GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

Fakultät für Physik

Bachelor's Thesis

Untersuchung eines neuen Ansatzes zum Truth Tagging für die $t\bar{t}H(bb)$ Analyse mithilfe eines neuronalen Graphnetzwerks

Investigation of a new approach for truth tagging within the $t\bar{t}H(bb)$ analysis via a graph neural network

prepared by

Dennis Rodermund

from Herzberg am Harz

at the II. Physikalischen Institut

Thesis number: II.Physik-UniGö-BSc-2022/03

Thesis period: 28th March 2022 until 4th July 2022

First referee: Prof. Dr. Arnulf Quadt

Second referee: Prof. Dr. Steffen Schumann

Abstract

In dieser Bachelorarbeit wird ein neuer Ansatz zum Truth Tagging für die $t\bar{t}H(bb)$ Analyse mithilfe eines neuronalen Graphnetzwerks untersucht. Analysiert wird die Performanz des neuronalen Netzes im Vergleich zu den etablierten Effizienz Karten und dem Direct Tag für verschiedene Jet-Regionen, Arbeitspunkten und Anzahl an getaggten Jets. Weiterhin wird der Einfluss von Hadron Variablen auf die Performanz des neuronalen Netzes untersucht sowie die Fähigkeit des neuronalen Netzes, bottom-Jet Multiplizitäten vorherzusagen.

Es wird gezeigt, dass das neuronale Netz auch mit geringer vorhandener Statistik umgehen kann und für unterschiedliche Anzahlen an Jets innerhalb eines Ereignisses vergleichbare Resultate erzielt. Weiterhin wird gezeigt, dass das neuronale Netz ein Problem bei der Vorhersage von light Jets mit einem p_T von unter 60 GeV hat. Die genaue Ursache dieses Problems konnte noch nicht ergründet werden. Zusätzlich wird gezeigt, dass das neuronale Netz ein geeigneter Kandidat für den Einsatz in einer möglichen $t\bar{t}H(bb)$ Analyse Region ist. Jedoch sind für eine abschließende Bewertung weitere Untersuchungen erforderlich.

Stichwörter: Physik, Teilchenphysik, Truth Tagging, neuronales Netzwerk, neuronales Graphnetzwerk, $t\bar{t}H(bb)$ Analyse

Abstract

In this thesis, a new approach for truth tagging within the $t\bar{t}H(bb)$ analysis via a graph neural network is analysed. It is analysed how the neural network performs in comparison to the established efficiency map and the direct tag for various jet regions, working points and number of tagged jets. Furthermore, the influence of hadron variables on the performance of the neural network and the ability of the neural network to predict bottom-jet multiplicities is investigated.

It is shown that the neural network can deal with low statistics and can produce similar results for various numbers of jets within an event. Furthermore it is shown that the neural network has a problem dealing with light jets which have a p_T below 60 GeV. The cause of this problem has not been found. Additionally, it is shown that it is likely that the neural network can be used for a possible $t\bar{t}H(bb)$ analysis region, but more investigations are required for a final conclusion.

Keywords: Physics, Particle Physics, Truth Tagging, Neural Network, Graph Neural Network, $t\bar{t}H(bb)$ analysis

Contents

1. Introduction

The goal of particle physics is to describe the fundamental components of our universe and their interactions. The currently most successful theory of particle physics is the Standard Model, describing the fundamental components of our universe as elementary particles and their interactions through the exchange of gauge bosons. One example of these interactions is the Yukawa coupling between the Higgs boson and the top quark.
The top quark was discovered by the DØ and CDF collaborations at Fermilab in the year 1995, making it the most recently discovered quark [1, 2]. The Higgs boson H was predicted by the Higgs mechanism in the year 1964 [3, 4] and discovered by the ATLAS and CMS experiments at the LHC in the year 2012 [5, 6], thus completing the Standard Model. One way to measure the Yukawa coupling between the Higgs boson and the top quark properly is to examine the cross section of the $t\bar{t}H(bb)$ interaction, which is a Higgs boson associated production of a top-antitop quark pair ($t\bar{t}H$), discovered by the ATLAS and CMS collaborations [7, 8], where the Higgs boson decays into a bottom-antibottom quark pair ($H \to b\bar{b}$). Evidence for this decay has been found in the year 2018 [9–11]. Since the Higgs boson couples directly to the top quark during this interaction, the cross section of this process is proportional to the strength of the Yukawa coupling squared. But this process has a high background, which has to be distinguished from the signal events. A part of this background has additional light and c-jets and is therefore reducible by applying b-tagging, as will be further discussed in chapter 2.4. b-tagging is an algorithm which provides a discriminator that the jet came from a b quark. The ATLAS collaboration currently uses the DL1r algorithm for b-tagging [12, 13].
A first step to do this is the so-called direct selection cut, in which regions of phase space with a higher signal purity get defined for further processing. These selection cuts are used to reduce background contamination and include b-tagging cuts for the $t\bar{t}H(bb)$ channel because of the four b-jets expected in the final state. Events with too few b-tags at a specific Working Point are being rejected to purify the signal, but therefore the number of events usable for later developments of multivariate classification techniques is strongly reduced, resulting in lower training statistics. This can give badly modelled variable distributions for backgrounds which are almost completely removed. Therefore,

another approach needs to be developed. One alternative approach is the so-called "Truth Tagging" [14], which has the goal to imitate the behaviour of direct tagging without rejecting events with few b-tags.
Within this thesis, a new approach of truth tagging by the means of artificial neural networks originally developed for a $VH(bb)$ analysis will be evaluated on $t\bar{t}H(bb)$ background samples. The results will then be compared to the direct cut and the truth tagging with efficiency maps, which is the currently established procedure of truth tagging. The approach of truth tagging by the means of artificial neural networks was first introduced by Francesco Di Bello et al. [15]. Truth tagging will be discussed further in chapter 5 after a brief summary of the standard model in chapter 2, an overview of the LHC and the ATLAS detector in chapter 3 and a general overview of the functionality of neural networks and graphs in chapter 4. After this, the tools used will be described in chapter 6, followed by a presentation of the results in chapter 7 and a discussion of them in chapter 8. In the end, a summary of the thesis will be given in chapter 9, together with an outlook for further research and applications.

2. The Standard Model of Particle Physics

The Standard Model (SM) is the currently accepted underlying theory of particle physics, describing the fundamental particles and their interactions except for gravity [16–19]. Even if there are many things the SM cannot explain, as discussed in more detail in chapter 2.5, it is currently the most successful scientific theory ever created.

2.1. Basics of the Standard Model

All particles can be divided into two groups: The bosons, which have an integer spin, and the fermions, which have a half numbered spin. All matter that we see consists of fermions, while the fundamental forces electromagnetism, strong interaction and weak interaction can be described by the exchange of bosons. The fermions can be divided further into the quarks, which interact through the strong interaction, and the leptons, which do not. Also, the fermions are ordered into three groups, called generations, which roughly distinguish the particles by their mass. The particles of the SM can be seen in Figure 2.1 together with their mass, electric charge and their spin [20].
All fermions, except neutrinos, and the W bosons interact via the electromagnetic force, which is described by the exchange of photons. The charge, which describes the sensitivity of the particles to the electromagnetic interaction, is the electric charge. All quarks interact via the strong force, which is described by the exchange of gluons, which also interact via the strong force themselves. The corresponding charge is the colour charge. All particles except the photon and the gluons interact via the weak force, which is described by the exchange of W and Z bosons. The sensitivity of particles to the weak interaction is described by the weak isospin. The quarks furthermore hadronise because of colour confinement, which states that quarks cannot occur isolated in a single state, but only in bound states with neutral colour charge, creating so called hadrons. Two different types of hadrons exist: The baryons, which are bound states with a half numbered spin, and

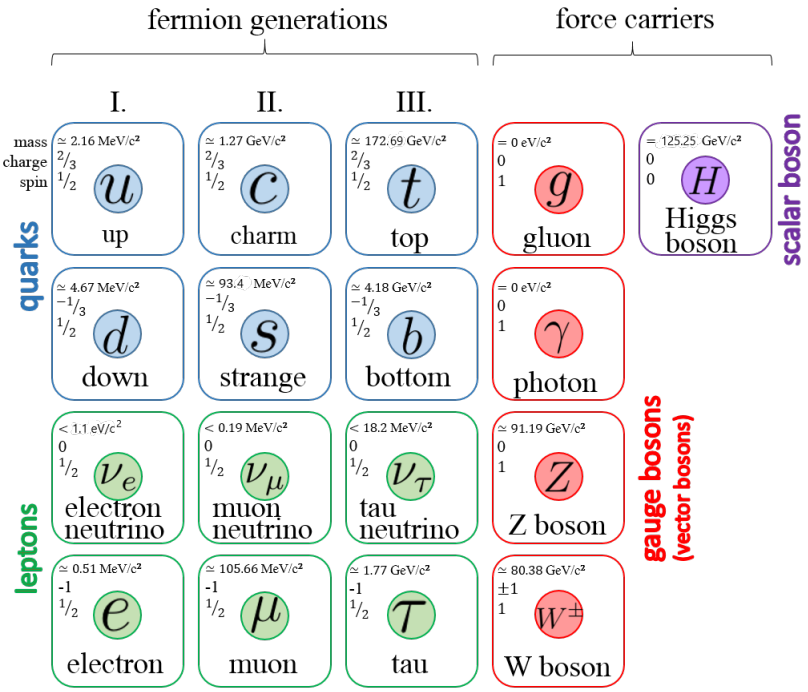

Figure 2.1.: The particles of the Standard Model of particle physics. Shown are the three fermion generations and the force carrying bosons together with their mass, electric charge and spin. The values for the particle properties are taken from [20]. Courtesy of Konrad Helms 2022.

the mesons, which have an integer spin. An example for baryons are the protons, which are used for collisions in the Large Hadron Collider.

2.2. Top Quark

Of special interest for this thesis is the top quark (t quark), the heaviest of the known elementary particles with a mass of $m_t = (172.69 \pm 0.30)\,\text{GeV}$ [20]. Since the typical timescale of hadronization lies within the order of $10^{-23}\,\text{s}$ [21] and the t quark has a decay width of $\Gamma = 1.42^{+0.19}_{-0.15}\,\text{GeV}$ [20], giving it a mean lifetime of $\tau \approx 5 \cdot 10^{-25}\,\text{s}$, the t quark is the only known hadron that decays before it hadronises. Therefore, it presents the unique opportunity to investigate properties of quarks in a non-bound state. The t quark has an electric charge of $+\frac{2}{3}\,\text{e}$ and is the weak isospin partner of the bottom quark (b quark). Because of its short life time, it cannot be measured directly with the ATLAS detector, but only through its decay products. The t quark mostly decays into a b or $\bar{b}$ quark together with a W^+ or W^- boson, respectively, as described by the CKM matrix [22, 23]. The W-

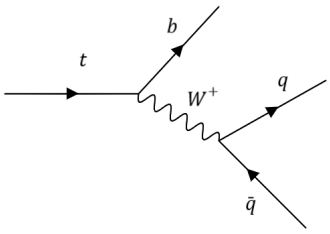

Figure 2.2.: Hadronic decay channel of the t quark

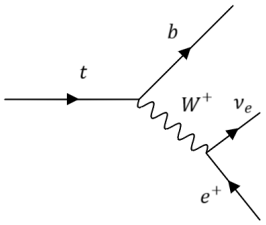

Figure 2.3.: Leptonic decay channel of the t quark

boson then furthermore decays either hadronically ($W \to q\bar{q}'$) or leptonically ($W \to \ell\nu_\ell$). Examples for the decay channels of the t quark can be seen in the figures 2.2 and 2.3. The decay into a pair of $q\bar{q}'$ is more difficult to measure in comparison to a leptonic decay: Because quarks cannot occur in non-bound states, the separation of quarks leads to the creation of particle jets. One of the main problems in hadron collider experiments is the identification of the flavour of the quark from which they originate. This procedure is called "Flavour Tagging", which will be further discussed in chapter 5.

2.3. The Higgs Boson

The Higgs boson is predicted by the Higgs mechanism, which was developed in 1964 [3, 4]. The Higgs mechanism solves the problem of how the W and the Z bosons acquire mass, since they were required massless by the gauge invariance in the SM, what was in contradiction to experimental results. The Higgs boson was predicted to be the quantized excitation of the Higgs field.
Since the Higgs boson couples to the mass of a particle, it couples with all SM particles except the photon, the gluon and the neutrinos. The t quark is the most relevant interaction partner for the Higgs boson since it has the highest mass of all known particles. Therefore, it is important to know the properties of the t quark if working in Higgs physics.

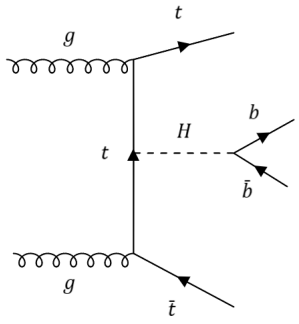

Figure 2.4.: Gluon fusion into a pair of top quarks and a Higgs boson.

The Higgs boson has a mass of $m_H = (125.25 \pm 0.17)\,\text{GeV}$ [20], a spin of zero and a decay width of $\Gamma = 3.2^{+2.8}_{-2.2}\,\text{MeV}$ [20]. With the ATLAS detector, the Higgs boson cannot be measured directly but instead via its decay products, for example a decay into a $b\bar{b}$ pair.

2.4. $t\bar{t}H$ Production Mode

The main goal in the investigation of $t\bar{t}H$ production is to measure the strength of the top-Higgs Yukawa coupling. One of the main production modes of $t\bar{t}H$ is shown in Figure 2.4. A pair of gluons merges into $t\bar{t}$ from which a Higgs boson is emitted. The Higgs boson thereupon decays into $b\bar{b}$. The notation for this decay channel is $t\bar{t}H(H \rightarrow b\bar{b})$. The $t\bar{t}H$ can decay fully leptonically, fully hadronic or semileptonically. The jets, which develop as a result of the W decay, can be any jets which are lighter than the W boson, making the analysis explicitly more difficult. But in most cases of hadronic decay, the W boson decays into a b jet together with a $\bar{u}$- or $\bar{c}$ jet or into a $\bar{b}$ jet and a u- or c jet, respectively, because the decay into d- and s jets have smaller CKM matrix elements.
This production mode is dominated by background processes in collision data. A background is everything that can be misinterpreted as the examined event. The dominant background is shown in Figure 2.5. A pair of gluons merge into $t\bar{t}$. After the merge one of the t quarks emits a gluon, which furthermore decays into $b\bar{b}$, resulting in the same final state particles as the examined $t\bar{t}H$ signal event. It is also possible that a gluon is emitted from an initial state parton. Backgrounds can be distinguished into reducible and irreducible backgrounds: Events, where the emitted gluon decays into light jets or c-jets can be cut away by applying b-tagging and are therefore reducible, whereas events with the same final state particles cannot be cut away and are therefore irreducible. To reduce this background within collision data via b-tagging, a direct selection cut is ap-

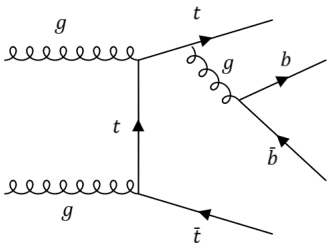

Figure 2.5.: Gluon fusion into a pair of top quarks, whereby an additional gluon gets emitted which decays into $b\bar{b}$, resulting in the same final state as $t\bar{t}H$.

plied. The concept of tagging will be discussed in more detail in chapter 5. But this approach has some difficulties, for example it lowers the number of usable events by a huge amount, and therefore needs to be improved, as discussed further in chapter 5. The $t\bar{t}H$ production mode with various decay modes of the Higgs boson was first observed by the ATLAS and CMS collaborations in the year 2018 [7, 8], but the $t\bar{t}H(H \to b\bar{b})$ decay has not been directly observed yet [9–11] and needs to be measured more precisely to extract the strength of the top-Higgs Yukawa coupling.

2.5. Limitations of the Standard Model

Although the SM is currently the scientific theory tested to the highest precision ever achieved, there are many things the SM cannot explain.
For example, there is no explanation in the SM how gravity works at microscopic scales. Also, there is no explanation for the phenomena of dark matter or why there is more matter within the universe than antimatter [24, 25]. Physics that tries to expand the SM to explain these phenomena is called "Physics beyond the Standard Model" (BSM). One example for a theory from the field of BSM is the Supersymmetry (SUSY), in which the assumption is made that every particle of the SM has a "Superpartner" with higher mass [26]. For example, within SUSY, the superpartners of WP, B^0 and π^0 would be candidates for the dark matter.

