

Rekonstruktion von $H \rightarrow \tau\tau$ Ereignissen mit maschinellem Lernen

Jurek Eisinger

Bachelorarbeit in Physik
angefertigt im Physikalischen Institut

vorgelegt der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
Bonn

September 2021

Ich versichere, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie die Zitate kenntlich gemacht habe.

Bonn, 28.09.2021
Datum

..... 
Unterschrift

1. Gutachter: Prof. Dr. Klaus Desch
2. Gutachter: Prof. Dr. Jochen Dingfelder

Danksagung

Ich möchte mich bei Prof. Dr. Klaus Desch für die Möglichkeit bedanken, meine Bachelorarbeit in seiner Arbeitsgruppe zu schreiben. Weiterhin möchte ich mich bei Priv.-Doz. Dr. Philip Bechtle und Dr. Christian Grefe für die hilfreichen Gespräche im Morning Coffee und die Unterstützung im Allgemeinen bedanken. Mein Dank gilt besonders Lena Herrmann für die zahlreichen beantworteten Fragen und die großartige Unterstützung! Danke auch dir Paul, für das Korrekturlesen. Und danke an die gesamte Arbeitsgruppe für das nette Arbeitsklima und das schöne Summer Jamboree. Mein abschließender Dank gilt meiner Familie und meinen Freunden.

Inhaltsverzeichnis

1 Einleitung	1
2 Physikalische Grundlagen	3
2.1 Das Standardmodell der Teilchenphysik	3
2.1.1 Wechselwirkungen	3
2.1.2 Quarks	4
2.1.3 Leptonen	5
2.1.4 Higgs-Boson	5
2.2 Das Tau-Lepton	6
2.3 Zerfallskinematik	6
2.3.1 Helizität und Polarisationsasymmetrie	6
2.3.2 Rekonstruktion unbekannter Größen	7
3 LHC und ATLAS	9
3.1 Der LHC	9
3.2 ATLAS	10
3.2.1 Innerer Detektor	11
3.2.2 Kalorimeter	11
3.2.3 Myon-Detektor	12
3.3 Das ATLAS-Koordinatensystem	12
3.4 Der Missing-Mass-Calculator	13
4 Künstliche neuronale Netze	15
4.1 Grundlagen des Maschinellen Lernens	15
4.2 Funktionsweise	16
4.3 Der Lernvorgang	17
4.4 Hyperparameter	19
4.5 Software	19
5 Daten	20
5.1 Simulationen	20
5.2 Allgemeines	20
5.3 Vorselektion	23
5.3.1 Z-Datensatz	23
5.3.2 Trainingsdatensatz	23

6	Ergebnisse vorheriger Arbeiten	24
6.1	Erklärung	24
6.2	Physikalischer Zusammenhang	26
7	Training und Netzarchitektur	30
7.1	Ein- und Ausgabedaten	30
7.2	Verarbeitung der Daten	33
7.3	Hyperparameter	34
7.4	Methoden der Dekorrelation	34
7.4.1	Adversarial neural network	34
7.4.2	Distance Correlation	36
8	Ergebnisse	38
8.1	Relative Massenauflösung	38
8.2	Mögliche Ursachen	42
8.3	Relative Massenauflösung in Intervallen der wahren Masse	45
8.4	Kalibrierung	47
8.5	Distance Correlation	49
8.6	Weiterführende Studien	52
9	Fazit	53
A	Anhang	54
Literatur		61
Abbildungsverzeichnis		65
Tabellenverzeichnis		68

KAPITEL 1

Einleitung

In der Physik werden grundlegende Naturphänomene untersucht. Mathematische Modelle werden konstruiert, um verschiedensten Fragestellungen in allen Bereichen der Physik nachzugehen. Diese Bachelorarbeit bewegt sich im Gebiet der Teilchenphysik. Die Teilchenphysik befasst sich mit elementaren Teilchen und deren Wechselwirkungen untereinander. Zur Analyse des Standardmodells der Teilchenphysik, das all diese elementaren Teilchen und Wechselwirkungen zusammenfasst, werden im größten Teilchenbeschleuniger der Welt, dem LHC (**L**arge **H**adron **C**ollider) Protonen mit enormen Geschwindigkeiten zum Kollidieren gebracht und deren Zerfallsprodukte unter Anderem im ATLAS-Detektor untersucht. So wurde 2012 das Higgs-Boson nachgewiesen [1]. Das Higgs-Boson ist noch heute einer der zentralen Forschungsschwerpunkte am LHC. Die Kopplung des Higgs-Bosons an elementare Teilchen ist ein zentraler Aspekt der Analyse. Es soll sich in dieser Arbeit auf den Zerfall des Higgs-Bosons in zwei Tau-Leptonen konzentriert werden. Das Higgs-Boson zerfällt in etwa 6 % [2] der Fälle in zwei Tau-Leptonen, was für das Studium der Interaktion des Higgs-Bosons mit Leptonen der geeignetste Zerfall ist. Allerdings zerfällt das Z-Boson ebenso in zwei Tau-Leptonen. Um das Mutterteilchen bei diesen Zerfallsprozessen ermitteln zu können, muss die invariante Masse des $\tau^-\tau^+$ -Systems (kurz: Di-Tau Masse) ermittelt werden. Dafür wird aktuell der Missing-Mass-Calculator genutzt [3]. Da dieser viel Rechnerkapazität erfordert, wird nach Alternativen der Massenrekonstruktion mit Künstlichen Neuronalen Netzen (KNNs) geforscht. Diese Analyse wurde in vorherigen Arbeiten durchgeführt und aus den Ergebnissen (die in Abschnitt 6 näher vorgestellt werden) ist die Hypothese entstanden, dass die vorhergesagte Di-Tau Massenverteilung abhängig von der Polarisationsasymmetrie A_{pol} (siehe Abschnitt 2) der Tau-Leptonen, sowie vom Spin des Mutterteilchens ist. Es soll in dieser Arbeit überprüft werden, welchen Anteil die Polarisationsasymmetrie der Tau-Leptonen am Einfluss auf die Massenvorhersage des neuronalen Netzes hat. Anschließend soll das Netz gegen diesen Einfluss desensibilisiert werden. Allerdings soll die Vorhersage des Netzes für die tatsächliche Anwendung nicht unabhängig von der Polarisationsasymmetrie sein. Die Vorhersage des Netzes soll unabhängig vom Spin des Mutterteilchens im Zerfall (also Higgs oder Z) sein. Die Dekorrelation dieser Größen kann zum jetzigen Zeitpunkt noch nicht durchgeführt werden. Es wird ein Datensatz für das Training des neuronalen Netzes benötigt, dessen Zerfall ein Mutterteilchen mit Spin null zu Grunde liegt. Ein solcher existiert aktuell nicht. Der in dieser Arbeit betrachtete Desensibilisierungsprozess kann als Vorstudie zur angestrebten Dekorrelation von Spin und Masse angesehen werden.

In Abschnitt 2 werden allgemein relevante physikalische Zusammenhänge dargestellt. Anschließend wird in Abschnitt 3 der LHC, sowie der ATLAS-Detektor beschrieben. In Abschnitt 4 wird das Prinzip

Kapitel 1 Einleitung

künstlicher neuronaler Netze eingeführt und auf wichtige Größen eingegangen. Die Verwendung des Trainingsdatensatz mit dem das neuronale Netz in dieser Arbeit trainiert werden soll wird in Abschnitt 5 motiviert und die Vorselektion beschrieben. In Abschnitt 6 werden Ergebnisse vorheriger Studien in diesem Gebiet besprochen. Hier wird die Idee dieser Bachelorarbeit tiefergehend motiviert. In Abschnitt 7 wird die Architektur, sowie verwendete Hyperparameter des neuronalen Netzes beschrieben. In Abschnitt 8 werden wichtige Darstellungsmöglichkeiten besprochen, die der Analyse der Vorhersage des neuronalen Netzes dienen. Die Ergebnisse werden zusammengefasst, interpretiert und Vorschläge für mögliche weitere Arbeiten formuliert. Schließlich werden in Abschnitt 9 die wichtigsten Punkte noch einmal aufgegriffen.

KAPITEL 2

Physikalische Grundlagen

In diesem Abschnitt werden physikalische Zusammenhänge vorgestellt, die relevant für die Arbeit sind. Es werden das Standardmodell der Teilchenphysik, insbesondere Higgs-Boson und Tau-Lepton, sowie zerfallskinematische Größen wie Helizität und Polarisationsasymmetrie eingeführt. In der Übersicht wird sich, wenn nicht anders gekennzeichnet, an [4] orientiert.

2.1 Das Standardmodell der Teilchenphysik

Das Standardmodell der Teilchenphysik (siehe Abbildung 2.1) fasst alle bis jetzt bekannten Elementarteilchen, sowie die zwischen ihnen herrschenden Wechselwirkungen zusammen. Diese Teilchen sind einerseits die Fermionen. Fermionen haben eine Spin von $1/2$ und es lässt sich jedem Fermion ein entsprechendes Antifermion mit gleicher Masse, aber (falls elektrisch geladen) entgegengesetzter elektrischer Ladung zuordnen. Andererseits sind im Standardmodell die Eichbosonen (auch Vektorbosonen genannt) mit einem Spin von 1 , und das Higgsboson (wegen seines Spin 0 wird es als Skalarboson bezeichnet) aufgeführt.

2.1.1 Wechselwirkungen

Neben der Gravitation, die im Standardmodell nicht berücksichtigt wird, sind drei elementare Wechselwirkungen bekannt. Jede dieser Wechselwirkungen wird durch einen Austausch der entsprechenden Eichbosonen beschrieben. Das Austauschteilchen der starken Wechselwirkung ist das Gluon. Es gibt acht verschiedene Gluonen, die jeweils eine Farbladung, zusammengesetzt aus einer Farbe und einer Antifarbe, tragen. Die Farbladung hat weder mit der optischen Farbe, noch mit elektrischer Ladung zu tun. Sie ist vielmehr die Eigenschaft eines Teilchens im Bezug auf die starke Wechselwirkung. Die Zustände, in denen sich ein stark wechselwirkendes Teilchen befinden kann, werden als rot, blau und grün gekennzeichnet. Allen diesen Farben werden entsprechende Antifarben (antirot, antiblau, antigrün) zugeordnet. Lediglich Gluonen und Quarks tragen Farbladung, allein sie können demnach stark wechselwirken. In Einklang mit der Optik wird ein Teilchen als farbneutral bezeichnet, das aus den drei Farben (rot, grün, blau), aus den drei Antifarben, oder aus Farben in Kombination mit der jeweiligen Antifarbe besteht. In der Natur werden ausschließlich farbneutrale Objekte beobachtet. Dieses Phänomen wird als Color-Confinement bezeichnet. Da Gluonen die Austauschteilchen der starken Wechselwirkung sind, hat diese aufgrund des Color-Confinements eine Reichweite, die auf ca. 1 fm

Standard-Modell der Elementarteilchen

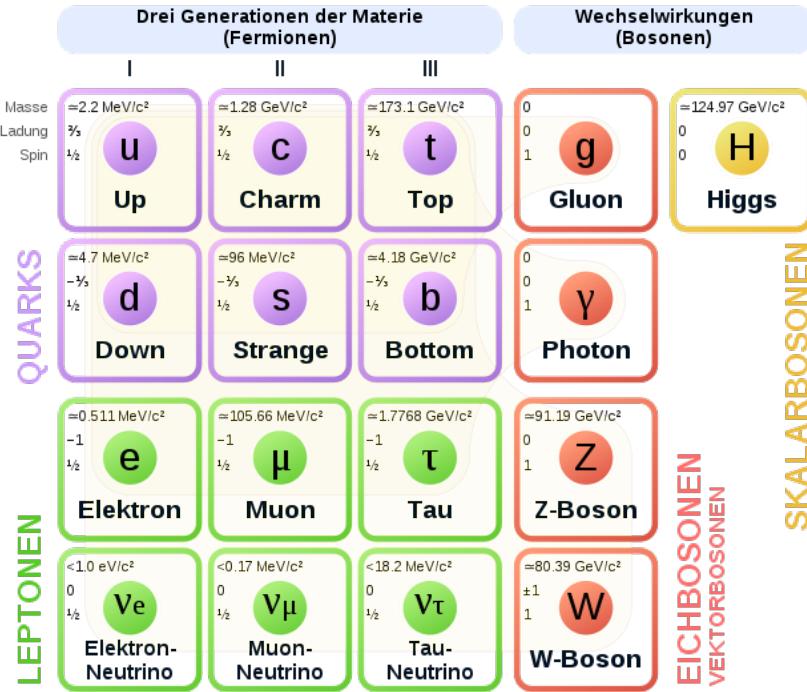


Abbildung 2.1: Standardmodell der Teilchenphysik [5]

beschränkt ist. Gluonen tragen keine elektrische Ladung und werden im Rahmen des Standardmodells als masselos angenommen.

Die Austauschteilchen der schwachen Wechselwirkung sind die W^\pm -Bosonen mit einer elektrischen Ladung von $\pm e$ mit der Elementarladung e und einem Spin von eins, sowie das elektrisch neutrale Z-Boson, ebenfalls mit einem Spin von eins. Sie haben Massen von $M_{W^\pm} = 80,38 \text{ GeV}$ [6] und $M_Z = 91,19 \text{ GeV}$ [6], sind damit also verhältnismäßig schwer. Nach der Heisenberg'schen Unschärferelation können die Größen Energie und Zeit nicht gleichzeitig beliebig genau gemessen werden. Dieser Zusammenhang ist auch als „Energie-Zeit-Unschärfe“ bekannt. Da Masse und Energie äquivalent sind, können die Austauschteilchen der schwachen Wechselwirkung nur sehr kurz existieren, weswegen deren Reichweite ausgesprochen klein ist. Schwache Ladung tragen Leptonen und Quarks.

Das Austauschteilchen der elektromagnetischen Wechselwirkung ist das Photon. Es trägt keine elektrische Ladung, ist masselos und hat einen Spin von eins. Quarks und Leptonen mit Ausnahme der Neutrinos tragen elektrische Ladung und können damit elektromagnetisch wechselwirken. Die elektromagnetische Wechselwirkung hat eine unendliche Reichweite.

2.1.2 Quarks

Die Quarks sind Spin-1/2-Teilchen und zählen damit zu den Fermionen. Es existieren sechs Arten (Flavours) von Quarks; Up-, und Downquark, Charm-, und Strangequark, Top-, und Bottomquark. Up-, Charm-, und Topquark haben eine elektrische Ladung von $\frac{2}{3}e$, Down-, Strange-, und Bottomquark haben

eine elektrische Ladung von $-\frac{1}{3}e$. Quarks können schwach, stark und elektromagnetisch wechselwirken. Aufgrund des Color-Confinements werden Quarks in der Natur nicht isoliert beobachtet. Sie treten einzig gebunden in Hadronen auf. Dies sind unter anderen Mesonen, also Quark-Antiquark-Paare ($q\bar{q}$) oder (Anti-)Baryonen, also Kombinationen aus drei (Anti-)Quarks (qqq oder \overline{qqq}).

2.1.3 Leptonen

Leptonen haben, ebenso wie Quarks, einen Spin von $1/2$ und lassen sich damit ebenso den Fermionen zuordnen. Elektron e , Myon μ und Tauon τ sowie deren entsprechende Neutrinos ν_e , ν_μ und ν_τ gehören zur Klasse der Leptonen. Hierbei tragen Elektron, Myon und Tauon jeweils die elektrische Ladung $-e$, deren entsprechende Antiteilchen tragen die Ladung $+e$. Die Neutrinos sind elektrisch neutral und werden im Rahmen dieser Arbeit als masselos angenommen. Allerdings ist seit der Entdeckung der Neutrinooszillation bekannt, dass die Masse der Neutrinos von Null verschieden sein muss [7]. Leptonen wechselwirken schwach und elektromagnetisch, falls sie eine elektrische Ladung tragen.

Quarks und Leptonen werden in 3 Generationen unterschieden, die jeweils einer Spalte in Abbildung 2.1 entsprechen.

2.1.4 Higgs-Boson

Das einzige, bis jetzt noch nicht vorgestellte Teilchen des Standardmodells ist das Higgs-Boson. Es entspricht einer quantenmechanischen Anregung des Higgs-Feldes. Das Feld ist dadurch charakterisiert, dass alle elementaren Teilchen ihre Masse erst durch Wechselwirkung mit diesem Feld erhalten. Die Kopplung eines Teilchens an das Higgs-Feld ist proportional zu seiner Masse. Es soll hier nicht weiter auf das Higgs-Feld eingegangen werden, für weiterführenden Literatur siehe z.B. [4]. Das Higgs-Boson trägt keine elektrische Ladung, hat Spin 0 und eine Masse von $125,10 \text{ GeV}$ [6]. In Abbildung 2.2 sind die Verzweigungsverhältnisse für einige Zerfallskanäle des Higgs-Bosons dargestellt. Der Zerfall des

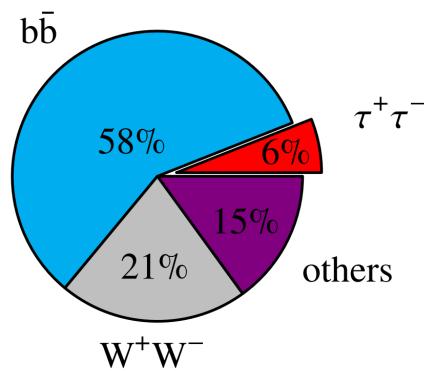


Abbildung 2.2: Verzweigungsverhältnisse für den Zerfall des Higgs-Bosons [8]

Higgs-Bosons in zwei Tau-Leptonen, auf den sich in dieser Arbeit konzentriert werden soll, macht also 6 % der gesamten Zerfälle aus. Das ist ein signifikanter Beitrag.

2.2 Das Tau-Lepton

Das Tau-Lepton (auch τ -Lepton oder Tauon) ist mit einer Masse von $(1\,776,86 \pm 0,12)$ MeV [6] das schwerste der Leptonen. Es zerfällt aufgrund seiner hohen Masse, und der damit einhergehenden kurzen Lebensdauer so schnell, dass nur dessen Zerfallsprodukte in Detektoren nachgewiesen werden können. Der Zerfall erfolgt sowohl hadronisch (Verzweigungsverhältnis: 65 %), als auch leptonisch (Verzweigungsverhältnis: 35 %) [9]. In Tabelle 2.1 sind die verschiedenen Zerfallskanäle des Tau-Leptons mit den entsprechenden Verzweigungsverhältnissen dargestellt, wobei auf die Auflistung der Tau-Neutrinos ν_τ in den Zerfallsprodukten verzichtet wurde. Es wird in 1- und 3-Prong-Zerfälle

Zerfallsmodus	Zerfallsprodukte	Verzweigungsverhältnis / %
hadronisch - 1 Prong		50,1 %
	π^-	10,8 %
	$\pi^-\pi^0$	25,5 %
	$\pi^-2\pi^0$	9,3 %
	$\pi^-3\pi^0$	1,1 %
hadronisch - 3 Prong		15,2 %
	$\pi^-\pi^-\pi^+$	9,3 %
	$\pi^-\pi^-\pi^+\pi^0$	4,6 %
	Andere	2,4 %
leptonisch		
	$e^-\bar{\nu}_e$	17,83 %
	$\mu^-\bar{\nu}_\mu$	17,41 %

Tabelle 2.1: Zerfälle des τ -Leptons, Daten aus [9]

unterschieden, die angeben, ob ein oder drei geladene Teilchen unter den Zerfallsprodukten sind. In jeden möglichen Zerfall des Tau sind aufgrund der Leptonenfamilienzahlerhaltung (siehe Abschnitt 2.3.2) Neutrinos involviert.

2.3 Zerfallskinematik

2.3.1 Helizität und Polarisationsasymmetrie

Die Helizität h ist definiert als die Projektion des Spins auf den Impuls eines Teilchens:

$$h = \frac{\vec{s} \cdot \vec{p}}{|\vec{s}| \cdot |\vec{p}|}$$

mit dem Spinvektor \vec{s} und dem Impulsvektor \vec{p} des Teilchens. Die Helizität ist ein Pseudoskalar; bei einer Raumspiegelung ändert sich das Vorzeichen. Die schwache Wechselwirkung ist paritätsverletzend, sie koppelt nur an linkshändige Fermionen und rechtshändige Antifermionen. Die Händigkeit eines Teilchens ist nicht unmittelbar gleichbedeutend mit der Helizität. Ist der Helizitätszustand eines Teilchens positiv, kann es dennoch einen geringen linkshändigen Beitrag besitzen. Die schwache Wechselwirkung kann demnach (wenn auch mit einer sehr kleinen Wahrscheinlichkeit) an das Teilchen koppeln. Für

ein Teilchen, dessen Masse gegenüber seiner Energie vernachlässigbar ist, beschreiben Helizität und Händigkeit die selbe Eigenschaft. Da sich diese Arbeit ausschließlich mit solchen Teilchen befasst, ist es im weiteren Verlauf ausreichend, Händigkeit und Helizität als gleichbedeutend zu verwenden.

Die Polarisationsasymmetrie ist allgemein definiert durch [10]:

$$A_{\text{Pol}} = \frac{\sigma_L - \sigma_R}{\sigma_L + \sigma_R}$$

wobei $\sigma_{R/L}$ der Wirkungsquerschnitt für die Entstehung rechts- oder linkshändiger Tau-Leptonen beim Zerfall ist. Für das Higgs-Boson ist im Zerfall keine Polarisationsasymmetrie zu erwarten, da es gleich stark an links- und rechtshändige Teilchen koppelt. Das Z-Boson als ein Austauschteilchen der schwachen Wechselwirkung koppelt dagegen unterschiedlich stark an links- und rechtshändige Fermionen. Später (Abschnitt 5) wird außerdem der Zerfall eines virtuellen Photons in zwei Tau-Leptonen betrachtet. Bei diesem Zerfall ist ebenfalls keine Polarisationsasymmetrie zu erwarten, da das Photon gleichstark an links- und rechtshändige Fermionen koppelt.

2.3.2 Rekonstruktion unbekannter Größen

Bei der Analyse von Teilchenzerfällen ist die Kenntnis physikalischer Erhaltungssätze essenziell. Im Folgenden sind einige Erhaltungsgrößen angeführt:

1. Energie
2. Spin
3. Ladung
4. Baryonenzahl
5. Leptonenzahl- und Leptonenfamilienzahlerhaltung

Um aus der invarianten Masse des $\tau^+ \tau^-$ -Systems auf die Masse des Mutterteilchens, also Higgs-Boson oder Z-Boson zu schließen, betrachtet man die Vier-Impuls-Erhaltung. Für den Zerfall des Higgs-, oder Z-Bosons sieht diese wie folgt aus:

$$m_{H/Z} = \sqrt{(P_{\tau^+} + P_{\tau^-})^2} \quad (2.1)$$

Mit den 4-Impulsen des Antitäus P_{τ^+} und des Taus P_{τ^-} . Das Tau-Lepton selbst zerfällt weiter, daher muss der Vier-Vektor der Tau-Leptonen ebenfalls rekonstruiert werden. In jeden Zerfall des Tau-Leptons sind Neutrinos involviert, die aufgrund ihrer geringen Wechselwirkungswahrscheinlichkeit nicht vom Detektor registriert werden können. Für die Vier-Impulserhaltung beim Zerfall des Tau-Leptons mit den Vier-Impulsen des Neutrinos P_ν und der registrierbaren Zerfallsprodukte P_{vis} gilt folglich folgender Zusammenhang:

$$m_\tau = \sqrt{(P_\nu + P_{\text{vis}})^2}$$

Daneben kann auch die Spinerhaltung für die Massenrekonstruktion hilfreich sein. Da in der Natur nur linkshändige Neutrinos, und rechtshändige Antineutrinos vorkommen, kann die Flugrichtung des Neutrinos in diesen Fällen geschlussfolgert werden (da Neutrinos im Standardmodell als masselos

angesehen werden).

Die Leptonen(familien)zahlerhaltung spielt bei der Eruierung erlaubter Zerfallsprozesse des Tau-leptons eine Rolle. Die Leptonenzahl L ist definiert als:

$$L = n_l - n_{\bar{l}}$$

mit der Anzahl der in den Zerfall involvierten Leptonen n_l und Antileptonen $n_{\bar{l}}$. Die Leptonenzahl wird weiter aufgegliedert in Leptonenfamilienzahlen für Elektronen, Myonen und Tauonen:

$$L_e = n_e - n_{\bar{e}} \quad L_\mu = n_\mu - n_{\bar{\mu}} \quad L_\tau = n_\tau - n_{\bar{\tau}}$$

KAPITEL 3

LHC und ATLAS

Die im Rahmen dieser Arbeit untersuchte Methode zur Rekonstruktion der Di-Tau Masse basiert auf Monte-Carlo-Simulationen von Proton-Proton-Kollisionen. Diese Proton-Proton-Kollisionen werden im ATLAS (**A** Toridial LHC ApparatuS)-Experiment am LHC (**L**arge **H**adron **C**ollider) aufgenommen. Es soll hier kurz auf den LHC und den ATLAS-Detektor eingegangen werden. Der Aufbau des ATLAS-Detektors wird beschrieben, die Datenaufnahme und -verarbeitung erklärt und es wird auf die Rekonstruktion der physikalischen Prozesse und Größen hinter den gemessenen Signalen eingegangen.

3.1 Der LHC

Der LHC ist mit einem Umfang von ca. 27 km [11] der größte und leistungsfähigste Teilchenbeschleuniger der Welt. Er verläuft unterirdisch in der französischsprachigen Schweiz und in Frankreich. Er ist in Abbildung 3.1 dargestellt.

Es werden die bei dieser Kollision entstehenden Teilchen einerseits hinsichtlich neuer Physik jenseits des Standardmodells, andererseits werden die elementaren Teilchen des Standardmodells untersucht. Der LHC ist ein Teil des CERN (**C**onseil **E**uropéen pour la **R**echerche **N**ucléaire)-Beschleunigerkomplexes, der aus insgesamt neun Teilchenbeschleunigern besteht. Dabei ist er das letzte Element in einer Kette von Beschleunigungen. Ein Teilchenstrahl, im für diese Arbeit relevanten Fall bestehend aus Protonen, wird in unterschiedlichen Beschleunigern auf immer höhere Geschwindigkeiten gebracht, bis er im LHC seine Maximalgeschwindigkeit erreicht. In einem der vier großen Teilchendetektoren ATLAS, ALICE (**A** Large **I**on **C**ollider **E**xperiment), CMS (**C**ompact **M**uon **S**olenoid) und LHCb (**L**arge **H**adron **C**ollider **b**eauty), deren Design auf die Untersuchung spezieller physikalischer Fragestellung ausgelegt sind, werden zwei gegenläufige Teilchenstrahlen mit einer maximalen Energie von jeweils 7 TeV zur Kollision gebracht. Die gesamte Kollisionsenergie beträgt dementsprechend 14 TeV. Der Protonenstrahl ist aufgeteilt in 2808 Pakete (engl. bunch), die jeweils etwa 10^{11} Protonen enthalten. Der Zeitabstand zwischen den Kollisionen der einzelnen Bunches liegt bei 25 ns. Die Strahlen zirkulieren mehrere Stunden im LHC, wobei die Protonen den LHC 11 245 mal in der Sekunde umrunden [11].

