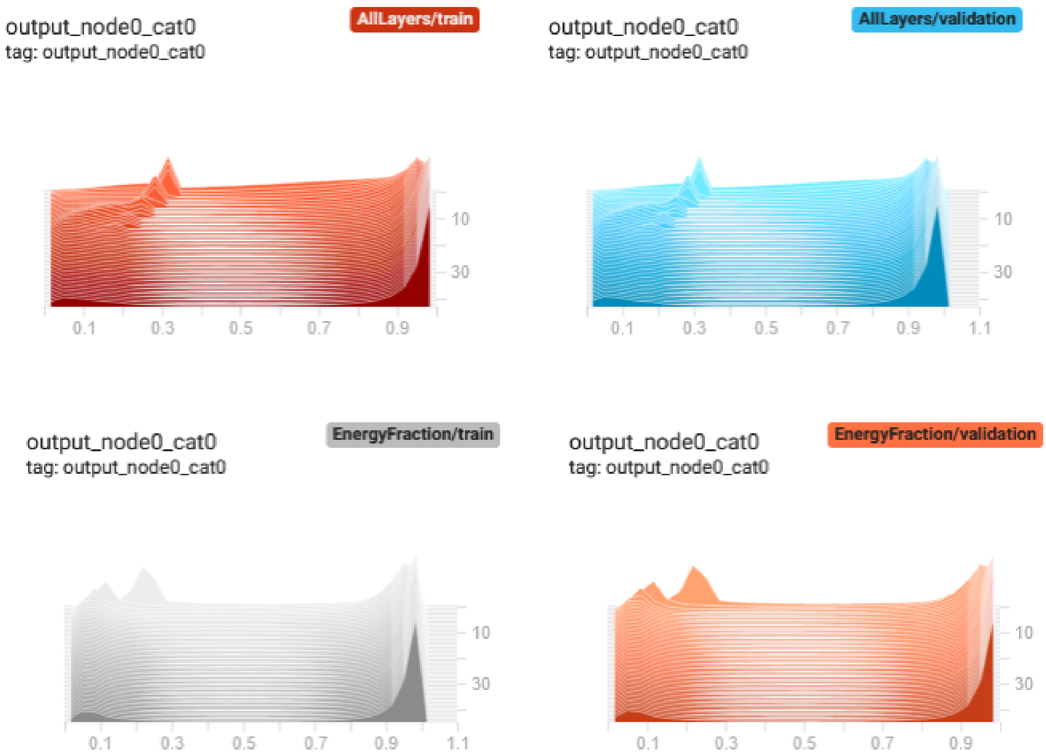

Figure 7: Classifier outputs for node 0, category 0 (signal).

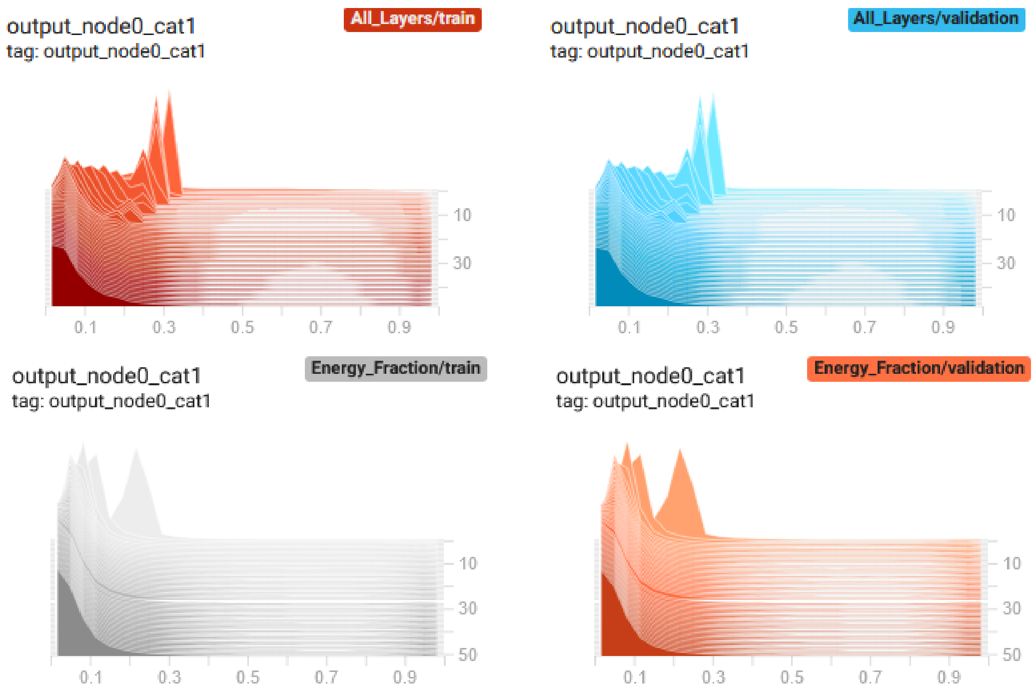

Figure 8: Classifier outputs for node 0, category 1 (hadronic top)