3. Experimental Setup

3.1. The LHC

The Large Hadron Collider (LHC) [27], run by CERN, is located in Switzerland near Geneva and is currently the largest and most powerful particle accelerator in the world. The LHC was installed in the same 26.7 km long tunnel with a tunnel cross section diameter of 3.7 m where the Large Electron-Positron Collider (LEP) was located [28]. The tunnels are located between 45 m and 170 m below the surface. The LHC is a proton-proton collider and therefore consists of two rings with counter-rotating particle beams. The protons are being injected through two transfer tunnels with a length of about 2.5 km, which connect the LHC with the CERN accelerator complex. There are four different interaction points used at the LHC, each one of them belonging to one experiment. Currently there are four experiments working at the LHC: ATLAS, CMS, LHCb and ALICE [29–32]. While ATLAS and CMS are multipurpose detectors, the focus of LHCb lies on precision measurements of CP violation and rare decays of B hadrons and ALICE is examining heavy ion physics and the quark-gluon plasma of the SM.

Currently, the LHC is starting Run 3, where its center-of-mass energy is increased from $\sqrt{s} = 13\,\text{TeV}$ in Run 2 to $\sqrt{s} = 13.6\,\text{TeV}$ [33].

3.2. The ATLAS Detector

After the collision of the particle beams at the interaction point, the particles produced in the interaction need to be identified and their properties like energy or transverse momentum need to be measured. For this purpose, the ATLAS (A Toroidal LHC ApparatuS) detector was developed. The goal of ATLAS is to test the SM and to search for physics beyond the SM [29].

The ATLAS detector has a height of 25 m and a length of 44 m. It is forward-backward symmetric with respect to the interaction point. It consists of four different concentric layers: The inner detector, the electromagnetic calorimeter, the hadron calorimeter and the muon spectrometer [29]. A cross-section of the ATLAS detector can be seen in Figure

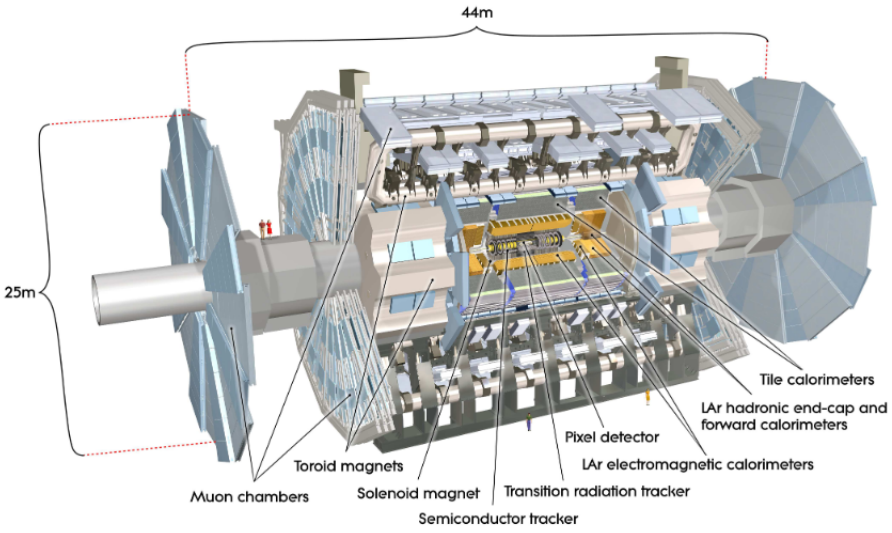

Figure 3.1.: Computer generated Cross-section of the ATLAS detector. Atlas Experiment © 2008 CERN.

3.1.

3.2.1. Mathematical Description

The ATLAS detector is described through a cylindrical coordinate system. The z-axis corresponds to the beam line. The x and y axis are perpendicular to the beam line. Furthermore, an azimuthal angle ϕ and a polar angle θ are used. Often, θ is not used directly, but instead the pseudorapidity η is used [29]. The pseudorapidity has the advantage that the difference of two pseudorapidities is invariant under a Lorentz transformation, which is important, if one wants to transform between the lab frame of reference and the center of mass system. η and θ are connected through

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right). \tag{3.1}$$

3.2.2. Inner Detector

The inner detector is immersed in a 2 T solenoidal field and consists of Pixel, a silicon pixel detector, a semi-conductor tracker (SCT) and a transition radiation tracker (TRT) [34, 35]. Within the inner detector, the particles are being tracked and with the transition radiation tracker the electrons are being distinguished from the hadrons. Because of the magnetic field, the charged particles get deflected via the Lorentz force, forcing the particles to

perform a circular movement. This deflection allows the calculation of their transverse momentum and the identification of the electric charge through a measurement of the track curvature. Since b-tagging relies on identifying a secondary vertex within the detector, the tracking of particles by the inner detector is crucial for b-tagging. The Pixel detector [34] has the task to precisely track the particles near the interaction point. It is made of silicon, which is divided into pixels to increase the precision. For run 2, the insertable b-layer (IBL) [36] was added to further increase the precision. The semi-conductor tracker tracks the particles at larger scales but with lower resolution. The transition radiation tracker [35] consists of straw tubes, filled with a gas mixture. The tubes offer tracking within a large volume but with a comparatively bad resolution and are used for cost efficiency reasons, mainly for particle identification and to separate electrons from pions.

3.2.3. Calorimeters

The electromagnetic calorimeters measure the energy and momentum of electrons and photons while the hadronic calorimeters measure the energy and momentum of hadrons. Within the calorimeters, particle showers evolve due to electromagnetic effects like pair production and Bremsstrahlung or due to strong interactions between hadrons and the calorimeter material. The energy of the particles can be calculated through the measurement of the generated charge within the calorimeter or through the arising scintillation light from particles traversing the calorimeter. Since the energy resolution of a calorimeter increases as $\frac{1}{\sqrt{E}}$, calorimeters are very useful for high energy physics. The electromagnetic calorimeter cover a pseudorapidity of $|\eta| < 3.2$ and is filled with high granularity liquid-argon (LAr). The hadronic calorimeters have a range of $|\eta| < 1.7$, but together with the LAr forward calorimeters, which can measure both electromagnetic and hadronic showers, the range of the calorimeters increases to $|\eta| = 4.9$.

3.2.4. Muon Spectrometer

Muons do not interact much within the calorimeters, therefore their energy and momentum is measured with the muon spectrometers, which consist of three layers of high precision tracking chambers. The idea of the muon spectrometers is the same as of the inner detector: A magnetic field is created by outer toroidal magnets, forcing the muons to perform a circular movement and therefore the calculation of their transverse momentum through a measurement of the track curvature is possible.

3.2.5. Trigger System

Within the LHC, there is an interaction every 25 ns, which is too much data to be processed. Therefore, the ATLAS detector has a trigger system [37], which filters the interesting data from the less relevant. The trigger system consists of a hardware-based component, called "Level 1 Trigger", and a software based component, called "High level Trigger". The level 1 triggers measure if particle quantities like energy, momentum or missing transverse energy of the event exceed a predefined barrier. If the events are accepted by the level 1 trigger, the events are passed on to the higher level trigger, which in turn uses information like the tracking information of the inner detector and the calorimeter deposits to decide if the event gets accepted or rejected. If the event is accepted by the higher level trigger, the data of the event gets passed on to external storage.

4. Neural Networks and Graphs

4.1. Structure of Neural Networks

Neural networks (NN) are a widely used tool for machine learning (ML). ML has the goal to create algorithms that are capable of 'learning'. Learning means that the considered algorithm gets a set of training data as input and practices to make predictions or decisions without an external user telling the program to do so.
A fully connected NN, which is the easiest form of an NN, consists of different layers, as shown in Figure 4.1. The layers can be split in three categories: The first layer is the input layer, in which the NN takes the input data it shall process. The last layer is the output layer, in which the NN outputs the result of its calculation. The layers between the first and the last layer, also called hidden layers, have the purpose to process the input data into the output result.

Each layer consists of different nodes n, which correspond to the neurons in the brain. Each node holds a number, which represents its activation a_n. The higher the number, the higher the activation of the node. Each node is connected to each node of the previous and of the following layer. A weight w can be assigned to every connection, representing the strength of this connection. The activation of a node within the layer j can therefore be described as the weighted sum over all nodes in the layer $j-1$. The output of a node, given an input, is described by an activation function σ, for example a linear activation function or a rectified linear unit (ReLU) function. The ReLU function is defined as

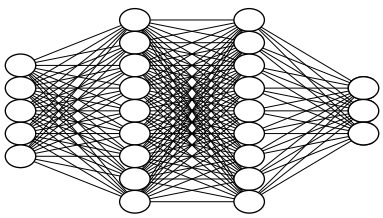

Figure 4.1.: The Structure of a Neural Network. Courtesy of Konrad Helms 2022.

$$f(x) = \max(0, x). \tag{4.1}$$

Furthermore, a bias b_n is assigned to every node to control how big the weighted sum has to be to activate the node:

$$a_n^{(j)} = \sigma\left(\sum_{i=1}^{N} w_{i,n}^{(j)} \cdot a_i^{(j-1)} + b_n^{(j)}\right) \tag{4.2}$$

4.2. Loss Functions

Each connection weight and bias can be seen as a variable. In the beginning, the weights and biases are randomly initialised. After the NN processed the first input, the resulting output gets compared to the expected output. To compare the result of the NN with the wished result, a cost function (CF) gets defined. The inputs of the CF are the output of the NN and the expected output. The CF thereafter returns a positive value, the loss. The loss indicates, how high the difference between the output of the NN and the expected output is. Within this thesis, the CrossEntropyLoss function is used.

4.3. Training of Neural Networks

The goal of the training of an NN is to find the minimum of the CF. The algorithm, by which the minimum of the CF gets calculated, is called the optimizer of the NN. The commonly used optimizers are using the gradient descent of the CF: At first, the NN gets filled with random variables and the CF gets calculated for this set of variables. Thereupon, the gradient ∇CF of the CF gets calculated. If m is the number of variables within the NN, then ∇CF is an m dimensional vector pointing in the direction of the largest increase of the CF. Therefore -∇CF points in the direction of the largest decrease of the CF. The variables of the NN get updated by going a step into the direction -∇CF and the CF with the new variables gets calculated. The size of the step is called the learning rate of the GNN. This procedure gets repeated until the training converges. It is also possible that the learning rate gets reduced after a few training steps to avoid overshooting of the gradient descent. The factor, by which the learning rate is reduced, is called the gamma factor.
It is important to not overdo the training of the NN, otherwise the NN gets too used to the training data and cannot work well with a different set of data. This case is called

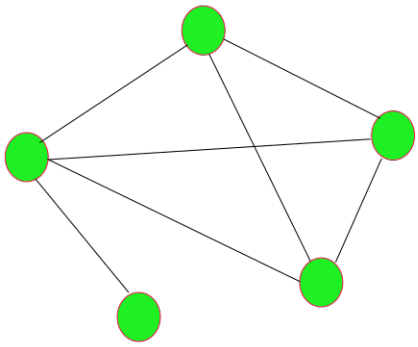

Figure 4.2.: Example of an undirected graph with five vertices

"overtraining". Furthermore, it is important to test the performance of the NN with a set of data which is not used doing the training to assure that the NN performs well on an before completely unknown set of data. After the training is complete, the validation of the NN takes place with another set of data.

4.4. Graphs

A graph is a data structure that consists of two different types of components: The vertices and the edges, which represent the relations between the vertices and are drawn as connecting lines between them [38]. An example for a graph structure can be seen in figure 4.2. Although there are many different types of graphs, for example directed graphs or mixed graphs, the only relevant type of graphs for this thesis is the graph with undirected but weighted edges. Graphs are very useful to describe sets of objects and correlations between these objects and are widely used in chemistry, traffic analysis or physics. The usage of graphs for this thesis will be further described in chapter 5.2.2.

4.5. Graph Neural Networks

A graph neural network (GNN) is a special kind of NN, that takes a graph as an input and operates on it [39]. A GNN has the advantage, that it can deal with inputs of variable sizes, whereas normal NNs can only deal with a fixed number of input parameters. Within this Thesis, a GNN will be used to classify the vertices of a graph and therefore predict the flavour of a jet, which is represented by this vertex. This will be further discussed in

chapter 5.2.2.
To classify a vertex, the GNN takes the information of the vertex itself and all vertices that are connected to the considered vertex via edges. If the vertex n has the status v during step k, then the status during step k+1 depends on the update function f via

$$v_n^{k+1} = f(v_n^k, g(v_i^k), \forall i \in V), \tag{4.3}$$

where V is the set of all vertices that are directly connected to the considered vertex and g is the aggregation function, which determines, how the information of the vertices v_i get processed.
Therefore, within the first actualisation step, only information of vertices that are directly connected to the considered vertex have an impact on the classification. But since those vertices are connected with other vertices themselves and will be updated by considering those information, the information of all vertices of the graph will be used to classify a certain vertex step by step, prioritising the information of vertices, that are directly connected to the considered vertex over those, which are not.

5. Flavour Tagging methods

Because of colour confinement, as described in chapter 2.1, highly energetic quarks produced in collisions form showers of hadrons, called jets. Therefore the problem arises to identify which quark-flavour is the origin of a certain jet. The procedure of assigning a jet to an origin quark is called "Tagging".
Of special interest for this bachelor thesis is the b-tagging since the decay of a t quark nearly always results in a b-jet. b hadrons have the advantage that they travel a certain distance before they decay, but they still decay inside the detector, resulting in a secondary vertex, which can be reconstructed using the inner detector of the ATLAS detector. Therefore, a major component in the process of b-tagging is to search for this secondary vertex [40]. For this thesis, two types of tagging are of special interest: The direct tagging and the truth tagging. While direct tagging is the only technique that can be used for detector data, truth tagging is used to enhance statistics of Monte-Carlo (MC) Simulations. For both methods, a working point (WP) is given. The WP indicates what proportion of b-jets has to be tagged correctly. By decreasing the proportion of correctly tagged b-jets, the proportion of mistagged light- and c-jets gets decreased to a larger extent, resulting in a higher purity of correctly tagged b-jets.

5.1. Direct tagging

By using direct tagging, every event that does not fit in the region defined by b-tagging multiplicity for a given WP is being rejected. Although this procedure is working very well in identifying the b-jets, the problem is that a lot of data is required to reach usable amounts of statistics due to the high event rejection rate. Therefore, a technique, which imitates the direct tagging behaviour, but does not reject events with a low b-tagging probability, is desirable.

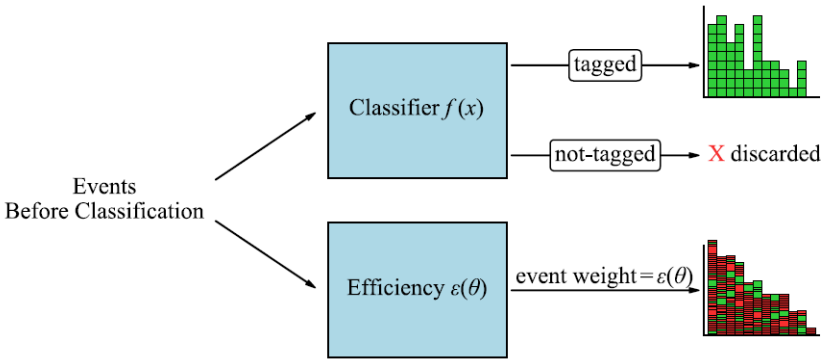

Figure 5.1.: Comparison of direct tagging and truth tagging. Courtesy of Francesco Di Bello et al. [15]. No changes to the image have been made.[1]

5.2. Truth tagging

The goal of truth tagging is to achieve the same tagging behaviour as with the direct tagging, but without the high event rejection rate. Therefore, instead of rejecting all events that do not fit the tagging requirements, every event gets a weight assigned corresponding to the probability that the event has the required number of jets which are tagged as b-jets. The higher the probability that the event fulfills a tagging selection cut, the higher is the weight of the event. The difference between direct tagging and truth tagging is illustrated in figure 5.1. Within truth tagging, every jet of an event gets a weight assigned. The jet weight is corresponding to the probability of the jet being tagged as a b-jet. These jet weights are then combined to an event weight by multiplying the probability of jets being tagged or not tagged, respectively, depending on the number of aimed tags. Different permutations of tagged jets are being considered by summing them up, where the kinematic variables of one of those permutations gets randomly chosen to represent all permutations. For example, for an event with two jets, the calculation of the event weight is given by

$$\epsilon_{\text{event}} = \begin{cases} (1-\epsilon_1)(1-\epsilon_2) & \text{if } n_{\text{tag}} = 0 \\ \epsilon_1(1-\epsilon_2) + (1-\epsilon_1)\epsilon_2 & \text{if } n_{\text{tag}} = 1 \\ \epsilon_1\epsilon_2 & \text{if } n_{\text{tag}} = 2 \end{cases} \tag{5.1}$$

Here, ϵ_1 and ϵ_2 are the weights of jet 1 and 2, respectively, and n_{tag} indicates the required number of tagged jets.
Two types of truth tagging are relevant for this thesis: The truth tagging by the usage of efficiency maps and the truth tagging with GNNs.