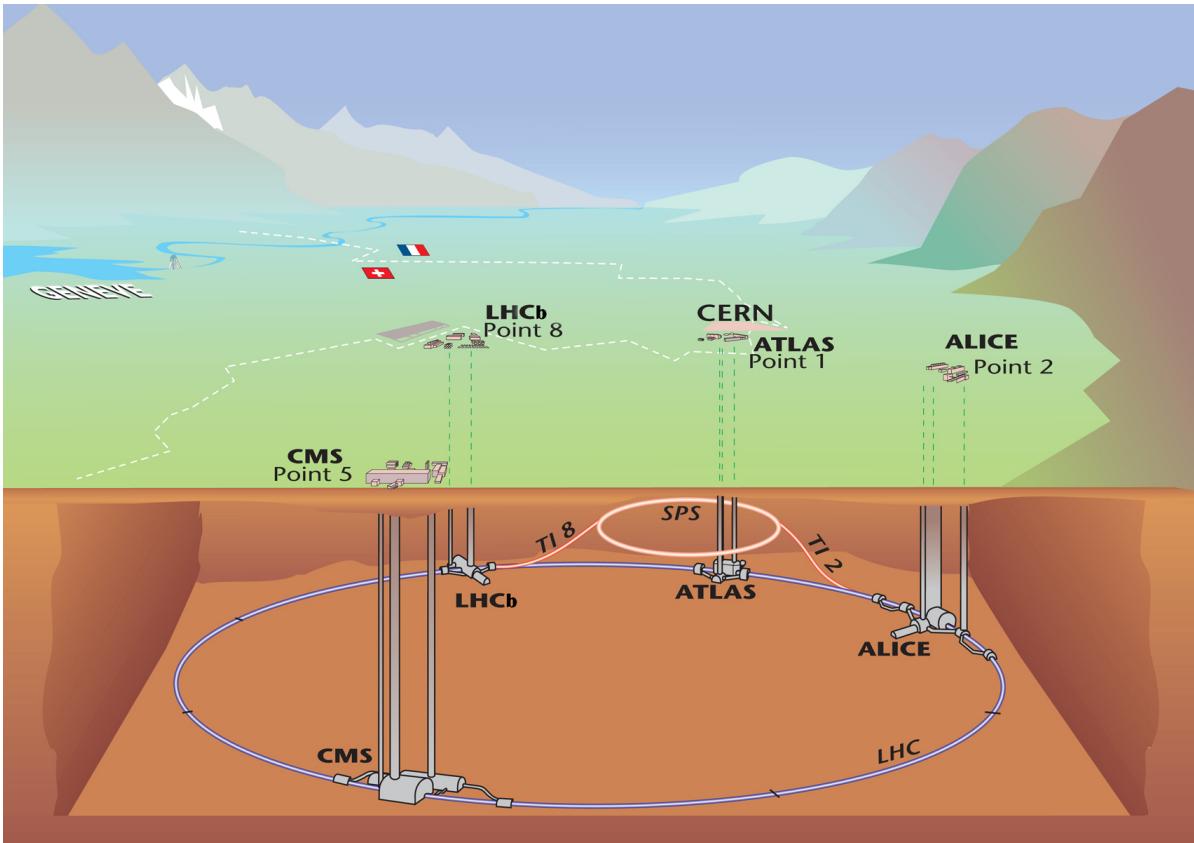


Abbildung 3.1: Der LHC [12]

Ein ausschlaggebender Faktor für die Bewertung der Qualität eines Teilchenbeschleunigers ist die Luminosität \mathcal{L} . Sie ist für ein Experiment zweier kollidierender Strahlen definiert als [4]:

$$\mathcal{L} = \frac{N_1 \cdot N_2 \cdot j \cdot f}{A}$$

mit N_1 der Anzahl an Protonen im ersten Strahl, N_2 der Anzahl an Protonen im zweiten Strahl, j der Anzahl an Bunches, f der Frequenz, mit der die Strahlen aufeinander treffen und A dem Strahlquerschnitt im Kollisionspunkt. Im LHC wird eine Luminosität von maximal $1,2 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ erreicht [13].

3.2 ATLAS

Der ATLAS-Detektor ist mit 46 m Länge und 25 m Durchmesser der größte Detektor am CERN [13]. Er wiegt etwa 7000 t und ist in Abbildung 3.2 illustriert. Mit dem ATLAS-Detektor wird physikalischen Fragestellungen von der Untersuchung des Higgs-Bosons bis hin zur Suche nach neuer Physik nachgegangen [11]. Der ATLAS-Detektor ist in Subsysteme unterteilt, die in einer Zwiebelschalenstruktur angeordnet sind. Diese werden von innen nach außen kurz vorgestellt.

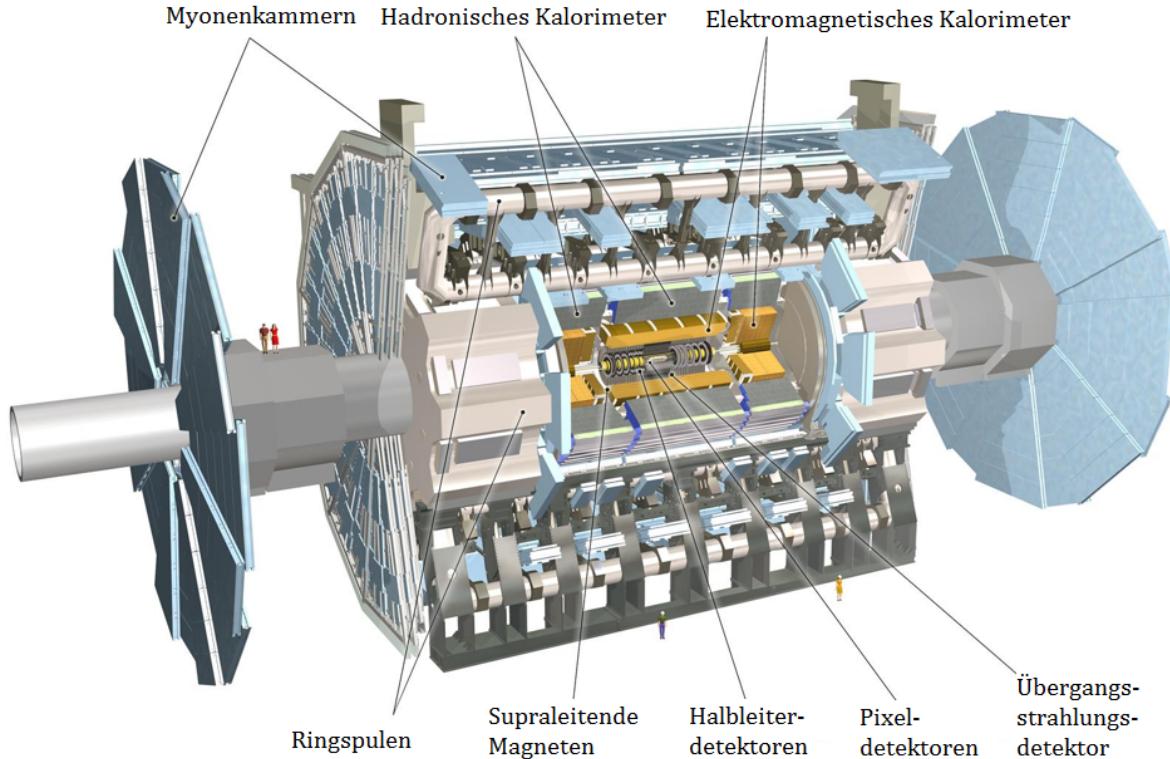


Abbildung 3.2: Aufbau des ATLAS-Detektors [14]

3.2.1 Innerer Detektor

Die Aufgabe des inneren Detektors [15], der so nah wie möglich am Wechselwirkungspunkt liegt, ist die genaue Orts- und Impulsmessung geladener Teilchen. Im gesamten inneren Detektor ist ein starkes Magnetfeld von etwa 2 T angelegt, welches die Bahnen der geladenen Teilchen krümmt. Bei bekanntem Magnetfeld kann aus der Stärke der Krümmung der Teilchenbahnen auf den Impuls der Teilchen geschlossen werden. Der innere Detektor besteht aus drei Komponenten; dem Pixeldetektor, dem Siziliumstreifen-Detektor und dem Übergangsstrahlendetektor (Transition-Radiation-Tracker). Der Pixeldetektor und der Siziliumstreifen-Detektor werden für verschiedene Aspekte der Ortsrekonstruktion der Teilchen genutzt. Der Übergangsstrahlendetektor dient der Teilchenidentifikation.

3.2.2 Kalorimeter

Die Aufgabe der Kalorimeter [16] ist es, die Energie der geladenen Teilchen zu messen. Das passiert durch Wechselwirkung der Teilchen mit dem Detektormaterial. Im besten Fall verliert das Teilchen seine gesamte Energie im Detektormaterial. Das aus einem elektromagnetischen und einem hadronischen Kalorimeter bestehende System liegt außerhalb des 2 T starken Magnetfeldes. Das elektromagnetische Kalorimeter bestimmt die Energie von Elektronen und Photonen, die beim Einfall in das Material elektromagnetische Schauer bilden. Myonen können hier nicht nachgewiesen werden, da sie zu schwer sind. Tauonen zerfallen, bevor sie in das Detektormaterial eindringen können. Es werden also nur deren Zerfallsprodukte (teilweise) registriert. Das hadronische Kalorimeter dient dem Nachweis stark

wechselwirkender Teilchen wie Protonen, Pionen etc.

3.2.3 Myon-Detektor

Da Myonen nicht stark wechselwirken und auch im elektromagnetischen Kalorimeter nicht registrierbar sind, werden sie im Myon-Detektor [17] weiter untersucht. Der Myon-Detektor befindet sich in einem Magnetfeld, das der Spurverlauf- und Impulsmessung dient. Es gibt ein Triggersystem, das physikalisch relevante Ereignisse, in denen Myonen auftreten, markiert.

3.3 Das ATLAS-Koordinatensystem

Die Ereignisse im ATLAS-Detektor werden mit Hilfe von Zylinderkoordinaten mit einer Erweiterung um den Winkel θ beschrieben. Die z -Achse zeigt in Strahlrichtung und $r = \sqrt{x^2 + y^2}$ gibt den Abstand zur Strahlachse an. Die x -Achse zeigt in Richtung des Zentrums des LHC, und die y -Achse zeigt nach oben. Der Azimutwinkel ϕ beschreibt die Ebene senkrecht zur Strahlrichtung und θ beschreibt die von z - und y -Achse aufgespannte Ebene. Das Koordinatensystem ist in Abbildung 3.3 dargestellt. Es wird

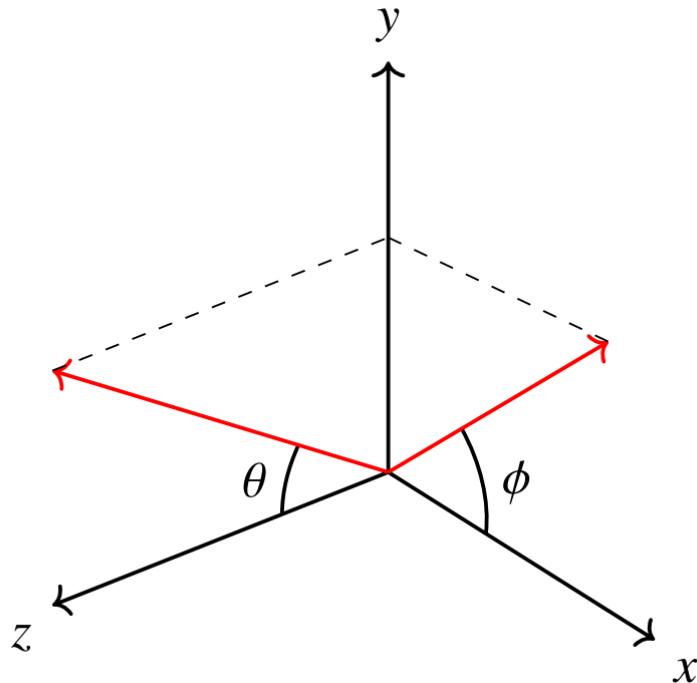


Abbildung 3.3: Das im ATLAS-Detektor verwendete Koordinatensystem [8]

ferner θ durch die Pseudorapidität η ersetzt. Diese Größen verbindet die folgende Relation [18]:

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

Die Pseudorapidität wird genutzt, da Differenzen lorentz-invariant, also unabhängig von der Wahl des Bezugssystems sind. Sie wird klein für Teilchen, die sich nach oben bewegen, und divergiert für Teilchen, die sich entlang des Strahls bewegen. Der Abstand ΔR zweier Vektoren im Raum ist dann gegeben durch [18]:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$$

Eine weitere wichtige Größe ist der Transversalimpuls p_T , sprich der Impuls eines Teilchens in der Ebene transversal zu seiner Bewegung. Er ist über folgenden Zusammenhang definiert [19]:

$$p_T = \sqrt{p_x^2 + p_y^2}$$

Der Longitudinalimpuls eines Teilchens zeigt in Strahlrichtung, entspricht also p_z . Es kann jetzt der Impuls eines Teilchens allgemein ausgedrückt werden als [8]:

$$\vec{p} = \begin{pmatrix} p_T \cos \phi \\ p_T \sin \phi \\ p_T \sinh \eta \end{pmatrix}$$

Im Gegensatz zu der kartesischen Darstellung:

$$\vec{p} = \begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix}$$

3.4 Der Missing-Mass-Calculator

Das grundlegende physikalische Problem im Zentrum dieser Arbeit ist die Rekonstruktion der Di-Tau Masse. Sie ist, wie in Abschnitt 2.2 erklärt, lediglich aus den Zerfallsprodukten der Tau-Leptonen bestimmbar. Für die Di-Tau Masse gilt:

$$m_{\tau\tau} = \sqrt{(P_{\text{vis}} + P_\nu)_{\tau\tau}^2}$$

Über die Berechnung der fehlenden Transversalenergie kann die Summe der (nicht registrierten) Neutrinoimpulse ermittelt werden. Allerdings kann keine Aussage über die individuellen Neutrinoimpulse getroffen werden. Für die Berechnung der fehlenden Transversalenergie wird ausgenutzt, dass der Transversalimpuls vor der Kollision zweier Protonen null beträgt. Die entsprechenden Transversalimpulse der Zerfallsprodukte nach der Kollision müssen sich dann zu null addieren. Falls das nicht der Fall ist, so sind Neutrinos für diese fehlende Energie in der Transversalebene verantwortlich. Der Vektor der fehlenden Transversalenergie kann allgemein beschrieben werden durch seinen Betrag $|\vec{E}_{T, \text{miss}}|$ und seinen Azimutwinkel Φ_{miss} [20]:

$$|\vec{E}_{T, \text{miss}}| = \sqrt{E_{x, \text{miss}}^2 + E_{y, \text{miss}}^2} \quad \Phi_{\text{miss}} = \arctan\left(\frac{E_{y, \text{miss}}}{E_{x, \text{miss}}}\right)$$

Die aktuell verwendete Methode zur Di-Tau Massenrekonstruktion ist der Missing-Mass-Calculator (MMC) [3]. Es wird für die Rekonstruktion der Masse ein perfekter Detektor angenommen, und dass außer den Neutrinos der Tau-Lepton Zerfälle keine weiteren auftreten. Kurz: die gesamte fehlende Transversalenergie wird ausschließlich durch die Neutrinos aus den Tau-Lepton Zerfällen verursacht. Es sind also 6 bis 8 Unbekannte Größen involviert, je nachdem, ob die Tau-Leptonen hadronisch oder leptonisch zerfallen (vgl. Tabelle 2.1). Diese Unbekannten Größen sind einerseits die x -, y - und z -Komponenten der Neutrinoimpulse, andererseits, und das nur im Fall eines leptonischen Zerfalls, die invariante Masse der beiden beim leptonischen Zerfall entstehenden Neutrinos. Insgesamt sind es dementsprechend mindestens sechs (kein Tau-Lepton zerfällt hadronisch) und höchstens acht (beide Tau-Leptonen zerfallen hadronisch) Unbekannte. Allerdings sind nur vier Gleichungen vorhanden, die diese Unbekannten miteinander verbinden. Es sind mehr Unbekannte als Freiheitsgrade vorhanden, daher lassen sich die Gleichungen nicht exakt lösen. Allerdings sind manche Lösungen des Systems wahrscheinlicher als andere. Diese Tatsache wird bei der Rekonstruktion der Di-Tau Masse mit Hilfe des MMCs ausgenutzt. Für weitere Literatur siehe z.B. [3]. Da der MMC viel Rechnerkapazität beansprucht, wird nach neuen Verfahren gesucht, um die Di-Tau Masse zu rekonstruieren. Ein Ansatz sind Künstliche Neuronale Netze, die im folgenden Kapitel beschrieben werden.

KAPITEL 4

Künstliche neuronale Netze

4.1 Grundlagen des Maschinellen Lernens

Die künstliche Intelligenz findet im Alltag immer mehr Anwendung. Seien es Suchmaschinen, Gesichts- oder Texterkennung oder der Einsatz in der Medizin, um beispielsweise Hautkrebs zu identifizieren [21]. Im Maschinellen Lernen, einer Teildisziplin der künstlichen Intelligenz, wird der Ansatz verfolgt, einem Computer das Lernen beizubringen. Die Aufgabe des Computers ist es, durch Erfahrungen eine Aufgabe besser zu lösen, wobei dieser nicht explizit programmiert wird [22]. Es soll die Fähigkeit des Computers ausgenutzt werden, Muster und Zusammenhänge in Daten zu erkennen, und diese in einer gewissen Form zu speichern, sodass Vorhersagen auf Basis von Eingaben formuliert werden können. Das Ziel ist es, dass der Computer diese Zusammenhänge automatisch und ohne die Intervention eines Menschen lernt. Der Begriff des Maschinellen Lernens wird in drei Kategorien unterschieden [22]:

1. **Überwachtes Lernen:** Es wird dem Computer ein Datensatz aus Eingabe- und entsprechenden Ausgabevariablen gegeben. Die Aufgabe des Computers ist es, einen Zusammenhang zwischen Ein- und Ausgabegrößen zu formulieren. Die Anwendung beläuft sich auf das Eingeben eines solchen Eingabedatensatzes, für den der entsprechende Ausgabedatensatz vom Computer vorhergesagt werden soll.
2. **Unüberwachtes Lernen:** Für diese Art des Lernens ist kein Ausgabedatensatz vorhanden. Der Computer verarbeitet die Eingabedaten, und verfolgt das Ziel, Muster und Zusammenhänge in den Daten zu erkennen. Die Ausgabe des Computers ist demnach im Voraus unbekannt.
3. **Bestärkendes Lernen:** Diese Art des Lernens beruht auf dem Konzept der Bewertung der Leistung des Netzes in Form von Belohnung oder Bestrafung. Es ist auch hier kein Ausgangsdatensatz vorhanden. Der Computer (, der in diesem Zusammenhang auch als Agent bezeichnet wird) interagiert mit seiner Umgebung, und lernt aus diesen Interaktionen.

Überwachtes Lernen kann weiter spezifiziert werden in Regressionsprobleme und Klassifikationsprobleme [22]. Bei der Regression ist die Ausgangsvariable kontinuierlich und reell. Bei der Klassifikation ist die Ausgangsvariable eine Kategorie (Klasse). In dieser Arbeit wird ein Regressionsproblem behandelt; es wird die Masse des Di-Tau Systems vorhergesagt, die eine kontinuierliche Größe ist. Die Klassifikation ist für das in Abschnitt 7.4.1 beschriebene adversarial neural network relevant.

4.2 Funktionsweise

Künstliche Neuronale Netze (KNNs) und das mit ihnen verbundene Deep Learning sind Teilgebiete des maschinellen Lernens. Sie sind dem menschlichen Gehirn nachempfunden und bestehen aus miteinander vernetzten künstlichen Neuronen. Ein Feed-Forward Neuronales Netz (siehe Abbildung 4.1) ist dadurch charakterisiert, dass die Information lediglich in eine Richtung (Information verläuft von Eingabe- zur Ausgabeschicht, wie in Abbildung 4.1 dargestellt) durch das Netz gegeben wird. Es wird zu keinem Zeitpunkt Information aus dem Netz in das Netz zurückgeschleust [23]. Im Rahmen dieser Arbeit wird ausschließlich mit Feed-Forward-KNNs gearbeitet.

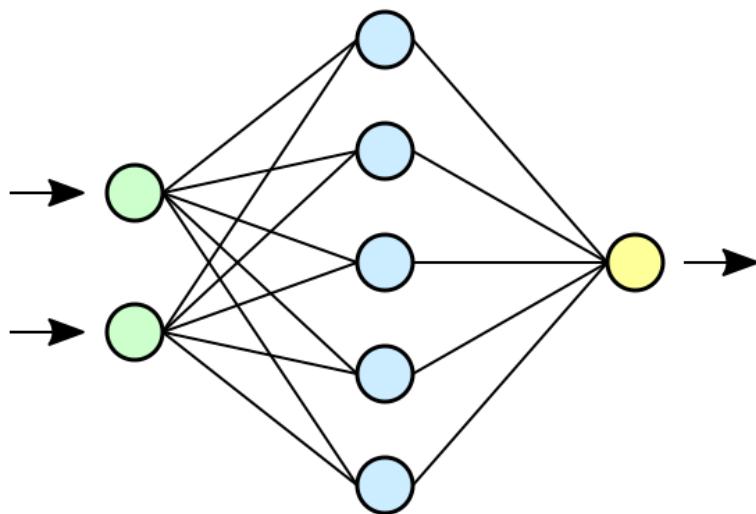


Abbildung 4.1: Feed-Forward Neuronales Netz, die Information wird in die Eingabeschicht (grün) gegeben, wird in der verdeckten Schicht (blau) weitergeleitet und die entsprechende Vorhersage wird in der Ausgabeschicht (gelb) ausgegeben. [24]

Im Wesentlichen besteht ein KNN aus mehreren Schichten; der Eingabeschicht (in Abb. 4.1 grün), den verdeckten Schichten (in Abb. 4.1 blau) und der Ausgabeschicht (in Abb. 4.1 gelb). Der Eingabedatensatz wird in die Eingabeschicht gegeben. Die Daten werden in den verdeckten Schichten weiter verarbeitet und in der Ausgabeschicht entsteht die Vorhersage. Die Anzahl der Neuronen in Ein- und Ausgabeschicht sind vom physikalischen Problem abhängig, wohingegen die Anzahl der verdeckten Schichten und insbesondere die Anzahl der Neuronen in diesen nicht im einfachen Zusammenhang mit dem physikalischen Problem steht. Es wird von Deep Learning gesprochen, sobald verdeckte Schichten vorliegen. Das ist bei allen Netzen der Fall, mit denen sich in dieser Arbeit auseinandergesetzt wird.

Jedes der künstlichen Neuronen bekommt Eingabegrößen der vorangehenden, mit ihm verbundenen Neuronen $\vec{x} = \sum_i x_i$. Jede der EingabevARIABLEN x_i wird mit einem individuellen Gewicht w_i multipliziert. Außerdem werden die Eingaben jedes weiteren Neurons ebenfalls mit individuellen Gewichten multipliziert. Jede der in Abbildung 4.1 dargestellten Verbindungen der Neuronen ist also individuell gewichtet. Ferner hat ein künstliches Neuron einen individuellen, konstanten Wert (engl.: Bias) b . Somit

gilt für z , den Wert, den das Neuron als Eingabe bekommt [25]:

$$z = \sum_i x_i \cdot w_i + b$$

Die Aktivierung eines Neurons wird bestimmt durch die verwendete Aktivierungsfunktion f_{akt} . Diese bestimmt, unter welchen Bedingungen Informationen vom entsprechenden Neuron weitergegeben werden. Die Ausgabe a eines Neurons sieht wie folgt aus [25]:

$$a = f_{\text{akt}}(z)$$

Das Ziel eines neuronalen Netzes, das aus einer großen Zahl solcher Neuronen besteht, ist es, eine allgemeine Funktion zu finden, die die Eingabeveriablen \vec{x} so genau wie möglich auf die Ausgabeveriablen \vec{y} abbildet. Im Rahmen dieser Arbeit wird ausschließlich das Überwachte Lernen angewandt; dem Netz werden Eingabedatensätze mit entsprechenden Ausgabedatensätzen geliefert, sodass das Netz eine Funktion (einen Zusammenhang) zwischen den Datensätzen lernen kann.

Neben der grundlegendsten Netzwerkarchitektur, dem Feed-Forward neuronalen Netz, gibt es fortgeschrittenere Strukturen und Konzepte, die auf die Lösung spezifischer Aufgabenstellungen ausgelegt sind. Für diese Arbeit von Relevanz sind das adversarial neural network, sowie die Methode der Distance Corellation, die in Abschnitt 7.4 erläutert werden

4.3 Der Lernvorgang

Der eigentliche Lernvorgang des Netzes entspricht der Anpassung der Gewichtungen und der konstanten Offsets jedes Neurons. Während des Lernvorgangs werden diese Anpassungen vorgenommen, um die Abweichung der Vorhersage des Netzes vom wahren Wert so gering wie möglich zu machen. Damit das Netz weiß, wie gut oder schlecht seine Leistung bezüglich der Vorhersage ist, wird eine Lossfunktion definiert. Diese Lossfunktion fasst die Abweichung aller vorhergesagten Größen $y_{\text{pred}, i}$ von den wahren Größen $y_{\text{true}, i}$ zusammen. Ein Beispiel ist die Mean-Squared-Error Lossfunktion, die in dieser Arbeit Anwendung findet [26]:

$$L(y_{\text{pred}}, y_{\text{true}}) = \frac{1}{m} \sum_i^m (y_{\text{pred}, i} - y_{\text{true}, i})^2 \quad (4.1)$$

wobei m die Anzahl an Ereignissen ist, die verarbeitet werden. Wird die Lossfunktion minimiert, so lernt das Netz, bessere Vorhersagen zu machen. Für die Minimierung der Lossfunktion wird das Konzept der Backpropagation [27] verwendet. Um zu ermitteln, wie stark bestimmte Gewichte und Biases angepasst werden sollen, betrachtet man den Gradienten $\vec{\nabla}$ der Lossfunktion abhängig von diesen Parametern. Allgemein gilt für die Variation der Lossfunktion L mit dem Gewicht w_{jk}^l , das das Signal zwischen j -tem und k -tem Neuron in der l -ten Schicht gewichtet:

$$\frac{\partial L}{\partial w_{jk}^l} = \frac{\partial L}{\partial z_j^l} \frac{\partial z_j^l}{\partial w_{jk}^l}$$

Außerdem ist für diesen Fall:

$$z_j^l = \sum_{k=1}^m w_{jk}^l a_k^{l-1} + b_j^l$$

wobei a_k^{l-1} die Ausgabe des vorherigen Neurons, bedingt durch seine Aktivationsfunktion, ist. Man sieht dann, dass:

$$\frac{\partial z_j^l}{\partial w_{jk}^l} = a_k^{l-1}$$

Also:

$$\frac{\partial L}{\partial w_{jk}^l} = \frac{\partial L}{\partial z_j^l} \cdot a_k^{l-1}$$

Analog kann die Rechnung für den Bias durchgeführt werden. Man erhält:

$$\frac{\partial L}{\partial b_j^l} = \frac{\partial L}{\partial z_j^l}$$

Oft wird weiter folgende Definition verwendet:

$$\delta_j^l = \frac{\partial L}{\partial z_j^l}$$

Die individuellen Weights w und Biases b werden im Verlauf des Trainings wie folgt angepasst:

$$w := w - \alpha \cdot \frac{\partial L}{\partial w} \quad \text{und} \quad b := b - \alpha \cdot \frac{\partial L}{\partial b}$$

bis die Voraussetzungen an die Vorhersage des Netzes erfüllt sind. Hier wurde die Learning-Rate α eingeführt, die bestimmt, wie schnell der Lernvorgang passiert. Dieser Faktor sollte überlegt gewählt werden; eine zu niedrige Learning-Rate hat viele Anpassungen der Parameter zur Folge, eine zu hohe Learning-Rate kann dafür sorgen, dass das Minimum nicht erreicht wird.