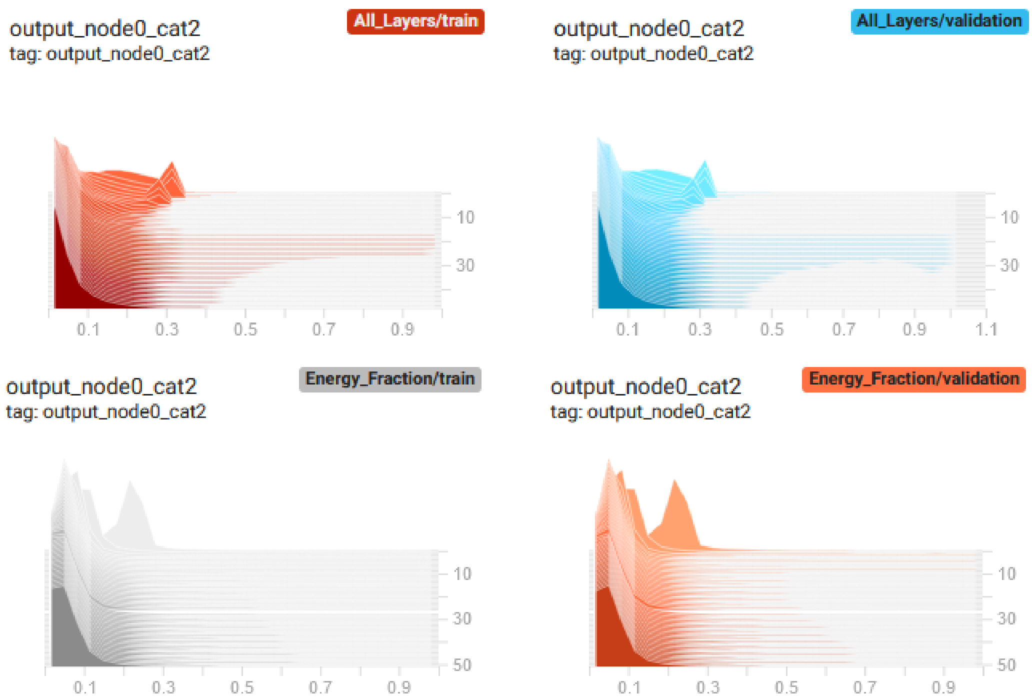

Figure 9: Classifier outputs for node 0, category 1 (QCD)

especially at lower signal efficiencies, as illustrated by the Receiver Operating Characteristics (ROC) curves in figure 6. At 60% signal efficiency, the hadronic top background efficiency is in the order of 10^{-2} for both validations and the QCD background efficiency is in the order of 10^{-3}, with the one for All Layers being marginally higher in both cases. The Area Under the Curve (AUC) plots show that the All Layers run reaches a larger AUC for leptonic top vs QCD than Energy fraction layers (0.955 at epoch 25 for All Layers validation against 0.945 of the Energy Fraction validation), while the opposite is true for leptonic top from hadronic top events (0.91 All Layers agaist 0.92). Both networks perform better at separating the signal from QCD background rather than from hadronic top background.

The values of the output node 0 for category 0 (signal) are displayed in figure 7. As epochs progress, the output approaches 1, which is expected if the network successfully recognized the signal. The values of the output node 0 for the two background categories are in figure 8 and 9. The outputs for both are close to 0 especially as epochs progess, sign that the network learns to reject these background events.

4 Conclusions

The ResNet50 network can clearly distinguish the leptonic top quarks from the background,so the results look promising.

Conventional (non-ML) methods require an already well identified and reconstructed lepton, which is a challenging task in cases where the quark jet and lepton start overlapping, such as highly boosted cases with a single fatjet. This CNN-based technique solves the problem as it does not need a precise lepton reconstruction and identification.

A possible improvement to be explored can be using a custom classifier for specific types of signal and background types only, rather than a multiclassifier like ResNet50. Adding different types of background, for example decaying leptonically and hadronically decaying W and Z bosons would also improve the classification in experiments in which the background types are more varied.

To study more in depth this classification method, a next step can be checking its performance on different bins of top quark p_T and comparing with traditional methods.

Acknowledgements

I would like to thank my supervisors Soham and Isabell for having me work on the project and being so helpful all the way through, the CMS group at DESY I've been a part of for these 8 weeks, and the entire Summer Student programme.

References

[1] S. Kanemura, M. Kikuchi, K. Yagyu *One-loop corrections to the Higgs self-couplings in the singlet extension* **http://arxiv.org/abs/1608.01582v2**(2017)

[2] T. S. Roy, A. H. Vijay *TIFR-TH/19-4 A robust anomaly finder based on autoencoders,* **arXiv:1903.02032**, (2020).

[3] M. Cacciari, G. P. Salam, G.Soyez, *The anti-kt jet clustering algorithm*, Annals of Physics, **JHEP 04**, 063 (2008).

[4] S. Bhattacharya, M. Guchait, A. Vijay, *Boosted Top Quark Tagging and Polarization Measurement using Machine Learning*, Annals of Physics, **JHEP 04**, 063 (2008).

[5] K. He, X. Zhang, S. Ren, J. Sun, *Deep Residual Learning for Image Recognition*, **arXiv:1512.03385**, (2015).

[6] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, *Aggregated Residual Transformations for Deep Neural Networks*, **arXiv:1512.03385**, (2015).

[7] V. Phung, E.Rhee *A High[U+2010]Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets*, **MDPI, https://doi.org/10.3390/app9214500** , (2019).