[1] A copy of the license to use this image is available under https://creativecommons.org/licenses/by/4.0/

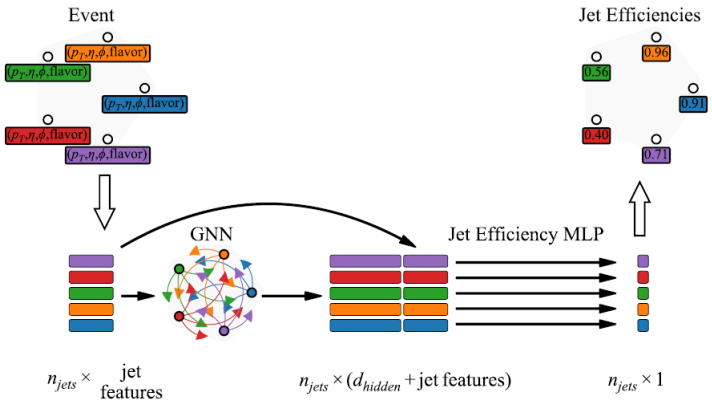

Figure 5.2.: Principle of truth tagging with graph neural networks. Courtesy of Francesco Di Bello et al. [15]. No changes to the image have been made.[2]

5.2.1. Efficiency maps

Efficiency maps are the commonly used technique for truth tagging so far. An efficiency map is a two-dimensional map, consisting of different bins. A jet gets assigned to a certain bin depending on its transverse momentum and pseudorapidity. After that, the tagging probability of that bin, which depends on the reconstruction of secondary vertices and combined track impact parameter information of reconstructed charged particles, gets assigned to the jet. The problems of this approach are that they neglect correlations like the angular distance ΔR between the jets and are limited by statistics for more than two variables. Therefore, another approach, which considers the correlations between the jets and can handle more variables, is in need of being developed.

5.2.2. Truth tagging with graph neural networks

Truth tagging by the means of graph neural networks aims to circumvent the limitations of the efficiency maps approach. This approach was first developed by Francesco Di Bello et al. [15]. The idea is to handle the events as a graph, where the jets are the vertices and the angular distances ΔR between the jets are represented by the edges. A GNN then takes the event as input and tries to predict the probabilities of the jets to be in the different WP bins by classifying the corresponding vertex of the graph. The outputs of the GNN are the weights of the jets, which then get combined to an event weight. This method is illustrated in figure 5.2. The specifics of the GNN used in this thesis will be further discussed in chapter 6.3.

[2] A copy of the license to use this image is available under https://creativecommons.org/licenses/by/4.0/

6. Methods

6.1. Sample and event selection

The used sample was a $t\bar{t}$ sample in the semileptonic decay channel from the Powheg + Pythia 8 MC generator [41, 42] within the five flavour scheme, which states that the b-quarks are assumed to be massless. The Atlfast-II generator was used for the detector simulation [43].
Every event has at least 5 jets and, if not otherwise indicated, a maximum of 8 jets. No event, which has a jet with a p_T of more than 1 TeV, was used.
During the preprocessing, the Sample was split into four parts based on the event number. Two of the parts were used for the training of the GNN, one for the evaluation after each epoch and the last part was used for the final validation after the training. Furthermore, a downsampling was applied to the heavy flavour classification (HF classification) [11] so that the sample has the same magnitude of $tt + b$, $tt + c$ and $tt +$ light events. However, this downsampling was applied only to the train sample and the evaluation sample. There was no downsampling applied to the final validation sample to make the validation sample more realistic.
Before the preprocessing, the used sample had 33,243,348 events in total. After the preprocessing, the train data set had 2,135,576 events, the evaluation data set had 1,018,111 events and the validation data set had 7,692,397 events.

6.2. Assignment of hadrons to jets

During the analysis, hadron variables were included to see, if they can improve the performance of the GNN. To assign a hadron to a jet, a cone around the jet was considered. If not otherwise indicated, the size of this cone was $\Delta R = 0.4$ and there was no cut for the hadron p_T. Within the cone, b-hadrons were prioritised over c-hadrons. If there was no b- or c-hadron within the cone, a light hadron was assigned by default. If there was more than one b- or c-hadron, either the hadron with the lowest ΔR to the jet was assigned (ΔR matching), or the hadron with the highest p_T was assigned (p_T matching).

6.3. GNN Specifics

The GNN used was originally developed for the ATLAS VHbb analysis by Nilotpal Kakati and needed to be retrained since it was adjusted to the $t\bar{t}H(bb)$ analysis. If not otherwise indicated, a learning rate of $l = 10^{-5}$ and a batch size of 1024 was used. The gamma factor was 0.1 and the used optimizer was "Adam". The CrossEntropyLoss function was used as loss function. The used activation function was a sequence of a linear function, a ReLU function, another linear function and a tangens hyperbolicus function.
The variables, on which the GNN was trained, were the transverse momentum, the pseudorapidity η, the azimuthal angle φ and the truth flavour of the jet and the hadron. For the hadrons, also the energy and the mass was used. Also the angular distances ΔR between the jets and the ΔR between the jets and the hadrons were used.
The predictions of the GNN were analysed for the WPs "no Tag", "85% WP", "77% WP", "70% WP" and "60% WP", where the % value indicates, how high the fraction of correctly tagged b-jets has to be. For the WP "no Tag", the considered jet does not even pass the 85% WP. If not otherwise indicated also jets which are tagged at a higher WP were considered in the histograms for a given WP to imitate the direct tagging. However, the output of the GNN assigns a jet to only one WP.
All results that are shown in this thesis were produced with a GNN that was trained for 5 epochs.

7. Results

The goal of this chapter is to compare the predictions of the GNN with those of the conventional p_T - η efficiency maps and the direct selection cut. A possible $t\bar{t}H(bb)$ analysis region will be reconstructed step by step and the influence of different parameters within this region will be analysed one by one.
First, the behaviour of different permutations of the sample splitting will be presented. Then, the influence of hadron variables on the predictions of the GNN will be presented. After that, the results for leading and subleading jets will be presented, followed by the results for different WPs and jet regions. Furthermore, the b-jet multiplicity for a given WP will be analysed. In the end, results for different numbers of tagged jets will be presented.
The aim of truth tagging is to reproduce the kinematic distributions of the sample with the direct cut applied so that the truth-tagged sample is a good approximation of a direct tagged sample with enough statistics. Since the reducible background of the $t\bar{t}H$ production mode has additional light and c-jets, as described in chapter 2.4, the focus of this thesis lies on the ability of the GNN to predict the kinematic distributions for light and c-jets. Within this chapter, the results will be presented for light flavour jets. Corresponding plots for charm jets can be seen in the Appendix. Investigated were mostly the 4th highest p_T jets in the 5-8 jet region since this region has the most statistics for light jets. If not otherwise indicated, the chosen WP is 70% since this is the WP that is used for a possible $t\bar{t}H(bb)$ analysis region.

7.1. Influence of different sample splitting

To examine the influence of the used sample on the predictions and to check for uncertainties introduced by the GNN through the random sample and initialisation, the sample was subdivided into four parts based on the event number. Two of those were used for the training of the GNN, one for the evaluation and one for the final validation after the training. Figure 7.1 shows the predictions of the GNN, if the four parts of the sample were cyclically permuted. Additionally, the mean of the predictions is shown in red. The

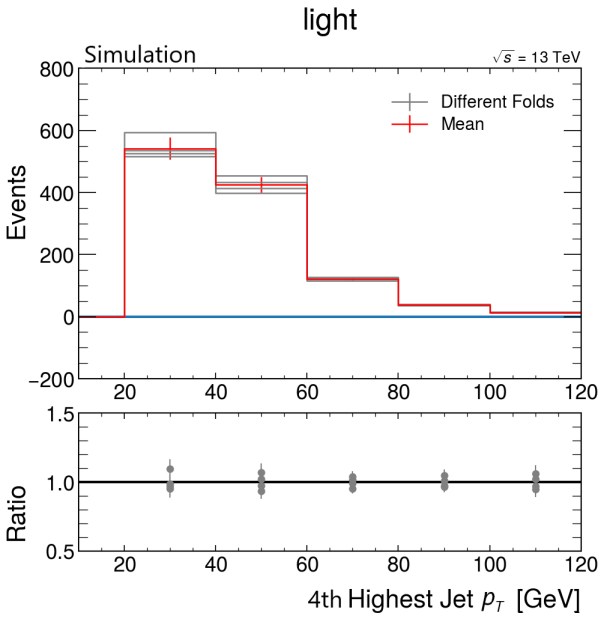

Figure 7.1.: Comparison of the GNN predictions for different permutations of the sample parts for the 4th highest p_T jet, if that jet is a truth light jet, within the 5-8 jet region at a 70% WP. The errorbars indicate the statistical uncertainties. The mean is shown in red.

results are shown for the prediction efficiency of light jets for the 4th highest p_T jet in the 5-8 jet region. The chosen WP was 70%. Also jets, that were tagged at a higher WP, were considered to imitate the direct tagging.
It can be seen that the mean lies within the statistical uncertainty of the different folds. Therefore the GNN predictions seem to be independent from the chosen permutation.

7.2. Evolution of average loss for training with and without hadron variables

After each actualisation of the GNN, the loss was calculated as described in chapter 4.2. Additionally, after each epoch, the GNN was evaluated on a set of data, which it was not trained on, and the loss from this evaluation was calculated, too. Therefore, two different evolutions of average losses can be investigated: The train loss evolution, which shows the loss taken after each iteration, and the evaluation loss evolution, which shows the loss after each epoch of training on an independent sample. Since the train loss was calculated using the train data sample and the evaluation loss was calculated using the evaluation data, differences between the behaviour of the evolution of losses indicate an overtraining of the GNN. The loss evolutions for the GNN with and without implemented

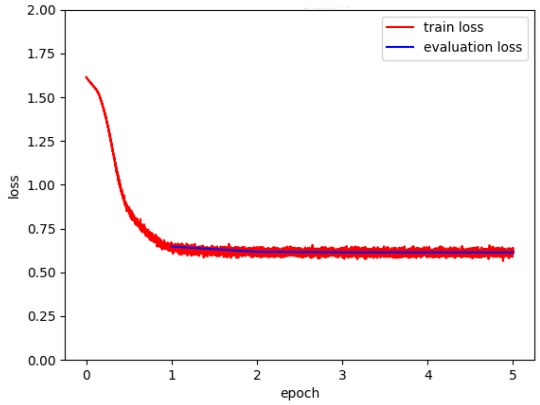

Figure 7.2.: Evolution of the average loss for the GNN with implemented hadron variables for a training of 5 epochs. The train loss is shown in red and the evaluation loss in blue.

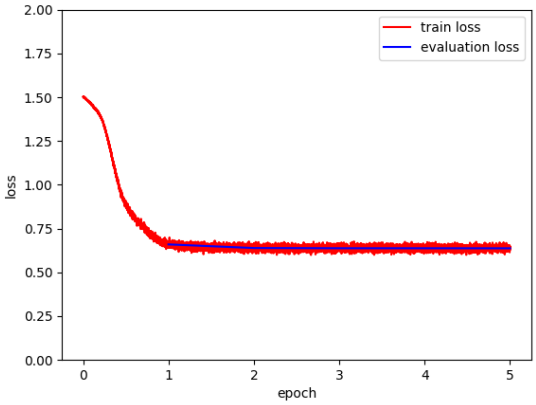

Figure 7.3.: Evolution of the average loss for the GNN without implemented hadron variables for a training of 5 epochs. The train loss is shown in red and the evaluation loss in blue.

hadron variables can be seen in the figures 7.2 and 7.3
It can bee seen that the loss evolutions for the GNN with and without hadron variables behave almost identically. After an exponential and smooth decrease of the train loss within the first epoch, there is no change after the first epoch except for fluctuations. The evaluation loss starts after the first epoch and, after a smooth decrease from the first to the second epoch, the minimum of the evaluation loss becomes the mean of the train loss. Therefore, an overtraining of the GNN is not observed.
Similiar results were obtained for all trainings of the GNN presented in this thesis.

7.3. Predictions with and without hadron variables

To analyse the influence of the hadron variables on the performance of the GNN, a training with implemented hadron variables was compared to a training without. The hadrons were assigned to the jets using the ΔR matching method as described in chapter 6.2. The results can be seen in the figures 7.4 and 7.5. The predictions were made for the 4th highest p_T jet in the 5-8 jet region. The chosen WP was 70%.
It can be seen that the efficiency map overall performs better than the GNN except for jets with a p_T above 100 GeV and implemented hadron variables. This will be further discussed in chapter 8.2. For the efficiency map the implementation of the hadron variables improves the results for jets with a p_T below 100 GeV, but the predictions for jets with a p_T above

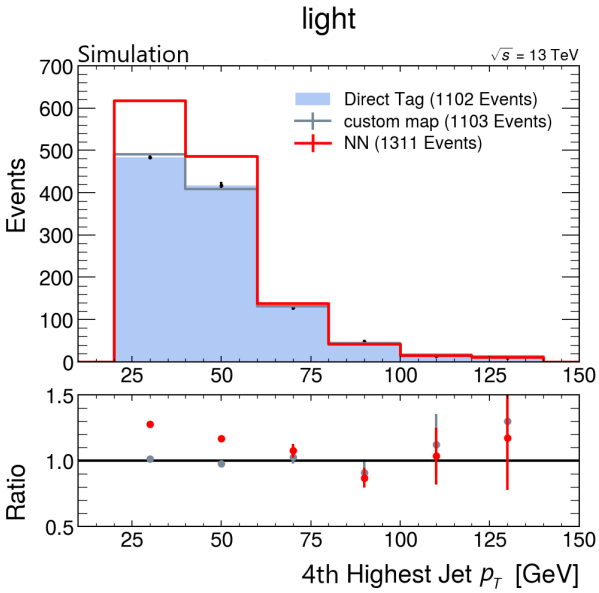

Figure 7.4.: Results for the GNN with implemented hadron variables for the 4th highest p_T jet, if that jet is a truth light jet, in the 5-8 jet region at a 70% WP. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainties.

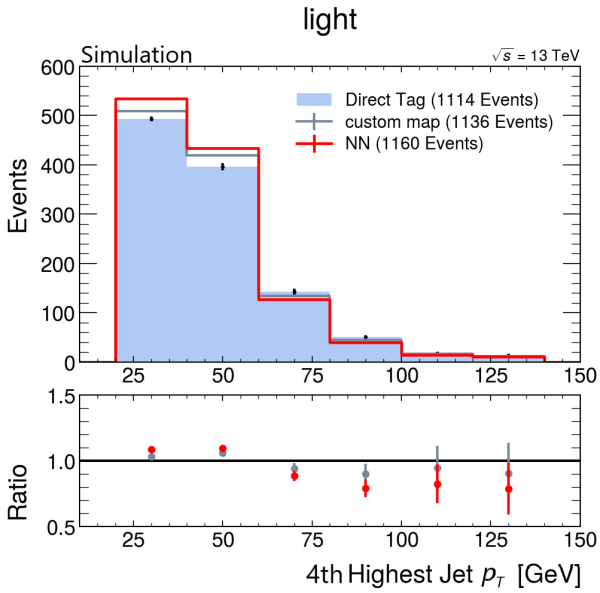

Figure 7.5.: Results for the GNN without implemented hadron variables for the 4th highest p_T jet, if that jet is a truth light jet, in the 5-8 jet region at a 70% WP. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainties.

100 GeV become worse.

On the other hand, the GNN performs better for jets with a p_T above 100 GeV if the hadron variables are implemented. However, the predictions of the GNN become worse for jets with a p_T below 60 GeV. This will be further discussed in chapter 8.2.

For the following results, the hadron variables are always implemented using the ΔR matching method.

7.4. Comparison between leading and subleading jets

To analyse how well the GNN performs for jets of different p_T within an event, the predictions of the GNN were analysed for the leading jet up to the 4th highest p_T jet. In the figures 7.6 - 7.9, the predictions of the GNN are compared to those of the direct tag and the efficiency maps. The predictions were made for light jets in the 5-8 jet region to improve the statistics and the chosen WP was 70%.

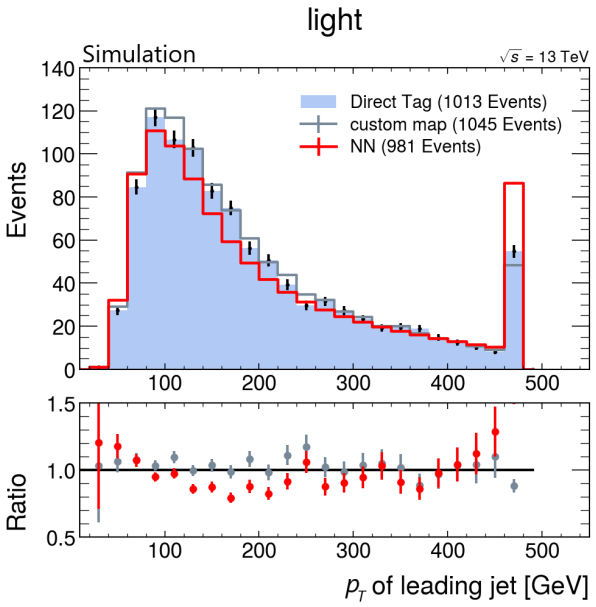

Figure 7.6.: Results for leading jets, if that jet is a truth light jet, in the 5-8 jet region for a 70%WP. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainties.