4.4 Hyperparameter

Hyperparameter werden vor Trainingsbeginn festgelegt, und steuern den Trainingsablauf. Einige Hyperparameter sind z.B. Learning-Rate, Aktivierungsfunktion und Anzahl verdeckter Schichten. Eine weitere, noch nicht vorgestellte Größe, ist die Epoche. Sie ist definiert als ein Durchlauf durch den gesamten Trainingsdatensatz. Die Anzahl der Epochen ist ebenso ein Hyperparameter. Ferner werden Optimierer genutzt. Optimierer sind Algorithmen, die genutzt werden, um Anpassungen der Gewichte vorzunehmen, um die Lossfunktion zu minimieren [28]. Der in dieser Arbeit genutzte Optimierer ist Adam [29]. Der Optimierer ist ebenfalls ein Hyperparameter.

Falls das neuronale Netz die gewünschte Genauigkeit schon vor Durchlauf aller Epochen erreicht hat, oder sich die Genauigkeit über eine bestimmte Anzahl an Epochen nicht mehr ändert, so gibt es die Möglichkeit, das Training vorzeitig zu beenden. Diese Methode sind die Callbacks [30]. Ein letzter wichtiger Hyperparameter ist die Batchsize. Die Batchsize gibt an, wie viele Ereignisse durch das Neuronale Netz geschickt werden, bis die Parameter des Netzes angepasst werden. Wären beispielsweise 10 000 Ereignisse für das Training vorhanden, und es würde eine Batchsize von 100 gewählt, so würden die Parameter des Netzes innerhalb einer Epoche 100 mal angepasst, und das immer, nachdem 100 Ereignisse durch das Netz geschickt worden sind.

4.5 Software

Die neuronalen Netze in dieser Arbeit werden nicht neu programmiert, sondern es wird auf schon vorhandene Bibliotheken zurückgegriffen. Es wird mit Tensorflow (Version 2.3) [31] über die Open-Source-Deep-Learning-Bibliothek Keras (Version 2.24) [32] gearbeitet.

KAPITEL 5

Daten

In diesem Abschnitt werden die Monte-Carlo Daten, die im Rahmen dieser Arbeit verwendet werden, beschrieben. Es wird der grobe Aufbau des Neuronalen Netzes besprochen, mit dem die Di-Tau Masse vorhergesagt werden soll. Ferner wird der Trainingsdatensatz besprochen und motiviert und die verwendeten Einschränkungen der Daten beschrieben.

5.1 Simulationen

Die in dieser Arbeit verwendeten Daten stammen aus Monte-Carlo-Simulationen. Daten aus Simulationen werden gegenüber tatsächlichen Messdaten für das Training des Netzes bevorzugt, da gemessene Daten durch neue Physik oder defekte Messgeräte beeinflusst werden können.

Für die Modellierung des Zerfallsprozesses werden Sherpa [33], Pythia [34] und Powheg [35] verwendet. Die simulierten Zerfälle müssen außerdem in einem simulierten ATLAS-Detektor registriert werden. Dazu wird Geant4 [36] genutzt, um die Detektorgeometrie des ATLAS-Detektors zu kopieren. Die simulierten Teilchen können so mit dem simulierten Material des Detektors interagieren.

5.2 Allgemeines

Als Ausgabegröße des Netzes wird die Di-Tau Masse genutzt. Als Eingabegrößen werden dem Netz physikalische Größen gegeben, die Informationen über den Zerfall und die Zerfallskinematik tragen. Es handelt sich dementsprechend um ein Regressionsproblem (siehe Abschnitt 4; die Masse des Di-Tau Systems ist kontinuierlich verteilt). Das allgemeine Ziel ist es, dass das Netz Strukturen und Zusammenhänge zwischen den Ein- und Ausgabedaten erkennt, deren Kenntnis es ihm ermöglichen, die Di-Tau Masse möglichst präzise vorherzusagen. Um dieses Ziel zu erreichen, muss geklärt werden, welcher Datensatz für das Training des Netzes verwendet werden soll. Der Trainingsdatensatz sollte allgemeine physikalische Eigenschaften der Higgs- und Z-Zerfälle wiederspiegeln, andererseits aber keine individuellen physikalischen Eigenschaften enthalten, da das Netz andernfalls Charakteristika des Trainingsdatensatzes lernen könnte, was die physikalisch sinnvolle Anwendbarkeit auf Higgs- und Z-Datensatz beeinträchtigt. Der Trainingsdatensatz muss eine breite Di-Tau Massenverteilung mitbringen, die sowohl die Masse des Higgs-Bosons, als auch die Masse des Z-Bosons einschließt. Weiterhin ist es günstig, wenn die entsprechende Di-Tau Massenverteilung flach ist. Ist die Massenverteilung nicht

flach, so ist der Lernvorgang des Netzes verfälscht und es sagt tendenziell stärker auftretende Massen der Di-Tau Verteilung eher voraus. Die Ansprüche werden durch einen Datensatz für den Zerfall eines virtuellen Photons γ^* in zwei Tau-Leptonen erfüllt: $\gamma^* \rightarrow \tau\tau$. Die simulierte Di-Tau Massenverteilung dieses Datensatzes deckt Massen von 60 GeV bis 7 000 GeV ab. Es wird allerdings nur ein Teil für das Training verwendet (siehe Abschnitt 5.3). Der Datensatz ist beschränkt auf den hadronischen Zerfall des γ^* , also:

$$\gamma^* \rightarrow \tau_{\text{had}} \tau_{\text{had}}$$

da leptonische Zerfälle die Anzahl an unbekannten Größen und somit die Komplexität der zu erlenden Zusammenhänge erhöhen. Ferner ist der ausschließliche Zerfall eines γ^* in zwei Tau-Leptonen unphysikalisch, da dieser immer vom Zerfall $Z \rightarrow \tau\tau$ überlagert wird. Diese Überlagerung ist als Drell-Yan Prozess [4] bekannt. Im weiteren Verlauf wird der Datensatz des $\gamma^* \rightarrow \tau\tau$ Zerfalls als Trainingsdatensatz bezeichnet. Die Unphysikalität des Trainingsdatensatzes kann die spätere Anwendung des Netzes auf Daten aus realistischen Zerfällen einschränken. Die Di-Tau Massenverteilung des Trainingsdatensatzes ist in Abbildung 5.1 unten dargestellt. Die angestrebte Flachheit der Massenverteilung wird durch die Anwendung entsprechender Gewichte erreicht. Dazu wird der gesamte Massenbereich auf der x -Achse des Histogrammes in 1 600 Intervalle aufgeteilt und es wird die Anzahl der zugehörigen Ereignisse N_i innerhalb des i -ten Massenintervalls bestimmt. Die entsprechenden Gewichte w_i , deren Anwendung die Flachheit der Massenverteilung gewährleistet, können definiert werden als:

$$w_i = \frac{1}{N_i}$$

wobei die Massenverteilung hiermit auf 1 normiert ist. Das i -te Gewicht wird beim Training des Netzes auf alle Events im i -ten Massenintervall angewandt. Um das Integral über die flach gewichtete Massenverteilung gleich groß zu machen wie das Integral über die ungewichtete Massenverteilung, wird jedes Gewicht durch die Summe aller Gewichte:

$$\sum_{i=0}^N w_i$$

geteilt, und mit der Gesamtanzahl N an Events multipliziert. Das ist vorteilhaft, wenn zum Beispiel die Entwicklungen der Lossfunktionen über die Epochen gewichtet und ungewichtet betrachtet werden; der Betrag des Losswertes hängt von der Größe der Gewichte ab. Da in dieser Arbeit die Abhängigkeit der Massenvorhersage von der Polarisationsasymmetrie untersucht werden soll, wird der vorhandene $\gamma^* \rightarrow \tau\tau$ Datensatz mit dem Programm Tauspinner [37] um eine Polarisationsasymmetrie erweitert. Die Ereignisse im Datensatz können durch weitere Gewichte so verändert werden, als hätten die Tau-Leptonen aus dem Zerfall des virtuellen Photons eine Polarisation, die der Polarisation der Tau-Leptonen aus dem Zerfall des Z gleicht. Physikalisch bedeutet das, dass das virtuelle Photon γ^* aus dem Trainingsdatensatz nicht mehr gleichstark an links- und rechtshändige Fermionen koppelt. Ein Zerfall des virtuellen Photons γ^* in ein linkshändiges Antitau oder ein rechtshändiges Tau wird beispielsweise unwahrscheinlicher, während der Zerfall in ein linkshändiges Tau und ein rechtshändiges Antitau wahrscheinlicher wird. Das ist unphysikalisch und tritt in der Natur so nicht auf, es können allerdings mit dem erzeugten Datensatz physikalische Eigenschaften des Z -Datensatzes imitiert werden. Der erzeugte Datensatz entspricht zunächst der Simulation des $\gamma^* \rightarrow \tau\tau$ Zerfalls mit einem zusätzlichen Satz an Gewichten. Werden die Gewichte auf die mit einem der Tau-Leptonen verknüpften Variablen angewandt, sind die Tau-Leptonen

Kapitel 5 Daten

aus dem Zerfall des virtuellen Photons mit einer Polarisationsasymmetrie behaftet. Durch Nicht-Anwendung der Gewichte steht ein Datensatz zur Verfügung, bei dem es keine Polarisationsasymmetrie der Tau-Leptonen gibt. Die Massenverteilungen mit Polarisationsasymmetrie ist in Abbildung 5.1 oben zu sehen.

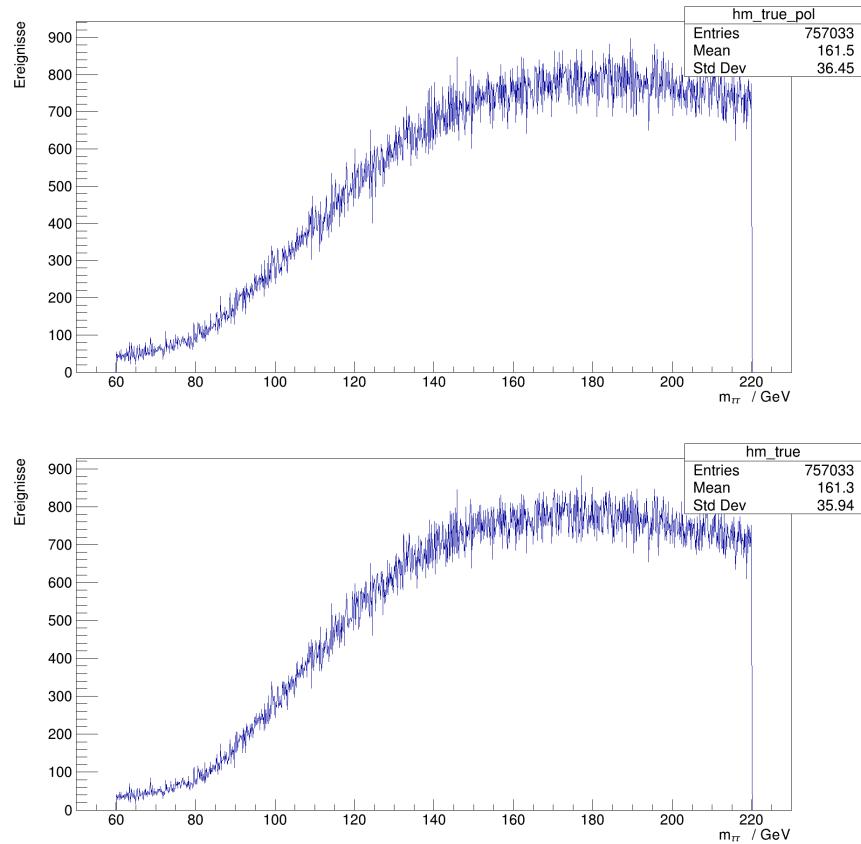


Abbildung 5.1: Massenverteilung des Trainingsdatensatzes mit (oben) und ohne Polarisationsasymmetrie (unten)

5.3 Vorselektion

5.3.1 Z-Datensatz

Die Vorselektion für den Z-Datensatz wird durch folgende Einschränkungen beschrieben und wurde aus [8] übernommen:

1. Ein oder drei geladene Tracks
2. Tau Identifikation: medium [38]
3. $0 \leq |\eta| \leq 1,37$ oder $1,52 \leq |\eta| \leq 2,5$
4. Unterschiedliche Ladung der zwei Tau-Leptonen
5. $E_{T, \text{miss}} \geq 20 \text{ GeV}$
6. $p_T^{\tau_0} \geq 40 \text{ GeV}$ und $p_T^{\tau_1} \geq 30 \text{ GeV}$
7. $0,8 < \Delta R(\tau_0, \tau_1) < 2,5$
8. $|\Delta\eta(\tau_0, \tau_1)| < 1,5$
9. Gültiger Zerfallsmodus

5.3.2 Trainingsdatensatz

Nicht alle Daten im Datensatz sind für das Training relevant. Die Massenverteilung muss ausreichend breit sein, um sowohl Higgs- als auch Z-Masse abzudecken. Allerdings ist die Breite des $\gamma^* \rightarrow \tau\tau$ Prozesses von 60 GeV bis 7 000 GeV dafür zu breit. Nur solche Events, deren Di-Tau Masse im Bereich von 60 GeV bis 220 GeV liegt, werden für das Training verwendet. Mit diesem Bereich sind Higgs- und Z-Masse ausreichend eingeschlossen. Die Vorselektion, die auf den Z-Datensatz angewandt wird, wird so auch auf den Trainingsdatensatz angewandt. Es werden außerdem nur solche Ereignisse gewählt, in denen zwei Tau-Leptonen auftreten, die entsprechenden wahren pdgIDs [39] sind also 15 oder -15.

KAPITEL 6

Ergebnisse vorheriger Arbeiten

Diese Bachelorarbeit baut auf den Ergebnissen, die in den Masterarbeiten „Optimization of a Regression-Based Deep Neural Network for Di-Tau Mass Reconstruction in ATLAS“ [8] von Lena Herrmann und „Estimating the Mass of Di-Tau Systems in the ATLAS Experiment Using Neural Network Regression“ [25] von Martin Werres herausgefunden wurden, auf. Die Ergebnisse sollen hier kurz besprochen, und insbesondere soll in dem Zusammenhang das zentrale Thema dieser Arbeit motiviert werden.

6.1 Erklärung

In [25] wurde untersucht, wie neuronale Netze auf das Problem der Rekonstruktion der Di-Tau Masse angewandt werden können und die Vorhersagen mit den Rekonstruktionen des MMC verglichen. Es wurde die Di-Tau Masse direkt als Ausgabe des Neuronalen Netzes genutzt. In [8] wurden, aufbauend auf [25], die Neutrinoimpulse als Ausgabe des Neuronalen Netzes gewählt und verschiedene Einschränkungen beim Training angewandt. In dem für diese Arbeit relevanten Aspekt, der Korrelation zwischen Massenvorhersage des Neuronalen Netzes und der Polarisationsasymmetrie der Tau-Leptonen, ließen sich aus [25] und [8] die selben Hypothesen formulieren. Es werden hier die Ergebnisse aus [8] vorgestellt, da die für diese Arbeit relevanten Aspekte gründlicher dokumentiert wurden. Aus dem vorhergesagten Vierimpuls der Neutrinos kann, wie in Abschnitt 2 beschrieben, über die 4-Impulserhaltung auf die zu Grunde liegende Di-Tau Masse geschlossen werden [8]:

$$m_{\tau\tau} = \sqrt{(P_{\text{vis}_0} + P_{\nu_0} + P_{\text{vis}_1} + P_{\nu_1})^2}$$

wobei die Indizes 0 und 1 jeweils für das höher- und niederenergetische Tau stehen. Das Netz wurde mit dem $\gamma^* \rightarrow \tau\tau$ Datensatz trainiert. Es wurde sowohl mit einem Datensatz des simulierten Higgs- als auch des simulierten Z-Zerfalls eine Vorhersage für die Masse des Mutterteilchens gemacht. Die resultierende Vorhersage ist in Abbildung 6.1 zu sehen. Die vorhergesagten Massen weichen mit 102,9 GeV für die Higgs Masse und 80,9 GeV für die Z-Masse stark von den wahren Werten ab. Als allerdings Events des Trainingsdatensatzes für die Vorhersage verwendet wurden, deren wahre Masse im Bereich der Higgs- und Z-Masse liegen, kam es zu den Vorhersagen in Abbildung 6.2. Die vorhergesagten Massen liegen hier für Higgs mit 126,02 GeV und Z mit 95,87 GeV deutlich näher an den wahren Massen. Daraus kann geschlussfolgert werden, dass die falschen Vorhersagen bei Anwendung der tatsächlichen Higgs- und Z- Datensätze keineswegs einem fehlerhaften Trainingsprozess, sondern physikalischen Unterschieden

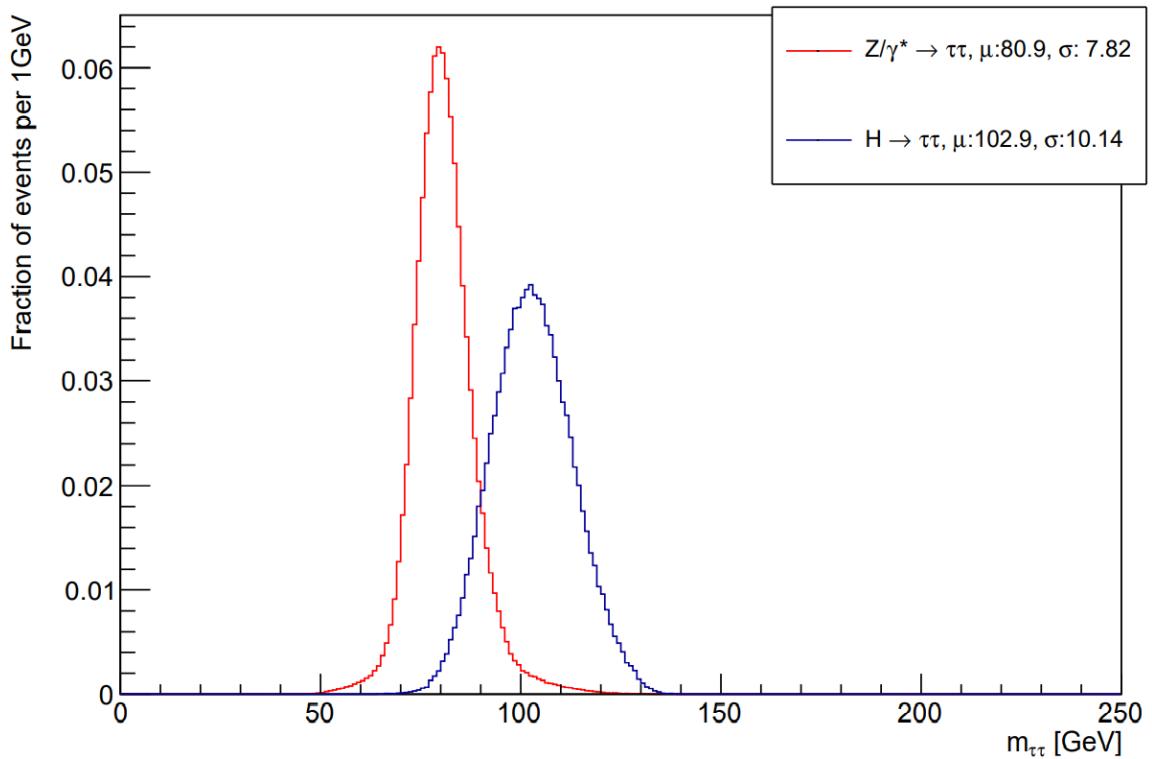


Abbildung 6.1: Vorhergesagte Massenverteilung für Higgs- und Z-Masse, die vorhergesagten Massen weichen stark von den wahren Massen ab [8]

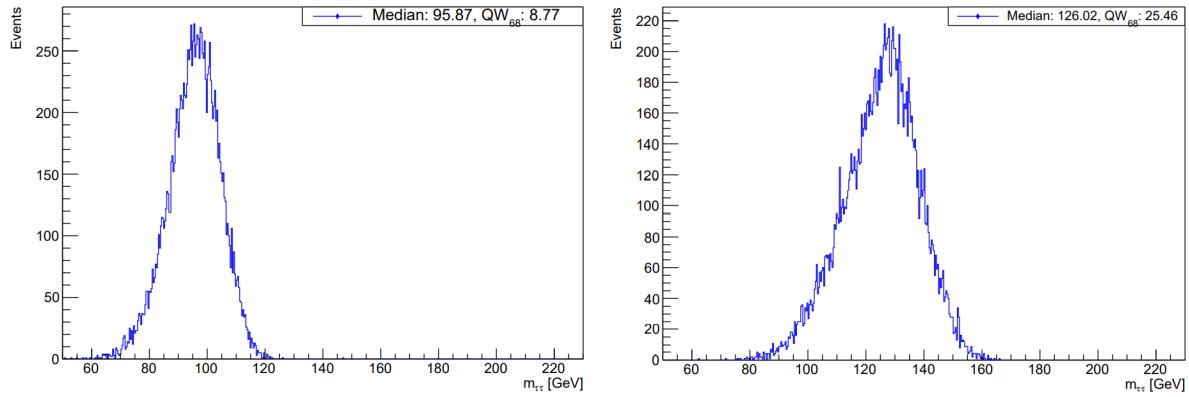


Abbildung 6.2: Vorhergesagte Massenverteilung für Ereignisse aus dem Trainingsdatensatz, deren wahre Masse im Bereich der Z- (links) und Higgs-Masse (rechts) liegt. Die Vorhersage ist deutlich besser, als mit den Daten aus Z- und Higgs-Datensatz. [8]

zwischen dem Trainingsdatensatz und dem Higgs-, bzw. Z-Datensatz geschuldet sind. Das kann einerseits der Spin sein; sowohl das Z-Boson als auch das virtuelle Photon haben einen Spin von eins, anders als das Higgs-Boson mit einem Spin von null. Andererseits sind sowohl die Tau-Leptonen aus dem Higgs-Zerfall, als auch die Tau-Leptonen aus dem Zerfall des virtuellen Photons unpolarisiert, da das Higgs-Boson und das Photon gleichstark an links- und rechtshändige Fermionen koppeln. Das Z-Boson als Austauschteilchen der schwachen Wechselwirkung koppelt unterschiedlich stark an links- und rechtshändige Fermionen. Die Hypothese, dass die Massenvorhersage des neuronalen Netzes abhängig von verschiedenen physikalischen Eigenschaften der Datensätze, also der Polarisationsasymmetrie der Tau-Leptonen und dem Spin des Mutterteilchens ist, ist der physikalische Ausgangspunkt dieser Bachelorarbeit. In [8] wurden die Neutrino Vier-Impulse vorhergesagt. Das kann einen Einfluss auf die Ausprägung des beschriebenen Effektes haben; da die Zerfallskinematik der Tau-Leptonen durch die Polarisationsasymmetrie beeinflusst wird, sind die Neutrino Vier-Impulse potentiell sensitiver auf die Polarisationsasymmetrie der Tau-Leptonen, als die Masse des Di-Tau Systems, die in dieser Arbeit als Ausgabe genutzt wird.

Die Vorhersage der Higgs-Masse mit dem Higgs-Datensatz zeigt eine größere Abweichung des Mittelwertes vom wahren Wert und, wie in Abbildung 6.1 zu sehen, eine größere Standardabweichung als die Vorhersage mit dem Z-Datensatz. Das legt nahe, dass der Trainingsdatensatz physikalisch dem Z-Datensatz ähnlicher als dem Higgs-Datensatz ist. An dieser Stelle ist die grundlegende Aufgabe, die Vorhersage der Higgs-Masse zu verbessern. Allerdings ist der physikalische Aspekt, der Higgs- und Trainingsdatensatz voneinander unterscheidet, der Spin. Mit einem Trainingsdatensatz, dessen Zerfall ein Mutterteilchen mit Spin 0 zu Grunde liegt, kann die Dekorrelation von Spin und Masse untersucht werden, um die Vorhersage mit dem Higgs-Datensatz zu verbessern. Der Trainingsdatensatz wäre in diesem Fall physikalisch näher am Higgs, da Spin und Polarisationsasymmetrie gleich sind. Ein solcher fehlt allerdings. Die Korrelation zwischen Polarisationsasymmetrie und Massenvorhersage kann allerdings untersucht werden, da, wie in Abschnitt 5 besprochen, die Polarisationsasymmetrie des Trainingsdatensatzes mit Hilfe des Tauspinnerprogrammes verändert werden kann. Anhand der Analyse der Vorhersage eines neuronalen Netzes mit dem Z-Datensatz soll diese Abhängigkeit untersucht, und in einem weiteren Schritt dekorreliert werden. Der Trainingsdatensatz mit Polarisationsasymmetrie ist physikalisch bedeutend näher am Z-Datensatz (gleicher Spin des Mutterteilchens und gleiche Polarisationsasymmetrie der Tau-Leptonen). Diese Dekorrelation dient der „Übung“, um später den Spin und die Masse zu dekorrelieren.

6.2 Physikalischer Zusammenhang

Hier soll die physikalisch Abhängigkeit zwischen Di-Tau Masse und Polarisationsasymmetrie der Tau-Leptonen erläutert werden. Die Asymmetrieeinformation wird dem Netz grundlegend in Form von Ladung, und der Information gegeben, welcher Teil der Energie (aus dem Ruhesystem des Tau-Leptons) nach vorne, und welcher Teil nach hinten ausgesandt wird. Es wird zunächst auf den Zerfall eines Tau-Leptons in ein Pion und ein Tau-Neutrino:

$$\tau^- \rightarrow \pi^- \nu_\tau$$

eingegangen. In diesem Zerfall übernimmt das Neutrino den Spin des Tau-Leptons, da das Pion einen Spin von null hat. Der Spin des Neutrinos muss antiparallel zu seiner Bewegungsrichtung ausgerichtet sein. Die Flugrichtung des Pions ist im Ruhesystem des Tau-Leptons antiparallel zu der des Neutrinos.

Für die zwei unterschiedlichen Helizitätszustände des Tau-Leptons können die Flugrichtungen und Spinausrichtungen in Abbildung 6.3 geschlussfolgert werden.

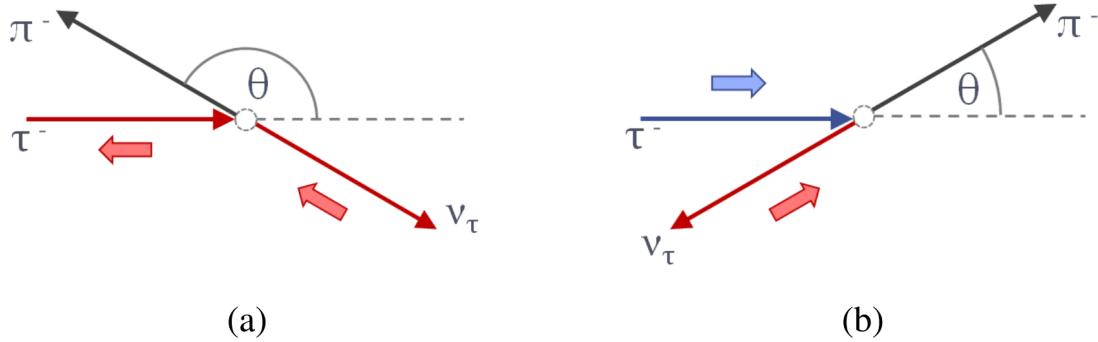


Abbildung 6.3: Flugrichtungen und Bewegungsausrichtungen für den Zerfall $\tau^- \rightarrow \pi^- \nu_\tau$ für den Fall eines (a) linkshändigen Tau-Leptons, (b) rechtshändigen Tau-Leptons. Die schmalen Pfeile geben die Bewegungsrichtung, die breiten Pfeile die Spinausrichtung an. [40]

Die Energie des Pions ist im Falle eines rechtshändigen Tau-Leptons höher als im Falle eines linkshändigen Tau-Leptons. Es kann also aus der Pionenergie auf die Helizität der Tau-Leptonen geschlossen werden. Ein weiterer Zerfall ist der eines Tau-Leptons in ein ρ -Meson, das weiter in zwei Pionen zerfällt, und ein Tau-Neutrino:

$$\tau^- \rightarrow \rho^- \nu_\tau \quad \text{und} \quad \rho^- \rightarrow \pi^- \pi^0$$

Dieser Zerfall hat das höchste Verzweigungsverhältnis unter den hadronischen Tau-Zerfällen (vgl. Tabelle 2.1). Hierbei ist der Zusammenhang zwischen sichtbarer Energie und Helizität der Tau-Leptonen nicht so einfach wie im ersten Zerfall. Das liegt daran, dass das ρ -Meson einen Spin von eins hat. Der Spin des ρ -Mesons kann entweder longitudinal oder transversal zu seiner Bewegungsrichtung ausgerichtet sein. Die Spinerhaltung fordert, dass sich der Spin des ρ -Mesons und der Spin des Neutrinos zum Spin des Tau-Leptons ergeben müssen. Es sind also für die beiden Helizitäten des Tau-Leptons die zwei Zerfälle in Abbildung 6.4 möglich. Es sind nur zwei von vier möglichen Flugrichtungen und Ausrichtungen der Spins dargestellt, allerdings werden solche Zerfälle präferiert, in denen das ρ -Meson in Flugrichtung ausgesandt wird [40]. Da das ρ -Meson in beiden Fällen in Abbildung 6.4 die selbe Energie hat, kann aus der Energie des ρ -Mesons alleine nicht auf die Helizität des Tau-Leptons geschlossen werden. Die Information über die Polarisation des ρ -Mesons ist nötig, um auf die Helizität des Tau-Leptons schließen zu können. Die Polarisation des ρ -Mesons kann durch die Analyse seines Zerfalles in zwei Pionen $\rho^- \rightarrow \pi^- \pi^0$ gefunden werden. Das Pion hat einen Spin von Null. Das bedeutet, dass der Spin des ρ -Mesons beim Zerfall in einen Bahndrehimpuls umgewandelt wird. Für ein transversal polarisiertes ρ -Meson bedeutet das eine Aussendung der Pionen in entgegengesetzte Richtungen, orthogonal zu der Ausrichtung des Spins des ρ -Mesons. Für eine longitudinale Polarisation bedeutet das allerdings, dass ein Pion, um die Spinerhaltung zu gewährleisten, in Flugrichtung, und das andere Pion antiparallel zur Flugrichtung ausgesandt wird. Dieser Vorgang ist in Abbildung 6.5 illustriert. Im Falle eines transversal polarisierten ρ -Mesons ist die Energie also gleichmäßig auf die Pionen verteilt, während das eine Pion bei einer longitudinalen Polarisation des ρ -Mesons mehr Energie hat als das andere. Die Energieasymmetrie der Pionen enthält die Information über die Helizität des Tau-Leptons. Die Energieasymmetrie Y ist

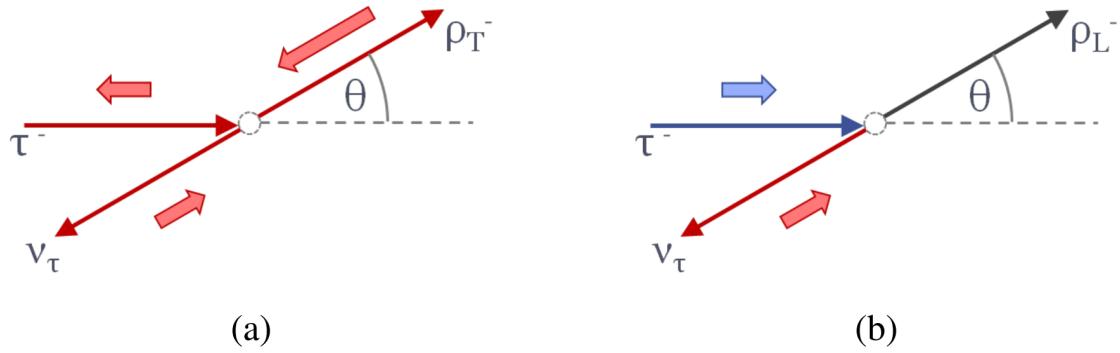


Abbildung 6.4: Flugrichtungen und Bewegungsausrichtungen für den Zerfall $\tau^- \rightarrow \rho^- \nu_\tau$ für den Fall eines (a) linkshändigen Tau-Leptons, (b) rechtshändigen Tau-Leptons. L und T signalisieren die longitudinale, bzw. transversale Polarisation des ρ -Mesons. Die schmalen Pfeile geben die Bewegungsrichtung, die breiten Pfeile die Spinausrichtung an. [40]

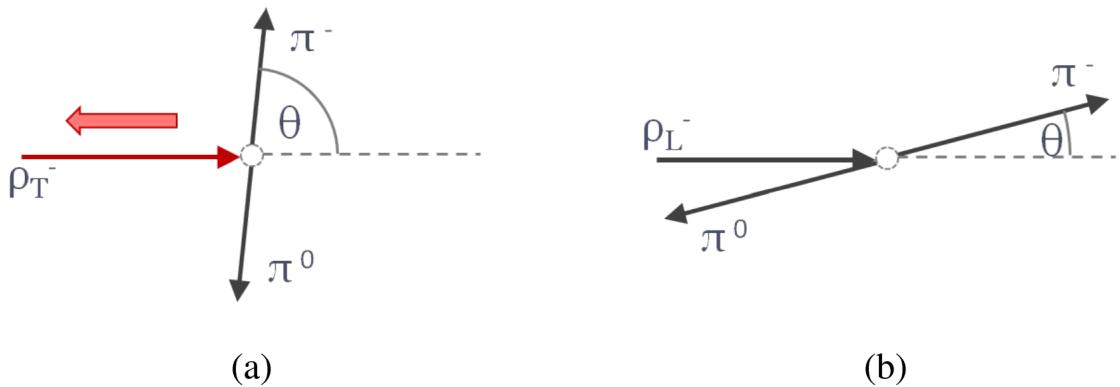


Abbildung 6.5: Verschiedene Flugrichtungen und Bewegungsausrichtungen für den Zerfall $\rho^- \rightarrow \pi^- \pi^0$ für den Fall eines (a) transversal und (b) longitudinal polarisierten ρ -Mesons. Die Energie wird in den beiden Fällen unterschiedlich auf die Pionen verteilt. Die schmalen Pfeile geben die Bewegungsrichtung, die breiten Pfeile die Spinausrichtung an. [40]

gegeben durch [40]:

$$Y = \frac{E_{\pi^-} - E_{\pi^0}}{E_{\pi^-} + E_{\pi^0}}$$

Die Information über die Asymmetrie in der sichtbaren Energie lässt also Rückschlüsse auf die Tau-Polarsationsasymmetrie ziehen.

KAPITEL 7

Training und Netzarchitektur

Das grundlegende Ziel ist zu untersuchen, wie stark die Polarisationsasymmetrie der Tau-Leptonen die Massenvorhersage beeinflusst und das Netz anschließend gegen diesen Einfluss zu desensibilisieren. Es wurden zwei Trainingsvorgänge durchgeführt; ein Training mit der simulierten Polarisationsasymmetrie, beim dem die Tauspinnergewichte angewandt wurden, und ein Training ohne Polarisationsasymmetrie, bei dem die Gewichte nicht angewandt worden sind. Im Folgenden meint Training mit und ohne Polarisationsasymmetrie das Training mit An- bzw. nicht Anwendung der Tauspinnergewichte während des Trainingsvorgangs. In diesem Abschnitt werden die Ein- und Ausgabedaten, sowie deren Verarbeitung erläutert. Es werden die verwendeten Hyperparameter und anschließend Methoden zur Dekorrelation vorgestellt.

7.1 Ein- und Ausgabedaten

Die Ausgabe des Neuronalen Netzes wird die Di-Tau Masse sein. Die Eingabegrößen enthalten die physikalische Information über den Zerfall, und beeinflussen den Trainingsprozess elementar. Es folgt eine Auflistung der EingabevARIABLEN mit jeweiliger physikalischer Begründung. Es werden die Bezeichnungen τ_0 und τ_1 verwendet. Diese stehen für die höher- und niederenergetischen Tau-Leptonen im Zerfall. Es werden ausschließlich die Ereignisse graphisch dargestellt, die die in Abschnitt 5.3 genannte Vorselektion erfüllen. Alle graphischen Darstellungen entsprechen den Ereignissen aus dem Trainingsdatensatz.

Ladung der Tau-Leptonen

Die Polarisationsasymmetrie beschreibt, wie in Abschnitt 6 beschrieben, den Zusammenhang zwischen Ladung und Spin der Tau-Leptonen. Ein großer Teil der Information bezüglich der Polarisationsasymmetrie der Tau-Leptonen wird dem neuronalen Netz durch die Ladung als Eingabegröße zur Verfügung gestellt. In Abbildung 7.1 ist die Ladung der Tau-Leptonen dargestellt.

Es sind auch Ereignisse zu sehen, bei denen die Ladung der Tau-Leptonen null, oder ± 2 beträgt. Das ist unphysikalisch, und liegt an fehlerhafter Rekonstruktion der Ereignisse.

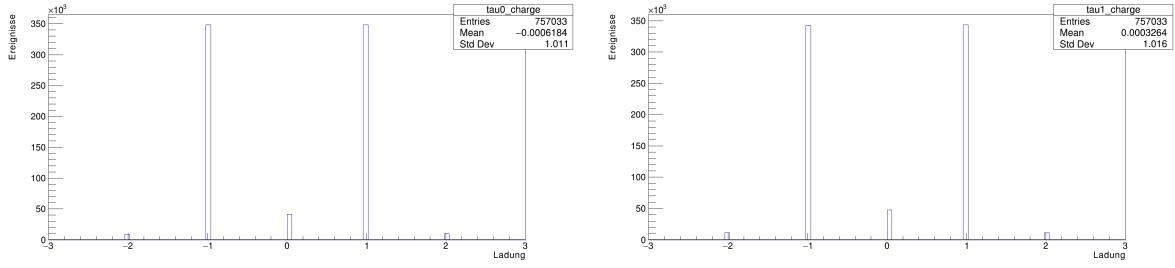


Abbildung 7.1: Ladung des höher- (links) und niederenergetischen (rechts) Tau-Leptons für alle Ereignisse aus dem Trainingsdatensatz, die die Vorselektion aus Abschnitt 5.3 erfüllen

Zerfallsprodukte der Tau-Leptonen

Die Zerfallsprodukte werden unterschieden in detektierbare, und nicht detektierbare Teilchen (Neutrinos). Einerseits werden dem Netz die Variablen p_T (Abbildung 7.2), η (Abbildung 7.3) und ϕ (Abbildung 7.4) der sichtbaren Zerfallsprodukte, andererseits der Betrag und der Azimutwinkel ϕ der fehlenden Transversalenergie (Abbildung 7.5) als Eingabegrößen zur Verfügung gestellt.

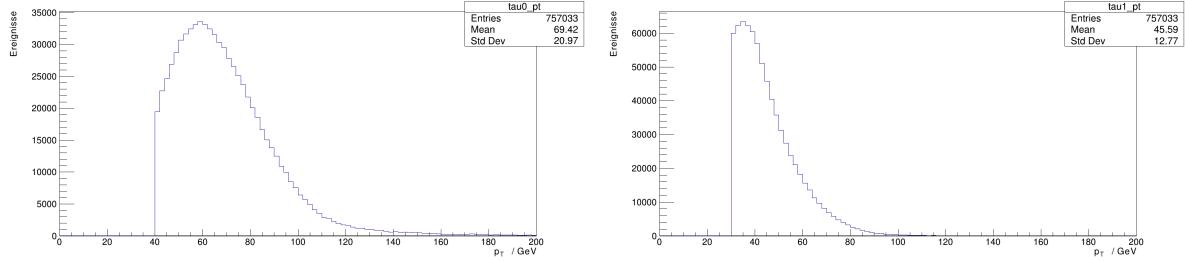


Abbildung 7.2: Transversalimpuls des höher- (links) und niederenergetischen (rechts) Tau-Leptons für alle Ereignisse aus dem Trainingsdatensatz, die die Vorselektion aus Abschnitt 5.3 erfüllen

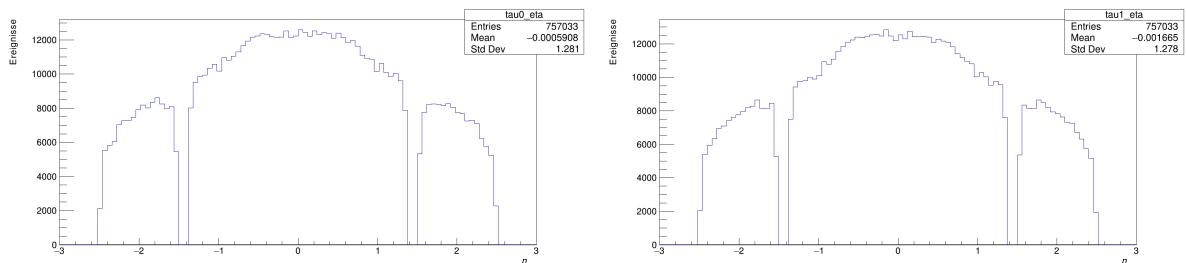


Abbildung 7.3: Pseudorapidität des höher- (links) und niederenergetischen (rechts) Tau-Leptons für alle Ereignisse aus dem Trainingsdatensatz, die die Vorselektion aus Abschnitt 5.3 erfüllen

Die fehlende Transversalenergie enthält alle bekannten Informationen über die Neutrinos. Geladene und neutrale Zerfallsprodukte der Tau-Leptonen werden in dieser Arbeit nicht unterschieden. Es ist allerdings zu erwarten, dass die zusätzliche Information über die einzelnen Vier-Impulse (der geladenen und

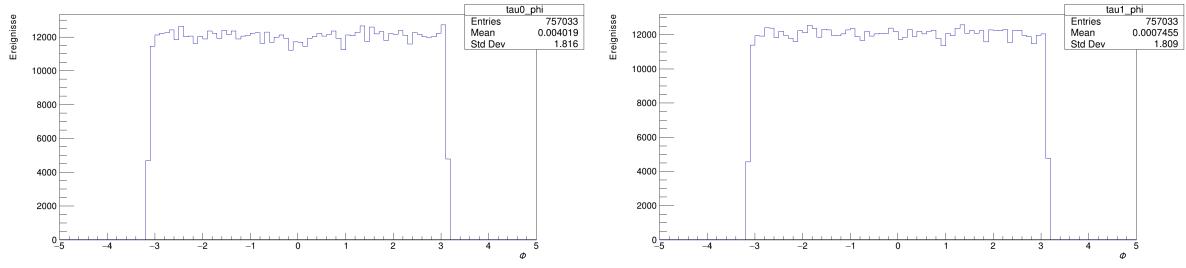


Abbildung 7.4: Azimutwinkel des höher- (links) und niederenergetischen (rechts) Tau-Leptons für alle Ereignisse aus dem Trainingsdatensatz, die die Vorselektion aus Abschnitt 5.3 erfüllen

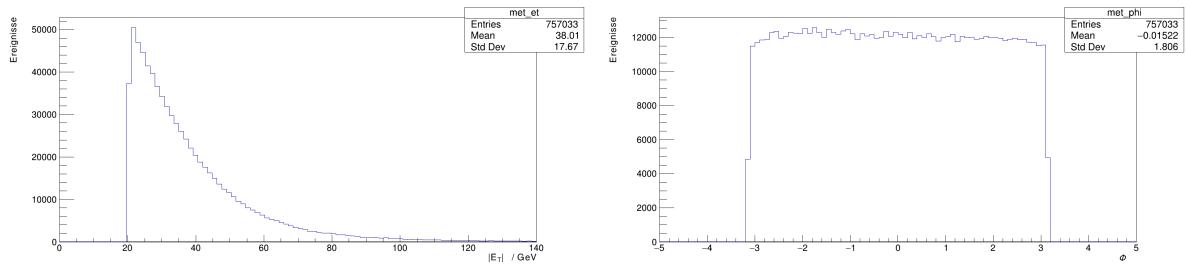


Abbildung 7.5: Betrag (links) und Azimutwinkel (rechts) der fehlenden Transversalenergie für alle Ereignisse aus dem Trainingsdatensatz, die die Vorselektion aus Abschnitt 5.3 erfüllen

neutralen Zerfallsprodukte) die Korrelation zwischen Massenvorhersage und Polarisationsasymmetrie noch deutlicher zum Vorschein bringt.

Jets

Jets sind Bündel aus Teilchen und entstehen durch Hadronisierung. Werden zwei gebundene Quarks bei der Kollision der Protonen im LHC voneinander weg beschleunigt, so wächst die zwischen ihnen gespeicherte Energie mit ihrem Abstand an. An einem gewissen Punkt ist die Energie groß genug, um ein neues Quark-Antiquark-Paar zu bilden. Dieser Vorgang wiederholt sich vielfach; es entsteht ein Bündel aus Teilchen. Die beim Zerfall der Tau-Leptonen entstehenden Jets können Information über den Boost des zu Grunde liegenden Mutterteilchens enthalten. Die Variablen p_T , η und ϕ der drei energiereichsten Jets (Abbildung 7.6 für den energiereichsten Jet) werden dem Netz als Eingabe gegeben. Die graphische Darstellung der zwei weiteren Jets ist im Anhang in Abbildung A.1 und A.2 zu finden. Wird ein Jet für ein bestimmtes Ereignis nicht rekonstruiert, so wird ein Eintrag bei null gesetzt. Der verwendete Eingabedatensatz ist nicht der mit der meisten physikalischen (insbesondere zerfallskinematischen) Information. Allerdings lassen sich die Hypothesen auch hiermit überprüfen, wie der weitere Verlauf zeigen wird. In weiterführenden Studien kann der Einfluss weiterer Eingabegrößen untersucht werden.

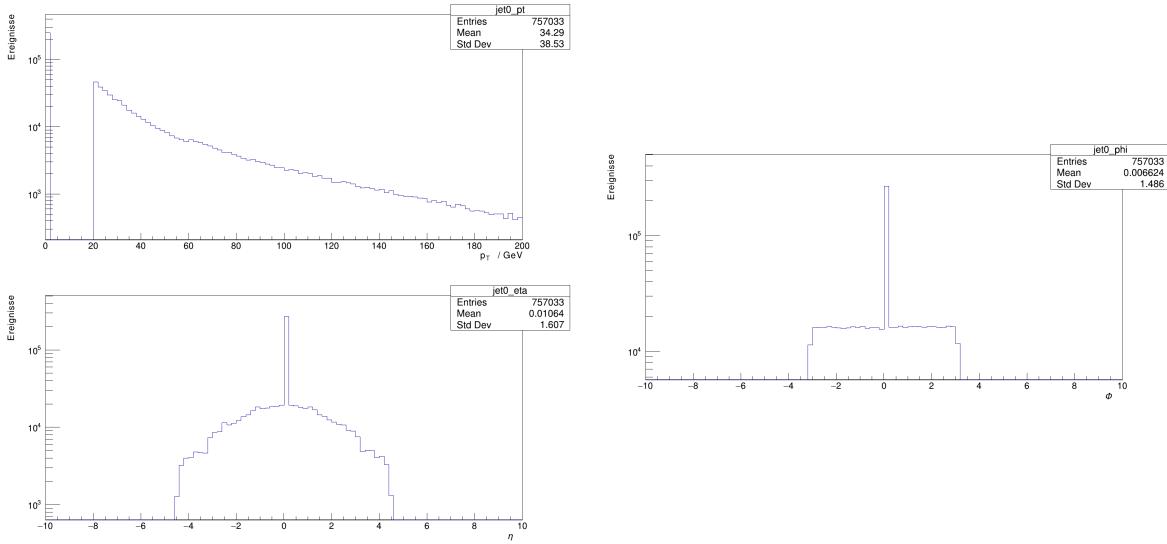


Abbildung 7.6: Transversalimpuls p_T , Pseudorapidität η und Azimutwinkel Φ für den Jet mit der höchsten Energie, für alle Ereignisse aus dem Trainingsdatensatz, die die Vorselektion aus Abschnitt 5.3 erfüllen, logarithmisch dargestellt. Wird ein Jet nicht rekonstruiert, entspricht das einem Eintrag bei null.

7.2 Verarbeitung der Daten

Für das Training des Netzes und die Vorhersagen werden unterschiedliche Teile des Trainingsdatensatzes verwendet. Es werden 80 % der im Datensatz vorhandenen Ereignisse für das Training des neuronalen Netzes verwendet und als Trainingsteil bezeichnet. 20 % werden für die Vorhersage genutzt und als Validationsteil bezeichnet. Die Verteilungen der Eingabeveriablen bewegen sich in sehr unterschiedlichen Wertebereichen. Um mögliche Faktoren zu eliminieren, die den Trainingsprozess einschränken könnten, werden die Daten vor dem Training skaliert, sodass die Standardabweichung der Verteilungen der Eingabeveriablen eins, und deren Mittelwert null beträgt [41]. Dazu wird der Mittelwert der jeweiligen Verteilung μ von jedem Wert innerhalb der Verteilung x subtrahiert, und durch die Standardabweichung σ geteilt. Es gilt dann für den resultierenden Wert z [41]:

$$z = \frac{x - \mu}{\sigma}$$

7.3 Hyperparameter

Es wurde eine Mean-Squared-Error Lossfunktion (Gleichung 4.1) für das Training beider Netze verwendet. Die Aktivierungsfunktion der Neuronen war die ReLU-Funktion (kurz für engl. Rectified Linear Activation Unit) [42]. Der Optimierer war Adam [29]. Es wurden vier verdeckte Schichten mit jeweils 16 Neuronen verwendet. Eine Batchsize von 32 wurde genutzt und die Learning-Rate war 0,001. Es wurden folgende Callbacks genutzt:

1. Sollte der Lernprozess des Netzes über eine definierte Anzahl an Epochen unverändert bleiben, so wird die Learning-Rate angepasst
2. Sollte der Wert des Loss NaN (engl. Not a Number), also nicht definiert werden, so wird das Training terminiert
3. Es werden der Loss und eine definierte Metrik während des Trainings aufgezeichnet

In Abbildung 7.7 ist die Entwicklung des Losswertes über die Epochen dargestellt. Allgemein sind Darstellungen dieser Art sehr hilfreich, um zu prüfen, wie das Netz trainiert hat. Der Losswert entspricht dem Wert der Lossfunktion ausgewertet für die Vorhersage des Netzes nach einer bestimmten Anzahl an Epochen. Der Losswert sollte mit der Anzahl an Epochen fallen, da das Training eines neuronalen Netzes mit der Minimierung einer Lossfunktion einhergeht. In Abbildung 7.7 sind zwei Kurven dargestellt; eine, die der Entwicklung des Losswertes des Trainingsteils entspricht, und eine Kurve, die der Entwicklung des Losswertes für den Validationsteil entspricht. Es werden diese Kurven gesondert betrachtet, da das Netz die Daten aus dem Validationsteil während des Trainingsvorgangs nicht verwendet, um die Gewichte anzupassen. Es wird die Vorhersage des Netzes auf Daten, die nicht im Trainingsvorgang vorkommen, betrachtet. Das ist relevant, da neuronale Netze in der praktischen Anwendung ausschließlich Vorhersagen für Daten machen, die nicht im Training verwendet wurden.

In Abbildung 7.7 ist zu erkennen, dass ab einer Epochenzahl von ca. 40 Übertraining (engl. Overtraining) stattfindet. Das bedeutet, dass das Netz an diesem Punkt nicht mehr allgemeine physikalische Eigenschaften, sondern solche spezifisch für den Trainingsteil des Trainingsdatensatzes lernt. Das Übertraining ist in Abbildung 7.7 daran zu erkennen, dass ab ca. 40 Epochen die Vorhersagen mit dem Trainingsteil immer besser werden, während die Vorhersagen auf dem Validationsteil gleich gut bleiben. Daher wurden für das Training im Rahmen dieser Arbeit stets 40 Epochen verwendet.

7.4 Methoden der Dekorrelation

In diesem Abschnitt werden das adversarial neural network und die Methode des Distance Correlation vorgestellt, die für die Dekorrelation verwendet werden können.

7.4.1 Adversarial neural network

Grundlegend besteht ein adversarial neural network aus zwei neuronalen Netzen wie in Abbildung 7.8 zu sehen.

Die Information verläuft in der Abbildung von links nach rechts. Das erste Netz (in Abbildung 7.8 links) ist das Regressionsnetz, das die Massenvorhersage formuliert. Die Vorhersage des ersten Netzes wird (gegebenenfalls mit weiteren Größen zusammen) in die Eingabeschicht des zweiten Netzes gegeben.