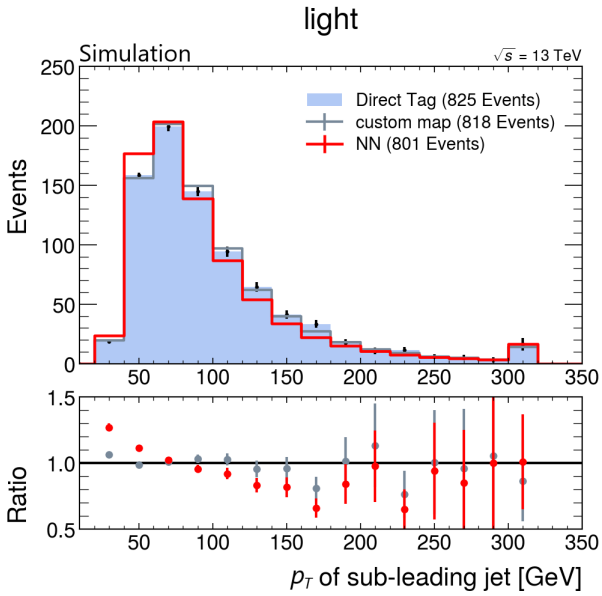

Figure 7.7.: Results for subleading jets, if that jet is a truth light jet, in the 5-8 jet region for a 70%WP. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainties.

It can be seen that for the leading jet the predictions of the NN are in most cases lower than those of the direct tag and lie outside of the statistical uncertainty of the direct tag, whereas the efficiency map agrees better with the predictions of the direct tag and lies within the uncertainty.

For the subleading jet, the predictions of the NN become better, especially for jets with a p_T above 200 GeV, but also the statistical uncertainty of the NN becomes very high in this region. But still, the predictions of the efficiency map are closer to those of the direct tag.

For the third and fourth leading jet, the predictions of the NN are very close to those of the direct tag and the efficiency map for jets with a p_T above 60 GeV, whereas the predictions of the NN are by far higher than those of the direct tag and the efficiency map for jets with a p_T below 60 GeV. The reason for this will be discussed further in chapter 8.2.

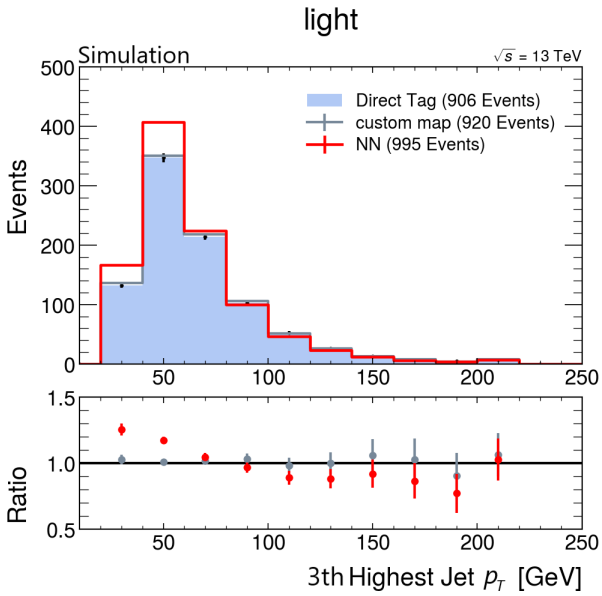

Figure 7.8.: Results for 3th leading jets, if that jet is a truth light jet, in the 5-8 jet region for a 70%WP. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainties.

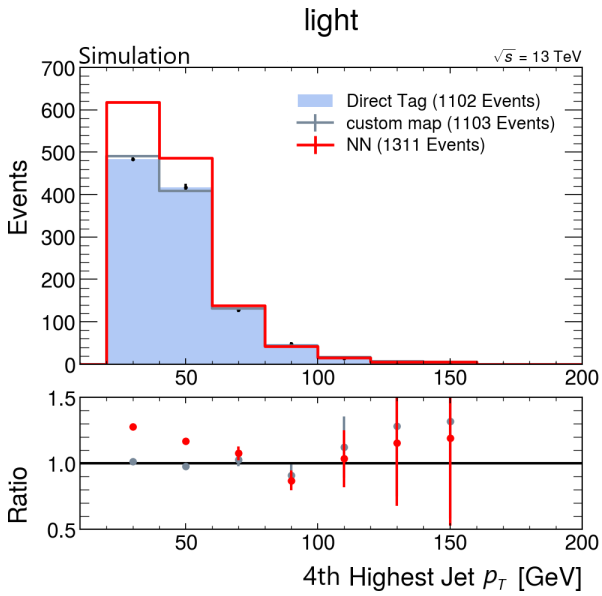

Figure 7.9.: Results for 4th leading jets, if that jet is a truth light jet, in the 5-8 jet region for a 70%WP. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainties.

7.5. Results for different working points

To analyse the dependency of the GNN performance on the different WPs that could be used for selection, predictions were made for the WPs "no Tag", in which case no tagging requirement was made, "85% WP", "77% WP", "70% WP" and "60% WP". In the figures 7.10 - 7.14 the predictions of the GNN in comparison to those of the direct tag and the efficiency map for different WPs are shown. Investigated was the tagging of light jets for the 4th highest p_T jet. The chosen jet region is the full 5-8 jet region.

For every WP, also jets, which were tagged at a higher WP, were considered to see where deviations occur. Therefore the plot for the "no Tag" WP can be seen as a sanity check whether the GNN, the efficiency map and the direct tag are considering the same number of total events. This is the case, as can be seen in figure 7.10.

For an increasing WP, it can be seen that the GNN predicts higher tagging probabilities than direct tag for light jets with a p_T below 60 GeV. While this also happens for the efficiency map, the effect is by far greater for the GNN. For jets with a p_T above 120 GeV, the statistical uncertainty of the GNN and the efficiency map become very large because

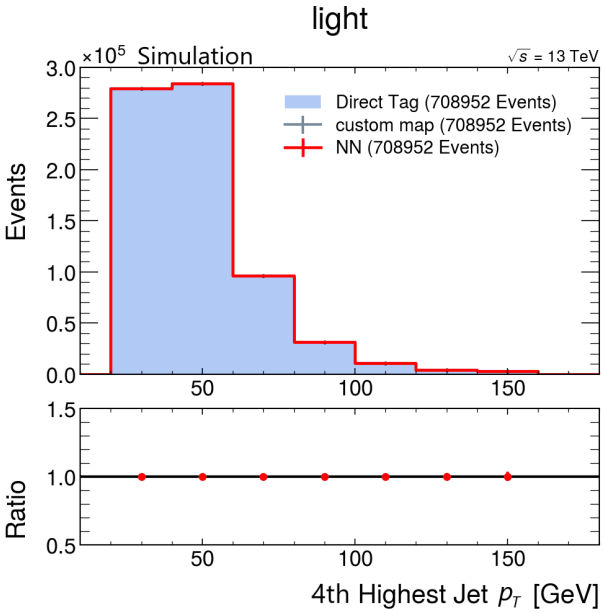

Figure 7.10.: Results for the "no Tag" WP in the 5-8 Jet region for 4th highest p_T jet, if that jet is a truth light jet. Also jets that are tagged at a higher WP are considered here. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainties.

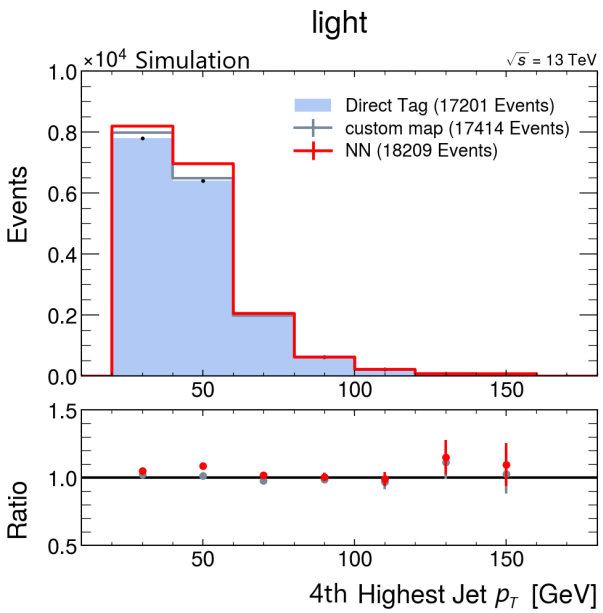

Figure 7.11.: Results for the 85% WP in the 5-8 Jet region for 4th highest p_T jet, if that jet is a truth light jet. Also jets that are tagged at a higher WP are considered here. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainties.

of the low statistics, but still agree with the direct tag. For jets with a p_T around 100 GeV, the predictions of the GNN and the efficiency map are within the uncertainty of each other and within the uncertainty of the direct tag predictions except for the 70% WP, where only the efficiency map lies within the uncertainty of the direct tag.

7.6. Results for different jet regions

To analyse the performance of the GNN on different jet multiplicity regions, predictions of the GNN were compared to those of the efficiency map and the direct tag for the 5, 6, 7, 8 and the full 5-8 jet regions. The results are shown in the figures 7.15 - 7.19. Investigated were the predictions for light jets for the 4th highest p_T jet. The chosen WP was 70%. Overall, it can be seen that the predictions of the GNN are clearly higher than those of the direct tag and the efficiency map for jets with a p_T below 60 GeV, whereas the predictions of the GNN are very accurate for jets with a p_T above 60 GeV within the 5,

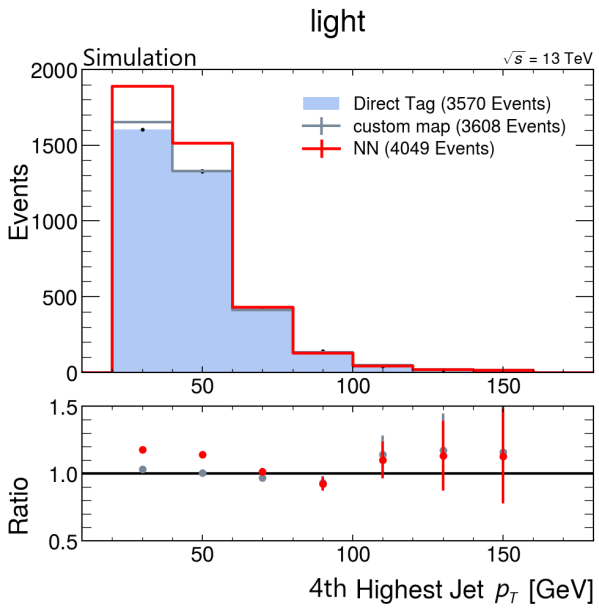

Figure 7.12.: Results for the 77% WP in the 5-8 Jet region for 4th highest p_T jet, if that jet is a truth light jet. Also jets that are tagged at a higher WP are considered here. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainties.

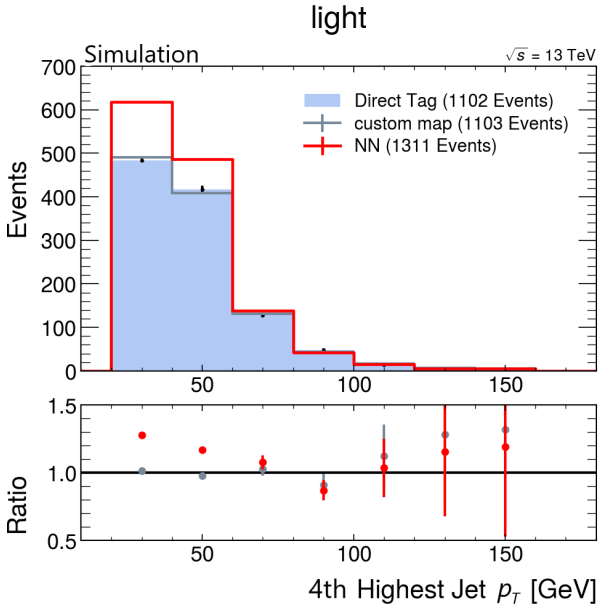

Figure 7.13.: Results for the 70% WP in the 5-8 Jet region for 4th highest p_T jet, if that jet is a truth light jet. Also jets that are tagged at a higher WP are considered here. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainties.

6 and 5-8 jet region. Here, the predictions mostly lie within the uncertainty of the direct tag. However, for the 5-8 jet region the predictions of the efficiency map are still more accurate than those of the GNN.

For the 7 jet and 8 jet regions, the predictions of the GNN and the efficiency map differ from those of the direct tag. In this regions, except for jets with a p_T around 50 GeV in the 8 jet region, the predictions of the efficiency map are closer to the direct tag than those of the GNN. The reasons for this behaviour will be discussed further in chapter 8.1.

7.7. B-jet multiplicity

During the analysis it turned out that the parameters for the assignment of the hadrons to the jets as described in chapter 6.2 are different from those that were used to assign the truth flavours used by b-tagging algorithms. The truth flavours were assigned by considering a cone of $\Delta R = 0.3$ and a cut for the hadron p_T was applied at 5 GeV. This

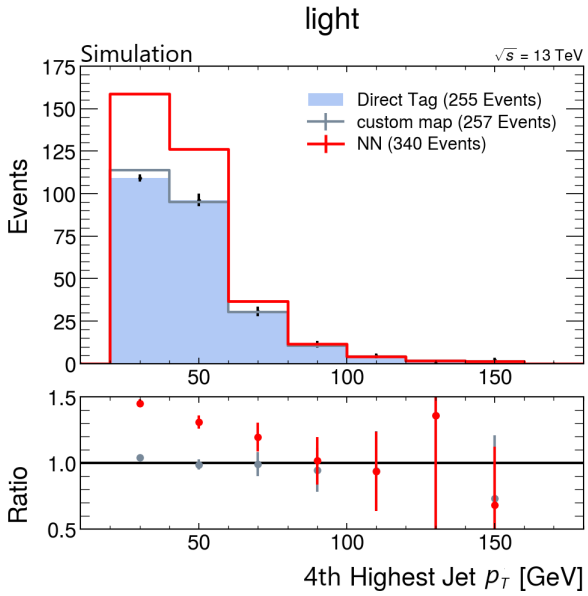

Figure 7.14.: Results for the 60% WP in the 5-8 Jet region for 4th highest p_T jet, if that jet is a truth light jet. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainties.

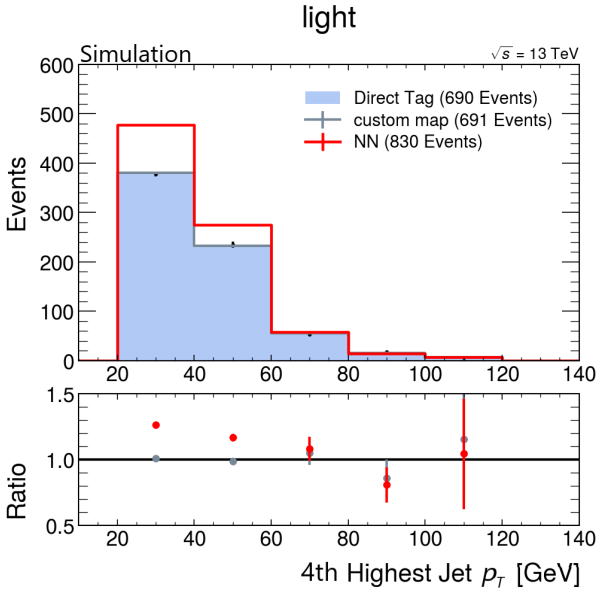

Figure 7.15.: Results for the prediction of 4th highest p_T jet, if that jet is a truth light jet, in the 5 Jet region for a 70% WP. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainty.

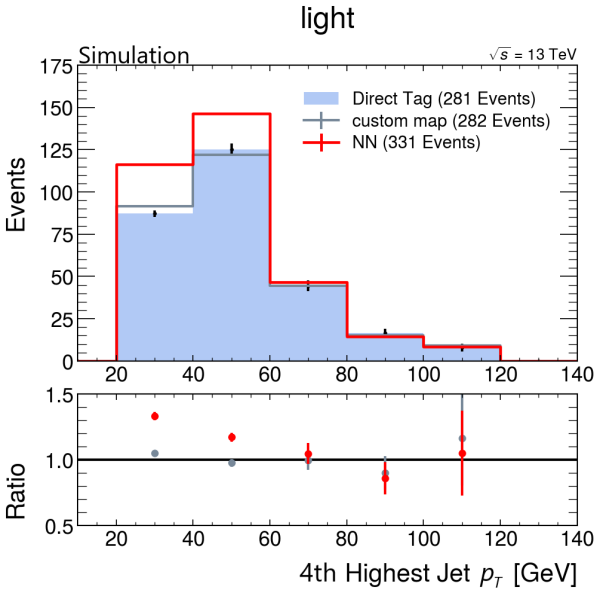

Figure 7.16.: Results for the prediction of 4th highest p_T jet, if that jet is a truth light jet, in the 6 Jet region for a 70% WP. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainty.