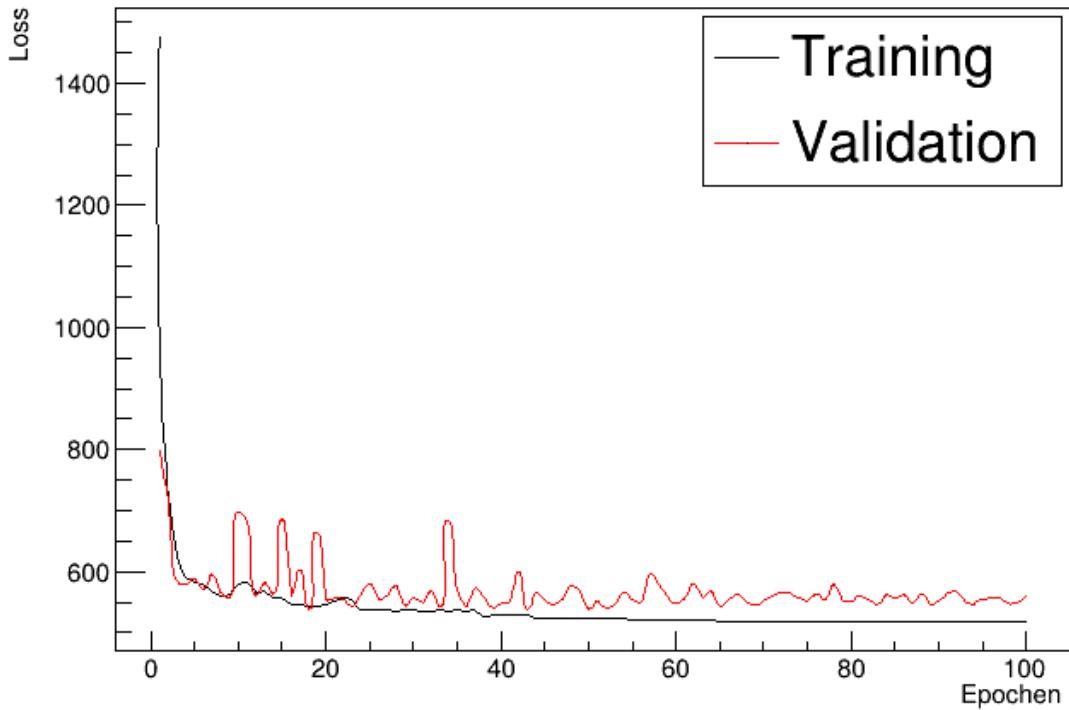


Abbildung 7.7: Verlauf der Losskurven für den Trainings- und Validationsteil, Training mit Polarisationsasymmetrie über 100 Epochen, ab einer Epochenzahl von ca. 40 kommt es zu Übertraining

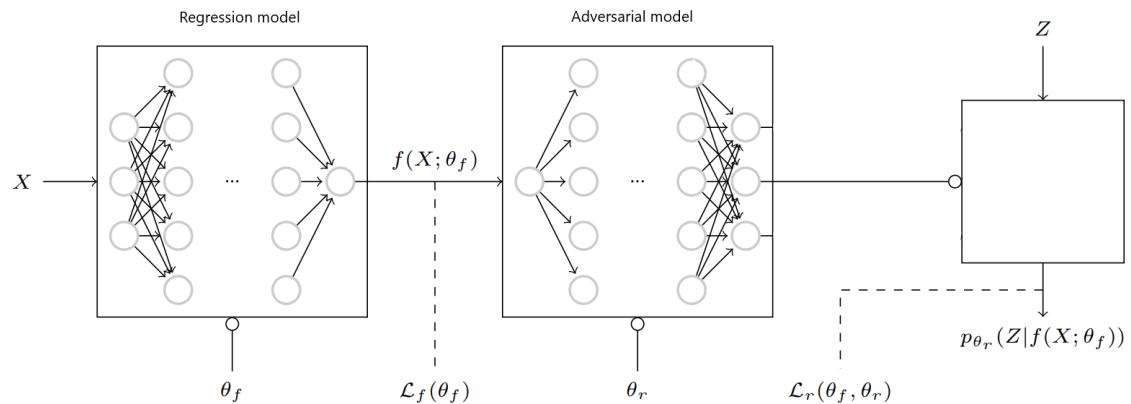


Abbildung 7.8: Adversarial neural network, bestehend aus dem Regressionsnetz (links) und dem adversarial Netz (rechts), wobei θ_f die Parameter des Regressions- und θ_r die Parameter des adversarial Netzes sind. Z ist der Störungsparameter, in unserem Fall die Helizität des τ^- . Diese wird vom adversarial Netz vorhergesagt, die Vorhersage wird mit steigender Dekorrelation schlechter. Bild aus [43], bearbeitet

Das Ziel des zweiten Netzes (adversarial Netz genannt) ist es, basierend auf der Massenvorhersage des ersten Netzes einen Parameter vorherzusagen, der im Zusammenhang mit der Polarisationsasymmetrie steht. Es wäre die Helizität des τ^- als Ausgabe des adversarial Netzes gewählt worden. Die Information über die Helizität eines Tau-Leptons reicht hier aus, da die Helizität des τ^+ aus der Helizität des τ^- unmittelbar geschlussfolgert werden kann.

Die Lossfunktion für die gesamte Struktur setzt sich aus der Lossfunktion des Regressionsnetzes L_{reg} und der Lossfunktion des adversarial Netzes L_{adv} zusammen und sieht wie folgt aus [43]:

$$L = L_{\text{reg}} - \lambda \cdot L_{\text{adv}}$$

wobei λ ein Faktor ist, der bestimmt, wie groß der Anteil der adversarial Lossfunktion L_{adv} an der Lossfunktion der gesamten Struktur L ist. Die Minimierung der Lossfunktion der gesamten Struktur L geht mit der Maximierung der Lossfunktion des adversarial Netzes L_{adv} einher. Die Vorhersage des ersten Netzes soll also besser werden, die Vorhersage des zweiten Netzes aber schlechter. Das ergibt intuitiv Sinn; wenn das adversarial Netz die Polarisationsasymmetrie richtig vorhersagen kann, dann gibt es eine Korrelation zwischen Massenvorhersage und Polarisationsasymmetrie. Wird die Vorhersage des adversarial Netzes immer schlechter, so wird auch die Korrelation zwischen Massenvorhersage und Polarisationsasymmetrie immer schwächer, kurz: Polarisationsasymmetrie und Massenvorhersage werden dekorreliert. Der Faktor λ entscheidet über den Anteil, den die Dekorrelation im Verhältnis zum Training des Regressionsnetzes vom gesamten Trainingsprozess beansprucht. Es können weitere Größen, über die Vorhersage des Regressionsnetzes hinaus, in das adversarial Netz gegeben werden. Siehe dafür Abschnitt 8.6.

7.4.2 Distance Correlation

Die Distance Correlation [44] (kurz: DisCo) ist eine weitere Methode, um Größen voneinander zu dekorrelieren. Der Vorteil gegenüber dem adversarial neural network ist, dass nur ein neuronales Netz, nämlich das Regressionsnetz benötigt wird. Die gesamte Dekorrelation findet während des Trainings statt und wird in einer Modifikation der Lossfunktion erreicht. Die gesamte Lossfunktion sieht wie folgt aus [44]:

$$L = L_{\text{reg}}(y_{\text{rek}}, y_{\text{true}}) + \lambda \cdot \text{dCorr}^2(y_{\text{true}}, h)$$

wobei L_{reg} die Lossfunktion des Regressionsnetzes ist und die Distance Correlation Funktion dCorr^2 angibt, wie sehr zwei Größen voneinander abhängig sind. Der Faktor λ beschreibt hier, analog zur Lossfunktion des adversarial neural networks, die Abwägung zwischen Dekorrelation und Qualität der Vorhersage des Regressionsnetzes. In der Lossfunktion des Regressionsnetzes tauchen die Vorhersage des Regressionsnetzes y_{rek} und der zugehörige wahre Wert y_{true} , in der Distance Correlation Funktion die Vorhersage des Regressionsnetzes und eine Größe h auf, die von y_{true} dekorreliert werden soll. In unserem Fall entspricht y_{true} der Masse und h der Helizität des τ^- . Um die Definition der Distance Correlation dCorr^2 zu verstehen, muss zuerst die Distance Covariance betrachtet werden [44]:

$$\text{dCov}^2(X, Y) = \int d^p s d^q t |f_{X,Y}(s, t) - f_X(s)f_Y(t)|^2 w(s, t)$$

f_X und f_Y sind die charakteristischen Funktionen zu den Größen $X \in \mathbb{R}^p$ und $Y \in \mathbb{R}^q$, die es zu dekorrelieren gilt, $f_{X,Y}$ ist die gemeinsame charakteristische Funktion der Variablen X und Y und $w(s, t)$

ist eine individuelle Gewichtsfunktion, auf die hier nicht näher eingegangen werden soll. Falls die Variablen X und Y unabhängig voneinander sind, so gilt:

$$f_{X,Y} = f_X \cdot f_Y$$

Die Funktion wird in diesem Fall also null. Die Distance Correlation lässt sich mit Hilfe der Distance Covariance definieren als [44]:

$$\text{dCorr}^2(X, Y) = \frac{\text{dCov}^2(X, Y)}{\text{dCov}(X, X) \cdot \text{dCov}(Y, Y)}$$

Sie kann nur Werte zwischen Null und eins annehmen. Das erleichtert ihre Handhabung deutlich; verliefe sie z.B. von $-\infty$ bis ∞ , so wäre deutlich schwerer abzuschätzen, wie groß der Einfluss auf die gesamte Lossfunktion ist. Insbesondere das Finden eines angemessen Wertes für λ wäre erheblich erschwert. Die Distance Correlation wird in Abschnitt 8.5 im Training verwendet. Es werden drei Werte für den Faktor λ getestet.

KAPITEL 8

Ergebnisse

In diesem Abschnitt werden die Ergebnisse besprochen. Es werden verschiedene Darstellungsmöglichkeiten erläutert, die dazu dienen verschiedene Aspekte der Vorhersage des Netzes zu untersuchen. Es werden zuerst Vorhersagen mit Netzen ohne die Methode der Distance Correlation untersucht. Dann wird auf die Methode der Distance Correlation für unterschiedliche Werte von $\lambda = 0, 1000, 10000$ (die Wahl dieser Werte wird motiviert) eingegangen. Anschließend werden die Ergebnisse kurz zusammengefasst und Ideen für weitere Untersuchungen vorgestellt.

Bei der Analyse der Ergebnisse wird stets die Vorhersage mit dem Trainingsdatensatz und dem Z-Datensatz unterschieden. Es finden grundsätzlich zwei verschiedene Trainings statt; ein Training mit, und ein Training ohne Polarisationsasymmetrie. Die entsprechende Vorhersage mit dem Validationsteil des Trainingsdatensatzes für das Training mit und ohne Polarisationsasymmetrie erfolgt dann auch unter entsprechender An- oder nicht Anwendung der Tauspinner Gewichte bei der Vorhersage.

8.1 Relative Massenauflösung

Um die Qualität der Vorhersage des neuronalen Netzes beurteilen zu können ist es essentiell, die Massenvorhersage im Verhältnis zur wahren Masse zu betrachten. Die Massenvorhersage des Netzes alleine kann, ohne Kenntnis der wahren Massenverteilung, nur sehr bedingt Aussagen über die Qualität der Vorhersage machen. Zunächst wird die Relative Auflösung der Massenvorhersage definiert. Für die relative Massenauflösung $A_{m, \text{rel}}$ gilt:

$$A_{m, \text{rel}} = \frac{m_{\text{rek}} - m_{\text{wahr}}}{m_{\text{wahr}}}$$

Hierbei ist m_{wahr} die wahre Masse, und m_{rek} ist die vom Netz vorhergesagte Masse. Im Idealfall entspricht die relative Massenauflösung einer Spitzte bei Null. In Abbildung 8.1 oben ist beispielhaft die vorhergesagte Massenverteilung mit dem Trainingsdatensatz für ein Training mit Polarisationsasymmetrie dargestellt. Idealerweise sollte diese Verteilung aussehen wie die in Abbildung 5.1 gezeigten wahren Massenverteilungen des Trainingsdatensatzes mit Polarisationsasymmetrie und ohne Polarisationsasymmetrie. Um die Abweichung zwischen der vorhergesagten und der wahren Massenverteilung darzustellen ist in Abbildung 8.1 unten die relative Massenauflösung zu sehen.

Das Training wurde viermal hintereinander mit der selben Anzahl an Epochen, jeweils mit und ohne Polarisationsasymmetrie durchgeführt. Insgesamt werden also acht Trainingsdurchläufe

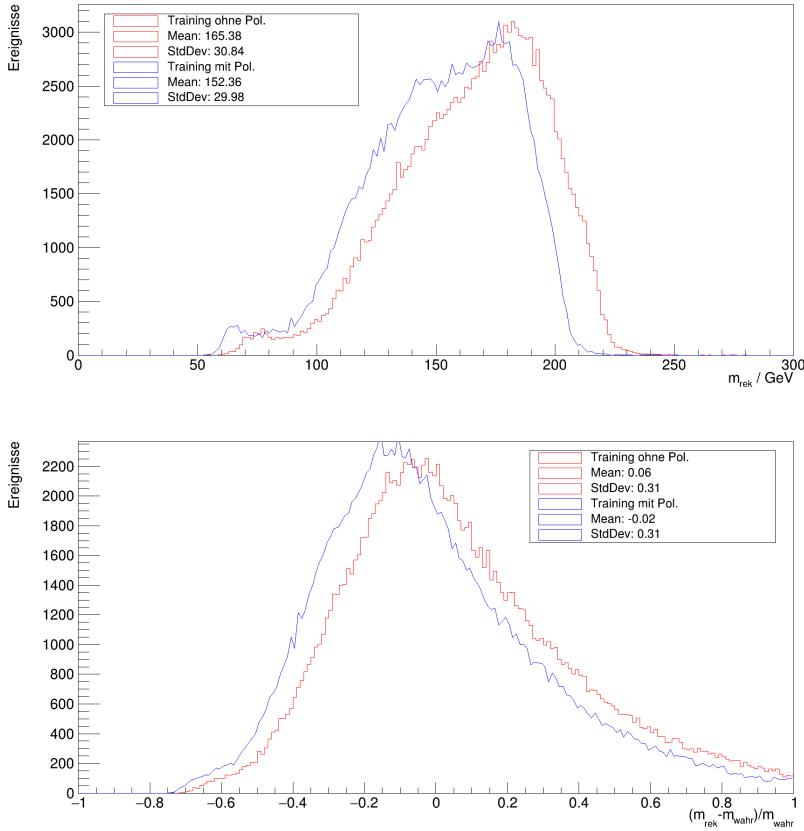


Abbildung 8.1: Massenvorhersage (oben) und relative Massenauflösung (unten) mit dem Trainingsdatensatz, Training Nr. 4 mit und ohne Polarisationsasymmetrie

(Training mit Polarisationsasymmetrie Nr. 1, 2, 3 und 4 und Training ohne Polarisationsasymmetrie Nr. 1, 2, 3 und 4) untersucht. In Abbildung 8.2 sind die resultierenden relativen Massenauflösungen für die Vorhersage mit Ereignissen aus dem Trainingsdatensatz, deren zugehörige Masse im Bereich der Z-Masse (Massenintervall: 85 bis 95 GeV) liegt (stets links), und mit allen Ereignissen aus dem Z-Datensatz (stets rechts) für die vier Trainingsdurchläufe dargestellt. Es wird ausschließlich dieser Ausschnitt des Trainingsdatensatzes für die Vorhersage genutzt, da das Intervall der Z-Masse das relevante Intervall für die Untersuchung ist. Es soll schließlich die Z-Masse möglichst präzise vorhergesagt werden.

Den gezeigten Abbildungen können zwei Aspekte entnommen werden:

1. Die Qualität der Vorhersage mit dem Z-Datensatz unterscheidet sich zwischen dem Training mit und ohne Polarisationsasymmetrie
2. Es entstehen für gleiche Trainingsdurchläufe unterschiedliche Vorhersagen

Auf diese Effekte wird im Folgenden eingegangen.

Die Qualität der Vorhersage des Trainings mit Polarisationsasymmetrie, ausgewertet auf dem Z-Datensatz ist im Mittel besser als die Qualität der Vorhersage des Trainings ohne Polarisationsasymmetrie.

Kapitel 8 Ergebnisse

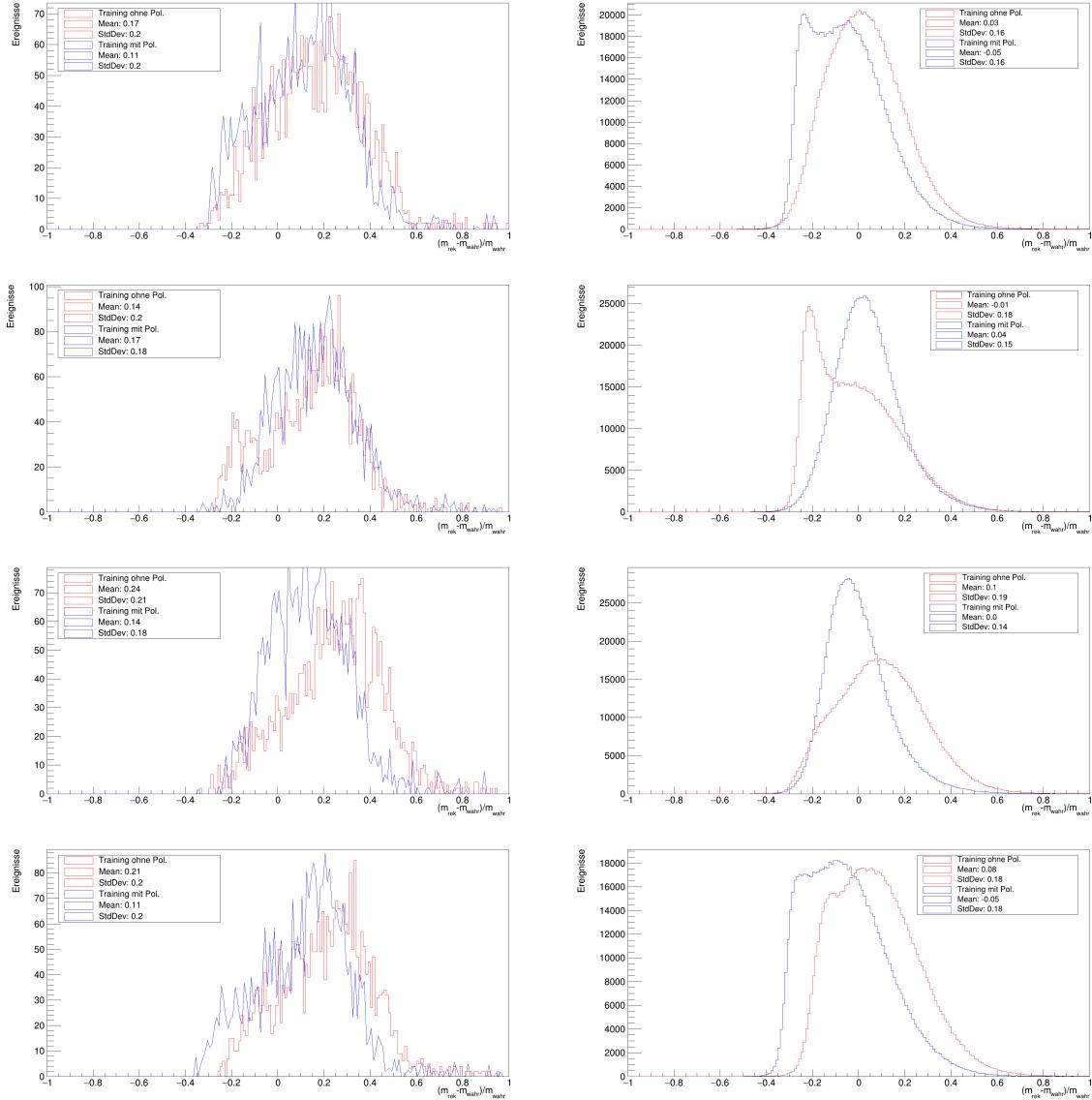


Abbildung 8.2: Relative Massenauflösung, bestimmt mit Daten aus dem Trainingsdatensatz im Bereich der Z-Masse (links), Z-Datensatz (rechts), Training Nr. 1, 2, 3, 4 absteigend, jeweils mit und ohne Polarisationsasymmetrie

Das sieht man am durchschnittlichen Mittelwert $\bar{\mu}$ und der durchschnittlichen Standardabweichung $\bar{\sigma}$ mit:

$$\bar{\mu}_{\text{mit Pol., Z}} = -0,015 \text{ GeV} \quad \bar{\sigma}_{\text{mit Pol., Z}} = 0,1575 \text{ GeV}$$

und ohne Polarisationsasymmetrie:

$$\bar{\mu}_{\text{ohne Pol., Z}} = 0,05 \text{ GeV} \quad \bar{\sigma}_{\text{ohne Pol., Z}} = 0,1775 \text{ GeV}$$

Auch mit dem bloßen Auge können klare Unterschiede zwischen Training mit und ohne Polarisationsasymmetrie, ausgewertet auf dem Z-Datensatz erkannt werden. Die Aussagekraft dieses Ergebnisses ist allerdings eingeschränkt, da nur vier Trainingsdurchläufe vorgenommen wurden. Dass das Training mit Polarisationsasymmetrie bessere Vorhersagen liefert, spiegelt, da der Trainingsdatensatz mit Polarisationsasymmetrie physikalisch näher am Z-Datensatz ist, wieder, dass die schlechte Vorhersage mit dem Z-Datensatz, die in [8] im Vergleich zur Qualität der Vorhersage mit dem Trainingsdatensatz gefunden wurde, tatsächlich auf physikalische Unterschiede zwischen Trainings- und Z-Datensatz zurückzuführen ist. Die zusätzliche Polarisationasymmetrie der Tau-Leptonen im Training führt zu einer besseren Vorhersage mit dem Z-Datensatz. Dieses Ergebnis ist eine Bestärkung der Hypothese, dass die Massenvorhersage des neuronalen Netzes abhängig von der Polarisationsasymmetrie der Tau-Leptonen ist.

Der zweite Punkt, der sich aus den Abbildung ableiten lässt ist, dass das Netz für gleiche Trainings unterschiedliche Vorhersagen mit dem Z und mit dem Trainingsdatensatz im Bereich der Z-Masse formuliert. Ein großes Problem an dieser Stelle sind die verhältnismäßig wenigen Ereignisse im Bereich der Z-Masse im Trainingsdatensatz (wie gut in Abbildung 8.2 zu erkennen ist). Das schränkt die Aussagekraft des Ergebnisses stark ein. Die Abweichungen der relativen Massenauflösungen für die verschiedenen Trainingsdurchläufe sind für den Z-Datensatz, verglichen mit dem Trainingsdatensatz im Bereich der Z-Masse für das Training ohne Polarisationsasymmetrie deutlicher erkennbar (siehe z.B. Abbildung 8.2, Training zwei: Es lässt sich in der relativen Massenauflösung mit dem Trainingsdatensatz links in der Verteilung eine Spitze erkennen, allerdings ist diese keineswegs so ausgeprägt wie die in der relativen Massenauflösung für den Z-Datensatz). Die schwächer ausgeprägten Diskrepanzen zwischen den verschiedenen Trainingsdurchläufen in der relativen Massenvorhersage für das Training ohne Polarisationsasymmetrie können unter Anderem darin begründet liegen, dass die Vorhersage für das Training ohne Polarisationsasymmetrie mit nicht polarisationsasymmetriebehafteten Daten formuliert wurde. Die relative Massenauflösung für das Training ohne Polarisationsasymmetrie mit dem Trainingsdatensatz ist hier also nicht im Stande, einen überzeugenden Vergleich zur Vorhersage des Z-Datensatzes für das Training ohne Polarisationsasymmetrie anzustellen.

Für das Training mit Polarisationsasymmetrie haben die vier verschiedenen Trainingsdurchläufe für die Auswertung mit dem Trainingsdatensatz keine so starken Unterschiede für die verschiedenen Trainingsdurchläufe in Mittelwert und Standardabweichung der relativen Massenauflösung zur Folge, wie für die mit dem Z-Datensatz bestimmten relativen Massenauflösungen. Der größte Unterschied in der relativen Massenauflösung für den Trainingsdatensatz liegt für den Mittelwert bei 0,06 GeV und für die Standardabweichung bei 0,02 GeV. Für die Auswertung mit dem Z-Datensatz hingegen liegt der größte Unterschied unter den vier relativen Massenauflösungen für den Mittelwert bei 0,09 GeV und für die Standardabweichung bei 0,04 GeV. Für das Training mit Polarisationsasymmetrie sind die Unterschiede in den relativen Massenauflösungen durch die verschiedenen Trainingsdurchläufe für den Z-Datensatz ausgeprägter als für den Trainingsdatensatz. Daher stehen der Übertragbarkeit des

Z-Datensatzes auf ein mit dem γ^* -Datensatz trainiertes Netz potentiell weitere Faktoren, abgesehen von der Polarisationsasymmetrie und dem Spin im Weg. Mit diesen möglichen Faktoren wird sich in Abschnitt 8.2 beschäftigt.

Für weitere Studien ist es möglich, mit den Netzen, die mit und ohne Polarisationsasymmetrie trainiert wurden, Vorhersagen für die Daten aus dem Trainingsdatensatz mit Polarisation zu machen. So könnten die resultierenden relativen Massenauflösungen besser mit denen verglichen werden, die durch Vorhersage mit dem Z-Datensatz entstanden sind. Die relative Massenauflösung für die vier Trainingsdurchläufe, die mit allen Ereignissen aus dem Trainingsdatensatz bestimmt worden ist, ist im Anhang in Abbildung A.3 dargestellt.

Um diesen Effekt weiter zu untersuchen, wurden die Gewichte für alle Eingabeveriablen zwischen dem jeweiligen Eingabeneuron und jedes Neurons in der ersten verdeckten Schicht betragsmäßig aufaddiert und der resultierende Werte in Abbildung 8.3 für die ersten beiden Trainingsdurchläufe dargestellt. Der Wert der Beträge aller Gewichte zwischen den ersten beiden Schichten für eine Eingabeveriable aufsummiert soll im Folgenden als Gewichtung bezeichnet werden. Es wird hier auf die Darstellung der weiteren Trainingsdurchläufe verzichtet, da der Effekt bereits so ausreichend illustriert werden kann. Die Gewichte der Trainingsdurchläufe Nr. 3 und 4 sind im Anhang in Abbildung A.4 dargestellt. Die Gewichte in den weiteren Schichten sind irrelevant für die Interpretation, da die Neuronen untereinander verbunden sind, und man somit die in Abbildung 8.3 dargestellten Verteilungen nur um einen (für jedes Training unterschiedlichen) konstanten Wert verschieben würde.

Die Gewichtungen sehen für die unterschiedlichen Trainingsdurchläufe (sowohl für das Training mit Polarisationsasymmetrie als auch für das Training ohne Polarisationsasymmetrie) unterschiedlich aus. Der Trainingsprozess läuft unterschiedlich ab. Diese Tatsache spiegelt sich auch in den verschiedenen Vorhersagen in Abbildung 8.2 wieder. Das Training an sich ist potentiell also instabiler als in vorherigen Arbeiten ([25] und [8]) angenommen. Im Folgenden sollen mögliche Ursachen für die Unterschiede besprochen werden.

8.2 Mögliche Ursachen

Prinzipiell kommen drei Ursachen für die unterschiedlichen Vorhersage mit Trainings- und Z-Datensatz in Frage:

1. Grundlegende physikalische Unterschiede zwischen den Datensätzen
2. Unterschiede in der Simulation der Datensätze
3. Unterschiede in der Vorselektion der Datensätze

Der einzige physikalische Aspekt, der den Trainingsdatensatz vom Z-Datensatz unterscheidet ist, dass die Polarisationsasymmetrie der Tau-Leptonen im Z-Datensatz nicht homogen über die Massenverteilung verläuft. Die Polarisationsasymmetrie im Trainingsdatensatz ist dagegen gleichmäßig über die Massenverteilung verteilt. Um mögliche, daraus resultierende Einflüsse zu untersuchen, werden alle Eingabeveriablen (außer den Jets) für den Z-Datensatz und für den Trainingsdatensatz mit Polarisationsasymmetrie im Bereich der Z-Masse miteinander verglichen. Die entsprechenden Darstellungen des Transversalimpulses p_T für das höher- und niederenergetische Tau-Lepton und des Betrages der fehlenden Transversalenergie $|E_T|$ sind in Abbildung 8.4 normiert dargestellt. Die Darstellung der weiteren Eingabeveriablen ist im Anhang in Abbildung A.5 zu sehen.

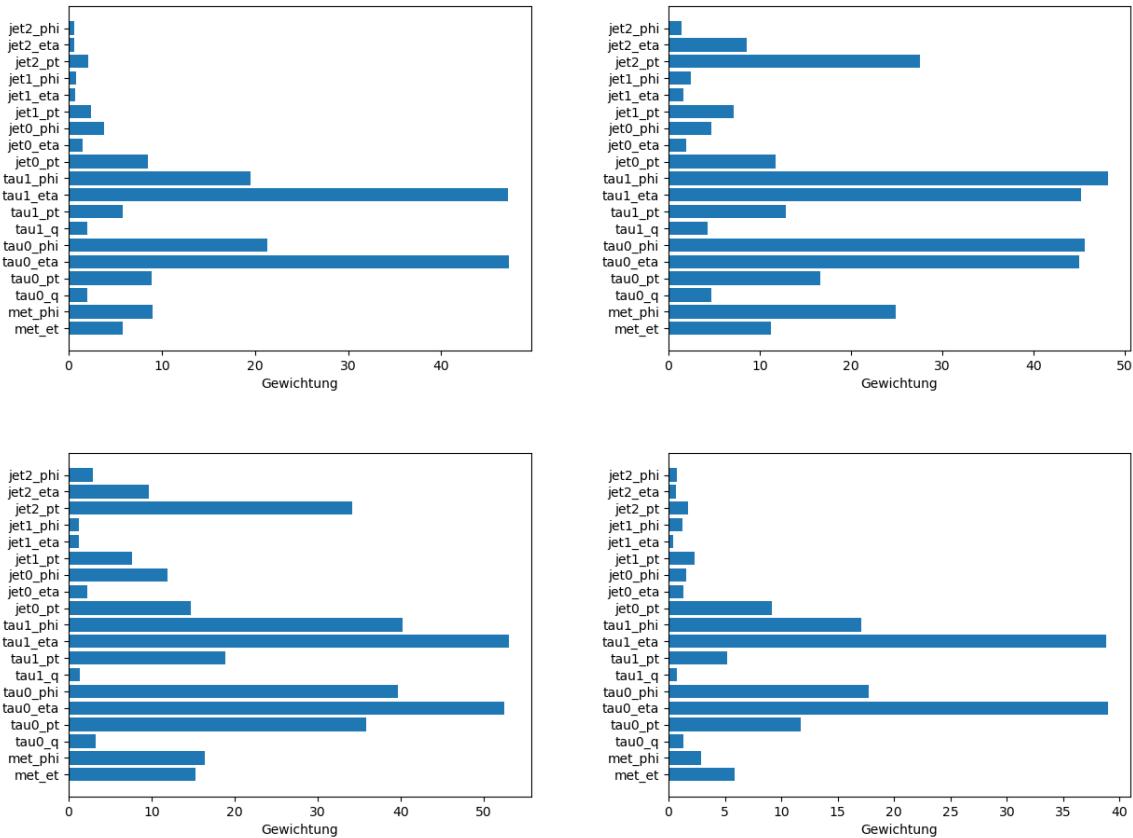


Abbildung 8.3: Gewichtungen mit (oben) und ohne (unten) Polarisationsasymmetrie zwischen den ersten beiden Schichten, Training Nr. 1 (links) und 2 (rechts)

Die Verteilung der Transversalimpulse und des Betrages der fehlenden Transversalenergie unterscheiden sich stark zwischen Trainings- und Z-Datensatz. Die Unterschiede sind für das niederenergetische Tau-Lepton schwächer. Dieser Aspekt kann für weitere Untersuchungen dieses Aspektes hilfreich sein. Die beiden Datensätze sind durch zwei unterschiedliche Simulationen entstanden; Sherpa [33] und Pythia [34]. Verschiedene Aspekte der unterschiedlichen Simulationstechniken können für Unterschiede zwischen den Datensätzen sorgen. Die weitere Untersuchung dieses Aspektes geht allerdings über den Rahmen dieser Bachelorarbeit hinaus.

Die Vorselektion, die auf die Datensätze angewandt wurden, sind in Abschnitt 5.3 gelistet. Es folgt eine kurze graphische Bestätigung der Vorselektion. In Abbildung 8.5 sind die Pseudorapidität η , und der Transversalimpuls p_T für das höher- (τ_0) und niederenergetische (τ_1) Tau-Lepton jeweils für alle Ereignisse aus dem Trainings- und aus dem Z-Datensatz dargestellt.

Die Variablen beider Tau-Leptonen unterliegen den Beschränkungen, die in Abschnitt 5 gelistet sind. Die Vorselektion unterscheidet sich also zwischen Trainings- und Z-Datensatz nicht und kann damit nicht die unterschiedlichen Vorhersagen mit dem Z-Datensatz für die verschiedenen Trainingsdurchläufe erklären.

Kapitel 8 Ergebnisse

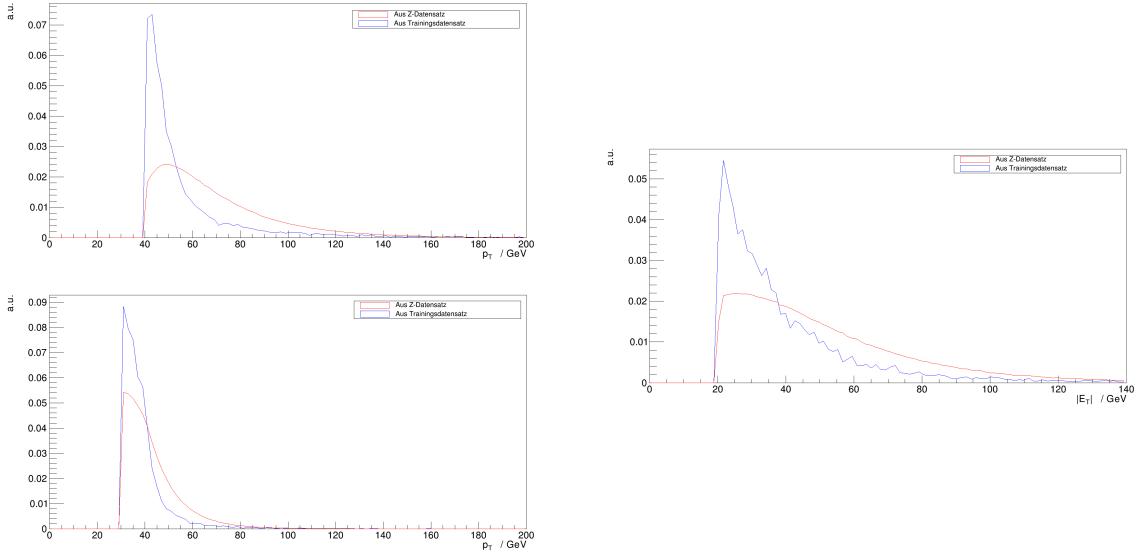


Abbildung 8.4: Normierte Verteilung des Transversalimpulses p_T (links) des höher- (oben) und niedrigerenergetischen (unten) Tau-Leptons, und des Betrages der fehlenden Transversalenergie $|E_T|$ (rechts) für die Ereignisse aus dem Trainingsdatensatz mit Polarisationsasymmetrie im Bereich der Z-Masse und Ereignisse aus dem Z-Datensatz, a.u.: arbitrary unit (deutsch: willkürliche Einheit)

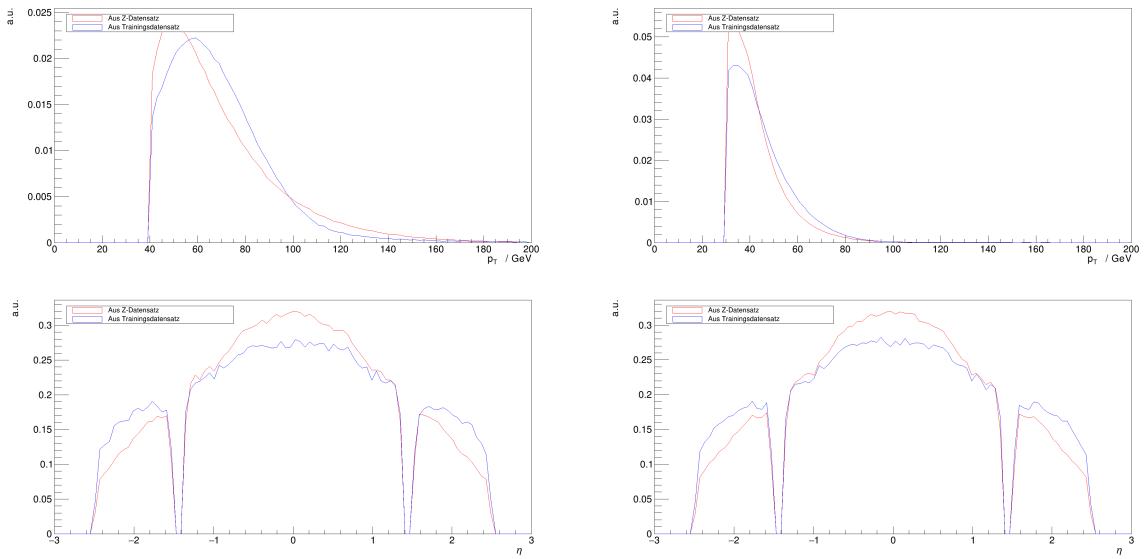


Abbildung 8.5: Bestätigung der Vorselektion; Transversalimpuls p_T (oben) und Pseudorapidität η (unten) für alle Ereignisse aus Trainings- und Z-Datensatz, für τ_0 (links) und τ_1 (rechts)

8.3 Relative Massenauflösung in Intervallen der wahren Masse

Die Darstellung der Massenauflösung in Intervallen der wahren Masse ist eine praktische Darstellungsmöglichkeit, die weitere Informationen über die Qualität der Vorhersage des neuronalen Netzes birgt. Um herauszufinden, ob die relative Massenauflösung abhängig von der wahren Masse variiert, wird die relative Massenauflösung in Intervallen der wahren Masse dargestellt. Die Darstellung soll hier Schritt für Schritt anhand des vierten Trainingsdurchlaufs mit Polarisationsasymmetrie, ausgewertet auf allen Ereignissen des Trainingsdatensatzes, konstruiert werden. Um diese Darstellung zu erreichen werden die Ereignisse zuerst in Intervallen der wahren Masse angeordnet. In diesem Beispiel sind es 16 Intervalle, die Intervalle sind also 10 GeV groß (insgesamt ist der Massenbereich 160 GeV breit). Es wird an die Massenverteilungen der jeweiligen Intervalle eine Gaußkurve angepasst. Dieser Vorgang ist exemplarisch für das Intervall von 90 bis 100 GeV in Abbildung 8.6 oben dargestellt. Die Mittelwerte und Standardabweichungen der Gaußanpassungen der relativen Massenauflösungen aller Massenintervalle werden abhängig von den Mittelwerten der betrachteten Intervalle der wahren Masse aufgetragen. Die Darstellung ist in Abbildung 8.6 mittig dargestellt. Wahre Massen zwischen 95 GeV und 215 GeV weisen einen klaren Abwärtstrend auf. Die Ereignisse, deren wahre Masse im Bereich 65 GeV und 85 GeV liegt, weichen von diesem Trend ab. Das kann unter Anderem daran liegen, dass in diesen Bereichen weniger Statistik vorliegt, d.h. es weniger Ereignisse mit Di-Tau Massen im Bereich 60 bis 70 GeV gibt, als z.B. für das Massenintervall 90 bis 100 GeV. Die Massenverteilung für das erste Intervall (60 bis 70 GeV) der wahren Masse ist in Abbildung 8.6 unten exemplarisch dargestellt.

In diesem Massenbereich sind weniger Ereignisse vorhanden. Außerdem ist die Verteilung deutlich asymmetrischer als in dem Intervall 95 GeV (Abbildung 8.6 oben). Die erhöhte Anzahl an Ereignissen hin zu größeren rekonstruierten Massen und damit einhergehende Asymmetrie der Verteilung beeinträchtigt die Anpassung der Gaußkurve. An dieser Stelle kann auf andere Größen zurückgegriffen werden, um die Physik hinter der Verteilung möglichst präzise wiederzuspiegeln.

Kapitel 8 Ergebnisse

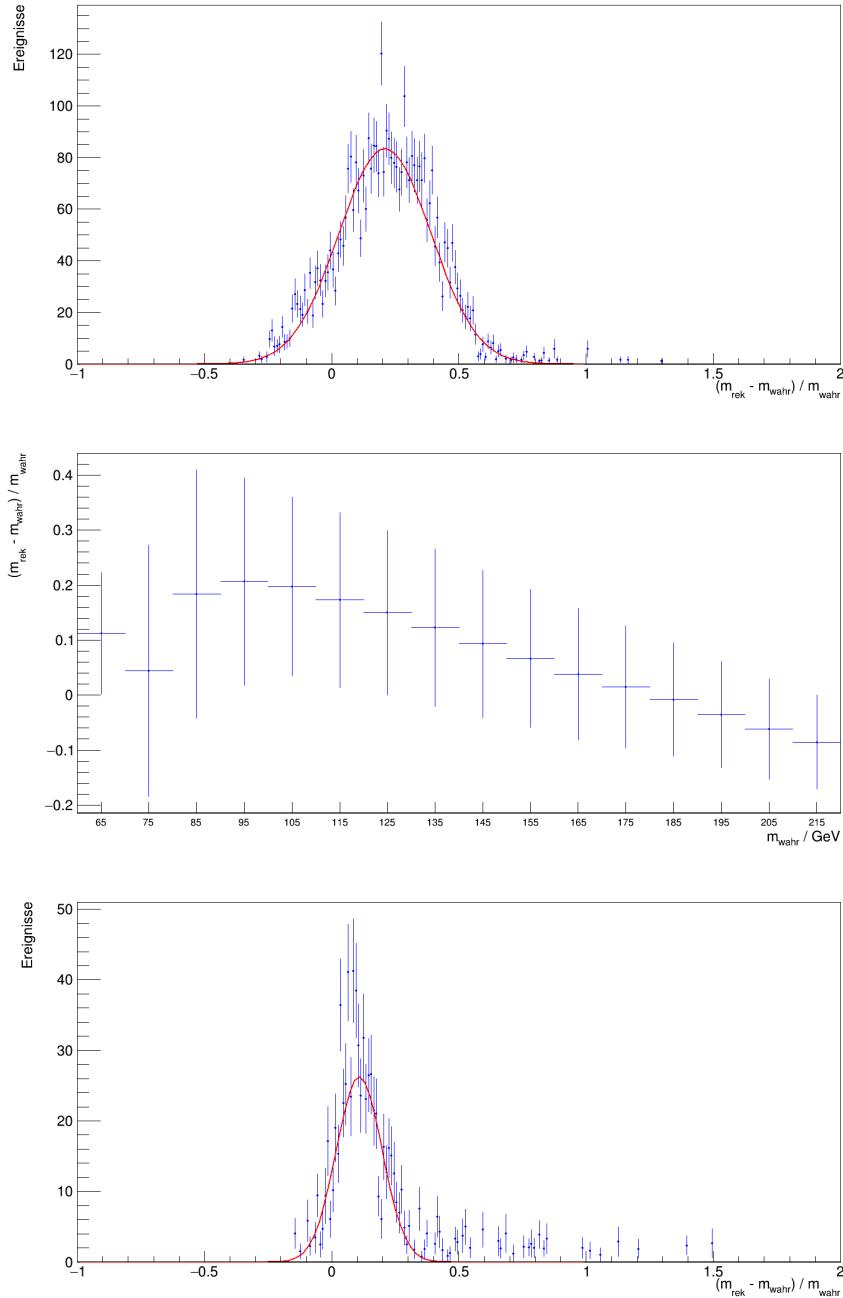


Abbildung 8.6: Relative Massenauflösung für das Massenintervall 90 bis 100 GeV (oben), 60 bis 70 GeV (unten) mit Gaußanpassung und relative Massenauflösung in Intervallen der wahren Masse (mitte), Training Nr. 4 mit Polarisationsasymmetrie

8.4 Kalibrierung

Ein weiterer wichtiger Schritt, um die Qualität der Vorhersage des Netzes zu untersuchen ist die Kalibrierung. Wie man in Abbildung 8.6 mittig gut erkennen kann, ist die Vorhersage für Ereignisse, deren wahre Masse im Bereich 85 bis 145 GeV liegt, größer als die wahre Masse. Für Ereignisse, deren wahre Masse im Bereich von 85 bis 215 GeV liegt, ist die relative Massenauflösung näher bei null, die Vorhersage also besser. Es kann an dieser Stelle eine Kalibrierung vorgenommen werden, um Über- bzw. Unterschätzungen des Netzes teilweise zu eliminieren. Diese Vorgehensweise wird im Folgenden erläutert. Es werden erneut Beispiele des vierten Trainingsdurchlaufes mit Polarisationsasymmetrie herangezogen. Die Vorhersagen werden mit allen Ereignissen aus dem Validationsteil des Trainingsdatensatzes erzeugt. Die beschriebene Kalibrierung besteht aus zwei Schritten.

In Abbildung 8.7 links ist der Median der wahren Masse m_{wahr} in Intervallen der rekonstruierten Masse m_{rek} dargestellt. Hierbei wird der Median genutzt, da die Verteilungen der wahren Masse in Intervallen eine asymmetrische Form wie in Abbildung 8.7 rechts dargestellt, haben. Der Median spiegelt einen Ausgleich zwischen Mittelwert und Position der Spitze wieder.

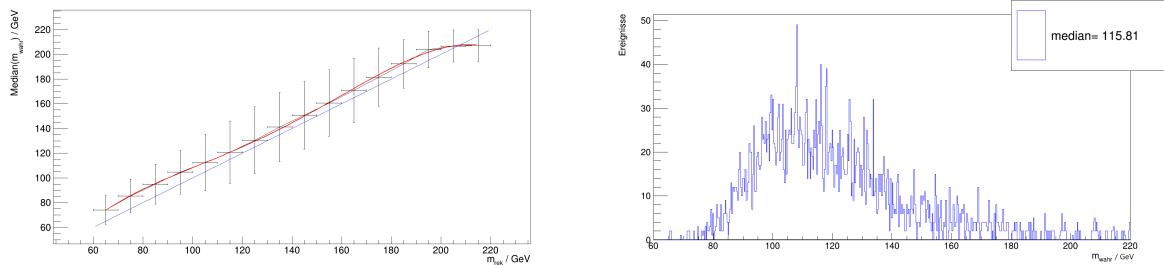


Abbildung 8.7: Median(m_{wahr}) in Intervallen der rekonstruierten Masse m_{rek} aufgetragen (links) und exemplarische Massenverteilung der wahren Masse im rekonstruierten Massenintervall 110 bis 120 GeV (rechts), Training Nr. 4 mit Polarisationsasymmetrie

Optimalerweise wäre in Abbildung 8.7 links eine Winkelhalbierende zu sehen, wie sie in blau dargestellt ist. Allerdings ist das an die Messwerte angepasste Polynom vierten Grades (in Abbildung 8.7 links rot) keine Winkelhalbierende. Das Polynom hat die Form:

$$m_{\text{wahr, polynom}} = a \cdot m_{\text{rek}}^4 + b \cdot m_{\text{rek}}^3 + c \cdot m_{\text{rek}}^2 + d \cdot m_{\text{rek}} + e$$

Um das Polynom möglichst nah an die Winkelhalbierende zu bekommen, wird der rekonstruierte Wert wie folgt korrigiert:

$$m_{\text{korr}, 1} = m_{\text{rek}} + \Delta m \quad (8.1)$$

Hierbei steht $m_{\text{korr}, 1}$ für den Wert der rekonstruierten Masse nach der ersten Kalibrierung. Es ist dann:

$$\Delta m = m_{\text{wahr, polynom}} - m_{\text{rek}}$$

Diese Korrektur wird für jedes betrachtete Intervall der rekonstruierten Masse durchgeführt. Wird die rekonstruierte Masse korrigiert, so kann der Median der wahren Masse in Intervallen der korrigierten Masse dargestellt werden, um zu prüfen, dass der Kalibrierungsschritt funktioniert hat, siehe Abbildung 8.8 links. Die resultierende Anpassungsgerade entspricht der Winkelhalbierenden, der Kalibrierungsschritt

war erfolgreich. Wird allerdings die korrigierte Masse in Intervallen der wahren Masse dargestellt, so erhält man Abbildung 8.8 rechts.

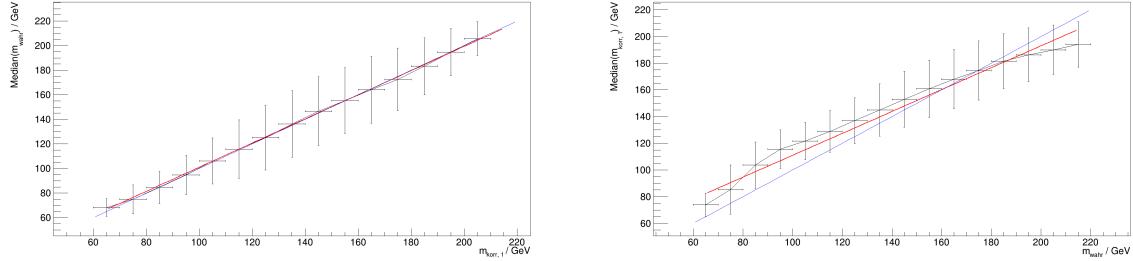


Abbildung 8.8: Median(m_{wahr}) in Intervallen der korrigierten Masse $m_{\text{korr}, 1}$ (links) und Median($m_{\text{korr}, 1}$) in Intervallen der wahren Masse m_{wahr} (rechts), Training Nr. 4 mit Polarisationsasymmetrie

Die Anpassungsgerade entspricht nicht der Winkelhalbierenden. Das liegt daran, dass nicht lediglich x- und y-Achse vertauscht wurden; auf der y-Achse ist der Median von Verteilungen dargestellt. Der Wert des Medians hängt davon ab, wie die Form der Massenverteilung im jeweiligen Massenintervall aussieht. Um auch die korrigierte Masse in Intervallen der wahren Masse in die Form einer Winkelhalbierenden zu bekommen wird ein zweiter Kalibrierungsschritt vorgenommen. Die an die Messdaten in Abbildung 8.8 rechts angepasste Gerade hat die Form:

$$m_{\text{korr}, 1} = p \cdot m_{\text{wahr, gerade}} + t$$

Um eine Steigung von 1 zu erreichen formt man um:

$$m_{\text{wahr, gerade}} = \frac{m_{\text{korr}, 1} - t}{p}$$

Mit diesen Werten kann eine zweite Kalibration analog zur ersten (Gleichung 8.1) durchgeführt werden. Mit diesem zweiten Kalibrationsschritt sieht der Median der korrigierten Masse in Intervallen der wahren Masse aus wie in Abbildung 8.9 oben dargestellt. Die Anpassungsgerade entspricht der Winkelhalbierenden. An diesem Punkt kann die relative Massenauflösung in Intervallen der wahren Masse in Abbildung 8.9 unten nach den beiden Kalibrationsschritten dargestellt werden.

Es ist zwar keine horizontale Linie bei Null zu sehen, allerdings ist die relative Auflösung nach der Kalibrierung deutlich besser als vor der Kalibrierung (Abbildung 8.6 mittig). Diese Art der Darstellung wird bei der Besprechung der Ergebnisse, die mit der Methode der Distance Correlation entstanden sind, relevant.

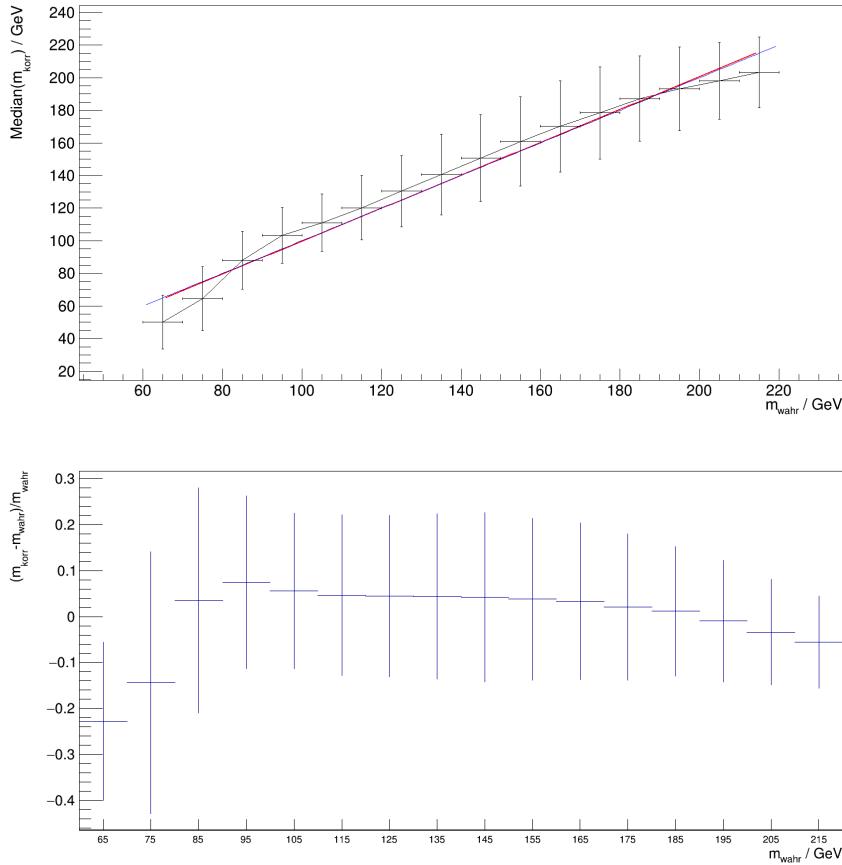


Abbildung 8.9: Median(m_{wahr}) in Intervallen der korrigierten Masse $m_{\text{korr},2}$ (oben) und relative Massenauflösung in Intervallen der wahren Masse (unten), nach zweiter Kalibrierung, Training Nr. 4 mit Polarisationsasymmetrie

8.5 Distance Correlation

In diesem Abschnitt werden die relativen Massenauflösungen und anschließenden kalibrierten relativen Massenauflösungen in Intervallen der wahren Masse besprochen, die mit der Methode der Distance Correlation gefunden wurden. Diese werden mit den Netzen ohne die Methode der Distance Correlation verglichen. Im Vorhinein kann schon gesagt werden, dass es sich lohnt, weitere Weitere für λ zu testen. Mit der implementierten Distance-Correlation Funktion in der Lossfunktion wurden drei Werte für den Faktor λ getestet; $\lambda = 0$, $\lambda = 1\,000$ und $\lambda = 10\,000$. Der Wert $\lambda = 0$ entspricht einem Training, in dem keine Dekorrelation stattfindet. Der Wert $\lambda = 1\,000$ wird gewählt, da sich der Losswert am Anfang des Trainings bei ca. einem Wert von 1 400 befindet. Somit würde die Dekorrelation mit einem Faktor von ungefähr $\frac{1000}{2400}$ ins Training eingehen, sollte der Betrag der Distance-Correlation Funktion in etwa 1 sein (wofür es allerdings keinerlei Indizien gibt; die Wahl des Wertes dient in erster Linie der Orientierung, wie stark der Einfluss der Dekorrelation auf die Vorhersage ist). Der Faktor $\lambda = 10\,000$ ist nach Einschätzung der Dekorrelation durch die Distance-Correlation mit dem Faktor $\lambda = 1\,000$ entstanden, um den Effekt der

Kapitel 8 Ergebnisse

Dekorrelation stärker beobachten zu können. In Abbildung 8.10 sind die relativen Massenauflösungen für den ersten und den dritten Trainingsdurchlauf, ausgewertet auf dem Z-Datensatz für die drei verwendeten Werte von λ dargestellt. Dass für alle Trainings mit unterschiedlichen Werten für λ das erste und dritte Training klar sichtbare Unterschiede aufweist, ist Zufall. Die Ergebnisse zweier Trainingsdurchläufe sind an dieser Stelle hinreichend, um den gewünschten Effekt zu sehen. Die relativen Auflösungen für die weiteren Trainingsdurchläufe sind im Anhang in Abbildung A.6 dargestellt.