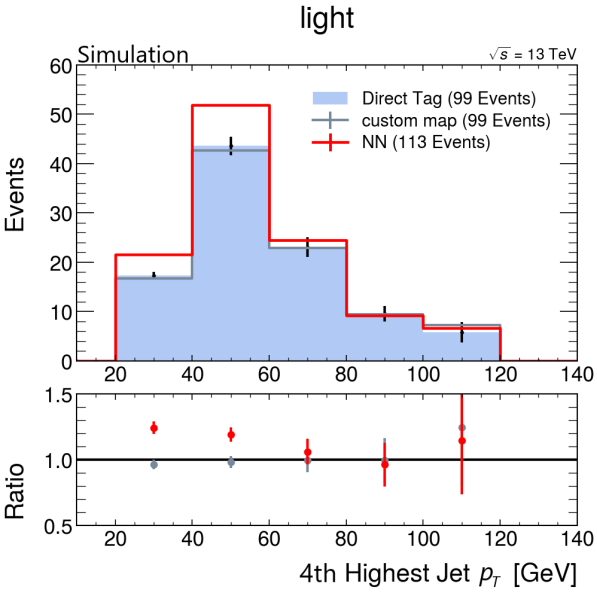

Figure 7.17.: Results for the prediction of 4th highest p_T jet, if that jet is a truth light jet, in the 7 Jet region for a 70% WP. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainty.

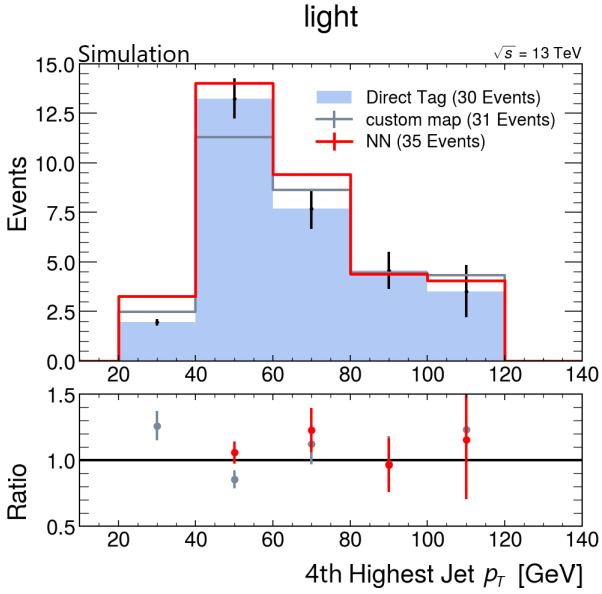

Figure 7.18.: Results for the prediction of 4th highest p_T jet, if that jet is a truth light jet, in the 8 Jet region for a 70% WP. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainty.

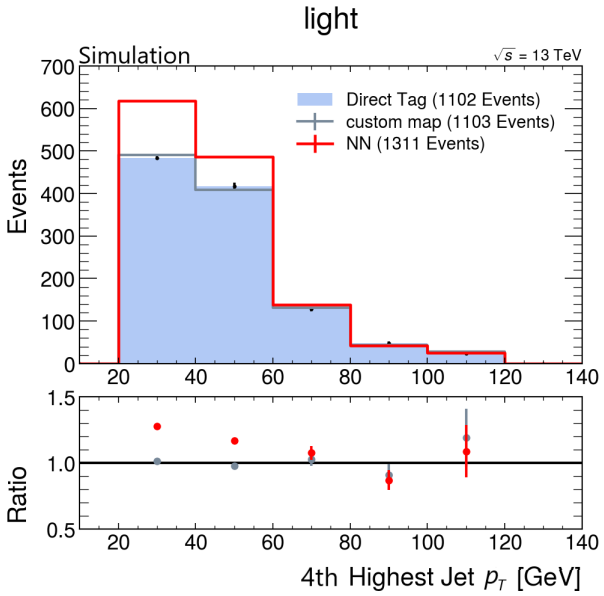

Figure 7.19.: Results for the prediction of 4th highest p_T jet, if that jet is a truth light jet, in the 5-8 Jet region for a 70% WP. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainty.

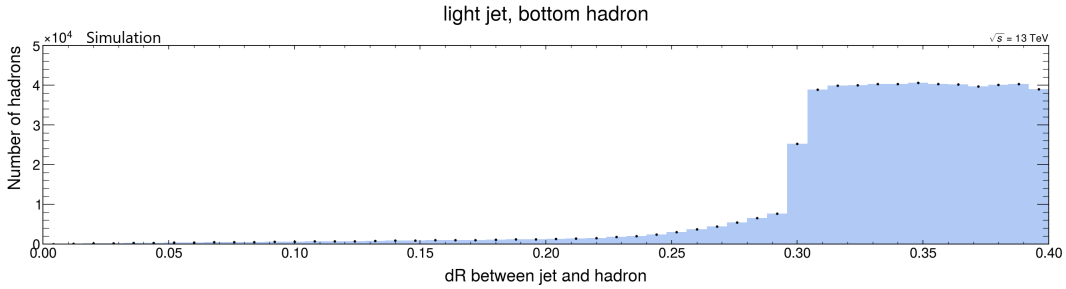

Figure 7.20.: Distribution of the angular distance ΔR between light jets and assigned b-hadrons. There is a jump at a angular distance of $\Delta R = 0.3$. Similiar results were observed for light jets with c-hadrons and c-jets with b-hadrons.

was noticed by analysing the distribution of the ΔR between the jets and the assigned hadrons. At a angular distance of $\Delta R = 0.3$, a jump in the distribution occurs, as can be seen in figure 7.20.

To analyse the effects of these different cuts, the b-jet multiplicity for the 70% WP in the 5-8 jet region was investigated for a ΔR cut at $\Delta R = 0.4$ and compared to the b-jet multiplicity if a ΔR cut was applied at $\Delta R = 0.3$ and a cut for the hadron p_T was applied at 5 GeV in the validation sample following the b-jet definition used by flavour tagging algorithms. Events with a jet, that does not satisfy the requirements, get rejected.

The results are shown in the figures 7.21 and 7.22. It can be seen that the b-jet multiplicity predicted by the NN, the efficiency map and the direct tag is almost identical for zero up to three tagged jets. It also does not matter whether the ΔR cut is applied at $\Delta R = 0.3$ or $\Delta R = 0.4$ and if a p_T cut is applied at 5 GeV or not. For four and five tagged jets, the NN predicts less events than the direct tag if the ΔR cut is applied at $\Delta R = 0.3$ and a p_T cut is applied at 5 GeV. Here, different permutations were considered as described in chapter 5.2 However, the NN predicts more events than the direct tag for four and five tagged jets if the ΔR cut is applied at $\Delta R = 0.4$ and there was no p_T cut applied. The efficiency map shows no change in its behaviour. For six tagged jets, the NN predicts far less events than the direct tag, no matter which cuts were applied. The efficiency map predictions fit to those of the direct tag but with a high uncertainty.

7.8. Comparison for different number of tags

A possible $t\bar{t}H(bb)$ analysis region could have at least three b tags at a 70% WP. To investigate how the GNN can deal with various numbers of tagged jets, the predictions of the NN were compared to those of the efficiency map and the direct tag for one, two

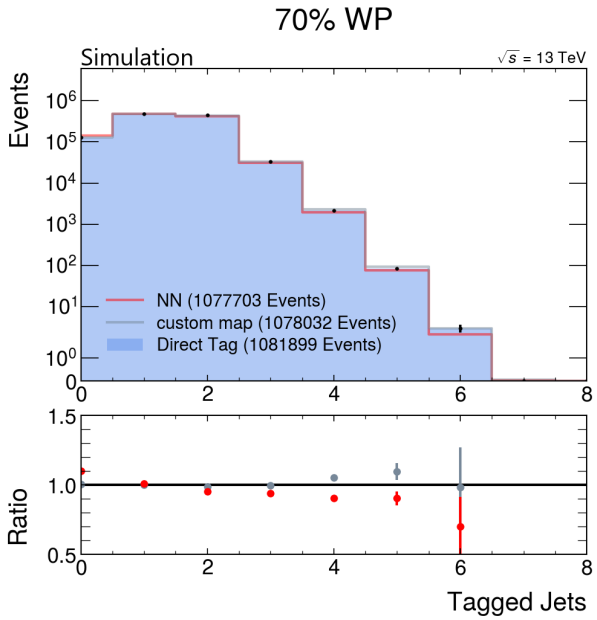

Figure 7.21.: B-jets multiplicity with a ΔR cut at $\Delta R = 0.3$ and a p_T cut at 5 GeV. The chosen WP is 70%. The errorbars indicate the statistical uncertainty. The y-axis is scaled logarithmic.

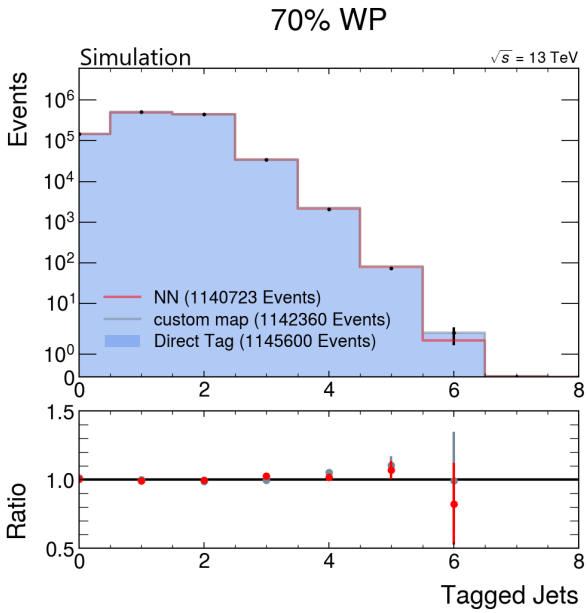

Figure 7.22.: B-jets multiplicity with a ΔR cut at $\Delta R = 0.4$ and without an p_T cut. The chosen WP is 70%. The errorbars indicate the statistical uncertainty. The y-axis is scaled logarithmic.

and three tagged jets. The hadrons were assigned to the jets using a cut for the angular distance at $\Delta R = 0.4$ and there was no cut for the hadron p_T. The predictions were made without considering the flavour of the jet to improve the statistics and make the results comparable. Furthermore, by considering all flavours the problem of the different cuts applied for the flavour assignment in the sample creation and the analysis as described in chapter 7.7 vanishes.

7.8.1. 1 tagged jet

The results for one tagged jet are shown in figure 7.23. Investigated were the predictions of the GNN, the efficiency map and the direct tag for the 5-8 jet region. The chosen WP is 70%. Investigated were the 4th highest p_T jets.
It can be seen that the predictions of the GNN and those of the efficiency map fit almost perfectly to those of the direct tag.

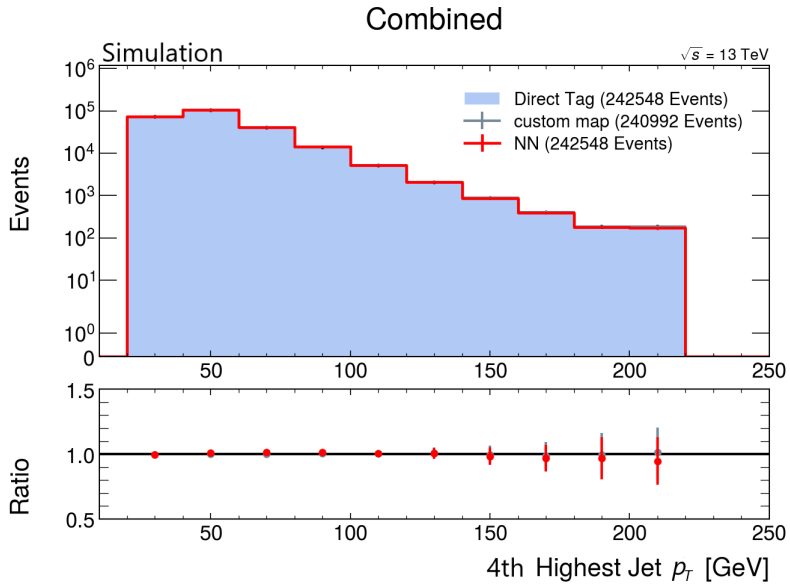

Figure 7.23.: Results for 1 tagged jet in the 5-8 jet region at a 70%WP. Investigated were the 4th highest p_T jets and no distinction was made for the jet flavour. The errorbars indicate the statistical uncertainty. The y-axis is scaled logarithmic. The last bin serves as overflow bin.

7.8.2. 2 tagged jets

The results for two tagged jets are shown in figure 7.24. Shown are the predictions of two tagged jets for the 5-8 jet region for a 70% WP. Investigated were the p_T of the 4th highest p_T jets. All possible permutations of tagged jets were considered as described in chapter 5.2.

It can be seen that the predictions of the NN and the efficiency map fit almost perfectly for jets with a p_T up to 150 GeV. For jets with a higher p_T, the predictions of the NN and the efficiency map differ more, but the NN and the efficiency map have a high statistical uncertainty because of the low statistics and the direct tag predictions lie within those statistical uncertainties.

7.8.3. 3 tagged jets

The results for three tagged jets are shown in figure 7.25. Shown are the predictions of the direct tag, the efficiency map and the GNN for the 5-8 jet region and a 70% WP. The predictions were made for the 4th highest p_T jets. All possible permutations of tagged jets were considered as described in chapter 5.2.

Overall the predictions of the NN fit to those of the direct tag. For jets with a p_T above 200 GeV the statistics becomes very low, resulting in a high uncertainty of the GNN and

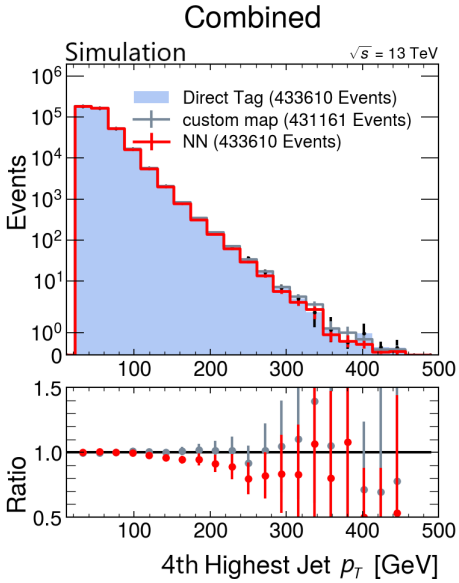

Figure 7.24.: Results for 2 tagged jets in the 5-8 jet region at a 70% WP. Investigated were the 4th highest p_T jets and no distinction was made for the jet flavours. The errorbars indicate the statistical uncertainty. The y-axis is scaled logarithmic.

the efficiency map, but the direct tag predictions lie within the uncertainty of the NN and the efficiency map for those jets. For jets with a p_T below 60 GeV, the predictions of the GNN and the efficiency map are higher than those of the direct tag. While this happens both for the GNN and the efficiency map, the differences between the GNN and the direct tag are higher than those of the efficiency map and the direct tag. This will be further discussed in chapter 8.2.

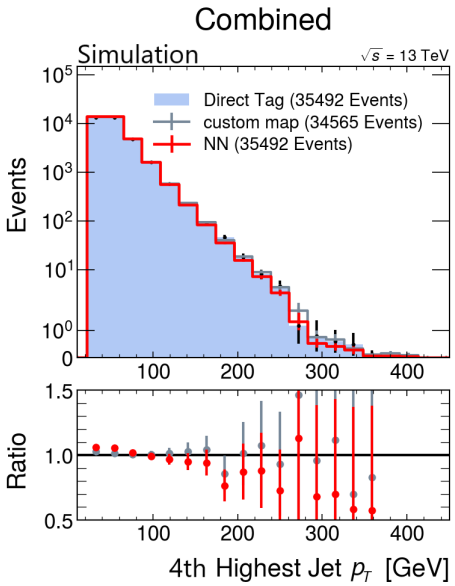

Figure 7.25.: Results for 3 tagged jets in the 5-8 jet region at a 70% WP. Investigated were the 4th highest p_T jets and no distinction was made for the jet flavours. The errorbars indicate the statistical uncertainty. The y-axis is scaled logarithmic.