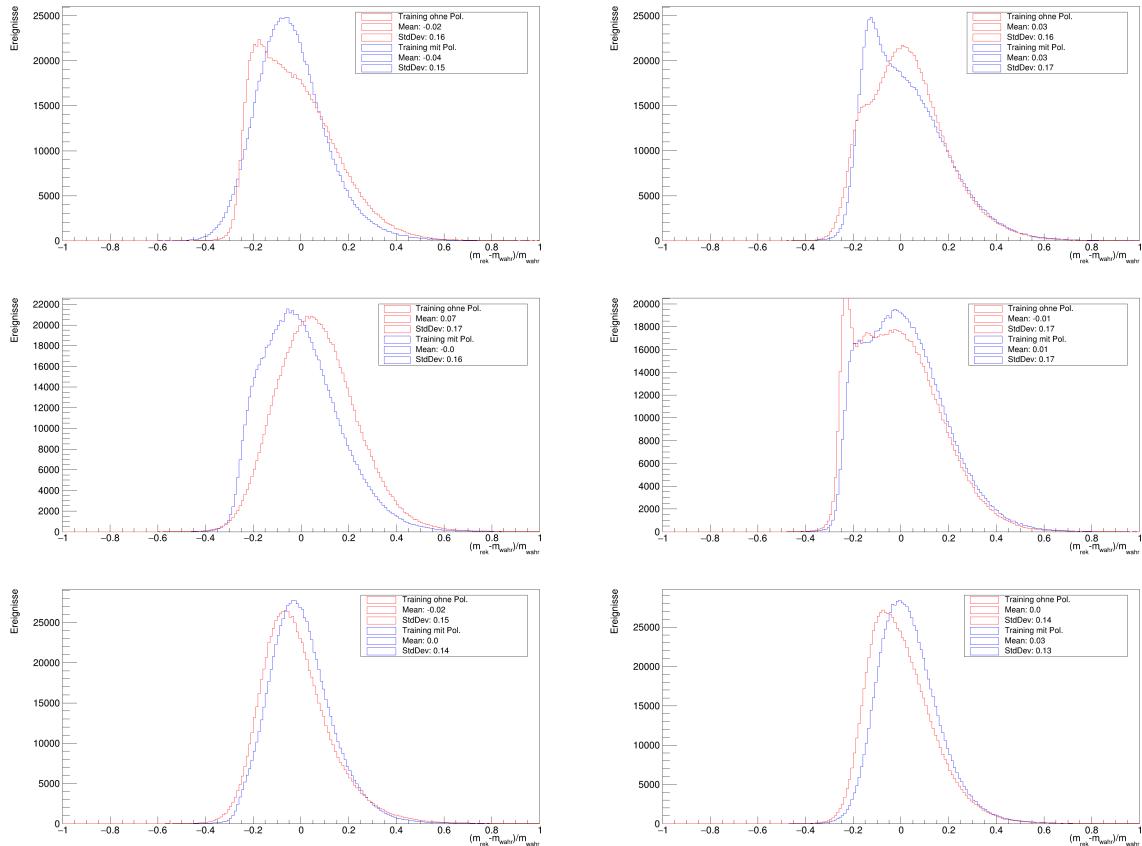


Abbildung 8.10: Relative Massenauflösung, DisCo-Training, $\lambda = 0$ (oben), $\lambda = 1000$ (mitte) und $\lambda = 10000$ (unten), jeweils Training Nr. 1 (links) und 3 (rechts) mit und ohne Polarisationsasymmetrie, der Effekt der Dekorrelation ist daran zu erkennen, dass sich die Kurven des Trainings mit und ohne Polarisationsasymmetrie mit steigendem λ ähnlicher werden.

Allein die Form der Kurven mit und ohne Polarisationsasymmetrie gleicht sich für einen Wert von $\lambda = 10000$ am stärksten.

Die Standardabweichungen werden im Mittel kleiner für steigende Werte von λ ; während die mittlere Standardabweichung $\bar{\sigma}$ der vier Trainingsdurchläufe für einen Wert von $\lambda = 0$ für das Training mit Polarisationsasymmetrie $\bar{\sigma}_{\text{pol.}}$ und das ohne Polarisationsasymmetrie $\bar{\sigma}_{\text{ohne pol.}}$ noch:

$$\bar{\sigma}_{\text{mit Pol.}} = 0,165 \text{ GeV} \quad \text{und} \quad \bar{\sigma}_{\text{ohne pol.}} = 0,168 \text{ GeV}$$

beträgt, sind es für einen Wert von $\lambda = 10\,000$ nur noch:

$$\bar{\sigma}_{\text{mit Pol.}} = 0,133 \text{ GeV} \quad \text{und} \quad \bar{\sigma}_{\text{ohne pol.}} = 0,143 \text{ GeV}$$

Für die Mittelwerte der Verteilungen kann keine Verbesserung festgestellt werden. Da sich die Kurvenform des Trainings mit und ohne Polarisationsasymmetrie mit steigendem Wert des Faktors λ angleichen, liegt nahe, dass die zusätzliche Distance-Correlation Funktion in der Lossfunktion dafür sorgt, dass die Massenvorhersage weniger stark abhängig von der Polarisationsasymmetrie ist, Massenvorhersage und Polarisationsasymmetrie der Tau-Leptonen also dekorreliert werden.

Da die Dekorrelation eine Korrelation voraussetzt, wird an dieser Stelle erneut die Hypothese bestärkt, dass eine Korrelation grundlegend existiert. Allerdings wurde lediglich erwartet, dass die Abweichungen zwischen den Kurven der relativen Massenauflösung für das Training mit und ohne Polarisation kleiner werden. Die beobachtete bessere Auflösung ist kein erwarteter Effekt. Potentielle Ursache für diesen Effekt ist, dass das Netz allgemein präferiert die Z-Masse vorhersagt, und so die Standardabweichung mit steigendem λ kleiner wird. Das würde die praktischen Anwendungsmöglichkeiten des Netzes stark einschränken, da die Vorhersage der Masse des Higgs-Bosons sehr schlecht wäre.

Im Folgenden werden die relativen Massenauflösungen in Intervallen der wahren Masse nach der Kalibrierung für die Trainingsdurchläufe mit Distance Correlation besprochen. Es wird ab diesem Punkt lediglich das Training mit Polarisationsasymmetrie untersucht. In Abbildung 8.11 sind diese für die drei verwendeten Werte für λ dargestellt.

Es sind jene Darstellungen gewählt worden, die die beste relative Auflösung im Bereich der Z-Masse aufweisen. Die restlichen Darstellungen sind im Anhang in Abbildung A.7 dargestellt. Der Kalibrierungsvorgang für das zweite Training des Netzes mit und ohne Polarisation für den Faktor $\lambda = 10\,000$ hat allerdings nicht funktioniert und wird daher im Anhang nicht dargestellt. Wie in den Abbildungen zu erkennen, nimmt die Qualität der Vorhersage im Bereich der Z-Masse mit größerem Wert für λ nicht ab, sondern wird, wie in Abbildung 8.11 zu sehen, geringfügig besser. Die Anzahl an Trainingsdurchläufen ist allerdings nicht hoch genug, um an dieser Stelle eine Verbindung zwischen der besseren Auflösung und dem steigenden Wert von λ zu ziehen. Aus der Darstellung kann die Hypothese aufgestellt werden, dass durch den Vorgang der Dekorrelation keine relative Auflösung eingebüßt wird. Diese Hypothese birgt Hoffnung, dass die Methode der Distance-Correlation in weiteren Studien untersucht werden, und schließlich dafür verwendet werden kann, die Spininformation des Mutterteilchens von der Massenvorhersage zu dekorrelieren.

In Abbildung 8.11 ist der Verlauf der Auflösungskurve für $\lambda = 10\,000$ für größere wahre Massen (ab dem Massenintervall mit dem Mittelwert 165 GeV bis ca. 195 GeV) deutlich schlechter als die Auflösungen für die weiteren Werte für λ . Auch für wahre Massenintervalle, deren Mittelwert 105 GeV bis einschließlich 125 GeV ist, weicht die Vorhersage mit dem Wert $\lambda = 10\,000$ stark von den Auflösungen der weiteren Werte ab, ist allerdings nicht schlechter. Dieser Effekt wird für weitergehende Studien von Relevanz sein können.

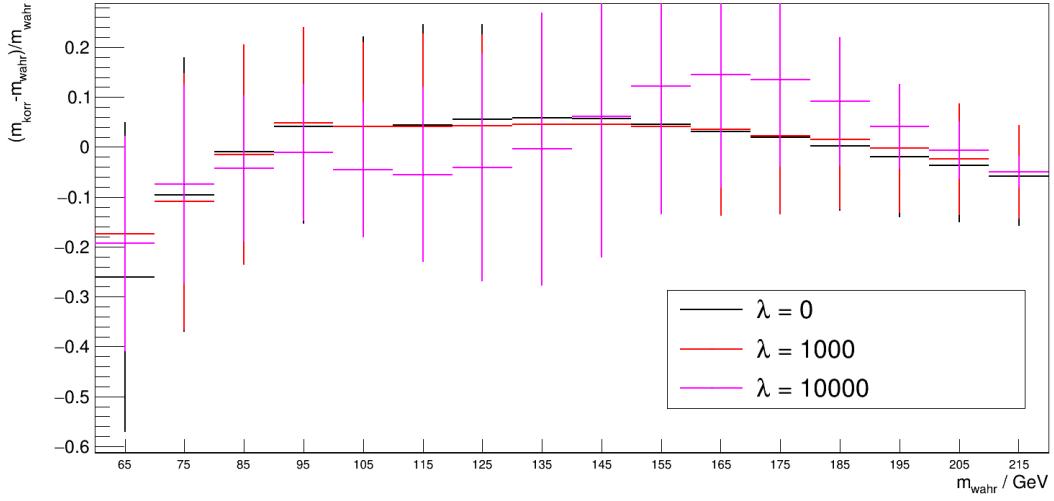


Abbildung 8.11: Relative Massenauflösung in Intervallen der wahren Masse, $\lambda = 0$, Training Nr. 1, $\lambda = 1000$, Training Nr. 1 und $\lambda = 10000$, Training Nr. 3; alle Trainingsdurchläufe mit Polarisationsasymmetrie. Die Auflösung im Bereich der Z-Masse nimmt für steigende Werte von λ nicht ab

8.6 Weiterführende Studien

Sobald ein Trainingsdatensatz vorliegt, dessen Zerfall ein Mutterteilchen mit Spin null zu Grunde liegt, können die Vorhersagen unterschiedlicher Trainings mit unterschiedlichen Trainingsdatensätzen miteinander verglichen und dekorreliert werden. In weiteren Studien kann außerdem das adversarial neural network (siehe Abschnitt 7.4.1) verwendet werden. Bei einer solchen Architektur ist es zum Beispiel möglich, dem adversarial Netz weitere Informationen über die Vorhersage des Regressionsnetzes hinaus zu geben. Das können auch wahre Werte sein, die für die praktische Vorhersage nicht zur Verfügung stehen, da lediglich das Regressionsnetz genutzt wird, um Vorhersagen zu formulieren. Eine weitere Möglichkeit ist, dem adversarial Netz die geladenen und neutralen Zerfallsprodukte des Tau-Leptons einzeln zu geben, während man dem Regressionsnetz diese addiert als Eingabe zur Verfügung stellt. Somit stehen dem Teil des Netzes, der für die Dekorrelation zuständig ist (das adversarial Netz) Daten zur Verfügung, die mehr Informationen über die Polarisationsasymmetrie der Tau-Leptonen enthalten. Es kann weiterhin ein Netz getestet werden, das die Helizität explizit vorherzusagen. Diese Art von Netz hat zwei Ausgangsneuronen, wobei ein Neuron die Vorhersage für die Masse, und ein weiteres die Vorhersage für die Helizität ausgibt.

Zur Weiterführung der in dieser Bachelorarbeit angestellten Untersuchungen kann die Auswirkung unterschiedlicher Eingabeveriablen, wie der Zerfallsmodus oder die geladenen und neutralen Zerfallsprodukte der Tau-Leptonen einzeln auf die Vorhersage des Netzes untersucht werden. Würde ein Datensatz vorliegen, dessen wahre Massenverteilung nach unten nicht auf 60 GeV beschränkt ist, so könnte der in Abschnitt 8.3 beobachtete Effekt in Abbildung 8.6 mittig weiter untersucht werden.

KAPITEL 9

Fazit

In dieser Bachelorarbeit wurde die Methode der Distance-Correlation angewandt, um die Di-Tau Massenvorhersage eines neuronalen Netzes von der Polarisationsasymmetrie der Tau-Leptonen zu dekorrelieren. Die Motivation dieser Anwendung findet sich in den aus den Ergebnissen aus [25] und [8] formulierten Hypothesen, dass Massenvorhersage und Polarisationsasymmetrie (sowie Massenvorhersage und Spin) korreliert sind. Das langfristige Ziel ist nicht, die Massenvorhersage des Netzes unabhängig von der Polarisationsasymmetrie zu machen. Das Ziel ist es langfristig, die Massenvorhersage unabhängig vom Spin des Mutterteilchens formulieren zu können, da das grundlegende Ziel ist, die Vorhersage der Higgs-Masse zu verbessern. Eine tiefergehende Motivation des Problems wird in Abschnitt 6 gegeben. In Abschnitt 8.1 konnte die Hypothese der Abhängigkeit der Massenvorhersage von der Polarisationsasymmetrie der Tau-Leptonen bestärkt werden. Allerdings führt das viermalige Ausführen des selben Trainings (mit der selben Anzahl an Epochen) zu unterschiedlichen Vorhersagen und insbesondere unterschiedlicher relativer Massenauflösung. Das Training des neuronalen Netzes ist an dieser Stelle sehr instabil, wie auch die Untersuchung der betragsmäßig aufaddierten Gewichte zwischen den ersten Schichten des Netzes gezeigt hat. Die relative Massenauflösung für den Trainingsdatensatz mit Polarisationsasymmetrie im Bereich der Z-Masse (85 bis 95 GeV) zeigte weniger Unterschiede für die vier verschiedenen Trainingsdurchläufe, als die relative Massenauflösung für den Z-Datensatz. Hierfür können weitere (bis jetzt nicht betrachtete) Unterschiede zwischen Trainings- und Z-Datensatz verantwortlich sein. Auf potentielle Unterschiede zwischen Trainings- und Z-Datensatz wird in Abschnitt 8.2 eingegangen. Allerdings ist die Aussagekraft der Ergebnisse hier stark durch verhältnismäßig wenige Ereignisse im Trainingsdatensatz eingeschränkt. Der Effekt der Dekorrelation wurde in Abschnitt 8.5 mit der Methode der Distance Correlation untersucht. Die Dekorrelation von Polarisationsasymmetrie und Masse macht sich dadurch bemerkbar, dass sich die betrachteten Kurven für die relative Massenauflösung für steigende Werte von λ angleichen. Das Verhalten der Dekorrelation kann an dieser Stelle für weitere Werte von λ untersucht werden. Es konnte weiterhin die Hypothese aufgestellt werden, dass die relative Massenauflösung für Ereignisse im Bereich der Z-Masse mit steigendem Wert von λ nicht abnimmt. Die Distance Correlation kann, aufbauend auf den in dieser Arbeit formulierten Hypothesen, weiter untersucht werden. In Abschnitt 8.6 wurden Anregungen und Ideen für weiterführende Studien, teilweise aufbauend auf den Ergebnissen dieser Bachelorarbeit, geboten.

ANHANG A

Anhang

Eingabeveriablen

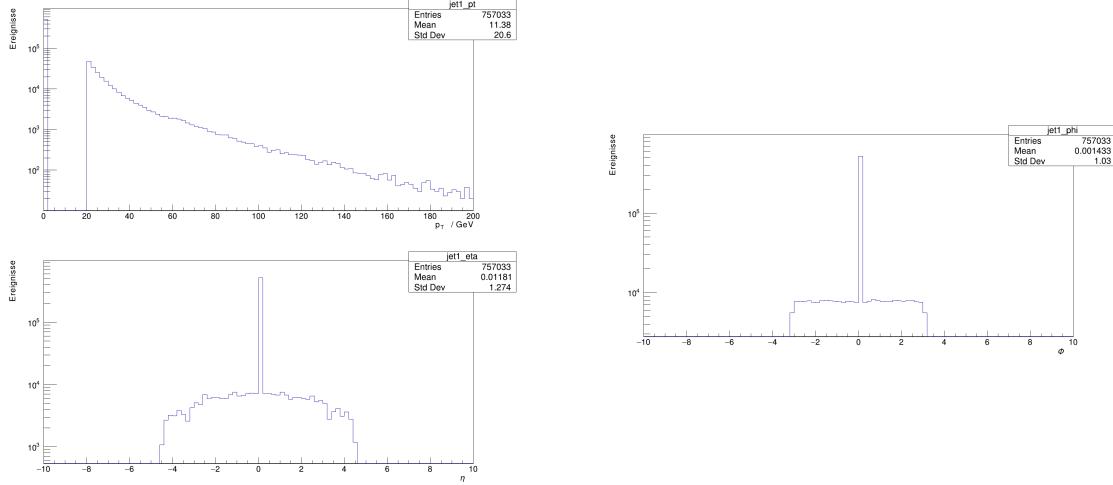


Abbildung A.1: Transversalimpuls p_T , Pseudorapidität η und Azimutwinkel Φ für den Jet mit der zweithöchsten Energie, für alle Ereignisse aus dem Trainingsdatensatz, die die Vorselektion aus Abschnitt 5.3 erfüllen, logarithmisch dargestellt.

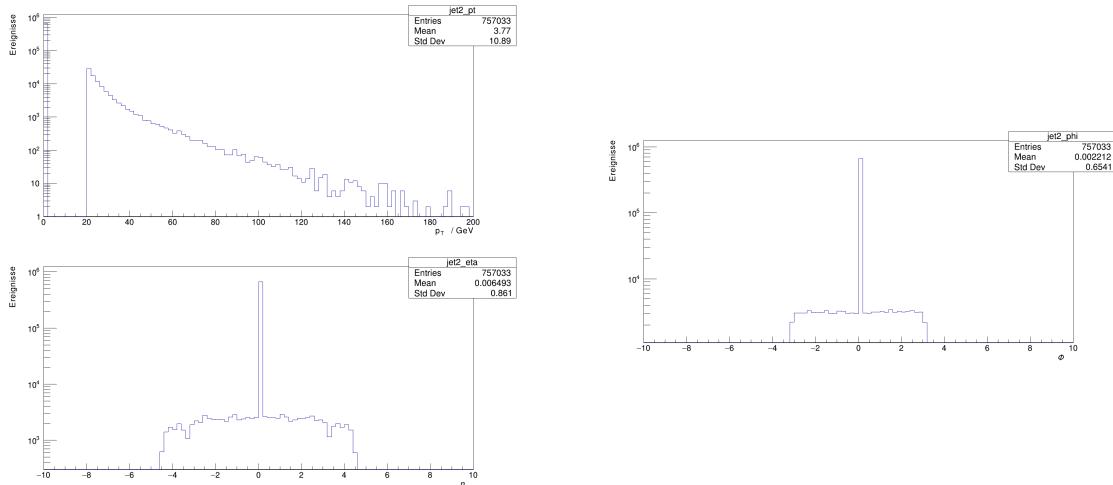


Abbildung A.2: Transversalimpuls p_T , Pseudorapidität η und Azimutwinkel Φ für den Jet mit der niedrigsten Energie, für alle Ereignisse aus dem Trainingsdatensatz, die die Vorselektion aus Abschnitt 5.3 erfüllen, logarithmisch dargestellt.

Relative Massenauflösung

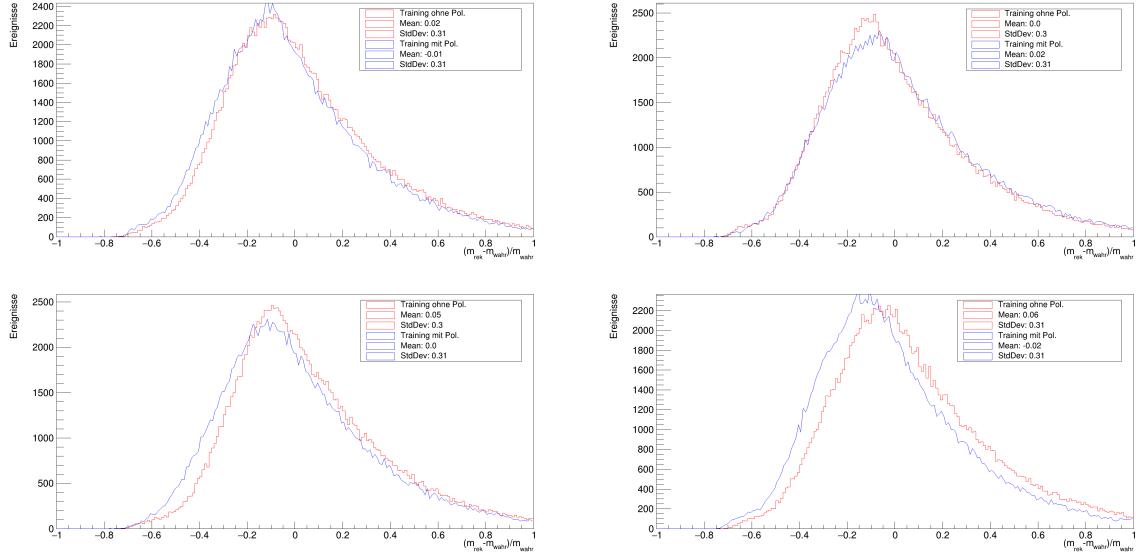


Abbildung A.3: Relative Massenauflösung, bestimmt mit allen Ereignissen aus dem Trainingsdatensatz, Training Nr. 1 (links oben), 2 (rechts oben), 3 (links unten), 4 (rechts unten) jeweils mit und ohne Polarisationsasymmetrie

Gewichte zwischen den ersten beiden Schichten

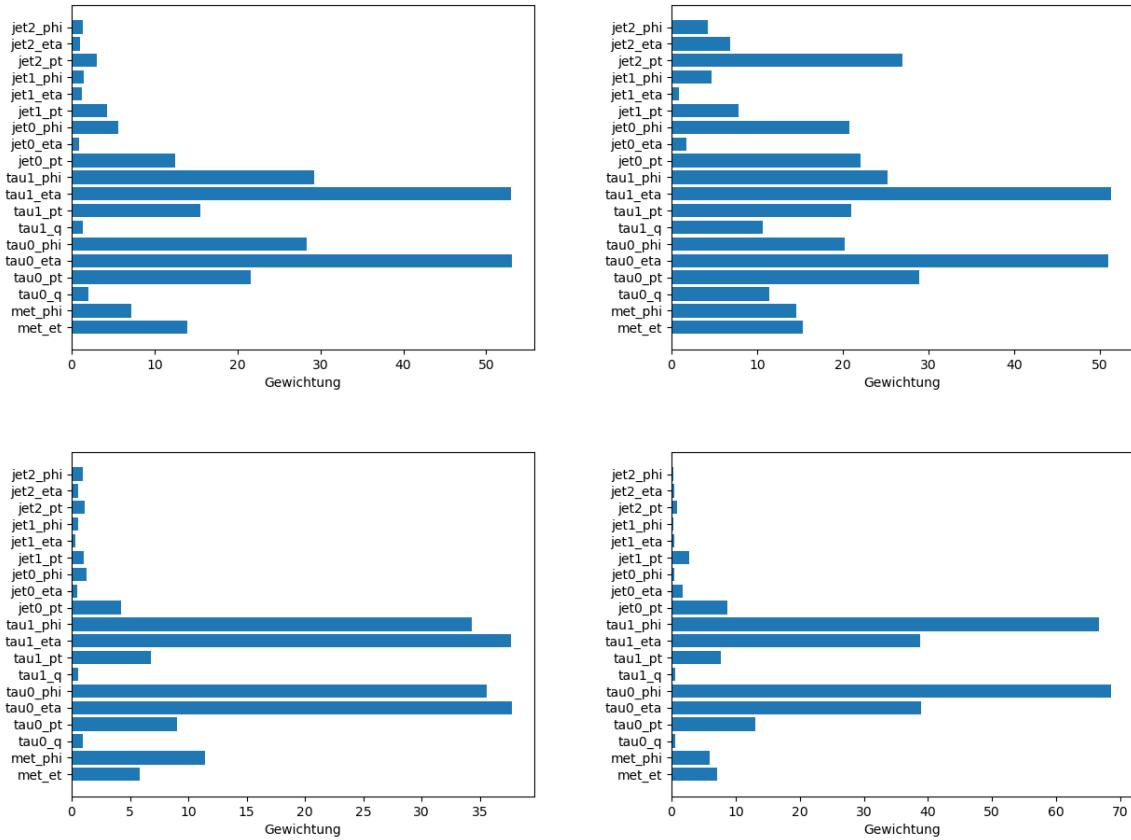


Abbildung A.4: Gewichtungen mit (oben) und ohne (unten) Polarisationsasymmetrie zwischen den ersten beiden Schichten, Training Nr. 3 (links) und 4 (rechts)

Mögliche Ursachen

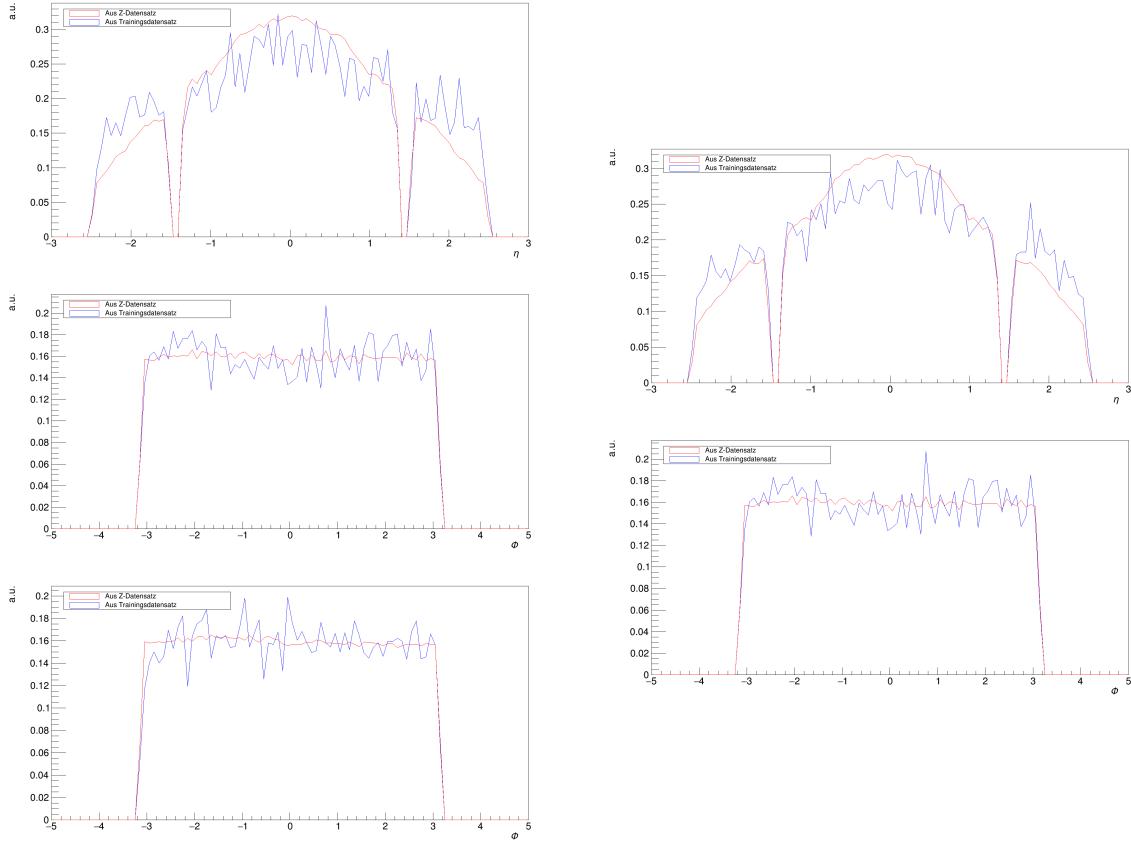


Abbildung A.5: Normierte Verteilung der Pseudorapidität η und des Azimutwinkels des höher- und niederenergetischen Tau-Leptons und des Azimutwinkels Φ der fehlenden Transversalenergie für die Ereignisse aus dem Trainingsdatensatz mit Polarisationsasymmetrie im Bereich der Z-Masse und Ereignisse aus dem Z-Datensatz