8. Comparing GNN Predictions to Conventional Efficiency Maps

8.1. Dependency on statistics and jet number

The GNN seems to be widely independent of the available statistics. For example, one can see this by looking at the results for different WPs in chapter 7.5. In this analysis, there are low statistics available for jets with a p_T above 100 GeV. But still the predictions of the GNN and the efficiency map lie within the uncertainty of the direct tag. Similar results can be seen for the results in the 7 and 8 jet region, shown in the figures 7.17 and 7.18, where low statistics are present as well.
One might think that the GNN performs worse the more jets are within an event because of the additional angular distances ΔR. In this case, the GNN has to deal with more variables, resulting in more possible sources of error. But by looking at the results in chapter 7.6, one can see that the predictions of the GNN, if one takes the statistical uncertainty of the GNN into consideration, are as good in the 7 and 8 jet region as in the 5 and 6 jet regions. The predictions of the efficiency map and the GNN always lie within the statistical uncertainties of each other for jets with a p_T above 60 GeV. Furthermore the predictions of the GNN lie within the uncertainty of the direct tag in most cases for jets with a p_T above 60 GeV. The differences for jets with a p_T below 60 GeV will be discussed in chapter 8.2.
To conclude, the GNN can deal with low available statistics and performs comparably well for various number of jets within an event.

8.2. Behaviour above and below 60 GeV

Overall, it can be noticed that the performance of the NN is by far worse for light jets with a p_T below 60 GeV compared to those with a p_T above 60 GeV. For example, this can be seen in the figures 7.14 and 7.15. While this also happens for the efficiency map in some cases, as can be seen in figure 7.12, for example, the effect is much larger for the

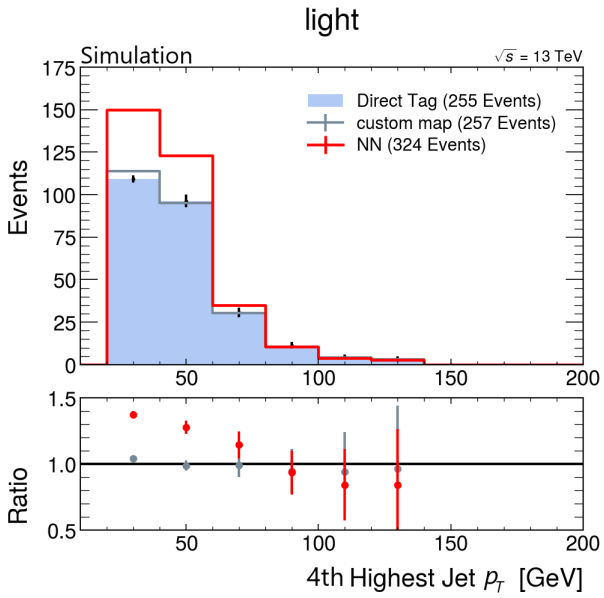

Figure 8.1.: Results by using the p_T matching method for an 60% WP in the full 5-8 jet region for 4th highest p_T jets, if that jet is a truth light jet. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainty.

GNN than for the efficiency map. A similar behaviour is observable for charm jets with a p_T between 50 GeV and 100 GeV, as can be seen in the appendix.
Since this behaviour does not occur to the same extent if the hadron variables are not implemented, as can be seen in figure 7.5, the problem seems to be the hadron variables. The specific problem of the hadron variables that causes this issue is not clear. One might think there is a problem with the assignment of the hadrons to the jets in this region. To check this, a different matching method, the p_T matching as described in chapter 6.2, was tested. The results for the same region as in figure 7.14 are shown in figure 8.1. It can be seen that the p_T matching method produces the same behaviour for the GNN as the ΔR matching method. Therefore the problem does not seem to be the ΔR matching itself.

Furthermore, different cuts for the hadron assignment were tested to check if this can improve the performance of the GNN. In the results shown in chapter 7, except chapter 7.7, there was no cut for the p_T of the hadron applied and the cut for the ΔR between the jet and the hadron was at $\Delta R = 0.4$, as described in chapter 6.2. However, when the sample was created, the truth flavours of the jets were assigned using a p_T cut of 5 GeV and a cut for the ΔR at 0.3. To check if this has led to the bad predictions of the GNN, the analysis was tested with the same cuts as during the creation of the sample. The results are shown in figure 8.2.
It can be seen that, while the direct tag and the efficiency map are predicting less events

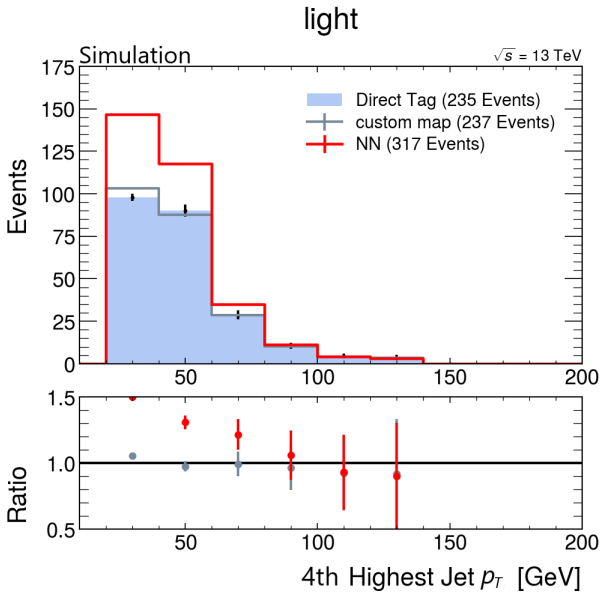

Figure 8.2.: Results by applying a p_T cut of 5 GeV and a cut of 0.3 for ΔR for an 60% WP in the 5-8 jet region for the 4th highest p_T jets, if that jet is a truth light jet. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainty.

than without the cut, the predictions of the NN do not change. The fact that the NN seems to not be affected by the applied cut could be a hint, that the NN is not working well. Furthermore, the problem that the NN predicts by far more events for jets with a p_T below 60 GeV than the direct tag and the efficiency map, still exists.
To conclude, the GNN can not deal well with light jets that have a p_T below 60 GeV. Similar problems occur for c-jets with a p_T between 50 GeV and 100 GeV, as can be seen in the appendix. The cause of this is not clear.

8.3. Results for a possible $t\bar{t}H(bb)$ analysis region

A possible analysis region for the $t\bar{t}H(bb)$ analysis could have at least 3 b-tags at a 70% WP for events with at least 5 jets. Unfortunately, since the GNN was only analysed for an exact number of tags, this region could not be investigated directly within this thesis. However, the results for exactly 3 tags at a 70 % WP within the 5-8 jet region represent a good approximation of this region. As can be seen in Figure 7.22, the number of events with 4 b-tags or more is low in comparison to the number of events with exactly 3 b-tags. Therefore, the region with at least 3 b-tags is dominated by the events with exactly 3 tags.
Figure 7.25 shows that the predictions of the GNN are in general very similar to those of

the direct tag. The predictions of the GNN and those of the efficiency map lie within the statistical uncertainty of the direct tag except for jets with a p_T below 60 GeV and jets with a p_T around 180 GeV, where only the predictions of the efficiency map lie within the statistical uncertainty of the GNN. For jets with a p_T below 60 GeV, the predictions of the GNN are higher than those of the direct tag. While this behaviour is a general problem of the GNN, as discussed in chapter 8.2, the effect for this region is lower than for the results discussed in chapter 8.2. But still, this is the greatest difference in the predictions of the NN, the efficiency map and of the direct tag.
A problem is that within this thesis, the truth flavours of the jets were assigned directly to the jets. However, within the $t\bar{t}H(bb)$ analysis, the HF classification is used, where the truth flavours are not assigned directly to the jets, but the number of jets with a certain flavour within an event gets counted. To minimise possible errors that can occur because of this difference, there was no distinction made for the jet flavour in figure 7.25. But it is still possible that the GNN behaves differently if the HF classification is used instead of the direct truth flavour assignment.
In conclusion it can be said that the GNN can imitate the behaviour of the direct tag for exactly three tags at a 70% WP in the 5-8 jet region except for jets with a p_T below 60 GeV and jets with a p_T around 180 GeV.

9. Summary and Outlook

The performance of a graph neural network (GNN) specialised on generating b-tagging efficiencies for truth tagging were presented for various jet regions, WPs, jets of different p_T within the event and number of tagged jets. Furthermore, the influence of hadron variables on the performance of the GNN and the ability of the GNN to predict b-jet multiplicities was analysed.
It was shown that the GNN can deal with comparatively low statistics and produces similar results for various numbers of jets between 5 jets and 8 jets. It was also shown that the GNN has a problem dealing with light jets that have a p_T below 60 GeV. Although it was observed that the implementation of the hadron variables aggravates this problem, the specific cause of this issue could not yet be found.
In the end, it was shown that the GNN can imitate the behaviour of the direct tag for three tags at a 70% WP in the 5-8 jet region, which is an approximation of a possible $t\bar{t}H(bb)$ analysis region, which requires at least three tags at a 70% WP and at least 5 jets within an event. However, this region could not be reconstructed fully. Furthermore, the assignment of truth flavours to the jets within this thesis varies with respect to the HF classification that is used in the $t\bar{t}H(bb)$ analysis.

Since the GNN performs well on an approximation of a possible $t\bar{t}H(bb)$ analysis region, it is likely that the approach of truth tagging with a graph neural network can be used for the $t\bar{t}H(bb)$ analysis. However, the GNN still needs to be tested on a complete reconstruction of a likely $t\bar{t}H(bb)$ analysis region and it needs to be investigated how the GNN performs with the HF classification instead of the direct assignment of the truth flavours. Furthermore, the cause of the problem that the GNN has with light jets that have a p_T below 60 GeV needs to be found.

Bibliography

[1] S. Abachi, et al. (DØ Collaboration), *Observation of the top quark*, Phys. Rev. Lett. **74**, 2632-2637 (1995).

[2] F. Abe, et al. (CDF Collaboration), *Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab*, Phys. Rev. Lett. **74**, 2626 (1995).

[3] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13**, 508 (1964).

[4] F. Englert, R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13**, 321 (1964).

[5] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716(1)**, 1 (2012).

[6] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716(1)**, 30 (2012).

[7] ATLAS Collaboration, *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, Phys. Lett. B **784**, 173 (2018).

[8] CMS Collaboration, *Observation of ttH Production*, Phys. Rev. Lett. **120**, 231801 (2018).

[9] ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector*, Eur. Phys. J. C **75** 7, 349 (2015).

[10] ATLAS Collaboration, *Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Phys. Rev. D **97** 7, 072016 (2018).

[11] ATLAS Collaboration, *Measurement of Higgs boson decay into b-quarks in associated production with a top-quark pair in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, CERN-EP-2021-**202** (Submitted to JHEP).

[12] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s}$* = 13 TeV, Eur. Phys. J. C **79** 11, 970 (2019).

[13] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, ATL-PHYS-PUB-2017-013 (2017).

[14] ATLAS Collaboration, *Tagging Rate Function B-Tagging*, ATL-PHYS-PUB-2007-011 (2007).

[15] F. A. Di Bello, et al., *Efficiency Parameterization with Neural Networks*, Comput. Softw. Big Sci. **5** 1, 14 (2021).

[16] C. N. Yang, R. L. Mills, *Conservation of Isotopic Spin and Isotopic Gauge Invariance*, Phys. Rev. **96**, 191 (1954).

[17] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19**, 1264 (1967).

[18] S. L. Glashow, *Partial-symmetries of weak interactions*, Nucl. Phys. **22(4)**, 579 (1961).

[19] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. C **680519**, 367 (1968).

[20] The Particle Data Group, R. L. Workman, et al., *The Review of Particle Physics (2022)*, PTEP **2022**, 083C01 (To be published in Prog. Theor. Exp. Phys. 2022).

[21] T. Humanic, *Extracting the hadronization timescale in $\sqrt{s}$ = 7 TeV proton-proton collisions from pion and kaon femtoscopy*, J. Phys. G: Nucl. Part. Phys. **41** 075105 (2014).

[22] M. Kobayashi, T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys., **49**, 2 (1973).

[23] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10**, 531 (1963).

[24] D. Clowe, et al., *A Direct Empirical Proof of the Existence of Dark Matter*, ApJ **648** L109 (2006).

[25] A. Sakharov, *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, JETP Letters **5** (1967).

[26] A. Gladyshev, D. Kazakov, *Supersymmetry and LHC*, Phys. Atom. Nuclei **70**, 1553 (2007).

[27] L. Evans, P. Bryant, *LHC Machine*, JINST **3**, S08001 (2008).

[28] S. Myers, *The LEP collider, from design to approval and commissioning*, CERN Yellow Reports: Monographs, CERN, Geneva (1991).

[29] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3**, S08003 (2008).

[30] CMS Collaboration, *The CMS Experiment at the CERN LHC*, JINST **3**, S08004 (2008).

[31] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3**, S08005 (2008).

[32] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, JINST **3**, S08002 (2008).

[33] CERN, *LHC Configuration and Operational Scenario for Run 3*, CERN-ACC-2021-0007 (2021).

[34] ATLAS Collaboration, *ATLAS inner detector: Technical design report. Vol.1* (1997).

[35] ATLAS Collaboration, *ATLAS Transition Radiation Tracker (TRT): Straw tubes for tracking and particle identification at the Large Hadron Collider*, Nucl. Instrum. Methods Phys. Res. A **845**, 257 (2017).

[36] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, CERN-LHCC-2010-013, ATLAS-TDR-19 (2010).

[37] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, Eur. Phys. J. C **77**, 317 (2017).

[38] R. Diestel, *Graph Theory*, Springer, 5. edition.

[39] F. Scarselli, et al., *The Graph Neural Network Model*, IEEE Transactions on Neural Networks, vol. **20**, 61 (2009).

[40] ATLAS Collaboration, *Performance of b-Jet Identification in the ATLAS Experiment*, JINST **11**, P04008 (2016).

[41] T. Sjöstrand, et al., *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191**, 159 (2015).

[42] S. Frixione, et al., *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11** 070 (2007).

[43] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70**, 823 (2010).

A. Further Plots

A.1. Influence of different sample splitting for charm jets

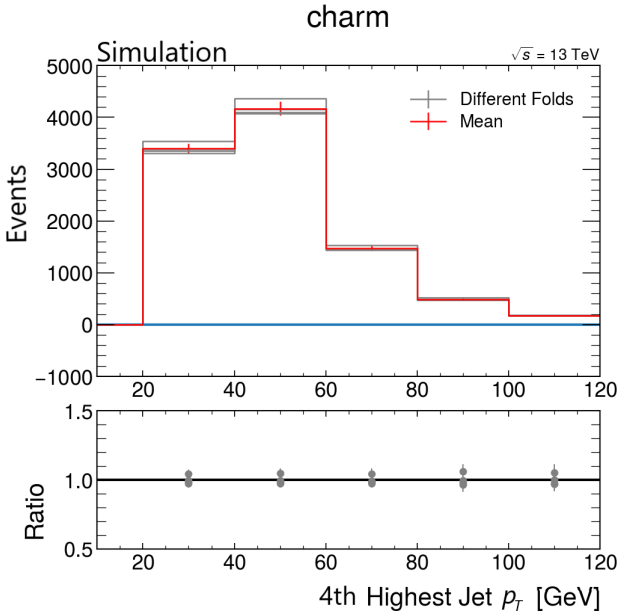

Figure A.1.: Comparison of the GNN predictions for different permutations of the sample parts for the 4th highest p_T truth charm jet in the 5-8 jet region at a 70% WP. The errorbars indicate the statistical uncertainties.

A.2. Predictions with and without hadron variables for charm jets

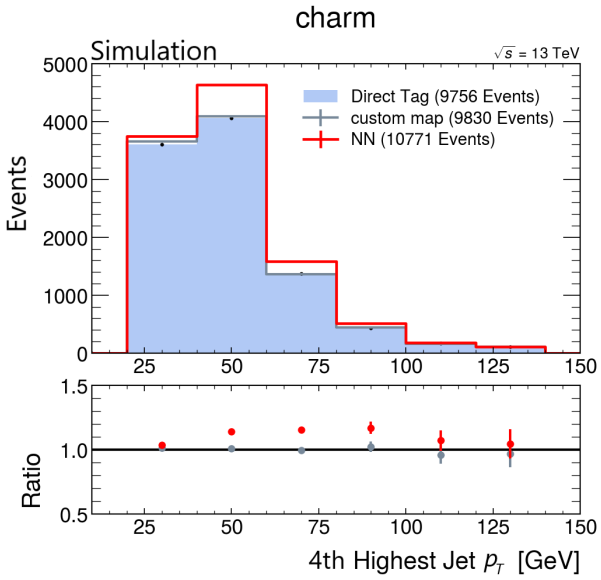

Figure A.2.: Results for the GNN with implemented hadron variables for the 4th highest p_T truth charm jet in the 5-8 jet region at a 70% WP. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainties.

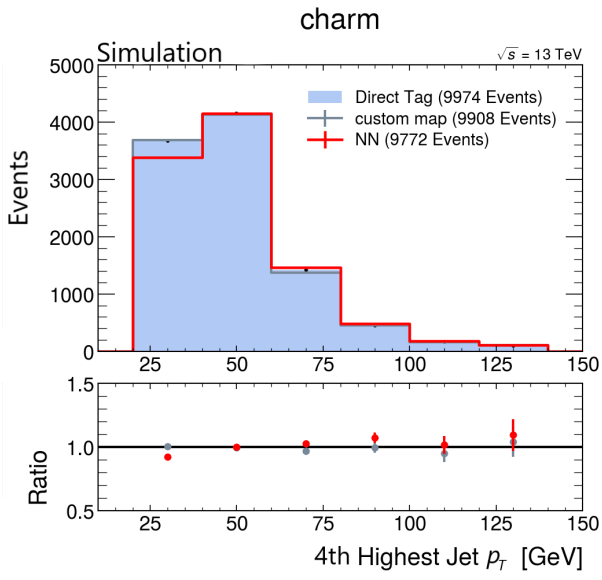

Figure A.3.: Results for the GNN without implemented hadron variables for the 4th highest p_T truth charm jet in the 5-8 jet region at a 70% W. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainties.

A.3. Comparison between leading and subleading jets for charm jets

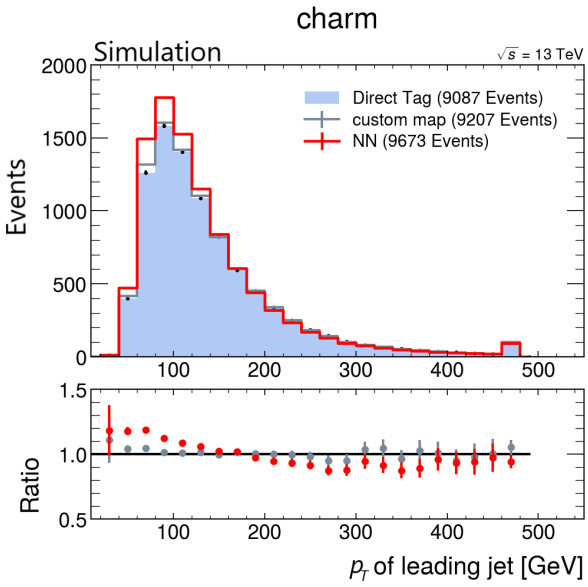

Figure A.4.: Results for the leading jet in the 5-8 jet region for a 70% WP. Shown are the predictions for truth charm jets. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainties.

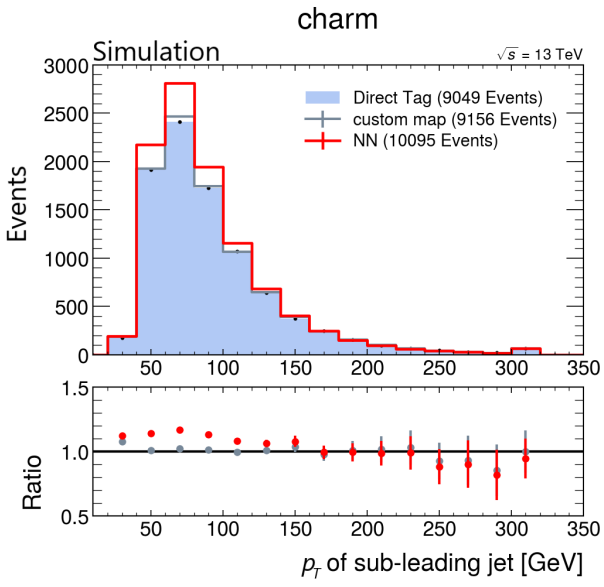

Figure A.5.: Results for the subleading jet in the 5-8 jet region for a 70% WP. Shown are the predictions for truth charm jets. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainties.

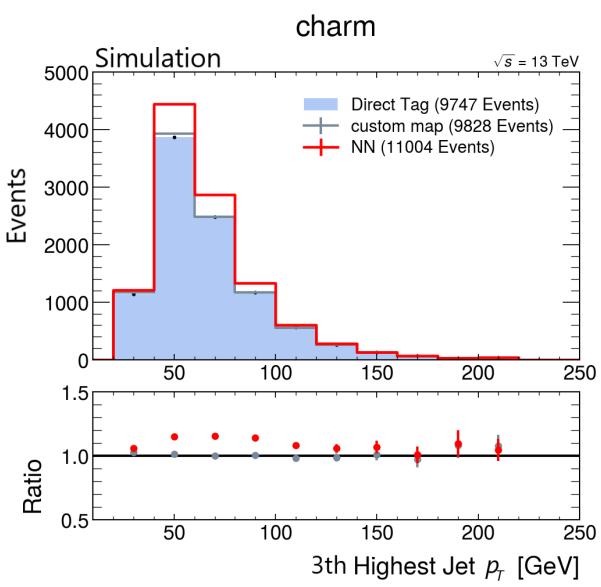

Figure A.6.: Results for the 3th leading jet in the 5-8 jet region for a 70% WP. Shown are the predictions for truth charm jets. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainties.

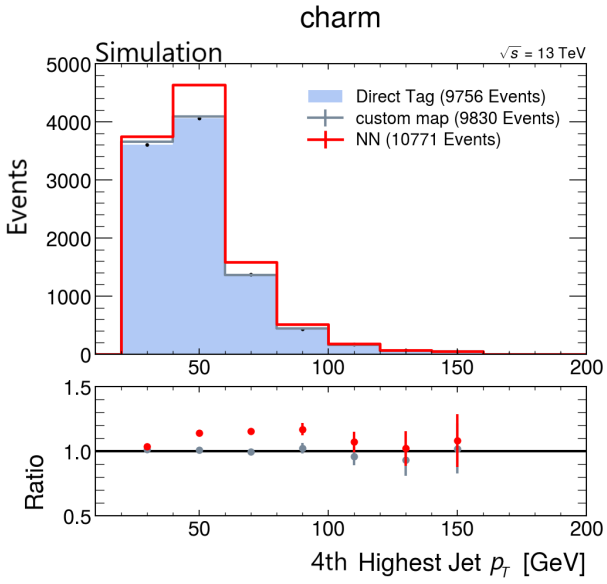

Figure A.7.: Results for the 4th leading jet in the 5-8 jet region for a 70% WP. Shown are the predictions for truth charm jets. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainties.

A.4. Results for different working points for charm jets

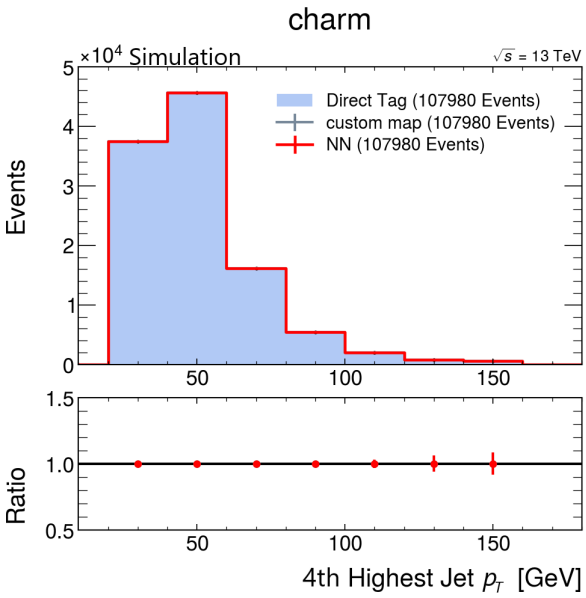

Figure A.8.: Results for the "no Tag" WP in the 5-8 Jet region for the 4th highest p_T truth charm jet. Also jets, that are tagged at a higher WP, are considered here. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainties.

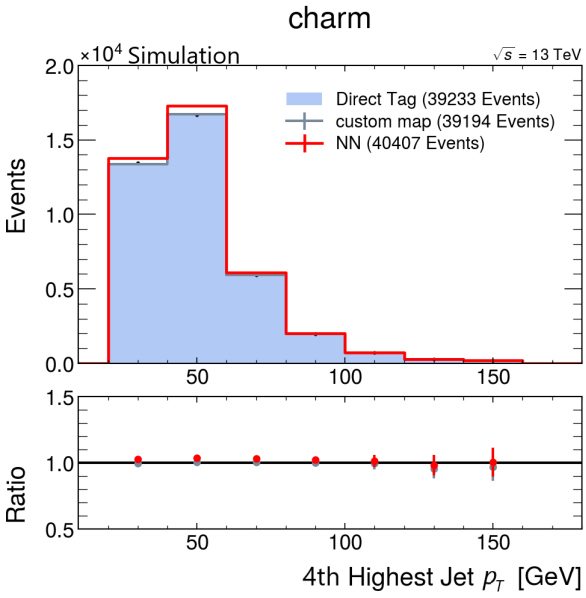

Figure A.9.: Results for the 85% WP in the 5-8 Jet region for the 4th highest p_T truth charm jet. Also jets, that are tagged at a higher WP, are considered here. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainties.

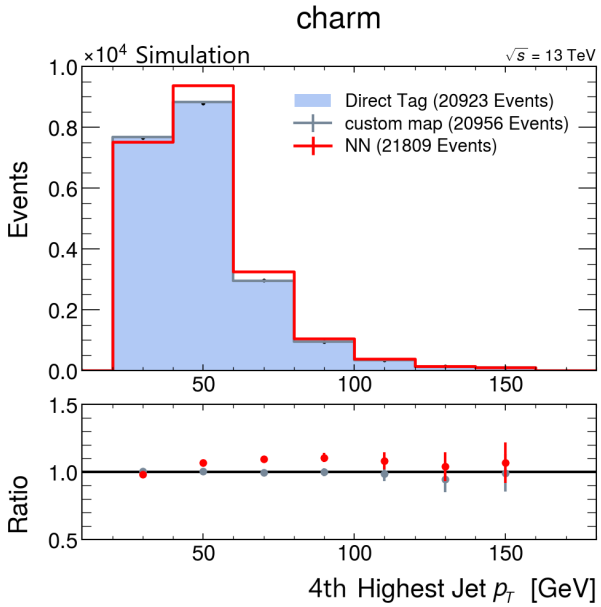

Figure A.10.: Results for the 77% WP in the 5-8 Jet region for the 4th highest p_T truth charm jet. Also jets, that are tagged at a higher WP, are considered here. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainties.

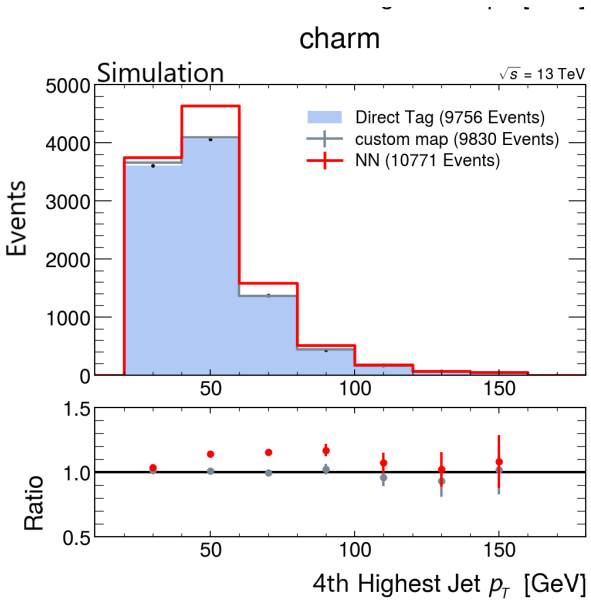

Figure A.11.: Results for the 70% WP in the 5-8 Jet region for the 4th highest p_T truth charm jet. Also jets, that are tagged at a higher WP, are considered here. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainties.

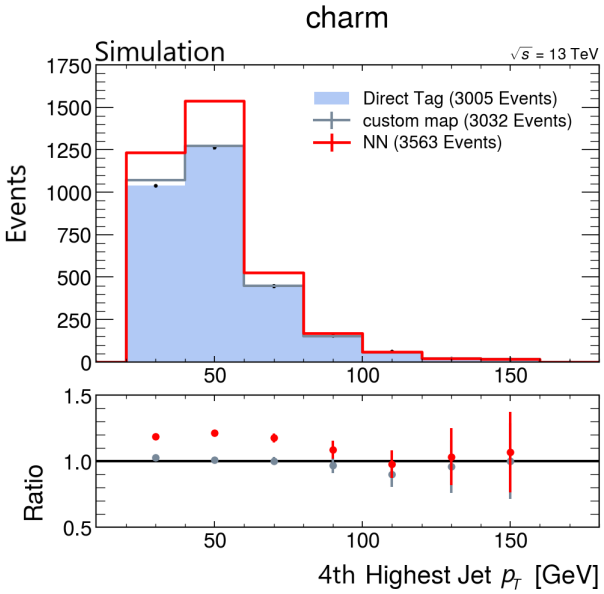

Figure A.12.: Results for the 60% WP in the 5-8 Jet region for the 4th highest p_T truth charm jet. Also jets, that are tagged at a higher WP, are considered here. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainties.

A.5. Results for different jet regions for charm jets

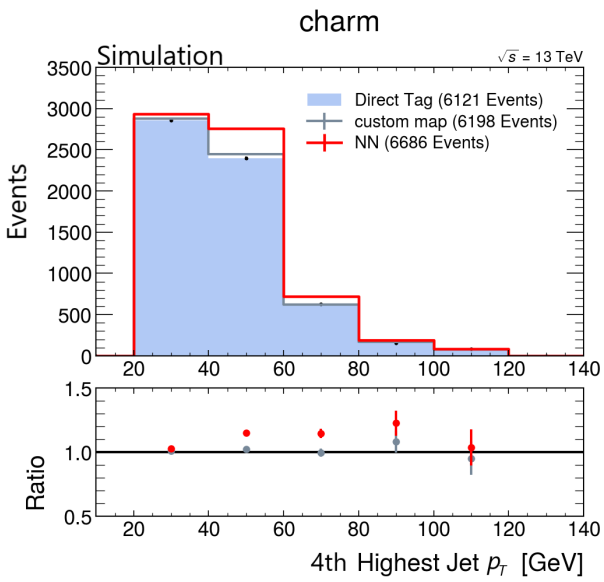

Figure A.13.: Results for the prediction of 4th highest p_T truth charm jets in the 5 Jet region for a 70%WP. The last bin serves as overflow bin. The error-bars indicate the statistical uncertainty.

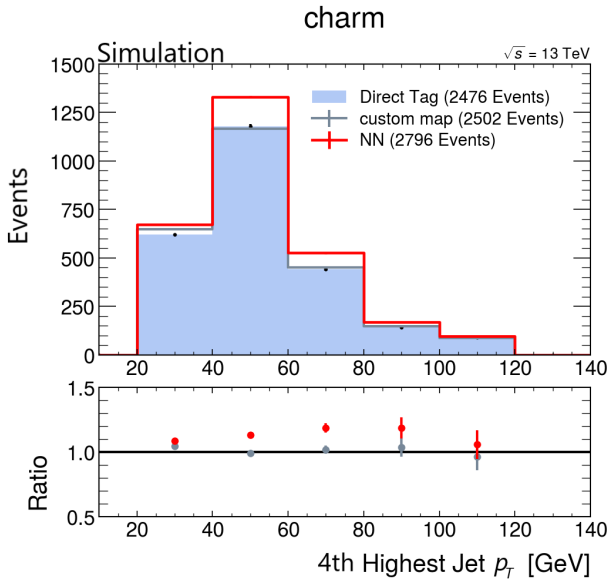

Figure A.14.: Results for the prediction of 4th highest p_T truth charm jets in the 6 Jet region for a 70%WP. The last bin serves as overflow bin. The error-bars indicate the statistical uncertainty.

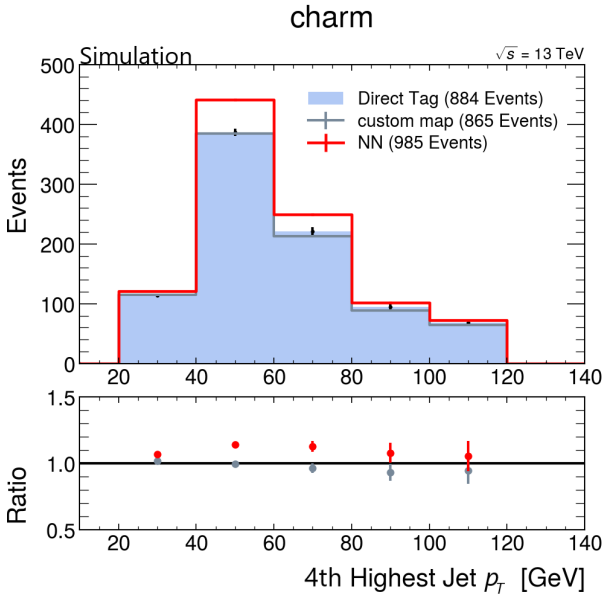

Figure A.15.: Results for the prediction of 4th highest p_T truth charm jets in the 7 Jet region for a 70%WP. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainty.