Relative Auflösung, DisCo

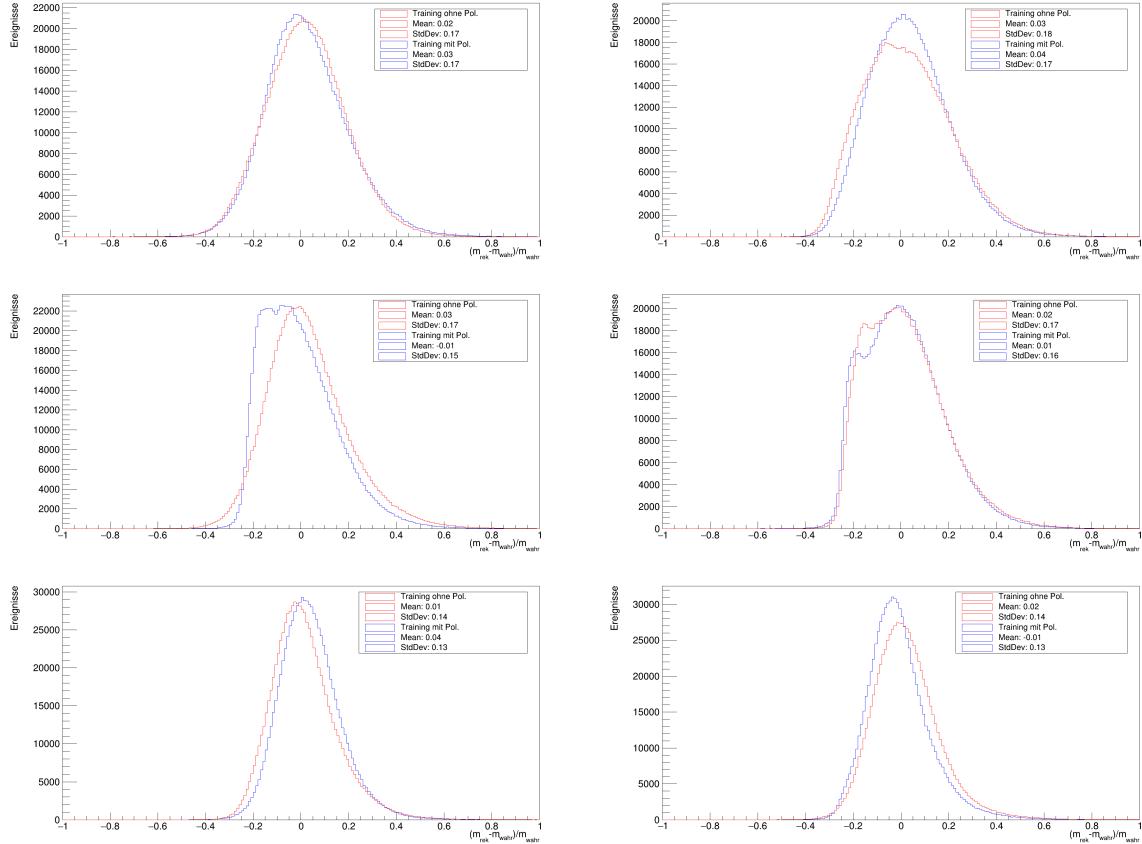


Abbildung A.6: Relative Massenauflösung, DisCo-Training, $\lambda = 0$ (oben), $\lambda = 1\,000$ (mitte) und $\lambda = 10\,000$ (unten), Training Nr. 2 (links) und 4 (rechts)

Relative Massenauflösung in Intervallen der wahren Masse, DisCo

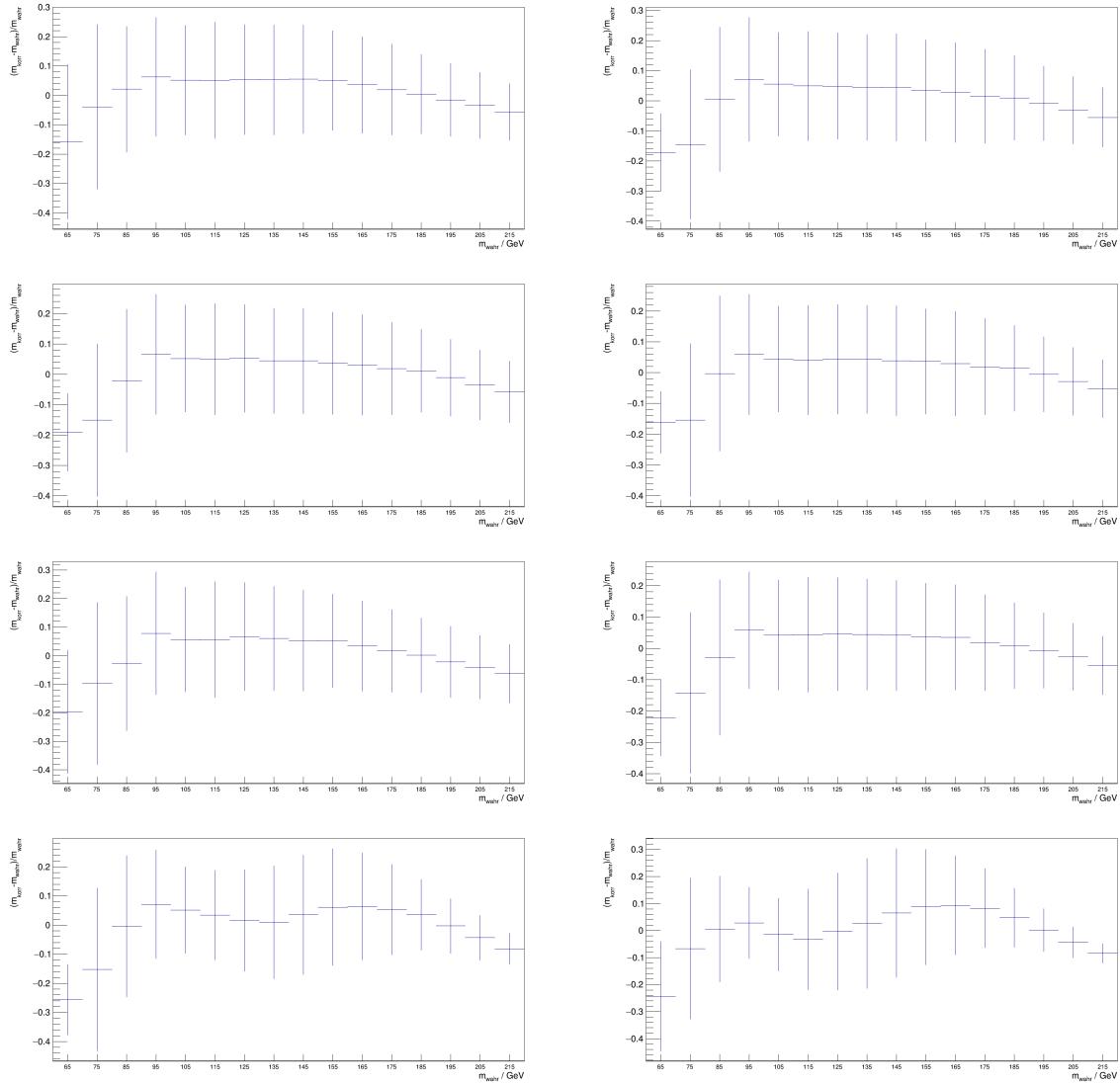


Abbildung A.7: Relative Massenauflösung in Intervallen der wahren Masse, für $\lambda = 0$ und Training Nr. 2, 3 und 4 (links oben, absteigend), für $\lambda = 1000$ und Training Nr. 2, 3 und 4 (rechts oben, absteigend) und für $\lambda = 10000$ (unten) und Training Nr. 1 (links) und 4 (rechts), Training mit Polarisationsasymmetrie

Literatur

- [1] G. Aad u. a., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Physics Letters B **716** (2012) 1, ISSN: 0370-2693, URL: <http://dx.doi.org/10.1016/j.physletb.2012.08.020>.
- [2] M. e. a. Tanabashi, *Status of Higgs Boson Physics*, (2019), URL: <https://pdg.lbl.gov/2019/reviews/rpp2019-rev-higgs-boson.pdf>.
- [3] A. Elagin, P. Murat, A. Pranko und A. Safonov, *A new mass reconstruction technique for resonances decaying to*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **654** (2011) 481, ISSN: 0168-9002, URL: <http://dx.doi.org/10.1016/j.nima.2011.07.009>.
- [4] B. P. et al., *Teilchen und Kerne*, Springer Spektrum, 2014, ISBN: 978-3-642-37821-8.
- [5] W. C. contributors, *Standardmodell der Teilchenphysik*, URL: https://commons.wikimedia.org/w/index.php?title=File:Standard_Model_of_Elementary_Particles-de.svg&oldid=521452056 (besucht am 05.08.2021).
- [6] P. Zyla u. a., *Review of Particle Physics*, PTEP **2020** (2020) 083C01.
- [7] T. Hofmann, *Neutrinooszillationen*, URL: http://www.physik.uni-regensburg.de/forschung/gebhardt/gebhardt_files/skripten/Neutrinooszillationen.pdf (besucht am 23.08.2021).
- [8] L. Herrmann, *Optimization of a Regression-Based Deep Neural Network for Di-Tau Mass Reconstruction in ATLAS*, Universität Bonn, 2020, URL: <https://web.physik.uni-bonn.de/group/view.php?&group=1&lang=de&c=t&id=121>.
- [9] K. Olive, *Review of Particle Physics*, Chinese Physics C **38** (2014) 090001, URL: <https://doi.org/10.1088/1674-1137/38/9/090001>.
- [10] J. D. Anderson, M. H. Austern und R. N. Cahn, *Tau polarization asymmetry as a probe of Z' couplings*, Phys. Rev. Lett. **69** (1 1992) 25, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.69.25>.
- [11] CERN, *LHC the guide*, URL: https://home.cern/sites/home.web.cern.ch/files/2018-07/CERN-Brochure-2017-002-Eng_0.pdf (besucht am 24.09.2021).
- [12] C. Bulletin, *LHC Illustration*, URL: <http://cds.cern.ch/journal/CERNBulletin/2008/38/News%20Articles/1125888> (besucht am 11.08.2021).

- [13] lhc-facts.ch, *LHC-ATLAS Detektor*,
URL: <http://www.lhc-facts.ch/index.php?page=atlas> (besucht am 11.08.2021).
- [14] I. Masterclass, *ATLAS Illustration*, URL:
<https://atlas.physicsmasterclasses.org/de/atlas.htm> (besucht am 11.08.2021).
- [15] CERN, *The Inner Detector*, URL:
<https://atlas.cern/discover/detector/inner-detector> (besucht am 24.09.2021).
- [16] CERN, *Calorimeter*,
URL: <https://atlas.cern/discover/detector/calorimeter> (besucht am 24.09.2021).
- [17] CERN, *Muon Spectrometer*,
URL: <https://atlas.cern/discover/detector/muon-spectrometer> (besucht am 24.09.2021).
- [18] E. Nickel, *Messung der Top-Quark-Masse mittels der Lepton-Transversalimpuls-Methode mit einem Datensatz von 5 fb^{-1} am ATLAS-Detektor*, BONN-IB-2020-11, Universität Bonn, 2012,
URL: <https://www.brock.physik.uni-bonn.de/files/BONN-IB-2014-04.pdf>.
- [19] U. Dresden, *Der ATLAS-Versuch*,
URL: https://tu-dresden.de/mn/physik/iktp/ressourcen/dateien/studium/praktikum/Versuchsbeschreibung_ATLAS_e214.pdf?lang=de (besucht am 11.08.2021).
- [20] M. Aaboud u. a., *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at*

$$\sqrt{s} = 13 \text{ TeV}$$

- s=13TeV*, The European Physical Journal C **78** (2018), ISSN: 1434-6052,
URL: <http://dx.doi.org/10.1140/epjc/s10052-018-6288-9>.
- [21] M. D. A. G. Avoxa, *Künstliche Intelligenz schlaegt Hautärzte*,
URL: <https://www.pharmazeutische-zeitung.de/kuenstliche-intelligenz-schlaegt-hautaerzte/> (besucht am 25.08.2021).
- [22] S. Loukas, *What is Machine Learning?*,
URL: <https://towardsdatascience.com/what-is-machine-learning-a-short-note-on-supervised-unsupervised-semi-supervised-and-aed1573ae9bb> (besucht am 11.08.2021).
- [23] T. Gupta, *Feedforward Neural Network*, URL: <https://towardsdatascience.com/deep-learning-feedforward-neural-network-26a6705dbdc7> (besucht am 24.08.2021).
- [24] W. C. contributors, *Neural Network*,
URL: https://commons.wikimedia.org/w/index.php?title=File:Neural_network.svg&oldid=571062292 (besucht am 12.08.2021).
- [25] M. Werres, *Estimating the Mass of Di-Tau Systems in the ATLAS Experiment Using Neural Network Regression*, Universität Bonn, 2019, URL: <https://web.physik.uni-bonn.de/group/view.php?group=1&lang=de&c=t&id=102>.

- [26] Peltarion, *Mean squared error*, URL: <https://peltarion.com/knowledge-center/documentation/modeling-view/build-an-ai-model/loss-functions/mean-squared-error> (besucht am 24.08.2021).
- [27] S. Kostadinov, *Understanding Backpropagation Algorithm*, URL: <https://towardsdatascience.com/understanding-backpropagation-algorithm-7bb3aa2f95fd> (besucht am 12.08.2021).
- [28] S. Doshi, *Various Optimization Algorithms For Training Neural Network*, URL: <https://towardsdatascience.com/optimizers-for-training-neural-network-59450d71caf6> (besucht am 10.09.2021).
- [29] J. Brownlee, *Gentle Introduction to the Adam Optimization Algorithm for Deep Learning*, URL: <https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/> (besucht am 28.08.2021).
- [30] J. Brownlee, *Use Early Stopping to Halt the Training of Neural Networks At the Right Time*, URL: <https://machinelearningmastery.com/how-to-stop-training-deep-neural-networks-at-the-right-time-using-early-stopping/> (besucht am 24.09.2021).
- [31] Tensorflow, *Tensorflow*, URL: <https://www.tensorflow.org/> (besucht am 24.08.2021).
- [32] Keras, *Keras*, URL: <https://keras.io/> (besucht am 24.08.2021).
- [33] T. Gleisberg u. a., *Event generation with SHERPA 1.1*, Journal of High Energy Physics **2009** (2009) 007, ISSN: 1029-8479, URL: <http://dx.doi.org/10.1088/1126-6708/2009/02/007>.
- [34] T. Sjöstrand, S. Mrenna und P. Skands, *A brief introduction to PYTHIA 8.1*, Computer Physics Communications **178** (2008) 852, ISSN: 0010-4655, URL: <http://dx.doi.org/10.1016/j.cpc.2008.01.036>.
- [35] S. Frixione, P. Nason und C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, Journal of High Energy Physics **2007** (2007) 070, ISSN: 1029-8479, URL: <http://dx.doi.org/10.1088/1126-6708/2007/11/070>.
- [36] S. A. et al., *Geant4—a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506** (2003) 250, ISSN: 0168-9002, URL: <https://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- [37] T. Przedzinski, E. Richter-Was und Z. Was, *Documentation of TauSpinner algorithms: program for simulating spin effects in τ -lepton production at LHC*, The European Physical Journal C **79** (2019), ISSN: 1434-6052, URL: <http://dx.doi.org/10.1140/epjc/s10052-018-6527-0>.
- [38] *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*, Techn. Ber., All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2019-033>: CERN, 2019, URL: <https://cds.cern.ch/record/2688062>.
- [39] CERN, *PdgID*, URL: <https://twiki.cern.ch/twiki/bin/view/Main/PdgId> (besucht am 26.09.2021).

- [40] L. Schildgen,
Measurement of the τ Lepton Polarisation in $Z \rightarrow \tau_{lep}\tau_{had}$ Decays with the ATLAS Detector,
Universität Bonn, 2015, URL:
<https://web.physik.uni-bonn.de/group/view.php?&group=1&lang=de&c=t&id=66>.
- [41] J. Brownlee, *How to Use StandardScaler and MinMaxScaler Transforms in Python*,
URL: <https://machinelearningmastery.com/standardscaler-and-minmaxscaler-transforms-in-python/> (besucht am 25.09.2021).
- [42] J. Brownlee, *A Gentle Introduction to the Rectified Linear Unit (ReLU)*,
URL: <https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/> (besucht am 28.08.2021).
- [43] G. Louppe, M. Kagan und K. Cranmer, *Learning to Pivot with Adversarial Networks*, 2017,
arXiv: [1611.01046 \[stat.ML\]](https://arxiv.org/abs/1611.01046).
- [44] G. Kasieczka und D. Shih, *Robust Jet Classifiers through Distance Correlation*,
Physical Review Letters **125** (2020), ISSN: 1079-7114,
URL: <http://dx.doi.org/10.1103/PhysRevLett.125.122001>.

Abbildungsverzeichnis

2.1	Standardmodell der Teilchenphysik [5]	4
2.2	Verzweigungsverhältnisse für den Zerfall des Higgs-Bosons [8]	5
3.1	Der LHC [12]	10
3.2	Aufbau des ATLAS-Detektors [14]	11
3.3	Das im ATLAS-Detektor verwendete Koordinatensystem [8]	12
4.1	Feed-Forward Neuronales Netz, die Information wird in die Eingabeschicht (grün) gegeben, wird in der verdeckten Schicht (blau) weitergeleitet und die entsprechende Vorhersage wird in der Ausgabeschicht (gelb) ausgegeben. [24]	16
5.1	Massenverteilung des Trainingsdatensatzes mit (oben) und ohne Polarisationsasymmetrie (unten)	22
6.1	Vorhergesagte Massenverteilung für Higgs- und Z-Masse, die vorhergesagten Massen weichen stark von den wahren Massen ab [8]	25
6.2	Vorhergesagte Massenverteilung für Ereignisse aus dem Trainingsdatensatz, deren wahre Masse im Bereich der Z- (links) und Higgs-Masse (rechts) liegt. Die Vorhersage ist deutlich besser, als mit den Daten aus Z- und Higgs-Datensatz. [8]	25
6.3	Flugrichtungen und Bewegungsausrichtungen für den Zerfall $\tau^- \rightarrow \pi^- \nu_\tau$ für den Fall eines (a) linkshändigen Tau-Leptons, (b) rechtshändigen Tau-Leptons. Die schmalen Pfeile geben die Bewegungsrichtung, die breiten Pfeile die Spinausrichtung an. [40]	27
6.4	Flugrichtungen und Bewegungsausrichtungen für den Zerfall $\tau^- \rightarrow \rho^- \nu_\tau$ für den Fall eines (a) linkshändigen Tau-Leptons, (b) rechtshändigen Tau-Leptons. L und T signalisieren die longitudinale, bzw. transversale Polarisation des ρ -Mesons. Die schmalen Pfeile geben die Bewegungsrichtung, die breiten Pfeile die Spinausrichtung an. [40]	28
6.5	Verschiedene Flugrichtungen und Bewegungsausrichtungen für den Zerfall $\rho^- \rightarrow \pi^-\pi^0$ für den Fall eines (a) transversal und (b) longitudinal polarisierten ρ -Mesons. Die Energie wird in den beiden Fällen unterschiedlich auf die Pionen verteilt. Die schmalen Pfeile geben die Bewegungsrichtung, die breiten Pfeile die Spinausrichtung an. [40]	28
7.1	Ladung des höher- (links) und niederenergetischen (rechts) Tau-Leptons für alle Ereignisse aus dem Trainingsdatensatz, die die Vorselektion aus Abschnitt 5.3 erfüllen	31
7.2	Transversalimpuls des höher- (links) und niederenergetischen (rechts) Tau-Leptons für alle Ereignisse aus dem Trainingsdatensatz, die die Vorselektion aus Abschnitt 5.3 erfüllen	31

7.3 Pseudorapidität des höher- (links) und niederenergetischen (rechts) Tau-Leptons für alle Ereignisse aus dem Trainingsdatensatz, die die Vorselektion aus Abschnitt 5.3 erfüllen	31
7.4 Azimutwinkel des höher- (links) und niederenergetischen (rechts) Tau-Leptons für alle Ereignisse aus dem Trainingsdatensatz, die die Vorselektion aus Abschnitt 5.3 erfüllen	32
7.5 Betrag (links) und Azimutwinkel (rechts) der fehlenden Transversalenergie für alle Ereignisse aus dem Trainingsdatensatz, die die Vorselektion aus Abschnitt 5.3 erfüllen	32
7.6 Transversalimpuls p_T , Pseudorapidität η und Azimutwinkel Φ für den Jet mit der höchsten Energie, für alle Ereignisse aus dem Trainingsdatensatz, die die Vorselektion aus Abschnitt 5.3 erfüllen, logarithmisch dargestellt. Wird ein Jet nicht rekonstruiert, entspricht das einem Eintrag bei null.	33
7.7 Verlauf der Losskurven für den Trainings- und Validationsteil, Training mit Polarisationsasymmetrie über 100 Epochen, ab einer Epochenzahl von ca. 40 kommt es zu Übertraining	35
7.8 Adversarial neural network, bestehend aus dem Regressionsnetz (links) und dem adversarial Netz (rechts), wobei θ_f die Parameter des Regressions- und θ_r die Parameter des adversarial Netzes sind. Z ist der Störungsparameter, in unserem Fall die Helizität des τ^- . Diese wird vom adversarial Netz vorhergesagt, die Vorhersage wird mit steigender Dekorrelation schlechter. Bild aus [43], bearbeitet	35
 8.1 Massenvorhersage (oben) und relative Massenauflösung (unten) mit dem Trainingsdatensatz, Training Nr. 4 mit und ohne Polarisationsasymmetrie	39
8.2 Relative Massenauflösung, bestimmt mit Daten aus dem Trainingsdatensatz im Bereich der Z-Masse (links), Z-Datensatz (rechts), Training Nr. 1, 2, 3, 4 absteigend, jeweils mit und ohne Polarisationsasymmetrie	40
8.3 Gewichtungen mit (oben) und ohne (unten) Polarisationsasymmetrie zwischen den ersten beiden Schichten, Training Nr. 1 (links) und 2 (rechts)	43
8.4 Normierte Verteilung des Transversalimpulses p_T (links) des höher- (oben) und niedenergetischen (unten) Tau-Leptons, und des Betrages der fehlenden Transversalenergie $ E_T $ (rechts) für die Ereignisse aus dem Trainingsdatensatz mit Polarisationsasymmetrie im Bereich der Z-Masse und Ereignisse aus dem Z-Datensatz, a.u.: arbitrary unit (deutsch: willkürliche Einheit)	44
8.5 Bestätigung der Vorselektion; Transversalimpuls p_T (oben) und Pseudorapidität η (unten) für alle Ereignisse aus Trainings- und Z-Datensatz, für τ_0 (links) und τ_1 (rechts)	44
8.6 Relative Massenauflösung für das Massenintervall 90 bis 100 GeV (oben), 60 bis 70 GeV (unten) mit Gaußanpassung und relative Massenauflösung in Intervallen der wahren Masse (mitte), Training Nr. 4 mit Polarisationsasymmetrie	46
8.7 Median(m_{wahr}) in Intervallen der rekonstruierten Masse m_{rek} aufgetragen (links) und exemplarische Massenverteilung der wahren Masse im rekonstruierten Massenintervall 110 bis 120 GeV (rechts), Training Nr. 4 mit Polarisationsasymmetrie	47
8.8 Median(m_{wahr}) in Intervallen der korrigierten Masse $m_{\text{korr}, 1}$ (links) und Median($m_{\text{korr}, 1}$) in Intervallen der wahren Masse m_{wahr} (rechts), Training Nr. 4 mit Polarisationsasymmetrie	48
8.9 Median(m_{wahr}) in Intervallen der korrigierten Masse $m_{\text{korr}, 2}$ (oben) und relative Massenauflösung in Intervallen der wahren Masse (unten), nach zweiter Kalibrierung, Training Nr. 4 mit Polarisationsasymmetrie	49

Abbildungsverzeichnis

8.10 Relative Massenauflösung, DisCo-Training, $\lambda = 0$ (oben), $\lambda = 1\,000$ (mitte) und $\lambda = 10\,000$ (unten), jeweils Training Nr. 1 (links) und 3 (rechts) mit und ohne Polarisationsasymmetrie, der Effekt der Dekorrelation ist daran zu erkennen, dass sich die Kurven des Trainings mit und ohne Polarisationsasymmetrie mit steigendem λ ähnlicher werden.	50
8.11 Relative Massenauflösung in Intervallen der wahren Masse, $\lambda = 0$, Training Nr. 1, $\lambda = 1\,000$, Training Nr. 1 und $\lambda = 10\,000$, Training Nr. 3; alle Trainingsdurchläufe mit Polarisationsasymmetrie. Die Auflösung im Bereich der Z-Masse nimmt für steigende Werte von λ nicht ab	52
A.1 Transversalimpuls p_T , Pseudorapidität η und Azimutwinkel Φ für den Jet mit der zweithöchsten Energie, für alle Ereignisse aus dem Trainingsdatensatz, die die Vorselektion aus Abschnitt 5.3 erfüllen, logarithmisch dargestellt.	55
A.2 Transversalimpuls p_T , Pseudorapidität η und Azimutwinkel Φ für den Jet mit der niedrigsten Energie, für alle Ereignisse aus dem Trainingsdatensatz, die die Vorselektion aus Abschnitt 5.3 erfüllen, logarithmisch dargestellt.	55
A.3 Relative Massenauflösung, bestimmt mit allen Ereignissen aus dem Trainingsdatensatz, Training Nr. 1 (links oben), 2 (rechts oben), 3 (links unten), 4 (rechts unten) jeweils mit und ohne Polarisationsasymmetrie	56
A.4 Gewichtungen mit (oben) und ohne (unten) Polarisationsasymmetrie zwischen den ersten beiden Schichten, Training Nr. 3 (links) und 4 (rechts)	57
A.5 Normierte Verteilung der Pseudorapidität η und des Azimutwinkels des höher- und niedrigerenergetischen Tau-Leptons und des Azimutwinkels Φ der fehlenden Transversalenergie für die Ereignisse aus dem Trainingsdatensatz mit Polarisationsasymmetrie im Bereich der Z-Masse und Ereignisse aus dem Z-Datensatz	58
A.6 Relative Massenauflösung, DisCo-Training, $\lambda = 0$ (oben), $\lambda = 1\,000$ (mitte) und $\lambda = 10\,000$ (unten), Training Nr. 2 (links) und 4 (rechts)	59
A.7 Relative Massenauflösung in Intervallen der wahren Masse, für $\lambda = 0$ und Training Nr. 2, 3 und 4 (links oben, absteigend), für $\lambda = 1\,000$ und Training Nr. 2, 3 und 4 (rechts oben, absteigend) und für $\lambda = 10\,000$ (unten) und Training Nr. 1 (links) und 4 (rechts), Training mit Polarisationsasymmetrie	60

Tabellenverzeichnis

2.1 Zerfälle des τ -Leptons, Daten aus [9]	6
---	---