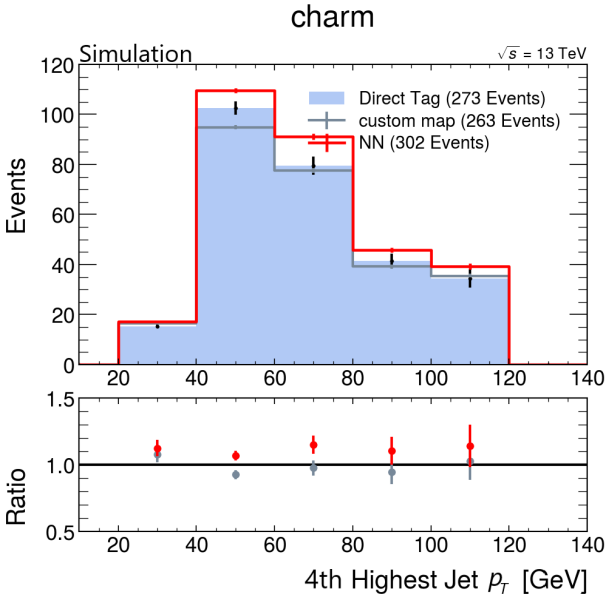

Figure A.16.: Results for the prediction of 4th highest p_T truth charm jets in the 8 Jet region for a 70%WP. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainty.

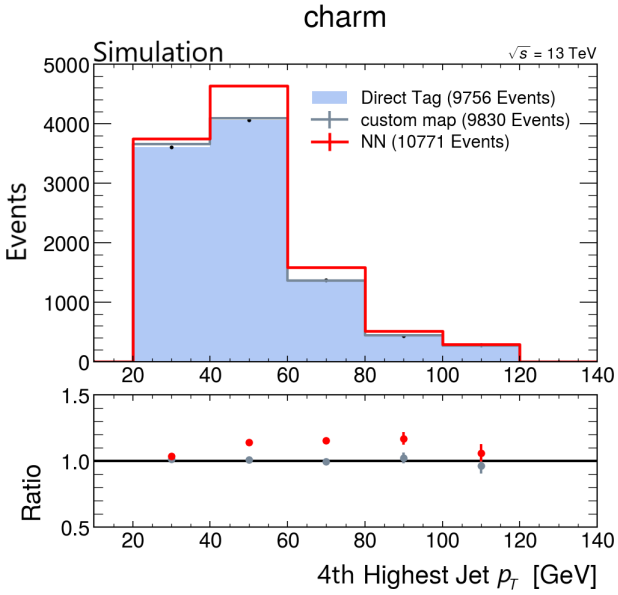

Figure A.17.: Results for the prediction of 4th highest p_T truth charm jets in the 5-8 Jet region for a 70%WP. The last bin serves as overflow bin. The errorbars indicate the statistical uncertainty.

A.6. B-jet multiplicity for other working points

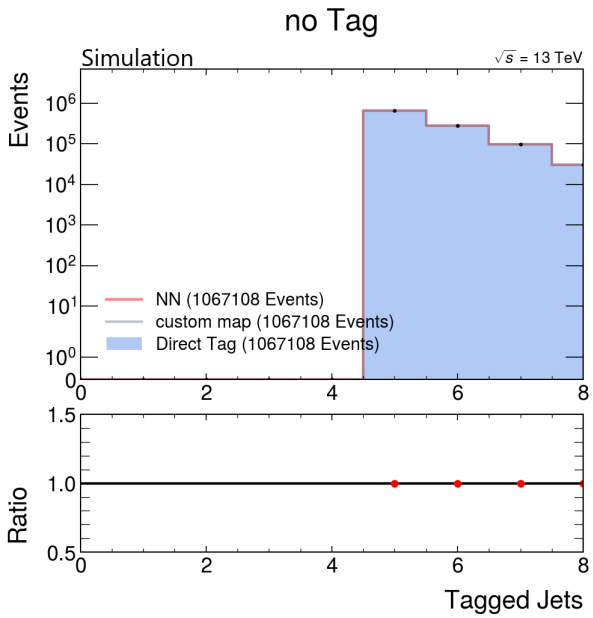

Figure A.18.: B-jets multiplicity with a ΔR cut at $\Delta R = 0.3$ and a cut for the hadron p_T at $p_T = 5\,\text{GeV}$. The chosen WP is "no Tag". The error-bars indicate the statistical uncertainty. The y-axis is scaled logarithmic.

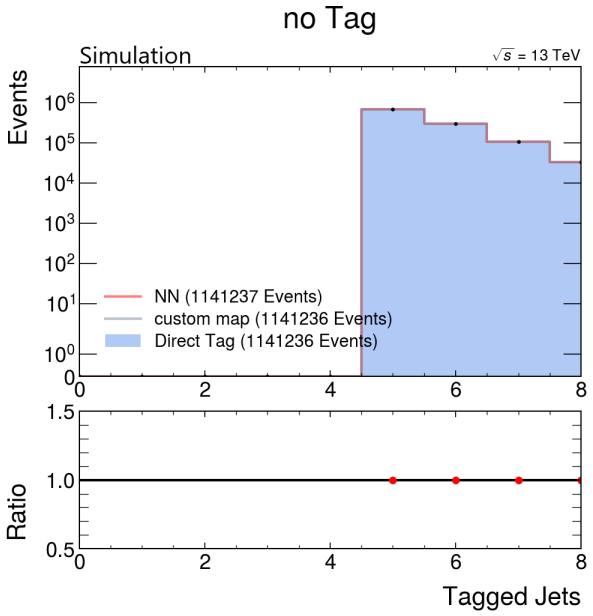

Figure A.19.: B-jets multiplicity with a ΔR cut at $\Delta R = 0.4$ and without an cut for the hadron p_T. The chosen WP is "no Tag". The error-bars indicate the statistical uncertainty. The y-axis is scaled logarithmic.

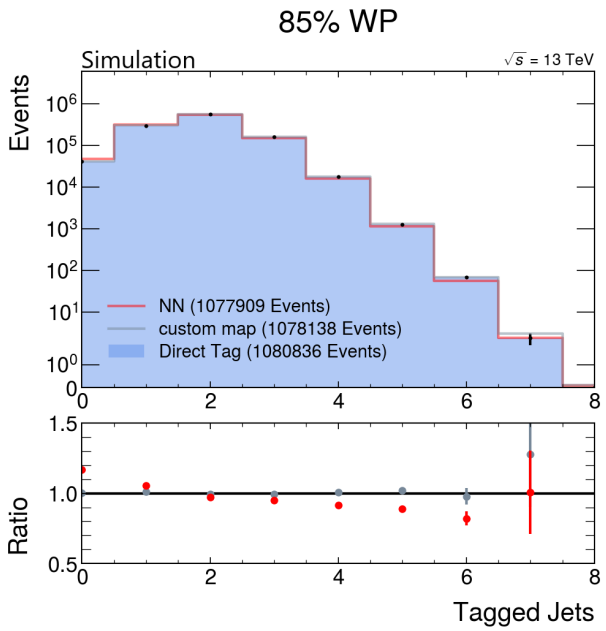

Figure A.20.: B-jets multiplicity with a ΔR cut at $\Delta R = 0.3$ and a cut for the hadron p_T at $p_T = 5\,\text{GeV}$. The chosen WP is 85%. The errorbars indicate the statistical uncertainty. The y-axis is scaled logarithmic.

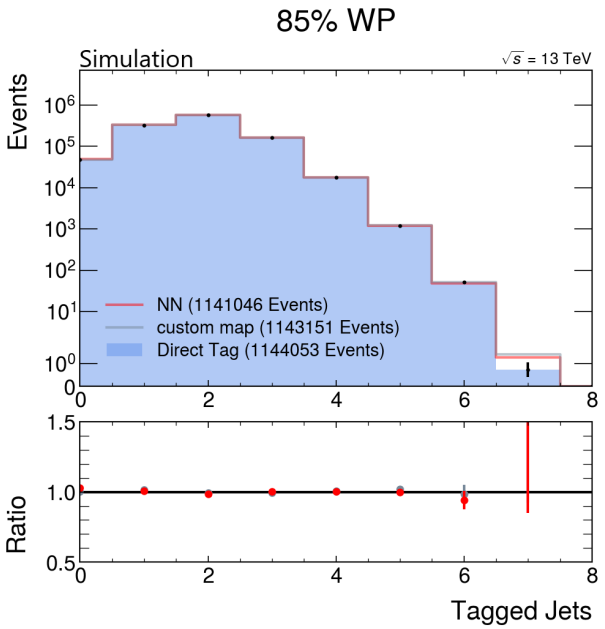

Figure A.21.: B-jets multiplicity with a ΔR cut at $\Delta R = 0.4$ and without an cut for the hadron p_T. The chosen WP is 85%. The errorbars indicate the statistical uncertainty. The y-axis is scaled logarithmic.

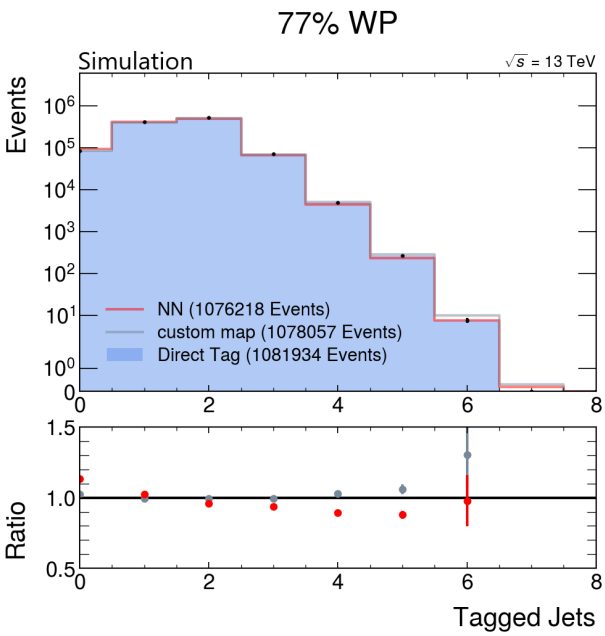

Figure A.22.: B-jets multiplicity with a ΔR cut at $\Delta R = 0.3$ and a cut for the hadron p_T at $p_T = 5\,\text{GeV}$. The chosen WP is 77%. The errorbars indicate the statistical uncertainty. The y-axis is scaled logarithmic.

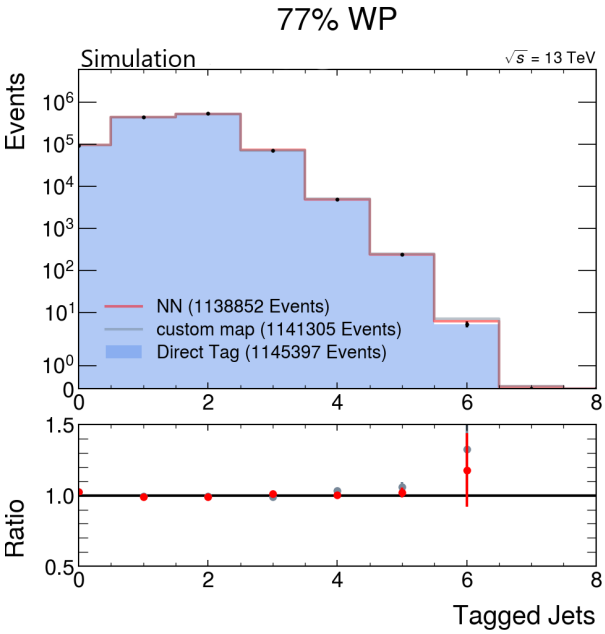

Figure A.23.: B-jets multiplicity with a ΔR cut at $\Delta R = 0.4$ and without an cut for the hadron p_T. The chosen WP is 77%. The errorbars indicate the statistical uncertainty. The y-axis is scaled logarithmic.

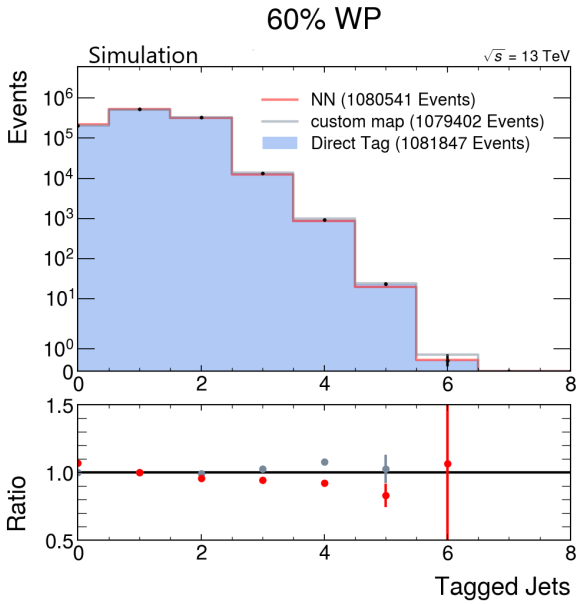

Figure A.24.: B-jets multiplicity with a ΔR cut at $\Delta R = 0.3$ and a cut for the hadron p_T at $p_T = 5\,\text{GeV}$. The chosen WP is 60%. The errorbars indicate the statistical uncertainty. The y-axis is scaled logarithmic.

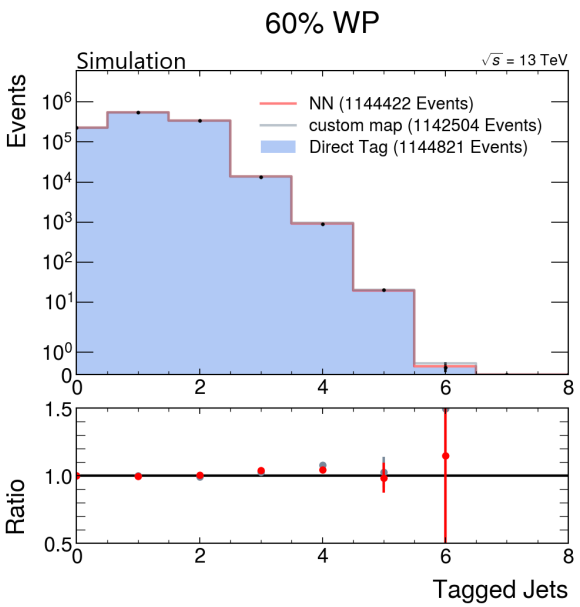

Figure A.25.: B-jets multiplicity with a ΔR cut at $\Delta R = 0.4$ and without an cut for the hadron p_T. The chosen WP is 60%. The errorbars indicate the statistical uncertainty. The y-axis is scaled logarithmic.

B. List of GNN input variables

Variable	Description	Name in the code
	Jet variables	
p_T	Momentum of the jet	jet_pt
η	Pseudorapidity of the jet	jet_eta
φ	Azimuthal angle of the jet	jet_phi
Flavour	Truth flavour of the jet	jet_truthflav
	Hadron variables	
p_T	Momentum of the hadron	jet_hadron_dRmatch_pt
η	Pseudorapidity of the hadron	jet_hadron_dRmatch_eta
φ	Azimuthal angle of the hadron	jet_hadron_dRmatch_phi
Energy	Energy of the hadron	jet_hadron_dRmatch_e
Mass	Mass of the hadron	jet_hadron_dRmatch_m
ΔR	Angular distance between jet and assigned hadron	jet_hadron_dRmatch_dR
Flavour	Truth flavour of the hadron	jet_hadron_dRmatch_flav
	Event variables	
ΔR	Angular distance ΔR between jets	deltaR

Table B.1.: Table of variables, which are used for the training of the GNN.

Danksagung

First of all, I want to thank Prof. Dr. Arnulf Quadt for the opportunity to write my bachelor thesis in his working group and for the supervision of my thesis. Also I want to say "thank you" to the whole group for their help and suggestions during the weekly meetings. Furthermore, I want to thank Dr. Elizaveta Shabalina for her support and advice, may it be for further analyses or solving some occured problems, and Prof. Dr. Steffen Schumann for his time to be my second referee.

And lastly, of course, a very big "thank you" to Chris for your inexhaustible support and advice during my whole thesis! I always knew that I could come to you whatever question I had, may it be at the beginning of the week or at saturday evening, and I am very grateful for that.

Erklärung nach §13(9) der Prüfungsordnung für den Bachelor-Studiengang Physik und den Master-Studiengang Physik an der Universität Göttingen:
Hiermit erkläre ich, dass ich diese Abschlussarbeit selbständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe.
Darüberhinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht auszugsweise, im Rahmen einer nichtbestandenen Prüfung an dieser oder einer anderen Hochschule eingereicht wurde.

Göttingen, den 9. Dezember 2022

(Dennis Rodermund)